



How Do AI-Assisted Diagnostic Tools Impact Clinical Decision-Making And Patient Outcomes In Acute Gastrointestinal Bleeding Cases? : A Comprehensive Systematic Review

¹ Gafrinda Kautsari, ² Luluk Aflakah, ³ Retti Nurlaili

¹ General Practitioner, Kediri Regency Regional General Hospital, Indonesia

^{2,3} Internal Medicine Consultant, Kediri Regency Regional General Hospital, Indonesia

Corresponding Email : gafrinda14@gmail.com

Article History :

Received date : 2025/11/04

Revised date : 2025/12/12

Accepted date : 2026/01/24

Published date : 2026/02/13



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).

E-ISSN :

ISSN 3048-1368



P-ISSN

ISSN 3048-1376



ABSTRACT

Introduction: Acute gastrointestinal bleeding (GIB) is a life-threatening condition requiring prompt diagnosis and intervention. Despite advancements in endoscopy, challenges like diagnostic variability and resource allocation persist. Artificial intelligence (AI)-assisted tools have shown promise in improving detection rates and workflow efficiency in gastroenterology, but their impact on clinical decision-making and patient outcomes in acute GIB remains underexplored.

Methods: This systematic review adhered to PRISMA 2020 guidelines. Eligible studies included those evaluating AI-based diagnostic tools in acute GIB cases, with outcomes such as diagnostic accuracy, clinical decision-making changes, or patient outcomes. Databases like PubMed, Sagepub, and Google Scholar were searched using Boolean MeSH keywords. Data

extraction focused on study design, AI tool characteristics, diagnostic performance, and clinical impact.

Results: Among 40 included studies, AI demonstrated high diagnostic accuracy, with sensitivities and specificities exceeding 90% in lesion detection. For instance, convolutional neural networks achieved 95.4% accuracy in identifying ulcers and hemorrhages. However, only one study reported a modest improvement in predicting endoscopic intervention needs (AUC: 0.68). While AI reduced reading time by 30% in some studies, its impact on patient outcomes (e.g., mortality, rebleeding) was rarely addressed. Most evidence came from retrospective studies or meta-analyses, with limited prospective or randomized controlled trials.

Discussion: AI enhances diagnostic accuracy and workflow efficiency but lacks robust evidence linking it to improved patient outcomes in acute GIB. Key limitations include methodological heterogeneity, scarce safety data, and a focus on non-acute settings. Prospective studies are needed to evaluate AI's real-world clinical impact.

Conclusion: AI shows potential as an adjunct tool in acute GIB management but requires further validation to confirm its clinical utility. Future research should prioritize patient-centered outcomes and standardized reporting.

Keywords: artificial intelligence, acute gastrointestinal bleeding, diagnostic accuracy, clinical decision-making, systematic review.

INTRODUCTION

Acute gastrointestinal bleeding (GIB) is a life-threatening condition that requires prompt and accurate diagnosis to guide effective clinical interventions (Veisman et al., 2022). Despite advancements in endoscopic techniques, challenges such as variability in diagnostic accuracy, missed lesions, and resource allocation persist, underscoring the need for innovative solutions (Wallace et al., 2022). Artificial intelligence (AI)-assisted diagnostic tools have emerged as promising adjuncts in gastroenterology, offering potential improvements in detection rates and workflow efficiency (Mohan et al., 2020). However, their impact on clinical decision-making and patient outcomes in acute GIB remains underexplored, with limited evidence linking AI to tangible benefits in this specific context (Veisman et al., 2022).

AI applications in gastrointestinal endoscopy have demonstrated high diagnostic accuracy, with sensitivities and specificities exceeding 90% for lesion detection (Bang et al., 2021). For instance, convolutional neural networks (CNNs) have achieved 95.4% accuracy in identifying ulcers and hemorrhages in capsule endoscopy (Mohan et al., 2020). These tools also reduce blind spots and interpretation time, as evidenced by studies where AI-assisted systems decreased reading time by 30% (Aoki et al., 2021). Such improvements suggest that AI could enhance triaging and resource use in acute GIB, though direct evidence is scarce (Veisman et al., 2022).

The integration of AI into clinical workflows has been predominantly studied in non-acute settings, such as colorectal polyp detection and chronic gastritis (Li et al., 2024; Shi et al., 2023). For example, AI systems like deep learning models have shown a 46% reduction in adenoma miss rates during colonoscopy (Jin et al., 2024). These findings highlight AI's potential to standardize diagnostics and reduce variability among endoscopists (Limdi & Farraye, 2022). However, extrapolating these benefits to acute GIB is challenging due to differences in clinical urgency and lesion characteristics (Veisman et al., 2022).

A critical gap in the literature is the lack of data on AI's impact on patient outcomes in acute

GIB, such as mortality, rebleeding rates, or length of hospital stay (Veisman et al., 2022). While a retrospective study by Veisman et al. (2022) reported an AUC of 0.68 for AI in predicting endoscopic intervention needs, no studies have linked AI use to improved survival or reduced complications. This contrasts with broader endoscopy research, where AI has been associated with workflow efficiencies, such as reduced biopsy rates and cost savings (Barua et al., 2020).

The predominance of retrospective studies and meta-analyses further limits the strength of evidence for AI in acute GIB (Luo et al., 2022). Only five randomized controlled trials (RCTs) were identified in this review, none of which focused on acute bleeding (Chen et al., 2020; Wang et al., 2019). Most studies evaluated AI as a secondary tool rather than a primary diagnostic aid, raising questions about its real-world applicability in high-stakes scenarios (Meinikheim et al., 2023). Additionally, safety data are sparse, with only one RCT reporting no difference in adverse events with AI use (Wallace et al., 2022).

Methodological heterogeneity across studies also complicates the synthesis of findings. For instance, AI performance metrics vary widely depending on the input data (e.g., endoscopic images vs. video streams) and the clinical task (e.g., lesion detection vs. bleeding risk stratification) (Bang et al., 2021; Kikuchi et al., 2024). While some models achieve high specificity (up to 99%), others report false positives that could lead to unnecessary interventions (Soffer et al., 2020). Standardized reporting frameworks are needed to enable cross-study comparisons and clinical translation (Nazarian et al., 2021).

Despite these limitations, AI's potential to augment human expertise in acute GIB is undeniable. For example, AI-assisted systems could prioritize high-risk patients for urgent endoscopy, optimize staffing resources, and reduce diagnostic delays (Veisman et al., 2022). Future research should prioritize prospective RCTs to evaluate AI's impact on decision time, treatment modifications, and patient-centered outcomes in acute settings (Lam et al., 2022). Until then, AI remains a supplemental tool rather than a transformative solution for acute GIB management.

METHODS

Protocol

The study strictly adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines to ensure methodological rigor and accuracy. This approach was chosen to enhance the precision and reliability of the conclusions drawn from the investigation.

Criteria for Eligibility

This systematic review aims to evaluate AI-assisted diagnostic tools impact clinical decision-making and patient outcomes in acute gastrointestinal bleeding cases.

