BioMedical Engineering OnLine

REVIEW

Open Access

Data-driven decision-making for precision diagnosis of digestive diseases

Song Jiang[1,2], Ting Wang[1,2] and Kun-He Zhang[1,2*]

*Correspondence:
khzhang@ncu.edu.cn

[1] Department of Gastroenterology, The First Affiliated Hospital of Nanchang University, No. 17, Yongwai Zheng Street, Nanchang 330006, China
[2] Jiangxi Institute of Gastroenterology and Hepatology, Nanchang 330006, China

Abstract

Modern omics technologies can generate massive amounts of biomedical data, providing unprecedented opportunities for individualized precision medicine. However, traditional statistical methods cannot effectively process and utilize such big data. To meet this new challenge, machine learning algorithms have been developed and applied rapidly in recent years, which are capable of reducing dimensionality, extracting features, organizing data and forming automatable data-driven clinical decision systems. Data-driven clinical decision-making have promising applications in precision medicine and has been studied in digestive diseases, including early diagnosis and screening, molecular typing, staging and stratification of digestive malignancies, as well as precise diagnosis of Crohn's disease, auxiliary diagnosis of imaging and endoscopy, differential diagnosis of cystic lesions, etiology discrimination of acute abdominal pain, stratification of upper gastrointestinal bleeding (UGIB), and real-time diagnosis of esophageal motility function, showing good application prospects. Herein, we reviewed the recent progress of data-driven clinical decision making in precision diagnosis of digestive diseases and discussed the limitations of data-driven decision making after a brief introduction of methods for data-driven decision making.

Keywords: Omics data, Data-driven decision, Precise diagnosis, Machine learning, Deep learning, Digestive diseases

Introduction

The concept of precision medicine has been introduced at the onset of the twenty-first century [1]. Precision medicine relies on data-driven decision-making that involves collecting massive amounts of data and organizing them to form information, and then integrating and refining the relevant information to form automated decision models via training and fitting [2]. Theoretically, with a sufficiently representative sample (data) and mathematical and statistical methods, it is possible for us to establish a model to produce prediction results that are very close to the true situation, which helps to predict the occurrence and progression of diseases and to assist in clinical diagnosis, personalized treatment and prognosis assessment [3, 4].

Human diseases involve complex and individualized pathophysiological dynamic changes, which generate big data of biology and medicine due to the increasing

application of clinical examination and high-throughput biotechnologies. Therefore, current data-driven decision-making is based on the analysis of large-scale heterogeneous data [5], which is a complex process, requiring constant data input, comparing the prediction results of models with real data, then feeding deviation information to the models, and self-improving in the continuous iterative process [6].

For simple data sets, traditional statistical methods may be suitable to build models for decision-making in disease diagnosis or prognosis prediction [7, 8]. However, traditional statistical methods were not sufficient to process the large-scale heterogeneous data. Therefore, data-driven decision-making was mainly implemented through machine learning (ML) algorithms. Due to its outstanding performance, ML has been used in an increasing number of studies to process big medical data [9, 10].

With the emergence and development of multiple omics technologies, data-driven decision-making has provided a mathematical basis for the analysis of omics data in precision medicine [11], including disease diagnosis [12], prognostic assessment [13], new drug development [14], remote patient monitoring [15], bioinformatics research [16], etc.

Herein, after a brief introduction of data-driven medical decision methods, we reviewed the progress of data-driven precision diagnosis based on omics data and clinical data in digestive disease. We searched the relevant literature in PubMed database for recent 5 years. Search terms are constructed from MeSH terms, including artificial intelligence, machine learning, digestive tract diseases, digestive tract tumors, and diagnosis. A total of 629 articles were retrieved and screened individually, the closely related articles were selected for intensive reading, and the representative articles were cited in this review.

Methods for data-driven decision-making

Data-driven decision making is achieved by ML algorithms. ML is a process in which computer learn from sample data without prior knowledge, including extracting features from the sample data, determining parameters, constructing a model and evaluating its performance, identifying and correcting deviation, and repeating the above process until the model performance cannot be improved [17]. The model can be used to predict the output values of independent external data sets [18].

Different data sets require different ML algorithms to process [7]. Traditional ML is mainly divided into unsupervised ML, supervised ML, and semi-supervised ML. Choosing an appropriate ML algorithm is critical to ensure the precision of data-driven decision-making.

Unsupervised ML is applicable for data sets without output values (labels), which can reveal hidden structures of data based on input features [19]. Main unsupervised ML methods include two types: dimensionality reduction (DR) and clustering. There are two common approaches for DR: Principal Component Analysis (PCA) [20] and t-Distributed Stochastic Neighbor Embedding (t-SNE) [21], while typical clustering algorithms include K-means clustering [22], hierarchical clustering [23], and spectral clustering [24].

Supervised ML is applicable for data sets with output values (labels), which trains a model with parameters identified during the training process to predict the output

values [25]. Main supervised learning algorithms include k-nearest neighbor algorithm (KNN) [26], generalized linear model (GLM) algorithms including ordinary least squares (OLS) [27], ridge regression [28], least absolute shrinkage and selection operator (LASSO) regression [29], and logistic regression (LR) [30], Naive Bayes [31], support vector machine (SVM) [32], and random forest (RF) [33].

Semi-supervised ML trains a model based on training data set with labels to predict an unlabeled data set, and labels the unlabeled data set according to the prediction value with the highest confidence (pseudo-labeling), then incorporates the unlabeled data set with pseudo-labeling into the training data set to retrain the model until the model's prediction results remain constant [34]. Common semi-supervised ML algorithms include Self-Training, Co-Training, Transductive SVM and so on [35].

Reinforcement learning (RL) is a subfield of ML focused on how agents can learn to make sequential decisions in an environment to maximize cumulative rewards [36]. Unlike traditional ML, RL involves an agent interacting with an environment, receiving feedback in the form of rewards or penalties based on its actions. RL has been widely used in medicine [37]. Classical RL algorithms include Q-learning, Policy gradients, deep Q-networks, Actor-Critic, and Monte Carlo [38].

Deep Learning (DL) algorithms, also known as deep neural networks, are a subfield of ML that focuses on training artificial neural networks (ANN) with multiple layers, which is a further development of traditional ML algorithms [39], which has been used to process enormous data sets and surpass many classical ML methods for processing natural language, documents, images data. Deep neural networks adjust internal parameters to minimize the loss function through iteration of the backpropagation process [40]. For backpropagation, a loss function is calculated based on the difference between model output and target output, and fed back through the system, which then adjust the parameters (or weights) in each layer of the neural network to minimize the error of each neuron and the error of the entire network. Repeating above process until the error between model output and target output is minimized to acceptable levels. The principal DL algorithms include convolutional neural networks (CNN) [41], recurrent neural networks (RNN) [42], generative adversarial networks (GANs) [43], and deep reinforcement learning (DRL) [44].

Data-driven precision diagnoses for digestive diseases

Data-driven decision-making has been widely applied in medical research. Figure 1 shows the schematic diagram of data-driven decision-making in the precision diagnosis of digestive diseases.

