BMJ Open Gastroenterology

Machine learning in gastrointestinal endoscopy: challenges and opportunities

Sergejs Lobanovs ,[1,2] Jekaterina Aleksejeva,[1,2] Alise Kitija Rūtiņa,[1,2] Eduards Krustiņš,[2,3] Jurijs Čižovs,[4] Dmitrijs Bļizņuks[4]

To cite: Lobanovs S, Aleksejeva J, Rūtiņa AK, *et al*. Machine learning in gastrointestinal endoscopy: challenges and opportunities. *BMJ Open Gastroenterol* 2025;**12**:e001923. doi:10.1136/bmjgast-2025-001923

Received 27 May 2025
Accepted 16 September 2025

[1]Faculty of Medicine, University of Latvia, Riga, Latvia
[2]Pauls Stradiņš Clinical University Hospital, Riga, Latvia
[3]Rīga Stradiņš University, Riga, Latvia
[4]Riga Technical University, Riga, Latvia

Correspondence to
Dr Sergejs Lobanovs; sergeyslobanovs@gmail.com

ABSTRACT

The integration of machine learning (ML) into medical diagnostics has significantly advanced endoscopic examinations for gastrointestinal diseases. By leveraging extensive datasets and sophisticated algorithms, ML technologies enhance diagnostic precision, detect subtle abnormalities, classify diverse pathologies and predict disease progression. However, their widespread adoption is hindered by the inherent heterogeneity of gastrointestinal diseases, technical limitations, limited generalisability across different populations and ethical challenges related to patient privacy, data security and algorithmic bias.

This review provides a comprehensive structural analysis of ML approaches in endoscopy, starting with an overview of the classical endoscopic methodology that relies on direct visualisation of the gastrointestinal tract for diagnosis and therapeutic interventions. Then, current ML applications that hold promise for reducing physician-dependent variability, improving diagnostic accuracy and streamlining procedural workflows were explored. Despite these advances, the effectiveness of ML models often remains constrained by the quality and diversity of training data, which can undermine both reliability and generalisability.

Ethical considerations – such as safeguarding patient information, upholding data security and mitigating biases embedded in algorithms – are integral to responsibly deploying ML in clinical settings. By examining these technical and ethical barriers, this work contributes to the evolving discourse on integrating advanced ML techniques into gastroenterology. Ultimately, our goal is to pave the way for more effective and reliable ML-driven endoscopic practices that will enhance disease detection, optimise patient care and benefit healthcare providers worldwide.

SUMMARY BOX

- ⇒ Enhanced diagnostics: machine learning (ML) improves gastrointestinal endoscopy by increasing lesion detection rates, providing accurate lesion characterisation, predicting disease outcomes, standardising examination quality and reducing operator-dependent variability.
- ⇒ Key applications: ML is applied in detection, diagnosis, therapeutic assistance and technical support, aiding both accuracy and efficiency.
- ⇒ Training data challenges: large, diverse, high-quality annotated datasets are essential, but current limitations include bias, rare disease underrepresentation and reliance on retrospective studies.
- ⇒ Annotation and image quality issues: manual labelling is slow, while artefacts, reflections and low-quality images hinder accuracy; tools like Label Studio and SAM help accelerate reliable annotation.
- ⇒ Model optimisation trade-offs: Accuracy must be balanced with speed, cost and computational efficiency, using techniques such as pruning, quantisation, transfer learning (TL) and knowledge distillation.
- ⇒ Trust, ethics and regulation: transparency via explainable artificial intelligence, patient privacy (eg, General Data Protection Regulation), algorithmic bias, liability and data security remain major barriers to safe adoption.
- ⇒ Integration into practice: real-time deployment is limited by technical infrastructure, workflow disruption and false positives; future success depends on standardised protocols, clinician training and interdisciplinary collaboration.

INTRODUCTION

Machine learning (ML), a subset of artificial intelligence (AI), refers to the development and application of algorithms that can learn from and make predictions or decisions based on data. In gastrointestinal endoscopy, ML algorithms automatically analyse endoscopic images or videos, detect relevant patterns and assist in diagnosing, characterising or localising gastrointestinal diseases. These systems can help endoscopists identify lesions such as polyps, cancers and other abnormalities, thereby improving diagnostic accuracy and reducing the risk of missed findings during procedures. In recent years, the integration of ML and artificial neural networks (ANNs) into medical diagnostics has revolutionised endoscopic examination, particularly in the context of gastrointestinal tract (GIT) diseases.[1 2] The advent of ML and ANN has introduced transformative possibilities for improving the precision and efficiency of disease diagnosis through endoscopy. These technologies harness the power of vast datasets and complex algorithms to enhance the accuracy of identifying subtle abnormalities, classify diverse pathology and even predict disease progression.[1 3 4]

A key advantage of ML integration is the reduction of variability in accuracy and sensitivity associated with physician expertise, thereby mitigating risks related to insufficient knowledge. It is crucial to emphasise that ML complements the expertise of physicians but does not replace them, leading to optimal and efficient outcomes.[4]

ML algorithms are currently utilised in four domains[3]:

Computer-assisted detection (CADe), aiding in lesion identification, for example, polyp or bleeding detection.[3 5]

Computer-assisted diagnosis (CADx), aiding in lesion characterisation and classification, for example, polyps, peptic ulcers, inflammatory bowel disease and cancers.[3 6]

Therapeutic assistance aids interventional endoscopists in swiftly and precisely making therapeutic decisions.[3] For instance, the complexity of deciding on additional surgical interventions after the endoscopic removal of T1 colorectal cancer (CRC) arises from the difficulty in accurately predicting lymph node metastasis (LNM) before surgery. ML can anticipate the presence of LNM, reducing the need for extra surgery. In comparison to existing protocols, ML notably diminishes the occurrence of unnecessary supplementary surgeries after endoscopic resection of T1 CRC while effectively identifying cases of LNM positivity.[7]

Technical support enhances gastrointestinal endoscopy performance, including scope insertion guidance.[3] For example, colonoscopy without sedation can be uncomfortable for patients. A new computer-assisted colonoscopy features an articulated, computer-controlled insertion tube that mimics the tip's shape and position during insertion, improving precision and comfort.[8]

A brief overview of representative studies in CADe, CADx, therapeutic assistance, technical support and quality assurance (QA) applications is provided in table 1. We will reference it throughout our discussion of the ML system lifecycle.

