

REVIEW

Open Access



# Application of artificial intelligence in the diagnosis of malignant digestive tract tumors: focusing on opportunities and challenges in endoscopy and pathology

Yinhu Gao<sup>1†</sup>, Peizhen Wen<sup>2\*†</sup>, Yuan Liu<sup>3†</sup>, Yahuang Sun<sup>4†</sup>, Hui Qian<sup>5</sup>, Xin Zhang<sup>6</sup>, Huan Peng<sup>4</sup>, Yanli Gao<sup>7</sup>, Cuiyu Li<sup>8</sup>, Zhangyuan Gu<sup>9</sup>, Huajin Zeng<sup>2</sup>, Zhijun Hong<sup>9</sup>, Weijun Wang<sup>6\*</sup>, Ronglin Yan<sup>5\*</sup>, Zunqi Hu<sup>6\*</sup> and Hongbing Fu<sup>6\*</sup>

## Abstract

**Background** Malignant digestive tract tumors are highly prevalent and fatal tumor types globally, often diagnosed at advanced stages due to atypical early symptoms, causing patients to miss optimal treatment opportunities. Traditional endoscopic and pathological diagnostic processes are highly dependent on expert experience, facing problems such as high misdiagnosis rates and significant inter-observer variations. With the development of artificial intelligence (AI) technologies such as deep learning, real-time lesion detection with endoscopic assistance and automated pathological image analysis have shown potential in improving diagnostic accuracy and efficiency. However, relevant applications still face challenges including insufficient data standardization, inadequate interpretability, and weak clinical validation.

**Objective** This study aims to systematically review the current applications of artificial intelligence in diagnosing malignant digestive tract tumors, focusing on the progress and bottlenecks in two key areas: endoscopic examination and pathological diagnosis, and to provide feasible ideas and suggestions for subsequent research and clinical translation.

**Methods** A systematic literature search strategy was adopted to screen relevant studies published between 2017 and 2024 from databases including PubMed, Web of Science, Scopus, and IEEE Xplore, supplemented with searches of early classical literature. Inclusion criteria included studies on malignant digestive tract tumors such as esophageal

<sup>†</sup>Yinhu Gao, Peizhen Wen, Yuan Liu and Yahuang Sun contributed equally to this work.

\*Correspondence:

Peizhen Wen  
wenpeizhen1996@163.com

Weijun Wang  
610207494@qq.com

Ronglin Yan  
yanronglin@smmu.edu.cn

Zunqi Hu  
xiaoqiye@163.com

Hongbing Fu  
fuhongbing@smmu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

cancer, gastric cancer, or colorectal cancer, involving the application of artificial intelligence technology in endoscopic diagnosis or pathological analysis. The effects and main limitations of AI diagnosis were summarized through comprehensive analysis of research design, algorithmic methods, and experimental results from relevant literature.

**Results** In the field of endoscopy, multiple deep learning models have significantly improved detection rates in real-time polyp detection, early gastric cancer, and esophageal cancer screening, with some commercialized systems successfully entering clinical trials. However, the scale and quality of data across different studies vary widely, and the generalizability of models to multi-center, multi-device environments remains to be verified. In pathological analysis, using convolutional neural networks, multimodal pre-training models, etc., automatic tissue segmentation, tumor grading, and assisted diagnosis can be achieved, showing good scalability in interactive question-answering. Nevertheless, clinical implementation still faces obstacles such as non-uniform data standards, lack of large-scale prospective validation, and insufficient model interpretability and continuous learning mechanisms.

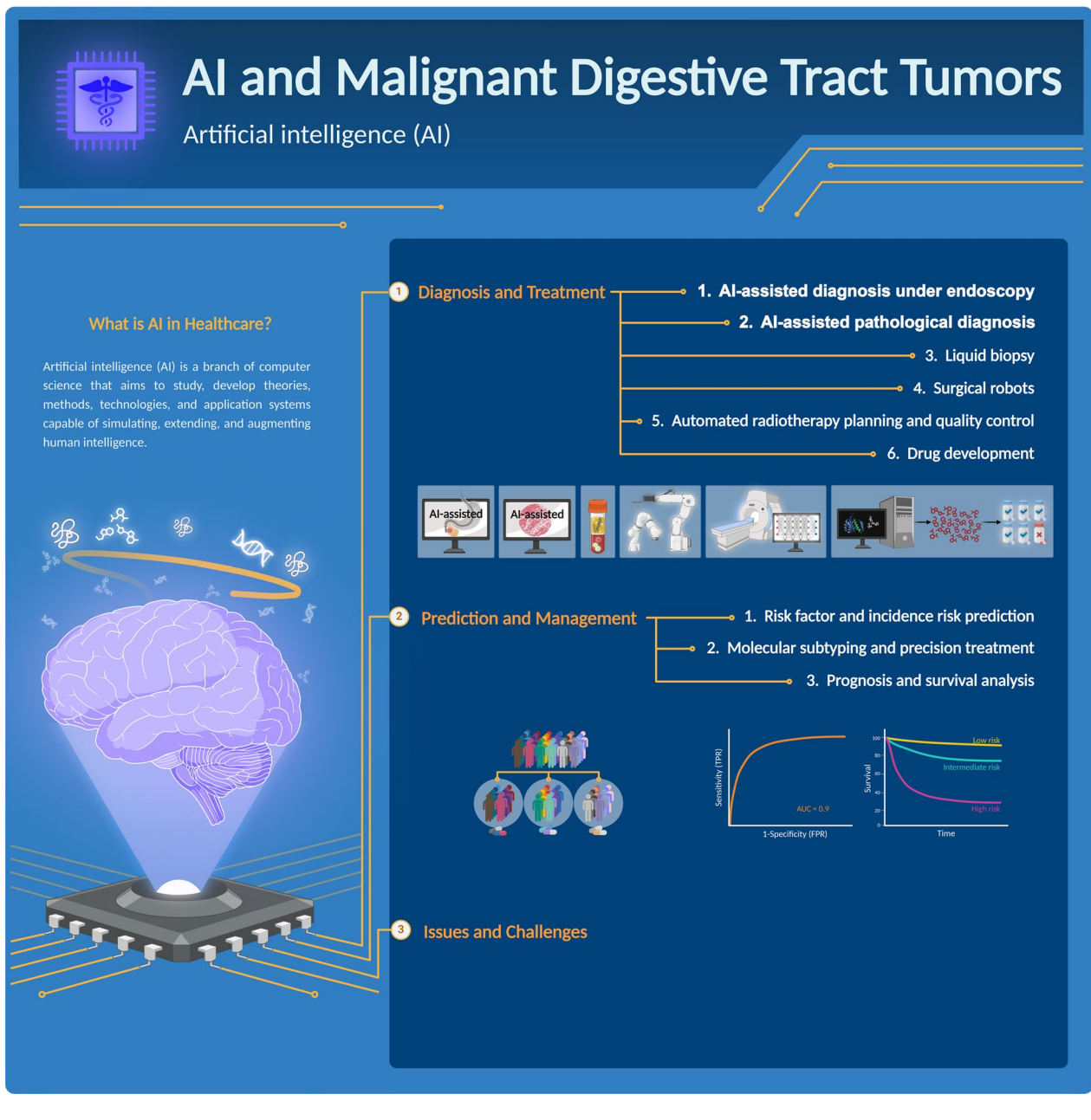
**Conclusion** Artificial intelligence provides new technological opportunities for endoscopic and pathological diagnosis of malignant digestive tract tumors, achieving positive results in early lesion identification and assisted decision-making. However, to achieve the transition from research to widespread clinical application, data standardization, model reliability, and interpretability still need to be improved through multi-center joint research, and a complete regulatory and ethical system needs to be established. In the future, artificial intelligence will play a more important role in the standardization and precision management of diagnosis and treatment of digestive tract tumors.

## Highlights

1. Early symptoms of malignant digestive tract tumors are often atypical, resulting in high misdiagnosis rates, which urgently calls for more precise diagnostic methods; artificial intelligence has demonstrated significant application potential in the two core areas of endoscopy and pathology.
2. Real-time endoscopic assisted detection systems driven by deep learning can significantly improve the detection rate of early lesions, reducing the risk of missed diagnoses due to physician inexperience or fatigue.
3. Pathological AI technologies based on multimodal and vision-language pre-training models can achieve automatic segmentation, grading, and interactive diagnosis of digital slides, providing objective quantitative basis for individualized treatment decisions.
4. The main obstacles to current AI applications in endoscopy and pathology include insufficient data standardization, poor model interpretability, and lack of large-scale prospective validation; multi-center collaboration and standardized regulation urgently need to be strengthened.
5. With the continued advancement of multidisciplinary integration and technological breakthroughs, artificial intelligence is expected to further improve early diagnosis and precise management of digestive tract tumors, enhancing patient prognosis and promoting the standardized development of diagnostic and treatment processes.

**Keywords** Malignant digestive tract tumors, Endoscopy, Pathology, Artificial intelligence, Deep learning, Diagnosis

## Graphical Abstract

**Introduction****Epidemiology and current clinical status of malignant digestive tract tumors**

Malignant digestive tract tumors are among the malignant tumors with high incidence and mortality rates globally. According to the Global Cancer Observatory (GLOBOCAN) 2020 estimates, new cases of malignant digestive tract tumors (including esophageal cancer

(3.1%), gastric cancer (5.6%), and colorectal cancer (10%) account for approximately 18.7% of all new malignant tumor cases, with colorectal cancer and gastric cancer ranking third and fifth in incidence, respectively [1]. China is a high-incidence area for malignant digestive tract tumors, with consistently high incidence and mortality rates. Esophageal cancer, gastric cancer, and colorectal cancer rank among the top six in incidence for

both men and women, constituting a major public health problem that seriously threatens population health and life [2].