Data-driven precision diagnosis based on radiomics

Radiomics is a rapidly developing field of diagnostic research, which extracts quantitative metrics (features) of medical images, such as heterogeneity and shape, to inform precision diagnosis. These features can work alone or integrate with demographic, histological, genomics or proteomics data for clinical problem solving [45]. The National Cancer Institute's Quantitative Research Network has framed the radiomics in five components: (1) image acquisition and reconstruction; (2) image segmentation and

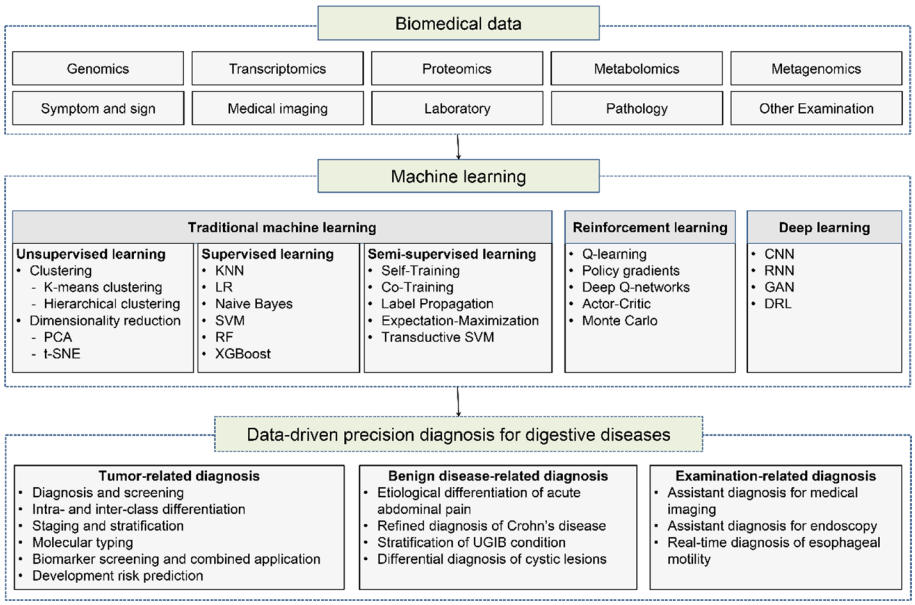

Fig. 1 Data-driven precision diagnosis for digestive diseases. *PCA* principal component analysis, *t-SNE* t-distributed stochastic neighbor embedding, *KNN* k-nearest neighbor algorithm, *LR* logistic regression, *SVM* support vector machine, *RF* random forest, *XGBoost* extreme gradient-boosting, *CNN* convolutional neural networks, *RNN* recurrent neural networks, *GANs* generative adversarial networks, *DRL* deep reinforcement learning, *UGIB* upper gastrointestinal bleeding

mapping; (3) feature extraction and quantification; (4) database building; and (5) analysis of individual data [46].

Radiomics has shown encouraging performance in the precision diagnosis of gastrointestinal tumors. Liu et al. [47] applied radiomics to predict c-kit gene mutations in gastrointestinal stromal tumor (GIST). They collected arterial phase, venous phase, delayed phase and tri-phase combined data from contrast-enhanced CT images of 106 GIST patients, selected features with LASSO regression and GLM and then constructed a classifier using multivariate LR; the classifier showed an accuracy of 0.808 in distinguishing GIST patients with or without mutations in exon 11 of c-kit gene. This study noninvasively analyzed specific gene mutations by radiomics to support precision medicine for GIST, but it was a retrospective study and further validation is needed.

In detection of hepatic metastases of colorectal cancer (CRC), a deep learning-based lesion detection algorithm (DLLD) for CT images showed a comparable sensitivity to abdominal radiologists (81.82% vs. 80.81%) [48]. Although the DLLD had higher false-positive rate than radiologists, it may serve as an adjunct to detect liver metastases. Ma et al. extracted and selected 485 radiomic features from portal venous CT images and constructed a LASSO–Logistic regression model, which can differentiate Borrmann type IV gastric cancer (GC) from primary gastric lymphoma (PGL) [49], with an accuracy of 81.43%.

Endoscopic images have been used for data-driven precision diagnosis of gastrointestinal diseases [50]. Yasar and colleagues developed a computerized decision support system (CDS) to assist in identifying the cancerous area of endoscopic images of biopsies [51]. They assessed the performance of image segmentation algorithms in CDS, such as region growing (RG), statistical region merging (SRM), statistical region merging with

region growing (SRMWRG), for detecting stomach cancerous areas, and found that RG produced the best performance, with sensitivity and specificity of 85.81% and 97.72%, respectively. CDS could help endoscopists identify cancerous areas that may have been missed and/or incompletely detected. However, data-driven precision diagnosis based on endoscopic images and videos lack of standardized imaging protocols and radiomics workflow.

Recent studies on data-driven precision diagnostics using radiomics data are summarized in Table 1.

Data-driven precision diagnostics based on genomics

With the rapid development of DNA sequencing technologies, especially whole exome sequencing (WES) and whole genome sequencing (WGS), the assessment of rare genetic mutations of complex diseases has become possible [85], facilitating the study on the pathogenesis of digestive diseases and disease diagnosis at the genetic level [86].

Genomics facilitate data-driven precise classification for GC subtype at the genetic level [87]. Based on The Cancer Genome Atlas (TCGA) database, TCGA Research Network proposed four molecular subtypes of gastric adenocarcinoma, namely, EBV-positive, microsatellite unstable, genomically stable, and chromosomally unstable tumors [88]. Ichikawa and colleagues performed a similar study, in which they identified at least one alteration in 435 cancer-related genes and 69 actionable genes of 207 patients by WES and classified GC into hypermutated and non-hypermutated tumors, and the latter was subdivided into six clusters by hierarchical clustering [89]. These molecular classifications pave the way for the molecular therapy of GC, but further studies with larger samples and multicenter clinical trials are needed.

CRC is a leading cause of cancer-related deaths globally [90], and early diagnosis plays a crucial role in improving the prognosis of patients [91]. Imperiale et al. detected multiple stool DNA targets (KRAS mutations, aberrant NDRG4 and BMP3 methylations) and used logistic-regression algorithm to build model for screening CRC [92], the combination of the stool DNA targets had a sensitivity of 92.3% for CRC and 42.6% for advanced adenomas, suggesting that multi-targeted fecal DNA screening may be an alternative test for patients who are intolerant to colonoscopy. However, the multitarget stool DNA test had more false positive results than fecal immunochemical test (FIT), and patients with positive multitarget stool DNA test require more endoscopy. Therefore, the improvement of the specificity of multitarget stool DNA test needs more attention.

Recent studies on data-driven precision diagnostics using genomics data are shown in Table 2.

Data-driven precision diagnostics based on transcriptomics

The transcriptome is the sum of all RNA transcripts of an organism, including coding RNA and non-coding RNA [101]. There were two critical technologies in this field: (1) microarrays [102] for quantifying a set of specific sequences and (2) RNA sequencing (RNA-Seq) [103], which analyzes RNA transcripts with high-throughput sequencing. Transcriptomics has been widely applied for biomedical research, such as disease diagnosis and staging [104].

