

DOI: <http://dx.doi.org/10.12996/gmj.2026.4723>

Artificial Intelligence in Peroral Endoscopic Myotomy: Current Evidence, Limitations, and Future Directions

Peroral Endoskopik Miyotomide Yapay Zeka: Güncel Kanıtlar, Kısıtlılıklar ve Geleceğe Yönelik Yaklaşımlar

© Ahmed Abdallah Salman¹, © Ahmed Abdallah²

¹Department of Internal Medicine, Cairo University, Faculty of Medicine, Cairo, Egypt

²Department of General Surgery, Cairo University, Faculty of Medicine, Cairo, Egypt

ABSTRACT

Peroral endoscopic myotomy (POEM) is an established treatment for achalasia and selected spastic esophageal motility disorders, but several aspects of care remain highly operator-dependent. These include early diagnosis, patient selection, procedural planning, myotomy length, intra-procedural safety, training assessment, and prediction of post-procedure reflux. Artificial intelligence (AI) may help address these limitations by integrating endoscopic video, high-resolution manometry, timed barium esophagram, EndoFLIP, clinical variables, procedural data, and follow-up outcomes into more personalized decision-support tools. This narrative review critically examines the evidence for AI across the POEM pathway and distinguishes direct achalasia or POEM evidence from adjacent and conceptual applications. Early AI applications are most likely to emerge in achalasia detection during diagnostic endoscopy, automated motility interpretation, video-based procedural assessment, and prediction of clinical response or post-POEM reflux. More ambitious applications, such as real-time tunnel orientation, vessel recognition, mucosal injury alerts, and personalized myotomy planning, remain largely investigational. Most AI-specific POEM applications remain at an early or conceptual stage and require multicenter validation, standardized outcomes, and careful clinical governance. AI should support, rather than replace, expert judgment, and an AI-derived reflux estimate cannot replace standard follow-up or objective reflux assessment.

ÖZ

Peroral endoskopik miyotomi (POEM), akalazya ve seçilmiş spastik özofagus motilite bozukluklarının tedavisinde yerleşik bir yöntem olmakla birlikte, tedavi sürecinin çeşitli aşamaları halen büyük ölçüde uygulayıcı deneyimine bağlıdır. Bu aşamalar; erken tanı, hasta seçimi, işlem planlaması, miyotomi uzunluğunun belirlenmesi, işlem sırasında güvenliğin sağlanması, eğitim düzeyinin değerlendirilmesi ve işlem sonrası reflü gelişiminin öngörülmesini kapsamaktadır. Yapay zeka (YZ); endoskopik video görüntüleri, yüksek çözünürlüklü manometri, zamanlı baryumlu özofagus grafisi, EndoFLIP, klinik değişkenler, işleme ilişkin veriler ve takip sonuçlarını bütünleştirerek daha kişiselleştirilmiş karar destek araçlarının geliştirilmesine katkı sağlayabilir. Bu anlatısal derlemede, POEM sürecinin farklı aşamalarında YZ kullanımına ilişkin kanıtlar eleştirel olarak incelenmekte ve doğrudan akalazya veya POEM'e ilişkin kanıtlar, komşu alanlardaki ve kavramsal düzeydeki uygulamalardan ayrılmaktadır. YZ'nin erken dönemdeki uygulamalarının; tanısal endoskopi sırasında akalazyanın saptanması, motilite bulgularının otomatik olarak yorumlanması, video tabanlı işlem değerlendirmesi ve klinik yanıtın veya POEM sonrası reflünün öngörülmesi alanlarında ortaya çıkması beklenmektedir. Tünel yönünün gerçek zamanlı olarak belirlenmesi, damarların tanınması, mukozal yaralanma uyarılarının oluşturulması ve kişiselleştirilmiş miyotomi planlaması gibi daha ileri uygulamalar ise büyük ölçüde araştırma aşamasındadır. POEM'e özgü YZ uygulamalarının çoğu halen erken

Cite this article as: Salman AA, Abdallah A. Artificial intelligence in peroral endoscopic myotomy: current evidence, limitations, and future directions. Gazi Med J. 2026;37(4):596-604

Address for Correspondence/Yazışma Adresi: Ahmed Abdallah Salman, Department of Internal Medicine, Cairo University, Faculty of Medicine, Cairo, Egypt
E-mail / E-posta: Awea84@kasralainy.edu.eg, Awea844@gmail.com
ORCID ID: orcid.org/0009-0006-4244-9377

Received/Geliş Tarihi: 24.05.2026

Accepted/Kabul Tarihi: 16.08.2026

Publication Date/Yayınlanma Tarihi: 15.10.2026



©Copyright 2026 The Author(s). Published by Galenos Publishing House on behalf of Gazi University Faculty of Medicine. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

©Telif Hakkı 2026 Yazar(lar). Gazi Üniversitesi Tıp Fakültesi adına Galenos Yayınevi tarafından yayımlanmaktadır. Creative Commons Atıf-GayriTicari-Türetilemez 4.0 (CC BY-NC-ND) Uluslararası Lisansı ile lisanslanmaktadır.

ABSTRACT

Keywords: Artificial intelligence, peroral endoscopic myotomy, achalasia, computer vision, high-resolution manometry, gastroesophageal reflux disease

Öz

veya kavramsal aşamadır ve çok merkezli doğrulama çalışmalarına, standardize edilmiş sonuç ölçütlerine ve dikkatli bir klinik yönetim yaklaşımına ihtiyaç duymaktadır. YZ, uzman klinik değerlendirmesinin yerine geçmekten ziyade onu desteklemelidir. Ayrıca YZ aracılığıyla elde edilen reflü tahmini, standart takip sürecinin veya objektif reflü değerlendirmesinin yerini alamaz.