Currently, clinical diagnosis and treatment of malignant digestive tract tumors cover multiple aspects, including endoscopic examination, imaging examination, pathological diagnosis, surgical treatment, radiotherapy and chemotherapy, molecular targeted therapy, and immunotherapy. Despite rapid developments in medical technology, clinical diagnosis and treatment still face many challenges due to the inherent heterogeneity and complexity of tumors: (1) Low early diagnosis rate: Early symptoms of malignant digestive tract tumors are atypical and easily overlooked by patients and doctors, resulting in most patients being diagnosed at advanced stages, missing the optimal treatment window. This highlights the importance and urgency of early screening and diagnosis [3]; (2) Difficulties in precise diagnosis: Some tumors lack specific manifestations under endoscopy, and pathological diagnosis is highly dependent on physician experience, making the risk of misdiagnosis and missed diagnosis non-negligible. Pathological diagnosis of malignant tumors faces multiple challenges, including morphological and functional heterogeneity, standardization of specimen collection and processing, and consistency of diagnostic standards [4]; (3) Poor treatment effects: Patients with advanced disease have poor prognosis, and traditional single treatment modalities have limited efficacy. There is an urgent need to explore new treatment targets and strategies from aspects such as molecular typing, immune microenvironment, and host characteristics to improve efficacy and prolong survival [5–7]; (4) Lack of standardized management: China still lacks standardized guidelines and processes for diagnosis, treatment, and follow-up of digestive tract tumors. The varied levels of diagnostic and treatment capabilities, equipment configuration, and talent cultivation in primary healthcare institutions affect patients' long-term management and quality of life.

Given these clinical challenges, there is an urgent need for innovative technologies and emerging methods to enable precise diagnosis and treatment of malignant digestive tract tumors and improve patient prognosis. Artificial intelligence (AI) technology, with its powerful data analysis and massive information processing capabilities, has the potential to solve clinical problems from multiple dimensions: by using AI algorithms to build efficient and accurate computer-aided diagnostic systems to standardize endoscopic and pathological diagnostic processes; combining liquid biopsy technology to improve early diagnostic accuracy (Appendix 1); integrating surgical robot technology to enhance surgical precision and safety (Appendix 2); optimizing automated

radiotherapy planning and quality control to reduce complications (Appendix 3); and accelerating drug development to promote precision medication (Appendix 4). Additionally, AI demonstrates enormous potential in prognosis prediction and management of malignant digestive tract tumors. By analyzing large amounts of clinical data, AI can more accurately identify risk factors and predict disease risk, providing early intervention recommendations for high-risk populations (Appendix 5). During treatment, AI-assisted molecular typing can provide more precise individualized treatment plans for patients (Appendix 6). Meanwhile, machine learning-based prognosis and survival analysis models can provide more reliable evidence for clinical decision-making, helping to improve long-term patient prognosis (Appendix 7). Therefore, in-depth exploration of AI applications throughout the diagnosis and treatment process of digestive tract malignant tumors, including diagnosis, surgery, radiotherapy, drug development, and prognosis management, has important scientific significance and practical value for improving diagnostic efficiency, enhancing prognosis, and promoting standardized management.

#### Overview of artificial intelligence technology development

Artificial Intelligence (AI) is an important branch of computer science dedicated to researching and developing theories, methods, technologies, and application systems that can simulate, extend, and expand human intelligence [8, 9]. Its core goal is to enable machines to perform perception, cognition, decision-making, and task execution similar to humans. The prototype of this concept can be traced back to 1950, when Alan Turing published "Computing Machinery and Intelligence," proposing the Turing test, which laid the foundation for AI research. The term "artificial intelligence" was formally proposed by John McCarthy and other scientists at the Dartmouth Conference in 1956 [10] (Fig. 1, Appendix 8).

From the late twentieth century to the early twenty-first century, AI research experienced multiple peaks and valleys. The rise of expert systems in the 1980s marked the first peak, but it fell into a trough due to knowledge acquisition bottlenecks and robustness issues. Since the twenty-first century, thanks to improved computational capabilities and the accumulation of massive data, machine learning (ML) and deep learning (DL) have driven the revival of AI. In 2006, Hinton et al. proposed deep belief networks, initiating a wave of deep learning research. In 2012, marked by Krizhevsky et al.'s breakthrough in the ImageNet competition, deep convolutional neural networks (CNN) demonstrated excellent image classification performance through the AlexNet model [11]. Deep learning algorithms extract high-level features through multi-layer non-linear transformations.



## The history of artificial intelligence: timeline

1950	- Alan Turing publishes "Computing Machinery and Intelligence," introducing the Turing test and opening the door to AI research.
1951	- Marvin Minsky and Dean Edmonds develop the first artificial neural network, SNARC.
1956	- John McCarthy and others coin the term "artificial intelligence" at a workshop.
1958	- Frank Rosenblatt develops the perceptron, laying the foundation for modern neural networks.
1959	- Arthur Samuel first uses the term "machine learning."
1964	- Daniel Bobrow develops STUDENT, an early natural language processing program.
1965	- Edward Feigenbaum and others develop Dendral, the first expert system.
1968	- Terry Winograd creates SHRDLU, the first multimodal AI.
1969	- Arthur Bryson and Yu-Chi Ho describe the backpropagation learning algorithm.
1980	- Symbolics Lisp machines are commercialized, marking an AI renaissance.
1981	- Danny Hillis designs parallel computers for AI.
1985	- Judea Pearl introduces Bayesian networks for causal analysis.
1989	- Yann LeCun and others demonstrate the use of convolutional neural networks (CNNs) in handwritten character recognition.
1997	- Sepp Hochreiter and Jürgen Schmidhuber propose Long Short-Term Memory (LSTM) networks.
2000	- University of Montreal researchers publish a paper on neural probabilistic language models.
2006	- Fei-Fei Li begins work on the ImageNet visual database.
2011	- Jürgen Schmidhuber and others develop the first CNN to achieve "superhuman" performance.
2012	- Geoffrey Hinton and others propose a deep CNN architecture, winning the ImageNet challenge.
2013	- DeepMind introduces deep reinforcement learning.
2014	- Ian Goodfellow and others invent generative adversarial networks (GANs).
2017	- Stanford researchers publish work on diffusion models.
-	- Google researchers introduce the concept of transformers in a paper.
2018	- OpenAI releases the GPT language model.
2019	- Microsoft launches the Turing Natural Language Generation model.
2020	- OpenAI releases GPT-3 language model.
-	- DeepMind's AlphaFold system wins the protein structure prediction contest.
2021	- OpenAI introduces the multimodal AI system Dall-E.
2022	- OpenAI releases ChatGPT in November.
2023	- OpenAI announces the multimodal language model GPT-4.

**Fig. 1** Timeline of major developments in artificial intelligence. This figure illustrates the historical timeline of artificial intelligence development from 1950 to 2023, highlighting key milestones including the introduction of fundamental concepts, breakthrough technologies, and significant achievements in AI research and applications. Notable events include the establishment of AI as a field in 1956, the development of neural networks, the emergence of deep learning, and recent advances in large language models

This breakthrough greatly enhanced AI's performance in perception and cognition, opening up new possibilities for solving complex tasks [12–15].

In the field of computer vision, convolutional neural networks (CNNs) have become a typical application of deep learning (Appendix 8) [16–20]. CNNs effectively

capture image features using convolutional layers and pooling layers by simulating the human visual system. Their local connectivity and weight sharing characteristics make them more efficient when processing large numbers of images, significantly improving the accuracy of image recognition and classification. In recent

years, other visual algorithms such as generative adversarial networks (GAN) have also made progress in medical image synthesis and data augmentation. In the field of natural language processing (NLP), the emergence of large language models (LLMs) marks a breakthrough in AI's language understanding and generation capabilities (Appendix 8). These models (such as ChatGPT [21], Claude [22], and the Llama series [23, 24]) learn rich language and world knowledge through self-supervised learning on massive text data, enabling them to understand and generate natural language and perform tasks such as question-answering, translation, and summarization. The emergence of LLMs has advanced the development of natural language processing technology, bringing new tools for medical text data analysis and clinical decision support.

In summary, these advances in artificial intelligence technology have brought revolutionary changes to various industries, particularly significant in the medical and health fields. The accumulation of massive medical big data has laid a solid foundation for the application of AI technology in the medical field. Through deep learning algorithms and large language models analyzing medical data, AI can assist in the entire process of disease prevention, diagnosis, treatment, and management. It has the potential to break through the limitations of traditional medical models and address challenges in current clinical practice such as standardization, precision, and intelligence [25–27].

#### **Recent progress of AI applications in the medical field**

With powerful data processing and analytical capabilities, AI shows enormous application potential in various aspects of healthcare and is expected to drive the transformation of medical practice toward precision medicine. Currently, the main applications of AI in the medical field include: (1) Medical image analysis: Deep learning-based image segmentation, classification, and detection algorithms can automatically identify organs and pathological structures in medical images, assisting in diagnosis and efficacy evaluation. In radiology, AI can be used for screening and diagnosing diseases such as nodules and tumors; in pathology, AI can achieve automatic analysis and diagnosis of tissue pathology slides. (2) Clinical decision support: Machine learning algorithms can extract features from massive clinical data such as electronic medical records, physician orders, and laboratory reports to establish disease diagnosis and prognosis prediction models, assisting clinical decision-making. For example, based on NLP technology, AI can automatically structure medical record information and intelligently recommend examination and medication plans [28–30]. (3) Drug development: AI can be applied

to drug molecule screening, virtual drug screening, drug repositioning, etc., accelerating the new drug development process. For example, deep learning algorithms can screen lead compounds with activity from massive molecular libraries; AI can predict drug toxicity, reducing the risk of clinical trial failure [31–35]. (4) Individualized treatment: Machine learning based on multi-omics data can perform molecular typing of patients, predict specific treatment responses, and optimize individualized plans [36–40]. For example, in tumor treatment, AI integrates multi-omics data such as genomics, transcriptomics, and proteomics to construct prognosis and efficacy prediction models, guiding precision therapy [41, 42]. (5) Intelligent health management: Wearable devices and mobile healthcare generate large amounts of health data, and AI can assess individual health status, enabling early disease warning and health management. For example, intelligent wearable ECG monitoring devices analyze ECG data in real-time, providing early warning of cardiovascular diseases such as arrhythmia [43, 44].