Table 1 Data-driven precision diagnosis in digestive diseases based on radiomics

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Liu, 2022	GIST	106	Abdominal CT image; VOI segmentation, image normalization and feature extraction	GLM/LASSO	Accuracy: 80.8%	[47]
Kim, 2021	CRC	502	Abdominal CT image; ROI segmentation, feature extraction	CNN/Transfer Learning	Sensitivity: 81.82%	[48]
Ma, 2017	GC	40	Abdominal CT image; VOI segmentation, feature extraction	LASSO	Accuracy: 81.43%	[49]
Yasar, 2019	GC	10	Endoscopic image; image-based segmentation	Clustering	Accuracy: 96.33%	[51]
Li, 2021	Crohn disease	167	Abdominal CT enterography; VOI segmentation, feature extraction and selection	LASSO	AUC: 0.816 (95%CI, 0.706–0.926)	[52]
Yuan, 2022	CRC	140	Abdominal CT image; manual contouring, image-based ResNet-3D base neuron	ResNet3D/SVM	AUC: 0.922 (95%CI, 0.912–0.944)	[53]
Wu, 2022	Hepatic cystic echinococcosis	967	Abdominal ultrasound image; artificial marker repair and ROI extraction, image-based classification	DCNN	Accuracy: 90.6%	[54]
Kundu, 2020	Multi-disease detection	50	WCE image; image ROI separation, probability density function	LDA/Hierarchical SVM	Accuracy: 97.39%	[55]
Klang, 2020	Crohn disease	49	WCE image; image-based classification	CNN	Accuracy: 95.4–96.7%	[56]
Dmitriev, 2020	Pancreatic cystic lesions	134	Abdominal CT image; graph-based segmentation	RF/CNN	Accuracy: 91.7%	[57]
Meng, 2022	Crohn disease	235	Abdominal CT enterography; image ROI separation, patch-based classification	3D DCNN	AUC: 0.808–0.839	[58]
Wang, 2023	GHAC	216	Abdominal CT image; image ROI segmentation and radiomics feature extraction	LASSO	AUC: 0.731–0.942	[59]
Shi, 2023	PMME	122	Chest CT image; image resampling, tumor segmentation and feature extraction	LASSO	AUC: 0.906–0.975	[60]

Table 1 (continued)

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Zhou, 2023	Crohn disease	316	CT enterography; VAT features extraction	PCA/LASSO/3D-CNN	AUC: 0.775 (95%CI, 0.683–0.868)	[61]
Sun, 2019	GC	100	Abdominal CT image; ROI segmentation and radiomics feature extraction	LASSO	AUC: 0.903	[62]
Lonseko, 2023	GI lesion	4880	GI endoscopic image; gastrointestinal lesion segmentation	GANs/CNN	Precision: 91.72% ± 4.05%	[63]
Jia, 2023	GIST	151	Abdominal CT image/EUS image; image segmentation, image normalization, and feature extraction	LASSO	AUC: 0.766–0.866	[64]
Guo, 2022	CRC	360	Abdominal imaging examination data; ROI segmentation and feature extraction	CNN/K-means clustering	AUC: 0.950	[65]
Du, 2023	gastric neoplasms	3449	WL and WM endoscopy image and video; ROI segmentation and feature extraction	CNN	Accuracy: 90.0%	[66]
Tang, 2023	GI tract diseases	1645	GI endoscopic image; classification and segmentation	TransMT-Net	Accuracy: 96.9%	[67]
Gong, 2023	gastric neoplasms	8993	GI endoscopic image; semantic segmentation	U-Net + +/CNN	Accuracy: 95.6%	[68]
Yang, 2023	Intestinal Metaplasia Gastritis Atrophy	21,420	Gastric endoscopic image; localization, patch-based classification	LAG/DTL	Accuracy: 97.1–99.2%	[69]
Ding, 2023	GI lesion	2565	Capsule endoscopy image and video; image-based classification	CNN/CRNN	Accuracy: 79.2–97.5%	[70]
Muniz, 2023	CRC	71	Micro-FTIR absorbance HSI from biopsy tissue; localization, voxel-based classification	FCNN/linear SVM	Accuracy: 96–99%	[71]
Du, 2023	GC	1273	Gastroscopic image; segmentation, co-spatial attention and channel attention	CSA–CA–TB–ResUnet	Accuracy: 91.2%	[72]
Yuan, 2023	GC	4315	Tongue image; patch-based classification	KNN/SVM/DT/APINet/TransFG/DeepLabV3 +	AUC: 0.830–0.920	[73]

Table 1 (continued)

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Faust, 2023	Celiac Disease	96	Duodenitis biopsy image; CLAHE, feature extraction	SVM/KNN/DT	Accuracy: 98.5–98.6%	[74]
Kim, 2023	CRC	889	CRC histopathologic slide; patch extraction and normalization, patch-based classification	CNN	Accuracy: 95.5%	[75]
Abdelrahim, 2023	Barrett's neoplasia	270	Gastroscopy image and video; image-based classification	CNN	Accuracy: 92.0–94.7%	[76]
Fockens, 2023	Barrett's neoplasia	4920	WL endoscopy image; segmentation, image-based classification	Efficient-Net-Lite1/ MobileNetV2 DeepLabV3 +	Sensitivity: 84–100%	[77]
Zhang, 2023	gastrointestinal disorders	315,767	Gastroscopy image and video; localization, video-based classification	DCNN	Accuracy: 73.1–85.2%	[78]
Zhou, 2023	gastric polyps/ gastric ulcers/ gastric erosions	227	Gastroscopic image; feature extraction, feature fusion, image-based classification	GoogLeNet/ ResNet/ ResNeXt/SVM/ RF	Accuracy: 81.7–82.5%	[79]
Fan, 2023	UC	332	Endoscopic image and video; feature extraction, image-based classification	CNN	Accuracy: 86.54%	[80]
Faghani, 2022	Barrett's esophagus	542	Esophagus histology slide; image-based classification	CNN	Sensitivity: 90–100%	[81]
Yang, 2022	upper GI diseases	9403	GI endoscopic image; image-based classification	VGG-11/ ResNet50/ DenseNet121	Accuracy: 91.8%	[82]
Yuan, 2022	ESCC	685	GI endoscopic image; feature extraction, patch-based classification	DCNN	Accuracy: 89.8–91.3%	[83]
Luo, 2022	CAG	4005	GI WL image; image-based classification	CNN	Accuracy: 85.4–91.6%	[84]

Full names of abbreviations are given in the Abbreviations section of the manuscript

Patients with different stages of CRC differ in terms of therapy and prognosis. Xu et al. assessed the diagnostic capacity of tumor-educated platelet RNA profiles in differentiating CRC from healthy donors and noncancerous intestinal diseases using binary particle swarm optimization (PSO) coupled with SVM, and their classifier showed better performance than clinically utilized serum biomarkers, with areas under receiver operating characteristic curve (AUROC) ranging from 0.915 to 0.928 [105]. The tumor-educated

Table 2 Data-driven precision diagnosis in digestive diseases based on genomics

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Ichikawa, 2017	GC	207	Tumor tissue WGS data; actionable gene-based classification	Hierarchical clustering	–	[89]
Imperiale, 2014	CRC	9989	Multitarget stool DNA testing data; multimarker-based classification	LR	Sensitivity: 92.3% Specificity: 84.6%	[92]
Luo, 2020	CRC	1822	Circulating tumor DNA methylation markers; multimarker-based classification	LASSO/RF	AUC: 0.870	[93]
Romagnoni, 2019	Crohn disease	5277	Genome-wide genotyping data; genetic variant-based classification	Penalized LR/GBT/ANN	AUC: 0.802	[94]
Chung, 2023	CMMRD	639	Low-pass genomic instability characterization (LOGIC) assay; classification based on genomic microsatellite signature	LR	Sensitivity: 100%	[95]
Zuo, 2022	PEAC	86	Tumor tissue WES and targeted bisulfite sequencing data; DNA methylation-based classification	RF/LASSO/SVM/XGBoost	AUC: 0.900–1.000	[96]
Wan, 2019	CRC	817	WGS data of plasma cfDNA; classification based on genetic features	PCA/SVM/LR	AUC: 0.920 (95% CI, 0.910–0.930)	[97]
Cakmak, 2023	CRC	115	SNP profiles of immune phenotypes; prediction based on SNPs	LR/RF/SVM/KNN	AUC: 0.960	[98]
Guo, 2023	CRC	173	Tissue RNA-seq data; WGCNA, classification based on key hub genes	LASSO	AUC: 0.821–1.000	[99]
Killcoyne, 2020	EC	412	Shallow WGS data; classification based on genomic copy numbers	Elastic-net regression	Sensitivity: 72.0% Specificity: 82.0%	[100]

Full names of abbreviations are given in the Abbreviations section of the manuscript

platelet RNA profile analysis offered a potential noninvasive alternative to early CRC screening, but it was nonspecific, and related to the occurrence and development of multiple types of cancer. Zhao and coworkers identified four hub genes (BGN, COMP, COL5A2 and SPARC) based on transcriptomics and single cell sequencing, which highly expressed in GC and had potential value in diagnosis, therapy and prognosis [106]. this work, the transcriptomics data came from Gene Expression Omnibus (GEO) and TCGA

databases, and thus the efficacy and generalization ability of the established diagnostic model require further verification.