Table 1 Representative machine learning applications in gastrointestinal endoscopy

Procedure type	Clinical application	ML approach	Evidence of benefit
Colonoscopy	Outcome prediction	XGBoost with clinical data	Accurate prediction of polyp recurrence after endoscopic mucosal resection[59]
	Polyp detection (CADe)	Deep CNN (ENDO-AID and EndoScreener) and RetinaNet	Increased adenoma detection rate by 7%–10%[60,61,62]
	Quality assessment	3D-ResNet	Automated bowel preparation scoring with high correlation to experts[63]
	Polyp segmentation	YOLACT-based models	Real-time polyp delineation during procedures[64]
Upper GI endoscopy	Bleeding risk prediction	XGBoost and CatBoost	Improved mortality risk stratification vs traditional scores[65]
	Barrett's neoplasia detection	Deep CNN with transfer learning	Outperformed general endoscopists in detection rates[66]
	Pylori detection	CNN-based CADx	Real-time evaluation without biopsy requirement[67]
Upper GI endoscopy/ colonoscopy	Polyp characterisation (CADx)	CNN with transfer learning, SVM and random forest	Improved optical diagnosis accuracy[68,69,70]
	GIT disorder classification	Naive Bayes, random forest, ensemble methods (NasNet and EfficientNet)	Improved prediction of polyp recurrence, bleeding outcomes[71]
Capsule endoscopy	Bleeding lesion detection	Deep CNN	Faster, more accurate detection vs standard reading[72]
	Multi-abnormality detection	YOLOv5 variants	Automated small bowel lesion identification[73,74]
EUS	Pancreatic mass differentiation	CNN with interpretable AI (SHAP)	Distinguished PDAC from other masses with high accuracy[75,76]
	Multimodal analysis	ResNet-50 with data fusion	Enhanced diagnostic accuracy combining multiple inputs[77]

AI, artificial intelligence ; CADe, computer-assisted detection; CADx, computer-assisted diagnosis; CNN, convolutional neural network; EUS, endoscopic ultrasound; ML, machine learning; PDAC, pancreatic ductal adenocarcinoma; SHAP, Shapley additive exPlanations; SVM, support vector machine; YOLACT, You Only Look At CoefficientTs.

Additionally, ML leverages patient data to create prognostic models that optimise treatment planning and execution, though the effectiveness of some models remains uncertain.[3 9] Harnessing advanced technologies, healthcare professionals aim to innovate and improve patient outcomes.

Despite these promising advancements, the application of ML and ANNs in endoscopic examinations is not without its challenges and limitations. Moreover, the heterogeneity of GIT diseases presents a significant obstacle. GIT diseases exhibit significant demographic diversity, with variations observed across different populations and regions.[10–12] The limited generalisability of ANNs across different diseases and patient populations raises questions about their reliability in real-world clinical settings.[10]

Ethical considerations also play a pivotal role in ML deployment in medical applications. Patient privacy, data security and the potential for algorithmic bias must be addressed to ensure that the benefits of these technologies outweigh their risks.[13]

This review aimed to explore and address these issues comprehensively. By shedding light on these aspects, this work contributes to the ongoing dialogue surrounding the integration of cutting-edge computational techniques into the practice of gastroenterology, ultimately paving the way for more effective disease detection and patient care.

Since modern ML systems are not merely programmes but cloud-based solutions built around an ML model, a high-level view of the ML lifecycle (figure 1) will be used to more clearly illustrate our findings.

TRAINING DATA

The section explores the critical aspects of training data, which form the foundation of effective ML applications in endoscopic examinations. It addresses patient data diversity, dataset quality, annotation efforts and challenges such as imaging limitations and false reflections.

Patient data diversity

Algorithms are frequently developed in individual centres, which can lead to diagnostic or selection biases and complicate broader implementation. Since each institution serves a specific patient population, algorithms may lose diagnostic accuracy when applied to other populations. To ensure effective integration into practice, it is vital to use data from diverse sources.[10]

Data from different countries often varies significantly, especially in GIT diseases influenced by ethnicity, geography, sociocultural factors, dietary habits and genetics.[11 12] Such variations challenge the generalisability of algorithms trained on one country's data, reducing accuracy in new regions. For instance, Korean patients with intestinal metaplasia of duodenal ulcers show a higher prevalence than patients in Colombia, the USA or South Africa. Similarly, intestinal metaplasia in the gastric corpus is more common among Korean populations.[14] These disparities highlight the importance of accounting for differences in data collection and patient characteristics across countries to ensure algorithm validity and reliability.[15]

Another GIT disease with regional variation is gastro-oesophageal reflux disease (GORD). GORD prevalence is the highest in North America and Europe, with limited data on its occurrence in the Indian subcontinent, Africa, South America and the Middle East. Studies indicate that African Americans and Asians have a lower risk of severe GORD complications like Barrett's oesophagus. While GORD pathophysiology differs among populations, factors observed in Western populations are also present in Asians to a lesser extent.[16]

Geographical variations also influence Crohn's disease presentation. In India, colonic involvement is more frequent, and fistulisation is less common. In Pakistan, perianal or fistulising disease and extraintestinal manifestations are rare compared with Western data.[17] Additionally, inflammation in Crohn's disease is more prevalent in African Americans, Hispanics and Asians than in white patients.[12]

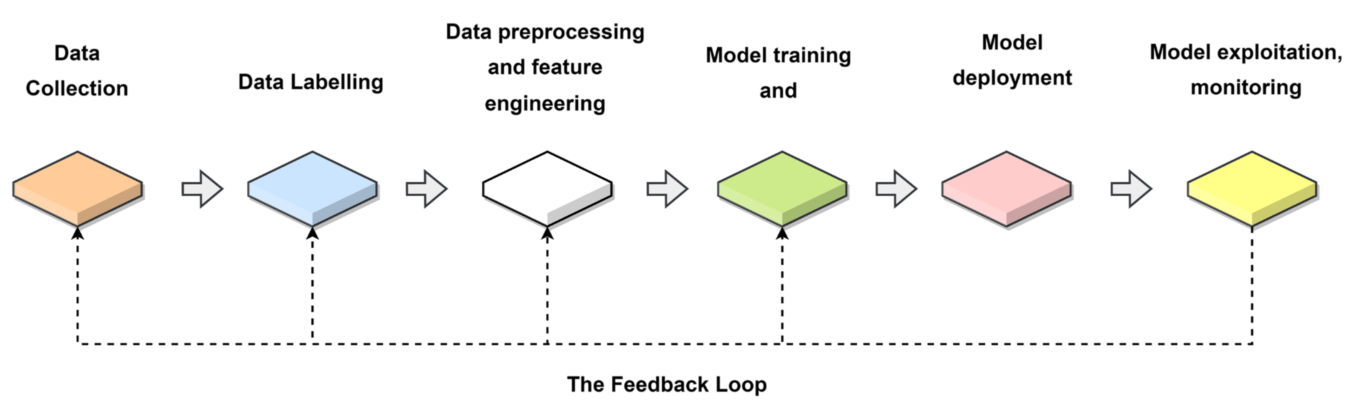

Figure 1 Machine learning lifecycle with feedback loop (high-level view).

