Artificial Intelligence in Gastrointestinal Disease Diagnosis: A Systematic Review of Endoscopy, Histology, and Radiology Applications

Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery

WILEY

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

ADVANCED REVIEW OPEN ACCESS

Artificial Intelligence in Gastrointestinal Disease Diagnosis: A Systematic Review of Endoscopy, Histology, and Radiology Applications

Sameena Pathan[1] | Alen Shahini[2] | Massimo Salvi[2] | Tanweer Ali[3] | Jovita Relasha Lewis[1] | Filippo Molinari[2] | U. Rajendra Acharya[4]

[1]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India | [2]Biolab, PolitoBIOMed Laboratory, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy | [3]Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India | [4]School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia

Correspondence: Massimo Salvi (massimo.salvi@polito.it)

Received: 8 April 2025 | **Revised:** 1 August 2025 | **Accepted:** 14 February 2026

Editor-in-Chief: Witold Pedrycz

Keywords: artificial intelligence | colon cancer | endoscopy images | gastric cancer | gastrointestinal disease

ABSTRACT

Gastrointestinal (GI) diseases remain among the leading causes of global mortality, with early detection directly linked to survival outcomes. While previous reviews have focused on single imaging modalities, this systematic review uniquely examines artificial intelligence applications across endoscopic, radiological, and histological approaches, reflecting actual clinical diagnostic pathways. This systematic review analyzes 76 high-quality studies (2016–2024) and provides the first comprehensive assessment of how AI performs across different imaging techniques for GI abnormality detection. This multi-modal perspective is particularly timely as healthcare systems move toward integrated diagnostic workflows. Our analysis reveals endoscopy as the most widely used modality ($n = 44$), particularly for *Helicobacter pylori*, colorectal polyps, and ulcerative colitis detection. Histological analysis emerges as the second most common approach ($n = 25$), especially for celiac disease and ulcerative colitis, while CT imaging ($n = 10$) primarily supports colorectal polyp detection. Deep learning methods significantly outnumber traditional machine learning techniques (68 vs. 8 studies), consistently achieving 90%–99% diagnostic accuracy across multiple disease categories. However, these systems face significant implementation barriers to clinical adoption. Most validation is still conducted in controlled, single-center settings using curated datasets that poorly reflect clinical complexity. Future studies must prioritize multicenter validation, standardized imaging protocols and preprocessing pipelines, and the integration of interpretable AI models capable of providing transparent diagnostic rationale. This review maps the current technical landscape while highlighting critical translational challenges that must be addressed to enable real-world impact.

This article is categorized under:

Technologies > Data Preprocessing

Technologies > Artificial Intelligence

1 | Introduction

Artificial intelligence (AI) has had a significant impact on the medical industry in recent years, as seen by the large number of papers on the subject (Dlamini et al. 2020; Trenfield et al. 2022). Public health (Morley et al. 2020; Schwalbe and Wahl 2020), patient diagnosis (S. Huang et al. 2020; Uyanik et al. 2025), medication (Paul et al. 2021), clinician-patient communication (Brizio et al. 2022), and telemedicine (Shaik et al. 2023) are just a few of the many possible uses for AI in healthcare (Shuvo et al. 2025). The impact of AI on gastrointestinal (GI) system analysis is the main topic of this study. Many diseases impact the GI tract, and increasing patient lifespan depends on early detection and treatment (Horiuchi et al. 2020).

AI systems have the ability to analyze endoscopic images and produce diagnostic ratings for diseases such as inflammatory bowel disease (IBD) (Stidham and Takenaka 2022). Similar applications are being developed for histopathological slides and radiological scans, which are discussed in the following section. Nonetheless, the interpretation of an exam might fluctuate significantly among physicians and even within the same physician over time. Consequently, samples are still taken during endoscopies to confirm diagnosis by pathology investigation (Takenaka et al. 2020). Thankfully, advances in medical technology are leading to early detection through increased public awareness, screening initiatives, and screening programs. Every year, more GI cancer cases are identified at an earlier stage.

TABLE 1 | Imaging modalities for detecting GI diseases.

Types of images	Diseases
Endoscopy	Erosion; polyp; ulcer; submucosal tumor; xanthoma, lesions in the proximal intestine
CT image	Pancreatic cancer; gastric cancer; inflammatory bowel
Histopathological image	Gastric disease

1.1 | Imaging Modalities

Three primary imaging modalities are used for diagnosing gastrointestinal diseases: endoscopy, computed tomography (CT), and histopathological examination. Each modality provides complementary information at different scales, from macroscopic visualization of organs to microscopic cellular details.

Endoscopy is the most common method for detecting GI diseases. This method involves inserting a thin, flexible tube with a tiny camera on the end into the GI tract. This allows doctors to directly visualize the esophagus, stomach, and intestines to look for ulcers, polyps, cancers, and other issues. White light endoscopy (WLE) is the primary technique used to screen for early gastric cancer. However, some studies showed that WLE misses 20%–40% of early gastric cancer cases, even though later stages are easier to spot (Hwang et al. 2018; Sugano et al. 2015).

CT scanning provides three-dimensional visualization of the digestive organs. This technique enables physicians to determine the precise location of tumors, abscesses, inflammation, and other anomalies, offering a macro-level perspective of the tissues and their relationships to surrounding structures.

Histological examination provides a cellular perspective of GI tissue at the microscale. The ability of pathologists to recognize abnormal alterations in tissue structure aids in the diagnosis of diseases such as inflammatory bowel disease, celiac disease, and gastritis. Histopathology remains the gold standard for definitive diagnosis of many gastrointestinal conditions, particularly for confirming malignancies.

The imaging modalities used to diagnose GI disorders are summarized in Table 1, which also highlights each modality's unique functions and contributions to the field of CAD. Figure 1 shows an illustrative example of each modality covered in this review.

1.2 | Role of AI in Gastrointestinal Diseases

The initial suggestions for GI analysis using computer-aided diagnosis (CAD) date back to the 1990s. Early approaches employed region-growing algorithms and pixel-based techniques to segment the outlines of the large intestine lumen

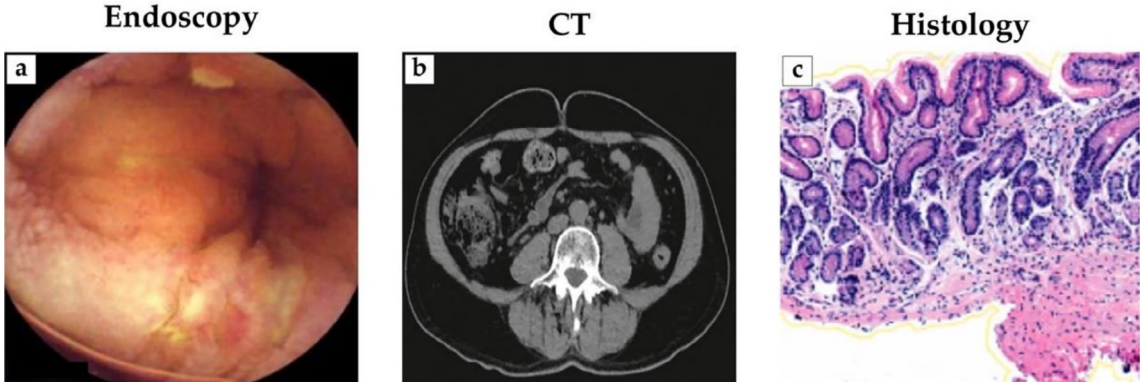

FIGURE 1 | Image modalities covered in this review. (a) Endoscopy of ulcers in the proximal intestine, (b) CT image of pancreatic cancer, and (c) Histology of a gastric mucosa.

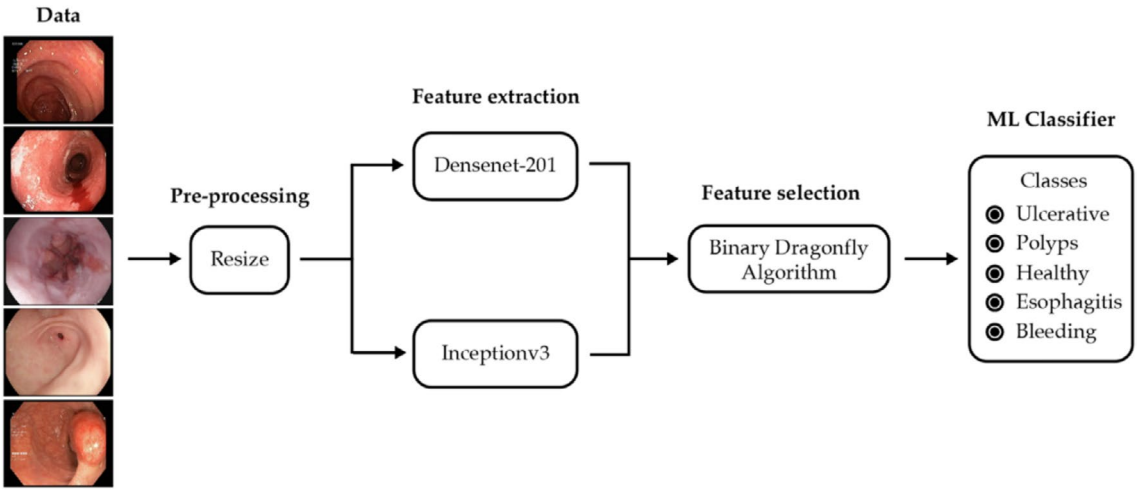

FIGURE 2 | Example of an AI-based framework for GI disease classification using endoscopic images. Common stages involve pre-processing of endoscopic images, extraction of discriminative features, and classification of lesions.

and identify lower GI tract diseases (Alagappan et al. 2018). By the late 1990s, most research concentrated on combining intelligent pattern classification with texture, color, or hybrid analytical techniques (Alagappan et al. 2018). In the last few years, several AI researchers have developed automated CAD models to identify GI anomalies by blending handcrafted features and deep convolutional neural networks (CNN) to recognize patterns in wireless capsule endoscopy (WCE) images (Charfi and El Ansari 2017; Khan et al. 2019; Naz et al. 2021; Saito et al. 2020). Various quantitative studies have utilized color features (Suman, Hussin, Malik, Ho, et al. 2017; Suman, Hussin, Malik, Pogorelov, et al. 2017), textures (Mohammad and Al-Razgan 2022), key point attributes (Tuba et al. 2017), and histograms of oriented gradients (Charfi and El Ansari 2017) to analyze GI disorders. Thanks to recent advances in DL, some researchers have also applied deep features for GI disease recognition. Figure 2 illustrates a typical DL framework for GI disease classification.

Given the increasing number of AI methods for GI disease detection and classification across various imaging modalities, a systematic review was conducted to comprehensively analyze and compare the existing techniques. The significance of this review lies in its broad examination of how AI is being applied across the full spectrum of GI imaging modalities. This is particularly timely, as healthcare systems worldwide are under increasing pressure to improve early detection rates while managing rising patient volumes. Understanding how AI can be effectively deployed across different imaging techniques is crucial for several reasons:

1. It enables healthcare providers to optimize diagnostic pathways by identifying where AI can have the greatest impact.
2. It helps researchers uncover gaps and opportunities across modalities, rather than focusing solely on individual techniques
3. It lays the groundwork for developing truly integrated AI systems that synthesize insights from multiple imaging sources.

This multimodal perspective is especially relevant given the ongoing innovation in medical imaging technologies, including the emergence of hybrid modalities (e.g., hyperspectral imaging) and the integration between different diagnostic tools.

1.3 | Related Review

In recent years, several reviews have been published on the application of AI methods for GI diseases. However, most of the previous works only focus on a single imaging modality or a specific pathology:

- Tziortziotis et al. (Tziortziotis et al. 2021) reviewed AI only for capsule endoscopy.
- Christou et al. (Christou and Tsoulfas 2021) provided an overview of the current applications of ML-based models in gastroenterology and hepatology, highlighting the opportunities they present.
- Gubatan et al. (Gubatan et al. 2021) summarized AI for assessing and predicting only IBD.
- Jin et al. (Jin et al. 2022) surveyed DL methods employed for gastroscopy images, covered only diseased and healthy conditions of the gastric tissue.
- Jahmunah et al. (Jahmunah et al. 2023) reviewed methods employed for detecting celiac disease across endoscopy, capsule endoscopy, and biopsy.
- Lewis et al. (Lewis et al. 2024) discussed AI methods used for GI disease, focusing specifically on endoscopic procedures.
- Song et al. (Song et al. 2023) provided a review of computational pathology advances, focusing primarily on methodological developments in whole-slide image analysis across multiple diseases.
- Bilal, Nimir, et al. (2023) presented an in-depth analysis of AI applications in colorectal cancer immunotherapy, with

particular emphasis on microsatellite instability prediction from histopathology.

Our systematic review takes a distinctly different approach by examining the integration of multiple imaging modalities in GI disease diagnosis. We specifically considered studies on gastric cancer, peptic ulcer illness, ulcerative colitis, *Helicobacter pylori* infection, colorectal cancer, esophageal cancer, and celiac disease. This selection represents a coherent group of GI conditions where the diagnostic process typically involves multiple imaging techniques. The studies included employed computer-aided diagnostics to automate the study of many imaging modalities, including CT scans, GI tract x-rays, histology slides, capsule endoscopy, and endoscopic video and photographs. This reflects real-world clinical workflows, where a comprehensive diagnosis often requires the use of several imaging techniques. The comparison of our proposed review paper with the previously published review papers is shown in Table 2.

This review makes several key contributions compared to prior work in this field. First, it provides a comprehensive overview of how AI is applied across multiple imaging modalities such as endoscopy, histopathology, and radiology. This multimodal approach mirrors the actual clinical diagnostic pathway, where different imaging techniques provide complementary information crucial for accurate diagnosis and treatment planning. Second, it examines the variety of AI approaches used, including both ML techniques and DL architectures. Third, this review addresses applications of classification and segmentation, highlighting the significance of data preprocessing in AI for GI illnesses. This demonstrates the various ways AI may support diagnosis and offer insights at various phases of the diagnostic process.

2 | Methods

2.1 | Literature Search Strategy

We conducted a systematic literature search on using AI for GI disease diagnosis with different imaging modalities, following PRISMA guidelines. By analyzing the application of AI techniques across different imaging modalities, our aim is to study trends in research focus, which modalities are most used for specific gastrointestinal diseases, and whether any studies have begun to explore multimodal approaches. A total of 1801 articles were gathered from PubMed, Google Scholar, IEEE Explore, and Scopus. Boolean searches were performed using terms like "endoscopic colorectal cancer," "gastric cancer endoscopy," "peptic ulcer," "*H. pylori*," "inflammatory bowel disease," "ulcerative colitis," "celiac disease" along with "artificial intelligence," "machine learning," and "deep learning."

Our search covered peer-reviewed publications from January 2016 to December 2024 relevant to gastroenterology. After removing duplicates, we screened 919 unique studies. This search yielded articles from both medical and technical perspectives. By "medical," we refer to articles that include case studies, medical reviews, case reports, and clinical findings. Since our focus

TABLE 2 | Comparisons of the proposed review paper with previously published reviews (2021–2025).

