

108021\_Auto\_Edited.docx

---

WORD COUNT

7185

TIME SUBMITTED

22-MAY-2025 10:55AM

PAPER ID

116362330

**Name of Journal:** *World Journal of Gastroenterology*

**Manuscript NO:** 108021

**Manuscript Type:** REVIEW

**Revolutionizing gastroenterology and hepatology with artificial intelligence: From precision diagnosis to equitable healthcare through interdisciplinary practice**

AI in Gastroenterology & Hepatology

Zhi-Li Chen, Chao Wang, Fang Wang

## **Abstract**

Artificial intelligence (AI) is driving a paradigm shift in gastroenterology and hepatology by delivering cutting-edge tools for disease screening, diagnosis, treatment, and prognostic management. Through deep learning, radiomics, and multimodal data integration, AI has achieved diagnostic parity with expert clinicians in endoscopic image analysis (*e.g.*, early gastric cancer detection, colorectal polyp identification) and non-invasive assessment of liver pathologies (*e.g.*, fibrosis staging, fatty liver typing) while demonstrating utility in personalized care scenarios such as predicting hepatocellular carcinoma recurrence and optimizing inflammatory bowel disease treatment responses. Despite these advancements challenges persist including limited model generalization due to fragmented datasets, algorithmic limitations in rare conditions (*e.g.*, pediatric liver diseases) caused by insufficient training data, and unresolved ethical issues related to bias, accountability, and patient privacy. Mitigation strategies involve constructing standardized multicenter databases, validating AI tools through prospective trials, leveraging federated learning to address data scarcity, and developing interpretable systems (*e.g.*, attention heatmap visualization) to enhance clinical trust. Integrating generative AI, digital twin technologies, and establishing unified ethical/regulatory frameworks will accelerate AI adoption in primary care and foster equitable healthcare access while interdisciplinary collaboration and evidence-based implementation remain critical for realizing AI's potential to redefine precision care for digestive disorders, improve global health outcomes, and reshape healthcare equity.

**Key Words:** Artificial intelligence; Precision medicine; Gastroenterology; Hepatology; Multimodal data integration; Deep Learning; Microbiome

Chen ZL, Wang C, Wang F. Revolutionizing gastroenterology and hepatology with artificial intelligence: From precision diagnosis to equitable healthcare through interdisciplinary practice. *World J Gastroenterol* 2025; In press

**Core Tip:** This review highlights artificial intelligence (AI)-driven innovations in gastroenterology and hepatology, demonstrating breakthroughs in endoscopic/image analysis, multi-omics integration, and precision therapy. AI achieves diagnostic parity with experts in detecting early cancers and fibrosis, while addressing challenges like data fragmentation and bias through standardized databases, federated learning, and explainable systems. The study emphasizes interdisciplinary collaboration and ethical frameworks to advance equitable healthcare access and redefine digestive disease management.

## INTRODUCTION

The development of AI applications in medicine dates back to the mid-20th century, and the combination of its technological evolution and medical needs has driven several key breakthroughs. Early explorations were represented by expert systems, such as the MYCIN system developed in 1975, which first realized rule-based diagnosis of infectious diseases[1]. This phase of research laid the foundation for subsequent AI logical reasoning in medicine, but the scope of application was limited due to the scale of data and computational power at that time. Machine learning technology emerged from the 1990s to the beginning of the 21st century. Machine learning algorithms (ML) such as support vector machines (SVMs) have emerged in medical data classification, and in a 2004 machine learning study exploring the impact of genetic polymorphisms on breast cancer risk, support vector machines (SVMs) became the optimal model at the time for distinguishing between breast cancer and normal controls by comprehensively analyzing the three key SNP loci with a prediction accuracy of 69%, revealing the multi-gene SNP combination's synergistic advantage for cancer risk assessment[2]. Meanwhile, research on artificial neural networks continued to advance, with Yann LeCun's team realizing handwritten character recognition through LeNet, a convolutional neural network (CNN) architecture, in 1998, providing the basis for medical image analysis[3]. After 2010, the explosive development of deep learning

technology completely changed the landscape of medical AI. 2012 AlexNet's breakthrough performance in the ImageNet competition pushed the widespread application of CNN in medical imaging[4].

The current medical AI technology system covers <sup>7</sup> machine learning (ML), deep learning (DL) and natural language processing (NLP) technologies, and the algorithms include <sup>4</sup> support vector machines (SVM), decision trees, random forests, logistic regression, convolutional neural networks (CNN), recurrent neural networks (RNN), long and short-term memory networks (LSTM), Transformer, generative adversarial networks (GAN) and other algorithms. (GAN) and other algorithms. Machine learning algorithms have played an important role in medical image analysis and disease diagnosis. For example, algorithms such as SVM and random forests excel at processing structured data and can be used for disease risk prediction and diagnostic support. In the lung nodule segmentation task, the Dice similarity coefficient (DSC) of the fully automated lung nodule segmentation model constructed on the basis of random forest can reach 0.986[5]. Deep learning models, especially Convolutional Neural Networks (CNN) and Transformer, have made significant progress in medical image analysis and multimodal data fusion. Among them, CNNs still dominate the field of medical image parsing[6]. In 2017, deep convolutional neural networks can reach a level comparable to experts in skin cancer classification[7]. Vision Transformer achieves global feature extraction through the self-attention mechanism and demonstrates performance beyond traditional CNNs in analysis such as CT 3D reconstruction and full-slice pathology images[8-11]. Natural Language Processing (NLP) technologies are important in <sup>5</sup> medical text processing and information extraction, which cover several healthcare scenarios such as clinical practice, hospital management, personal care, public health and drug development[12]. NLP facilitates the translation of cancer treatment from laboratory to clinic by mining unstructured text data in electronic medical records. It empowers oncologists to construct evidence-based research frameworks in case identification, staging, and outcome quantification. In this way, NLP lays a

technological foundation for the advancement of an efficient and precise cancer diagnosis and treatment system[13].