In model training, diverse datasets that reflect demographic and socio-economic variations or create customised models for specific patient cohorts must be incorporated. Individual patient characteristics significantly influence disease outcomes and prognosis. For instance, the Forrest classification for peptic ulcers assesses recurrent bleeding risk but overlooks patient-specific factors. Factors such as age, smoking, *Helicobacter pylori* presence, co-morbidities (eg, type 2 diabetes), co-therapy with corticosteroids or anticoagulants, use of aspirin or non-aspirin antiplatelet treatments (clopidogrel or ticlopidine), previous history of dyspepsia, peptic ulcers, or ulcer bleeding, as well as alcohol consumption and use of non-steroidal anti-inflammatory drugs, can significantly increase the risk of peptic ulcer bleeding. Highlighting the necessity of incorporating clinical data into diagnostic algorithms and conducting additional investigations.[18–21]

Quality of training data and limitations during endoscopic examination

ML-based tools rely on diagnoses or image labels made by humans for training, thus introducing bias from varying classification criteria and experiences. Final diagnoses can differ among endoscopists in complex cases. Relying on a single endoscopist is particularly problematic, underscoring the need to involve multiple experts in dataset creation.[10] Even experienced endoscopists can have differing opinions. This underscores the importance of consensus processes, such as collaborative decision-making, for reliable ML-assisted gastroscopy.

Gastroscopy, combined with pathology, is a key method for diagnosing upper GIT diseases. However, increased use has raised issues like incomplete examinations and misdiagnoses. ML offers solutions by standardising procedures and aiding diagnosis.[22] ML applications improve examination quality by reducing blind spots and enhancing detection. For example, an ML model by Choi *et al* identified eight gastrointestinal regions with 97.58% accuracy and 89.20% completeness.[22]

The second problem is that most studies are retrospective, which introduces biases and limitations. Although such studies aid in developing ML models, they are prone to biases, such as selection bias, which can overstate results.[23] Historical data may not adequately represent diverse patient populations seen in clinical practice, and spectrum bias—evaluating a diagnostic test on an unrepresentative group—is a key challenge in ML for endoscopy.[10 24] Additionally, retrospective studies fail to address strategies for managing uninterpretable or low-quality images, common in real-world clinical settings.[10 23]

Another major limitation is data availability. ML models often fail to generalise well when applied to cases significantly different from their training data. Rare conditions are frequently under-represented, reducing accuracy in real-world scenarios. Gathering enough images for each pathology is particularly challenging. For example, small intestine diseases are rare, limiting data for algorithm development and complicating rare disease diagnosis.[25] If algorithms are trained on early-stage ailments such as adenomas, their accuracy in detecting advanced stages remains uncertain.[15] Less than 1% of diminutive adenomas (<6 mm) exhibit high-grade dysplasia.[26] Rare diseases, such as colonic lymphangioma, are more common in children but occasionally in adults or benign gastric tumours, such as gastric lipomas (seen in only 3% of cases), further underscoring data challenges.[27 28] Training must cover all possible scenarios to improve recognition and treatment.

Large high-quality annotated image datasets for effective machine learning training

The use of ML in gastroscopy requires a large dataset of high-quality annotated images for effective training and deployment. Here are the key points regarding the need for over one thousand images for ML-assisted gastroscopy:

Data quantity: Training robust and generalisable ML models for gastroscopy typically requires a large dataset, often in the range of thousands to tens of thousands of annotated images. This is necessary to capture the diverse range of normal anatomy, pathologies and imaging conditions encountered in real-world gastroscopy.[29 30]

Annotation quality: The images used for training must be carefully annotated by expert gastroenterologists to provide accurate ground truth labels. This allows the ML model to learn the visual patterns associated with different diseases and conditions.[10 30]

Diversity: The training dataset should include a diverse set of images representing different patient demographics, disease stages, imaging perspectives and other variables to ensure that the ML model can perform well across a wide range of real-world scenarios.[29 31]

Validation and testing: In addition to the training dataset, separate validation and test datasets of over one thousand images each are typically required to properly evaluate the ML model's performance and generalisation.[10 30]

Iterative refinement: As the ML model is deployed and used in clinical practice, the dataset can be iteratively expanded and refined with new annotated images to further improve the model's capabilities over time.[29 31]

In summary, building an effective ML system for gastroscopy requires a large, high-quality dataset of over one thousand annotated images to train, validate and continuously improve the model's performance to meet the demands of real-world clinical practice.

Continuous and effective training dataset annotation

High annotation speed is critical due to the large dataset volume and the need for diversity to avoid overfitting during model training.[32] Maintaining quality requires experienced endoscopy specialists, who often lack time to annotate entire datasets, reducing the available pool of annotators and slowing annotation speed.

Labelling tools, such as Label Studio, can enhance annotation efficiency for ML-assisted endoscopy. It supports video annotation with object tracking that interpolates bounding boxes between keyframes, significantly accelerating the process.[33] ML integrations enable 'pre-labelling' via model predictions, allowing annotators to review and refine suggested annotations.[33] Active learning prioritises complex or ambiguous samples for human annotation, focusing efforts on valuable data. The platform's configurability allows custom workflows tailored to project needs, streamlining annotation. Integration with cloud storage (eg, AWS S3 and Google Cloud Storage) facilitates direct cloud-based annotation, improving efficiency for large datasets. In summary, Label Studio's features—video annotation, ML integration, active learning, custom workflows and cloud compatibility—boost annotation speed for ML-assisted endoscopy.[33]

The segment anything model (SAM) is another tool that can greatly accelerate segmentation in the gastroscopy domain. SAM is a general image segmentation foundation model that suggests masks for human refinement, potentially speeding annotation.[34 35] However, medical images differ from natural images due to complex anatomy and uncertain boundaries, limiting SAM's direct applicability.

With adaptation and fine-tuning, SAM can enhance medical image annotation. For instance, guiding segmentation with manual hints, such as bounding boxes, is more effective than its default 'everything' mode. SAM can also identify ambiguous regions for focused human annotation, employing an 'active learning' strategy to optimise limited resources.

Additionally, SAM could expedite building large, annotated datasets critical for training robust gastroscopy ML models, especially for video data. While direct use in gastroscopy is challenging, careful integration of SAM into workflows can significantly improve annotation efficiency.