References	Duration	Disease type	# Papers	Image modalities	Processing methods
Tziortziotis et al. (2021)	2016–2021	Crohn's disease, celiac disease, intestinal infections	18	Endoscopy	Classification tasks
Christou and Tsoulfas (2021)	2014–2020	Gastric cancer, Crohn's disease, peptic ulcer	116	Endoscopy, Histology, MRI	Classification tasks
Gubatan et al. (2021)	2015–2020	Inflammatory bowel disease	58	Endoscopy, Histology	Classification and segmentation tasks
Jin et al. (2022)	2018–2020	Gastric cancer, *H. pylori*, precancerous conditions	40	Endoscopy	Pre-processing, classification and segmentation tasks
Jahmunah et al. (2023)	2008–2022	Celiac disease	32	Endoscopy	Pre-processing and classification tasks
Lewis et al. (2024)	2017–2023	Gastric cancer, *H. pylori*, Ulcerative colitis, Celiac disease	33	Endoscopy	Pre-processing, classification and segmentation tasks
Proposed Review	2016–2024	Gastric cancer, gastritis, *H. pylori*, colorectal polyps, Crohn's disease, ulcerative colitis, celiac disease	61	Endoscopy, Histology, CT	Pre-processing, classification and segmentation tasks

is on the role of AI in medical diagnosis, we excluded medical studies that lacked an objective or quantitative analysis. In our screening process, we also excluded reviews, letters, and conference abstracts. This left 114 eligible papers. To ensure the quality and rigor of the research, we further refined our selection by only including articles in top-quartile journals (Q1) based on impact factors. This ensured the quality and rigor of the research we reviewed. After applying this threshold, 76 papers remained for full-text review and data extraction.

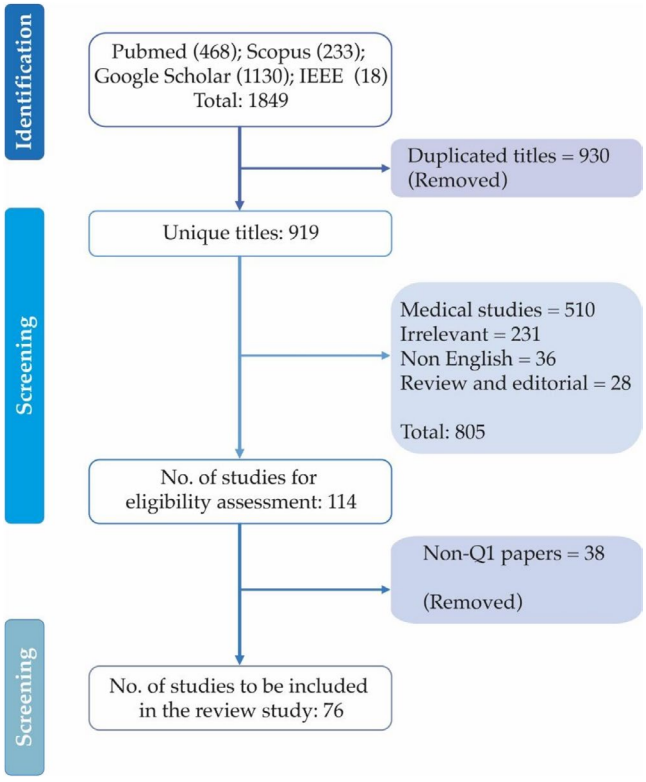

FIGURE 3 | PRISMA flowchart used for the article selection.

Figure 3 shows the PRISMA flowchart followed for the systematic screening process, including the number of articles identified, included, and excluded at each stage.

3 | Results

Artificial intelligence is showing promising applications in modeling and predicting GI diseases. Two main techniques are being used: machine learning and deep learning.

3.1 | Pre-Processing

Image preprocessing is an important first step when applying ML and DL to medical images. It can enhance image quality and improve the performance of subsequent analysis algorithms. For instance, endoscopy images can suffer from issues like air bubbles, uneven exposure, or opaque fluids that can obscure early signs of gastrointestinal diseases. Accuracy is also affected by the physician's skill and experience. To address this, common preprocessing techniques in AI pipelines include adjusting brightness and contrast at the pixel level, geometric transformations, and filtering to restore image details (Jahmunah et al. 2023). The goal is to optimize the visibility of anatomical structures and gastrointestinal features to enable more precise and reliable disease assessment. In another work, Cogan et al. (2019) developed a framework called MAPGI that provides preprocessing tools like cropping, sharpening edges, and boosting contrast through color mapping, scaling, and filtering. Figure 4a shows original and contrast-enhanced endoscopy images using MAPGI—the original pylorus image has very low pixel values, which contrast enhancement corrects. MAPGI then applies filters like low-pass, high-pass, and band-pass specifically to the luminance channel of images to reduce noise while preserving color information (Figure 4b).

In the context of histological images, pre-processing addresses different challenges than those in endoscopy. Since

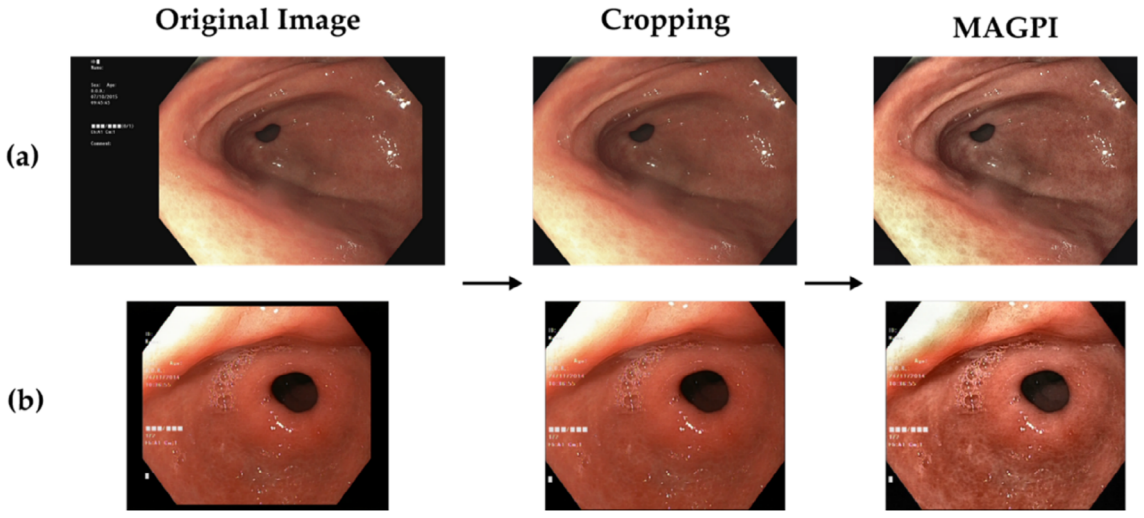

FIGURE 4 | Pre-processing: Effect of the MAGPI filter on endoscopy images. Comparison between the original image (left), cropped image (center), and image with improved contrast (right).

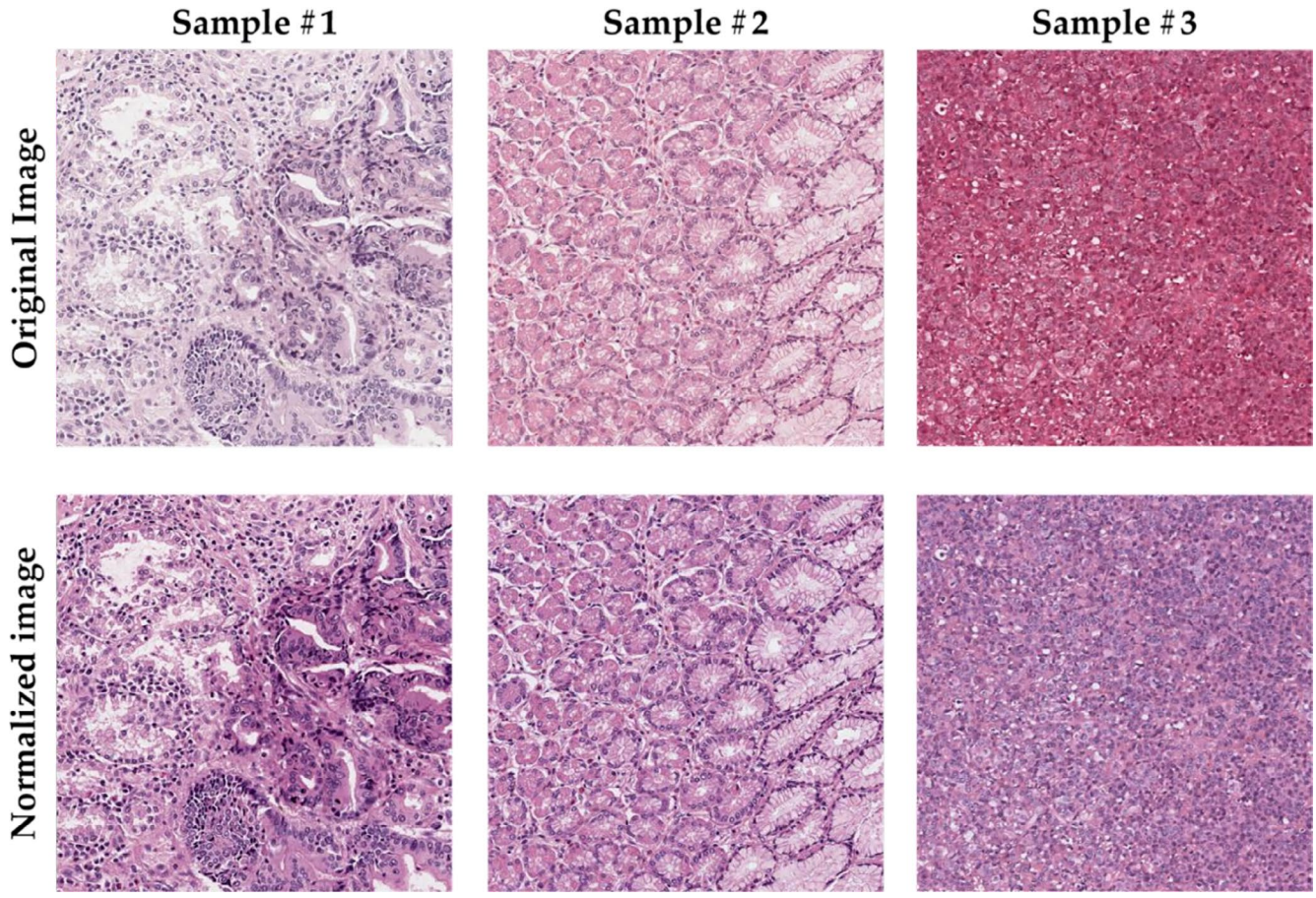

FIGURE 5 | Pre-processing: Effect of stain normalization of histological images. The normalized version (right) demonstrates how such techniques reduce variability, allowing AI models to focus on relevant histological patterns rather than technical artifacts.

histopathology slides are typically stained to highlight cellular components, variations in staining techniques can significantly impact image characteristics and subsequent analysis. Pre-processing tools for histological images are essential for standardizing these variations, thereby improving the consistency of AI tools across different laboratories and institutions. For example, stain normalization techniques (Salvi et al. 2023) are often used to adjust for color variations in histological samples, which can arise due to differences in staining protocols, sample preparation methods, or scanner settings. These normalization approaches ensure that AI algorithms focus on genuine pathological differences rather than technical variations in sample preparation. An example of stain normalization in H&E-stained gastric mucosa is shown in Figure 5, where original slides with distinct color profiles are normalized to a common template, demonstrating the impact of this preprocessing step on visual consistency.

For CT imaging, pre-processing techniques focus primarily on enhancing tissue contrast and reducing noise to improve diagnostic accuracy. Other researchers, such as Chen et al. (2019) have developed techniques to rescale soft tissue intensities in CT images, thereby enhancing contrast to reveal more diagnostic organ details. By optimizing the visibility of anatomical structures, preprocessing facilitates the accurate evaluation and diagnosis of gastrointestinal diseases.

3.2 | Traditional Machine Learning Approaches

ML is a crucial field within AI that draws from computer science, statistics, and biology. The main goal of ML is to uncover patterns in data to gain insights and enable algorithms to learn from experience to improve performance. For GI diseases, ML can identify patterns and risk factors associated with digestive conditions in large datasets. For example, decision trees can determine which symptoms predict specific diseases such as Crohn's or ulcerative colitis. Regression models can also provide personalized risk scores based on individual characteristics and lifestyle, aiding early detection and treatment planning.

This section summarizes common ML approaches for GI diseases. Studies have shown ML can effectively diagnose various GI diseases through endoscopy, CT, and histology images, as outlined in Table A1. The choice of algorithm depends on the data type and objective, as each application requires specific techniques.

3.2.1 | Feature Extraction and Classification

Feature extraction is the most important step in ML pipelines. It transforms raw data into meaningful numerical features that preserve the underlying structure. The goal is to identify and

extract the most salient features from complex datasets, simplifying subsequent analysis and classification.

In GI imaging, feature extraction is vital for quantifying the visual signs of diseases such as *H. pylori* infection from endoscopy images. Important endoscopic features associated with *H. pylori* include gastric atrophy, metaplasia, nodularity, redness, edema, enlarged folds, and sticky mucus (Alhajlah 2022; Glover et al. 2021; Kato 2016). Conversely, features like red streaks, hematin, polyps, and regular vessel patterns indicate the absence of infection. By extracting these visual patterns as numerical features, feature extraction enables objective quantification of disease characteristics.

For example, Zhang et al. (2023) used a gradient boosting tree model for feature extraction and classification of *H. pylori* infection. Thanks to the identification of mucosal features, their model achieved an accuracy of 91.9% in distinguishing infection status while remaining interpretable—important factors for clinician acceptance. With an AUC of 0.96, the proposed model demonstrated high sensitivity (90.3%) and high specificity (93.6%), ensuring accurate identification of *H. pylori* infection while minimizing misdiagnosis and unnecessary treatment. This demonstrates the value of feature extraction in developing effective ML models for GI diagnosis.

In recent years, several authors have focused their studies on celiac disease. This pathology has been increasingly addressed using AI-based techniques, particularly with the analysis of biopsy images. Celiac disease is characterized by the sign of villous atrophy. An automated approach for analyzing biopsy images was introduced by Koh et al. (2021). Their research employed ML techniques to automatically detect and categorize villous atrophy, utilizing a modified version of the Marsh classification system. The authors achieved an accuracy of 88.89%, a sensitivity of 89.7%, and a specificity of 86.7% in classifying two-class villous abnormalities using Hematoxylin and Eosin (H&E) stained biopsy images, demonstrating the potential for automating biopsy image interpretation with AI. Faust et al. (2023) investigated whether ML models could assist in distinguishing normal, celiac disease, and non-celiac duodenitis based on the characteristics of the small intestinal lamina propria. Their SVM model achieved an accuracy of 98.53% in differentiating normal controls from celiac disease, and 98.55% in distinguishing normal controls from non-celiac duodenitis. The proposed model, while maintaining excellent sensitivity (97.7%), also achieved an excellent specificity value (99%). Similarly, but using video capsule endoscopy images, Stoleru et al. (2022) employed an SVM algorithm for automated detection of celiac disease, achieving an accuracy of 94.1%, a sensitivity of 96%, and a precision of 94%.

Other studies have used techniques like random forests (Sucipto et al. 2023) and support vector machines (Mesejo et al. 2016) for feature-based classification of diseases like ulcerative colitis, polyps, and other GI abnormalities from endoscopy images.

3.3 | Deep Learning Approaches

DL has now become the standard paradigm for medical image analysis, and this trend is equally evident in the field of GI imaging. Initially, Convolutional Neural Networks (CNNs) gained widespread adoption thanks to their strong performance across a variety of image-based tasks. More recently, however, the landscape has expanded to include a broader range of architectures and training approaches, each offering specific advantages for GI disease analysis. In the following subsections, we will focus on three key DL applications in GI imaging: training approaches, classification methodologies, and segmentation strategies. The complete list of papers reported in this section is provided in Table A2.