Global cancer statistics for 2022 show that cancers of the digestive tract are one of the major threats to human life and health, with esophageal, gastric, liver, and colorectal cancers having the seventh, fifth, third, and second highest mortality rates in the world, respectively[14, 15]. And due to the lack of specific symptoms and early diagnostic markers in the early stages, there is a large number of potential patients who remain undiagnosed, resulting in the majority of patients often being diagnosed at a late stage[16-18]. Improving early detection of gastroenterological cancers is urgently needed. Benign ulcerative colorectal diseases (UCDs), such as ulcerative colitis, Crohn's disease, ischemic colitis, and intestinal tuberculosis, still require accurate diagnostic typing and appropriate therapeutic strategies by clinicians due to the fact that such diseases have similar phenotypes and but different etiologies and therapeutic strategies, however, AI may be able to accomplish the nuanced typing of the diseases with the advantage of its algorithm[19]. Therefore, there is an important need for the introduction of artificial intelligence (AI) techniques in gastroenterological diseases. Moreover, current studies have shown that AI technologies exhibit great potential in the early detection, diagnosis, treatment planning, and prognosis prediction of gastrointestinal cancers, *e.g.*, AI is able to improve the sensitivity and specificity of tumor screening through medical image analysis and support intraoperative assessment of the depth of tumor invasion and prediction of the treatment response to optimize personalized treatment plans[20]. Furthermore, applying AI to pathology has significantly improved diagnostic efficiency and accuracy, surpassing the performance of human experts[18]. The application of artificial intelligence in gastrointestinal diseases is summarized as shown in **Figure 1**. Although AI technology is reshaping the diagnosis and treatment paradigm of digestive diseases, and its application has crossed over from single-modality assisted diagnosis to multidimensional, chain-wide intelligent decision-making systems, there are still some unresolved issues, such as the crisis of trust of patients in AI, the issue of patient privacy, and ethical norms. In

summary, this paper will systematically review the breakthrough progress of AI in the diagnosis and treatment of digestive diseases, focusing on the dissection of its translational value in early warning, precise staging, treatment optimization and prognostic management, as well as elaborate the current problems and possible future solutions for the large-scale promotion of AI in digestive diseases, and further explore the key challenges of multimodal data fusion, interpretability enhancement, data privacy protection and clinical landing of the future development direction. We expect that this review will provide the latest research progress and unique contributions on AI in gastroenterology and hepatology, break through the original limitations through emerging AI tools or algorithms such as multimodal data integration, interpretable AI systems, or new ethical frameworks adopted by AI, and provide a comprehensive overview for clinical researchers and AI enthusiasts, as well as insights and ideas to further ground the clinical translation of AI.

### **BREAKTHROUGHS IN ARTIFICIAL INTELLIGENCE FOR GASTROINTESTINAL DISEASE DIAGNOSIS AND TREATMENT**

Recent advancements in artificial intelligence have catalyzed transformative applications across gastrointestinal (GI) disease diagnostics and therapeutics, with machine learning (ML) and deep learning (DL) emerging as distinct yet complementary paradigms. As summarized in **Table 1**, ML algorithms—including support vector machines (SVM), random forests, and gradient boosting—have demonstrated robust performance in structured data analysis, particularly for risk stratification and treatment response prediction. For instance, SVM-based models achieved 98.5% accuracy in differentiating celiac disease from non-celiac duodenitis by analyzing duodenal histopathological features, while ensemble methods improved recurrence risk stratification in Crohn’s disease (AUC: 0.84).

In parallel, **Table 2** highlights the ascendancy of DL architectures—notably convolutional neural networks (CNNs) and vision transformers—in processing high-dimensional imaging and multimodal data. These systems have achieved human-expert

parity in endoscopic image interpretation, with CNN models attaining 99% sensitivity for polyp detection and transformer-based frameworks enabling 3D lesion quantification in CT enterography ( $\kappa = 0.83$ ). The integration of attention mechanisms and federated learning has further enhanced model generalizability across heterogeneous datasets, addressing longstanding challenges in rare disease analysis.

### *AI applications for inflammatory gastrointestinal diseases*

Artificial intelligence demonstrates multidimensional value in the management of ulcerative colitis (UC) and Crohn's disease (CD). In the field of UC, deep learning-based histologic scoring systems (e.g., PHRI) are highly correlated with endoscopic scores (Mayo, PICA<sub>SSO</sub>) and clinical outcomes (ICC = 0.84), with AI-assisted diagnostic sensitivity of 78% and specificity of 91.7%[77]. By quantifying features such as neutrophil extravasation, a convolutional neural network (CNN) can predict Nancy's histologic index score ( $\kappa = 0.91$ ) and accurately determine histologic remission (97% accuracy)[78, 79]. AI also predicts the risk of recurrence by the “vascular healing” status in endoscopic images, with a significantly higher recurrence rate in the vasoactive group (23.9%) than in the vascular healing group (3.0%)[80], and a recurrence rate of up to 45% in patients with a rectal cuprocyte mucus area (GCR)  $\leq 12\%$ [81].

In addition, the UC-SCALE algorithm enables topological visualization of disease severity by automating the assessment of inflammation distribution in full-length colonoscopy videos (86.5% accuracy,  $\kappa = 0.813$ )[82].

For CD, a multimodal machine learning model integrating magnetic resonance small bowel imaging (MRE) and biochemical markers noninvasively assessed terminal ileal endoscopic activity with an AUC of 0.84[83]; and a deep learning model (SA-AbMILP) predicted GHAS scores from histologic images (65%-89% accuracy), which was in high agreement with the pathologist's findings[84]. For the prediction of postoperative recurrence, the EfficientNet-b5 model analyzed whole slide images of surgical specimens with an AUC as high as 0.995 and found adipocyte atrophy and mast cell infiltration to be key features of recurrence[85].