Reflections of internal tissues and false results

The problem of reflections of internal tissues during the ML segmentation model for gastroscopy images is a common challenge in the development of deep learning (DL)-based systems for gastroscopy-assisted diagnosis. Reflections, also known as artefacts, can occur because of the interaction between the endoscope and the internal tissues of the stomach, leading to distorted or unclear images. This can negatively affect the accuracy of ML models in detecting and segmenting lesions and anatomical structures. To address this issue, researchers have employed the following techniques:

Data augmentation: This involves artificially increasing the size of the training dataset by applying random transformations to the images, such as flipping, rotation and zooming. This helps the model become more robust to variations in image quality and artefacts.[36 37]

Image preprocessing: Techniques such as contrast-limited adaptive histogram equalisation (CLAHE) can be used to enhance image contrast and visibility, reducing the effect of artefacts.[36]

Network architectures: The use of advanced network architectures, such as transformers, can help the model better handle the complexities of medical images and artefacts.[37]

Object detection and segmentation algorithms: Algorithms, such as You Only Live Once and U-Net, have been shown to be effective in detecting and segmenting lesions and anatomical structures in gastroscopic images, even in the presence of artefacts.[37]

Artefact detection: Some models are designed to specifically detect and remove artefacts from the images before performing segmentation or detection tasks.[37]

By applying these techniques, ML models enhance performance in gastroscopy diagnosis despite reflections and artefacts. False results from endoscopic tools can compromise ML-assisted gastroscopy accuracy. High-quality training data are vital for reliable medical diagnosis, making precision essential for effective ML algorithms. Validation plays a crucial role in confirming model efficacy. Training data quality profoundly affects ML performance; ensuring accuracy and minimising tool-induced artefacts is critical for model reliability. Addressing inaccuracies in clinical contexts ensures dependable ML-assisted diagnostic systems, optimising patient care with actionable outcomes. Thus, rigorous data quality and validation are fundamental to developing robust ML-supported endoscopic tools.

RESOLVING MODEL TRAINING TRADE-OFFS

In endoscopy, as in other fields, ML model training involves balancing trade-offs. One side is prediction accuracy, whereas the other includes performance, costs and research time. There are no universal solutions; each case depends on the problem's context. This section outlines methods for ML engineers to manage model properties, focusing on performance and cost vs accuracy.

Performance and accuracy trade-off

Several techniques are used to optimise the performance of ML models in gastroscopy applications, reducing computational load and preserving diagnostic performance. These approaches suit real-time applications in gastroscopy and can involve uniform, non-uniform or knowledge distillation-based methods.[22 36] To balance the speed and accuracy of ML models in gastroscopy, researchers use various strategies:

Dropout: Dropout randomly deactivates neurons during training to curb overfitting, boosting generalisation and resilience to image-quality variations. By forcing reliance on multiple neurons, dropout augments performance on unseen data.[22 36]

Model pruning and quantisation: Pruning removes redundant neurons or connections, shrinking network size without severely harming accuracy. Quantisation lowers numerical precision from floating-point to lower-bit integers, cutting memory usage and computational overhead. These techniques reduce the model's size and computational requirements, making it faster and more efficient while maintaining its diagnostic capabilities.[22 36]

Knowledge distillation: Knowledge distillation trains a smaller model to mimic the behaviour of a larger, more accurate model. This approach can reduce the computational requirements while maintaining the accuracy of the model.[22 36]

TL: TL uses pre-trained models and fine-tuning them for specific tasks. It leverages the knowledge gained from larger datasets and improves the performance of the model on the specific task of gastroscopy.[22 36]

Data augmentation: Data augmentation involves artificially increasing the size of the training dataset by applying random transformations to the images. This approach can improve the robustness of the model to variations in image quality and artefacts, enhancing its performance in real-world scenarios.[22 36]

Real-time semantic segmentation: Real-time semantic segmentation involves training models to segment lesions and anatomical structures in real-time, allowing for faster and more accurate diagnoses during gastroscopy procedures.[36] The bilateral segmentation network combined with TL, CLAHE and data augmentation can provide high sensitivity and good accuracy for gastric intestinal metaplasia detection and segmentation.[36]

By combining these techniques, ML models can be tuned to balance speed and accuracy, making them more suitable for real-time applications in gastroscopy.

Cost and accuracy trade-off

ML in gastroscopy requires large, high-quality annotated datasets for effective training. However, the cost of creating such a dataset can be significant. A potential solution is to utilise junior endoscopists whose tariffs are acceptable; however, the quality may not be as precise. Here are the key points regarding the trade-off between cost and precision in ML-assisted gastroscopy:

Junior endoscopists: High-quality images are essential yet expensive. Junior endoscopists can generate data at lower expense; however, their annotations may be less comprehensive. Training on junior-generated data may diminish model accuracy and reliability.

Training data: ML models need vast annotated datasets for robust performance, often incurring substantial costs.

Balancing cost and precision: Combining junior and senior endoscopists or using cost-effective annotation methods can help.

Involving expert and junior endoscopists in annotation quantifies experience levels and refines ML models. For example, the ML-Endo model, trained on expert endoscopic submucosal dissection cases, shows high performance on validation data and can be applied to junior endoscopists with various skill levels.[38]

NON-ALGORITHMIC CHALLENGES

The challenges of creating ML models go beyond data quality, preprocessing and robustness. Even at the training phase, ensuring sufficient technical infrastructure can be problematic. Some endoscopy tasks, such as real-time image classification or segmentation, may exceed desktop or mobile capabilities. When deploying models, interpretability issues arise; medical staff must grasp how outcomes relate causally to the original data. Finally, ethical and legal concerns affect both (a) patients, who may object to or misunderstand data processing and storage, and (b) medical staff, ML engineers and others with access to the data. The involvement of many parties can undermine confidentiality, fostering distrust among patients who provide health data.

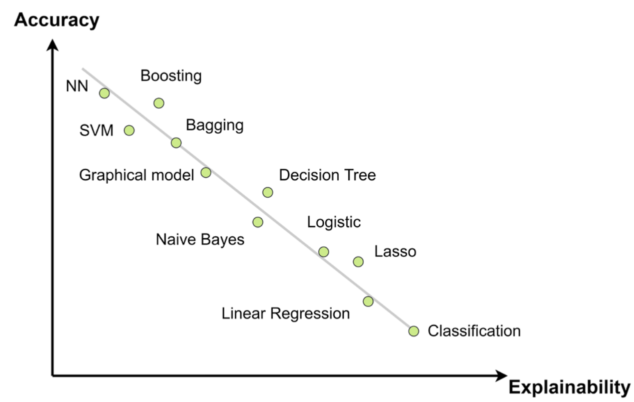

Figure 2 Trade-off between model accuracy and explainability in machine learning. NN, neural network; SVM, support vector machine

Machine learning model explainability

Explainability is crucial for ML adoption in gastroscopy: clinicians must understand predictions to trust and use them effectively.[39 40] 'Black box' models hamper clinical deployment.[39–41] SHapley Additive exPlanations (SHAP) and saliency maps reveal key features in gastroscopy images, enhancing ML transparency[39 40 42] (figure 2). A study used hybrid CNN and SHAP for interpretable gastrointestinal cancer classification.[41 42] Saliency maps can also be integrated into model training (eg, image concatenation).[43–45]

Because lesions vary in shape and size, eXplainable artificial intelligence (XAI) is vital for uncovering complex visual patterns. Incorporating XAI fosters clinician trust and adoption; however, explanation alone does not guarantee correctness; physicians must make final decisions. In summary, XAI (eg, attention maps) is crucial for ML acceptance and clinical deployment in gastroscopy.