Anahtar Sözcükler: Yapay zeka, peroral endoskopik miyotomi, akalazyza, bilgisayarlı görü, yüksek çözünürlüklü manometri, gastroözofageal reflü hastalığı

INTRODUCTION

POEM as a Platform for Precision Endoscopy

Peroral endoscopic myotomy (POEM) has transformed the management of achalasia and selected spastic esophageal motility disorders. Since its introduction as a natural-orifice endoscopic alternative to laparoscopic Heller myotomy, POEM has moved from an experimental third-space endoscopy technique to an established therapeutic option for idiopathic achalasia, type III achalasia, distal esophageal spasm, jackhammer esophagus, and selected cases of failed prior therapy (1,2). Its main appeal lies in its ability to provide a flexible, endoluminal myotomy without external incisions, while allowing the endoscopist to tailor the length and orientation of the myotomy according to the underlying motility pattern.

Contemporary guidelines and reviews support POEM as a safe and effective treatment for achalasia, with outcomes comparable to established alternatives such as laparoscopic Heller myotomy and generally superior efficacy to pneumatic dilation in many clinical settings (2-4). However, POEM remains a highly operator-dependent procedure. Important decisions are still based on endoscopist experience, center practice, and interpretation of heterogeneous pre-procedural data. These include whether POEM is the best treatment for a given patient, how long the esophageal and gastric myotomy should be, whether an anterior or posterior approach is preferable, how to manage sigmoid or advanced achalasia, and how to balance symptom relief against the risk of post-POEM gastroesophageal reflux disease (GERD) (2,4).

This shift creates a strong rationale for artificial intelligence (AI). In achalasia care, clinicians already integrate multiple data streams, including symptoms, endoscopic findings, high-resolution manometry (HRM), timed barium esophagram (TBE), functional lumen imaging probe (FLIP; commonly performed with the EndoFLIP system) measurements, prior interventions, anatomical complexity, and patient-specific reflux risk. However, these variables are rarely combined in a systematic or quantitative way. AI may help convert this fragmented information into clinically useful predictions, supporting earlier diagnosis, improved patient selection, personalized procedural planning, objective intra-procedural guidance, and tailored follow-up (5). In this review, more precise terms such as machine learning, clinical prediction model, deep learning, computer vision, or transformer model are used when the specific method is known.

The earliest AI applications relevant to POEM are likely to emerge before and after the procedure rather than during the procedure

itself. For example, computer vision models may assist in recognizing subtle endoscopic features of achalasia, while clinical prediction models may help estimate the probability of symptom response, recurrence, or post-POEM reflux. A recent proof-of-concept study showed that a model using upper endoscopy videos could detect achalasia with diagnostic accuracy comparable to expert endoscopists, supporting the concept that routine endoscopic video contains clinically useful information that may be under-recognized during standard visual assessment (6).

Nevertheless, AI in POEM remains at an early stage. Most available work relates to achalasia diagnosis, general AI in endoscopy, or outcome prediction rather than real-time AI-guided myotomy. Therefore, the purpose of this review is not to claim that AI is ready to perform or direct POEM, but to critically examine where it may add value across the POEM pathway. The central argument is that POEM is an ideal model for precision therapeutic endoscopy: it is technically complex, data-rich, outcome-driven, and vulnerable to variation in decision-making. AI may eventually help standardize this pathway while preserving the endoscopist's central role in judgment, technique, and patient-centered care.

Methods: Narrative Search Strategy and Evidence Classification

PubMed, MEDLINE, Embase, Scopus, Web of Science Core Collection, and Google Scholar were searched from database inception through April 2026. Search concepts combined "achalasia," "peroral endoscopic myotomy" or "POEM" with "artificial intelligence," "machine learning," "deep learning," "computer vision," "clinical prediction model," "risk score," "nomogram," "technical difficulty," "adverse event," "complication," "clinical failure," "poor responder," "high-resolution manometry," "functional lumen imaging probe," "FLIP," "EndoFLIP," "outcome prediction," "reflux," "training," or "video assessment." Relevant studies published through the April 2026 cutoff were considered, and reference lists of relevant guidelines, reviews, and eligible studies were also examined.

English-language, peer-reviewed human studies directly addressing achalasia or POEM were prioritized. Conventional regression-based POEM risk scores were included when they addressed treatment response, technical difficulty, adverse events, or reflux because they are methodological precursors to AI-enabled prediction; they were described separately and were not treated as deployable AI systems. Studies from endoscopic submucosal dissection (ESD), minimally invasive surgery, or general endoscopy were included only when they addressed an analogous video or workflow task and were labeled

as adjacent evidence. Reports were selected purposively for clinical relevance and methodological detail. This was a narrative review; duplicate independent screening, protocol registration, formal risk-of-bias assessment, and meta-analysis were not undertaken.

Evidence was classified as clinically evaluated when a complete system or conventional prediction model had been tested in a patient cohort, reader study, or clinical prediction cohort; proof of concept when evaluation was image-level, preclinical, a small feasibility study, or limited to a component task rather than a complete clinical model; adjacent when evidence came from a different procedure; and conceptual when no direct POEM evaluation was available. Retrospective design and internal validation were recorded as limitations rather than overriding the clinically evaluated label for completed prediction models. Conventional regression-based scores were identified explicitly so that clinical evaluation was not mistaken for AI-specific integration or readiness for routine care. These labels describe the maturity of evaluation and do not imply deployability.

Current Evidence Base and Maturity

The strongest direct evidence concerns recognition of achalasia during diagnostic endoscopy. Kaklamanos et al. (6) assembled 1,229 videos from 1,203 patients investigated for dysphagia at one center. Their transformer model used 1,036 videos from 1,010 patients for training, 98 patients and videos for validation, and 95 patients and videos for testing, with HRM as the reference standard. In the held-out test set, model accuracy was 0.926, F1 score 0.821, sensitivity 0.800, and negative predictive value 0.947; two expert endoscopists had accuracy 0.905, F1 score 0.710, sensitivity 0.550, and negative predictive value 0.893 (6). The single-center, enriched dysphagia cohort, exclusions, and small number of achalasia cases in the test set limit generalizability to routine unselected endoscopy.