Although medical AI has made significant progress, its application in clinical practice still faces many challenges, including data standardization and sharing, algorithm interpretability and robustness, ethics, and legal issues. Future research and application of medical AI requires integration of multiple disciplines such as medicine, information, and management, establishment of industry standards and application specifications, and the establishment of a sound clinical validation and evaluation system to ensure that AI technology benefits patients safely and effectively.

#### **Applications of artificial intelligence in the diagnosis and treatment of malignant digestive tract tumors**

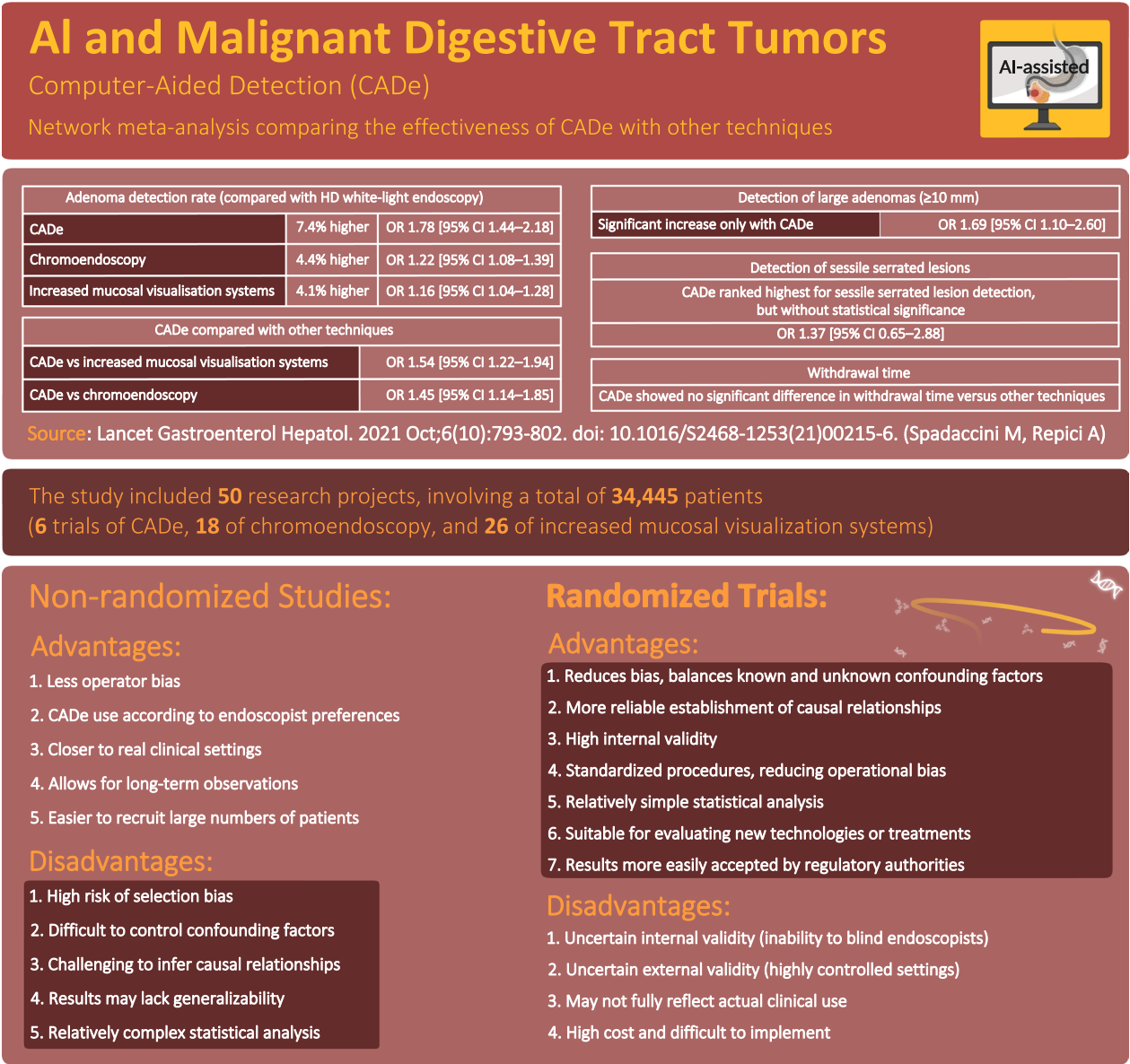
##### **AI-assisted endoscopic diagnosis**

Endoscopic examination is an important method for diagnosing malignant digestive tract tumors, but the identification of early lesions under endoscopy requires doctors to have high experience and technical expertise, with a miss rate as high as 10–20%. In recent years, artificial intelligence technology represented by deep learning has been widely applied in the field of endoscopy, which is expected to overcome the limitations of human eye recognition and improve the detection rate and diagnostic accuracy of digestive tract tumors.

**Upper Digestive Tract Tumors:** Upper digestive tract tumors mainly include esophageal cancer and gastric cancer. Under traditional endoscopy, these early lesions present atypically and are easily overlooked. AI-assisted diagnostic systems based on endoscopic images and videos can achieve automated, real-time detection and recognition of lesions, compensating for the deficiency

of insufficient physician experience. Horie Y et al. developed a deep learning-based esophageal cancer screening system that automatically detects esophageal squamous cell carcinoma and adenocarcinoma by intelligently analyzing endoscopic images, with a sensitivity as high as 98% [45]. Additionally, Tang D et al. constructed an in-situ diagnostic model for early esophageal squamous cell carcinoma based on real-time deep convolutional neural networks, which can accurately classify lesions under

endoscopy with an accuracy of 95.4% [46]. For gastric cancer, Horiuchi Y et al. developed an AI detection system for early gastric cancer and atrophic gastritis, which can determine the presence of lesions within 0.02 s by analyzing each frame of gastroscopy videos, with diagnostic accuracy and sensitivity of 85.3% and 95.4%, respectively [47]. Besides detecting early lesions, AI can also assist in determining the invasion depth of gastric cancer, providing references for surgical planning. Nagao



et al. used convolutional neural networks combined with transfer learning to train on endoscopic images and developed a diagnostic model for gastric cancer invasion depth, achieving accuracy and sensitivity of 94.5% and 84.4%, respectively [48].

**Lower Digestive Tract Tumors:** Lower digestive tract tumors are mainly colorectal cancer. Colonoscopy is the gold standard for screening and diagnosing colorectal cancer, but due to factors such as multiple colonic mucosal folds and narrow lumen, the miss rate can be as high as 22%. AI technology can fully utilize the information from each frame of colonoscopy videos, scan the intestine from all angles, identify small lesions easily overlooked by the human eye, and reduce missed diagnoses. Wallace MB et al. applied deep learning to construct a computer-aided diagnostic system for colonoscopy, showing that AI can reduce the miss rate of colorectal tumors by approximately twofold [49]. To achieve real-time prompting of lesions, Misawa M et al. developed a real-time detection system for colorectal polyps that can automatically mark suspicious polyp areas by analyzing colonoscopy video images, with a sensitivity of 90.0% [50]. Urban G et al. used convolutional neural networks for computer-aided analysis of colonoscopy images, confirming that AI can significantly improve adenoma detection rates with a detection accuracy as high as 96.4% [51]. Furthermore, AI can be combined with new endoscopic technologies to achieve more precise diagnosis. The AI model EndoBRAIN developed by Kudo SE et al. efficiently distinguishes between neoplastic and non-neoplastic colorectal lesions by analyzing endoscopic cytological staining images and Narrow-Band Imaging (NBI) images, with an accuracy of [52], providing a new approach for "optical biopsy."