There are distinct expression patterns in the transcriptomics of various tumors, including hepatocellular carcinoma (HCC) [107]. Identification of biomarkers from tumor transcriptomics could contribute to data-driven tumor diagnosis. Using different techniques to select features from large-scale transcriptomics data, Kaur et al. identified three biomarkers (FCN3, CLEC1B and PRC1) with independent diagnostic value for HCC [108] and developed diagnostic models based on the three genes with various ML algorithms (Naive Bayes, KNN, RF and LR), with diagnostic accuracies ranging from 93 to 98% and AUROCs ranging from 0.97 to 1.0 for the training and validation data sets. This study provided an alternative method for the non-invasive diagnosis of HCC; however, the research data were also derived from GEO and TCGA databases, and further validation studies are needed for the diagnostic models.

Recent reports on data-driven precision diagnostics using transcriptomics data are shown in Table 3.

Data-driven precision diagnostics based on proteomics

In the context of precision medicine, disease therapy requires individualized strategies based on latent molecular signatures to overcome the challenges arising from heterogeneity. Biological specimens, such as blood, contain abundant proteins that provide reliable information about physiological and pathological state of body [116]. Proteomics, focuses on the large-scale analysis of proteins within biological system, has promising applications in the diagnosis and personalized management of gastrointestinal diseases [117].

Esophageal cancer (EC) is one of the highly invasive cancers and the leading cause of cancer-related deaths [118]. The lack of clinically relevant molecular subtypes for EC hinders development of effective therapeutic strategies. To explore the molecular subtypes of EC, Liu et al. performed proteomics and phosphorylated proteomics profiling in 124 pairs of EC tumors and paraneoplastic tissues based on mass spectrometry (MS) [119]. Using the PCA and hierarchical clustering, they classified the EC cohort into two molecular subtypes based on protein signatures: S1 and S2. Two typical protein signatures, ELOA and SCAF4, exhibited significantly higher expression levels in the subtype S1 than in the subtype S2, and the SVM classifier developed with these two protein features yielded an AUC of 0.976 in distinguishing these two subtypes. This study provided a basis for clarifying clinically relevant molecular subtypes of EC, which could help guide subtype-based clinical treatment. However, this is a monocenter study and a multicenter trial with a large sample is still needed to validate the results.

Proteomics analysis of clinical specimens facilitates identifying protein markers and establishing non-invasive diagnostic approaches. Komor et al. performed stool proteomics to identify biomarkers for the detection of high-risk adenoma and CRC [120]. In their study, colorectal adenoma tissue samples were characterized by low-coverage WGS to determine high-risk adenomas based on specific DNA copy number changes, a LASSO regression model was built with protein biomarkers identified from proteomics data

Table 3 Data-driven precision diagnosis in digestive diseases based on transcriptomics

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Xu, 2022	CRC	322	Transcriptomics data of patient platelets; classification based on DEGs	SVM/PSO	AUC: 0.915–0.928	[105]
Zhao, 2021	GC	6	Transcriptomics data sets of gastric tissues; classification based on hub DEGs	Ridge regression	AUC: 0.797–0.930	[106]
Kaur, 2020	HCC	3981	Large-scale transcriptomic profiling data sets of HCC; classification based on three DEGs	Naive Bayes/RF/LR	AUC: 0.970–1.000	[108]
Sallis, 2018	EoE	193	Transcriptomics data of esophageal biopsy tissues; classification based on mRNA transcript patterns	RF/PCA	AUC: 0.985	[109]
Samadi, 2022	CRC	3523	Transcriptomic data sets from GEO database; classification based on the integration of mRNA, miRNA and lncRNA	RF/SVM/LASSO/XGBoost/CNN/BPNN	AUC: 0.885–0.999	[110]
Maurya, 2021	CRC	695	TCGA mRNA data set of CRC tissues, classification based on DEGs	LASSO/RF/KNN/ANN	Accuracy: 100%	[111]
Long, 2019	CRC	311	RNA-seq data sets of CRC from TCGA and GTEx cohorts, classification based on DEGs	RF/KNN/Naive Bayes	Accuracy: 99.8%	[112]
Sallis, 2018	EoE	215	Transcriptomics data of esophageal biopsy tissues; classification based on mRNA patterns	PCA/RF	AUC: 0.990	[113]
Su, 2022	CRC	521	TCGA transcriptomic data of CRC tissues, classification based on DEGs	RF/SVM/LASSO/DT	Accuracy: 99.81%	[114]
Lu, 2022	UC	267	Transcriptomic data sets of UC from GEO database; classification based on DEGs	LR	AUC: 0.721–0.850	[115]

Full names of abbreviations are given in the Abbreviations section of the manuscript

to differentiating healthy controls from patients with high-risk adenoma and CRC, the model exhibited an AUC of 0.711. Their study provided a completely noninvasive and new method for detecting high-risk adenomas that develop into CRC, but its sensitivity was low and might lead to missed diagnoses.

Recent reports on data-driven precision diagnostics using proteomics data are shown in Table 4.

Table 4 Data-driven precision diagnosis in digestive diseases based on proteomics

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Liu, 2020	EC	248	MS-based proteomic and phosphoproteomic profiles of tumor and adjacent tissues; subtyping EC based on a protein signature	PCA/clustering/SVM	AUC: 0.976	[119]
Komor, 2021	Colorectal adenomas	281	Stool proteomics data; classification based on a panel of protein biomarkers	LASSO	AUC: 0.711	[120]
Bhardwaj, 2020	CRC	259	Quantitative data of 275 plasma proteins by PEA; classification based on selected protein features	LASSO	AUC: 0.920	[121]
Kalla, 2021	IBD	552	Quantitative data of 460 serum proteins by PEA; classification based on six proteins with age and sex	LR	Accuracy: 79.8%	[122]
Demirhan, 2023	GC	64	N-glycomics data of tumor and adjacent tissues; classification by differentially expressed N-glycans	MLP	AUC: 0.980	[123]
Fan, 2022	GC	255	Urine proteomics data; classification by 4 differentially expressed urine proteins	OPLS–DA	AUC: 0.810–0.920	[124]
Bergemalm, 2021	UC	451	Quantitative data of 92 plasma proteins by PEA; preclinical prediction by a panel of up-regulated proteins	PCA/LR	AUC: 0.920	[125]
Zhao, 2020	Acute appendicitis	568	Urinary proteomics data; classification based on a 10-protein signature	RF/SVM/Naive Bayes	Accuracy: 81.2–83.6%	[126]
Song, 2020	GC	60	Label-free global proteomics data of tumor and control tissues; classification based on a four-protein signature	RF	AUC: 0.886–0.996	[127]
Shen, 2019	GC	150	Targeted proteomics data of serum by PEA; classification based on 19 proteins	Elastic-net regression	AUC: 0.990	[128]
Chatziioannou, 2018	NEC	86	Serum proteomics profiles; classification based on two panels of three proteins	OPLS–DA	AUC: 0.999	[129]

Full names of abbreviations are given in the Abbreviations section of the manuscript

Data-driven precision diagnostics based on metabolomics

Metabolomics refers to comprehensive and simultaneous analysis of metabolites in biological samples and estimate their effective changes triggered by various conditions for instance, diet, lifestyle, genetic or environmental factors [130]. Due to inherent sensitivity of metabolomics, subtle changes of biological pathways can be detected, providing insight into the mechanisms hidden under various physiological conditions and abnormal processes [131].