Shiwaku et al. (7) conducted a multicenter reader study involving 83 endoscopists from 27 centers. Readers assessed 50 selected videos, including 25 with achalasia, before and after computer-aided detection. Sensitivity increased from 74.2% to 91.2% overall and from 66.9% to 91.9% among less-experienced readers (7). The artificial disease prevalence, repeated case set, and video-reading design do not establish performance in a live workflow. Both diagnostic studies are vulnerable to spectrum and selection bias when results from enriched cohorts and selected videos are extrapolated to routine endoscopy.

Evidence for automated physiological testing is less mature. HRM remains the diagnostic reference standard under Chicago Classification version 4.0 (8), and existing machine-learning work has mainly addressed component tasks rather than complete clinical interpretation (9). Czako et al. (10) used single-center, markedly imbalanced image datasets and an internal image-level split without external patient-level validation of a complete motility diagnosis. Carlson et al. (11) studied FLIP panometry in 180 treatment-naive patients with achalasia, using 140 for training and 40 for testing. Test accuracy was 78% for spastic versus non-spastic achalasia and 55% for all three HRM subtypes, supporting a preliminary phenotype model rather than a replacement for HRM.

Conventional regression-based POEM prediction models provide direct clinical precursor evidence. Urakami et al. (12) developed a poor-response score in 244 patients and internally validated

it by bootstrapping [area under the curve (AUC) 0.78]. Liu et al. (13) developed a major-adverse-event score in 3,135 patients with bootstrap internal validation (AUC 0.795), while a separate 3,385-patient study used a 1,693-patient training cohort and 1,692-patient internal validation cohort to develop a technical-difficulty nomogram (validation AUC 0.743) (14). These retrospective models are not deployable AI systems and lack independent prospective validation.

The principal direct machine-learning outcome-prediction study included 1,824 treatment-naive patients across multiple Japanese centers, 1,778 of whom underwent POEM. The area under the receiver operating characteristic curve was 0.70 for persistent symptoms and 0.61 for reflux esophagitis, indicating moderate and poor discrimination, respectively (15). No independent geographic validation or prospective decision-impact evaluation was reported. Figure 1 maps these direct findings alongside proof-of-concept, adjacent, and conceptual applications.

AI-assisted Diagnosis and Patient Selection Before POEM

Accurate diagnosis and careful patient selection are central to successful POEM. Although POEM is highly effective for achalasia, it is not appropriate for every patient with dysphagia or esophageal outflow obstruction. The decision to proceed with POEM requires integration of symptoms, endoscopic findings, HRM, TBE, FLIP data, previous interventions, anatomical complexity, patient preference, and reflux risk. Current diagnostic frameworks, including the Chicago Classification version 4.0, provide a structured approach to esophageal motility disorders, but interpretation can remain challenging in borderline cases, spastic disorders, opioid-associated dysmotility, and patients with prior therapy (8,16).

AI may improve the earliest stage of the POEM pathway by supporting recognition of achalasia during routine upper endoscopy. Achalasia can be missed because endoscopic findings are often subtle, including retained saliva or fluid, mild esophageal dilation, abnormal contractions, resistance at the esophagogastric junction (EGJ), or a tight-appearing lower esophageal sphincter. Recent computer-vision work suggests that AI can identify achalasia from upper endoscopy videos with performance comparable to expert endoscopists. This is clinically relevant because many patients experience delayed diagnosis and may receive repeated endoscopy or empirical reflux therapy before manometry confirms the disorder (6,7).

AI may also support interpretation of HRM, which remains the diagnostic cornerstone for achalasia and other esophageal motility disorders. Manometric interpretation requires specialist expertise and may be affected by artifacts, catheter position, patient tolerance, and interobserver variability. Machine-learning tools could help classify motility patterns, identify achalasia subtypes, distinguish true spastic disease from artifact, and flag complex studies for expert review. This is important for POEM because manometric phenotype influences procedural planning, particularly in type III achalasia and spastic disorders, where a longer proximal myotomy is often required (9,10).

Patient selection should extend beyond confirming the diagnosis. A 244-patient regression-based score used the pretreatment Eckardt score, previous treatment, sigmoid esophagus, and esophageal dilation grade to estimate poor response after POEM, but it was only

