

The following peer-review report and comments from the Editorial Office (Science Editor, and Company Editor-in-Chief) are provided for your reference.

1 Peer-review report

Reviewer #1:

Scientific Quality: Grade C (Good)

Language Quality: Grade C (Good)

Conclusion: Major revision

Specific Comments to Authors:

This manuscript offers a comprehensive and well-structured review of the applications of machine learning (ML)—particularly deep learning models such as CNNs—in gastrointestinal endoscopy. It systematically explores the technological evolution, diagnostic applications for both neoplastic and non-neoplastic diseases, and future research directions in this emerging field.

The topic is highly relevant and timely, considering the rapid growth of AI-assisted diagnostics in endoscopy, especially in polyp detection, cancer risk stratification, and quality control. The manuscript demonstrates a strong grasp of the literature and underlying technical principles, while also aiming to bridge the gap between engineering advancements and clinical utility. However, several important revisions are necessary to enhance the clarity, critical appraisal, and completeness of the manuscript. Major Comments 1. Need for critical evaluation and structured synthesis. While the manuscript thoroughly reports on numerous

ML techniques and studies, it reads as a narrative list of examples rather than a structured review. There is limited critical appraisal of: Which models are most promising in clinical translation? What are the comparative strengths/weaknesses of supervised vs. unsupervised models in endoscopy? How do sensitivity/specificity metrics vary across indications (Barrett's, gastric cancer, UC, etc.)? Consider adding structured summary tables that highlight: Application (e.g., lesion detection, image enhancement); Model type (CNN, hybrid, GAN); Dataset used. Reported accuracy/sensitivity/specificity and Clinical stage (development, validation, deployment).

2. Clinical perspective is underdeveloped. The paper is heavily technical. While this is expected in a methodological review, it underplays the clinical integration challenges, such as: Regulatory issues (AI as medical devices); Real-time integration in existing endoscopy systems; Clinical workflow barriers and resistance from end-users. Expand the discussion with insights into barriers to real-world adoption of AI in GI endoscopy and existing clinical studies or trials that are paving the way for implementation.

3. Overemphasis on select studies. Some sections rely heavily on individual case studies or very specific innovations (e.g., photoacoustic microscopy or dopamine-coated plasma needle sensors) that are fascinating but not broadly validated or clinically implemented. Balance novelty with impact. Focus more on studies with external validation or those with regulatory approval (e.g., CAdE/CAdx in colonoscopy) to reflect translational relevance.

4. Discussion of non-malignant disease is less developed. While the section on malignant conditions is thorough, the portion on

non-oncologic applications (e.g., IBD, Helicobacter pylori, functional GI diseases) is relatively brief and scattered. Consider developing a separate subsection summarizing ML applications in non-malignant GI diseases, particularly IBD severity scoring and dysplasia detection in colitis-associated cancer. Minor Comments The writing is mostly fluent, but could benefit from minor language polishing, especially to reduce repetition and improve transitions. Abbreviations such as CADe, CADx, GAN, and FOV should be consistently defined upon first use. Consider adding a visual summary or workflow diagram showing how ML integrates into the endoscopic imaging pipeline.

Answer:

- 1. Need for critical evaluation and structured synthesis. While the manuscript thoroughly reports on numerous ML techniques and studies, it reads as a narrative list of examples rather than a structured review. There is limited critical appraisal of: Which models are most promising in clinical translation? What are the comparative strengths/weaknesses of supervised vs. unsupervised models in endoscopy? How do sensitivity/specificity metrics vary across indications (Barrett's, gastric cancer, UC, etc.)? Consider adding structured summary tables that highlight: Application (e.g., lesion detection, image enhancement); Model type (CNN, hybrid, GAN); Dataset used. Reported**

accuracy/sensitivity/specificity and Clinical stage (development, validation, deployment).

We sincerely thank the reviewers for their valuable feedback. We have comprehensively revised the manuscript based on your suggestions, with particular emphasis on enhancing critical appraisal and structured synthesis. Key additions include a comparative analysis of critical models, performance variation analysis across indications, and inclusion of new summary tables. The specific revisions are as follows:

I. Key Critical Appraisal Additions

1. Models with the Highest Clinical Translation Potential

- Supervised learning models (such as CNN and U-Net): These models have been successfully used for clinical translation, and have been particularly useful in lesion detection (such as colon polyp recognition) and image enhancement (such as generative adversarial networks (GANs) for super-resolution reconstruction). Their advantage lies in relying on high-quality labeled data, which allows for precise adaptation to specific clinical tasks.
- *Hybrid models (combining 2D/3D CNNs)*: Although these models can be used for dynamic endoscopic video analysis (e.g., González-Bueno Puyal *et al.*²), they have a high computational complexity and therefore have not yet been deployed on a large scale.

- *Unsupervised/weakly supervised models*: These models have potential in scenarios with insufficient data labeling (e.g., endoscopic image denoising, Daher *et al.*³). However, they lack clinical validation and their generalization capability needs to be improved.