Gastrointestinal system is the most central metabolic organ [132], and changes of intestinal bacterial content (intestinal microecological dysbiosis) and disruption of intestinal epithelial barrier can induce or exacerbate disease [133]. Jiménez and colleagues analyzed metabolite spectra of cancerous and para-carcinoma tissues from CRC patients using high-resolution magic angle spinning nuclear magnetic resonance (HR–MAS–NMR) and showed significant biochemical differences between two types of tissues [134], the metabolic profile of tumor tissues can distinguish tumors at different T and N stages, suggesting that it may have value in tumor staging. However, the sample size of the study was small, and further validation studies are warranted.

Lipid omics is a branch of metabolomics that targets lipid metabolites and has been used to identify biomarkers for tumor. Yuan et al. performed a lipidomic analysis in 525 serum samples and developed a diagnostic model containing 12 lipid biomarkers and age and gender by ML [135], which performed well for detecting esophageal squamous cell carcinoma (ESCC) with AUC of 0.958, 0.966 and 0.818 and sensitivities of 90.7%, 91.3% and 90.7% in the training, validation and independent validation cohorts, respectively. However, despite its good diagnostic efficiency, the model contains many variables and needs further optimization to improve its utility.

Metabolomics may play an important role in the differential diagnosis based on clinical symptoms. Takis et al. performed proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy of serum to extract individual metabolic fingerprints in two groups of patients who suffered from different acute abdominal pain (epigastric pain vs. diffuse abdominal pain) [136] and showed that metabolomics fingerprint could distinguish two groups of patients with high accuracy (> 90%); further analysis demonstrated that metabolomics fingerprint could distinguish the etiology of abdominal pain in the two groups with accuracies of > 70% and > 85%. These findings indicate that serum metabolomics may help emergency physicians to diagnose acute abdominal pain precisely. Non-targeted MRI-based metabolomics for the diagnosis of acute GI diseases has the advantages of being rapid, accurate and non-invasive, but its practical value needs to be further investigated.

Recent reports on data-driven precision diagnostics using metabolomics data are shown in Table 5.

Data-driven precision diagnostics based on metagenomics

Intestinal microbiome is a microbial ecosystem that expresses 100 times greater number of genes than human hosts and plays a critical role in human health and disease pathogenesis [144]. Next generation sequencing technologies, such as 16S rRNA, internal transcribed spacer (ITS) sequencing, metagenomics sequencing and viral sequencing [145], have been widely applied to the study of intestinal microbiome. Traditional techniques

Table 5 Data-driven precision diagnosis in digestive diseases based on metabolomics

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Jiménez, 2013	CRC	26	Metabolic profiles of tumor and adjacent tissues by NMR spectroscopy; classification based on discriminant metabolites	OPLS–DA	AUC: 0.910	[134]
Yuan, 2021	ESCC	525	Serum lipidomics data; classification based on a panel of 12 lipid biomarkers, age and gender	SVM/PCA	AUC: 0.818–0.966	[135]
Takis, 2018	Diffuse abdominal pain	64	Serum metabolomics data by NMR spectroscopy; classification by metabolomics fingerprint	OPLS–DA/PCA	Accuracy > 90%	[136]
Wang, 2023	ESCC	1104	Serum metabolomics data by LC–MS; classification based on digital images of metabolome profiles	CNN	AUC: 0.950	[137]
Yang, 2022	CRC	99	LC–MS-based plasma lipidomics data; classification based on 14 lipids	PLS/RF/SVM/KNN	Accuracy: 72.6–100%	[138]
Huang, 2021	GC	400	Untargeted metabolomics data of plasma; classification based on 6 metabolites with clinical indicators	LR/RF	AUC: 0.830	[139]
Yu, 2023	GC	301	Serum metabolomics data by MS; classification based on 12 differential metabolites	PCA/SVM/RF/LASSO	AUC: 0.893	[140]
Matsumoto, 2023	GC	101	Hydrophilic metabolites quantified by LC–TOFMS; classification based on 3 metabolites	SVM	AUC: 0.885–0.915	[141]
Pan, 2022	GC	280	Target bile acid metabolomics data of serum; classification based on 6 bile acids	RF/LASSO/OPLS–DA	AUC: 0.940–1.000	[142]
Zhao, 2022	ESCC	239	Multi-platform metabolomics data of serum; classification based on 5 metabolites	RF/LASSO/PCA	AUC: 0.873 (95% CI, 0.825–0.925)	[143]

Full names of abbreviations are given in the Abbreviations section of the manuscript

for metagenomics depend on prior knowledge [146, 147] and are unable to annotate sequences not available in database [148]. In recent years, innovative approaches based on traditional ML and DL algorithms have emerged to analyze metagenomics data [149].

For example, unsupervised or supervised learning models were widely applied for classification or clustering of samples based on annotation matrices [150, 151].

Metagenomics-based precision medicine has become a hot topic in gastrointestinal disease research. Nonalcoholic fatty liver disease (NAFLD) is an important etiology of chronic liver disease, which can lead to liver cirrhosis (LC), HCC and liver-related death [152]. Loomba et al. used gut microbial metagenomics to distinguish liver fibrosis levels in NAFLD patients [153]. They characterized the composition of gut microbiome by metagenomics sequencing of DNA extracted from stool samples and constructed a RF classifier containing 40 features that distinguished liver fibrosis between stages 0–2 and stages 3–4 with an AUC of 0.936. This study, which detects the level of liver fibrosis in NAND from the perspective of intestinal microbiome, is an interesting study that deserves further validation. Yang et al. performed a metagenomics analysis of the intestinal microbiome of 52 CRC patients and 55 healthy family members and found significant differences between the gut microbiomes of CRC patients and healthy family members and constructed an RF classifier with 22 microbial genes that could accurately distinguish CRC patients from healthy controls with an AUC of 0.905, 0.811, 0.859 in Chongqing, Hong Kong and French cohorts, respectively [154], which may be valuable for the early CRC diagnosis. However, it is not known whether this method can distinguish CRC from benign intestinal diseases.

Recent reports on data-driven precision diagnostics using metagenomics data are shown in Table 6.

Data-driven precision diagnosis based on clinical data

Daily clinical practice generates medical big data involving disease history, laboratory examinations, medical images, pathology, therapy, etc. ML algorithms can mine more information from medical big data to facilitate precision diagnosis. For example, the development of diagnostic models based on clinical big data sets can provide clinicians with data-driven decision-making advice, thereby facilitating the evolution from guideline-oriented medicine to individualized precision medicine.

Laboratory data are frequently used in data-driven diagnostics based on clinical data. Li et al. developed diagnostic models based on the data of traditional laboratory examinations to detect CRC [163]. They extracted laboratory data, including liver enzymes, lipids, complete blood counts and tumor biomarkers from electronic medical records of patients with CRC and healthy controls, and applied five ML algorithms (LR, RF, KNN, SVM and Naive Bayes) to develop diagnostic models for CRC, in which the LR model performed best for identifying CRC, with AUC 0.865, sensitivity 89.5%, specificity 83.5%, PPV 84.4%, and NPV 88.9%.