AI significantly improves the efficiency and accuracy of inflammatory gastrointestinal disease screening. In capsule endoscopy (CE), CNN models can quickly identify ulcers, erosions, and vascular malformations (with 96.9%-99% sensitivity) with processing speeds up to 90 frames/second[86, 87]; a sequential CNN model for CD ulcer severity has an accuracy of 91% in identifying grade 3 ulcers[88]. In the field of colonoscopy, the FRCNN-AA-CIF model based on the attention mechanism and context fusion has a leakage rate of only 4.22% and a mAP of 0.817[89]; the ResNet50 migration learning model has an accuracy of 99.8% in the polyp classification task, which is superior to the traditional methods[90, 91]. In imaging, 3D-CNN distinguished colon cancer from acute diverticulitis by CT images (sensitivity 83.3%, specificity 86.6%), and AI support enabled radiologists to increase the sensitivity to 85.6%[92]. For data-scarce scenarios, CT colon segmentation adopts a guided sequential scenario training strategy, which requires only 10 cases of annotation to achieve a Dice coefficient of 97.15%, and its cross-layer feature comparison learning mechanism enables polyp detail retention to exceed 98.28%[93]. The most clinically translatable value is the hybrid architecture of 3D-CNN and Random Forest: By modeling spatial continuity, this model achieves a physician-level agreement of  $\kappa = 0.83$  for the severity classification of 7.5-mm-level lesions in CTE images, with a 91.51% accuracy in localizing moderate-to-severe lesions[94].

AI has also made breakthroughs in gastritis, celiac disease and small bowel disease. Multi-stage semantic segmentation model (MSP AI-G) has an accuracy of 91% in gastric biopsy diagnosis, with significant cross-institutional validation robustness[95]; CNN can differentiate between *Helicobacter pylori* gastritis and autoimmune gastritis (with 100% diagnostic concordance)[96]. For celiac disease (CD), the SVM model distinguishes between CD and non-celiac duodenitis (NCD) based on images of the duodenal lamina propria with 98.5% accuracy[97]. These innovations signal that AI is breaking through the traditional qualitative diagnostic framework and driving the transition of the diagnostic paradigm from qualitative detection to accurate quantitative

assessment through interpretable feature engineering and anatomical constraint algorithms.

### ***Recent advances in AI-enabled upper gastrointestinal tumor diagnosis and treatment***

Deep learning-based endoscopic image analysis system significantly improves the efficiency and detection rate of upper gastrointestinal tumors. The study showed that the sensitivity of the GRAIDS system in diagnosing upper gastrointestinal cancer by analyzing more than 1 million endoscopic images (0.942) was comparable to that of expert endoscopists (0.945) and significantly better than that of non-experts ( $P < 0.0001$ ), and its negative predictive value (0.978) was close to the level of experts (0.980), which demonstrated that AI can effectively assist grassroots hospitals in improving diagnostic capabilities[98]. For the determination of the depth of infiltration of early gastric cancer, the diagnostic model F1 value of AI collaborating with endoscopists (0.776) was higher than that of AI alone (0.768) or physician majority vote (0.662), confirming that human-machine synergy can compensate for the limitations of a single method[99]. In addition, AI-assisted white light endoscopy reduced the leakage rate of gastric tumors from 27.3% to 6.1% ( $P = 0.015$ ), which significantly improved the efficiency of identifying microscopic lesions[100]. In the evaluation of Barrett's esophagus for heterozygous hyperplasia, the AI system surpassed most pathologists in diagnostic accuracy (76.4%) of whole slide images, with an AUC of 0.94 and a sensitivity of 0.92 for the prediction of heterozygous hyperplasia[101]. For the prediction of peritoneal recurrence in gastric cancer, the deep learning model based on preoperative CT images had an AUC of 0.843 in external validation, and the predictive results of this model further indicated that adjuvant chemotherapy was associated with improved disease-free survival in stage II and stage III disease[102]. In addition, the AutoML model (GBM algorithm) based on the SEER database predicted liver metastases of gastrointestinal mesenchymal tumors with an AUC of 0.795, and feature significance analysis showed that tumor size and location were key predictors[103]. These findings indicate that the deep integration of AI and endoscopy technology is reshaping the diagnosis and treatment landscape of

upper gastrointestinal tumors. Through high-precision image recognition and real-time decision support, AI can not only alleviate the pressure of the shortage of professionals in primary care, but also promote the standardization of the diagnostic process and reduce the risk of misdiagnosis caused by differences in human experience.

Artificial intelligence fused with multi-omics further breaks through the limits of molecular typing, metastatic mechanisms and immunotherapy response prediction in gastric and esophageal cancer. Based on the dynamic features of mitochondrial ATP metabolism, membrane potential and lactate/pyruvate/glucose metabolism, the researchers constructed a MitoScore quantitative assessment system by integrating 10 machine learning algorithms, which can accurately differentiate the immune-metabolic subtypes of gastric cancer. Data analysis showed that high MitoScore subgroups presented abnormal activation of glycolytic pathways and were significantly associated with tumor aggressiveness phenotype and poor patient prognosis[104]. A support vector machine (SVM) model was constructed to predict the response of gastric cancer patients to Sintilimab combined with SOX treatment based on the screening of six key markers (DUOX2, HSPB1, S100A14, C1QA, TGFB1, and LTF) by combining the mIHC data and the deep-learning feature extraction technique, which The AUC values were 0.93 and 0.84 in the exploratory cohort ( $n = 107$ ) and validation cohort ( $n = 46$ ), respectively, showing good predictive efficacy[105]. In addition, the MetImage technology based on metabolomics encoded LC-MS data into images, and the AI model constructed using the convolutional neuron network algorithm screened esophageal squamous carcinoma with a sensitivity of 85%, specificity of 92%, and AUC of 0.95[106].

AI plays an important role in the optimization of treatment strategies for upper gastrointestinal tumors. A study based on a counterfactual random forest model using predictors of recurrence (mitotic count, tumor size, and tumor site) and imatinib duration to infer the likelihood of recurrence in a given patient at 7 years for each imatinib treatment duration suggested that gastric-derived GIST (< 15.9 cm and low mitotic count) does not require imatinib treatment, avoiding the 29%-35% of patients who receive ineffective treatment[26]. Moreover, the best model for machine learning to

predict imatinib adherence (LGBM) showed cognitive function and the presence or absence of therapeutic drug monitoring as key influencing factors[107].