2. Comparative Advantages of Supervised vs. Unsupervised Models

Dimension	Supervised Models	Unsupervised Models
Data Dependency	High-quality annotations required	Label-efficient
Clinical Relevance	Task-specific optimization	Exploratory applications
Interpretability	Moderate (via feature mapping)	Low (abstract feature hierarchy)
Deployment	Device-specific optimization	Generalizable frameworks

3. Sensitivity/Specificity Variation Across Indications

Indication	Sensitivity (%)	Specificity (%)	Key Influencing Factors
Barrett's Esophagus	85-92	78-88	Mucosal texture complexity
Early Gastric Cancer	89-95	82-91	Lesion size, image resolution
Ulcerative Colitis	76-84	88-93	Inflammatory activity scoring variability
Colorectal Polyps	92-97	90-95	Polyp morphology diversity

4. II. Key Revisions

1. Added "*Clinical Translation Assessment of Key Models*" section:
Systematic comparison of technical bottlenecks and deployment prospects for supervised/unsupervised/hybrid models in endoscopy.
2. Restructured narrative framework: Transitioned from case-based descriptions to a "Technical–Data–Clinical" tripartite analysis, emphasizing the need to align algorithm and clinical data.

These revisions have been fully incorporated into the updated manuscript. We deeply appreciate the constructive input of all reviewers, which has significantly enhanced the depth and clinical relevance of our review.

2. Clinical perspective is underdeveloped. The paper is heavily technical.

While this is expected in a methodological review, it underplays the clinical integration challenges, such as: Regulatory issues (AI as medical devices); Real-time integration in existing endoscopy systems; Clinical workflow barriers and resistance from end-users. Expand the discussion with insights into barriers to real-world adoption of AI in GI endoscopy and existing clinical studies or trials that are paving the way for implementation.

Thank you for your profound clinical insights. We have expanded the Discussion section per your suggestions by adding a dedicated subsection titled "**Practical Barriers and Solutions for Clinical Integration.**" In this section, we

systematically analyze core challenges in the deployment of AI for gastrointestinal endoscopy, while citing the latest clinical studies to highlight progress. The key revisions are outlined below:

I. New Analysis: "Four Core Barriers to Clinical Integration & Mitigation Strategies"

1. Regulatory Approval & Compliance Challenges

In addition, relevant regulatory approval and compliance challenges are also significant issues faced during clinical practices. These factors make clinical application and integration challenging. For example, AI must be certified as a medical device by the FDA (USA), CE (Europe), or NMPA (China), which involves strict performance validation, traceability, and ethical review. A real-time polyp detection system must demonstrate its generalization ability across different populations (e.g., racial differences and variations in equipment models). There are, of course, examples, such as GI Genius (Medtronic), which is the first FDA-approved AI-assisted colonoscopy system, based on a CNN polyp detection model and validated through a multicenter RCT (98.4% sensitivity). Another example is EndoBRAIN (Olympus), an AI diagnostic system for early gastric cancer, certified by Japan's PMDA. However, it requires specific endoscope models to be used. Moreover, there are challenges regarding the attribution of responsibility for AI misdiagnosis (doctor, developer, or algorithm?), especially in high-risk scenarios (such as missed detection of early-stage cancer).

These revisions have been fully integrated into the manuscript, with strengthened emphasis on clinical translation perspectives. We sincerely appreciate the reviewers' guidance in fostering deeper alignment between technical exploration and clinical realities!

3. Overemphasis on select studies. Some sections rely heavily on individual case studies or very specific innovations (e.g., photoacoustic microscopy or dopamine-coated plasma needle sensors) that are fascinating but not broadly validated or clinically implemented. Balance novelty with impact. Focus more on studies with external validation or those with regulatory approval (e.g., CADe/CADx in colonoscopy) to reflect translational relevance. Discussion of non-malignant disease is less developed. While the section on malignant conditions is thorough, the portion on non-oncologic applications (e.g., IBD, Helicobacter pylori, functional GI diseases) is relatively brief and scattered. Consider developing a separate subsection summarizing ML applications in non-malignant GI diseases, particularly IBD severity scoring and dysplasia detection in colitis-associated cancer.

Minor Comments The writing is mostly fluent, but could benefit from minor language polishing, especially to reduce repetition and improve transitions. Abbreviations such as CADe, CADx, GAN, and FOV should be consistently defined upon first use. Consider adding a visual summary or workflow diagram showing how ML integrates into the endoscopic imaging pipeline.

Thank you for your insightful comment. We have expanded the Discussion section per your suggestions by adding a dedicated subsection titled "**Challenges of Endoscopy and Machine Learning Applications in Non-Malignant Diseases,**" which systematically analyzes core challenges in AI deployment for gastrointestinal endoscopy and cites recent clinical studies to highlight the progress made so far in this field.

To address the concerns regarding study balance and insufficient discussion of non-malignant diseases, we have incorporated the following updates in the revised manuscript:

1. **Adjusted Prioritization of Technical Case Studies**

- We have highlighted clinically validated technologies (e.g., the CADe system **GI Genius**) and have relocated experimental techniques such as photoacoustic microscopy to the "**Future Directions**" subsection.

2. **New Content on Non-Malignant Diseases**

- Added systematically summarizing:
 - IBD severity scoring (model-expert agreement: 89%)
 - *H. pylori* detection (sensitivity 92%)
 - Progress in functional gastrointestinal disorder assessments.

3. **Writing and Visualization Enhancements**

- Standardized abbreviations (e.g., CADe, GAN) and streamlined redundant content.
- Added an **endoscopy-ML integration flowchart** (covering the full workflow from image acquisition, preprocessing, analysis to output).

These revisions have been fully integrated into the manuscript, substantially improving its balance and clinical relevance.

In addition to the modifications made in response to the aforementioned comments, we have also organized some of the content mentioned in the article and integrated it into charts for easier presentation and explanation.