Combining multiple types of clinical data may be necessary for data-driven diagnosis in certain conditions. Hu et al. performed a precision diagnostic study in patients initially diagnosed as gastric GIST [164]. They collected multiple types of preoperative data of the patients, including hematological indicators, features of enhanced CT and ultrasonic gastroscopy, and then developed and validated a diagnostic model for differentiating GIST from other confusing tumors by extreme gradient-boosting (XGBoost) algorithm, with an accuracy of 73%.

Table 6 Data-driven precision diagnosis in digestive diseases based on metagenomics

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Loomba, 2017	NAFLD	86	Gut metagenomics data of stool; classification based on a fecal metagenomic signature	RF/SVM/clustering	AUC: 0.936	[153]
Yang, 2020	CRC	534	Fecal metagenomics data; classification based on fecal microbiomics biomarkers	Clustering/RF	AUC: 0.811–0.930	[154]
Bang, 2019	CRC	404	Gut microbiome data from 16 S rRNA sequencing; classification based on gut microbiome	SVM/KNN/LogitBoost	Accuracy: 96.84%	[155]
Dai, 2018	CRC	526	Gut metagenomics data; classification based on seven CRC-enriched bacterial markers	PCA/SVM	AUC: 0.820–0.84	[156]
Abbas, 2019	IBD	973	Gut metagenomics data of biopsy samples from QIITA database; classification based selected features by NBBD	RF	AUC: 0.760–0.800	[157]
Syama, 2023	CRC/IBD	1849	Gut metagenomics data sets of CRC and IBD; classification based on gut metagenomics data by boosting GraphSAGE	GCN	AUC: 0.900–0.930	[158]
Lee, 2022	IBD/CRC/LC	644	Gut metagenomics data sets; classification based on metagenome features	RF/SVM/PCR/LASSO/XGBoost/	AUC: 0.840–0.980	[159]
Forbes, 2018	UC	102	Gut metagenomics data; classification based on abundant taxonomic biomarkers of gut microbiota	Naive Bayes/RF/PCA	AUC: 0.900–0.930	[160]
Liang, 2020	CRC	1012	Fecal metagenomics data; classification based on combining several gut microbial gene markers with FIT	LR	Sensitivity: 93.8% Specificity: 81.2%	[161]
Hollister, 2019	IBS	45	Fecal metagenomics and metabolomics data; classification based on fecal metagenomic and metabolic markers	RF/LASSO/SVM/Naive Bayes	AUC: 0.930	[162]

Full names of abbreviations are given in the Abbreviations section of the manuscript

The use of routine clinical examination data to build valuable diagnostic models should be valued, as the data are derived from routine clinical work without additional testing.

Recent reports on data-driven precision diagnostics using clinical data are shown in Table 7.

Data-driven precision diagnostics based on integrated omics

Due to nonlinear interactions and joint effect of multiple factors generated from biological systems, it became difficult to discern true biological signal from random noise. Noise may come from biological systems, analytical platforms, and various data-specific analytical workflows, which complicates the integration of data across omics. Nevertheless, integrated omics and clinical data provide more comprehensive and valid information that facilitate precision medicine [174].

A combined multi-omics analysis can provide a better molecular classification of tumors. Liu et al. used clustering approach to analyze the data sets of gene copy number alterations (CNAs), DNA methylation, mRNA and miRNA and divided 256 HCC samples into five subgroups, each showing distinct survival rates and molecular signature [175].

Integrated analysis of multiple omics can provide better diagnostic performance. Al-Harazi et al. established and validated a new network-based approach to analyze CRC [176], in which they performed an integrated analysis of whole genome gene expression profile and CNAs data sets to construct a gene interaction network for each significantly altered gene, and then these gene interaction networks were clustered to form gene interaction subnetwork markers. Using these subnetwork markers, a SVM classifier based on 15 subnetwork markers were developed, which showed over 98% accuracy in detecting CRC patients, providing better value for disease diagnosis compared to individual gene markers.

Diagnostic methods based on multi-omics can reveal the heterogeneity of gastrointestinal tumors, which facilitates physician to more fully understand the genetic differences of individual patients and develop targeted therapies. However, they are cumbersome in steps, difficult to collect data and generalize.

Recent reports on data-driven precision diagnostics using multi-omics data are shown in Table 8.

Limitations and prospects of data-driven decision making

With the development of biological technology and computer science, the cost of acquiring omics data and time required to analyze and process them have been significantly reduced. The application of ML algorithms to study the intrinsic patterns and correlations of medical data for data-driven disease diagnosis and prediction has become a research hotspot. Many clinical trials of data-driven clinical decision-making systems based on ML and medical big data have been registered. For instance, Wallace MB et al. compared adenoma miss rate of colonoscopy with GI-Genius (Medtronic), which has been currently approved as a medical device in both the United States and the European Union, and found that AI reduced adenoma miss rate by about twofold [187]. Another randomized controlled trial to develop and validate the Gastrointestinal

Table 7 Data-driven precision diagnosis in digestive diseases based on clinical data

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Li, 2021	CRC	1164	Laboratory test data from electronic medical records; classification based on four laboratory indicators	LR/RF/KNN/SVM/Naïve Bayes	AUC: 0.849 (95%CI, 0.840–0.860)	[163]
Hu, 2021	GIST	124	Clinical examination data of pre-operation; classification based on CT and EUS features	XGBoost	AUC: 0.770 (95%CI, 0.570–0.900)	[164]
Shung, 2020	UGIB	2357	Clinical and laboratory indicators; classification based on variables of demography, comorbidity, clinical feature and laboratory	XGBoost	AUC: 0.90 (95%CI, 0.87 — 0.93)	[165]
Wang, 2021	Esophageal motility function	229	Esophageal HRM data sets; predicting esophageal motility function over HRM features	Conv3D/BiConvLSTM	Accuracy: 91.32%	[166]
Zhu, 2020	GC	709	Demographic and laboratory indicators from electronic medical records; classification based on a panel of independent predictors	GBDT	Accuracy: 83.0%	[167]
Phan-Mai, 2023	Complicated Appendicitis	1950	Medical record data; classification based on indicators of demography, blood test, and ultrasound of the appendix	SVM/DT/LR/KNN/ANN/GB	AUC: 0.64–0.89	[168]
Nemlander, 2023	CRC	2681	PHC data; classification based on diseases diagnosed in PHC consultations and consultation number	SGB/LR	AUC: 0.830 (95%CI, 0.790–0.870)	[169]
Popa, 2022	EMD	157	Esophageal HRM images; classification based on the images	CNN	Accuracy: 93.0%	[170]
Fan, 2022	GC	574	Medical record data; classification based on age, sex and classical serum tumor markers	LR/RF	Accuracy: 86.8%	[171]

Table 7 (continued)

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Kou, 2022	EMD	1741	Esophageal HRM data set; classification based on raw multi-swallow data	CNN/ANN/XGBoost/Bayes	Accuracy: 88.0–93.0%	[172]
Ho, 2023	EC	819	Questionnaire data from the SPIT and RISQ data sets; classification based on 17 features selected from questionnaire response	LDA/GLMNET/SVM/RF/KNN/CART/GLM	AUC: 0.710–0.920	[173]

Full names of abbreviations are given in the Abbreviations section of the manuscript

Artificial Intelligence Diagnostic System (GRAIDS) for the diagnosis of upper gastrointestinal cancers has been conducted in six hospitals of different tiers in China [188], and the results showed that GRAIDS had high diagnostic accuracy in detecting upper gastrointestinal cancer, with sensitivity similar to that of endoscopists, better than that of non-expert endoscopists.