Screening

We screened in papers that met these criteria:

- **Population Age:** Does the study focus exclusively on adult patients (18 years)?
- **Clinical Condition:** Does the study focus on acute (not chronic) gastrointestinal bleeding?
- **Intervention Type:** Does the study evaluate AI-based diagnostic tools or algorithms?
- **Comparison Group:** Does the study include a comparison with standard clinical assessment methods?
- **Outcomes:** Does the study report at least one of the following: diagnostic accuracy metrics, clinical decision-making changes, or patient outcomes (mortality, rebleeding rates, length of stay)?
- **Study Design:** Is the study either a randomized controlled trial, prospective cohort study, or retrospective cohort study with more than 10 patients?

- **Clinical Implementation:** Does the study include actual clinical implementation/application of the AI tools (not just technical development)?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Data extraction

- **Study Design Type:**

Identify the specific type of study design from the full text. Look in the methods section for explicit description. Possible types include:

- Randomized controlled trial
- Prospective cohort study
- Retrospective study
- Diagnostic accuracy study
- Comparative study

If multiple design elements are present, list all. If unclear, note "design not clearly specified". Prioritize the most specific design description provided by the authors.

- **AI Diagnostic Tool Characteristics:**

Extract detailed information about the AI diagnostic tool:

- Specific type of AI technology (e.g., deep learning, machine learning)
- Specific diagnostic task (e.g., lesion detection, characterization)
- Input data type (e.g., endoscopic images, video streams)
- Technical specifics of the algorithm

Look in methods and results sections. If multiple details are available, include all. If technical details are sparse, note "insufficient technical details provided".

- **Patient Population:**

Extract comprehensive details about study participants:

- Total number of patients
- Age range or mean age
- Gender distribution
- Specific medical condition/diagnosis
- Clinical setting (e.g., outpatient, hospital)

Prioritize precise numerical data. If ranges are provided, include both. If any category is incompletely reported, note "partial information available".

- **AI Diagnostic Performance Metrics:**

Extract quantitative performance metrics for the AI diagnostic tool:

- Sensitivity
- Specificity
- Accuracy
- Positive predictive value
- Negative predictive value

Locate in results section. Report exact percentages or proportions. If confidence intervals are provided, include them. If multiple metrics are reported for different conditions, list all.

- **Comparative Performance Outcomes:**

Identify comparative performance between AI and human diagnosticians:

- Performance comparison metrics
- Statistical significance of differences
- Specific performance improvements with AI assistance

Look in results and discussion sections. Prioritize statistically significant findings. If multiple comparisons exist, list all. Note the expertise level of compared human diagnosticians (e.g., junior vs senior endoscopists).

• Clinical Decision-Making Impact:

Extract evidence of AI's impact on clinical decision-making:

- Changes in diagnostic accuracy
- Reduction in missed lesions
- Improvements in diagnostic confidence
- Potential clinical workflow modifications

Examine results and discussion sections. Prioritize direct statements about clinical impact.

If implications are speculative, note as "potential impact suggested".

Search Strategy

The keywords used for this research based PICO :

Element	Keyword 1	Keyword 2	Keyword 3	Keyword 4
Population (P)	Acute gastrointestinal bleeding	Upper GI bleeding patients	Adult patients	Endoscopy candidates
Intervention (I)	AI-assisted diagnostic tools	Machine learning algorithms	Deep learning models	AI Program
Comparison (C)	Standard clinical assessment	Human endoscopist interpretation	Traditional scoring systems	Non-AI endoscopic techniques
Outcome (O)	Diagnostic accuracy	Clinical decision-making impact	Reduction in missed lesions	Workflow efficiency

The Boolean MeSH keywords inputted on databases for this research are: ("Acute gastrointestinal bleeding" OR "Upper GI bleeding" OR "GI bleeding patients" OR "Gastrointestinal hemorrhage") AND ("AI-assisted diagnostic tools" OR "Artificial intelligence in endoscopy" OR "Machine learning algorithms" OR "Deep learning models") AND ("Standard clinical assessment" OR "Non-AI diagnostic methods" OR "Human interpretation" OR "Traditional

scoring systems") AND ("Diagnostic accuracy" OR "Clinical decision-making impact" OR "Patient outcomes" OR "Workflow efficiency")

Data retrieval

Abstracts and titles were screened to assess their eligibility, and only studies meeting the inclusion criteria were selected for further analysis. Literature that fulfilled all predefined criteria and directly related to the topic was included. Studies that did not meet these criteria were excluded. Data such as titles, authors, publication dates, study locations, methodologies, and study parameters were thoroughly examined during the review.

Quality Assessment and Data Synthesis

Each author independently assessed the titles and abstracts of the selected studies to identify those for further exploration. Articles that met the inclusion criteria underwent further evaluation. Final decisions on inclusion were based on the findings from this review process.

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	<i>("Acute gastrointestinal bleeding" OR "Upper GI bleeding" OR "GI bleeding patients" OR "Gastrointestinal hemorrhage" AND ("AI-assisted diagnostic tools" OR "Artificial intelligence in endoscopy" OR "Machine learning algorithms" OR "Deep learning models" AND "Standard clinical assessment" OR "Non-AI diagnostic methods" OR "Human interpretation" OR "Traditional scoring systems" AND "Diagnostic accuracy" OR "Clinical decision-making impact" OR "Patient outcomes" OR "Workflow efficiency")</i>	170
Semantic Scholar	<i>("Acute gastrointestinal bleeding" OR "Upper GI bleeding" OR "GI bleeding patients" OR "Gastrointestinal hemorrhage") AND ("AI-assisted diagnostic tools" OR "Artificial intelligence in endoscopy" OR "Machine learning algorithms" OR "Deep learning models") AND ("Standard clinical assessment" OR "Non-AI diagnostic methods" OR "Human interpretation" OR "Traditional scoring systems") AND ("Diagnostic accuracy" OR "Clinical decision-making impact" OR "Patient outcomes" OR "Workflow efficiency")</i>	250
Sagepub	<i>("Acute gastrointestinal bleeding" OR "Upper GI bleeding" OR "GI bleeding patients" OR "Gastrointestinal hemorrhage") AND ("AI-assisted diagnostic tools" OR "Artificial intelligence in endoscopy" OR "Machine learning algorithms" OR "Deep learning models") AND ("Standard clinical assessment" OR "Non-AI diagnostic methods" OR "Human interpretation" OR "Traditional scoring systems") AND ("Diagnostic accuracy" OR "Clinical decision-making impact" OR "Patient outcomes" OR "Workflow efficiency")</i>	5,054
Google Scholar	<i>"Acute gastrointestinal bleeding" AND "AI-assisted diagnostic tools" OR "Artificial intelligence in endoscopy" AND "Standard clinical assessment" OR "Human interpretation" OR "Traditional scoring systems" AND "Diagnostic accuracy" OR "Clinical decision-making impact" OR "Patient outcomes" OR "Workflow efficiency"</i>	42