3.3.1 | Training Approaches

Deep learning models require large training datasets, which presents a particular challenge in medical imaging where annotated data is often scarce. Transfer Learning has become a fundamental strategy, leveraging models pre-trained on large datasets from other domains and then fine-tuning them with limited target data. This approach has proven particularly effective in GI imaging applications. Many studies have successfully used transfer learning techniques, such as GastroNet (Yasmin et al. 2023) and other models pre-trained on GI images or general datasets like ImageNet (Recht et al. 2019). This approach has enabled accurate, efficient disease detection models despite limited training data (Caires Silveira et al. 2022; Khan et al. 2022; Zhang et al. 2017).

Foundation models represent the latest advancement in this field. Works by Vorontsov et al. (2024), Wang, Zhao, et al. (2024), and Xu et al. (2024) have demonstrated the potential of large-scale pre-trained models to generalize across multiple cancer types, including colorectal and gastric cancers. These models, typically based on Vision Transformer architectures, show promise in capturing complex tissue patterns and providing more robust diagnostic predictions.

A key driver behind the success of foundation models is self-supervised learning (SSL), which allows models to learn robust visual features from large collections of unlabeled data that can later be fine-tuned on limited labeled examples (Sanderson and Matuszewski 2024). In digital pathology, Vision Transformers pre-trained with SSL techniques (e.g., DINO and MAE) have shown strong performance in histology tasks such as tissue classification and gland segmentation (Chen, Ding, et al. 2024; Yang et al. 2025). Similar trends are emerging in endoscopy, where self-supervised models trained directly on endoscopic video frames have demonstrated improved generalization and reduced reliance on manual annotation (Guo et al. 2024; Liu and Zuo 2023; Nezhad et al. 2025).

3.3.2 | Classification

Deep learning classification approaches have been applied across various GI diseases (Guo et al. 2024), with each pathology benefiting from analysis of different imaging modalities (Horie et al. 2019; Misawa et al. 2021).

In colorectal diseases, polyp detection and classification have seen significant advances across different imaging modalities (Bilal, Tsang, et al. 2023; Huang et al. 2023). In endoscopic analysis,

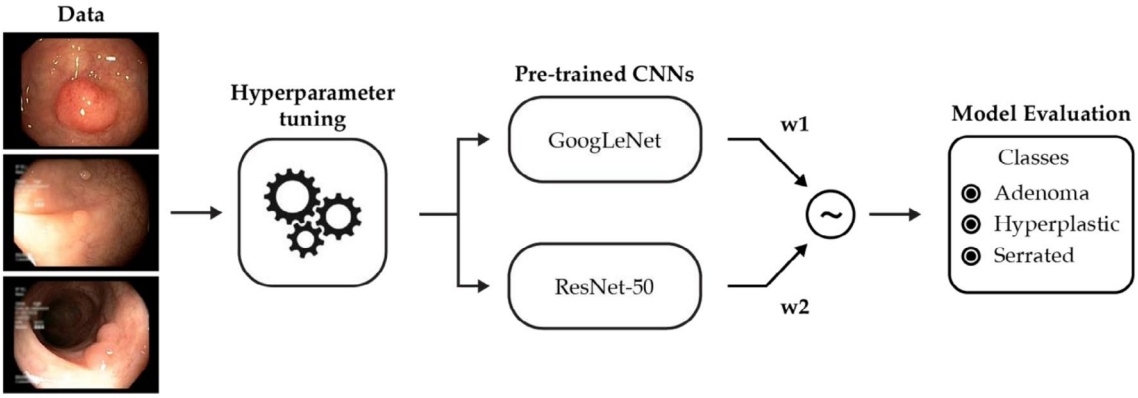

FIGURE 6 | An outline of the weighted-average ensemble classifier for polyp classification in endoscopic images.

Korbar et al. (2017) achieved 93% accuracy using transfer learning with NBI lighting, while (Rahman et al. 2021) reached 98.5% accuracy using a stacking ensemble of Xception, ResNet-101, and VGG-19. Similar approaches by Younas et al. (2023) and Zachariah et al. (2020) consistently exceeded 85%–90% accuracy. For histopathological analysis, Yu et al. (2021) achieved an AUC of 0.972 using Inception V3 on 13,111 whole slide images, while Graham et al. (2023) demonstrated exceptional performance using graph neural networks with an AUC of 0.98. In CT-based detection, Ghosal et al. (2024) achieved an accuracy of 95.83% and AUC of 0.956 using CNN-based approaches, demonstrating the potential of deep learning across all imaging modalities for colorectal disease assessment. Figure 6 illustrates how integrating transfer learning, ensemble modeling, and augmented training examples contributed to improved polyp classification.

In gastric cancer, detection has been significantly improved through the application of DL across different diagnostic modalities. In endoscopic evaluation, Ikenoyama et al. (2021) and Zhou et al. (2023) developed CNN-based systems achieving accuracy rates from 79.5% to 88.3%, demonstrating faster detection times than human endoscopists while maintaining comparable diagnostic accuracy. For histological analysis, (Iwaya et al. 2023) achieved 96.7% accuracy using ResNet-based architectures. Recent multimodal approaches by Chen, Ding, et al. (2024) and Gao et al. (2024) have shown improved prediction by integrating features from both CT and histological images, achieving AUC values ranging from 0.821 to 0.846.

Endoscopic assessment of ulcerative colitis has seen significant advances through works by Fan et al. (2023), Wang et al. (2023), Stidham and Takenaka (2022), and Takabayashi et al. (2024), with accuracy rates typically exceeding 85%. Notably, Qi et al. (2023) introduced UC-former, a Vision Transformer-based model achieving 90.8% accuracy in predicting Mayo endoscopic scores, outperforming senior endoscopists. In histological evaluation, Iacucci et al. (2023) and Kohli et al. (2024) developed models for assessing disease activity from biopsy samples, achieving accuracy rates up to 79% with high AUC values (0.972). For Crohn's disease, Poudel et al. (2020) demonstrated high performance using CNNs, achieving a sensitivity of 92.2% and precision of 86.8%.

Detection of *H. pylori* has primarily focused on endoscopic and histological imaging. Recent endoscopic approaches have demonstrated robust performance, with accuracy rates ranging from 87% to 92% (Ibrahim et al. 2024; Lin, Hsu, et al. 2023; Seo et al. 2023). While Shichijo et al. (2019) achieved a slightly lower accuracy of 81.6%, their study provided important insights into model generalization. In histological analysis, Lin, Chen, et al. (2023) achieved excellent results with a sensitivity of 93.3% and specificity of 90.1% (AUC: 0.97) using a two-tiered deep learning approach.

DL-based methods for celiac disease have shown high performance across different imaging modalities (Syed et al. 2019), (Scheppach et al. 2023). For example, recent work by Wang, Shi, et al. (2024) achieved outstanding results using a hybrid approach combining CNNs and Transformers, reaching an accuracy of 98.38% with high specificity (99.04%) and precision (99.38%) on endoscopic images. For histological analysis, Wei et al. (2019) and Kowsari et al. (2020) demonstrated accuracy rates exceeding 87% using various CNN architectures.

Similarly, the classification of esophageal cancer has greatly benefited from DL architectures. Kumagai et al. (2022) reported strong performance (AUC: 0.92) using Vision Transformers on 8097 histological images, while Horie et al. (2019) achieved 82% accuracy using CNNs on endoscopic images, demonstrating the potential of different architectural approaches in capturing complex tissue patterns.

Recent developments in foundation models have further advanced classification capabilities across all these pathologies. Work by Vorontsov et al. (2024) achieved impressive results across multiple cancer types using a Vision Transformer-based architecture trained on 89,417 whole slide images. Similar advances by Wang, Zhao, et al. (2024) and Xu et al. (2024) have demonstrated the potential of large-scale pre-trained models to generalize across multiple diseases and imaging modalities, suggesting a promising direction for future developments in GI disease classification.

3.3.3 | Segmentation

One important application of deep learning in GI imaging is tissue segmentation, which involves separating an image into distinct regions based on their visual characteristics. Deep learning segmentation models have achieved human-level performance

on various tasks (Inamdar et al. 2023; Salvi et al. 2020), with approaches varying in complexity based on specific clinical needs and imaging modalities (Tang et al. 2023).

At the most basic level, object detection approaches focus on identifying and localizing discrete abnormalities using bounding boxes. Although it lacks pixel-level precision, it enables real-time assistance during endoscopic procedures, particularly for detecting small lesions (< 5 mm) that may otherwise be missed. Models such as GastroNet (Yasmin et al. 2023) have demonstrated strong performance in detecting polyps and other mucosal abnormalities. Gong et al. (2022) showed its value in differentiating early versus advanced gastric cancer using magnifying endoscopy with narrow-band imaging. Similarly, Barash et al. (2021) applied CNN-assisted analysis in capsule endoscopy for Crohn's disease monitoring, achieving a sensitivity of 66.2% and a specificity of 84.6%.

Moving to higher granularity, semantic segmentation assigns a class label to each pixel in the image, enabling precise delineation of tissue types without distinguishing between multiple instances of the same category. Khan et al. (2022) demonstrated this approach by adapting the Mask-RCNN architecture to segment ulcerative lesions. Their model, trained on both private and clinical datasets including images of ulceration, bleeding, polyps, and normal mucosa, effectively delineated pathological tissue regions. In histological analysis, Jiang, Ding, et al. (2023) achieved a Dice score of 78.77% for gastric cancer segmentation using Vision Transformers, demonstrating the potential of newer architectures in this domain.

Instance segmentation advances beyond semantic segmentation by not only classifying each pixel but also distinguishing between different instances of the same class. This capability is particularly valuable in histopathology, where multiple pathological features may coexist in the same tissue section. Lin, Chen, et al. (2023) applied this approach to localize *H. pylori* gastritis in histological slides, achieving an average precision of 57.96%. For colorectal polyp detection, studies by Younas et al. (2023) have demonstrated the effectiveness of instance segmentation in distinguishing between multiple polyps in the same image.

At the highest level of complexity, panoptic segmentation integrates the strengths of semantic and instance segmentation to offer a comprehensive representation that includes both pixel-level classification and instance differentiation. This unified approach is especially advantageous in complex histological analyses, where both the global tissue architecture and individual cellular components have diagnostic relevance. Najdawi et al. (2023) applied this approach to colorectal biopsy slides to identify histological features of ulcerative colitis. Their models achieved strong agreement with expert pathologists, reaching an average accuracy of 97%.

Recent work by Sanderson and Matuszewski (2024) has demonstrated the potential of combining multiple segmentation approaches in a single framework. Their study on self-supervised learning achieved impressive results across different segmentation tasks, with Dice scores ranging from 89.6% to 92.7% and IoU values of 84.4% to 86.7%. This multi-task approach suggests that future segmentation solutions might benefit from combining different levels of granularity within unified frameworks.

The development of robust, automated segmentation solutions through DL continues to be an active area of research, with the potential to standardize disease assessment when validated on large clinical datasets.

4 | Discussion

Research on GI illnesses has substantially advanced in recent years thanks to AI approaches (Gubatan et al. 2021; Jin et al. 2022). Texture, shape, color, and intensity are among the imaging properties that are analyzed in CT, histology, and endoscopy. While other imaging modalities such as MRI show promise for specialized applications, particularly in inflammatory bowel disease evaluation, this review focused on these three primary diagnostic techniques that represent the current standard of care in routine clinical practice. To help clinicians during procedures, quantitative approaches such as texture analysis, shape analysis, and color/intensity analysis can be used to derive significant patterns from in vivo images. Personalized screening, diagnosis, and therapy selection for common gastrointestinal disorders such as IBD, irritable bowel syndrome (IBS), and colon cancer have been demonstrated to be possible with the use of both ML and DL. The ML and DL techniques used to analyze GI pictures are summarized in Tables A1 and A2.

4.1 | Summary of Main Findings

Figure 7 provides insights into annual publication trends for papers reviewed in this study based on their use of ML or DL. DL applications have grown substantially in recent years, peaking in 2023. Most articles with ML approaches were published between 2023 and 2024.

Figure 8 illustrates this trend, showing the distribution of ML and DL studies and imaging modalities across various GI diseases. DL techniques have been more frequently applied in most disease categories, particularly for colorectal polyps, ulcerative colitis, *H. pylori*, and gastric cancer, where DL studies significantly outnumber ML approaches. Several factors contribute to this shift. First, DL architectures are often preferable for image tasks due to their ability to automatically learn features directly from pixels. Second, recent advances in DL models and larger healthcare datasets have accelerated progress across many domains, including GI disease analysis, as evidenced by publication trends. Figure 8 confirms the dominance of endoscopy across most conditions, particularly for *H. pylori* detection, colorectal polyps, and ulcerative colitis. Histological imaging emerges as the second most common modality, with significant usage in celiac disease and ulcerative colitis. CT imaging, while less frequent, shows notable application in colorectal polyps' detection. Our analysis reveals not only which modalities are most commonly used for each condition, but also how they complement each other. Endoscopy is widely used in real-time detection tasks (e.g., colorectal polyps or ulcerative colitis), whereas histological imaging plays a crucial role in confirming diagnoses for conditions such as celiac disease and gastric cancer. This multimodal perspective offers practical guidance for researchers and clinicians in selecting appropriate imaging strategies for specific GI conditions.

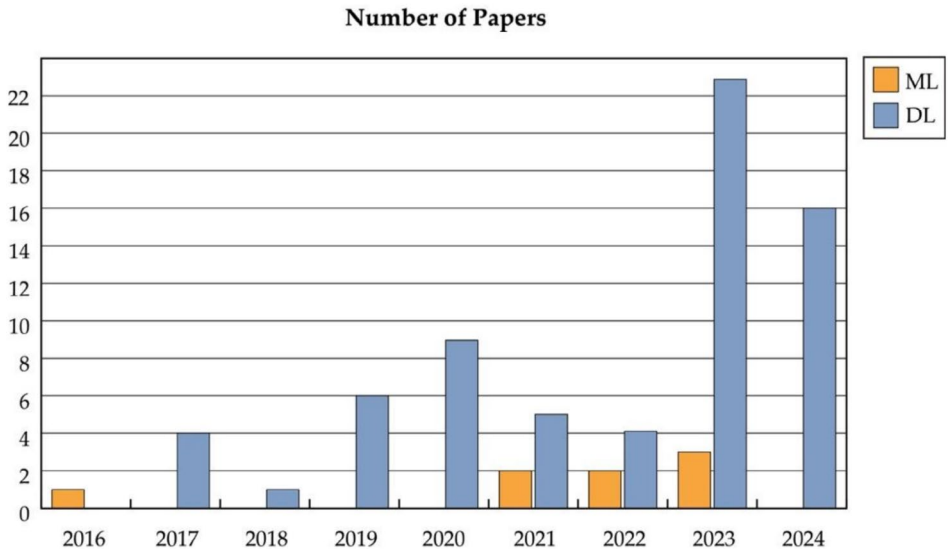

FIGURE 7 | Number of studies conducted on GI with AI each year from 2016 to 2024.