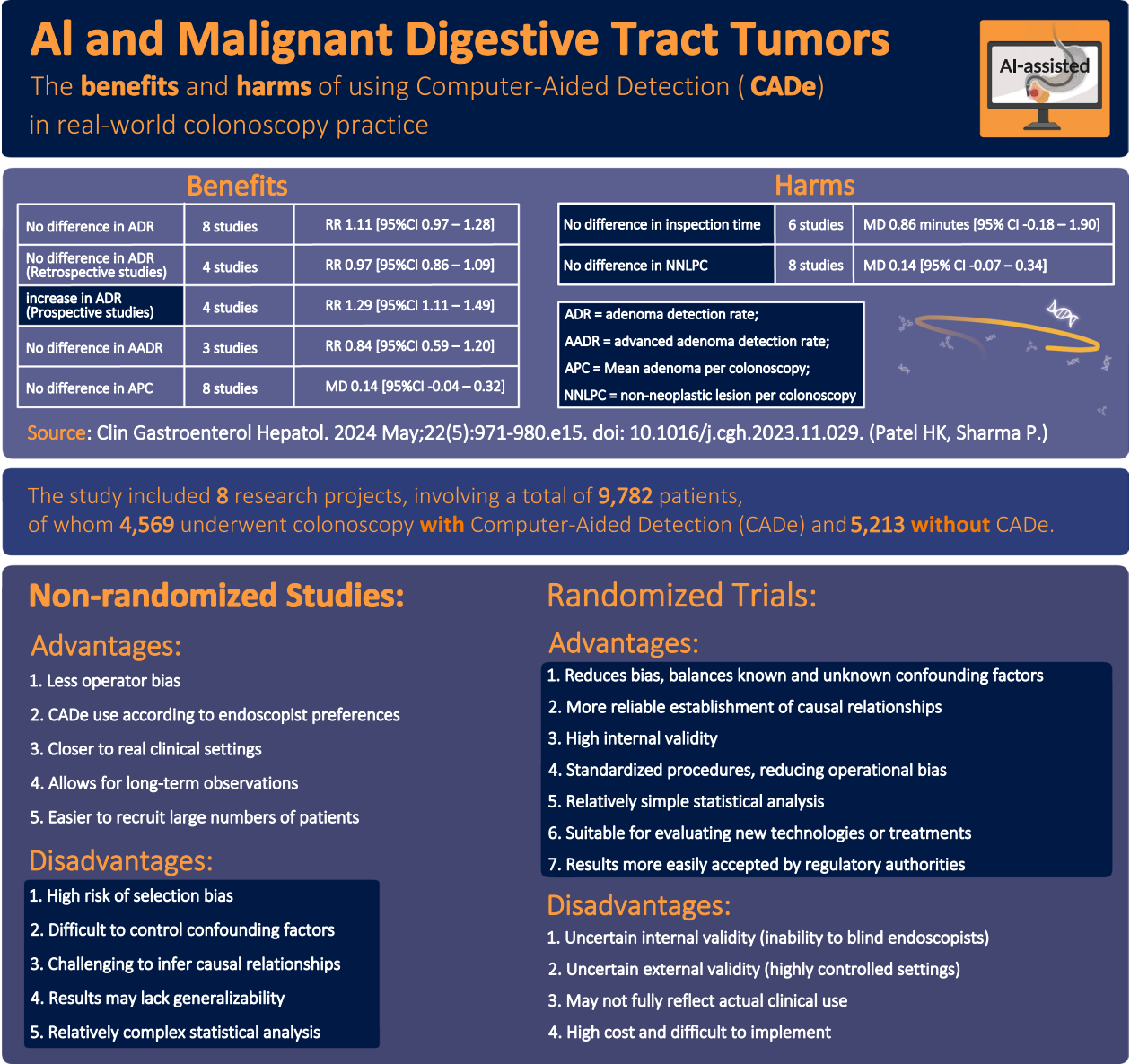
With the emergence of the above AI technology achievements, some AI systems have begun to enter clinical practice in the field of endoscopy. For example, GI Genius is the first real-time AI-assisted detection device approved by regulators for colonoscopy, which has proven to improve the detection rate of colorectal polyps in actual applications [53]. Spadaccini et al. conducted a network meta-analysis to evaluate the relative effectiveness of computer-aided detection (CADe) compared to other advanced endoscopic technologies in colorectal tumor detection [54]. The study included 50 randomized controlled trials involving 34,445 participants, using a frequentist framework and random effects model for systematic review (Fig. 2). The results showed that CADe significantly outperformed other technologies in adenoma detection rate and large adenoma ( $\geq 10$  mm) identification: compared to high-definition white-light endoscopy, the adenoma detection rate increased by 7.4%, and was also significantly better than enhanced

mucosal visualization (such as NBI) systems and chromoendoscopy. In the detection of serrated lesions, although CADe showed a trend of superiority over other strategies, the advantage did not reach statistical significance. This study is the first systematic review that directly compares the effectiveness of CADe with other advanced endoscopic technologies through network meta-analysis, providing strong empirical support for the application of CADe in clinical practice.

The core algorithms currently used for AI-assisted endoscopic diagnosis are mostly based on convolutional neural networks (CNN), such as ResNet, VGG, YOLO, or Faster R-CNN. These models are typically pre-trained on large endoscopic image datasets and then undergo transfer learning for specific lesions (such as colorectal polyps, early gastric cancer). In feature engineering, early methods relied on manually extracting image features such as texture and edges, while the current mainstream approach is to adopt end-to-end automatic feature extraction through deep learning, using multiple convolutional kernels to refine information on lesion patterns, mucosal textures, and color differences. In real-time detection scenarios, models often adopt lightweight networks combined with attention mechanisms (Attention Module) or real-time object detection frameworks such as SSD/YOLO to ensure a balance between inference speed and recognition accuracy.

Endoscopic AI models commonly use metrics such as Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Area Under the ROC Curve (AUC) to quantify performance. In actual screening, more attention is paid to changes in miss rates and Adenoma Detection Rate (ADR). To improve clinical feasibility, some studies additionally monitor false positive prompt rates to evaluate the operational burden on physicians. Although positive results have been achieved in single-center or small-scale multi-center studies, the cross-institutional generalizability of algorithms still needs further validation through large-scale, multi-geographic region clinical trials. Additionally, data bias and algorithmic inequality issues have gradually received attention, and validation across multi-center multi-ethnic populations will help reduce inconsistent algorithm performance across different populations.

We note that there is currently a lack of specific data on adverse consequences caused by errors or failures of AI endoscopy systems in the diagnosis and treatment of digestive tract tumors (Appendix 9). Understanding the potential risks of AI systems is crucial for their safe and effective application. Despite limited direct data on adverse impacts, some research results are inconsistent with the mainstream view that "AI is definitely beneficial." The latest meta-analysis by Patel HK et al.



**Fig. 3** Comparative analysis of CADe effectiveness in non-randomized studies. Meta-analysis results from 8 non-randomized controlled studies involving 9,782 patients, comparing outcomes between CADe-assisted (n=4,569) and standard colonoscopy (n= 5,213). The analysis found no significant differences in adenoma detection rate (ADR), advanced adenoma detection rate (AADR), mean adenomas per colonoscopy (APC), inspection time, and non-neoplastic lesions per colonoscopy (>NNLPC) in retrospective studies

(Fig. 3) included 8 non-randomized controlled studies (9,782 patients in total), comparing the CADe-assisted group (n=4569) with the standard colonoscopy group (n=5213). The results showed that in retrospective studies, the use of CADe did not significantly improve the detection rate of colorectal tumors, nor did it increase the burden of colonoscopy operation. There was no statistically significant difference between the two groups in indicators such as Adenoma Detection Rate (ADR), Advanced Adenoma Detection Rate (AADR), Adenomas Per Colonoscopy (APC), examination time, and Non-Neoplastic Lesions Per Colonoscopy (>NNLPC) [55]. However, this conclusion may be too conservative. The analysis has limitations such as limited sample size and uneven technical levels. Meanwhile, the learning curve effect of CADe, its potential advantages in high-risk populations, and its contribution to improving diagnostic consistency and reducing human errors have not been fully evaluated. It should be emphasized that CADe should be viewed as a supplement to physicians’



expertise, and its performance is expected to continuously improve with data accumulation and algorithm optimization.

Although artificial intelligence has shown significant potential in endoscopic diagnosis of malignant digestive tract tumors, its true value in clinical practice still needs further validation. Current limitations are mainly reflected in: (1) Most studies are still retrospective analyses, and data from single centers limit the generalizability of models; (2) There is a lack of prospective large-scale study evidence, and the effectiveness and safety of AI diagnostic models need to be further validated through multi-center randomized controlled trials; (3) The lack of unified standards for endoscopic image acquisition and processing affects the accuracy and reproducibility of AI diagnosis. To fully leverage the role of AI in improving colorectal cancer screening and diagnosis while minimizing potential risks, future research should focus on: (1) Conducting more large-scale, long-term randomized controlled trials to comprehensively evaluate the effectiveness and safety of CADE in real clinical environments; (2) Establishing industry standards for endoscopic image acquisition and use, constructing a medical endoscopic image knowledge base to provide high-quality data support for AI algorithm development; (3) Formulating quality control standards and ethical regulatory frameworks for endoscopic AI products, clarifying the responsibility boundaries between endoscopists and AI systems, ensuring the reliability of AI-assisted diagnosis and trust between doctors and patients.

### AI-assisted pathological diagnosis

Pathological examination, as the gold standard for diagnosing malignant digestive tract tumors, has long played a crucial role in clinical practice. However, traditional pathological diagnosis is highly dependent on the subjective experience of pathologists, inevitably leading to problems such as low diagnostic efficiency, poor accuracy, and large inter-observer variations, which are not conducive to timely and accurate identification of malignant tumors. With the rapid development of artificial intelligence (AI) technology, its application in the field of pathological diagnosis has provided new possibilities for solving these problems. Through objective quantitative analysis of pathological images, AI technology can not only assist in pathological diagnosis, improving diagnostic efficiency and accuracy, but also significantly reduce doctors' workload.

The development of AI-assisted pathological diagnosis for digestive tract tumors has evolved from basic image analysis to vision-language fusion, and then to multimodal interaction. **Basic Image Analysis Stage:** Research mainly focused on classification, segmentation, and

grading tasks of pathological images. Iizuka et al. developed an AI diagnostic system for colorectal cancer pathological images that can automatically identify normal tissue, adenoma, and adenocarcinoma, providing a powerful tool for rapid screening of suspicious cases [56]. The automated feature global delivery connection network (FGDC-net) proposed by Shi P et al. achieved significant results in the nuclear segmentation task of H&E stained images, providing a basis for nuclear atypia analysis [57]. The AI classification and grading method for colorectal cancer developed by Awan R et al. uses deep convolutional neural networks and achieved 97% binary classification accuracy (normal tissue vs. cancer tissue) and 91% three-category classification accuracy (normal, low-grade neoplasia, high-grade neoplasia) in histological grading tasks, providing an important basis for precision treatment [58].

**Vision-Language Fusion Stage:** As research deepened, large-scale vision-language pre-trained foundation models gradually became a new direction for pathological AI. The emergence of the CONCH (CONtrastive learning from Captions for Histopathology) model marked the entry of pathological AI into a new stage of vision-language fusion (Fig. 4) [59]. This model obtained rich visual-language representations through contrastive learning pre-training on more than 1.17 million pairs of pathological image-text pairs, enabling better understanding and utilization of language information in pathology reports. CONCH performed excellently in various downstream tasks, including zero-shot classification (classifying unseen categories of pathological images without additional training), cross-modal retrieval (using text to search for related images or images to search for descriptive text), image segmentation, and image description. Especially in the diagnosis of rare diseases, CONCH combined with weakly supervised learning showed significant potential, providing new ideas for addressing diseases with scarce data.