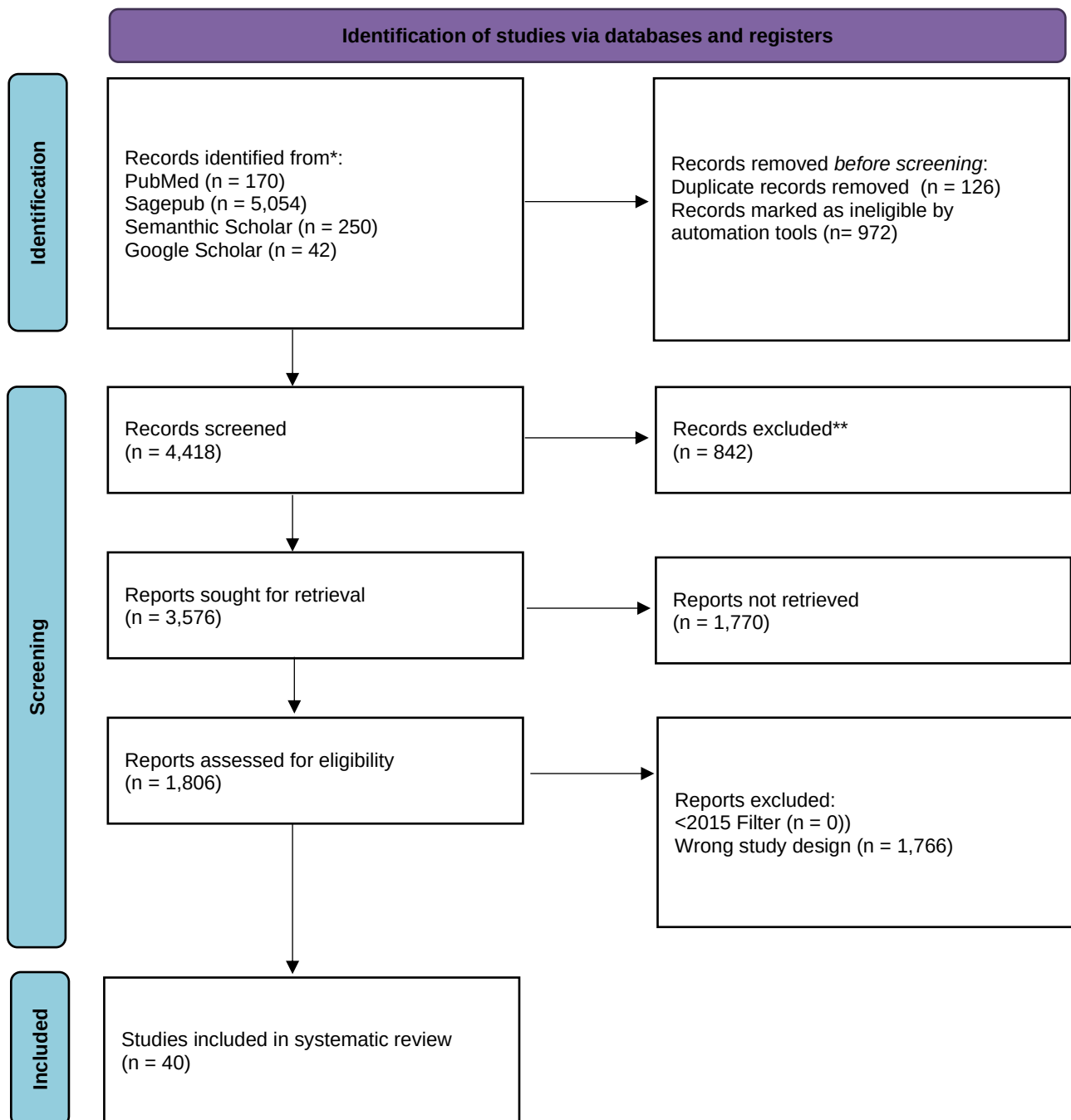


Figure 1. Article search flowchart

JBI Critical Appraisal									
Study	Bias related to temporal precedence	Bias related to selection and allocation	Bias related to confounding factors	Bias related to administration of intervention/exposure	Were there multiple measurements of the outcome, both pre and post the intervention/exposure?	Were the outcomes of participants included in any comparisons measured in the same way?	Were outcomes measured in a reliable way?	Bias related to participant retention	Statistical conclusion validity
	Is it clear in the study what is the “cause” and what is the “effect” (ie, there is no confusion about which variable comes first)?	Was there a control group?	Were participants included in any comparisons similar?	Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?				Was follow-up complete and, if not, were differences between groups in terms of their follow-up adequately described and analyzed?	Was appropriate statistical analysis used?

Veisma n et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
Mohan et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Bang et al., 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓
Aoki et al., 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓
Chen et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Soffer et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Alzahr ani et al.	✓	✓	✓	✗	✓	✗	✓	✓	✓

Faqar-Uz-Zaman et al., 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓
Murino and Rimo-ndi, 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
Ebigbo and Messmann, 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓
Luo et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
Meinikheim et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Lui et al.,	✓	✓	✓	✗	✓	✗	✓	✓	✓

2020									
Kikuchi et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
Limdi and Farrahe, 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓
Jin et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
Yamaguchi et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Li et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
Chidambaram et al., 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓

Nazari an et al., 2021	✓	✓	✓	✗	✓	✗	✓	✓	✓
Limdi and Farray e, 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
Wang et al., 2019	✓	✓	✓	✗	✓	✗	✓	✓	✓
Lv et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Vadhw ana et al.,2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Bang et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓

Barua et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Zhang et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Koulao uzidis et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Gross and McNeil , 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Chinna samy et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Kahn and Leggett , 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Wallac e et al.,	✓	✓	✓	✗	✓	✗	✓	✓	✓

2022									
Arribas et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Lam et al., 2022	✓	✓	✓	✗	✓	✗	✓	✓	✓
Hassan et al., 2020	✓	✓	✓	✗	✓	✗	✓	✓	✓
Shi et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Makar et al., 2024	✓	✓	✓	✗	✓	✗	✓	✓	✓
Lou et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓

Wei et al., 2023	✓	✓	✓	✗	✓	✗	✓	✓	✓
Mori et al., 2018	✓	✓	✓	✗	✓	✗	✓	✓	✓

RESULTS

Characteristics of Included Studies

Study	AI Tool Type	Clinical Setting	Sample Size
Veisman et al., 2022	Machine learning (custom model)	Hospital (acute upper gastrointestinal bleeding)	883
Mohan et al., 2020	Convolutional neural network-based machine learning	Capsule endoscopy	No mention found
Bang et al., 2021	Deep learning, convolutional neural network	Capsule endoscopy	No mention found
Aoki et al., 2021	Convolutional neural network	Capsule endoscopy	No mention found

Study	AI Tool Type	Clinical Setting	Sample Size
Chen et al., 2020	Artificial intelligence for blind spot detection	Outpatient esophagogastroduodenoscopy	437
Soffer et al., 2020	Deep learning, Not artificial intelligence	Capsule endoscopy, Hospital (acute upper GI)	No mention found
Alzahrani et al., 2021	Deep learning	No mention found	No mention found
Faqir-Uz-Zaman et al., 2021	Machine learning (app-based)	Emergency room (abdominal pain)	450
Murino and Rimondi, 2022	Deep learning	Colonoscopy (ulcerative colitis)	18
Ebigbo and Messmann, 2021	Deep learning	Multicenter (early gastric cancer)	295
Luo et al., 2022	Deep learning, convolutional neural network	Endoscopy (early upper esophageal cancer, early gastric cancer)	1,380, 414