Artificial intelligence across the peroral endoscopic myotomy pathway

Evidence maturity and clinical readiness at each stage

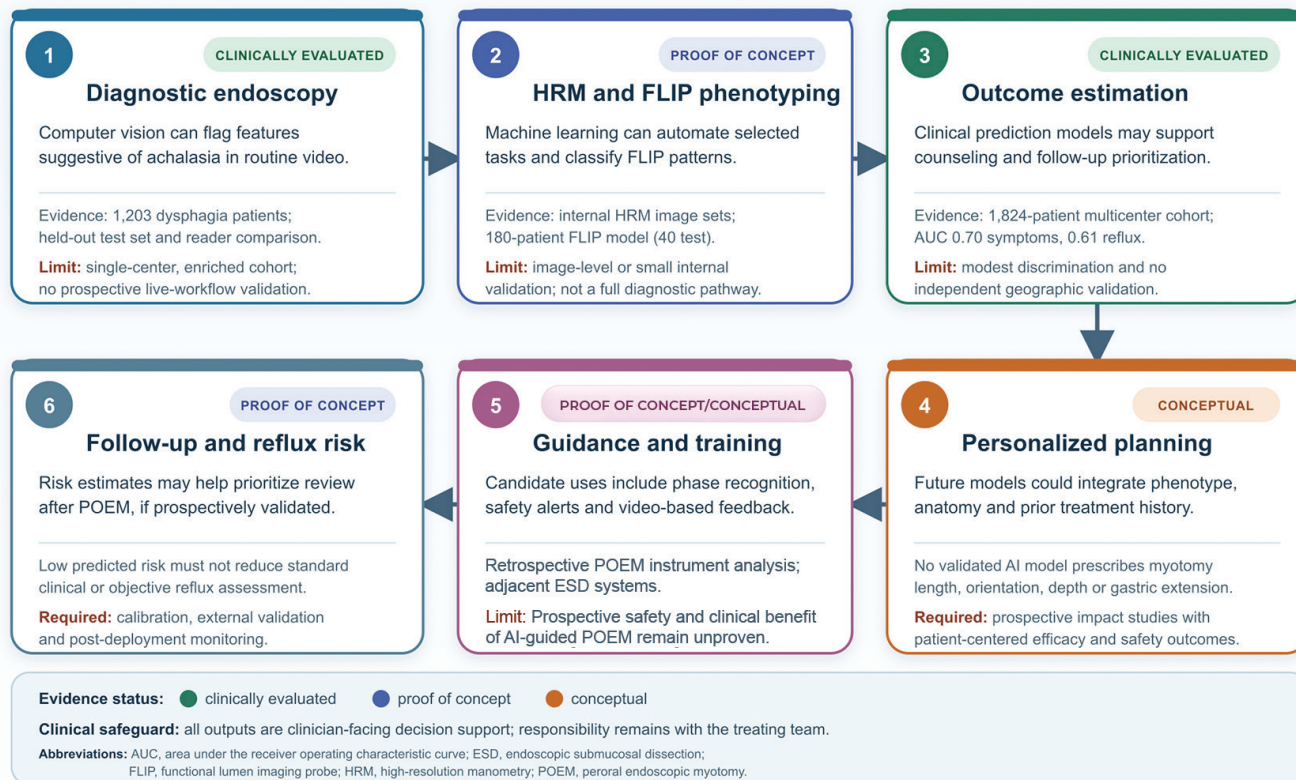


Figure 1. Evidence maturity across the proposed artificial intelligence-assisted peroral endoscopic myotomy pathway. “Clinically evaluated” denotes testing in a patient, reader, or prediction cohort and does not indicate readiness for routine care. Direct clinical evidence includes diagnostic endoscopy, retrospective machine-learning outcome estimation, and conventional regression-based POEM risk scores that are methodological precursors rather than deployable AI systems. Automated high-resolution manometry and functional lumen imaging probe interpretation remain proof of concept; personalized myotomy prescription, real-time intra-procedural guidance, and automated training assessment remain conceptual for peroral endoscopic myotomy. A low predicted reflux risk must not replace standard clinical follow-up or objective reflux assessment.

AI: Artificial intelligence, AUC: Area under the receiver operating characteristic curve, ESD: Endoscopic submucosal dissection, FLIP: Functional lumen imaging probe, HRM: High-resolution manometry, POEM: Peroral endoscopic myotomy. (This original figure was created by the authors for this review.)

internally validated and did not compare POEM with alternative treatments (12). A clinically useful model would ideally estimate the likely benefit of POEM compared with pneumatic dilation, laparoscopic Heller myotomy with fundoplication, botulinum toxin injection, redo myotomy, or conservative management. Contemporary guidelines support POEM as a safe and effective treatment for achalasia, with particular value in type III achalasia because the myotomy can be extended proximally according to the spastic segment. However, treatment selection remains complex in patients with advanced sigmoid achalasia, severe esophageal dilation, previous interventions, frailty, high reflux risk, or discordant symptoms (2,3,16).

Functional assessment may further strengthen AI-assisted selection. FLIP can assess EGJ distensibility and provide complementary information before, during, and after myotomy. Persistently low distensibility after treatment has been associated with poorer symptomatic outcomes, while excessive disruption of the antireflux

barrier may increase reflux risk. Future clinical prediction models could combine distensibility metrics with HRM, TBE, endoscopic findings, and symptom scores to identify patients most likely to benefit from POEM and estimate the degree of myotomy needed to achieve adequate outflow relief without unnecessary disruption (11,17).

A practical AI-assisted pre-POEM workflow could include three levels. First, computer vision could flag suspected achalasia during diagnostic endoscopy. Second, automated analysis of manometry and functional testing could support disease phenotyping and identify complex or discordant cases. Third, clinical prediction models could estimate treatment response, technical difficulty, recurrence risk, and post-POEM reflux risk. Conventional regression scores already provide internally validated precursor evidence for several of these tasks, but they are not AI systems and have not shown prospective decision impact. An AI-assisted approach would not replace clinical judgment, but it could standardize decision-

making and reduce variation between centers, particularly where access to expert motility interpretation is limited (6,7,9-15).

At present, AI-specific pre-POEM assessment remains early and should be viewed as investigational rather than practice-changing. Most AI studies are proof of concept, retrospective, or focused on diagnosis rather than treatment selection; conventional POEM risk scores are clinically evaluated but remain internally validated precursors. The key question is not simply whether a model can classify achalasia, but whether it improves outcomes by helping clinicians select the right patient, intervention, and procedural strategy.

AI for Procedural Planning: Toward Personalized Myotomy

Procedural planning is one of the most attractive but least developed applications of AI in POEM. Although the procedure is often described in standard technical steps, the optimal strategy varies substantially between patients. The endoscopist must decide the orientation of the tunnel, the entry site, the length of esophageal and gastric myotomy, the depth of myotomy, and the degree of extension required across the EGJ. These decisions are usually based on endoscopic judgment, manometric subtype, anatomical features, and local expertise rather than objective prediction models (18).

The most important planning question is myotomy length. In type I and type II achalasia, shorter myotomy strategies may be adequate in selected patients, while type III achalasia and spastic esophageal disorders often require a longer myotomy that extends proximally to cover the spastic segment. This is a natural area for AI because a future model could integrate HRM, TBE, FLIP findings, endoscopic anatomy, symptom pattern, and prior treatment history to estimate the minimum effective myotomy length for each patient (19).