Technical challenges in machine learning model real-time usage

ML and neural networks, especially DL, show promise in real-time endoscopy for detecting gastrointestinal lesions, predicting disease outcomes and supporting real-time diagnosis.[46–50] However, a major challenge is the dependence on large, high-quality training data. DL models require extensive, well-annotated datasets.[33] Acquiring such data is difficult due to privacy concerns

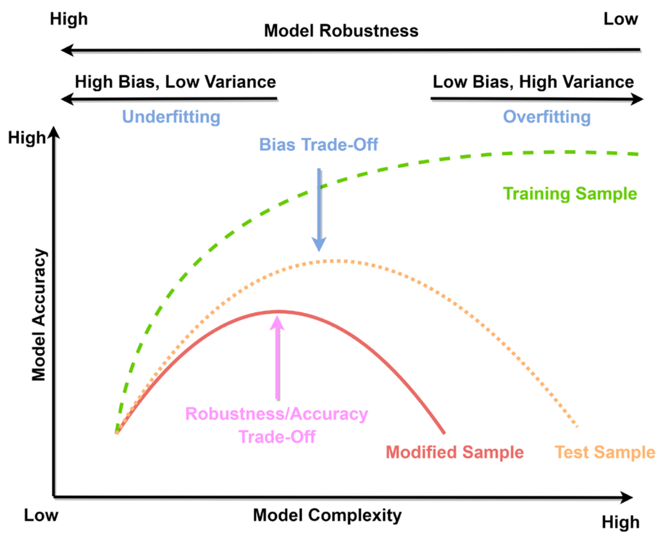

Figure 3 Balancing model complexity and accuracy in machine learning

and the labour-intensive nature of manual annotation. Variability in endoscopic procedures and image quality can also cause data inconsistencies, affecting model performance.[34 51]

DL models risk overfitting,[33] so those trained on one population or equipment may not generalise to others (figure 3). This limitation requires extensive validation for each new setting.[52] They also often function as 'black boxes', making decision processes hard to interpret.[36] This lack of transparency can hinder clinical adoption, as clinicians may be reluctant to trust tools without insight into how conclusions are reached.[33 35]

Integrating these technologies into clinical workflows is challenging, requiring careful design of user interfaces, decision-support timing and minimal disruption to procedures. Widespread adoption demands tools that complement, not hinder, clinical work.[38] Current ML interfaces are designed for research, not routine clinical use, necessitating additional equipment. Adopting ML also requires extra training for endoscopists and staff to handle false positives, especially amid visual challenges, such as inadequate bowel prep or excessive bubbles.[7]

ETHICAL AND LEGAL CONCERNS

The integration of ML in endoscopic examinations raises ethical and legal concerns. Large patient datasets (images, videos and clinical information) are required to train ML models, making data protection and privacy crucial. Endoscopic data storage, sharing and access must be secure and anonymised, especially when building cross-institutional databases or collecting rare disease data.[13] In accordance with the General Data Protection Regulation (GDPR) the following requirements must be met:

- Researchers must obtain a valid legal basis, often explicit, informed consent for processing patient data for ML training. Patients must be clearly informed about how their endoscopic data will be used in ML training and who will have access to it. Transparent communication of benefits, risks and limitations is necessary for informed consent. Clinicians must decide how and when to educate patients about ML complexities, including algorithms, data inputs, biases and limitations. The processing activities must be documented, and patients must be able to access and amend their data. Patients retain the right to access, rectify, erase and restrict processing of their data.[53]
- Only data that are adequate, relevant and necessary should be collected for training the ML model. Data cannot be used for purposes other than those explicitly described in the consent, unless additional consent is obtained. Whenever possible, data should be anonymised. Data must be securely stored and transferred to avoid unauthorised access. Contracts with vendors or third parties must clearly outline data usage, security obligations and explicitly prohibit unauthorised reidentification and repurposing of data.[53]

In addition to the GDPR requirements and broader ethical considerations described above, recent international consensus guidelines on the clinical application of AI in gastrointestinal endoscopy[54] provide practical recommendations for safe and effective implementation. These guidelines emphasise the necessity of rigorous preclinical and clinical validation in diverse patient populations, transparent performance reporting and continuous post-deployment monitoring to ensure sustained safety and accuracy. They also highlight the importance of clinician oversight, clear delineation of responsibilities and training for endoscopists in the interpretation of AI outputs.[54]

ML raises liability questions when diagnostic errors occur. Clear legal frameworks should define responsibility for suboptimal decisions resulting from inaccurate algorithms. Doctors, healthcare systems and device manufacturers share accountability, influencing compensation, adoption and potential hesitancy by developers. All parties should understand their obligations when deploying ML in endoscopy.[47 55] Addressing these concerns is vital for responsible ML adoption, adhering to ethical principles, regulatory guidelines, transparency and a patient-centric approach.

ML integration also affects professional roles. Adequate training is essential for clinicians to use ML effectively. Medical schools should incorporate ML and data science into curricula to prepare future providers.[56 57] Finally, governmental and professional approval is key for ML adoption. Some fear it may replace traditional methods, so each country must resolve legal questions before broad implementation. Gastroenterological organisations should establish guidelines for ML application.[7]