***AI enabled colorectal tumor diagnosis and treatment: Recent advances and clinical translation***

AI Becomes a Real-Time Aid in Colonoscopy Traditional colonoscopy suffers from the limitation of a high rate of adenoma leakage (15%-30%), which has been significantly improved by deep learning-based computer-aided detection (CADe) systems through real-time polyp identification. For example, a multicenter randomized controlled trial (RCT) showed an approximately twofold reduction in adenoma missed diagnosis (AMR) in the AI-assisted colonoscopy group compared to the control group (15.5% vs. 32.4%), with a particularly significant advantage in the detection of tiny polyps ( $\leq 5$  mm) and nonpolypoid lesions (OR = 0.34-0.24)[108]. Another U.S. multicenter study (CADeT-CS trial) further validated that the AMR and SSL leakage rates in the AI-assisted group decreased to 20.12% and 7.14%, respectively, while the number of first-pass adenomas detected (APCs) increased significantly (1.19 vs. 0.90)[109]. In addition, the GI Genius system improved adenoma detection rate (ADR) by 8.3% in a large RCT (COLO-DETECT) without the need for extended operating time[110]. These results suggest that AI has become a central tool for optimizing the quality of colonoscopy by reducing the number of perceptual errors.

Digital pathology combined with AI algorithms demonstrated high accuracy in tissue classification and tumor detection. A multi-class tissue segmentation model developed in one study achieved a composite Dice score of 0.895 in colorectal biopsies with a tumor detection sensitivity of 0.987[111]. For lymph node metastasis (LNM) prediction, a deep learning model based on H&E-stained images had an area under the curve (AUC) of 0.764 in stage T1 colorectal cancer, which reduces unnecessary surgery by 15.1%[112]. In addition, the Deep-immune score, which quantifies the immune microenvironment by AI, was significantly associated with patient survival, with a 5-year survival rate of up to 87.4% in the high-scoring group, in which a high Deep-

immune score was associated with high levels of CD3T cell infiltration in the stromal region[113]. More importantly, AI simplifies the MSI screening process. The MSIntuit pre-screening tool based on H&E images has a sensitivity of 96%-98%, reducing the need for immunohistochemical validation by almost 50%[114]. Label-free infrared imaging combined with convolutional neural network (CNN) has an AUC of 0.90 for classification of MSI, which provides rapid support for precise treatment[115]. Multi-omics integration models (*e.g.*, TMO-Net) fusing multimodal data from genomics and pathology optimize cancer typing accuracy, while non-invasive screening techniques based on respiratory volatomics achieve more than 89% sensitivity (AUC = 0.91) by detecting VOC markers[116, 117]. These technologies not only improve diagnostic efficiency, but also provide support for individualized treatment at the molecular level.

Imaging and multimodal data integration, AI in imaging analysis optimizes treatment prediction by integrating multimodal data. Multimodal AI models integrating imaging, pathology and genomic data provide new tools for individualized treatment of colorectal tumors. The RAPIDS (RAdioPathomics Integrated preDiction System) system combines MRI radiomics with H&E pathologic features to predict pathological complete remission (pCR) from neoadjuvant radiotherapy for rectal cancer, with an AUC of 0.872 in the validation cohort, which was significantly better than the unimodal model ( $P < 0.0001$ )[118]. AI has significantly improved the prediction of treatment response by integrating imaging histology and clinical data. One study utilized random forest (RF) and gradient boosting algorithms to predict response to radiotherapy for colorectal cancer with an accuracy of 93.8%[119]. Another multi-omics analysis combined with spatial interaction mapping developed the CCIM-Net model, which effectively predicts chemosensitivity and guides combination therapy targeting FOLR2 macrophages[120].

The application of AI technology to surgical planning for colorectal tumors has also seen breakthroughs. An automated CT-based tumor segmentation model accurately assessed the total tumor volume (TTV) of colorectal liver metastases (CRLM) with an intraclass correlation coefficient (ICC) of 0.98[121]. For colorectal cancer liver metastases

(CRLM), the game-theory-based Shapley's additive interpretation (SHAP) AI model recommended individualized margin widths (6-12 mm), with an AUC of 0.78 in the validation cohort, confirming the association between an optimal margin width of 7 mm and a significant prolongation of survival in an external cohort[122]. In addition, the MLP model for predicting lymph node metastasis in the inferior mesenteric artery (AUC = 0.873) was significantly better than expert judgment (AUC = 0.509), which may reduce unnecessary clearance[123]. Machine learning algorithms performed well in survival prediction. Logistic regression models combined with clinical variables (e.g., distant metastases, number of lymph nodes) predicted 1- and 5-year survival with AUCs of 0.850 and 0.872, respectively[124]. Genomic profiling identified 32 key genes by interpretable AI (XAI) and predicted stage II colorectal cancer recurrence with an AUC of 0.952[125].