AI may also help reduce unnecessary procedural variation. A recent meta-analysis showed substantial heterogeneity in reported myotomy length across POEM studies, with longer pooled myotomy lengths in type III achalasia than in non-spastic achalasia. This variation reflects differences in disease phenotype, operator preference, and evolving procedural strategy. A validated prediction model could help distinguish patients who need extended myotomy from those in whom shorter myotomy may achieve adequate symptom relief while potentially reducing procedural time, tissue disruption, and reflux risk (20).

Functional assessment could make procedural planning more objective. FLIP provides real-time information on EGJ distensibility and may help assess whether the myotomy has achieved adequate physiological disruption. Studies have shown that distensibility changes during POEM can be measured incrementally and may correlate with clinical outcomes. In the future, a model could combine baseline and intra-procedural distensibility data with manometric phenotype to guide whether additional proximal or gastric extension is required (21,22).

Another planning issue is the risk of post-POEM reflux. Unlike laparoscopic Heller myotomy, POEM is not routinely combined with fundoplication, and reflux remains a major limitation. Proposed predictors remain plausible but heterogeneous candidates. In a 92-patient cohort, loss of inspiratory augmentation was associated with GERD in the evaluable HRM subgroup. Shorter intra-abdominal lower esophageal sphincter length and type III EGJ morphology

showed nonsignificant trends; myotomy length and orientation were not significantly associated (23). These findings require external validation and cannot support routine model-based planning.

AI may also assist in predicting technical difficulty. Direct conventional evidence already exists: a regression nomogram developed in 3,385 patients and internally validated in 1,692 identified age, disease duration, sigmoid esophagus, mucosal edema, submucosal fibrosis, and tunnel length as predictors (14). This model is not an AI system and lacks independent prospective validation, but it establishes that technical-difficulty prediction is not entirely conceptual. Future models could integrate procedural videos, endoscopic images, radiology, and clinical variables to support operator, center, equipment, and backup-plan selection (18).

AI-guided prescription of myotomy length, orientation, depth, or gastric extension remains conceptual, whereas conventional models for technical difficulty and major adverse events have undergone internal clinical evaluation (13,14). No validated AI tool can prescribe a POEM strategy in routine practice. Future studies should develop clinically actionable models using multicenter datasets, standardized procedural variables, objective reflux testing, and long-term clinical outcomes, followed by external and prospective validation. The aim should not be to replace expert endoscopists, but to make planning more reproducible, transparent, and personalized (18-22).

AI for Intra-procedural Guidance and Training

AI-assisted intra-procedural guidance represents the most ambitious future application in POEM. Unlike diagnostic endoscopy, POEM requires continuous recognition of dynamic procedural steps, including mucosal incision, submucosal tunnel creation, myotomy, hemostasis, extension across the EGJ, and mucosal closure. In theory, computer vision could provide real-time support by identifying the procedural phase, maintaining orientation within the tunnel, highlighting vessels, warning of mucosal injury, and prompting reassessment before closure. The main potential applications across the POEM pathway are summarized in Table 1.

The rationale for intra-procedural AI is strengthened by progress in adjacent therapeutic endoscopy. In ESD, computer-vision systems have been developed for workflow recognition, anatomical structure delineation, and real-time orientation support. Although these models cannot be directly transferred to POEM, they show that AI can analyze complex therapeutic endoscopy videos and provide cognitive assistance beyond simple lesion detection (24,25).

The supporting evidence remains preliminary and procedure-specific. Cao et al. (24) developed an ESD workflow-recognition system from 201,026 frames across 47 cases, performed a small external evaluation, and demonstrated real-time operation in an animal study. Scheppach et al. (25) trained anatomical segmentation on 1,011 still images from 30 ESDs and evaluated real-time use in 12 porcine procedures. These studies demonstrate technical feasibility in an adjacent procedure, not safety or benefit during human POEM.

A key intra-procedural target is recognition of the correct submucosal plane. Loss of orientation, difficult tunneling, fibrosis, bleeding, and inadvertent mucosal injury can increase technical difficulty and procedural risk. A model could be trained to distinguish the mucosal layer, submucosal tunnel, circular muscle fibers, longitudinal muscle, vessels, and areas of poor visibility. Zhu et al. (26) reported

Table 1. Evidence maturity of potential AI applications across the POEM pathway.

Stage of care	Potential AI application	Expected clinical value	Evidence status	Representative study or data type	Validation status	Current limitation
Diagnosis	Detection of achalasia from endoscopic images or videos	Earlier recognition and referral for manometry	Clinically evaluated	Kaklamano et al. (6); Shiwaku et al. (7)	Single-center held-out test; multicenter reader study	Selected, enriched videos; no prospective live-workflow validation
Patient selection and risk stratification	Prediction of poor response, technical difficulty, major adverse events, and benefit from POEM	Better counseling, case selection, operator assignment, and procedural preparation	Clinically evaluated conventional prediction models; AI-specific integration and external validation remain unavailable	Urakami et al. (12); Liu et al. (13,14); Takahashi et al. (15)	Retrospective bootstrap, split-sample internal, or multicenter development evaluation	No independent prospective validation; conventional regression models are not deployable AI systems or comparative treatment selectors
Procedural planning	Prediction of optimal myotomy length, orientation, and gastric extension	More personalized and reproducible POEM strategy	Conceptual for AI-prescribed planning	Clinical POEM and FLIP cohorts; no direct AI planning model	No validated AI system	No evidence that an AI-prescribed plan improves efficacy or safety
Intra-procedural guidance	Instrument analysis, tunnel orientation, vessel detection and mucosal injury alerts	Improved safety and procedural consistency	Direct POEM instrument proof of concept; navigation/safety remain conceptual	Zhu et al. (26); adjacent ESD studies (24,25)	Retrospective POEM video analysis; no prospective safety/benefit validation	Different anatomy and error costs; false-alert risk
Training	Automated video-based assessment of procedural steps and technical errors	Objective competency assessment and feedback	Automated competency scoring unvalidated	POEM assessment tool: 12 assessors and 10 videos (30)	Initial human-rating validation only	The assessment tool is not AI; automated scoring remains unvalidated
Follow-up	Prediction of clinical recurrence and post-POEM reflux	Tailored surveillance and acid-suppression strategy	Clinically evaluated prediction cohorts; no deployable AI system	Urakami et al. (12); Takahashi et al. (15); Kim et al. (23)	Retrospective internal or development-cohort evaluation; no decision-impact study	Heterogeneous predictors; low predicted risk cannot replace standard follow-up or objective assessment

AI: Artificial intelligence, ESD: Endoscopic submucosal dissection, FLIP: Functional lumen imaging probe, POEM: Peroral endoscopic myotomy.

retrospective instrument-based analysis of POEM videos; general surgical AI reviews provide adjacent evidence (27).