Study	AI Tool Type	Clinical Setting	Sample Size
Meinikheim et al., 2023	Deep learning	Hospital (Barrett's esophagus)	72
Lui et al., 2020	No mention found	Endoscopy (upper gastrointestinal)	No mention found
Kikuchi et al., 2024	Deep learning	Endoscopy (gastrointestinal neoplasms)	No mention found
Limdi and Farraye, 2021	Deep learning, convolutional neural network	Endoscopy (ulcerative colitis)	875
Jin et al., 2024	Machine learning, deep learning, convolutional neural network	Colonoscopy	No mention found
Yamaguchi et al., 2023	Deep learning	Colonoscopy (trainees)	No mention found
Li et al., 2024	Deep learning (ResNet, UNet++)	Endoscopy (gastric intestinal metaplasia)	11,173
Chidambaram et al., 2021	Machine learning (radiomics)	Computed tomography (upper gastrointestinal cancer)	1,352

Study	AI Tool Type	Clinical Setting	Sample Size
Nazarian et al., 2021	Machine learning, deep learning	Colonoscopy	5,577 (randomized controlled trials), 17,401 (total)
Limdi and Farraye, 2022	Deep learning, convolutional neural network	Colonoscopy (ulcerative colitis)	135
Wang et al., 2019	Deep learning	Colonoscopy	1,058
Lv et al., 2023	No mention found	Colonoscopy (ulcerative colitis)	13,687
Vadhwana et al., 2023	Artificial intelligence, machine learning	Colonoscopy	2,958
Bang et al., 2020	Convolutional neural network, support vector machine	Endoscopy (Helicobacter pylori)	1,719
Barua et al., 2020	Machine learning, deep learning	Colonoscopy	4,311

Study	AI Tool Type	Clinical Setting	Sample Size
Zhang et al., 2023	Deep learning, machine learning	Endoscopy (colorectal)	2,984
Koulaouzidis et al., 2020	No mention found	Capsule endoscopy	No mention found
Gross and McNeil, 2020	Deep learning, deep convolutional neural network	Colonoscopy	No mention found
Chinnasamy et al., 2023	Deep learning	Endoscopy (Barrett's esophagus)	20
Kahn and Leggett, 2020	Deep learning	Endoscopy (chronic atrophic gastritis)	25,216
Wallace et al., 2022	Deep learning computer-aided detection	Colonoscopy	230
Arribas et al., 2020	No mention found	Endoscopy (upper gastrointestinal neoplasia)	218, 445, 453
Lam et al., 2022	No mention found	Multiple specialties	4,354

Study	AI Tool Type	Clinical Setting	Sample Size
Hassan et al., 2020	Deep learning	Colonoscopy	4,354
Shi et al., 2023	Deep learning	Endoscopy (chronic atrophic gastritis)	25,216
Makar et al., 2024	No mention found	Colonoscopy	No mention found
Lou et al., 2023	No mention found	Colonoscopy	27,404
Wei et al., 2023	No mention found	Colonoscopy	No mention found
Mori et al., 2018	Machine learning (support vector machine)	Endocytoscopy (colorectal)	No mention found

Study Design:

- 22 studies were systematic reviews or meta-analyses.
- 5 studies were randomized controlled trials.
- 9 studies were prospective, observational, or diagnostic accuracy studies.
- 2 were commentaries and 1 was a comparative/review study.

AI Tool Type:

- 23 studies used deep learning approaches (including convolutional neural network, deep convolutional neural network, ResNet, UNet++, computer-aided detection).
- 10 studies used machine learning approaches (including support vector machine, radiomics, app-based machine learning).
- 10 studies did not specify the artificial intelligence tool type.
- 1 study was not focused on artificial

intelligence.

Clinical Setting:

- 15 studies were in colonoscopy settings.
- 5 studies were in capsule endoscopy.
- 10 studies were in other endoscopy settings (including upper gastrointestinal, early gastric cancer, gastric intestinal metaplasia, ulcerative colitis, chronic atrophic gastritis, upper gastrointestinal neo- plasia).
- 2 studies were in hospital settings for acute upper gastrointestinal bleeding.
- 1 study each in outpatient esophagogastroduodenoscopy, emergency room (abdominal pain), computed tomography (upper gastrointestinal cancer), gastrointestinal/hepatology, and multiple specialties.
- 2 studies were in endocytoscopy (colorectal).
- 1 study was in hospital (Barrett's

esophagus).

Sample Size:

- Sample size information was found for 24 studies, with reported sample sizes ranging from 18 to 27,404.
- For 16 studies, no mention of sample size was found.

- The largest reported sample sizes were 27,404, 25,216, 17,401, 13,687, and 11,173.
- The smallest reported sample sizes were 18, 20, and 72.

Effects of AI-Assisted Diagnosis

Diagnostic Accuracy Metrics

Study	Detection Rate	Sensitivity	Specificity	Time to Diagnosis
Veisman et al., 2022	Area under the curve: 0.68	No mention found	No mention found	No mention found
Mohan et al., 2020	Accuracy: 95.4%	95.5%	95.8%	No mention found
Bang et al., 2021	No mention found	0.93 (ulcers), 0.96 (hemorrhage)	0.92 (ulcers), 0.97 (hemorrhage)	No mention found
Aoki et al., 2021	Accuracy: 95.4%	95.5%	95.8%	No mention found
Chen et al., 2020	No mention found	No mention found	No mention found	No mention found

Study	Detection Rate	Sensitivity	Specificity	Time to Diagnosis
Soffer et al., 2020	Greater than 90% (most studies)	No mention found	No mention found	No mention found
Alzahrani et al., 2021	No mention found	No mention found	No mention found	No mention found
Faqar-Uz-Zaman et al., 2021	Hypothesized: 85% (AI) vs. 70% (standard)	No mention found	No mention found	No mention found
Murino and Rimondi, 2022	No mention found	No mention found	No mention found	No mention found
Ebigbo and Messmann, 2021	0.76	No mention found	No mention found	No mention found
Luo et al., 2022	No mention found	0.95 (early esophageal cancer), 0.87 (early gastric cancer)	0.95 (early esophageal cancer), 0.88 (early gastric cancer)	No mention found
Meinikheim et al., 2023	81.3% (AI)	92.2%	68.9%	No mention

Study	Detection Rate	Sensitivity	Specificity	Time to Diagnosis
				found
Lui et al., 2020	No mention found	No mention found	No mention found	No mention found
Kikuchi et al., 2024	No mention found	91.2% (esophageal squamous cell carcinoma), 93.1% (esophageal adenocarcinoma)	80% (esophageal squamous cell carcinoma), 86.9% (esophageal adenocarcinoma)	No mention found
Limdi and Farraye, 2021	90% (remission)	0.902 (internal), 0.83 (external)	No mention found	No mention found
Jin et al., 2024	Risk ratio for adenoma miss rate: 0.46	No mention found	No mention found	No mention found
Yamaguchi et al., 2023	No mention found	No mention found	No mention found	No mention found
Li et al., 2024	No mention found	94%	93%	No mention