#### *AI-driven innovations in hepatic oncology: From multiscale diagnostics to precision therapeutics*

The degree of liver fibrosis is the most sensitive clinical warning sign for metabolic dysfunction-associated steatotic liver disease (MASLD)-associated hepatocellular carcinoma, and significant and advanced liver fibrosis not only increases the risk of hepatic and extra-hepatic complications, but is also significantly associated with liver-related mortality[126, 127]. Therefore, dynamic monitoring of liver fibrosis progression (e.g., liver stiffness value testing) has become a core strategy for early screening of hepatocellular carcinoma, and how to achieve accurate assessment of fibrosis staging remains an urgent clinical challenge. In recent years, artificial intelligence (AI) technologies have significantly improved the objectivity and sensitivity of assessment through multimodal innovations: The performance of deep learning models in ultrasound steatosis grading (AUC = 0.85) is not unlike the judgment of radiologists[128]; at the histological level, an AI-based measurement tool (AIM-MASH) effectively predicts the survival of patients with fibrosis through reproducible necroinflammation grading ( $\kappa = 1$ ) and pathologists' consensus highly concordant ( $\kappa =$

0.62-0.74), effectively predicted progression-free survival of fibrosis patients ( $P < 0.05$ ), providing a highly sensitive method for clinical trial endpoint assessment[129]; in the analysis of pathomechanisms, AI-based digital pathology combined with SHG/TPEF microscopy revealed the dynamics of treatment-induced regression of perisinusoidal fibrosis, breaking through the blindness of the traditional scoring for fibrosis regression[130]. For non-invasive techniques, machine learning models of imaging histology combined with diffusion-weighted imaging (DWI) can accurately identify liver fibrosis (AUC = 0.973) and early cirrhosis[131], whereas multiparametric quantitative ultrasound (MP-QUS) improves the diagnostic performance of fibrosis staging to an average AUC = 0.891 through feature engineering and algorithm optimization[132]. These technologies provide high-precision tools for the full management of liver disease through standardization, quantification and dynamic tracking capabilities, promoting fibrosis monitoring from static staging towards dynamic individualized interventions.

Artificial intelligence has shown no less ability than human experts in diagnostic imaging of liver tumors. Ultrasound (US)-based deep learning models significantly improve the accuracy of benign and malignant liver tumors identification by integrating patient background and blood biomarkers. For example, a multimodal deep learning model combining B-mode ultrasound images with clinical data improved the diagnostic accuracy from 68.52% to 96.30% (AUC 0.994) in unimodal mode[133]. In addition, the application of AI in CT and MRI image analysis is equally prominent. A generative adversarial network (GAN)-based model effectively mitigates the data imbalance problem by synthesizing high-fidelity CT images of liver tumors, improving the accuracy of classification models by 21%-34%[134]. In the MRI segmentation task, the three-dimensional convolutional neural network (3D CNN) performs well in semi-automatic segmentation of hepatocellular carcinoma (HCC) tumors, and especially performs best in diffusion-weighted imaging (DWI) and T1-weighted imaging, with Dice similarity coefficients (DSCs) up to 0.778[135]. The AI-driven software also detects liver metastases missed by radiologists on contrast-enhanced CT with a sensitivity of

70.8% and a false-positive rate of only 0.48/case[136]. These techniques not only improve the efficiency of the diagnosis, but also reduce human error.

AI has also demonstrated strong classification and prediction capabilities in liver cancer histopathology image analysis. Deep learning models based on the attention mechanism (*e.g.*, SENet) achieved 95.27% accuracy in the task of classifying the degree of differentiation of hepatocellular carcinoma, which is significantly better than traditional manual reading[137]. The CLAM model predicted immunogenetic signature activation by whole slide images with AUCs of 0.78-0.91, and pathologic analysis showed that the predicted areas were enriched with lymphocytes and neutrophils[138]. Microvascular infiltration (MVI) is a major risk factor for overall postoperative mortality and recurrence in hepatocellular carcinoma (HCC). Models combining imaging histology and deep learning perform well in microvascular infiltration (MVI) prediction. A deep learning model based on image-pathology fusion (Swin Transformer) was effective in predicting the VETC patterns of perivascular envelope tumor clusters, and its radiological-pathological histology column-line diagrams had a C-index of 0.67 in the external test set[139]. The random forest model based on preoperative CT had an accuracy of 96.8% (sensitivity 95.2%) in the test set[140], while the column-line graph model combining clinical features with deep learning features had an AUC of up to 0.940[141]. In addition, the multitask learning framework significantly improved prognostic stratification by simultaneously predicting MVI and VETC *via* 3D CNN with AUCs of 0.917 and 0.860, respectively[142].

AI shows potential to outperform traditional statistical methods in prognostic modeling. The machine-learning-based SMART-HCC score integrates nine clinical features, including liver stiffness, and its 5-year predictive AUC is  $\geq 0.89$ , which is superior to existing scoring systems[143]. Deep learning column line drawings combined with MRI imaging histologic features and clinical variables had an AUC of 0.949 for predicting early recurrence after HCC[144]. The randomized survival forest (RSF) model, by incorporating risk factors such as microvascular infiltration and satellite nodules, predicted early recurrence with a C-index of 0.896 in the training

group 0.798 in the validation group, which was significantly better than the Cox proportional risk model[145]. In addition, the knowledge-enhanced dual-style visual transformer (DSViT) improves the interpretability and performance of recurrence prediction by fusing multi-period CT images with domain knowledge[146]. Migration learning and multimodal strategies were used to optimize HCC risk prediction in patients with MASLD, significantly alleviating the problem of data scarcity and gender bias[147]. These models provide precise tools for individualized treatment and follow-up.

AI plays a key role in mining emerging molecular markers for liver cancer and optimizing immunotherapy strategies. Based on single-cell transcriptome analysis and machine learning, S100A10 was identified as a core gene for HCC diagnosis and immunotherapy response, and its expression was positively correlated with the stem cell marker POU5F1[148]. Stem-related classifiers (9-gene model, PPARGC1A, FTCD, CFHR3, MAGEA6, CXCL8, CABYR, EPO, HMMR, and UCK2) predicted the state of the tumor immune microenvironment and the efficacy of immune checkpoint inhibitors, with a significant enrichment of Treg cells and immune-suppressing pathways in high-risk patients[149]. These findings provide new ideas for targeting and stratifying immunotherapy in HCC.

Using AI to plan radiotherapy and manage post-liver transplantation is also maturing. The hierarchical feature fusion network (FA-Net) generated radiotherapy dose distributions close to the clinical standard (HI = 0.31, CI = 0.87) by integrating CT images with organ contours[150]. For recurrence prediction after liver transplantation, the deep learning model had an AUC of 0.86 based on preoperative factors (*e.g.*, tumor diameter and AFP level), with a 5-year recurrence-free survival rate of 92.6% in the low-risk group[151].