AI may also help standardize recognition and documentation of adverse events. Mucosal injury, bleeding, insufflation-related events, pain, and aspiration are recognized complications during or after POEM, although major adverse events are uncommon in experienced centers. A conventional regression score developed in 3,135 patients predicted major adverse events with bootstrap internal validation, but it was not an AI system and was not independently validated (13). Video-based models could automatically flag bleeding episodes, prolonged hemostasis, suspected mucosal breach, repeated loss of tunnel orientation, or difficult closure. Such systems must be evaluated against established prevention, recognition, and management standards before clinical use (28).

Training is likely to be a more realistic near-term application than real-time procedural control. POEM has a significant learning curve and requires both cognitive and technical skills, including patient selection, tunnel creation, controlled myotomy, hemostasis, complication recognition, and closure. Existing training guidance emphasizes structured curricula, expert supervision, simulation,

animal models, and competency-based assessment rather than reliance on case numbers alone (29).

Recent work on video-based POEM assessment provides an important foundation for this direction. The POEM assessment tool was developed as a structured human rating instrument, not an AI system. In its initial validation, 12 assessors rated 10 procedural videos, with variable inter-rater reliability across domains (30). It may provide a competency framework for future automated assessment, but automated scoring has not yet been prospectively validated.

In the future, the most practical model may be a staged approach. Early tools could support retrospective video review and training assessment. More advanced systems could undergo silent-mode evaluation before carefully bounded phase-recognition or safety alerts are introduced. Only after robust prospective and human-factors validation should AI be considered for real-time decision support during active myotomy.

AI Prediction of Post-POEM Outcomes and Reflux

Prediction of post-POEM outcomes is clinically relevant because technical success does not always guarantee durable symptomatic

benefit. Most patients improve after POEM, but a minority develop persistent symptoms, recurrent dysphagia, need for retreatment, or clinically significant reflux. Conventional regression and machine-learning cohorts provide direct POEM-specific prediction evidence, although current models remain insufficient for routine decision-making (12,15). Outcome assessment relies on symptom scores, endoscopy, pH testing, TBE, HRM, and FLIP, but these data are often interpreted separately rather than integrated into a single personalized risk estimate.

AI could help predict clinical response by combining baseline and procedural variables. Relevant candidates include age, symptom duration, Eckardt score, achalasia subtype, esophageal diameter, sigmoid morphology, prior pneumatic dilation or Heller myotomy, lower esophageal sphincter pressure, integrated relaxation pressure, FLIP distensibility, myotomy length, gastric extension, and intra-procedural complications. Existing conventional and machine-learning models show that such estimation is feasible, but their internal or development-cohort performance is insufficient for individual treatment decisions without independent prospective validation (12,15).

Post-POEM reflux is particularly important because objective reflux can occur even when symptoms are mild or absent. Reported candidate predictors are heterogeneous rather than established. In a 92-patient retrospective cohort, loss of inspiratory augmentation was associated with GERD in the evaluable HRM subgroup; intra-abdominal lower esophageal sphincter length, EGJ morphology, myotomy length, and orientation were not significantly associated (23). Other cohorts have proposed different baseline and procedural variables, and differences in reflux definitions and testing further limit comparability. No reflux-prediction model is ready for routine use, and objective follow-up remains necessary (4,31).

The value of a reflux-prediction model would be greatest if it informed counseling and surveillance without weakening established care. A high predicted risk could support careful counseling, avoidance of excessive gastric extension, consideration of sling-fiber-preserving techniques or emerging antireflux POEM concepts, and structured post-procedure surveillance (32). Conversely, a low predicted risk must not be used to omit standard clinical follow-up, acid-suppression decisions, endoscopy, or objective reflux assessment when otherwise indicated.

Longitudinal prediction is another important opportunity. Some patients remain asymptomatic early after POEM but later develop recurrent symptoms, reflux esophagitis, abnormal acid exposure, or need for retreatment. A future model could update risk estimates using serial Eckardt scores, medication use, endoscopic findings, pH testing, TBE, and functional measurements. This could support risk-adapted follow-up, provided that standard surveillance is not withheld on the basis of an unvalidated estimate.

However, outcome prediction models must avoid overfitting and clinically misleading conclusions. Many POEM studies are retrospective, single-center, and heterogeneous in follow-up duration, reflux definitions, pH testing protocols, and outcome thresholds. Models trained on such data may perform well internally but fail when applied to different populations, operators, devices, and follow-up pathways. Future models should therefore use standardized definitions of clinical success, objective reflux testing,

adequate follow-up, external validation, and transparent reporting and risk-of-bias standards (33,34).

The ideal AI-assisted POEM pathway would begin before the procedure and continue into follow-up. AI could flag suspected achalasia during endoscopy, support manometric phenotyping, predict benefit from POEM, assist with procedural planning, provide video-based quality assessment, and estimate the risk of recurrence or reflux after treatment. Such a system would not replace the endoscopist but could create a more consistent and data-driven care pathway. Figure 1 summarizes this proposed workflow, while Table 1 summarizes the evidence status of the major applications.