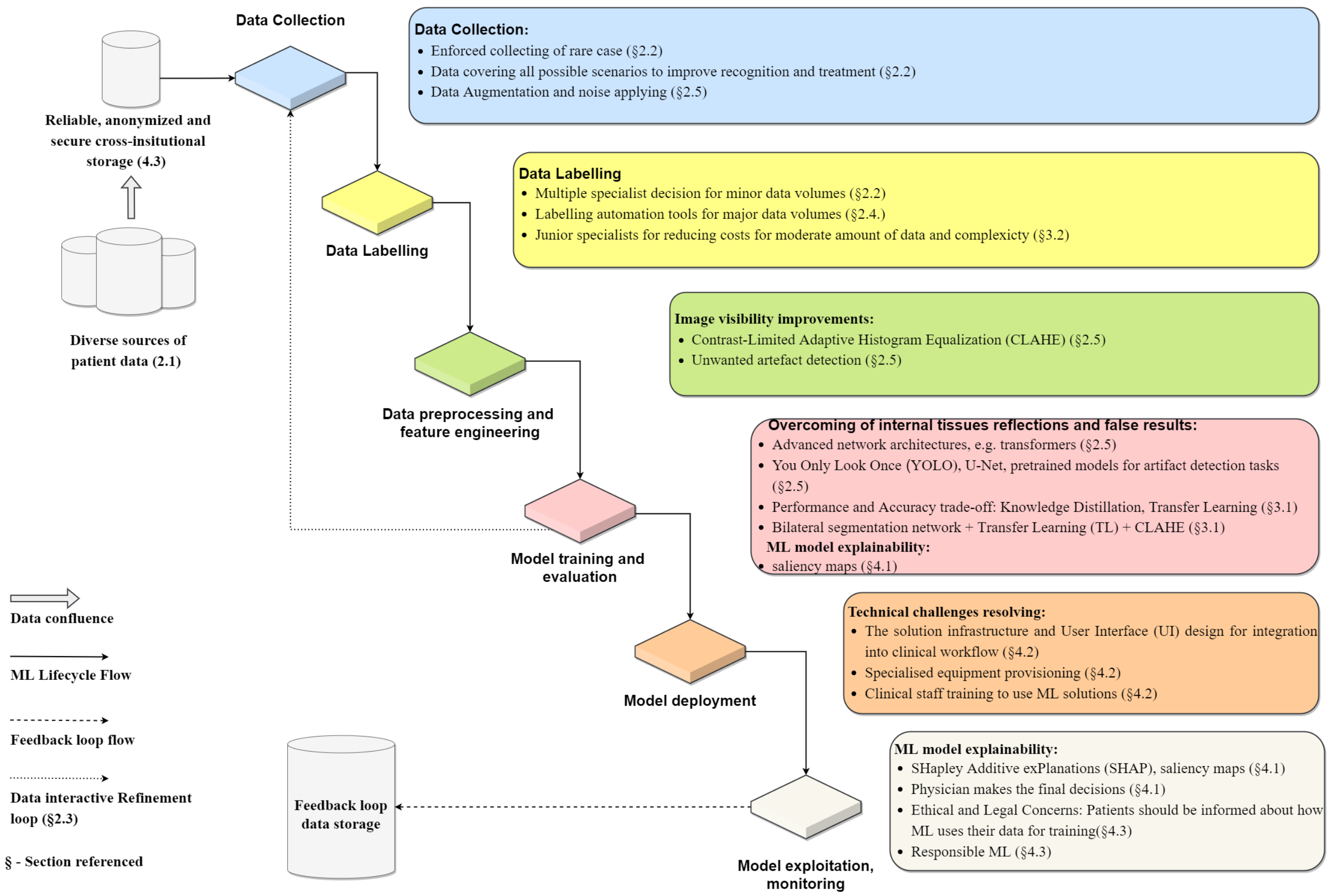

Figure 4 Proposed enhancements of machine learning lifecycle steps in gastrointestinal endoscopy. ML, machine learning

OVERVIEW

Summarising the methodological, ethical and legal considerations discussed above, table 1 systematises key ML application areas in gastrointestinal endoscopy, indicating data types, employed models and clinical scenarios.

MACHINE LEARNING LIFECYCLE IN GASTROINTESTINAL ENDOSCOPY

Throughout this review, we have considered numerous specialised solutions aimed at addressing distinct challenges within various stages of the ML lifecycle in gastrointestinal endoscopy. These individual improvements, while effective on their own, are even more impactful when integrated into a cohesive framework.

figure 4 illustrates our proposed enhancements to the ML lifecycle, tailored specifically for gastrointestinal endoscopy. This visualisation consolidates various isolated improvements into an integrated approach, designed to optimise diagnostic accuracy, enhance workflow efficiency and address key technical and ethical limitations inherent in this domain.

CONCLUSION

The integration of ML in endoscopy offers substantial promise for improving diagnostic accuracy, streamlining procedures and enhancing patient outcomes. However, challenges remain, notably the scarcity of high-quality, diverse datasets needed to train algorithms, along with ethical concerns over privacy, bias and security. Technical hurdles such as real-time data processing and the interpretability of ML outputs add further complexity. Balancing innovation with regulatory standards also complicates widespread adoption.

Overcoming these obstacles requires collaborative efforts among healthcare professionals, researchers and technology developers. Standardised data collection protocols, transparent algorithmic decision-making and strong interdisciplinary communication are key to responsibly advancing ML in endoscopy. By proactively addressing these challenges, ML can become a powerful tool to complement healthcare professionals, ultimately leading to more accurate diagnoses and better patient care.

It is important to acknowledge the parallel development of large language models (LLMs) in gastroenterology, which represent a specialised type of ML that processes and generates human-like text. Recent examples include GastroGPT, a proof-of-concept specialty-specific clinical LLM designed for gastroenterology tasks that demonstrated superior performance compared with general-purpose LLMs in clinical task evaluation.[58] LLM applications fall outside the scope of our current analysis focusing on computer vision and image analysis in endoscopy; however, these text-based AI systems represent a complementary approach that may eventually converge with image-based ML systems to provide comprehensive clinical decision support.

In this comprehensive review of modern methods aimed at overcoming the challenges associated with applying ML in endoscopy and the opportunities that arise in this regard, we identified key limitations of the existing approaches. We developed an enhanced ML lifecycle tailored specifically to address the domain-specific requirements of tasks in gastrointestinal endoscopy. It provides an initial framework that can be used to guide the development of new ML systems in gastrointestinal endoscopy.

Contributors SL is the first author and was responsible for writing and editing the medical section of the original draft, as well as image preparation. JA contributed to writing the medical section of the original draft and editing. AKR also contributed to writing the medical section of the original draft and editing. The manuscript was critically revised by EK and checked for completeness. JČ and DB were responsible for writing the technical section of the original draft. All authors reviewed and approved the final version of the manuscript. We used OpenAI ChatGPT only for language and document-preparation tasks. It helped with grammar, spelling, clarity and tone, consistent use of terms/abbreviations, light formatting (headings, tables and figure captions) and checking citation callouts. No AI was used for study design, data collection/analysis, figure creation, or interpretation. No patient-identifiable data were shared. All AI suggestions were manually reviewed by the authors, who take full responsibility for the final text.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iD