**Deep integration of gut microbes and machine learning: Decoding cross-disease signatures to drive precision therapies**

In the study of gut microbiology in inflammatory bowel disease, machine learning techniques have systematically deconstructed the gut-type-specific pathogenicity network of ulcerative colitis (UC). Cluster analysis of 16S rRNA data from 11 cohorts by a deep neural network (DNN) defined for the first time the three enterotypes of *Enterobacteriaceae* (ET-B), *Trichoderma* (ET-L), and *Clostridiales* (ET-C): Among them, patients with ET-B had abnormally elevated abundance of *Rummatococcus* (*R. gnavus*), which was positively correlated with the proliferation of *C. difficile* ( $P < 0.01$ ), and the Machine learning-guided metabolic pathway analysis revealed that *O. splanchnicus* and *B. uniformis* exerted a protective effect by activating the AMPK signaling pathway[152]. A sparse partial least squares discriminant analysis (sPLS-DA) was further used to construct a prediction model for active UC, which maintained more than 90% accuracy even when only 5% of the feature volume was used, and to establish *Bifidobacterium bifidum*/*Haemophilus parainfluenzae* as a negative/positive marker of disease activity[153]. Machine learning algorithm-based analysis of multicohort fecal macrogenomic data revealed that microbial gene markers in Crohn's disease (CD) patients had significantly better diagnostic performance than species and single nucleotide variants (SNVs). The gene diagnostic model constructed by machine learning performed optimally in cross-geographic validation (mean AUC 0.91) and targeted the key gene (*celB/manY*) of the phosphotransferase system (PTS), whose specificity was experimentally confirmed, revealing the potential of microbial functional genes for AI-driven precision diagnosis[154].

In terms of diagnostic technology innovation, CatBoost algorithm increased the colorectal precancer (AA) detection accuracy to 87.27% by integrating bacterial-viral two-dimensional features, identified *Prevotella sp900557255* and phage *Felixounavirus* as early warning markers, and the combined typing strategy enabled the prediction accuracy to exceed 98%[155]; while the iterative random forest model revealed the characteristics of pancreatic cancer metastasis-associated flora, and found that changes in the abundance of six genera, including *Anaero stipes hadrus*, were significantly correlated with the enrichment of Gram-negative bacteria (OR = 2.34,  $P = 0.007$ )[156]. A

neural network model based on 20 characteristic microorganisms combined with nanopore sequencing technology enabled rapid detection of hepatic encephalopathy (HE) (84% specificity), which drove the optimization of diagnostic and treatment protocols in 40% of cases[157]; while hepatocellular carcinoma (HCC) studies constructed a prognostic model with an AUC of 81% by integrating microbiome-transcriptome data through randomized forests, which reveals that *Mycobacterium anisopliae* spp. mediated tumor through bile acids key mechanism of immune microenvironment remodeling[158]. It is worth emphasizing that machine learning-enabled cross-disease meta-analysis identified significant overlap of flora characteristics between Crohn's disease and colorectal cancer, and between Parkinson's disease and type 2 diabetes mellitus (Jaccard's index > 0.65), which provides a new paradigm for cross-disease therapeutic target discovery[159]. This paradigm-shifting integration of gut microbiome-artificial intelligence symbiosis, as exemplified in **Figure 2**, not only deciphers conserved microbial signatures across diseases (*e.g.*, bile acid-immune axis remodeling in HCC) but also establishes a computational framework for translating multi-omics interactions into clinically actionable biomarkers and therapeutic blueprints.

### **Artificial intelligence-driven: An innovative antibiotic R&D paradigm to crack the drug-resistant bacteria crisis**

The threat to human health posed by Antimicrobial Resistance (AMR) has evolved from a theoretical warning to an urgent public health crisis. The Lancet research team modeled that by 2050, <sup>3</sup>1.91 million deaths globally will be directly attributed to AMR, with another 8.22 million deaths significantly associated with it[160]. However, in the face of “key priority pathogens”[161] represented by carbapenem-resistant Enterobacteriaceae (CRE) and third-generation cephalosporin-resistant Enterobacteriaceae (3GCRE), the antibiotic development pipeline is Severe lag: Only 13 new antibiotics targeting priority pathogens have been approved globally so far in 2017[160, 161], of which only 5 have demonstrated *in vitro* activity against CRE/3GCRE,

and some of them are difficult to be widely used due to rapidly induced resistance or dose-limiting toxicity[162, 163]. This paradox highlights the need to approach the gastrointestinal ecosystem, the source of infection. The gut is not only a major colonization site for pathogenic enterobacteria, but also a key biological interface for them to acquire drug resistance, spread and initiate systemic infections.

Excitingly, driven by artificial intelligence, researchers have constructed numerous ways to quickly find new and efficient antibiotics. In response to the threat of multidrug-resistant ESKAPE pathogens, traditional high-throughput screening (HTS) is becoming insufficient due to the limitations of bacterial penetration barriers and resistance mechanisms[164]. In this context, the integration of deep learning and big data technologies significantly improves the efficiency of antibiotic discovery. For example, virtual screening of 1.4 billion compounds by out-of-distribution generalization techniques in the GNEprop model identified 82 antimicrobially active molecules with a 90-fold increase in hit rate compared to conventional HTS, and most of the structures were significantly different from those of known antibiotics[165]. Similarly, the AMPSphere platform, constructed based on global microbiome data, utilized machine learning to mine 863000 non-redundant antimicrobial peptides (AMPs), of which 79% of the synthetically validated peptides target drug-resistant bacteria through membrane disruption mechanisms, providing an open repository for antibiotic development[166].