Safety, Generalizability, and Governance

Dataset shift is a central safety problem. A model trained in a high-volume expert center may fail when case mix, disease prevalence, endoscope manufacturer, processor, image enhancement, catheter, FLIP protocol, operator technique, or annotation practice changes. Class imbalance may conceal poor sensitivity for uncommon but consequential events such as mucosal injury. Validation should therefore include geographically and technically distinct centers, temporal testing, calibration plots, and prespecified subgroup performance by sex, age, ethnicity, disease subtype, previous treatment, and procedural complexity.

Fairness cannot be inferred from a single overall accuracy. Underrepresented groups may receive less reliable estimates, and missing data may encode unequal access to specialist testing. Explainability should be matched to the clinical task. False alerts can distract the operator, whereas false reassurance can delay recognition of a complication. Automation bias, alert fatigue, and the effect of AI on trainee skill acquisition require direct human-factors evaluation.

Governance must address lawful data use, privacy protection for endoscopy video and physiological data, cybersecurity, version control, audit logs, regulatory responsibility, and clear assignment of accountability when the clinician and system disagree. Institutions should define who may override an output, how failures are reported, and how performance is monitored after deployment. Prospective surveillance should assess drift, near misses, subgroup harm, and workflow effects, with a predefined withdrawal or recalibration process.

Prediction-model reports should follow the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis plus AI statement, and risk of bias should be assessed with the prediction model risk of bias assessment tool plus AI (33,34). Early live clinical studies should address the developmental and exploratory clinical investigations of decision-support systems driven by AI guidance, while the fairness, universality, traceability, usability, robustness, and explainability for AI framework informs the wider lifecycle (35,36). Table 1 summarizes the evidence maturity, validation status, and principal limitation of each application.

Novelty of This Review

The novelty of this narrative review is its stage-specific evidence-maturity framework across the complete POEM pathway. It distinguishes direct achalasia or POEM evidence from adjacent therapeutic-endoscopy research and conceptual proposals, and it

links each application to its validation status, principal limitation, and required clinical safeguards. This approach provides a clinically oriented research agenda without implying that conceptual procedural tools are ready for patient care.

CONCLUSION

AI-specific applications across the POEM pathway remain at an early stage, although conventional regression models provide internally validated precursor evidence for poor response, technical difficulty, and major adverse events. The most realistic near-term AI uses are clinician-facing detection support during diagnostic endoscopy, bounded assistance with HRM or FLIP interpretation, retrospective video review, and cautiously calibrated outcome stratification. Even these uses require external and prospective validation in representative populations before routine adoption.

Personalized myotomy prescription, real-time tunnel navigation, vessel or mucosal-injury alerts, and automated training assessment remain speculative for POEM. Progress requires multicenter datasets, standardized outcomes, calibration and subgroup analyses, human-factors and clinical-impact studies, regulatory clarity, and post-deployment monitoring. AI should augment expert judgment and must not weaken established diagnostic, procedural, or reflux follow-up safeguards.

Acknowledgments

Artificial intelligence was used only for language editing and grammatical corrections. The authors reviewed and approved the final manuscript and take full responsibility for its content.

Footnotes

Authorship Contributions

Concept: A.A.S., Design: A.A.S., A.A., Data Collection or Processing: A.A.S., A.A., Analysis or Interpretation: A.A.S., A.A., Literature Search: A.A.S., A.A., Writing: A.A.S., A.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