**Multimodal Interaction Stage:** The latest research progress has introduced more advanced multimodal generative AI pathology assistants, such as PathChat (Fig. 5) [60]. These systems not only analyze pathological images but also understand and generate relevant natural language, answering pathological diagnostic questions conversationally, achieving truly interactive diagnostic assistance. PathChat was constructed by combining a specially trained pathological image visual encoder [61] with a pre-trained large language model (such as Llama 2), and fine-tuned on over 450,000 diverse image-text instructions (containing 999,202 rounds of Q&A). Evaluations showed that PathChat performed excellently compared to other multimodal models: in multiple-choice diagnostic tasks with pathological images, PathChat

# AI and Malignant Digestive Tract Tumors

## CONCH (CONtrastive learning from Captions for Histopathology)



### Data Sources:

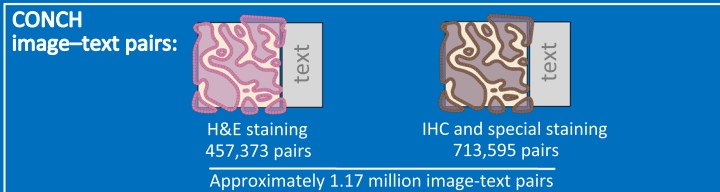
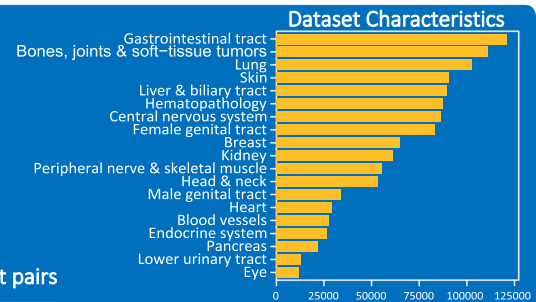
- Educational resources (EDU)
- PubMed Central Open Access Dataset (PMC OA)
- Other unlabeled sources

### Processing Steps:

- Object Detector: Trained to detect histopathology images
- Caption Splitter: Trained to split captions referring to multiple images
- Matcher: Matches detected images with their corresponding captions

### Dataset Scale:

- Unfiltered dataset: Approximately 1.79 million image-text pairs
- Human-only dataset (CONCH): excluding non-human related image-text pairs



Source: Nat Med. 2024 Mar;30(3):863-874. doi: 10.1038/s41591-024-02856-4. (Lu MY, Mahmood F.)

### Visual-language pretraining setup:

- Consists of an **image encoder**, a **text encoder**, and a **multimodal text decoder**
- Trained using contrastive and generative objectives
- The contrastive objectives align the outputs of the image and text encoders
- The generative objective maximizes the likelihood of generating the correct text

### Key performance:

- In zero-shot classification tasks, CONCH achieved over 90% accuracy on multiple datasets.
- It significantly outperformed baseline models in cross-modal retrieval tasks.
- CONCH also demonstrated excellent performance in zero-shot segmentation tasks.
- In few-shot learning scenarios, it exhibited high label efficiency.

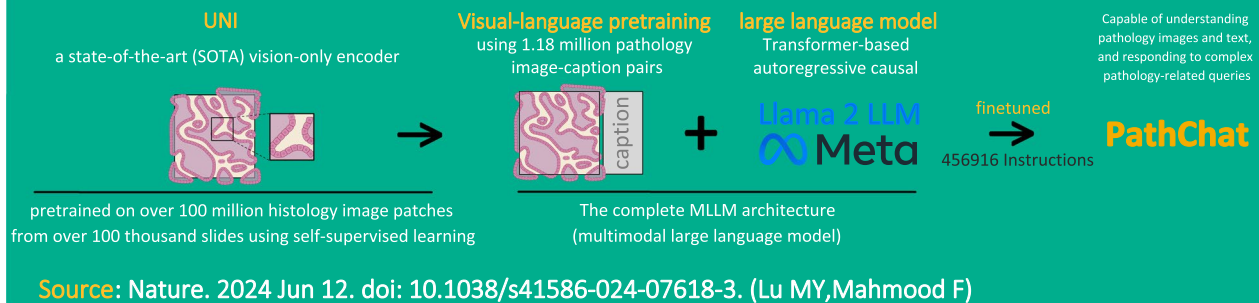
**Fig. 4** Overview of CONCH (CONtrastive learning from Captions for Histopathology). Illustration of the CONCH model architecture and dataset composition, comprising approximately 1.17 million image-text pairs, including 457,373 H&E staining pairs and 713,595 IHC and special staining pairs. The figure shows the data processing pipeline, including object detection, caption splitting, and image-text matching, along with key performance metrics in zero-shot classification and cross-modal retrieval tasks

achieved an accuracy of 78.1%, significantly higher than the comparative models LLaVA 1.5 (51.3%), LLaVA-Med (55.3%), and ChatGPT-4 (24.3%); when both images and clinical background information were provided, PathChat's accuracy increased to 89.5%, 39.0%, 60.9%, and 26.9% higher than the above models, respectively. In open-ended Q&A tasks, PathChat's answers were more favored by pathologists, with an accuracy of 78.7%, about 48% higher than LLaVA1.5 and LLaVA-Med, and 26.4% higher than ChatGPT-4. Particularly in "microscopic

examination" and "diagnostic" type questions that require careful examination of histological images, PathChat's performance was outstanding, with its answers considered more accurate and helpful by experts. This type of multimodal interactive AI system provides strong technical support for the "AI+pathologist" collaborative diagnostic model, allowing pathologists to engage in multiple rounds of dialogue with AI, clarify doubts, obtain more information, and ultimately make more accurate diagnoses. PathChat demonstrated high flexibility, effectively

# AI and Malignant Digestive Tract Tumors

**PathChat**, a vision-language generalist AI assistant for human pathology



## Multiple-choice diagnostic questions:

- Image-only input:  
78.1% accuracy (52.4% higher than LLaVA 1.5, 63.8% higher than LLaVA-Med, 53.8% higher than ChatGPT-4)
- Image with clinical context input:  
89.5% accuracy (39.0% higher than LLaVA 1.5, 60.9% higher than LLaVA-Med, 26.9% higher than ChatGPT-4)

## Open-ended question answering:

- More preferred by pathologists  
(48.9% higher than LLaVA 1.5, 48.1% higher than LLaVA-Med, 26.4% higher than ChatGPT-4)
- 78.7% accuracy on the subset with pathologist consensus evaluation

## Flexibility:

- Effectively combines visual features and clinical background information to provide more accurate diagnoses

## Interactive capability:

- Supports multi-turn dialogue, able to adjust diagnostic approach based on additional information

**Fig. 5** Architecture and performance of PathChat, a multimodal AI pathology assistant. Detailed representation of PathChat's architecture, combining a UNI visual-language pretrained model with Llama 2 LLM, fine-tuned on 456,916 instructions. The figure highlights PathChat's superior performance in multiple-choice diagnostic questions (78.1–89.5% accuracy) and open-ended question answering (78.7% accuracy), significantly outperforming other models like LLaVA 1.5, LLaVA-Med, and ChatGPT-4

combining visual features and clinical context information to provide diagnostic advice, and supporting judgment adjustments based on new information. In complex diagnostic processes (for example, when facing tumors of unknown primary origin requiring multiple rounds of immunohistochemical testing), this interactive, multi-round reasoning capability is particularly valuable.

**Fine-grained Feature Extraction and Interpretability:** Pathological AI commonly uses architectures based on fully convolutional networks (FCN), U-Net, or Swin

Transformer to extract slice-level or patch-level features from digital slides. These models can extract fine-grained features such as nuclear morphology, staining intensity, and tissue structure patterns in stages, and output specific diagnostic results in classification or segmentation heads. To enhance interpretability, some studies introduce visualization attention maps or class activation mapping (CAM) within the model, directly marking the areas of most concern to the model on pathological images to help pathologists understand the basis of AI

decisions. For example, in the diagnosis of submucosal invasion of gastric cancer, the model can mark the location of suspicious submucosal infiltration bands, indicating high-risk areas.

**Validation Metrics and Evidence-based Medical Evidence:** The evaluation of pathological AI typically includes multiple levels: in segmentation tasks, overlap is quantified using the Dice coefficient and Intersection over Union (IoU); in classification or grading tasks, diagnostic consistency is evaluated using metrics such as accuracy, F1 score, and AUC. There are differences in the setting of the "gold standard" across different studies, with some using the consensus of senior pathology experts as a reference, while others refer to molecular diagnosis or long-term prognosis. Most existing literature is retrospective in design, lacking large-scale prospective, multi-center clinical trials. To provide pathological AI with a more solid evidence-based medical foundation, embedded validation in real clinical workflows is needed, observing its impact on diagnostic efficiency, miss rates, and clinical decision-making accuracy.

Although AI-assisted pathological diagnosis has made significant technological progress, its clinical translation and application still face many challenges. The primary issue is data quality and standardization: the consistency of pathological slice quality and staining directly affects AI diagnostic accuracy, and differences in slice preparation and diagnostic standards between different hospitals and regions require the inclusion of multi-center, multi-source data when building AI models to improve model generalizability. Secondly, model interpretability still needs to be strengthened: clinical physicians need to understand the basis of AI diagnoses to increase trust and adoption. Next, knowledge update mechanisms: medical knowledge and diagnostic standards are constantly evolving, and AI models need to have the ability to continuously learn and update to adapt to emerging diagnostic standards and treatment methods. Additionally, clinical validation and quality control are also key: prospective studies are needed to validate the actual benefits of AI pathological diagnosis, and corresponding quality control and responsibility tracing mechanisms need to be established to ensure that AI systems are applied to clinical practice in a long-term, safe, and reliable manner.

Looking to the future, AI-assisted pathological diagnostic systems are expected to further integrate multimodal data, including whole-slide digital pathological images, multi-omics data such as genomics and transcriptomics, and patient clinical information, achieving more comprehensive and precise diagnoses. The deep integration of AI systems with clinical tools such as digital pathology slide scanners and electronic medical record systems will provide pathologists with a seamless intelligent assistance

experience. This integration will not only improve the accuracy and efficiency of diagnosis but also promote the development of personalized treatment plans, opening new prospects for the precise diagnosis and treatment of malignant digestive tract tumors.