However, there are still some issues that need to be addressed in the medical application of data-driven decision making. Although many reports on various models of data-driven decision making have been reported, few of them are applied in clinical practice. One of the reasons may be that the low quality of the data sets used to build ML models affects their practical application. Low-quality data sets can seriously impact the accuracy of data-driven decisions, the so-called garbage in, garbage out. Thereby, a prerequisite for effective data-driven decision making is to build high quality, well-constructed data sets. High-quality data sets can improve the predictive ability of ML algorithms and meanwhile reduce the size of data sets required for training models and the complexity of data representation. In addition, the ML models built for data-driven decision making need to be rigorously evaluated and optimized, which also requires new high-quality data sets to validate their application value and generalization performance.

The ML-based data-driven methodology still has some limitations. First, a critical drawback of DL algorithms is that it requires large amounts of data to train deep neural networks, and such scaled data sets are usually unachievable for many medical studies. Second, the interpretation of complex ML algorithms remains problematic. Third, considering the demand for large-scale data sets for data-driven, it is usually a challenge to integrate data sets across different platforms, languages and countries. Besides, the annotation of data sets from different sources differs, thus a uniform, standardized and publicly accepted data annotation system is required. An important point to remember is that classical ML algorithms require much less data than DL-based strategies; therefore, analyzing non-big data by appropriate classical ML algorithms can also be useful in precision medicine.

Although a data-driven diagnostic system can facilitate clinical decision-making, it can only provide physicians with complementary advice to assist them in noticing

Table 8 Data-driven precision diagnosis in digestive diseases based on integrated omics

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Liu, 2016	HCC	256	CNAs, DNA methylation, mRNA, and miRNA data from TCGA; subtyping HCC by multi-omics data	PCA/LR/Clustering	AUC: 0.780–1.000	[175]
Al-Harazi, 2021	CRC	89	Whole-genome gene expression profiling and CNA data sets from GEO database; classification based on the cores of 15 subnetwork markers	SVM/PCA/Clustering	Accuracy: 98.0%	[176]
Hoshino, 2022	CRC	24	Radiomics data of CT image and DNA sequencing data of tumor mutation burden; prediction of tumor mutation burden based on the image features	RF/XGBoost	Accuracy: 68.2%	[177]
Gawel, 2019	CRC	160	Public proteomics and transcriptomics data sets of tumor and adjacent tissues; classification based on nine secreted protein markers	Random elastic net	Sensitivity: 90.0% Specificity: 92.0%	[178]
Gai, 2023	CAG/GC	319	Fecal metabonomics and microbiota profiles data; classification based on 2 fecal metabolites and 2 gut microbes	SVM/RF	AUC: 0.88 Accuracy: 85.7%	[179]
Huang, 2022	CRC	743	Genomic and epigenetic profiles data sets of tissues from TCGA and GEO databases; classification based on DNA methylation and mutation burden data	LASSO/SVM/PCA/LR	AUC: 0.857–1.000	[180]
Cao, 2020	CRC	1214	Pathomics, genomic and transcriptomic data sets; classification base on pathomics signature	Residual CNN/XGBoost/Naive Bayes	AUC: 0.850–0.885	[181]
Gonzalez, 2022	Crohn disease	182	Fecal metaproteomics, metagenomics, metabolomics, and host genetics data; prediction of CD location based on a multi-omics feature set from metabolomics and metaproteomics	RF/LR/ExtraTrees/DT/Naive Bayes/KNN/SVC/MLPC/Voting Classifier/Adaboost	AUC: 0.94	[182]

Table 8 (continued)

First author, year	Disease	*n*	Data source and specific task	ML method	Diagnostic performance	Refs.
Adel-Patient, 2023	EoE	32	Tissue transcriptomics, tissue and blood immunologic components, and plasma metabolomics data sets; classification based on combining plasma metabolomics and cytokine biomarkers	PLS–DA/PCA	AUC: 0.929	[183]
Xing, 2023	CRC	212	Tissue transcriptomics and plasma metabolomics data; classification based on combining metabolomics and RNA-seq data	PLS–DA/PCA	AUC: 0.904–0.923	[184]
Kel, 2019	CRC	202	Full genome gene-expression data and genomic CpG island methylation data from tumor and gut epithelial tissues; classification based on 6 hypermethylated gene markers	F-Match/CMAcorrel/SVM/master-regulator search algorithm	Accuracy: 92.3%	[185]
Ding, 2019	CRC	315	Transcriptomics and proteomics data of CRC; classification based on secreted biomarkers	SVM	Accuracy: 85.9%	[186]

Full names of abbreviations are given in the Abbreviations section of the manuscript

problems they tend to overlook, not replace them in making diagnostic decisions. Excessive dependency of advice from a data-driven decision-making system is detrimental to the training of young physicians. Due to advances in science and technology, traditional physical examinations have been reduced and replaced by examinations performed by machines in modern medical practice, which led patients to doubt the competence of their physicians, and this distrust will increase if the patients are informed that the diagnosis comes from the computer.

Therefore, many aspects need to be improved before data-driven diagnostic systems become available for routine clinical application, including the establishment of high-quality data sets, standardization of data sets from different sources, selection of appropriate ML algorithms, improvement of relevant laws and regulations, and education for physicians and patients.

Conclusion

Mining the clinical value of medical data to build a data-driven medical decision-making system is a current research hotspot, which is important for large-scale medical data that are difficult for the human brain to process. In the data processing, there is no clear boundary between ML and traditional statistical approaches [189]. In general, traditional statistical models may perform better than ML algorithms for simple data sets, while for complex data sets and specific objectives, ML algorithms are required. Studies on data-driven medical decision making in digestive diseases have mainly focused on tumors, including detection and screening, molecular typing, staging, stratification, intra- and inter-class discrimination, as well as risk prediction. There are also reports on data-driven diagnosis and therapy for gastrointestinal non-tumor diseases, such as etiology differentiation of acute abdominal pain, precise diagnosis of Crohn's disease, stratification of UGIB, and real-time diagnosis of esophageal motility. Although data-driven clinical decision-making can contribute the precision of diagnosis of digestive diseases, there are still some limitations that need to be improved, including the establishment of high-quality data sets, standardization of data sets from different sources, selection of suitable ML algorithms, completion of relevant laws and regulations, relevant education for physicians and patients. However, it is believed that as relevant research continues to progress, data-driven clinical decision-making systems will be increasingly used in clinical practice and will become important assistants to clinicians and contribute to precision medicine.

Abbreviations

ANN	Artificial neural networks
APINet	Attentive pairwise interaction neural network
AUROC	Areas under receiver operating characteristic curve
BiConvLSTM	Bidirectional convolutional long–short-term-memory
BPNN	Back propagation neural network
CAG	Chronic atrophic gastritis
CART	Classification and regression tree
CDS	Computerized decision support system
CLAHE	Contrast Limited Adaptive Histogram Equalization
CMMRD	Constitutional mismatch repair deficiency
CNAs	Copy number alterations
CNN	Convolutional neural networks
Conv3D	Three-dimensional convolution
CRC	Colorectal cancer
CRNN	Convolutional recurrent neural network
CSA-CA–TB–ResUnet	Co-spatial attention and channel attention-based triple-branch ResUnet
3D DCNN	Three-dimensional deep convolutional neural networks
3D-CNN	Three-dimensional convolutional neural networks
DCNN	Deep convolutional neural networks
DEG	Differentially expressed gene
DL	Deep learning
DLLD	Deep learning-based lesion detection algorithm
DR	Dimensionality reduction
DRL	Deep reinforcement learning
DT	Decision trees
DTL	Dual transfer learning
EA	Ensemble algorithms
EC	Esophageal cancer
EMD	Esophageal motility disorders
EoE	Eosinophilic oesophagitis
ESCC	Esophageal squamous cell carcinoma
EUS	Endoscopic ultrasonography
FCNN	Fully connected neural network
FIT	Fecal immunochemical test
FTIR	Fourier transform infrared