- Inoue H, Sato H, Ikeda H, Onimaru M, Sato C, Minami H, et al. Peroral endoscopic myotomy: a series of 500 patients. *J Am Coll Surg*. 2015; 221: 256-64.
- Calabrese EC, Kindel T, Slater BJ, Marks J, Swanstrom L, Saxena P, et al. 2024 update to SAGES guidelines for the use of peroral endoscopic myotomy (POEM) in the treatment of achalasia. *Surg Endosc*. 2025; 39: 4027-37.
- Kohn GP, Dirks RC, Ansari MT, Clay J, Dunst CM, Lundell L, et al. SAGES guidelines for the use of peroral endoscopic myotomy (POEM) for the treatment of achalasia. *Surg Endosc*. 2021; 35: 1931-48.
- Werner YB, Hakanson B, Martinek J, Repici A, von Rahden BHA, Bredenoord AJ, et al. Endoscopic or surgical myotomy in patients with idiopathic achalasia. *N Engl J Med*. 2019; 381: 2219-29.
- Bencardino S, Lodola I, Centanni L, Mandarino FV, Fanizza J, Furfaro F, et al. Artificial intelligence in advanced endoscopic imaging: transforming optical diagnosis in gastroenterology. *Front Med (Lausanne)*. 2026; 12: 1719145.
- Kaklamanos E, Kristinsdottir K, Wittbrodt M, Li M, Pitisuttithum P, Kou W, et al. From pixels to peristalsis: comparing achalasia diagnosis with artificial intelligence with expert endoscopists. *Clin Gastroenterol Hepatol*. 2026; 24: 1542-9.
- Shiwaku H, Misawa M, Inoue H, Jiang K, Oda M, Familiari P, et al. Computer-aided detection for esophageal achalasia (with video). *Dig Endosc*. 2025; 37: 851-61.
- Yadlapati R, Kahrilas PJ, Fox MR, Bredenoord AJ, Prakash Gyawali C, Roman S, et al. Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0©. *Neurogastroenterol Motil*. 2021; 33: e14058. Erratum in: *Neurogastroenterol Motil*. 2024; 36: e14179.
- Fass O, Rogers BD, Gyawali CP. Artificial intelligence tools for improving manometric diagnosis of esophageal dysmotility. *Curr Gastroenterol Rep*. 2024; 26: 115-23.
- Czako Z, Surdea-Blaga T, Sebestyen G, Hangan A, Dumitrascu DL, David L, et al. Integrated relaxation pressure classification and probe positioning failure detection in high-resolution esophageal manometry using machine learning. *Sensors (Basel)*. 2021; 22: 253.
- Carlson DA, Kou W, Rooney KP, Baumann AJ, Donnan E, Triggs JR, et al. Achalasia subtypes can be identified with functional luminal imaging probe (FLIP) panometry using a supervised machine learning process. *Neurogastroenterol Motil*. 2021; 33: e13932.
- Urakami S, Abe H, Tanaka S, Kawara F, Toyonaga T, Ariyoshi R, et al. Development of a preoperative risk-scoring system for predicting poor responders to peroral endoscopic myotomy. *Gastrointest Endosc*. 2021; 93: 398-405.
- Liu X, Yao L, Cheng J, Xu M, Chen S, Zhong Y, et al. Landscape of adverse events related to peroral endoscopic myotomy in 3135 patients and a risk-scoring system to predict major adverse events. *Clin Gastroenterol Hepatol*. 2021; 19: 1959-66.e3.
- Liu XY, Geng ZH, Chen WF, Xu MD, Chen SY, Zhong YS, et al. A prediction model and nomogram for technical difficulty of peroral endoscopic myotomy. *Surg Endosc*. 2023; 37: 2781-8.
- Takahashi K, Sato H, Shimamura Y, Abe H, Shiwaku H, Shiota J, et al. Achalasia phenotypes and prediction of peroral endoscopic myotomy outcomes using machine learning. *Dig Endosc*. 2024; 36: 789-800.
- Khashab MA, Vela MF, Thosani N, Agrawal D, Buxbaum JL, Abbas Fehmi SM, et al. ASGE guideline on the management of achalasia. *Gastrointest Endosc*. 2020; 91: 213-27.e6.
- Rohof WO, Hirsch DP, Kessing BF, Boeckstaens GE. Efficacy of treatment for patients with achalasia depends on the distensibility of the esophagogastric junction. *Gastroenterology*. 2012; 143: 328-35.
- Hasan A, Low EE, Fehmi SA, Yadlapati R. Evolution and evidence-based adaptations in techniques for peroral endoscopic myotomy for achalasia. *Gastrointest Endosc*. 2022; 96: 189-96.
- Srinivasan V, Shah ED. Mini-review: tailored per-oral endoscopic myotomy for type III achalasia. *Neurogastroenterol Motil*. 2023; 35: e14700.
- Vespa E, Barchi A, Mandarino FV, Fasulo E, Fratto MC, Passaretti S, et al. Standard length of peroral endoscopic myotomy (POEM) for achalasia: a systematic review and meta-analysis. *Dis Esophagus*. 2024; 37: doae069.
- Knowles TB, Jackson AS, Chang SC, Schembre DB, Farivar AS, Aye RW, et al. Changes in distensibility index during an incremental POEM myotomy. *J Gastrointest Surg*. 2022; 26: 1140-6.
- Knight W, Kandiah K, Vrakopoulou Z, White A, Barbieri L, Tewari N, et al. Early outcomes following EndoFLIP-tailored peroral endoscopic myotomy (POEM): the establishment of POEM services in two UK centers. *Dis Esophagus*. 2023; 36: doac110.

23. Kim MJ, Jeon SY, Park JC, Youn YH, Park H. Loss of inspiratory augmentation as a predictive factor for the development of gastroesophageal reflux disease after peroral endoscopic myotomy in patients with achalasia. *J Neurogastroenterol Motil.* 2026; 32: 129-35.
24. Cao J, Yip HC, Chen Y, Scheppach M, Luo X, Yang H, et al. Intelligent surgical workflow recognition for endoscopic submucosal dissection with real-time animal study. *Nat Commun.* 2023; 14: 6676.
25. Scheppach MW, Yip HC, Chen Y, Yang H, Cao J, Chua T, et al. Feasibility of real-time artificial intelligence-assisted anatomical structure recognition during endoscopic submucosal dissection. *Endosc Int Open.* 2025; 13: a26158008.
26. Zhu Y, Du L, Fu PY, Geng ZH, Zhang DF, Chen WF, et al. An automated video analysis system for retrospective assessment and real-time monitoring of endoscopic procedures (with video). *Bioengineering (Basel).* 2024; 11: 445.
27. Arakaki S, Takenaka S, Sasaki K, Kitaguchi D, Hasegawa H, Takeshita N, et al. Artificial intelligence in minimally invasive surgery: current state and future challenges. *JMA J.* 2025; 8: 86-90.
28. Nabi Z, Reddy DN, Ramchandani M. Adverse events during and after per-oral endoscopic myotomy: prevention, diagnosis, and management. *Gastrointest Endosc.* 2018; 87: 4-17.
29. Dacha S, Aihara H, Anand GS, Byrne KR, Chahal P, James T, et al. Core curriculum for peroral endoscopic myotomy (POEM). *Gastrointest Endosc.* 2021; 93: 539-43.
30. Yang D, Draganov PV, Pohl H, Aihara H, Jeyalingam T, Khashab M, et al. Development and initial validation of a video-based peroral endoscopic myotomy assessment tool. *Gastrointest Endosc.* 2024; 99: 177-85.
31. Repici A, Fuccio L, Maselli R, Mazza F, Correale L, Mandolesi D, et al. GERD after per-oral endoscopic myotomy as compared with Heller's myotomy with fundoplication: a systematic review with meta-analysis. *Gastrointest Endosc.* 2018; 87: 934-43.e18.
32. Inoue H, Ueno A, Shimamura Y, Manolakis A, Sharma A, Kono S, et al. Peroral endoscopic myotomy and fundoplication: a novel NOTES procedure. *Endoscopy.* 2019; 51: 161-4.
33. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024; 385: e078378. Erratum in: *BMJ.* 2024; 385: q902.
34. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* 2025; 388: e082505.
35. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al.; DECIDE-AI expert group. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ.* 2022; 377: e070904.
36. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al.; FUTURE-AI Consortium. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ.* 2025; 388: e081554. Erratum in: *BMJ.* 2025; 388: r340.