## Problems and challenges in clinical applications

### Data quality and standardization

The introduction of artificial intelligence technology into the diagnosis and treatment of malignant digestive tract tumors requires high-quality data resources as a foundation. However, there is currently a lack of unified standards in the collection, storage, and annotation of tumor big data, which seriously affects data quality and constrains the development and application of intelligent models.

First, data collection lacks standardization: information systems across different medical institutions vary greatly, with inconsistent data recording formats and terminology. Non-structured data such as imaging and pathology lack unified collection specifications, and data quality obtained under different equipment and parameters varies significantly. There is also a lack of mechanisms for collecting long-term follow-up data on patients, resulting in short time spans and incomplete information. Second, data annotation lacks standardization: medical data annotation requires professional knowledge and experience, and the varying levels of different annotators can easily lead to bias and errors. Especially for non-structured data, there is a lack of unified annotation guidelines and quality control processes, resulting in strong subjectivity in annotation and difficulty in ensuring consistency. The manual annotation of massive medical data is labor-intensive and inefficient, becoming a bottleneck for AI development. Additionally, the integration of multi-omics data faces challenges such as complex non-linear interactions, data imbalance, batch effects, and dimensionality disaster [62, 63]. The lack of effective standardization methods for omics data seriously affects the development of AI models based on multi-omics big data. Duan R et al. evaluated the accuracy, robustness, and computational efficiency of ten integration methods in cancer molecular subtype classification based on multi-omics data from nine cancers in the TCGA database, and discussed the impact of different omics data types and their combinations on classification results [64]. Furthermore, the lack of data sharing platforms between medical institutions makes cross-institutional data aggregation difficult, with small sample sizes from single centers limiting large-scale validation of AI models. The so-called "data silos" phenomenon seriously hinders the research progress and translational application of AI in the medical field. At the same time, medical data



involves patient privacy, and the lack of clear data privacy protection policies and secure sharing mechanisms also limits data circulation.

To address these issues, there is an urgent need to establish a unified system of standards for the collection, storage, and access of medical big data, standardizing the acquisition processes and quality control measures for medical records, imaging, pathology, omics, and other data. Natural language processing and other technologies should be integrated to establish efficient professional annotation platforms and collaborative teams, improving the efficiency and quality of data annotation. Intelligent omics data preprocessing and analysis workflows should be developed to achieve standardized integration and deep mining of omics data. National medical big data sharing platforms should be built, promoting cross-institutional data sharing and cooperation under the premise of ensuring patient privacy and data security, providing large-scale, high-quality data support for AI model development.

### **Interpretability and reliability of algorithm models**

The application of AI in the diagnosis and treatment of malignant digestive tract tumors requires not only that models have high-precision predictive capabilities but also that clinicians can understand and trust their decision-making basis [65, 66]. However, current AI algorithms still face many challenges in terms of interpretability and reliability. First, the model "black box" problem is prominent [67]. Mainstream deep learning models have complex structures and massive parameters, making their decision-making processes difficult to explain in an intuitive way. This lack of transparency makes it difficult for physicians to understand the diagnoses and predictions given by AI, thereby affecting trust and acceptance of AI systems [68–71]. Second, the generalization ability of models needs improvement. Most AI models are trained on specific datasets with insufficient sample representativeness and diversity, limiting the applicability of models to real-world data from different populations and different hospitals. How to ensure that models can work reliably across different populations, regions, and time periods is an urgent problem to be solved. In addition, insufficient model robustness is also a challenge. Medical data often contains noise, annotation errors, and missing values, and current AI models are relatively sensitive to these data problems, prone to overfitting, and thus performing unstably in practical applications. The lack of model update and iteration mechanisms is also a significant problem. Changes in disease spectra, new diagnostic guidelines, and new drugs

all require AI models to continuously learn and update. However, there is currently a lack of effective mechanisms for continuous learning and updating of models, making it difficult for AI to adapt to the rapid iteration of medical knowledge; when attempting continuous learning, phenomena such as "catastrophic forgetting" and "knowledge hallucination" may be encountered [72, 73].

To improve the interpretability and reliability of AI models, a multi-pronged approach is needed. First, strengthen research on explainable AI technologies and develop models that can provide explanations for decision-making bases. For example, use inherently interpretable models (such as decision trees [74] or linear models [75]), apply post-hoc explanation methods (such as LIME or SHAP [76, 77]), introduce attention mechanisms to highlight key areas of model focus [78], combine multimodal data to provide richer explanations, and distill knowledge from complex models into simpler ones through knowledge distillation. Second, establish a model quality management and evaluation system covering the entire process from data acquisition and preprocessing to model training, testing, and deployment, and assess models from multiple dimensions such as statistical performance, clinical effectiveness, and ethical impact to ensure that AI systems are safely and reliably applied to patients. Third, fully utilize multi-center data and advanced privacy-preserving learning technologies such as federated learning [79] and transfer learning (Appendix 8) [80] to achieve cross-center collaborative training without directly sharing raw data, increasing training data diversity and improving model generalization capabilities. At the same time, combine active learning and incremental learning strategies to establish mechanisms for continuous learning and updating of models, timely incorporation of new diagnostic and treatment knowledge, and maintaining the advanced nature of model predictions. Some algorithms that can effectively mitigate catastrophic forgetting (such as Elastic Weight Consolidation EWC [81] and experience replay [82, 83]) can be introduced to ensure that models do not forget old knowledge while continuously learning new knowledge.

Taking pathology AI assistants like PathChat as examples, the following new methods can be adopted to improve their interpretability: 1) Attention visualization techniques that directly highlight model focus areas on pathological images; 2) Concept extraction, mapping features learned by the model to histological concepts familiar to pathologists; 3) Counterfactual explanations, exploring which changes in input would alter model output [84, 85]; 4) Language-vision alignment techniques, ensuring one-to-one correspondence between model



visual features and professional medical terminology [86]; 5) Knowledge graph integration, making model reasoning processes traceable to existing medical knowledge systems [87, 88]; 6) Multimodal explanation generation, simultaneously utilizing images, text, and other clinical data to provide comprehensive diagnostic bases [89]; 7) Interactive explanation systems, allowing users to engage in multiple rounds of interaction with the model to explore decision-making bases [90, 91]; 8) Case-comparative learning, explaining model diagnoses by comparing current cases with historical ones [92]; 9) Uncertainty quantification, clearly indicating the confidence level of model diagnoses and explaining sources of uncertainty [93]; 10) Model distillation, transferring knowledge from complex models to simple ones to improve interpretability [94]; 11) Neural-symbolic fusion, combining the representational capabilities of neural networks with the interpretability of symbolic systems [95]. In summary, improving the interpretability and transparency of AI models is a complex challenge requiring multidisciplinary collaboration across algorithms, medicine, and ethics. The development of these new technologies will help enhance the interpretability of medical AI systems and strengthen the trust of physicians and patients in AI-assisted diagnosis.

### **Ethics, safety, and regulations**

In the process of promoting the clinical translation of AI technology, ethical, safety, and regulatory factors are equally crucial. The application of AI in healthcare may raise ethical issues such as patient privacy breaches, unfair healthcare due to algorithmic bias, and attribution of responsibility for AI decisions, which need to be fully considered in technology development and application. Clear patient privacy protection policies should be established, using techniques such as data de-identification and federated learning to reduce privacy risks, and ensuring that algorithmic decisions do not systematically discriminate against certain groups due to training data bias. In terms of safety, rigorous testing of AI systems' ability to handle extreme situations is necessary to prevent dangerous outputs under abnormal inputs, and contingency plans for AI system failures should be formulated. On the regulatory front, regulatory bodies in various countries are gradually developing assessment and approval guidelines for medical AI. For instance, the U.S. Food and Drug Administration (FDA) has already approved the marketing of multiple AI diagnostic software products, and China is also exploring tiered management and admission standards for medical AI products. Establishing clear regulatory frameworks and industry standards helps regulate the development of AI

medical products and promotes their safe and effective entry into clinical use.

### **Algorithm bias and unfair healthcare**

Although AI shows potential in assisting diagnosis and treatment decisions, performance differences of AI models across different populations and environments have been widely observed. This bias may stem from imbalanced training data, implicit human biases in the data, and selection bias in the model development process. If not corrected promptly, algorithmic bias will lead to unfair treatment of certain patient groups, potentially further exacerbating existing health inequality issues. Taking pathology and endoscopy as examples, different equipment, staining methods, and operational techniques all introduce domain shift, making it difficult for AI systems to adapt to samples outside the training set. To ensure the fairness of AI clinical applications, researchers recommend introducing debiasing algorithms and strict bias detection mechanisms throughout the model development process [96]. In the future, data quality control and standardization processes should be further improved, cross-institutional joint research strengthened, diverse datasets collected, and bias issues effectively mitigated through statistical correction methods and improved model interpretability, allowing AI technology to benefit all patient groups more fairly.

### **Insufficient clinical applicability validation**

Like traditional medical devices, AI models also need to undergo thorough and rigorous clinical validation before they can truly enter clinical pathways. At the current stage, many AI studies achieve good results on retrospective datasets, but once transferred to real clinical scenarios (such as intraoperative real-time detection, dynamic pathological workflows), their performance often fluctuates due to factors such as on-site environment, operating habits, and individual patient differences. Most studies also lack direct observation of patient endpoints (such as survival time, complication rates), remaining only at the level of diagnostic accuracy. To truly assess the impact of AI on patient prognosis and medical resource allocation, longitudinal data needs to be collected through prospective, multi-center clinical trials or real-world studies (RWS) to test its robustness and generalizability in different scenarios.