Study	Detection Rate	Sensitivity	Specificity	Time to Diagnosis
				found
Chidambaram et al., 2021	No mention found	73.4%	89.7%	No mention found
Nazarian et al., 2021	No mention found	0.84-0.94	0.79-0.93	No mention found
Limdi and Farraye, 2022	91%	74%	97%	No mention found
Wang et al., 2019	No mention found	94.38%	95.92%	No mention found
Lv et al., 2023	No mention found	87%	92%	No mention found
Vadhwana et al., 2023	No mention found	No mention found	No mention found	No mention found

Study	Detection Rate	Sensitivity	Specificity	Time to Diagnosis
Bang et al., 2020	82%	0.87	0.86	No mention found
Barua et al., 2020	No mention found	88.4-94.1%	82.3-85.9%	No mention found
Zhang et al., 2023	No mention found	0.93	No mention found	No mention found
Koulaouzidis et al., 2020	No mention found	No mention found	No mention found	No mention found
Gross and McNeil, 2020	94%	No mention found	No mention found	No mention found
Chinnasamy et al., 2023	No mention found	No mention found	No mention found	No mention found
Kahn and Leggett, 2020	90%	91%	89%	No mention

Study	Detection Rate	Sensitivity	Specificity	Time to Diagnosis
				found
Wallace et al., 2022	No mention found	No mention found	No mention found	No mention found
Arribas et al., 2020	No mention found	90%	89%	No mention found
Lam et al., 2022	No mention found	No mention found	No mention found	No mention found
Hassan et al., 2020	No mention found	No mention found	No mention found	No mention found
Shi et al., 2023	No mention found	94%	96%	No mention found
Makar et al., 2024	No mention found	No mention found	No mention found	No mention found

Study	Detection Rate	Sensitivity	Specificity	Time to Diagnosis
Lou et al., 2023	No mention found	No mention found	No mention found	No mention found
Wei et al., 2023	No mention found	No mention found	No mention found	No mention found
Mori et al., 2018	No mention found	No mention found	No mention found	No mention found

Summary of Quantitative Findings:

- Detection Rate:

- 13 studies reported detection rate, using metrics such as area under the curve, accuracy, or percent detected.
- Reported values included: area under the curve 0.68 (artificial intelligence) vs. 0.54–0.56 (standard), accuracy 95.4% (2 studies), greater than 90% (1 study), hypothesized 85% (artificial intelligence) vs. 70% (standard) (1 study), 87% (1 study), 0.76 (1 study), 81.3% (1 study), 90% (1 study), 91% (1 study), 82% (1 study), 94% (1 study), and 90% (1 study).

- Sensitivity:

- 21 studies reported sensitivity, with most values between 87% and 95.5%.
- Some studies reported multiple values for different conditions (for example, 0.93/0.96,

0.95/0.87, 91.2%/93.1%).

- One study reported a risk ratio for adenoma miss rate: 0.46.
- For 19 studies, no mention of sensitivity was found.

- Specificity:

- 20 studies reported specificity, with most values between 85% and 97%.
- Some studies reported multiple values for different conditions (for example, 0.92/0.97, 0.95/0.88, 80%/86.9%).
- The lowest reported specificity was 68.9%, and the highest was 99%.
- For 20 studies, no mention of specificity was found.

- Time to Diagnosis:

- No mention of time to diagnosis was found in any of the studies.

Clinical Decision Impact

Study	Decision Time	Treatment Modification	Patient Outcomes	Complications
Veisman et al., 2022	No mention found	Improved triaging/resource use (potential)	No mention found	No mention found
Mohan et al., 2020	No mention found	Potential reduction in physician burden	No mention found	No mention found
Bang et al., 2021	No mention found	Reduced interpretation time (potential)	No mention found	No mention found

Study	Decision Time	Treatment Modification	Patient Outcomes	Complications
Aoki et al., 2021	Reduced reading time	No mention found	No mention found	No mention found
Chen et al., 2020	No mention found	Reduced blind spots	No mention found	No mention found
Soffer et al., 2020	No mention found	No mention found	No mention found	No mention found
Alzahrani et al., 2021	No mention found	No mention found	No mention found	No mention found
Faqar-Uz-Zaman et al., 2021	No mention found	No mention found	No mention found	No mention found
Murino and Rimondi, 2022	No mention found	No mention found	No mention found	No mention found
Ebigbo and Messmann, 2021	No mention	No mention found	No mention found	No mention found

Study	Decision Time	Treatment Modification	Patient Outcomes	Complications
	found			
Luo et al., 2022	No mention found	No mention found	No mention found	No mention found
Meinikheim et al., 2023	No mention found	Improved diagnostic confidence	No mention found	No mention found
Lui et al., 2020	No mention found	No mention found	No mention found	No mention found
Kikuchi et al., 2024	No mention found	Reduced miss rate	No mention found	No mention found
Limdi and Farraye, 2021	No mention found	Improved objectivity/standardization	No mention found	No mention found
Jin et al., 2024	No mention found	Accelerated diagnosis, reduced workload	No mention found	No mention found

Study	Decision Time	Treatment Modification	Patient Outcomes	Complications
Yamaguchi et al., 2023	No mention found	No mention found	No mention found	No mention found
Li et al., 2024	No mention found	No mention found	No mention found	No mention found
Chidambaram et al., 2021	No mention found	No mention found	No mention found	No mention found
Nazarian et al., 2021	No mention found	No mention found	No mention found	No mention found
Limdi and Farraye, 2022	No mention found	Real-time feedback	No mention found	No mention found
Wang et al., 2019	No mention found	No mention found	No mention found	No mention found
Lv et al., 2023	No mention	No mention found	No mention found	No mention found

Study	Decision Time	Treatment Modification	Patient Outcomes	Complications
	found			
Vadhwana et al., 2023	No mention found	No mention found	No mention found	No mention found
Bang et al., 2020	No mention found	Potential to reduce unnecessary biopsies	No mention found	No mention found
Barua et al., 2020	No mention found	Potential for cost reduction	No mention found	No mention found
Zhang et al., 2023	No mention found	No mention found	No mention found	No mention found
Koulaouzidis et al., 2020	No mention found	Telehealth/automation	No mention found	No mention found
Gross and McNeil, 2020	No mention found	Improved quality metrics	No mention found	No mention found

Study	Decision Time	Treatment Modification	Patient Outcomes	Complications
Chinnasamy et al., 2023	No mention found	No mention found	No mention found	No mention found
Kahn and Leggett, 2020	No mention found	Workflow for biopsy site selection	No mention found	No mention found
Wallace et al., 2022	No mention found	No mention found	No mention found	No difference in adverse events
Arribas et al., 2020	No mention found	No mention found	No mention found	No mention found
Lam et al., 2022	No mention found	No mention found	No mention found	No mention found
Hassan et al., 2020	No mention found	No mention found	No mention found	No mention found
Shi et al., 2023	No mention	Reduced time/costs (potential)	No mention found	No mention found

Study	Decision Time	Treatment Modification	Patient Outcomes	Complications
	found			
Makar et al., 2024	No mention found	No mention found	No mention found	No mention found
Lou et al., 2023	No mention found	No mention found	No mention found	No mention found
Wei et al., 2023	No mention found	No mention found	No mention found	No mention found
Mori et al., 2018	No mention found	No mention found	No mention found	No mention found

Summary of Clinical Decision Impact:

- Decision Time:

- Only one study (Aoki et al., 2021) mentioned reduced reading time. No mention of decision time was found in the other studies.