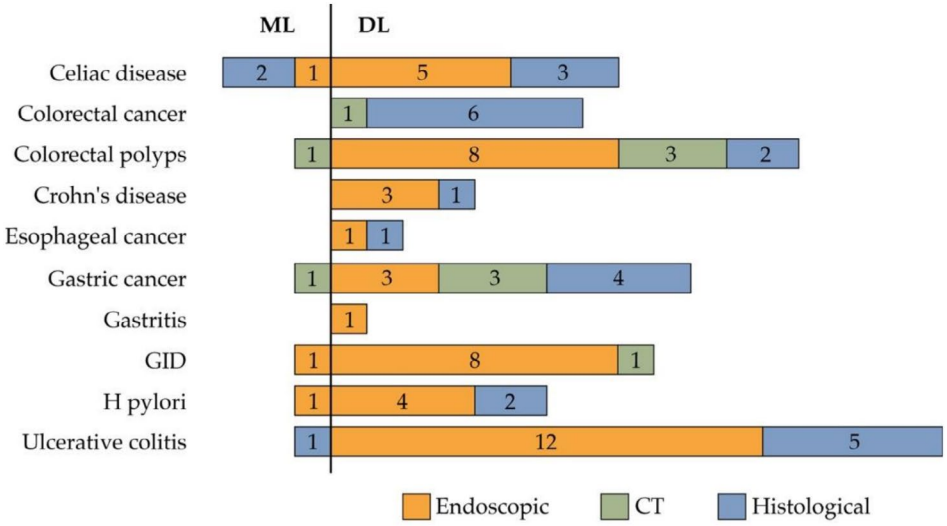

FIGURE 8 | Number of studies per detection target and AI technique. If the study addressed multiple gastrointestinal diseases, it was labeled as "GID" (gastrointestinal diseases).

Figure 9 shows the number of studies on AI for GI disease diagnosis from 2016 to 2024, categorized by imaging modality. The graph reveals an increase in publications over time, indicating a growing interest in applying AI techniques in this domain. Endoscopy-based studies ($n = 44$) constitute the majority, highlighting the prominence of this imaging modality for GI analysis. Histology ($n = 25$) and radiology ($n = 10$) contribute as well, though to a lesser extent compared to endoscopy.

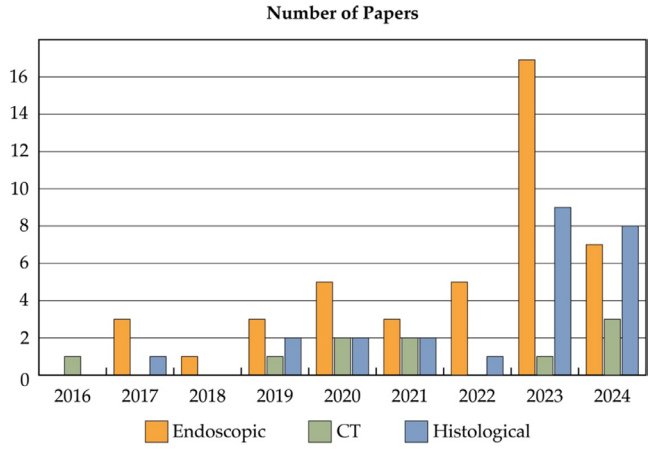

FIGURE 9 | Number of studies conducted on GI from 2016 to 2024, according to imaging technique.

4.2 | Development of ML Methods for GI Disease

Based on the reviewed literature, ML techniques are frequently used for GI disease feature extraction and classification. Typical feature extraction procedures involve examining texture, shape, and color patterns from medical

images such as radiology scans, histology slides, or endoscopy (Koh et al. 2021; Mesejo et al. 2016; Mohammad and Al-Razgan 2022).

For instance, texture analysis makes use of ML algorithms to identify distinctive visual patterns in pictures that could indicate diseases like cancer, polyps, or ulcers, as they exhibit certain textural anomalies. Because tumors can have uneven shapes, shape analysis measures the geometry of lesions or regions of interest. Shape analysis quantifies the geometry of lesions or regions of interest since irregular shapes can correlate with tumors. Color and intensity analysis examines variations in hue, brightness, contrast, and other factors, which can reveal areas of inflammation, bleeding, or abnormal tissue growth. To perform diagnosis predictions or prognostic classifications, the identified patterns are fed into ML classifiers such as random forests, SVMs, or KNN after feature extraction.

Figure 10 summarizes the distribution of ML classifiers applied in the reviewed GI disease studies. SVM emerged as the most used approach, followed by XGBoost and random forests.

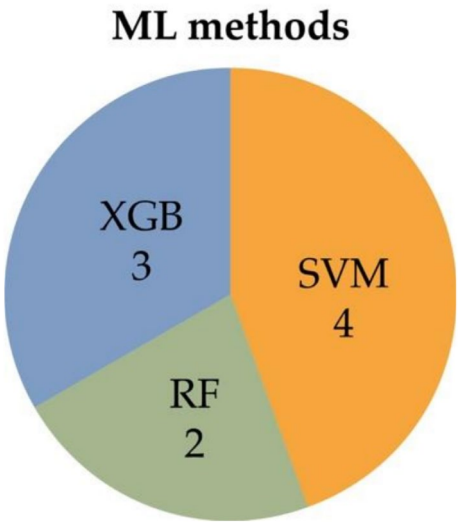

FIGURE 10 | ML methods used for the analysis of GI diseases.

4.3 | Development of DL Methods for GI Disease

The analysis of papers included in this review reveals an evolving landscape in DL architectures for GI disease diagnosis (Figure 11). While CNNs remain the most widely used approach, there is a clear trend toward architectural diversification, particularly with the emergence of Transformers and Graph Neural Networks (GNNs). Generic CNN architectures continue to provide a strong baseline for GI imaging analysis across radiological, histopathological, and endoscopic modalities, as outlined in Table A2. ResNet variants have shown particular success, with their residual connections enabling deeper networks and more robust feature extraction. For instance, Iwaya et al. (2023) achieved 96.7% accuracy in gastric cancer detection using ResNet-based architectures.

However, recent years have seen significant advances in alternative architectures. Vision Transformers have demonstrated impressive capabilities, particularly in capturing long-range dependencies in complex medical images. Notable examples include UC-former achieving 90.8% accuracy in ulcerative colitis evaluation Qi et al. (2023) and Kumagai et al. (2022) reporting strong performance (AUC: 0.92) using Vision Transformers for esophageal cancer diagnosis. Hybrid approaches combining CNNs and Transformers have shown exceptional promise, as demonstrated by Wang, Shi, et al. (2024) reaching 98.38% accuracy in celiac disease diagnosis.

Graph Neural Networks represent another emerging trend, particularly effective in analyzing structural relationships in histopathological images. Graham et al. (2023) achieved an AUC of 0.98 in colorectal cancer detection using GNNs, demonstrating their potential for capturing complex tissue architectures. This architectural diversity, combined with advances in foundation models (Vorontsov et al. 2024), suggests a shift toward more sophisticated approaches that can better handle the complexity of GI disease diagnosis.

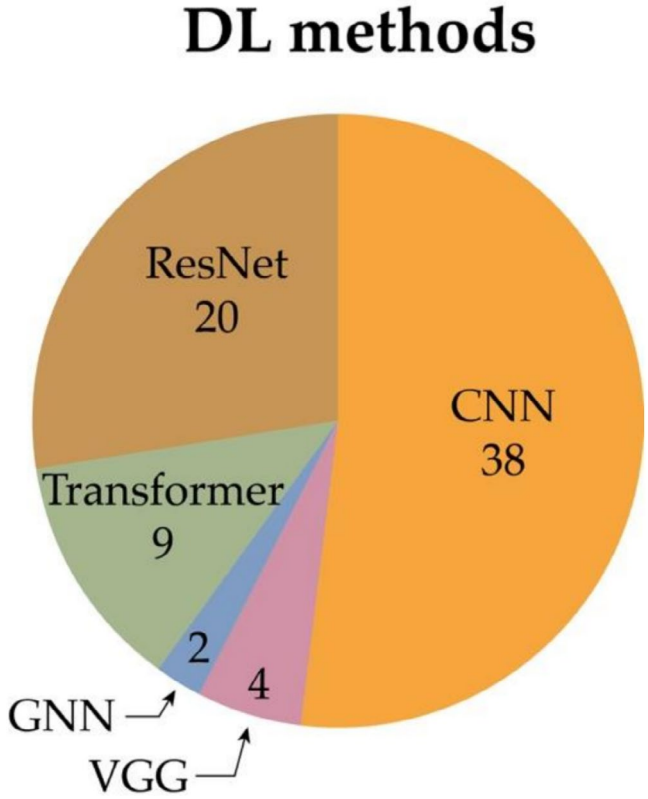

FIGURE 11 | DL methods used for the analysis of GI diseases. CNN refers to where custom or modified architectures were used without naming a standard model. Studies that explicitly mention well-known architectures, such as VGG or ResNet, are listed separately.

Several GI image datasets are now openly accessible to aid in the evaluation and development of DL models for tasks involving classification, segmentation, and diagnosis. All the publicly available datasets used by the reviewed papers, along with the disease classes they cover and their sources, are analyzed and compiled in Table 3. All of them consist of endoscopic images, except for one, which consists of histological images (Hu et al. 2022). They include a variety of conditions, including

TABLE 3 | GI publicly available datasets.

Modality	Year	Dataset	Source	Classes
Endoscopy	2015	CVC-ClinicDB (Bernal et al. 2015)	Hospital Clínic de Barcelona (Spain)	Polyps
Endoscopy	2017	Kvasir (Pogorelov, Randel, and Griwodz 2017)	Vestre Viken—Baerum Hospital (Norway)	Ulcerative colitis, pylorus, the cecum, esophagitis, polyps, colored and lifted polyps, and colored resection margins
Endoscopy	2017	Nerthus (Pogorelov, Randel, De Lange, et al. 2017)	Baerum Hospital Trust (Norway)	Inflammatory bowel disease (BBPS-0, BBPS-1, BBPS2, and BBPS-3)
Endoscopy	2019	EAD2019 (Ali et al. 2019)	Six data centers, including Oxford University Medical Centre (UK); ICL Cancer Institute & Boulogne-Billancourt Hospital (France); Istituto Oncologico Veneto (Italy); University Hospital Vaudois (Switzerland); Botkin Clinical City Hospital (Russia)	Pixel saturations, motion blur, specular reflections, bubbles, and debris.
Endoscopy	2020	HyperKvasir (Borgli et al. 2020)	Vestre Viken—Baerum Hospital (Norway)	Retroflex abdomen, pylorus, short-segment Barrett's, esophagitis A, B through D, ileum terminal, cecum, rectum retroflex, inflammatory bowel disease, ulcerative colitis, hemorrhoids, and polyps.
Endoscopy	2021	Kvasir-Capsule (Smedsrud et al. 2021)	Vestre Viken—Baerum Hospital (Norway)	Ileocecal valve, pylorus, papal ampulla, clean and normal mucosa, diminished mucosa view blood fresh, erythema, erosions, ulcers, polyps, lymphangiectasia, and angi ectasia.
Histology	2021	GasHisSDB (Hu et al. 2022)	Longhua Hospital Shanghai, University of Traditional Chinese Medicine (China)	Gastric cancer
Endoscopy	2023	TMC-UMC (Wang et al. 2023)	Tongji Hospital, Wuhan, Hubei, China	Ulcerative colitis

bleeding throughout the gastrointestinal tract from the esophagus to the colon, polyps, ulcers, and inflammatory bowel disease (Bernal et al. 2015; Borgli et al. 2020; Pogorelov, Randel, De Lange, et al. 2017; Pogorelov, Randel, and Griwodz 2017; Smedsrud et al. 2021). HyperKvasir (Borgli et al. 2020) and Kvasir (Pogorelov, Randel, and Griwodz 2017), two of the most extensive collections, offer multi-class GI images from Norwegian hospitals. Some datasets, such as those on inflammatory bowel disease (Nerthus) (Pogorelov, Randel, De Lange, et al. 2017) and capsule endoscopy (Kvasir-Capsule) (Smedsrud et al. 2021), concentrate on specific applications. GasHisSDB (Hu et al. 2022) is a publicly available dataset designed to support the automated diagnosis of gastric cancer. The emergence of such resources has begun enabling robust DL approaches for advanced GI disease diagnosis.

Additionally, DL has shown promise in identifying neoplastic tissue in early-stage gastric cancer (Bang et al. 2021). Additionally, during endoscopic examinations, clinicians can receive real-time decision support and feedback from DL models. Anatomical landmark-identifying models are able to analyze images as they are taken, pointing out areas like polyps that might need more examination or treatment. The objective of this real-time augmentation is to optimize procedural workflows and concentrate endoscopists' attention on important areas.

Nonetheless, the size and quality of the training data continue to have a significant impact on how well DL techniques perform. If not appropriately addressed, variables such as image resolution, variations in imaging devices, and irregularities in the acquisition protocol can also affect the generalizability of the model.

4.4 | Diagnosis Performance

The analysis of ML and DL models across various GI pathologies reveals significant variations in performance metrics, including sensitivity, specificity, precision, and accuracy (Table 4). While DL models often show high sensitivity and accuracy, the wide ranges in metrics such as specificity and precision highlight the challenges of achieving consistent performance across studies. For instance, diseases like Celiac Disease and Gastric Cancer exhibit substantial differences between studies, suggesting that model performance may be influenced by factors such as dataset size, disease characteristics, and model architecture. These aspects underscore the importance of considering the trade-offs between false positives and false negatives in clinical practice, as they can significantly impact treatment decisions. Furthermore, the absence of certain metrics in some diseases, such as precision for colorectal polyps, points to the need for more comprehensive reporting in future studies.

4.5 | Clinical Implications and Open Challenges

The management of GI diseases may be greatly advanced by the deployment of AI technology. Standardizing diagnosis with AI may reduce variability among practitioners. AI can also improve public health by optimizing screening through better resource allocation and disease risk stratification. But effectively tackling a few major obstacles is essential to incorporating AI into healthcare workflows:

1. *Limited public datasets*: Developing robust AI models demands large, high-quality annotated datasets. Yet, there is a scarcity of public data, particularly in GI histology and radiology. Endoscopic datasets are more common but still lack comprehensive labeling and clinical context. To accelerate progress, we recommend the creation of open-access, curated datasets targeting specific GI diseases. These should include diverse imaging types (e.g., white-light, narrow-band, CT, and histopathology), patient demographics, and disease stages to promote generalizability.

2. *Need for multicenter data*: As highlighted in Table 3, most existing public datasets are derived from a single device or organization, with the exception of EAD2019 (Ali et al. 2019), which focuses on artifacts. This lack of variability, constrained by specific acquisition devices and local protocols, limits the generalization of the model to different real-world scenarios. For example, CT scans from different vendors vary in contrast and resolution, while histology slides may differ in staining quality. Collaborative multicenter initiatives and standardized data collection frameworks are crucial to enable model robustness and external validation.

3. *Image quality and standardization*: As shown for other imaging modalities, digital image standardization could aid in lowering data variability while enhancing the generalization of AI models (Seoni et al. 2024). Specifically, each modality covered in this review presents its own open challenges related to data quality:
   - Endoscopy: Image variability due to lighting, motion artifacts, and device differences can hinder model reliability. AI algorithms trained on high-quality, ideal images may fail in real-time clinical settings. Standardized imaging protocols and preprocessing pipelines (e.g., motion compensation and color normalization) are essential.
   - Histology: Variability in staining techniques, scanner resolution, and section thickness across labs poses a major challenge for histopathological AI. Domain adaptation and stain normalization methods should be integrated into training pipelines to reduce these effects.
   - CT Imaging: differences in acquisition protocols, slice thickness, and contrast usage across centers can impair model transferability. Harmonization approaches, such as image resampling and protocol standardization, are needed to enable broader applicability.