### **Clinical validation and evidence-based support**

The true value of AI systems ultimately needs to be demonstrated through clinical validation. Like traditional drugs and devices, AI algorithms should undergo thorough prospective validation before clinical application to assess their actual impact on clinical decisions and

patient outcomes. However, most AI research currently remains at the stage of comparing technical performance indicators, lacking assessment of key clinical endpoints such as patient survival and quality of life, especially with a serious deficiency of prospective randomized controlled trials (RCTs), making it difficult to determine the effectiveness of AI algorithms in real clinical environments. As of 2022, approximately 43% of medical AI devices approved by the FDA had not published clinical validation data, and less than 5% had been validated through RCTs [97, 98], a situation that has drawn attention from regulatory authorities. In the future, to promote widespread clinical acceptance of AI technology, embedded clinical trials should be actively designed and conducted, incorporating AI-assisted diagnosis and treatment into actual clinical pathways for comprehensive assessment of its real efficacy. Meanwhile, data science education for healthcare professionals should be strengthened to enhance their understanding of the role and limitations of AI; patient communication should be enhanced to clearly explain the advantages and potential risks of AI technology, increasing patient acceptance and trust. Additionally, medical journals and academic conferences should encourage the publication of negative research results to reduce publication bias, and regulatory authorities should consider making prospective clinical validation of key AI algorithms a necessary condition for approval or reimbursement. Only when the effectiveness and safety of AI are fully proven by high-quality clinical evidence can its widespread application in the medical field truly be realized.

### Research gaps

Reviewing current literature, the following systematic research gaps still exist in AI for diagnosis and treatment of malignant digestive tract tumors: First, there is a lack of large-scale clinical trials to validate model generalizability across real-world multi-center, multi-ethnic data, which is insufficient to support universally applicable conclusions for populations in different regions; Second, research on algorithmic interpretability mostly remains at the technical level, lacking comprehensive assessment of social factors such as actual application feedback from clinical physicians, doctor-patient communication, and responsibility determination; Third, research on multi-modal AI based on multi-omics (genomics, transcriptomics, proteomics, etc.) is still not systematic, making it difficult to comprehensively explore the relationship between molecular typing and treatment plan selection; Fourth, standardized evaluation systems remain incomplete, particularly lacking longitudinal follow-up on patients' long-term survival outcomes, quality of life, and cost-effectiveness analysis.

In summary, although AI shows broad prospects in the diagnosis and treatment of malignant digestive tract tumors, to achieve the leap from research to clinical practice, the above challenges must be recognized and addressed.

It is worth noting that, according to the latest research statistics, less than 2% of models in the medical AI field actually surpass the prototype stage and enter routine clinical use [99]. In other words, the vast majority of AI systems still remain at the prototype development and validation stage, with very few being able to be applied in real clinical environments. In response to this situation, we have conducted Technology Readiness Level (TRL) assessments for each of the AI applications discussed in Appendix 10, to quantify the translational maturity of each technology and clarify its current stage. This will help identify which technologies are approaching clinical readiness (e.g., endoscopic AI has reached the TRL 8–9 stage) and which are still in early research and development (e.g., most prognostic prediction models are around TRL 4–5), thus providing reference for future research and resource investment.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06428-z>.

Additional file 1.

### Acknowledgements

We express our gratitude to all individuals and organizations that contributed to the preparation and writing of this review article.

### Author contributions

Yinhu Gao: Conceptualization, literature review, writing—original draft. Peizhen Wen: Conceptualization, study design, writing—original draft. Yuan Liu: Literature search and analysis, writing—original draft. Yahuang Sun: Literature search and analysis, writing—original draft. Hui Qian: Literature analysis, content review. Xin Zhang: Figure preparation, content review. Huan Peng: Figure preparation, content review. Yanli Gao: Data analysis, editing and polishing. Cuiyu Li: Data analysis, editing and polishing. Zhangyuan Gu: Data analysis, editing and polishing. Huajin Zeng: Data analysis, editing and polishing. Zhijun Hong: Data analysis, editing and polishing. Weijun Wang: Project administration, supervision, final version approval. Ronglin Yan: Project administration, supervision, final version approval. Zunqi Hu: Project administration, supervision, final version approval. Hongbing Fu: Project conception and design, supervision, final version approval. All authors have read and approved the final version of this manuscript for publication.

### Funding

This work was supported by the Key Research and Development Program of Shaanxi Province (No. 2023-YBSF-053), and the National Natural Science Foundation of China (Nos. 82070620, 82400774).

### Data availability

This review did not generate any new original data. All data analyzed in this article are from published literature and can be accessed through the citations provided in the reference list. The analysis results and figures are based on the summary and synthesis of these published literature. For specific analysis processes or more detailed information, please contact the corresponding authors.

## Declarations

### Consent for publication

All authors consent to the publication of this manuscript. This manuscript does not contain data from any individual person; therefore, additional consent for publication is not required.

### Competing interests

All authors declare that they have no competing interests related to this work. The conduct, writing, and publication of this research were not influenced by any commercial or financial interests.

### Author details

<sup>1</sup>Department of Gastroenterology, Shaanxi Province Rehabilitation Hospital, Xi'an, Shaanxi, China. <sup>2</sup>Department of General Surgery, Changzheng Hospital, Navy Medical University, 415 Fengyang Road, Shanghai 200003, China. <sup>3</sup>Shanghai Lung Cancer Center, Shanghai Chest Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. <sup>4</sup>Division of Colorectal Surgery, Changzheng Hospital, Navy Medical University, 415 Fengyang Road, Shanghai 200003, China. <sup>5</sup>Department of Gastroenterology, Changzheng Hospital, Naval Medical University, 415 Fengyang Road, Shanghai 200003, China. <sup>6</sup>Department of Gastrointestinal Surgery, Changzheng Hospital, Navy Medical University, 415 Fengyang Road, Shanghai 200003, China. <sup>7</sup>Infection Control Office, Shaanxi Province Rehabilitation Hospital, Xi'an, Shaanxi, China. <sup>8</sup>Department of Radiology, The First Hospital of Nanchang, the Third Affiliated Hospital of Nanchang University, Nanchang 330008, Jiangxi, China. <sup>9</sup>Tongji University School of Medicine, Tongji University, Shanghai 200092, People's Republic of China.

Received: 20 November 2024 Accepted: 25 March 2025

Published online: 09 April 2025

## References

- Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209–49.
- Chen W, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115–32.
- Dekker E, et al. Colorectal cancer. *Lancet*. 2019;394(10207):1467–80.
- Garcia-Buitrago M, Montgomery EA. Current concepts in gastrointestinal pathology. *Pathology*. 2022;54(2):145–6.
- Wang H, et al. Immune-based combination therapy for esophageal cancer. *Front Immunol*. 2022;13:1020290.
- Kelly RJ. Emerging multimodality approaches to treat localized esophageal cancer. *J Natl Compr Canc Netw*. 2019;17(8):1009–14.
- Zhou C, Zhang J. Immunotherapy-based combination strategies for treatment of gastrointestinal cancers: current status and future prospects. *Front Med*. 2019;13(1):12–23.
- Hamet, P. and J. Tremblay, *Artificial intelligence in medicine*. Metabolism, 2017. **69s**: S36–s40.
- Wang H, et al. Scientific discovery in the age of artificial intelligence. *Nature*. 2023;620(7972):47–60.
- Cordeschi R. AI turns fifty: revisiting its origins. *Appl Artif Intell*. 2007;21(4–5):259–79.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84–90.
- Jiang Y, et al. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun (Lond)*. 2020;40(4):154–66.
- Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Brief Bioinform*. 2022;23:2.
- Wu S, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc*. 2020;27(3):457–70.
- Routhier E, Mozziconacci J. Genomics enters the deep learning era. *PeerJ*. 2022;10: e13613.
- Yadav SS, Jadhav SM. Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*. 2019;6(1):113.
- Karimi D, Salcudean SE. Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Trans Med Imaging*. 2020;39(2):499–513.
- Kshatri SS, Singh D. Convolutional neural network in medical image analysis: a review. *Arch Comput Methods Eng*. 2023;30(4):2793–810.
- Kumar A, et al. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J Biomed Health Inform*. 2017;21(1):31–40.
- Abut S, Okut H, Kallail KJ. Paradigm shift from Artificial Neural Networks (ANNs) to deep Convolutional Neural Networks (DCNNs) in the field of medical image processing. *Expert Syst Appl*. 2024;244: 122983.
- OpenAI, et al. *GPT-4 Technical Report*. 2023. [arXiv:2303.08774](https://arxiv.org/abs/2303.08774) <https://doi.org/10.48550/arXiv.2303.08774>.
- The Claude 3 Model Family: Opus, Sonnet, Haiku*.
- Touvron, H., et al. *Llama 2: Open Foundation and Fine-Tuned Chat Models*. 2023. [arXiv:2307.09288](https://arxiv.org/abs/2307.09288) <https://doi.org/10.48550/arXiv.2307.09288>.
- Grattafiori, A., et al. *The Llama 3 Herd of Models*. 2024. [arXiv:2407.21783](https://arxiv.org/abs/2407.21783) <https://doi.org/10.48550/arXiv.2407.21783>.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
- Bhinder B, et al. Artificial intelligence in cancer research and precision medicine. *Cancer Discov*. 2021;11(4):900–15.
- Shimizu H, Nakayama KI. Artificial intelligence in oncology. *Cancer Sci*. 2020;111(5):1452–60.
- Hossain E, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. *Comput Biol Med*. 2023;155: 106649.
- Levis M, et al. Leveraging unstructured electronic medical record notes to derive population-specific suicide risk models. *Psychiatry Res*. 2022;315: 114703.
- Zhu E, et al. A unified framework of medical information annotation and extraction for Chinese clinical text. *Artif Intell Med*. 2023;142: 102573.
- Gupta R, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers*. 2021;25(3):1315–60.
- You Y, et al. Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct Target Ther*. 2022;7(1):156.
- Jiménez-Luna J, et al. Artificial intelligence in drug discovery: recent advances and future perspectives. *Expert Opin Drug Discov*. 2021;16(9):949–59.
- Zhu H. Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annu Rev Pharmacol Toxicol*. 2020;60:573–89.
- Ji S, et al. Pharmaco-proteogenomic characterization of liver cancer organoids for precision oncology. *Sci Transl Med*. 2023;15(706):3358.
- Liu Z, et al. Gene interaction perturbation network deciphers a high-resolution taxonomy in colorectal cancer. *Elife*. 2022;11:1.
- Liu Z, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun*. 2022;13(1):816.
- Wang L, et al. Comprehensive machine-learning survival framework develops a consensus model in large-scale multicenter cohorts for pancreatic cancer. *Elife*. 2022;11:1.
- Liu Z, et al. Integrative analysis from multi-center studies identifies a consensus machine learning-derived lncRNA signature for stage II/III colorectal cancer. *EBioMedicine*. 2022;75: 103750.
- Xu H, et al. Artificial intelligence-driven consensus gene signatures for improving bladder cancer clinical outcomes identified by multi-center integration analysis. *Mol Oncol*. 2022;16(22):4023–42.
- Argelaguet R, et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 2018;14(6): e8124.
- Argelaguet R, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol*. 2020;21(1):111.
- Majumder S, et al. Noncontact wearable wireless ECG systems for long-term monitoring. *IEEE Rev Biomed Eng*. 2018;11:306–21.
- Wang N, et al. Energy-efficient intelligent ECG monitoring for wearable devices. *IEEE Trans Biomed Circuits Syst*. 2019;13(5):1112–21.
- Horie Y, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc*. 2019;89(1):25–32.