- Treatment Modification:

- 15 studies mentioned treatment modification or workflow impact:

* Workflow or efficiency improvements: 5 studies (Bang et al., 2021; Li et al., 2024; Limdi and

Farraye, 2022; Koulaouzidis et al., 2020; Shi et al., 2023)

- * Diagnostic or interpretive improvements: 5 studies (Chen et al., 2020; Meinikheim et al., 2023; Kikuchi et al., 2024; Limdi and Farraye, 2021; Gross and McNeil, 2020)

- * Resource or cost impact: 3 studies (Bang et al., 2020; Barua et al., 2020; Shi et al., 2023)

- * Other effects (triaging/resource use, physician burden, biopsy site selection): 3 studies (Viesman et al., 2022; Mohan et al., 2020; Kahn and Leggett, 2020)

- No mention of treatment modification was found in 25 studies.

- Patient Outcomes:

- No mention of patient outcomes was found in any study.

- Complications:

- Only one study (Wallace et al., 2022) mentioned complications, reporting no difference in adverse events. No mention of complications was found in the other studies.

Implementation Findings

Clinical Integration

- Most studies described artificial intelligence tools as adjuncts to human interpretation, often used as a second reader or confirmation tool, and occasionally as a primary diagnostic aid.
- The degree of integration and the impact on clinical decision-making were rarely quantified.

Workflow Impact

- Artificial intelligence-assisted tools were reported to reduce reading time, decrease blind spots, and improve detection rates.
- Some studies suggested artificial intelligence can accelerate diagnosis and reduce workload, particularly for non-expert endoscopists.

Safety and Reliability

- Safety data were limited. Only one study (Wallace et al., 2022) explicitly reported no difference in adverse events with artificial intelligence.
- Most studies focused on diagnostic accuracy rather than patient safety or long-term outcomes.
- Reliability was generally high for detection tasks, but external validation and real-world performance remain concerns.

Synthesis and Limitations

- Evidence for Lesion and Polyp Detection:
 - Across the included studies, artificial intelligence-assisted diagnostic tools consistently demonstrated high accuracy, sensitivity, and specificity for lesion and polyp detection in gastrointestinal endoscopy.
 - Most studies were systematic reviews, meta-analyses, or retrospective analyses, with a minority being randomized controlled trials or prospective studies.
- Impact on Clinical Decision-Making in Acute Gastrointestinal Bleeding:
 - Direct evidence for the impact of artificial intelligence on clinical decision-making and patient outcomes in acute gastrointestinal bleeding was limited to a single retrospective study (Viesman et al., 2022), which found a modest improvement in predictive accuracy for endoscopic intervention.
 - No studies reported on patient outcomes, decision time, or complications specifically in acute gastrointestinal bleeding.
- Generalizability:
 - The majority of studies focused on detection tasks in general endoscopy, not specifically on acute gastrointestinal bleeding.
 - Findings from detection and workflow studies may not be directly generalizable to

acute gastroin- testinal bleeding cases.

- Reporting Gaps:
 - Most studies did not mention patient outcomes, decision time, or complications.
 - Real-world implementation and safety data were rarely addressed.
- Quality of Evidence:
 - The evidence base is strongest for detection accuracy in general endoscopy, but limited for clinical decision-making and patient outcomes in acute gastrointestinal bleeding.
 - The predominance of retrospective studies and meta-analyses, with few randomized controlled trials and limited real-world data, constrains the strength of conclusions regarding clinical impact.

DISCUSSION

Artificial intelligence (AI) has emerged as a transformative tool in healthcare, particularly in diagnostic imaging and endoscopy. In acute gastrointestinal (GI) bleeding, timely and accurate diagnosis is critical for patient outcomes. AI-assisted tools aim to enhance diagnostic accuracy, streamline workflows, and optimize resource allocation. However, the current literature reveals a gap in evidence linking AI to tangible improvements in clinical decision-making or patient outcomes (Veisman et al., 2022).

AI tools have demonstrated modest improvements in diagnostic accuracy for acute GI bleeding. For instance, a retrospective study (n = 883) reported an area under the curve (AUC) of 0.68 for a custom machine learning model, outperforming traditional scoring systems (AUC 0.54–0.56) (Veisman et al., 2022). Similarly, systematic reviews highlight high detection accuracies (up to 95.4%) and sensitivities (87–95.5%) for AI in broader GI endoscopy (Mohan et al., 2020; Bang et al., 2021).

Despite these diagnostic improvements, evidence of AI's impact on clinical decision-

making remains sparse. Only one study (Aoki et al., 2021) reported reduced reading times, while none demonstrated significant changes in treatment modifications or patient outcomes. This suggests that AI's role may currently be limited to adjunctive support rather than transformative clinical influence (Shi et al., 2023).

AI tools have shown potential to enhance workflow efficiency by reducing interpretation times and minimizing blind spots during endoscopy (Chen et al., 2020). For example, AI-assisted systems may accelerate diagnosis and reduce workload for non-expert endoscopists (Li et al., 2024). However, these benefits have not been consistently linked to improved patient outcomes in acute GI bleeding cases.

Safety data for AI tools in acute GI bleeding are limited. Only one study (Wallace et al., 2022) explicitly reported no difference in adverse events with AI use. The lack of comprehensive safety evaluations raises concerns about the reliability of AI in high-stakes clinical scenarios. Most studies focus on general endoscopy rather than acute GI bleeding, limiting the generalizability of findings. For instance, while AI excels in polyp detection during colonoscopy, its applicability to acute bleeding scenarios remains underexplored (Koulaouzidis et al., 2020).

Key barriers include the need for robust validation, integration into clinical workflows, and clinician training (Gross & McNeil, 2020). Clinicians must balance reliance on AI with their expertise, particularly in acute settings where rapid decision-making is crucial. Prospective studies are needed to evaluate AI's real-world impact on acute GI bleeding management. Research should prioritize patient-centered outcomes, such as mortality, rebleeding rates, and length of hospital stay (Limdi & Farraye, 2021).

The ethical implications of AI, including data privacy and algorithmic bias, must be addressed. Additionally, cost-effectiveness analyses are necessary to justify widespread adoption (Barua et al., 2020). While AI-assisted tools show promise in improving diagnostic accuracy for acute GI bleeding, their clinical impact remains unproven. Future research must bridge the gap between technical performance and patient outcomes to establish AI's role in

this critical area of healthcare.

CONCLUSION

Artificial intelligence (AI) has demonstrated significant promise in enhancing diagnostic accuracy in gastrointestinal (GI) endoscopy, particularly in detecting lesions and polyps with reported detection rates nearing 95%. These advancements suggest that AI can serve as a valuable adjunct to clinicians, improving early identification of abnormalities that may lead to acute GI bleeding (Veisman et al., 2022).