Sergejs Lobanovs http://orcid.org/0009-0007-9896-0817

REFERENCES

1. Wu J, Chen J, Cai J. Application of Artificial Intelligence in Gastrointestinal Endoscopy. *J Clin Gastroenterol* 2021;55:110–20.
2. Kröner PT, Engels MM, Glicksberg BS, *et al*. Artificial intelligence in gastroenterology: A state-of-the-art review. *World J Gastroenterol* 2021;27:6794–824.
3. El Hajjar A, Rey JF. Artificial intelligence in gastrointestinal endoscopy: general overview. *Chin Med J (Engl)* 2020;133:326–34.
4. Parasher G, Wong M, Rawat M. Evolving role of artificial intelligence in gastrointestinal endoscopy. *World J Gastroenterol* 2020;26:7287–98.
5. Barua I, Vinsard DG, Jodal HC, *et al*. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. *Endoscopy* 2021;53:277–84.
6. Kim DS, Jung Y, Rhee HS, *et al*. Usefulness of the Forrest Classification to Predict Artificial Ulcer Rebleeding during Second-Look Endoscopy after Endoscopic Submucosal Dissection. *Clin Endosc* 2016;49:273–81.
7. Ichimasa K, Kudo S-E, Mori Y, *et al*. Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer. *Endoscopy* 2018;50:230–40.
8. Eickhoff A, van Dam J, Jakobs R, *et al*. Computer-assisted colonoscopy (the NeoGuide Endoscopy System): results of the first human clinical trial ("PACE study"). *Am J Gastroenterol* 2007;102:261–6.
9. Feng Q, May MT, Ingle S, *et al*. Prognostic Models for Predicting Overall Survival in Patients with Primary Gastric Cancer: A Systematic Review. *Biomed Res Int* 2019;2019:5634598.
10. Abadir AP, Ali MF, Karnes W, *et al*. Artificial Intelligence in Gastrointestinal Endoscopy. *Clin Endosc* 2020;53:132–41.
11. Bhopal RS, Cezard G, Bansal N, *et al*. Ethnic variations in five lower gastrointestinal diseases: Scottish health and ethnicity linkage study. *BMJ Open* 2014;4:e006120.
12. Barnes EL, Loftus EV Jr, Kappelman MD. Effects of Race and Ethnicity on Diagnosis and Management of Inflammatory Bowel Diseases. *Gastroenterology* 2021;160:677–89.
13. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 2021;22:122.
14. Gutierrez O, Kim JG, *et al*. Geographic differences in the distribution of intestinal metaplasia in duodenal ulcer patients. *Am J Gastroenterol* 2001;96:666–72.
15. Hann A, Troya J, Fitting D. Current status and limitations of artificial intelligence in colonoscopy. *United European Gastroenterol J* 2021;9:527–33.
16. Sharma P, Wani S, Romero Y, *et al*. Racial and geographic issues in gastroesophageal reflux disease. *Am J Gastroenterol* 2008;103:2669–80.
17. Menon R, Riera A, Ahmad A. A global perspective on gastrointestinal diseases. *Gastroenterol Clin North Am* 2011;40:427–39.
18. Zullo A, Hassan C, Campo SMA, *et al*. Bleeding peptic ulcer in the elderly: risk factors and prevention strategies. *Drugs Aging* 2007;24:815–28.
19. Kang JM, Kim N, Lee BH, *et al*. Risk factors for peptic ulcer bleeding in terms of Helicobacter pylori, NSAIDs, and antiplatelet agents. *Scand J Gastroenterol* 2011;46:1295–301.
20. Peng Y-L, Leu H-B, Luo J-C, *et al*. Diabetes is an independent risk factor for peptic ulcer bleeding: a nationwide population-based cohort study. *J Gastroenterol Hepatol* 2013;28:1295–9.
21. Andersen IB, Jørgensen T, Bonnevie O, *et al*. Smoking and alcohol intake as risk factors for bleeding and perforated peptic ulcers: a population-based cohort study. *Epidemiology (Sunnyvale)* 2000;11:434–9.
22. Chen H, Liu S, Huang S, *et al*. Applications of artificial intelligence in gastroscopy: a narrative review. *J Int Med Res* 2024;52:3000605231223454.
23. Mori Y, Kudo S, Mohmed HEN, *et al*. Artificial intelligence and upper gastrointestinal endoscopy: Current status and future perspective. *Dig Endosc* 2019;31:378–88.
24. Mascarenhas M, Afonso J, Andrade P, *et al*. Artificial intelligence and capsule endoscopy: unravelling the future. *Ann Gastroenterol* 2021;34:300–9.
25. Namikawa K, Hirasawa T, Yoshio T, *et al*. Utilizing artificial intelligence in endoscopy: a clinician's guide. *Expert Rev Gastroenterol Hepatol* 2020;14:689–706.
26. Vleugels JLA, Hassan C, Senore C, *et al*. Diminutive Polyps With Advanced Histologic Features Do Not Increase Risk for Metachronous Advanced Colon Neoplasia. *Gastroenterology* 2019;156:623–34.
27. Zeng X-Y, Ding Y, Lu Q-Q, *et al*. A Rare Case of Colonic Lymphangioma in an Adult Woman. *Indian J Surg* 2023;85:664–6.
28. Sullivan IW, Hota P, Dass C. Gastric lipomas: a case series and review of a rare tumor. *BJR Case Rep* 2019;5:20180109.
29. Shin K, Lee JS, Lee JY, *et al*. An Image Turing Test on Realistic Gastroscopy Images Generated by Using the Progressive Growing of Generative Adversarial Networks. *J Digit Imaging* 2023;36:1760–9.
30. Buendgens L, Cifci D, Ghaffari Laleh N, *et al*. Weakly supervised end-to-end artificial intelligence in gastrointestinal endoscopy. *Sci Rep* 2022;12:4829.
31. Guo F, Meng H. Application of artificial intelligence in gastrointestinal endoscopy. Arab journal of gastroenterology: the official publication of the Pan. *Arab Association of Gastroenterology* 2024;25.
32. Dietterich TG. Overfitting and undercomputing in machine learning. *ACM Comput Surv* 1995;27:326–7.
33. Tkachenko M, Malyuk M, Holmanyuk A, *et al*. Label studio: data labeling software. 2024. Available: https://github.com/HumanSignal/label