46. Tang D, et al. A novel deep learning system for diagnosing early esophageal squamous cell carcinoma: a multicenter diagnostic study. *Clin Transl Gastroenterol*. 2021;12(8): e00393.
47. Horiuchi Y, et al. Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging. *Dig Dis Sci*. 2020;65(5):1355–63.
48. Nagao S, et al. Highly accurate artificial intelligence systems to predict the invasion depth of gastric cancer: efficacy of conventional white-light imaging, nonmagnifying narrow-band imaging, and indigo-carmin dye contrast imaging. *Gastrointest Endosc*. 2020;92(4):866–873.e1.
49. Wallace MB, et al. Impact of artificial intelligence on miss rate of colorectal neoplasia. *Gastroenterology*. 2022;163(1):295–304.e5.
50. Misawa M, et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology*. 2018;154(8):2027–2029.e3.
51. Urban G, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*. 2018;155(4):1069–1078.e8.
52. Kudo SE, et al. Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms. *Clin Gastroenterol Hepatol*. 2020;18(8):1874–1881.e2.
53. Cherubini A, Dinh NN. A review of the technology, training, and assessment methods for the first real-time ai-enhanced medical device for endoscopy. *Bioengineering (Basel)*. 2023;10:4.
54. Spadaccini M, et al. Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis. *Lancet Gastroenterol Hepatol*. 2021;6(10):793–802.
55. Patel HK, et al. Lack of Effectiveness of computer aided detection for colorectal neoplasia: a systematic review and meta-analysis of nonrandomized studies. *Clin Gastroenterol Hepatol*. 2024;22(5):971–980.e15.
56. Iizuka O, et al. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci Rep*. 2020;10(1):1504.
57. Shi P, et al. Nuclei segmentation of HE stained histopathological images based on feature global delivery connection network. *PLoS ONE*. 2022;17(9): e0273682.
58. Awan R, et al. Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images. *Sci Rep*. 2017;7(1):16852.
59. Lu MY, et al. A visual-language foundation model for computational pathology. *Nat Med*. 2024;30(3):863–74.
60. Lu MY, et al. A multimodal generative AI copilot for human pathology. *Nature*. 2024;634(8033):466–73.
61. Chen RJ, et al. Towards a general-purpose foundation model for computational pathology. *Nat Med*. 2024;30(3):850–62.
62. Athieniti E, Spyrou GM. A guide to multi-omics data collection and integration for translational medicine. *Comput Struct Biotechnol J*. 2023;21:134–49.
63. Pinu FR, et al. Systems biology and multi-omics integration: viewpoints from the metabolomics research community. *Metabolites*. 2019;9:4.
64. Duan R, et al. Evaluation and comparison of multi-omics data integration methods for cancer subtyping. *PLoS Comput Biol*. 2021;17(8): e1009224.
65. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform*. 2021;113: 103655.
66. Hatherley J, Sparrow R, Howard M. The virtues of interpretable medical AI. *Camb Q Healthc Ethics*. 2024;33(3):323–32.
67. Kalmykov, V.L. and L.V. Kalmykov XXAI: Towards eXplicitly eXplainable Artificial Intelligence. 2024. [arXiv:2401.03093](https://doi.org/10.48550/arXiv.2401.03093) <https://doi.org/10.48550/arXiv.2401.03093>.
68. Durán, J.M. and K.R. Jongsma, *Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI*. *J Med Ethics*, 2021.
69. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206–15.
70. Sidak D, et al. Interpretable machine learning methods for predictions in systems biology from omics data. *Front Mol Biosci*. 2022;9: 926623.
71. Gimeno, M., K. Sada DR, and A. Rubio, *Precision oncology: a review to assess interpretability in several explainable methods*. *Brief Bioinform*, 2023. **24**: 4.
72. Hasselmo ME. Avoiding Catastrophic Forgetting. *Trends Cogn Sci*. 2017;21(6):407–8.
73. Kirkpatrick J, et al. Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci U S A*. 2017;114(13):3521–6.
74. Podgorelec V, et al. Decision trees: an overview and their use in medicine. *J Med Syst*. 2002;26(5):445–63.
75. Wallisch C, et al. Review of guidance papers on regression modeling in statistical series of medical journals. *PLoS ONE*. 2022;17(1): e0262918.
76. Raptis S, Ilioudis C, Theodorou K. From pixels to prognosis: unveiling radiomics models with SHAP and LIME for enhanced interpretability. *Biomed Phys Eng Express*. 2024;10:3.
77. Sathyan A, Weinberg AI, Cohen K. Interpretable AI for bio-medical applications. *Complex Eng Syst*. 2022;2:4.
78. Cao R, et al. CFANet: context feature fusion and attention mechanism based network for small target segmentation in medical images. *Sensors (Basel)*. 2023;23:21.
79. Ray NK, Puthal D, Ghai D. Federated Learning. *IEEE Consumer Electronics Magazine*. 2021;10(6):106–7.
80. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data*. 2016;3(1):9.
81. Ovsianovskiy, A., et al. *Elastic Weight Consolidation Improves the Robustness of Self-Supervised Learning Methods under Transfer*. 2022. [arXiv:2210.16365](https://doi.org/10.48550/arXiv.2210.16365) <https://doi.org/10.48550/arXiv.2210.16365>.
82. Li C, et al. SLER: Self-generated long-term experience replay for continual reinforcement learning. *Appl Intell*. 2021;51(1):185–201.
83. Lan, Q., et al. *Memory-efficient Reinforcement Learning with Value-based Knowledge Consolidation*. 2022. [arXiv:2205.10868](https://doi.org/10.48550/arXiv.2205.10868) <https://doi.org/10.48550/arXiv.2205.10868>.
84. Xu, A. and T. Wu *Generally-Occurring Model Change for Robust Counterfactual Explanations*. 2024. [arXiv:2407.11426](https://doi.org/10.48550/arXiv.2407.11426) <https://doi.org/10.48550/arXiv.2407.11426>.
85. Cottin A, et al. MS-CPFI: a model-agnostic counterfactual perturbation feature importance algorithm for interpreting black-box multi-state models. *Artif Intell Med*. 2024;147: 102741.
86. Liu, J., et al. *Enhancing Vision-Language Model with Unmasked Token Alignment*. 2024. [arXiv:2405.19009](https://doi.org/10.48550/arXiv.2405.19009) <https://doi.org/10.48550/arXiv.2405.19009>.
87. Varshney D, et al. Knowledge graph assisted end-to-end medical dialog generation. *Artif Intell Med*. 2023;139: 102535.
88. Lan Y, et al. Path-based knowledge reasoning with textual semantic information for medical knowledge graph completion. *BMC Med Inform Decis Mak*. 2021;21(9):335.
89. Sun, Z., et al., *A scoping review on multimodal deep learning in biomedical images and texts*. *ArXiv*, 2023.
90. Pfeuffer N, et al. Explanatory Interactive Machine Learning. *Bus Inf Syst Eng*. 2023;65(6):677–701.
91. Beauxis-Aussalet E, et al. The role of interactive visualization in fostering trust in AI. *IEEE Comput Graph Appl*. 2021;41(6):7–12.
92. Ying H, et al. CoREx: contrastive learning for representing terms via explanations with applications on constructing biomedical knowledge graphs. *J Am Med Inform Assoc*. 2024;31(9):1912–20.
93. Caldeira J, Nord B. Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms. *Mach Learn*. 2021;2(1): 015002.
94. Song Y, et al. Medical image classification: Knowledge transfer via residual U-Net and vision transformer-based teacher-student model with knowledge distillation. *J Vis Comun Image Represent*. 2024;102:10.
95. Kim S, et al. Integration of neural network-based symbolic regression in deep learning for scientific discovery. *IEEE Trans Neural Netw Learn Syst*. 2021;32(9):4166–77.
96. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: Implications for clinical decision-making. *PLoS Digit Health*. 2024;3(11): e0000651.
97. Chouffani-El-Fassi S, et al. Not all AI health tools with regulatory authorization are clinically validated. *Nat Med*. 2024;30(10):2718–20.
98. Chouffani-El-Fassi S, et al. Author Correction: Not all AI health tools with regulatory authorization are clinically validated. *Nat Med*. 2024;30(11):3381.
99. Schouten JS, et al. From bytes to bedside: a systematic review on the use and readiness of artificial intelligence in the neonatal and pediatric intensive care unit. *Intensive Care Med*. 2024;50(11):1767–77.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.