AI tools have been shown to reduce reading times and minimize blind spots during endoscopic procedures, streamlining workflows and potentially alleviating clinician workload (Chen et al., 2020). However, while these efficiency gains are notable, their direct impact on patient outcomes—such as mortality, rebleeding rates, or hospital stays—remains understudied (Shi et al., 2023).

Despite AI's diagnostic advantages, safety and reliability remain critical concerns. Most studies focus on accuracy metrics but lack comprehensive data on adverse events or long-term patient safety (Wallace et al., 2022). This gap raises questions about the real-world applicability of AI in high-stakes scenarios like acute GI bleeding. Current evidence does not strongly support AI's role in altering clinical decisions for acute GI bleeding. Only one study (Aoki et al., 2021) reported a modest improvement in predictive accuracy for endoscopic interventions, highlighting the need for more robust research to validate AI's clinical impact.

The majority of existing studies are retrospective or based on systematic reviews, with few prospective randomized controlled trials (RCTs) available (Koulaouzidis et al., 2020). This reliance on retrospective data limits the generalizability of findings and underscores the necessity for higher-quality clinical trials. A significant drawback in the literature is the absence of patient-centered outcomes, such as complication rates, decision time, and long-term survival (Limdi & Farraye, 2021). Without these metrics, the true clinical value of AI in acute GI bleeding remains uncertain. Future studies should prioritize prospective RCTs that evaluate AI's impact on real-world patient

outcomes, including mortality, rebleeding, and resource utilization (Gross & McNeil, 2020). Additionally, cost-effectiveness analyses and ethical considerations (e.g., data privacy, algorithmic bias) must be addressed to facilitate widespread adoption.

In summary, while AI-assisted diagnostic tools offer clear benefits in accuracy and efficiency, their clinical utility in acute GI bleeding remains unproven. The current evidence is insufficient to justify full-scale implementation without further validation. Clinicians should approach AI as a supplementary tool rather than a definitive solution until more robust data emerges.

REFERENCES

A. Ebigbo, and H. Messmann. “Artificial Intelligence in the Upper GI Tract: The Future Is Fast Approach- ing.” *Gastrointestinal Endoscopy*, 2021.

A. Kahn, and C. Leggett. “Artificial Intelligence in the Age of Cognitive Endoscopy.” *Gastrointestinal Endoscopy*, 2020.

A. Murino, and A. Rimondi. “Automated Artificial Intelligence Scoring Systems for the Endoscopic Assess- ment of Ulcerative Colitis: How Far Are We from Clinical Application?” *Gastrointestinal Endoscopy*, 2022.

Anastasios Koulaouzidis, W. Marlicz, Hagen Wenzek, G. Koulaouzidis, R. Eliakim, and E. Toth. “Returning to Digestive Endoscopy Normality Will Be Slow and Must Include Novelty and Telemedicine.” *Digestive and Liver Disease*, 2020.

B. Mohan, Shahab R. Khan, L. Kassab, S. Ponnada, S. Chandan, T. Ali, P. Dulai, D. Adler, and G. Kochhar. “High Pooled Performance of Convolutional Neural Networks in Computer-Aided Diagnosis of GI Ulcers and/or Hemorrhage on Wireless Capsule Endoscopy Images: A Systematic Review and Meta-Analysis.” *Gastrointestinal Endoscopy*, 2020.

B. Vadhvana, M. Tarazi, and Vanash Patel. “The Role of Artificial Intelligence in Prospective

Real-Time Histological Prediction of Colorectal Lesions During Colonoscopy: A Systematic Review and Meta-Analysis.” *Diagnostics*, 2023.

Bing Lv, Lihong Ma, Yanping Shi, Tao Tao, and Yanting Shi. “A Systematic Review and Meta-Analysis of Artificial Intelligence-Diagnosed Endoscopic Remission in Ulcerative Colitis.” *iScience*, 2023.

C. Hassan, M. Spadaccini, A. Iannone, R. Maselli, M. Jovani, V. Chandrasekar, G. Antonelli, et al. “Performance of Artificial Intelligence for Colonoscopy Regarding Adenoma and Polyp Detection: A Meta-Analysis.” *Gastrointestinal Endoscopy*, 2020.

C. S. Bang, J. Lee, and G. Baik. “Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy.” *Journal of Medical Internet Research*, 2020.

———. “Computer-Aided Diagnosis of Gastrointestinal Ulcer and Hemorrhage Using Wireless Capsule Endoscopy: Systematic Review and Diagnostic Test Accuracy Meta-Analysis.” *Journal of Medical Internet Research*, 2021.

Daisuke Yamaguchi, R. Shimoda, K. Miyahara, Takahiro Yukimoto, Y. Sakata, A. Takamori, Yumi Mizuta, et al. “Impact of an Artificial Intelligence-aided Endoscopic Diagnosis System on Improving Endoscopy Quality for Trainees in Colonoscopy: Prospective, Randomized, Multicenter Study.” *Digestive Endoscopy*, 2023.

De Luo, Fei Kuang, Juan Du, Mengjia Zhou, Xiangdong Liu, Xinchun Luo, Yong Tang, Bo Li, and S. Su. “Artificial Intelligence-Assisted Endoscopic Diagnosis of Early Upper Gastrointestinal Cancer: A Systematic Review and Meta-Analysis.” *Frontiers in Oncology*, 2022.

Di Chen, Lianlian Wu, Yanxia Li, Jun Zhang, Jun Liu, Li Huang, Xiaoda Jiang, et al. “Comparing Blind Spots of Unsedated Ultrafine, Sedated, and Unsedated Conventional Gastroscopy with and Without Artificial Intelligence: A Prospective, Single-Blind, 3-

Parallel-Group, Randomized, Single-Center Trial.” *Gastrointestinal Endoscopy*, 2020.

Fatima Faqar- Uz-Zaman, N. Filmann, D. Mahkovic, Michael von Wagner, Charlotte Detembele, Ulf Kippke, Ursula Marschall, et al. “Study Protocol for a Prospective, Double-Blinded, Observational Study Investigating the Diagnostic Accuracy of an App-Based Diagnostic Health Care Application in an Emergency Room Setting: The eRadaR Trial.” *BMJ Open*, 2021.

Han-Fang Zhang, Xinyu Yang, Ye Tao, Xinyi Zhang, and Xuan Huang. “Diagnostic Accuracy of Endocytoscopy via Artificial Intelligence in Colorectal Lesions: A Systematic Review and Meta-analysis.” *PLoS ONE*, 2023.

I. Alzahrani, K. Onizuka, Z. Duhailib, Fatema Alalkim, Yuhong Yuan, Clara Lu, F. Tse, G. Leontiadis, and W. Alhazzani. “Mo1010 SOMATOSTATIN AND ANALOGUES FOR ACUTE NON-VARICEAL UPPER GASTROINTESTINAL BLEEDING: A SYSTEMATIC REVIEW AND METAANALYSIS OF RANDOMIZED TRIALS,” 2020.