4. *Interpretability and transparency in AI models*: Ensuring interpretability and transparency is essential for building clinician trust and minimizing the risk of biased or unsafe decisions (Loh et al. 2022). High-uncertainty cases must be clearly identified and assessed using explainability techniques (Seoni et al. 2023). To this end, we recommend integrating established interpretability tools such as LIME (Local Interpretable Model-Agnostic Explanations) or SHAP (SHapley Additive exPlanations) to support more transparent clinical decision-making.

TABLE 4 | Range of values between the maximum and minimum of the diagnosis performance of ML and DL models for different diseases.

Disease	Technique	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
Celiac disease	ML ($n=3$)	89.7–97.7	86.7–99	94	88.9–98.5
	DL ($n=8$)	87.3–100	76–100	88–99.4	84–98.4
Colorectal cancer	ML ($n=0$)	—	—	—	—
	DL ($n=7$)	—	—	—	95.8
Colorectal polyps	ML ($n=1$)	72.7	85.9	—	82.4
	DL ($n=13$)	77.7–100	88–99.8	74–99	75.1–99.5
Crohn's disease	ML ($n=0$)	—	—	—	—
	DL ($n=4$)	66.2–92.2	84.6	71.2–86.8	77.1–95.7
Esophageal cancer	ML ($n=0$)	—	—	—	—
	DL ($n=2$)	83–84.8	82–93.9	—	82–91.2
Gastric cancer	ML ($n=1$)	99	—	92.6	93.4
	DL ($n=10$)	58.4–97.7	81.4–94.6	26–97.6	78.8–96.7
Gastritis	ML ($n=0$)	—	—	—	—
	DL ($n=1$)	94.5	94	—	94.2
GID	ML ($n=1$)	99.4	—	99.8	99.8
	DL ($n=9$)	59.6–99.3	76.2–99.42	16–99.3	77.1–99.4
H pylori	ML ($n=1$)	90.3	93.6	0–0	91.9
	DL ($n=6$)	65.7–100	81–90.7	52.9–90.7	81.6–90.2
Ulcerative colitis	ML ($n=1$)	—	—	—	86
	DL ($n=17$)	82.3–98.9	85–99.8	83.2–98.9	64.8–99.5

Note: If the study addressed multiple gastrointestinal diseases, it was labeled as "GID" (gastrointestinal diseases).

5. *Multi-scale and temporal coherence integration*: Current approaches typically operate at a single scale or on isolated acquisitions, overlooking the rich contextual information available across different scales and temporal sequences. Specifically:
   - Endoscopy: Current classification approaches are predominantly trained and validated on individual frames, failing to leverage the temporal continuity available in endoscopic videos. Integrating information across multiple frames of the same video through techniques like temporal majority voting or attention-based frame aggregation could significantly improve classification robustness. This is particularly relevant in clinical scenarios where motion artifacts, varying viewing angles, and temporary occlusions in single frames could lead to misclassification.
   - Histology: Most methods analyze individual tiles or fixed-scale regions independently, without considering the hierarchical nature of tissue organization. There is a pressing need for multi-scale approaches that can effectively aggregate information from tile-level features to whole-slide image scores, while preserving the biological meaning of the hierarchical tissue structure. This could involve developing new mechanisms that learn to weigh the importance of different regions at different magnification levels, or hierarchical architectures that explicitly model the relationships between scales.
   - CT imaging: Most approaches process 2D slices independently or use simple 3D convolutions, without fully exploiting the spatial coherence between consecutive slices. There is a need for methods that can effectively model the full 3D context while maintaining computational efficiency. This could involve developing architectures that explicitly enforce spatial consistency constraints or incorporating anatomical/prior knowledge about the expected 3D structure of GI organs.

Multicenter evaluations offer a more realistic assessment of AI tools in diverse clinical settings, as they allow for the comparison of model performance across different environments. This is particularly important for ensuring that AI models are not overly specialized to a specific institution, but rather generalize well to broader, real-world clinical scenarios. In this context, data harmonization techniques have proven to be powerful tools in mitigating the variability inherent in datasets that contain images acquired using different protocols and devices (Seoni et al. 2024). Techniques such as the MAGPI filter (Cogan et al. 2019) for endoscopy, stain normalization (Salvi et al. 2021) for histology, and grayscale normalization for CT scans are just a few examples of how harmonization can improve data consistency across multiple sites.

While most studies focus on single imaging modalities, recent research has begun exploring multimodal approaches that

combine different imaging techniques. Initial work by Jiang, Luo, et al. (2023) applied separate CNNs to endoscopic and histologic images for distinct tasks, demonstrating the feasibility of cross-modal analysis. Building on this foundation, Chen, Chen, et al. (2024) and Gao et al. (2024) achieved improved gastric cancer prediction by integrating features from both CT and histological images, with AUC values ranging from 0.821 to 0.846. This multimodal perspective is particularly promising as it mirrors the clinical diagnostic process, where different imaging modalities provide complementary information.

However, effectively combining information across modalities remains challenging, requiring specific architectural designs and consideration of each modality's characteristics. Future directions should focus on developing more integrated multimodal approaches that can simultaneously process and combine information from different imaging sources. Such AI-based multimodality techniques may enhance diagnostic robustness, reduce uncertainty, and offer more comprehensive decision support to increase the clinical applicability of AI models in GI diagnosis.

5 | Conclusion

This systematic review demonstrates the significant potential of AI, particularly DL approaches, in enhancing the analysis of GI diseases across endoscopy, histology, and CT imaging modalities. Our findings reveal that DL-based methods obtained the highest diagnostic accuracies, with performance metrics frequently exceeding 90% across multiple disease categories. The integration of these technologies shows promise for improving early detection of conditions like gastric cancer and colorectal polyps, potentially transforming clinical practice through more standardized, objective assessment tools. Pre-processing techniques have proven crucial for optimizing model performance, while classification and segmentation tasks have emerged as the primary applications of AI in GI disease analysis. However, significant challenges remain to be addressed before widespread clinical implementation. These include the need for larger multicenter datasets, standardized image acquisition protocols, and interpretable AI models that instill confidence in clinicians. Future research should focus on creating transparent and explainable AI systems validated through rigorous multicenter trials to establish real-world clinical utility.

Author Contributions

Sameena Pathan: data curation (equal), writing – original draft (equal). **Alen Shahini:** data curation (equal), visualization (equal). **Massimo Salvi:** supervision (equal), writing – original draft (equal). **Tanweer Ali:** supervision (supporting), writing – review and editing (supporting). **Jovita Relasha Lewis:** data curation (supporting), writing – review and editing (supporting). **Filippo Molinari:** methodology (supporting), writing – review and editing (supporting). **U. Rajendra Acharya:** conceptualization (lead), supervision (supporting), writing – review and editing (supporting).

Acknowledgments

Open access publishing facilitated by Politecnico di Torino, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data is available from corresponding author upon reasonable request.

Related WIREs Articles

A survey on artificial intelligence in histopathology image analysis

A Review on Medical Image Segmentation: Datasets, Technical Models, Challenges and Solutions

References

Alagappan, M., J. R. G. Brown, Y. Mori, and T. M. Berzin. 2018. "Artificial Intelligence in Gastrointestinal Endoscopy: The Future is Almost Here." *World Journal of Gastrointestinal Endoscopy* 10, no. 10: 239–249. https://doi.org/10.4253/wjge.v10.i10.239.

Alhajlah, M. 2022. "Robust Ulcer Classification: Contrast and Illumination Invariant Approach." *Diagnostics* 12, no. 12: 2898. https://doi.org/10.3390/diagnostics12122898.

Ali, S., F. Zhou, C. Daul, et al. 2019. "Endoscopy Artifact Detection (EAD 2019) Challenge Dataset." https://doi.org/10.48550/ARXIV.1905.03209.

Bang, C. S., J. Y. Ahn, J.-H. Kim, Y.-I. Kim, I. J. Choi, and W. G. Shin. 2021. "Establishing Machine Learning Models to Predict Curative Resection in Early Gastric Cancer With Undifferentiated Histology: Development and Usability Study." *Journal of Medical Internet Research* 23, no. 4: e25053. https://doi.org/10.2196/25053.

Barash, Y., L. Azaria, S. Soffer, et al. 2021. "Ulcer Severity Grading in Video Capsule Images of Patients With Crohn's Disease: An Ordinal Neural Network Solution." *Gastrointestinal Endoscopy* 93, no. 1: 187–192. https://doi.org/10.1016/j.gie.2020.05.066.

Bernal, J., F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño. 2015. "WM-DOVA Maps for Accurate Polyp Highlighting in Colonoscopy: Validation vs. Saliency Maps From Physicians." *Computerized Medical Imaging and Graphics* 43: 99–111. https://doi.org/10.1016/j.compmedimag.2015.02.007.

Bilal, M., M. Nimir, D. Snead, G. S. Taylor, and N. Rajpoot. 2023. "Role of AI and Digital Pathology for Colorectal Immuno-Oncology." *British Journal of Cancer* 128, no. 1: 3–11. https://doi.org/10.1038/s41416-022-01986-1.

Bilal, M., Y. W. Tsang, M. Ali, et al. 2023. "Development and Validation of Artificial Intelligence-Based Prescreening of Large-Bowel Biopsies Taken in the UK and Portugal: A Retrospective Cohort Study." *Lancet Digital Health* 5, no. 11: e786–e797. https://doi.org/10.1016/S2589-7500(23)00148-6.

Borgli, H., V. Thambawita, P. H. Smedsrud, et al. 2020. "HyperKvasir, a Comprehensive Multi-Class Image and Video Dataset for Gastrointestinal Endoscopy." *Scientific Data* 7, no. 1: 283. https://doi.org/10.1038/s41597-020-00622-y.

Brizio, A., V. Faure, and F. Baudino. 2022. "Medical Semiotics in the Telemedicine Era: The Birth of Telesemiotics." *International Journal of Medical Informatics* 157: 104573. https://doi.org/10.1016/j.ijmedinf.2021.104573.

Caires Silveira, E., C. F. Santos Corrêa, L. Madureira Silva, B. Almeida Santos, S. Mattos Pretti, and F. Freire De Melo. 2022. "Recognition of Esophagitis in Endoscopic Images Using Transfer Learning." *World Journal of Gastrointestinal Endoscopy* 14, no. 5: 311–319. https://doi.org/10.4253/wjge.v14.i5.311.

Charfi, S., and M. El Ansari. 2017. "Computer-Aided Diagnosis System for Ulcer Detection in Wireless Capsule Endoscopy Videos." In *2017*

International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 1–5. IEEE. https://doi.org/10.1109/ATSIP.2017.8075590.

Chen, R. J., T. Ding, M. Y. Lu, et al. 2024. "Towards a General-Purpose Foundation Model for Computational Pathology." *Nature Medicine* 30, no. 3: 850–862. https://doi.org/10.1038/s41591-024-02857-3.

Chen, T., S. Liu, Y. Li, et al. 2019. "Developed and Validated a Prognostic Nomogram for Recurrence-Free Survival After Complete Surgical Resection of Local Primary Gastrointestinal Stromal Tumors Based on Deep Learning." *eBioMedicine* 39: 272–279. https://doi.org/10.1016/j.ebiom.2018.12.028.

Chen, Z., Y. Chen, Y. Sun, et al. 2024. "Predicting Gastric Cancer Response to Anti-HER2 Therapy or Anti-HER2 Combined Immunotherapy Based on Multi-Modal Data." *Signal Transduction and Targeted Therapy* 9, no. 1: 222. https://doi.org/10.1038/s41392-024-01932-y.

Christou, C. D., and G. Tsoulfas. 2021. "Challenges and Opportunities in the Application of Artificial Intelligence in Gastroenterology and Hepatology." *World Journal of Gastroenterology* 27, no. 37: 6191–6223. https://doi.org/10.3748/wjg.v27.i37.6191.

Cogan, T., M. Cogan, and L. Tamil. 2019. "MAPGI: Accurate Identification of Anatomical Landmarks and Diseased Tissue in Gastrointestinal Tract Using Deep Learning." *Computers in Biology and Medicine* 111: 103351. https://doi.org/10.1016/j.compbiomed.2019.103351.

Dlamini, Z., F. Z. Francies, R. Hull, and R. Marima. 2020. "Artificial Intelligence (AI) and Big Data in Cancer and Precision Oncology." *Computational and Structural Biotechnology Journal* 18: 2300–2311. https://doi.org/10.1016/j.csbj.2020.08.019.

Fan, Y., R. Mu, H. Xu, et al. 2023. "Novel Deep Learning–Based Computer-Aided Diagnosis System for Predicting Inflammatory Activity in Ulcerative Colitis." *Gastrointestinal Endoscopy* 97, no. 2: 335–346. https://doi.org/10.1016/j.gie.2022.08.015.

Faust, O., S. De Michele, J. E. Koh, et al. 2023. "Automated Analysis of Small Intestinal Lamina Propria to Distinguish Normal, Celiac Disease, and Non-Celiac Duodenitis Biopsy Images." *Computer Methods and Programs in Biomedicine* 230: 107320. https://doi.org/10.1016/j.cmpb.2022.107320.

Gao, P., Q. Xiao, H. Tan, et al. 2024. "Interpretable Multi-Modal Artificial Intelligence Model for Predicting Gastric Cancer Response to Neoadjuvant Chemotherapy." *Cell Reports Medicine* 5, no. 12: 101848. https://doi.org/10.1016/j.xcrm.2024.101848.

Ghosal, S., D. Das, J. K. Rai, A. S. Pandaw, and S. Verma. 2024. "iScan: Detection of Colorectal Cancer From CT Scan Images Using Deep Learning." *ACM Transactions on Computing for Healthcare* 5, no. 3: 1–22. https://doi.org/10.1145/3676282.

Glover, B., J. Teare, and N. Patel. 2021. "Assessment of *Helicobacter pylori* Status by Examination of Gastric Mucosal Patterns: Diagnostic Accuracy of White-Light Endoscopy and Narrow-Band Imaging." *BMJ Open Gastroenterology* 8, no. 1: e000608. https://doi.org/10.1136/bmjgast-2021-000608.

Gong, L., M. Wang, L. Shu, et al. 2022. "Automatic Captioning of Early Gastric Cancer Using Magnification Endoscopy With Narrow-Band Imaging." *Gastrointestinal Endoscopy* 96, no. 6: 929–942.e6. https://doi.org/10.1016/j.gie.2022.07.019.

Graham, S., F. Minhas, M. Bilal, et al. 2023. "Screening of Normal Endoscopic Large Bowel Biopsies With Interpretable Graph Learning: A Retrospective Study." *Gut* 72, no. 9: 1709–1721. https://doi.org/10.1136/gutjnl-2023-329512.

Gubatan, J., S. Levitte, A. Patel, T. Balabanis, M. T. Wei, and S. R. Sinha. 2021. "Artificial Intelligence Applications in Inflammatory Bowel Disease: Emerging Technologies and Future Directions." *World Journal of Gastroenterology* 27, no. 17: 1920–1935. https://doi.org/10.3748/wjg.v27.i17.1920.

Guo, H., S. A. Somayajula, R. Hosseini, and P. Xie. 2024. "Improving Image Classification of Gastrointestinal Endoscopy Using Curriculum Self-Supervised Learning." *Scientific Reports* 14, no. 1: 6100. https://doi.org/10.1038/s41598-024-53955-8.

Horie, Y., T. Yoshio, K. Aoyama, et al. 2019. "Diagnostic Outcomes of Esophageal Cancer by Artificial Intelligence Using Convolutional Neural Networks." *Gastrointestinal Endoscopy* 89, no. 1: 25–32. https://doi.org/10.1016/j.gie.2018.07.037.