It is clear that deep learning models show unique advantages in discovering novel antibiotic structures. The research team screened structurally unique halicin from Drug Repurposing Hub, which showed potent bactericidal activity against both carbapenem-resistant *Enterobacteriaceae* and *Acinetobacter baumannii*[167]. To further break through the limitations of the “black box model”, interpretable graphical neural networks were used to resolve chemical substructures related to antibiotic activity, successfully targeting lead compounds from 283 candidate molecules to inhibit methicillin-resistant *S. aureus* and vancomycin-resistant *enterococci*, which significantly reduced pathogen load in a mouse infection model[168]. Innovative strategies such as “molecular

extinction” have further expanded antibiotic sources. Ancient proteome mining techniques based on the panCleave random forest model screened stable, non-toxic antimicrobial peptides from extinct molecules and validated their efficacy in a model of *Acinetobacter baumannii* infection[169]. Against this persistent drug-resistant bacterium, machine learning has screened for the narrow-spectrum compound abaucin, which achieves precise control in a mouse wound model by interfering with the lipoprotein transport mechanism[170]. These breakthroughs not only provide a new pathway of “source blocking - precise intervention” to cope with the antimicrobial resistance crisis, but also promote the transition of medical research from single-pathogen targeting to multi-omics intelligent regulation system, and its technological framework lays a methodological foundation for the future development of structural innovative drugs and personalized flora therapy.

## **DISCUSSION**

Artificial intelligence technology has shown remarkable potential in the diagnosis and treatment of gastrointestinal diseases, especially in the field of gastrointestinal and liver tumors, where several major breakthroughs have been achieved. However, technology diffusion still faces common challenges: First, the lack of interpretability caused by the black-box nature of the algorithms undermines clinical trust; second, the lack of data and heterogeneity (including endoscopic device differences and cross-center data bias) lead to limitations in model generalization; and third, the chain of evidence for clinical translation is still incomplete as existing studies generally lack large-scale randomized controlled trials and prospective validation[111, 171-175]. In addition, the validation of molecular markers discovered by AI technology needs to expand the sample size, and the technology implementation needs to address the issue of geographic medical appropriateness[148, 149]. The breakthrough of these bottlenecks will determine the process of AI technology leapfrogging from experimental results to clinical routine applications.

In response to current challenges in model interpretability, for Gradient-weighted Class Activation Mapping (Grad-CAM) and Shapley value analysis were integrated into diagnostic systems to assist clinicians in understanding the AI decision basis[176]. Grad-CAM generates a heat map to visualize the basis of model decision-making by quantifying the gradient of the target category probability relative to the final convolutional layer feature map. Its implementation is divided into three steps: First, calculate the gradient of the target category scores with respect to a specific convolutional layer feature map to obtain the importance weights of each channel; second, perform channel-weighted summation of the feature map to generate a coarse-grained heatmap; and finally, upsample the heatmap to the input image size by bilinear interpolation to highlight the key regions. The method does not require modification of the model structure and is suitable for all types of convolutional neural networks. In biomedical research, Grad-CAM can localize molecular pathways driving classification (*e.g.*, the KRAS signaling pathway in HPV status prediction) or regions of differential brain function (*e.g.*, the default mode network of schizophrenic patients) to provide biologically plausible validation of model decisions. Its advantage lies in combining gradient information with spatial features to generate intuitive and interpretable heat maps while maintaining high resolution and class specificity, making it an important tool for connecting deep learning “black boxes” with domain knowledge[177, 178]. An interpretable three-stage lightweight deep learning framework (PSE-CNN-PCA-DELM) achieves 97.24% classification accuracy and 99.38% ROC-AUC metrics in a multi-level classification task of gastrointestinal diseases by integrating XAI (interpretable AI) technologies, such as Grad-CAM visualization and heat map analysis. With a model volume of only 14.88MB and extremely fast inference performance of 59ms, the framework breaks through the dependence of traditional AI models on high-performance arithmetic, and can be efficiently deployed in edge devices such as portable ultrasound machines and low arithmetic terminals, which is especially suitable for primary care scenarios and remote areas with a lack of resources, and can help narrow the geographic disparity of healthcare resources through the sinking of

transparent and trusted AI-assisted diagnostic capabilities, contributing to promote AI-driven healthcare fairness provides a practical technical solution[19]. At the application level, AI has penetrated into multiple scenarios such as automatic ultrasound scanning, OCT retinal stratification analysis, and digital pathology nuclear fission counting, in which the EndoMind system has achieved 94.6% sensitivity in real-time cancer detection in gastrointestinal endoscopy, which significantly reduces the leakage of early lesions[179]. Interpretable AI technology fused with lightweight design is driving the transformation of smart healthcare to credible and practical with high precision and transparent decision-making, reshaping the human-machine collaborative diagnosis and treatment model.

In addressing the lack of data and heterogeneity, multi-task networks (*e.g.*, TransMT-Net) combined with active learning can maintain high accuracy with small samples (96.94% classification accuracy and 77.76% segmentation DSC)[180]; and the course self-supervised learning framework utilizes unlabeled images to improve classification performance (73.39% F1 score)[181]. In addition, the migration learning strategy significantly alleviates the dilemma of annotated data scarcity in the medical field. In a deep migration learning study for catheter-dependent congenital heart disease (CHD) screening, the DDCHD-DenseNet model based on a two-stage migration strategy achieves the highest sensitivity for critical CHD detection by integrating the multi-center data of 6698 images and 48 videos 0.973 and specificity 0.985, and its hierarchical architectural design significantly improves cross-center generalization, providing an innovative solution for computer-aided hierarchical diagnosis of fetal heart defects in low-resource areas and model scalability[182]. And the breakthrough application of generative adversarial networks is reflected in two aspects: On the one hand, it can synthesize augmented training data, and on the other hand, it can realize cross-modality image conversion (*e.g.*, MRI to CT) by CycleGAN, which can effectively support multi-center studies[183, 184]. In one of the largest federated machine learning (FL) studies to date, researchers integrated data from 6314 patients from 71 healthcare sites across 6 continents to construct an automated tumor boundary detection model.