34 Zhang Y, Shen Z, Jiao R. Segment anything model for medical image segmentation: Current applications and future directions. *Comput Biol Med* 2024;171:108238.
35 Zhang C, Cui Y, Lin W, *et al*. Segment anything for videos: a systematic survey. *Preprint arXiv* [Preprint] 2024.
36 Siripoppohn V, Pittayanon R, Tiankanon K, *et al*. Real-time semantic segmentation of gastric intestinal metaplasia using a deep learning approach. *Clin Endosc* 2022;55:390–400.
37 Jin Z, Gan T, Wang P, *et al*. Deep learning for gastroscopic images: computer-aided techniques for clinicians. *Biomed Eng Online* 2022;21:12.
38 Cao J, Yip H-C, Chen Y, *et al*. Intelligent surgical workflow recognition for endoscopic submucosal dissection with real-time animal study. *Nat Commun* 2023;14:6676.
39 Chaddad A, Peng J, Xu J, *et al*. Survey of Explainable AI Techniques in Healthcare. *Sensors (Basel)* 2023;23:634.
40 Bhati D, Neha F, Amiruzzaman M. A Survey on Explainable Artificial Intelligence (XAI) Techniques for Visualizing Deep Learning Models in Medical Imaging. *J Imaging* 2024;10:239.
41 Gipiškis R, Tsai C W, Kurasova O. Explainable AI (XAI) in Image Segmentation in Medicine, Industry, and Beyond: A Survey. 2024. Available: https://arxiv.org/abs/2405.01636
42 Bin Moin M, Tuj Johora Faria F, Saha S, *et al*. Exploring explainable ai techniques for improved interpretability in lung and colon cancer classification. 2024. Available: https://arxiv.org/abs/2405.04610
43 Sobhaninia Z, Abharian N, Karimi N, *et al*. n.d. Endoscopy Classification Model Using Swin Transformer and Saliency Map.
44 Chen H, Wang S, Ding Y, *et al*. Saliency-Based Bleeding Localization for Wireless Capsule Endoscopy Diagnosis. *Int J Biomed Imaging* 2017;2017:1–8.
45 Khan MA, Sahar N, Khan WZ, *et al*. GestroNet: A Framework of Saliency Estimation and Optimal Deep Learning Features Based Gastrointestinal Diseases Detection and Classification. *Diagnostics (Basel)* 2022;12:2718.
46 Min JK, Kwak MS, Cha JM. Overview of Deep Learning in Gastrointestinal Endoscopy. *Gut Liver* 2019;13:388–93.
47 Yamada M, Saito Y, Imaoka H, *et al*. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci Rep* 2019;9:14465.
48 Nie Z, Xu M, Wang Z, *et al*. A Review of Application of Deep Learning in Endoscopic Image Processing. *J Imaging* 2024;10:275.
49 Kellerman R, Bleiweiss A, Samuel S, *et al*. P269 Spaciotemporal machine learning analysis of complete small bowel capsule endoscopy videos for prediction of outcomes in Crohn's disease. *Journal of Crohn's and Colitis* 2022;16:i308–9.
50 van der Sommen F, de Groof J, Struyvenberg M, *et al*. Machine learning in GI endoscopy: practical guidance in how to interpret a novel field. *Gut* 2020;69:2035–45.
51 Bektaş M, Burchell GL, Bonjer HJ, *et al*. Machine learning applications in upper gastrointestinal cancer surgery: a systematic review. *Surg Endosc* 2023;37:75–89.
52 Ishioka M, Hirasawa T, Tada T. Detecting gastric cancer from video images using convolutional neural networks. *Dig Endosc* 2019;31:e34–5.
53 European Union. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 (general data protection regulation). Official Journal of the European Union; 2016. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj
54 Force AAT, Parasa S, Berzin T, *et al*. Consensus statements on the current landscape of artificial intelligence applications in endoscopy, addressing roadblocks, and advancing artificial intelligence in gastroenterology. *Gastrointest Endosc* 2025;101.
55 Maliha G, Gerke S, Cohen IG, *et al*. Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation. *Milbank Q* 2021;99:629–47.
56 Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ Digit Med* 2018;1:54.
57 Sapci AH, Sapci HA. Artificial Intelligence Education and Tools for Medical and Health Informatics Students: Systematic Review. *JMIR Med Educ* 2020;6:e19285.
58 Simsek C, Ucdal M, de-Madaria E, *et al*. GastroGPT: Development and controlled testing of a proof-of-concept customized clinical language model. *Endosc Int Open* 2025;13:a26372163.
59 Kagawa Y. Clinical implications of a machine learning model predicting colorectal polyp recurrence after endoscopic mucosal resection. *World J Gastroenterol* 2025;31:107197.
60 Glissen Brown JR, Mansour NM, Wang P, *et al*. Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial). *Clin Gastroenterol Hepatol* 2022;20:1499–507.
61 Park DK, Kim EJ, Im JP, *et al*. A prospective multicenter randomized controlled trial on artificial intelligence assisted colonoscopy for enhanced polyp detection. *Sci Rep* 2024;14:25453.
62 Lau LHS, Ho JCL, Lai JCT, *et al*. Effect of Real-Time Computer-Aided Polyp Detection System (ENDO-AID) on Adenoma Detection in Endoscopists-in-Training: A Randomized Trial. *Clin Gastroenterol Hepatol* 2024;22:630–41.
63 Feng L, Xu J, Ji X, *et al*. Development and validation of a three-dimensional deep learning-based system for assessing bowel preparation on colonoscopy video. *Front Med* 2023;10:1296249.
64 Gelu-Simeon M, Mamou A, Saint-Georges G, *et al*. Deep learning model applied to real-time delineation of colorectal polyps. *BMC Med Inform Decis Mak* 2025;25:206.
65 Boros E, Pintér J, Molontay R, *et al*. New machine-learning models outperform conventional risk assessment tools in Gastrointestinal bleeding. *Sci Rep* 2025;15:6371.
66 Fockens KN, Jong MR, Jukema JB, *et al*. Barrett's Oesophagus Imaging for Artificial Intelligence (BONS-AI) consortium. A deep learning system for detection of early Barrett's neoplasia: a model development and validation study. *Lancet Digit Health* 2023;5:e905–16.
67 Shen Y, Chen A, Zhang X, *et al*. Real-Time Evaluation of Helicobacter pylori Infection by Convolution Neural Network During White-Light Endoscopy: A Prospective, Multicenter Study (With Video). *Clin Transl Gastroenterol* 2023;14:e00643.
68 Elamin S, Johri S, Rajpurkar P, *et al*. From data to artificial intelligence: evaluating the readiness of gastrointestinal endoscopy datasets. *J Can Assoc Gastroenterol* 2025;8:S81–6.
69 Aso MC, Sostres C, Lanas A. Artificial Intelligence in GI endoscopy: what to expect. *Front Med (Lausanne)* 2025;12:1588873.
70 Ramamurthy K, George TT, Shah Y, *et al*. A Novel Multi-Feature Fusion Method for Classification of Gastrointestinal Diseases Using Endoscopy Images. *Diagnostics (Basel)* 2022;12:2316.
71 Siddiqui S, Khan JA, Algamdi S. Deep ensemble learning for gastrointestinal diagnosis using endoscopic image classification. *PeerJ Comput Sci* 2025;11:e2809.
72 Spada C, Piccirelli S, Hassan C, *et al*. AI-assisted capsule endoscopy reading in suspected small bowel bleeding: a multicentre prospective study. *Lancet Digit Health* 2024;6:e345–53.
73 Jiang B, Dorosan M, Leong JWH, *et al*. Development and validation of a deep learning system for detection of small bowel pathologies in capsule endoscopy: a pilot study in a Singapore institution. *Singapore Med J* 2024;65:133–40.
74 Li L, Yang L, Zhang B, *et al*. Automated detection of small bowel lesions based on capsule endoscopy using deep learning algorithm. *Clin Res Hepatol Gastroenterol* 2024;48:102334.
75 Kuwahara T, Hara K, Mizuno N, *et al*. Artificial intelligence using deep learning analysis of endoscopic ultrasonography images for the differential diagnosis of pancreatic masses. *Endoscopy* 2023;55:140–9.
76 Yi N, Mo S, Zhang Y, *et al*. An endoscopic ultrasound-based interpretable deep learning model and nomogram for distinguishing pancreatic neuroendocrine tumors from pancreatic cancer. *Sci Rep* 2025;3383.
77 Cui H, Zhao Y, Xiong S, *et al*. Diagnosing Solid Lesions in the Pancreas With Multimodal Artificial Intelligence: A Randomized Crossover Trial. *JAMA Netw Open* 2024;7:e2422454.