Horiuchi, Y., T. Hirasawa, N. Ishizuka, et al. 2020. "Performance of a Computer-Aided Diagnosis System in Diagnosing Early Gastric Cancer Using Magnifying Endoscopy Videos With Narrow-Band Imaging (With Videos)." *Gastrointestinal Endoscopy* 92, no. 4: 856–865. https://doi.org/10.1016/j.gie.2020.04.079.

Hossain, T., F. M. J. M. Shamrat, X. Zhou, et al. 2024. "Development of a Multi-Fusion Convolutional Neural Network (MF-CNN) for Enhanced Gastrointestinal Disease Diagnosis in Endoscopy Image Analysis." *PeerJ Computer Science* 10: e1950. https://doi.org/10.7717/peerj-cs.1950.

Hu, W., C. Li, X. Li, et al. 2022. "GasHisSDB: A New Gastric Histopathology Image Dataset for Computer Aided Diagnosis of Gastric Cancer." *Computers in Biology and Medicine* 142: 105207. https://doi.org/10.1016/j.compbiomed.2021.105207.

Huang, J., S. N. Saw, T. He, W. Feng, and C. K. Loo. 2023. "A Masked Graph Neural Network Model for Real-Time Gastric Polyp Detection in Healthcare 4.0." *Journal of Industrial Information Integration* 34: 100467. https://doi.org/10.1016/j.jii.2023.100467.

Huang, S., J. Yang, S. Fong, and Q. Zhao. 2020. "Artificial Intelligence in Cancer Diagnosis and Prognosis: Opportunities and Challenges." *Cancer Letters* 471: 61–71. https://doi.org/10.1016/j.canlet.2019.12.007.

Hwang, Y.-J., N. Kim, H. S. Lee, et al. 2018. "Reversibility of Atrophic Gastritis and Intestinal Metaplasia After *Helicobacter Pylori* Eradication - a Prospective Study for up to 10 Years." *Alimentary Pharmacology & Therapeutics* 47, no. 3: 380–390. https://doi.org/10.1111/apt.14424.

Iacucci, M., T. L. Parigi, R. del Amor, et al. 2023. "Artificial Intelligence Enabled Histological Prediction of Remission or Activity and Clinical Outcomes in Ulcerative Colitis." *Gastroenterology* 164, no. 7: 1180–1188. https://doi.org/10.1053/j.gastro.2023.02.031.

Ibrahim, A. U., F. Dirilenoğlu, U. P. Hacisalihoğlu, A. Ilhan, and O. Mirzaei. 2024. "Classification of *H. pylori* Infection From Histopathological Images Using Deep Learning." *Journal of Imaging Informatics in Medicine* 37, no. 3: 1177–1186. https://doi.org/10.1007/s10278-024-01021-0.

Ikenoyama, Y., T. Hirasawa, M. Ishioka, et al. 2021. "Detecting Early Gastric Cancer: Comparison Between the Diagnostic Ability of Convolutional Neural Networks and Endoscopists." *Digestive Endoscopy* 33, no. 1: 141–150. https://doi.org/10.1111/den.13688.

Inamdar, M. A., U. Raghavendra, A. Gudigar, et al. 2023. "A Novel Attention-Based Model for Semantic Segmentation of Prostate Glands Using Histopathological Images." *IEEE Access* 11: 108982–108994. https://doi.org/10.1109/ACCESS.2023.3321273.

Iwaya, M., Y. Hayashi, Y. Sakai, et al. 2023. "Artificial Intelligence for Evaluating the Risk of Gastric Cancer: Reliable Detection and Scoring of Intestinal Metaplasia With Deep Learning Algorithms." *Gastrointestinal Endoscopy* 98, no. 6: 925–933. https://doi.org/10.1016/j.gie.2023.06.056.

Jahmunah, V., J. En Wei Koh, V. K. Sudarshan, et al. 2023. "Endoscopy, Video Capsule Endoscopy, and Biopsy for Automated Celiac Disease Detection: A Review." *Biocybernetics and Biomedical Engineering* 43, no. 1: 82–108. https://doi.org/10.1016/j.bbe.2022.12.002.

Jiang, X., Y. Ding, M. Liu, Y. Wang, Y. Li, and Z. Wu. 2023. "BiFTransNet: A Unified and Simultaneous Segmentation Network for Gastrointestinal Images of CT & MRI." *Computers in Biology and Medicine* 165: 107326. https://doi.org/10.1016/j.compbiomed.2023.107326.

Jiang, X., X. Luo, Q. Nan, Y. Ye, Y. Miao, and J. Miao. 2023. "Application of Deep Learning in the Diagnosis and Evaluation of Ulcerative Colitis Disease Severity." *Therapeutic Advances in Gastroenterology* 16: 17562848231215579. https://doi.org/10.1177/17562848231215579.

Jin, Z., T. Gan, P. Wang, et al. 2022. "Deep Learning for Gastroscopic Images: Computer-Aided Techniques for Clinicians." *Biomedical Engineering Online* 21, no. 1: 12. https://doi.org/10.1186/s12938-022-00979-8.

Kato, M. 2016. "Endoscopic Findings of *H. pylori* Infection." In *Helicobacter pylori*, edited by H. Suzuki, R. Warren, and B. Marshall, 157–167. Springer Japan. https://doi.org/10.1007/978-4-431-55705-0_10.

Khan, M. A., K. Muhammad, S.-H. Wang, et al. 2022. "Gastrointestinal Diseases Recognition: A Framework of Deep Neural Network and Improved Moth-Crow Optimization With DCCA Fusion." *Human-centric Computing and Information Sciences* 12: 325–345. https://doi.org/10.22967/HCIS.2022.12.025.

Khan, M. A., M. Rashid, M. Sharif, K. Javed, and T. Akram. 2019. "Classification of Gastrointestinal Diseases of Stomach From WCE Using Improved Saliency-Based Method and Discriminant Features Selection." *Multimedia Tools and Applications* 78, no. 19: 27743–27770. https://doi.org/10.1007/s11042-019-07875-9.

Kim, J. E., Y. H. Choi, Y. C. Lee, et al. 2023. "Deep Learning Model for Distinguishing Mayo Endoscopic Subscore 0 and 1 in Patients With Ulcerative Colitis." *Scientific Reports* 13, no. 1: 11351. https://doi.org/10.1038/s41598-023-38206-6.

Koh, J. E. W., S. De Michele, V. K. Sudarshan, et al. 2021. "Automated Interpretation of Biopsy Images for the Detection of Celiac Disease Using a Machine Learning Approach." *Computer Methods and Programs in Biomedicine* 203: 106010. https://doi.org/10.1016/j.cmpb.2021.106010.

Kohli, K., A. Khan, F. Müller, et al. 2024. "P357 Automated Assessment of Histological Disease Activity in Ulcerative Colitis Using a Deep Learning Algorithm." *Journal of Crohn's and Colitis* 18, no. Supplement_1: i761–i762. https://doi.org/10.1093/ecco-jcc/jjad212.0487.

Komeda, Y., H. Handa, T. Watanabe, et al. 2017. "Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience." *Oncology* 93, no. Suppl. 1: 30–34. https://doi.org/10.1159/000481227.

Korbar, B., A. M. Olofson, A. P. Miraflor, et al. 2017. "Deep Learning for Classification of Colorectal Polyps on Whole-Slide Images." *Journal of Pathology Informatics* 8, no. 1: 30. https://doi.org/10.4103/jpi.jpi_34_17.

Kowsari, K., R. Sali, L. Ehsan, et al. 2020. "HMIC: Hierarchical Medical Image Classification, A Deep Learning Approach." *Information* 11, no. 6: 318. https://doi.org/10.3390/info11060318.

Kumagai, Y., K. Takubo, T. Sato, et al. 2022. "AI Analysis and Modified Type Classification for Endocytoscopic Observation of Esophageal Lesions." *Diseases of the Esophagus* 35, no. 9: doac010. https://doi.org/10.1093/dote/doac010.

Lewis, J. R., S. Pathan, P. Kumar, and C. C. Dias. 2024. "AI in Endoscopic Gastrointestinal Diagnosis: A Systematic Review of Deep Learning and Machine Learning Techniques." *IEEE Access* 12: 163764–163786. https://doi.org/10.1109/ACCESS.2024.3483432.

Li, H., Z. Wang, Z. Guan, et al. 2024. "UCFN Net: Ulcerative Colitis Evaluation Based on Fine-Grained Lesion Learner and Noise Suppression Gating." *Computer Methods and Programs in Biomedicine* 247: 108080. https://doi.org/10.1016/j.cmpb.2024.108080.

Lin, C.-H., P.-I. Hsu, C.-D. Tseng, et al. 2023. "Application of Artificial Intelligence in Endoscopic Image Analysis for the Diagnosis of a Gastric Cancer Pathogen-*Helicobacter pylori* Infection." *Scientific Reports* 13, no. 1: 13380. https://doi.org/10.1038/s41598-023-40179-5.

Lin, Y., C. Chen, C. Lee, C. Yeh, and Y. Jeng. 2023. "Two-Tiered Deep-Learning-Based Model for Histologic Diagnosis of *Helicobacter* Gastritis." *Histopathology* 83, no. 5: 771–781. https://doi.org/10.1111/his.15018.

Liu, Y., and S. Zuo. 2023. "Self-Supervised Monocular Depth Estimation for Gastrointestinal Endoscopy." *Computer Methods and Programs in Biomedicine* 238: 107619. https://doi.org/10.1016/j.cmpb.2023.107619.

Loh, H. W., C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya. 2022. "Application of Explainable Artificial Intelligence for Healthcare: A Systematic Review of the Last Decade (2011–2022)." *Computer Methods and Programs in Biomedicine* 226: 107161. https://doi.org/10.1016/j.cmpb.2022.107161.

Malik, H., A. Naeem, A. Sadeghi-Niaraki, R. A. Naqvi, and S.-W. Lee. 2024. "Multi-Classification Deep Learning Models for Detection of Ulcerative Colitis, Polyps, and Dyed-Lifted Polyps Using Wireless Capsule Endoscopy Images." *Complex & Intelligent Systems* 10, no. 2: 2477–2497. https://doi.org/10.1007/s40747-023-01271-5.

Mesejo, P., D. Pizarro, A. Abergel, et al. 2016. "Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy." *IEEE Transactions on Medical Imaging* 35, no. 9: 2051–2063. https://doi.org/10.1109/TMI.2016.2547947.

Misawa, M., S. Kudo, Y. Mori, et al. 2021. "Current Status and Future Perspective on Artificial Intelligence for Lower Endoscopy." *Digestive Endoscopy* 33, no. 2: 273–284. https://doi.org/10.1111/den.13847.

Mohammad, F., and M. Al-Razgan. 2022. "Deep Feature Fusion and Optimization-Based Approach for Stomach Disease Classification." *Sensors* 22, no. 7: 2801. https://doi.org/10.3390/s22072801.

Morley, J., C. C. V. Machado, C. Burr, et al. 2020. "The Ethics of AI in Health Care: A Mapping Review." *Social Science & Medicine* 260: 113172. https://doi.org/10.1016/j.socscimed.2020.113172.

Najdawi, F., K. Sucipto, P. Mistry, et al. 2023. "Artificial Intelligence Enables Quantitative Assessment of Ulcerative Colitis Histology." *Modern Pathology* 36, no. 6: 100124. https://doi.org/10.1016/j.modpat.2023.100124.

Naz, J., M. Sharif, M. Yasmin, M. Raza, and M. A. Khan. 2021. "Detection and Classification of Gastrointestinal Diseases Using Machine Learning." *Current Medical Imaging Formerly Current Medical Imaging Reviews* 17, no. 4: 479–490. https://doi.org/10.2174/1573405616666200928144626.

Nezhad, S. A., G. Tajeddin, T. Khatibi, and M. Sohrabi. 2025. "Self-Supervised Learning Framework for Efficient Classification of Endoscopic Images Using Pretext Tasks." *PLoS One* 20, no. 5: e0322028. https://doi.org/10.1371/journal.pone.0322028.

Ozawa, T., S. Ishihara, M. Fujishiro, Y. Kumagai, S. Shichijo, and T. Tada. 2020. "Automated Endoscopic Detection and Classification of Colorectal Polyps Using Convolutional Neural Networks." *Therapeutic Advances in Gastroenterology* 13: 1756284820910659. https://doi.org/10.1177/1756284820910659.

Patino-Barrientos, S., D. Sierra-Sosa, B. Garcia-Zapirain, C. Castillo-Olea, and A. Elmaghraby. 2020. "Kudo's Classification for Colon Polyps Assessment Using a Deep Learning Approach." *Applied Sciences* 10, no. 2: 501. https://doi.org/10.3390/app10020501.

Paul, D., G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade. 2021. "Artificial Intelligence in Drug Discovery and Development." *Drug Discovery Today* 26, no. 1: 80–93. https://doi.org/10.1016/j.drudis.2020.10.010.

Pogorelov, K., K. R. Randel, T. De Lange, et al. 2017. "Nerthus: A Bowel Preparation Quality Video Dataset." In *Proceedings of the 8th ACM on Multimedia Systems Conference*, 170–174. ACM. https://doi.org/10.1145/3083187.3083216.

Pogorelov, K., K. R. Randel, and C. Griwodz. 2017. "Kvasir: A Multi-Class Image-Dataset for Computer Aided Gastrointestinal Disease Detection." In *ACM Multimedia Systems (MMSys '17)*. ACM.

Polat, G., H. T. Kani, I. Ergenc, Y. Ozen Alahdab, A. Temizel, and O. Atug. 2023. "Improving the Computer-Aided Estimation of Ulcerative Colitis Severity According to Mayo Endoscopic Score by Using

Regression-Based Deep Learning." *Inflammatory Bowel Diseases* 29, no. 9: 1431–1439. https://doi.org/10.1093/ibd/izac226.

Poudel, S., Y. J. Kim, D. M. Vo, and S.-W. Lee. 2020. "Colorectal Disease Classification Using Efficiently Scaled Dilation in Convolutional Neural Network." *IEEE Access* 8: 99227–99238. https://doi.org/10.1109/ACCESS.2020.2996770.

Qi, J., G. Ruan, Y. Ping, et al. 2023. "Development and Validation of a Deep Learning-Based Approach to Predict the Mayo Endoscopic Score of Ulcerative Colitis." *Therapeutic Advances in Gastroenterology* 16: 17562848231170945. https://doi.org/10.1177/17562848231170945.

Rahman, M. M., M. A. H. Wadud, and M. M. Hasan. 2021. "Computerized Classification of Gastrointestinal Polyps Using Stacking Ensemble of Convolutional Neural Network." *Informatics in Medicine Unlocked* 24: 100603. https://doi.org/10.1016/j.imu.2021.100603.

Recht, B., R. Roelofs, L. Schmidt, and V. Shankar. 2019. "Do ImageNet Classifiers Generalize to ImageNet?" In *Proceedings of the 36th International Conference on Machine Learning*, 5389–5400. PMLR. https://proceedings.mlr.press/v97/recht19a.html.

Rymarczyk, D., W. Schultz, A. Borowa, et al. 2024. "Deep Learning Models Capture Histological Disease Activity in Crohn's Disease and Ulcerative Colitis With High Fidelity." *Journal of Crohn's and Colitis* 18, no. 4: 604–614. https://doi.org/10.1093/ecco-jcc/jjad171.

Saito, H., T. Aoki, K. Aoyama, et al. 2020. "Automatic Detection and Classification of Protruding Lesions in Wireless Capsule Endoscopy Images Based on a Deep Convolutional Neural Network." *Gastrointestinal Endoscopy* 92, no. 1: 144–151. https://doi.org/10.1016/j.gie.2020.01.054.