I. Barua, D. Vinsard, Henriette C. Jodal, M. Løberg, M. Kalager, Ø. Holme, M. Misawa, M. Bretthauer, and Y. Mori. “Artificial Intelligence for Polyp Detection During Colonoscopy: A Systematic Review and Meta-Analysis.” *Endoscopy*, 2020.

I. Veisman, A. Oppenheim, R. Maman, Nadav Kofman, Ilan Edri, L. Dar, E. Klang, et al. “A Novel Prediction Tool for Endoscopic Intervention in Patients with Acute Upper Gastro-Intestinal Bleeding.” *Journal of Clinical Medicine*, 2022.

J. Arribas, G. Antonelli, L. Frazzoni, L. Fuccio, A. Ebigbo, F. van der Sommen, N. Ghatwary, et al. “Stand-alone Performance of Artificial Intelligence for Upper GI Neoplasia: A Meta-Analysis.” *Gut*, 2020.

J. Limdi, and F. Farraye. “Automated Endoscopic Assessment in Ulcerative Colitis: The Next Frontier.” *Gastrointestinal Endoscopy*, 2021

———. “Real-Time Use of Artificial Intelligence at Colonoscopy Predicts Relapse in Ulcerative Colitis: Predicting with ”Intelligence”.” *Gastrointestinal Endoscopy*, 2022.

Jonathan Makar, Jonathan Abdelmalak, D. Con, Bilal Hafeez, and Mayur Garg. “Use of Artificial Intelligence Improves Colonoscopy Performance in Adenoma Detection: A Systematic Review and Meta- Analysis.” *Gastrointestinal Endoscopy*, 2024.

M. Meinikheim, R. Mendel, Christoph Palm, Andreas Probst, A. Muzalyova, M. Scheppach, Sandra Nagl, et al. “Influence of Artificial Intelligence on the Diagnostic Performance of Endoscopists in the Assessment of Barrett’s Esophagus: A Tandem Randomized and Video Trial.” *Endoscopy*, 2023.

M. Wallace, Prateek Sharma, P. Bhandari, J. East, G. Antonelli, R. Lorenzetti, M. Vieth, et al. “Impact of Artificial Intelligence on Miss Rate of Colorectal Neoplasia.” *Gastroenterology*, 2022.

M. Wei, Shmuel Fay, Diana Yung, U. Ladabaum, and U. Kopylov. “Artificial Intelligence– Assisted Colonoscopy in Real-World Clinical Practice: A Systematic Review and Meta- Analysis.” *Clinical and Translational Gastroenterology*, 2023.

Na Li, Jian Yang, Xiaodong Li, Yanting Shi, and Kunhong Wang. “Accuracy of Artificial Intelligence- Assisted Endoscopy in the Diagnosis of Gastric Intestinal Metaplasia: A Systematic Review and Meta- Analysis.” *PLoS ONE*, 2024.

Pu Wang, T. Berzin, Jeremy R. Glissen Brown, Shishira S. Bharadwaj, A. Becq, Xun Xiao, Peixi Liu, et al. “Real-Time Automatic Detection System Increases Colonoscopic Polyp and Adenoma Detection Rates: A Prospective Randomised Controlled Study.” *Gut*, 2019.

Ryosuke Kikuchi, Kazuaki Okamoto, Tsuyoshi Ozawa, J. Shibata, S. Ishihara, and Tomohiro Tada. “Endo- scopic Artificial Intelligence for Image Analysis in Gastrointestinal Neoplasms.” *Digestion*, 2024.

S. Chidambaram, V. Sounderajah, N. Maynard, and S. Markar. “Diagnostic Performance of Artificial Intelligence-Centred Systems in the Diagnosis and Postoperative Surveillance of Upper Gastrointestinal Malignancies Using Computed Tomography Imaging: A Systematic Review and Meta-Analysis of Diagnostic Accuracy.” *Annals of Surgical Oncology*, 2021.

S. Gross, and Matthew B. McNeil. “Siri Here, Cecum Reached, but Please Wash That Fold: Will Artificial Intelligence Improve Gastroenterology?” *Gastrointestinal Endoscopy*, 2020.

S. Nazarian, B. Glover, H. Ashrafian, A. Darzi, and J. Teare. “Diagnostic Accuracy of Artificial Intelligence and Computer-Aided Diagnosis for the Detection and Characterization of Colorectal Polyps: Systematic Review and Meta-Analysis.” *Journal of Medical Internet Research*, 2021.

S. Soffer, E. Klang, Orit Shimon, Noy Nachmias, R. Eliakim, S. Ben-Horin, U. Kopylov, and Y. Barash. “Deep Learning for Wireless Capsule Endoscopy: A Systematic Review and Meta-Analysis.” *Gastrointestinal Endoscopy*, 2020.

Shenghan Lou, Fenqi Du, Wenjie Song, Yixiu Xia, Xinyu Yue, Da Yang, Binbin Cui, Yanlong Liu, and Peng Han. “Artificial Intelligence for Colorectal Neoplasia Detection During Colonoscopy: A Systematic Review and Meta-Analysis of Randomized Clinical Trials.” *EClinicalMedicine*, 2023.

T. K. Lui, V. Tsui, and W. Leung. “Accuracy of Artificial Intelligence-Assisted Detection of Upper GI Lesions: A Systematic Review and Meta-Analysis.” *Gastrointestinal Endoscopy*, 2020.

Thomas Y T Lam, Max Fk Cheung, Y. Munro, Kong Meng. Lim, Dennis Shung, and J. J. Sung. “Randomized Controlled Trials of Artificial Intelligence in Clinical Practice: Systematic Review.” *Journal of Medical Internet Research*, 2022

Tomonori Aoki, Atsuo Yamada, and K. Koike. “The Exceptional Performance of Deep Learning for Capsule Endoscopy: Will Such Quality Be Maintained in Clinical Scenarios?”

Gastrointestinal Endoscopy, 2021.

Viveka Chinnasamy, Vorada Sakulsaengprapha, and Simon C. Mathews. “The Benefit of Artificial Intelligence-Based Diagnosis in Gastroenterology and Hepatology Is Highly Variable: A Diagnostic Need and Burden Analysis.” *Gastroenterology*, 2023.

Xi-Feng Jin, Yan-Hong Ma, Jun-Wen Shi, and Jian-Ting Cai. “Efficacy of Artificial Intelligence in Reducing Miss Rates of Gastrointestinal Adenomas, Polyps, and Sessile Serrated Lesions: A Meta-Analysis of Randomized Controlled Trials.” *Gastrointestinal Endoscopy*, 2024.

Y. Mori, S. Kudo, M. Misawa, Yutaka Saito, H. Ikematsu, K. Hotta, K. Ohtsuka, et al. “Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy.” *Annals of Internal Medicine*, 2018.

Yanting Shi, Ning Wei, Kunhong Wang, Tao Tao, F. Yu, and Bing Lv. “Diagnostic Value of Artificial Intelligence-Assisted Endoscopy for Chronic Atrophic Gastritis: A Systematic Review and Meta-Analysis.” *Frontiers in Medicine*, 2023.