The framework effectively integrates heterogeneous datasets from multiple sources without sharing the original data by weighting and averaging the encrypted model parameters across sites *via* a central aggregation server. Compared with the model trained only on publicly available datasets, the federated learning model achieves significant breakthroughs in surgical target area detection (33% enhancement) and overall tumor extent identification (23% enhancement), which directly corroborates the critical role of distributed data aggregation in breaking through the data size limitations of a single center. In the study, a pre-training model initialization strategy was adopted to accelerate convergence, and data diversity was gradually improved through phased training (from the initial model at 16 sites to the final consensus model at 71 sites), successfully constructing a cross-regional privacy-preserving collaborative network in rare disease scenarios, which provides a scalable technological paradigm for solving the problem of scarcity of diagnostic resources in low-resource regions[185]. In a multicenter study of prostate MRI segmentation, three institutions (University of California, Los Angeles, State University of New York Upstate Medical Center, and National Cancer Institute) integrated their respective private datasets (100 cases of T2-weighted MRI images each) through federated learning, and aggregated the model weights by weighted averaging without sharing the original data, so that the federated model on the external dataset (343 cases) demonstrated significantly higher performance than single-agency models[186]. Furthermore, in breast cancer diagnosis, federated learning demonstrates high diagnostic efficacy by integrating data from multiple sources and protecting privacy, enabling model accuracy of 98.9% and 95.3%[187, 188]. However, no study has applied federated learning to the diagnosis and treatment of gastrointestinal diseases. For this reason, as shown in **Figure 3**, we design a federated learning-based collaborative framework, which relies on a central aggregation server and optimizes the encryption parameter ( $\Delta\theta_i$ ) by using the weighted average method, in order to integrate multi-site heterogeneous datasets, and then generates a more generalized consensus model  $\theta^{(k+1)}$ , which can both ensure privacy and security of gastrointestinal data and provide a scalable technology paradigm for cross-center AI

collaboration. process ensures the privacy and security of gastrointestinal data while providing a scalable technical paradigm for cross-center AI collaboration. In the future, there is a need to promote multi-center collaboration, standardized data collection, and explore the deep integration of AI with multiple emerging technologies[189, 190]. These cutting-edge machine learning technologies have significantly promoted the scientific process of rare disease diagnosis and treatment by breaking through data scarcity and collaboration barriers: Migration learning adapts multi-center heterogeneous data with hierarchical architecture, generative adversarial network breaks through the double limitation of image modality and sample size, and federated learning builds a cross-regional privacy-preserving collaboration network, which jointly cracks the dilemma of the scarcity of diagnostic resources in low-resource areas.

With the clinical validation and scale deployment of AI-assisted decision-making tools (e.g., the UC-SCALE standardized scoring system)[82, 191], complex disease screening technologies are being embedded into different tiers of the healthcare system in an algorithmically standardized manner to promote equal coverage of diagnostic capabilities from the bottom of the technology. In resource-constrained scenarios, the technology integration strategy of “lightweight + localization + privacy protection” is especially critical: Lightweight neural network combined with edge computing architecture realizes offline inference, ensuring stable operation on terminal devices in underdeveloped regions; the federated learning framework supports the construction of cross-institutional diagnosis and treatment knowledge, avoiding the risk of leaking the original data; and the rare case data synthesized by the generative adversarial network complements the regional database. Together, the three can build an “end - side - cloud” collaborative computing architecture to provide sustainable technology empowerment for primary care. These technological advances not only improve the efficiency of diagnosis and treatment, but also reconstruct the medical ecology in terms of deep logic: Through algorithmic optimization of resource allocation model, innovation of medical knowledge discovery mechanism, and promotion of adaptive evolution of ethical framework, AI is upgrading from an auxiliary tool to an

infrastructure for health equity. In the future, with the improvement of the technology-ethics synergy, machine learning is expected to become the core driving force to break through the plight of global health inequality, and promote the meta-goal of "health equity" from theoretical conception to practical realization.

## CONCLUSION

### **4 Conclusion**

Artificial Intelligence has demonstrated revolutionary potential in the fields of gastroenterology and hepatology, significantly improving the diagnostic accuracy and personalized treatment of GI diseases through image analysis, pathology, multimodal data integration and molecular marker mining. However, the technology landing and further development still face challenges such as data heterogeneity, insufficient data volume, insufficient algorithm interpretability and ethical regulation.

# 2%

SIMILARITY INDEX

---

### PRIMARY SOURCES

---

- |   |   |                 |
|---|---|-----------------|
| 1 | <a href="https://discovery.researcher.life">discovery.researcher.life</a><br>Internet   | 38 words — < 1% |
| 2 | <a href="https://profiles.stanford.edu">profiles.stanford.edu</a><br>Internet   | 26 words — < 1% |
| 3 | Zhou-Li Cheng, Shuyuan Zhang, Zhenning Wang, Aixia Song et al. "Pathogen-derived glyoxylate inhibits Tet2 DNA dioxygenase to facilitate bacterial persister formation", Cell Metabolism, 2025<br>Crossref | 15 words — < 1% |
| 4 | <a href="https://portal.brint.com">portal.brint.com</a><br>Internet   | 15 words — < 1% |
| 5 | Amruta Shankar Chougule, Aparna Annaso Patil, Indrajeet Gupta, Shashidhar Gurav, Govind Singh Patel. "chapter 17 Enhancing Healthcare Efficiency Through Edge Computing", IGI Global, 2025<br>Crossref    | 14 words — < 1% |
| 6 | <a href="https://pdfcookie.com">pdfcookie.com</a><br>Internet   | 13 words — < 1% |
| 7 | <a href="https://pharmaceuticalintelligence.com">pharmaceuticalintelligence.com</a><br>Internet   | 13 words — < 1% |
| 8 | <a href="https://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a><br>Internet   | 13 words — < 1% |

---

9

John M H Bennett, Samir Mehta, Michael Rhodes.  
"Surgery for morbid obesity", Postgraduate  
Medical Journal, 2007

Crossref

12 words — < 1%

---

EXCLUDE QUOTES ON

EXCLUDE SOURCES OFF

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES < 12 WORDS