Salvi, M., A. Caputo, D. Balmativola, et al. 2023. "Impact of Stain Normalization on Pathologist Assessment of Prostate Cancer: A Comparative Study." *Cancers* 15, no. 5: 1503. https://doi.org/10.3390/cancers15051503.

Salvi, M., A. Mogetta, K. M. Meiburger, et al. 2020. "Karpinski Score Under Digital Investigation: A Fully Automated Segmentation Algorithm to Identify Vascular and Stromal Injury of Donors' Kidneys." *Electronics* 9, no. 10: 1644. https://doi.org/10.3390/electronics9101644.

Salvi, M., F. Molinari, S. Iussich, et al. 2021. "Histopathological Classification of Canine Cutaneous Round Cell Tumors Using Deep Learning: A Multi-Center Study." *Frontiers in Veterinary Science* 8: 640944. https://doi.org/10.3389/fvets.2021.640944.

Sanderson, E., and B. J. Matuszewski. 2024. "A Study on Self-Supervised Pretraining for Vision Problems in Gastrointestinal Endoscopy." *IEEE Access* 12: 46181–46201. https://doi.org/10.1109/ACCESS.2024.3381517.

Scheppach, M. W., D. Rauber, J. Stallhofer, et al. 2023. "Detection of Duodenal Villous Atrophy on Endoscopic Images Using a Deep Learning Algorithm." *Gastrointestinal Endoscopy* 97, no. 5: 911–916. https://doi.org/10.1016/j.gie.2023.01.006.

Schwalbe, N., and B. Wahl. 2020. "Artificial Intelligence and the Future of Global Health." *Lancet* 395, no. 10236: 1579–1586. https://doi.org/10.1016/S0140-6736(20)30226-9.

Seo, J. Y., H. Hong, W.-S. Ryu, D. Kim, J. Chun, and M.-S. Kwak. 2023. "Development and Validation of a Convolutional Neural Network Model for Diagnosing *Helicobacter pylori* Infections With Endoscopic Images: A Multicenter Study." *Gastrointestinal Endoscopy* 97, no. 5: 880–888.e2. https://doi.org/10.1016/j.gie.2023.01.007.

Seoni, S., V. Jahmunah, M. Salvi, P. D. Barua, F. Molinari, and U. R. Acharya. 2023. "Application of Uncertainty Quantification to Artificial Intelligence in Healthcare: A Review of Last Decade (2013–2023)." *Computers in Biology and Medicine* 165: 107441. https://doi.org/10.1016/j.compbiomed.2023.107441.

Seoni, S., A. Shahini, K. M. Meiburger, et al. 2024. "All You Need Is Data Preparation: A Systematic Review of Image Harmonization Techniques in Multi-Center/Device Studies for Medical Support Systems." *Computer Methods and Programs in Biomedicine* 250: 108200. https://doi.org/10.1016/j.cmpb.2024.108200.

Shaik, T., X. Tao, N. Higgins, et al. 2023. "Remote Patient Monitoring Using Artificial Intelligence: Current State, Applications, and Challenges." *WIREs Data Mining and Knowledge Discovery* 13, no. 2: e1485. https://doi.org/10.1002/widm.1485.

Sharma, A., R. Kumar, and P. Garg. 2023. "Deep Learning-Based Prediction Model for Diagnosing Gastrointestinal Diseases Using Endoscopy Images." *International Journal of Medical Informatics* 177: 105142. https://doi.org/10.1016/j.ijmedinf.2023.105142.

Shen, Y., A. Chen, X. Zhang, et al. 2023. "Real-Time Evaluation of *Helicobacter pylori* Infection by Convolution Neural Network During White-Light Endoscopy: A Prospective, Multicenter Study (With Video)." *Clinical and Translational Gastroenterology* 14, no. 10: e00643. https://doi.org/10.14309/ctg.0000000000000643.

Shichijo, S., Y. Endo, K. Aoyama, et al. 2019. "Application of Convolutional Neural Networks for Evaluating *Helicobacter pylori* Infection Status on the Basis of Endoscopic Images." *Scandinavian Journal of Gastroenterology* 54, no. 2: 158–163. https://doi.org/10.1080/00365521.2019.1577486.

Shuvo, S. B., S. S. Alam, S. U. Ayman, et al. 2025. "Application of Wavelet Transformation and Artificial Intelligence Techniques in Healthcare: A Systemic Review." *WIREs Data Mining and Knowledge Discovery* 15, no. 2: e70007. https://doi.org/10.1002/widm.70007.

Smedsrud, P. H., V. Thambawita, S. A. Hicks, et al. 2021. "Kvasir-Capsule, a video capsule endoscopy dataset." *Scientific Data* 8, no. 1: 142. https://doi.org/10.1038/s41597-021-00920-z.

Song, A. H., G. Jaume, D. F. K. Williamson, et al. 2023. "Artificial Intelligence for Digital and Computational Pathology." *Nature Reviews Bioengineering* 1, no. 12: 930–949. https://doi.org/10.1038/s44222-023-00096-8.

Stidham, R. W., and K. Takenaka. 2022. "Artificial Intelligence for Disease Assessment in Inflammatory Bowel Disease: How Will It Change Our Practice?" *Gastroenterology* 162, no. 5: 1493–1506. https://doi.org/10.1053/j.gastro.2021.12.238.

Stoleru, C.-A., E. H. Dulf, and L. Ciobanu. 2022. "Automated Detection of Celiac Disease Using Machine Learning Algorithms." *Scientific Reports* 12, no. 1: 4071. https://doi.org/10.1038/s41598-022-07199-z.

Sucipto, K., A. Khosla, M. Drage, et al. 2023. "Quantitative and Explainable Artificial Intelligence (Ai)-powered Approaches to Predict Ulcerative Colitis Disease Activity From Hematoxylin and Eosin (H&E)-stained Whole Slide Images (WSI)." *Inflammatory Bowel Diseases* 29, no. Supplement_1: S22–S23. https://doi.org/10.1093/ibd/izac247.042.

Sugano, K., J. Tack, E. J. Kuipers, et al. 2015. "Kyoto Global Consensus Report on *Helicobacter Pylori* Gastritis." *Gut* 64, no. 9: 1353–1367. https://doi.org/10.1136/gutjnl-2015-309252.

Suman, S., F. Hussin, A. Malik, et al. 2017. "Feature Selection and Classification of Ulcerated Lesions Using Statistical Analysis for WCE Images." *Applied Sciences* 7, no. 10: 1097. https://doi.org/10.3390/app7101097.

Suman, S., F. A. B. Hussin, A. S. Malik, et al. 2017. "Detection and Classification of Bleeding Region in WCE Images Using Color Feature." In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, 1–6. ACM. https://doi.org/10.1145/3095713.3095731.

Syed, S., M. Al-Boni, M. N. Khan, et al. 2019. "Assessment of Machine Learning Detection of Environmental Enteropathy and Celiac Disease in Children." *JAMA Network Open* 2, no. 6: e195822. https://doi.org/10.1001/jamanetworkopen.2019.5822.

Takabayashi, K., T. Kobayashi, K. Matsuoka, et al. 2024. "Artificial Intelligence Quantifying Endoscopic Severity of Ulcerative Colitis in

Gradation Scale." *Digestive Endoscopy* 36, no. 5: 582–590. https://doi.org/10.1111/den.14677.

Takenaka, K., K. Ohtsuka, T. Fujii, et al. 2020. "Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis." *Gastroenterology* 158, no. 8: 2150–2157. https://doi.org/10.1053/j.gastro.2020.02.012.

Tang, S., X. Yu, C. F. Cheang, et al. 2023. "Transformer-Based Multi-Task Learning for Classification and Segmentation of Gastrointestinal Tract Endoscopic Images." *Computers in Biology and Medicine* 157: 106723. https://doi.org/10.1016/j.compbiomed.2023.106723.

Trenfield, S. J., A. Awad, L. E. McCoubrey, et al. 2022. "Advancing Pharmacy and Healthcare With Virtual Digital Technologies." *Advanced Drug Delivery Reviews* 182: 114098. https://doi.org/10.1016/j.addr.2021.114098.

Tuba, E., M. Tuba, and R. Jovanovic. 2017. "An Algorithm for Automated Segmentation for Bleeding Detection in Endoscopic Images." In *2017 International Joint Conference on Neural Networks (IJCNN)*, 4579–4586. IEEE. https://doi.org/10.1109/IJCNN.2017.7966437.

Tziortziotis, I., F.-M. Laskaratos, and S. Coda. 2021. "Role of Artificial Intelligence in Video Capsule Endoscopy." *Diagnostics (Basel)* 11, no. 7: 1192. https://doi.org/10.3390/diagnostics11071192.

Uyanik, H., A. Sengur, M. Salvi, R. Tan, J. H. Tan, and U. R. Acharya. 2025. "Automated Detection of Neurological and Mental Health Disorders Using EEG Signals and Artificial Intelligence: A Systematic Review." *WIREs Data Mining and Knowledge Discovery* 15, no. 1: e70002. https://doi.org/10.1002/widm.70002.

Vorontsov, E., A. Bozkurt, A. Casson, et al. 2024. "A Foundation Model for Clinical-Grade Computational Pathology and Rare Cancers Detection." *Nature Medicine* 30, no. 10: 2924–2935. https://doi.org/10.1038/s41591-024-03141-0.

Wang, G., S. Zhang, J. Li, et al. 2023. "CB-HRNet: A Class-Balanced High-Resolution Network for the Evaluation of Endoscopic Activity in Patients With Ulcerative Colitis." *Clinical and Translational Science* 16, no. 8: 1421–1430. https://doi.org/10.1111/cts.13542.

Wang, X., H. Qian, E. J. Ciaccio, et al. 2020. "Celiac Disease Diagnosis From Videocapsule Endoscopy Images With Residual Learning and Deep Feature Extraction." *Computer Methods and Programs in Biomedicine* 187: 105236. https://doi.org/10.1016/j.cmpb.2019.105236.

Wang, X., J. Zhao, E. Marostica, et al. 2024. "A Pathology Foundation Model for Cancer Diagnosis and Prognosis Prediction." *Nature* 634, no. 8035: 970–978. https://doi.org/10.1038/s41586-024-07894-z.

Wang, Y., T. Shi, F. Gao, S. Tian, and L. Yu. 2024. "Celiac Disease Diagnosis From Endoscopic Images Based on Multi-Scale Adaptive Hybrid Architecture Model." *Physics in Medicine and Biology* 69, no. 7: 75014. https://doi.org/10.1088/1361-6560/ad25c1.

Wei, J. W., A. A. Suriawinata, L. J. Vaickus, et al. 2020. "Evaluation of a Deep Neural Network for Automated Classification of Colorectal Polyps on Histopathologic Slides." *JAMA Network Open* 3, no. 4: e203398. https://doi.org/10.1001/jamanetworkopen.2020.3398.

Wei, J. W., J. W. Wei, C. R. Jackson, B. Ren, A. A. Suriawinata, and S. Hassanpour. 2019. "Automated Detection of Celiac Disease on Duodenal Biopsy Slides: A Deep Learning Approach." *Journal of Pathology Informatics* 10, no. 1: 7. https://doi.org/10.4103/jpi.jpi_87_18.

Wimmer, G., A. Vécsei, M. Häfner, and A. Uhl. 2018. "Fisher Encoding of Convolutional Neural Network Features for Endoscopic Image Classification." *Journal of Medical Imaging* 5, no. 3: 1. https://doi.org/10.1117/1.JMI.5.3.034504.

Xia, J., T. Xia, J. Pan, et al. 2021. "Use of Artificial Intelligence for Detection of Gastric Lesions by Magnetically Controlled Capsule Endoscopy." *Gastrointestinal Endoscopy* 93, no. 1: 133–139.e4. https://doi.org/10.1016/j.gie.2020.05.027.

Xu, H., N. Usuyama, J. Bagga, et al. 2024. "A Whole-Slide Foundation Model for Digital Pathology From Real-World Data." *Nature* 630, no. 8015: 181–188. https://doi.org/10.1038/s41586-024-07441-w.

Yang, Z., T. Wei, Y. Liang, et al. 2025. "A Foundation Model for Generalizable Cancer Diagnosis and Survival Prediction From Histopathological Images." *Nature Communications* 16, no. 1: 2366. https://doi.org/10.1038/s41467-025-57587-y.

Yasmin, F., M. M. Hassan, M. Hasan, et al. 2023. "GastroNet: Gastrointestinal Polyp and Abnormal Feature Detection and Classification With Deep Learning Approach." *IEEE Access* 11: 97605–97624. https://doi.org/10.1109/ACCESS.2023.3312729.

Younas, F., M. Usman, and W. Q. Yan. 2023. "A Deep Ensemble Learning Method for Colorectal Polyp Classification With Optimized Network Parameters." *Applied Intelligence* 53, no. 2: 2410–2433. https://doi.org/10.1007/s10489-022-03689-9.

Yu, G., K. Sun, C. Xu, et al. 2021. "Accurate Recognition of Colorectal Cancer With Semi-Supervised Deep Learning on Pathological Images." *Nature Communications* 12, no. 1: 6311. https://doi.org/10.1038/s41467-021-26643-8.

Zachariah, R., J. Samarasena, D. Luba, et al. 2020. "Prediction of Polyp Pathology Using Convolutional Neural Networks Achieves 'Resect and Discard' Thresholds." *American Journal of Gastroenterology* 115, no. 1: 138–144. https://doi.org/10.14309/ajg.0000000000000429.

Zhang, M., J. Pan, J. Lin, et al. 2023. "An Explainable Artificial Intelligence System for Diagnosing *Helicobacter pylori* Infection Under Endoscopy: A Case–Control Study." *Therapeutic Advances in Gastroenterology* 16: 17562848231155023. https://doi.org/10.1177/17562848231155023.

Zhang, R., Y. Zheng, T. W. C. Mak, et al. 2017. "Automatic Detection and Classification of Colorectal Polyps by Transferring Low-Level CNN Features From Nonmedical Domain." *IEEE Journal of Biomedical and Health Informatics* 21, no. 1: 41–47. https://doi.org/10.1109/JBHI.2016.2635662.

Zhang, Y., F. Li, F. Yuan, et al. 2020. "Diagnosing Chronic Atrophic Gastritis by Gastroscopy Using Artificial Intelligence." *Digestive and Liver Disease* 52, no. 5: 566–572. https://doi.org/10.1016/j.dld.2019.12.146.

Zhou, B., X. Rao, H. Xing, Y. Ma, F. Wang, and L. Rong. 2023. "A Convolutional Neural Network-Based System for Detecting Early Gastric Cancer in White-Light Endoscopy." *Scandinavian Journal of Gastroenterology* 58, no. 2: 157–162. https://doi.org/10.1080/00365521.2022.2113427.

Zhou, T., G. Han, B. N. Li, et al. 2017. "Quantitative Analysis of Patients With Celiac Disease by Video Capsule Endoscopy: A Deep Learning Method." *Computers in Biology and Medicine* 85: 1–6. https://doi.org/10.1016/j.compbiomed.2017.03.031.

Appendix A

TABLE A1 | ML methods applied to gastrointestinal images.

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Mesejo et al. (2016)	CT images	Colorectal polyps	76 videos	ML: SVM, RF, random subspace	Sensitivity: 72.7% Specificity: 85.9% Accuracy: 82.4%
Bang et al. (2021)	CT images	Gastric Cancer	2703 U-EGCs used for training	ML: 18 classifier models used for curative prediction	Sensitivity: 99% Precision: 92.6% Accuracy: 93.4%
Koh et al. (2021)	Histological images	Celiac disease	91 biopsy images	ML: XGB	Sensitivity: 89.7% Specificity: 86.7% Accuracy: 88.9%
Mohammad and Al-Razgan (2022)	WCE (endoscopic)	GID	3000 images per class	ML: SVM	Sensitivity: 99.4% Precision: 99.8% Accuracy: 99.8%
Stoleru et al. (2022)	Capsule endoscopy	Celiac disease	109 videos	ML: SVM	Sensitivity: 96% Precision: 94% Accuracy: 94.1%
Faust et al. (2023)	Histological images	Celiac disease	284 whole slide images	ML: SVM	Sensitivity: 97.7% Specificity: 99% Accuracy: 98.5%
Sucipto et al. (2023)	Histological images	Ulcerative colitis	820 whole slide images	ML: RF	Accuracy RF: 87%
Zhang et al. (2023)	Endoscopic images	H pylori	5106 endoscopic images	ML: XGB	Sensitivity: 90.3% Specificity: 93.6% Accuracy: 91.9% AUC: 0.96

Abbreviations: AUC: area under the curve; BMFA: bidirectional marginal fisher analysis; CNN: convolutional neural network; CT: computer tomography; DL: deep learning; GID: gastrointestinal disease; GNN: graph neural network; ML: machine learning; SCC: Spearman's correlation coefficient; RF: random forest; SVM: support vector machine; ViT: vision transformer; WCE: wireless capsule endoscopy; XGB: extreme gradient boosting.

TABLE A2 | DL methods applied to gastrointestinal images.

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Komeda et al. (2017)	Endoscopic images	Colorectal polyps	1800 images	DL: CNN	Accuracy: 75.1%
Korbar et al. (2017)	Histological images	Colorectal polyps	239 whole slide images	DL: ResNet	Sensitivity: 88.3% Precision: 89.7% Accuracy: 93%
Zhang et al. (2017)	Endoscopic images	Colorectal polyps	1930 images	DL: CNN	Polyp detection and classification Sensitivity: 97.6%, 87.6% Precision: 99.4%, 87.3% AUC: 1.00, 0.86 Accuracy: 98%, 85.9%
Zhou et al. (2017)	WCE (endoscopic)	Celiac disease	21 videos	DL: CNN	Sensitivity: 100% Specificity: 100%
Wimmer et al. (2018)	Endoscopic images	Celiac disease	1661 endoscopic images	DL: CNN	Accuracy: 92.5%
Chen et al. (2019)	Contrast-enhanced CT images	GID	80 patients	DL: ResNet	AUC: 0.947
Cogan et al. (2019)	Endoscopic images	GID	1880 images	DL: NASNet, Inception-v4, and Inception-ResNet v2	Sensitivity: 93.9% Precision: 93.8% Specificity: 99.1% Accuracy: 97.3%
Horie et al. (2019)	Endoscopic images	Esophageal cancer	9546 images	DL: CNN	Sensitivity: 83% Specificity: 82% Accuracy: 82%
Shichijo et al. (2019)	Endoscopic images	H pylori	23,699 images	DL: CNN	Positive vs. Negative vs. Eradicated Accuracy: 77.4% Positive vs. Eradicated: Sensitivity: 65.7% Precision: 52.9% Accuracy: 81.6%
Syed et al. (2019)	Histological images	Celiac disease	3118 images	DL: CNN	Accuracy: 93.4%
Wei et al. (2019)	Histological images	Celiac disease	1230 whole slides images	DL: ResNet	Sensitivity: 87.3% Precision: 88% Accuracy: 87.7%
Kowsari et al. (2020)	Histological images	Celiac disease	491 whole slide images	DL: Hierarchical CNN	Sensitivity: 88.7% Precision: 91.1%
Ozawa et al. (2020)	Endoscopic images	Colorectal polyps	16,418 images	DL: CNN	Sensitivity: 92% Precision: 86%

(Continues)

TABLE A2 | (Continued)

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Patino-Barrientos et al. (2020)	CT images	Colorectal polyps	600 images	DL: VGG	Sensitivity: 86% Precision: 81% Accuracy: 83%
Poudel et al. (2020)	Endoscopic images	Colorectal polyps, Crohn's disease, and ulcerative colitis	7515 images	DL: ResNet50	Colorectal dataset Sensitivity: 92.8% Precision: 93.2% KVASIR Sensitivity: 92.2% Precision: 86.8% Accuracy: 95.7%
Takenaka et al. (2020)	Endoscopic images	Ulcerative colitis	40,718 images	DL: Deep neural network	Sensitivity: 92.4% Specificity: 93.5% Precision: 93.7% Accuracy: 92.9%
Wang et al. (2020)	Endoscopic images	Celiac disease	2140 images	DL: ResNet	Sensitivity: 97.2% Specificity: 97.2% Accuracy: 95.6%
Wei et al. (2020)	Histological images	Colorectal polyps	326 whole slides images	DL: ResNet	Sensitivity: 77.7% Specificity: 91.6% Accuracy: 87%
Zachariah et al. (2020)	CT images	Colorectal polyps	6223 images	DL: Inception-ResNet-v2	Sensitivity: 91% Specificity: 88% Precision: 74% Accuracy: 89%
Zhang et al. (2020)	Endoscopic images	Gastritis	5470 images	DL: CNN	Sensitivity: 94.5% Specificity: 94% Accuracy: 94.2%
Barash et al. (2021)	Capsule endoscopy	Crohn's disease	17,640 CE images	DL: CNN	Sensitivity: 66.2% Specificity: 84.6% Precision: 71.2% Accuracy: 77.1%
Ikenoyama et al. (2021)	Endoscopic images	Gastric cancer	2940 images	DL: CNN	Sensitivity: 58.4% Specificity: 87.3% Precision: 26.0% AUC: 0.757

(Continues)

TABLE A2 | (Continued)

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Rahman et al. (2021)	CT images	Colorectal polyps	26,512 images	DL: ResNet-101, VGG 19, Xception Ensemble	Sensitivity: 96.2% Precision: 92.1% Specificity: 99% Accuracy: 98.5% AUC: 0.9912
Xia et al. (2021)	Magnetic controlled Capsule Endoscopy	GID	201,365 MCE images	DL: CNN	Sensitivity: 96.2% Specificity: 76.2% Precision: 16% Accuracy: 77.1%
Yu et al. (2021)	Histological images	Colorectal cancer	13,111 whole slide images	DL: Inception V3	AUC: 0.972
Gong et al. (2022)	Endoscopy with narrow-band imaging	Gastric cancer	1886 images	DL: CNN	Sensitivity: 72.2% Precision: 80.9% Specificity: 85.5% Accuracy: 79.5%
Khan et al. (2022)	WCE (endoscopic)	GID	38,000 images	DL: CNN	Accuracy: 99.4%
Kumagai et al. (2022)	Histological images	Esophageal cancer	8097 images	DL: ViT	Sensititivty: 84.8% Specificity: 93.9% Accuracy: 91.2% AUC: 0.92
Stidham and Takenaka (2022)	Endoscopic images	Crohn's disease, Ulcerative colitis	16,514 images	DL: CNN	AUROC: 0.966
Bilal, Tsang, et al. (2023)	Histological images	Colorectal cancer	6590 whole slide images	DL: Deep CNN	AUC: 0.9548
Fan et al. (2023)	Endoscopic images	Ulcerative colitis	5875 endoscopic images and 20 full-length videos	DL: ResNet, CNN	Vascular pattern, erosions and ulcers, bleeding Sensitivity: 86.8%, 82.9%, 77.2% Precision: 86.4%, 86.2%, 77.1% Specificity: 94.7%, 94.6%, 92.7% Accuracy: 90.7%, 84.6%, 77.7%
Graham et al. (2023)	Histological images	Colorectal cancer	6591 whole slide images	DL: GNN	AUC: 0.98
Huang et al. (2023)	Endoscopic images	Colorectal polyps	758 images	DL: Masked GNN	MSE: 0.016 MAE: 0.071 RMSE: 0.129

(Continues)

TABLE A2 | (Continued)

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Iacucci et al. (2023)	Histological images	Ulcerative colitis	535 whole slide images	DL: VGG16	Sensitivity: 89% Specificity: 85% Accuracy: 79%
Iwaya et al. (2023)	Histological images	Gastric cancer	5753 histological images	DL: ResNet	Sensitivity: 97.7% Specificity: 94.6% Precision: 97.6% Accuracy: 96.7%
Jiang, Luo, et al. (2023)	Endoscopic images, Histological images	Ulcerative colitis	12,257 endoscopic images and biopsy results	DL: CNN	Endoscopic remission and severity Sensitivity: 98.4%, 83.7% Precision: 96.4%, 85.6% Accuracy: 97%, 90.2%
Jiang, Ding, et al. (2023)	CT images	Gastric cancer	3779 images	DL: ViT	Dice: 78.77%
Kim et al. (2023)	Endoscopic images	Ulcerative colitis	984 endoscopic images	DL: CNN	Sensitivity: 91.7% Precision: 91.7% Accuracy: 90% AUROC: 0.86
Lin, Hsu, et al. (2023)	Endoscopic images	H pylori	302 endoscopic images	DL: CNN	Sensitivity: 100% Specificity: 81% Precision: 82% Accuracy: 90% AUROC: 0.88
Lin, Chen, et al. (2023)	Histological images	H pylori	885 whole-slide images	DL: CNN	Sensitivity: 93.3% Specificity: 90.1% AUC: 0.97
Najdawi et al. (2023)	Histological images	Ulcerative colitis	1264 whole slide images	DL: CNN	Accuracy: 97%
Polat et al. (2023)	Endoscopic images	Ulcerative colitis	11,276 endoscopic images	DL: CNN	—
Qi et al. (2023)	Endoscopic images	Ulcerative colitis	15,120 endoscopic images	DL: ViT (Vision Transformer)	Accuracy: 83.7%
Scheppach et al. (2023)	Endoscopic images	Celiac disease	1704 endoscopic images	DL: ResNet	Sensitivity: 90% Specificity: 76% Accuracy: 84%
Seo et al. (2023)	Endoscopic images	*H. pylori*	9786 endoscopic images	DL: CNN	Sensitivity: 86% Specificity: 88% Accuracy: 87%
Sharma et al. (2023)	Endoscopic images	Gastrointestinal disease	1200 endoscopic images	DL: ResNet50	Accuracy: 99.2%

(Continues)

TABLE A2 | (Continued)

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Shen et al. (2023)	Endoscopic images	*H. pylori*	143,830 endoscopic images	DL: ResNet34	Sensitivity: 91.5% Specificity: 88.8% Accuracy: 89.9%
Tang et al. (2023)	Endoscopic images	GID	1645 images	DL: Transformer	Sensitivity: 96.53% Precision: 96.52% Accuracy: 96.94% DSC: 76.21% IoU: 66.35%
Wang et al. (2023)	Endoscopic images	Ulcerative colitis	3191 endoscopic images	DL: CNN	Sensitivity: 92.9% Specificity: 95.4% Precision: 93.4% Accuracy: 95.1%
Yasmin et al. (2023)	Endoscopic images	Colorectal polyps	1000 images	DL: MobileNet v2, MobileNet v2, and Resnet50	Sensitivity: 100% Precision: 99%
Younas et al. (2023)	Endoscopic images	Colorectal polyps	76 colonoscopy videos	DL: CNN ensemble model	UCI, Piccolo Datasets Sensitivity: 97.2%, 81.1% Precision: 95.5%, 82.4% Accuracy: 96.3%, 81.2%
Zhou et al. (2023)	White light endoscopy	Gastric cancer	1243 images	DL: CNN	Sensitivity: 84.5% Specificity: 90.5% Precision: 83.2% Accuracy: 88.3%
Chen, Ding, et al. (2024)	CT images, Histological images	Gastric cancer	429 images	DL: CNN	Auc: 0.821
Gao et al. (2024)	CT images, Histological images	Gastric cancer	4284 patients	DL: ResNet	Sensitivity: 77.5% Specificity: 81.4% Precision: 89.6% Accuracy: 78.8% AUC: 0.846
Ghosal et al. (2024)	CT images	Colorectal cancer	296 images	DL: CNN	Accuracy: 95.83% AUC: 0.956 Dice: 62.14% IoU: 69.75%
Guo et al. (2024)	Endoscopic images	GID	10,662 labeled images 110,079 unlabeled images	DL: ResNet	Sensitivity: 75% Specificity: 99.42% Precision: 73.68% Accuracy: 88.92%

(Continues)

TABLE A2 | (Continued)

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Hossain et al. (2024)	Endoscopic images	GID	6000 images	DL: CNN	Sensitivity: 99.3% Precision: 99.3% Accuracy: 99.3%
Ibrahim et al. (2024)	Histological images	*H. pylori*	204 histopathological images	DL: ResNet	Sensitivity: 90.7% Specificity: 90.7% Precision: 90.7% Accuracy: 90.2%
Kohli et al. (2024)	Histological images	Ulcerative colitis	299 whole slides images	DL: CNN	Accuracy: 64.8% AUC: 0.972
Li et al. (2024)	Endoscopic images	Ulcerative colitis	6693 endoscopic images	DL: ResNet	Sensitivity: 84% Precision: 85.6% Accuracy: 85.5%
Malik et al. (2024)	Endoscopic images	Ulcerative colitis	4437 Wireless capsule endoscopy images	DL: VGG19	Sensitivity: 98.9% Specificity: 99.8% Precision: 98.9% Accuracy: 99.5%
Rymarczyk et al. (2024)	Histological images	Crohn's disease and ulcerative colitis	Crohn's disease: 2935 whole slide images Ulcerative colitis: 3496 whole slide images	DL: CNN	Accuracy: 87%
Sanderson and Matuszewski (2024)	Endoscopic images	GID	10,662 labeled images 110,079 unlabeled images	DL: ResNet50, Transformer	Classification Sensitivity: 59.6% Precision: 72.3% Accuracy: 78% Object detection Average precision 50: 92.1% Segmentation Dice: 89.6%, 92.7% IoU: 84.4%, 86.7% Sensitivity: 90.2%, 93.3% Precision: 92.1%, 92.6%
Takabayashi et al. (2024)	Endoscopic images	Ulcerative colitis	13,826 pairs of endoscopic images	DL: CNN	SCC: 0.89
Vorontsov et al. (2024)	Histological images	Colorectal cancer and gastric cancer	89,417 whole slide images for training 22,932 whole slide images for testing	DL: ViT, DiNO v.2	AUC: 0.958 AUC: 0.849

(Continues)

TABLE A2 | (Continued)

References	Diagnostic modality	Detection task	Data size	Classifier	Outcome/Results
Wang, Shi, et al. (2024)	Endoscopic images	Celiac disease	4002 endoscopic images	DL: CNN, ViT	Sensitivity: 98% Specificity: 99% Precision: 99.4% Accuracy: 98.4%
Wang, Zhao, et al. (2024)	Histological images	Colorectal cancer	60,530 whole slide images for training 19,491 whole slide images for testing	DL: ViT	AUC: 0.871
Xu et al. (2024)	Histological images	Colorectal cancer and gastric cancer	171,189 whole slide images	DL: ViT	—

Abbreviations: AUC: area under the curve; CNN: convolutional neural network; CT: computer tomography; DL: deep learning; EGC: early gastric cancer; GID: gastrointestinal disease; GNN: graph neural network; MAE: mean absolute error; MCE: magnetically controlled capsule endoscopy; ML: machine learning; MSE: mean square error; RMSE: root mean square error; SCC: spearman's correlation coefficient; SVM: support vector machine; VGG: visual geometry group; ViT: vision transformer; WCE: wireless capsule endoscopy.