GANs	Generative adversarial networks
GB	Gradient boosting
GBDT	Gradient boosting decision tree
GBT	Gradient boosted trees
GC	Gastric cancer
GCN	Graph convolutional network
GEO	Gene Expression Omnibus
GHAC	Gastric hepatoid adenocarcinoma
GI	Gastrointestinal
GIST	Gastrointestinal stromal tumor
GLM	Generalized linear model
GLMNET	Generalized linear model NETworks
GO	Gene Ontology
GRAIDS	Gastrointestinal Artificial Intelligence Diagnostic System
GraphSAGE	Graph SAmple and aggregate
GTEx	Genotype-Tissue Expression
HCC	Hepatocellular carcinoma
HSI	Hyperspectral imaging
^{1}H-NMR	Proton nuclear magnetic resonance
HRM	High-resolution manometry
HR-MAS-NMR	High-resolution magic angle spinning nuclear magnetic resonance
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
ITS	Internal transcribed spacer
KNN	K-nearest neighbor algorithm
LAG	Local attention grouping
LASSO	Least absolute shrinkage and selection operator
LC	Liver cirrhosis
LC–MS	Liquid chromatography with mass spectrometry
LC–TOFMS	Liquid chromatography–time-of-flight mass spectrometry
LDA	Linear discriminant analysis
LOGIC	Low-pass genomic instability characterization
LR	Logistic regression
ML	Machine learning
MLP	Multilayer perceptron algorithm
MLPC	Multilayer perceptron classifier
MS	Mass spectrometry
NAFLD	Nonalcoholic fatty liver disease
NBBD	Network-Based Biomarker Discovery
NMR	Nuclear magnetic resonance
OLS	Ordinary least squares
OPLS–DA	Orthogonal projection to latent structures–discriminant analysis
PCA	Principal component analysis
PCR	Principal component regression
PEA	Proximity extension assay
PEAC	Pulmonary enteric adenocarcinoma
PGL	Primary gastric lymphoma
PHC	Primary health care
PLS	Partial least squares
PLS–DA	Partial least square–discriminant analysis
PMME	Primary malignant melanoma of the esophagus
PSO	Particle swarm optimization
RF	Random forest
RG	Region growing
RISQ	Predicting risk of disease using detailed questionnaires
RL	Reinforcement learning
RNA-Seq	RNA sequencing
RNN	Recurrent neural networks
ROI	Region of interest
SGB	Stochastic gradient boosting
SNPs	Single nucleotide polymorphisms
SPIT	The saliva to predict risk of disease using transcriptomics and epigenetics
SRM	Statistical region merging
SRMWRG	Statistical region merging with region growing
SVC	Support Vector Classification
SVM	Support vector machine
TCGA	The Cancer Genome Atlas database
TransFG	Transformer architecture for fine-grained recognition
t-SNE	T-distributed stochastic neighbor Embedding
UC	Ulcerative colitis
UGIB	Upper gastrointestinal bleeding

VAT	Visceral adipose tissue
VOI	Volume of interest
WCE	Wireless capsule endoscopy
WES	Whole exome sequencing
WGCNA	Weighted gene co-expression network analysis
WGS	Whole genome sequencing
WL	White light
WM	Weak-magnifying
XGBoost	Extreme gradient-boosting

Author contributions
K-HZ, SJ and TW conceived and designed the paper; SJ drafted the manuscript; and K-HZ and TW revised the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (No. 82160494).

Availability of data and materials
Not applicable.

Declarations

Ethical approval and consent to participate
This article does not include any human participant studies conducted by any of the authors.

Human and animal ethics
This study did not include any human subjects or animals.

Consent for publication
This article contains no identifying information, so it is inapplicable.

Competing interests
There are no potential conflicts of interest reported by any of the authors.

Received: 14 January 2023 Accepted: 15 August 2023
Published online: 01 September 2023

References

1. Disease NRCUCoAFfDaNTo: Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. In Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington (DC): National Academies Press (US); 2011. [The National Academies Collection: Reports funded by National Institutes of Health].
2. Grossglauser M, Saner H. Data-driven healthcare: from patterns to actions. Eur J Prev Cardiol. 2014;21(2 Suppl):14–7.
3. Rutledge RB, Chekroud AM, Huys QJ. Machine learning and big data in psychiatry: toward clinical applications. Curr Opin Neurobiol. 2019;55:152–9.
4. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, Spreafico R, Hafler DA, McKinney EF. From big data to precision medicine. Front Med (Lausanne). 2019;6:34.
5. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. Cell. 2020;181(1):92–101.
6. Luo G. MLBCD: a machine learning tool for big clinical data. Health Inf Sci Syst. 2015;3:3.
7. Ma JL, Wang R, Zhang FK, Jia JD, Ou XJ, Zhang T, Wang Y, Duan WJ, Zhao XY, You H, Ma H. A noninvasive diagnostic model of liver fibrosis using serum markers in primary biliary cirrhosis. Zhonghua Nei Ke Za Zhi. 2012;51(8):618–22.
8. Sperger J, Shah KS, Lu M, Zhang X, Ungaro RC, Brenner EJ, Agrawal M, Colombel JF, Kappelman MD, Kosorok MR. Development and validation of multivariable prediction models for adverse COVID-19 outcomes in patients with IBD. BMJ Open. 2021;11(11): e049740.
9. Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920–30.
10. Bastanlar Y, Ozuysal M. Introduction to machine learning. Methods Mol Biol. 2014;1107:105–28.
11. Lavender CA, Shapiro AJ, Day FS, Fargo DC. ORSO (Online Resource for Social Omics): a data-driven social network connecting scientists to genomics datasets. PLoS Comput Biol. 2020;16(1): e1007571.
12. Ji J, Ling XB, Zhao Y, Hu Z, Zheng X, Xu Z, Wen Q, Kastenberg ZJ, Li P, Abdullah F, et al. A data-driven algorithm integrating clinical and laboratory features for the diagnosis and prognosis of necrotizing enterocolitis. PLoS ONE. 2014;9(2): e89860.
13. Karim MA, Samad A, Adhikari UK, Kader MA, Kabir MM, Islam MA, Hasan MN. A multi-omics analysis of bone morphogenetic protein 5 (BMP5) mRNA expression and clinical prognostic outcomes in different cancers using bioinformatics approaches. Biomedicines. 2020;8(2):19.
14. Mazandu GK, Chimusa ER, Rutherford K, Zekeng EG, Gebremariam ZZ, Onifade MY, Mulder NJ. Large-scale data-driven integrative framework for extracting essential targets and processes from disease-associated gene data sets. Brief Bioinform. 2018;19(6):1141–52.
15. Sapci AH, Sapci HA. Digital continuous healthcare and disruptive medical technologies: m-Health and telemedicine skills training for data-driven healthcare. J Telemed Telecare. 2019;25(10):623–35.

16. Schneider MV. Bioinformatics: scalability, capabilities and training in the data-driven era. Brief Bioinform. 2019;20(2):735–6.
17. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol. 2019;188(12):2222–39.
18. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, Siebourg-Polster J, Steiert B, Zhang JD. An introduction to machine learning. Clin Pharmacol Ther. 2020;107(4):871–85.
19. Jafari M, Wang Y, Amiryousefi A, Tang J. Unsupervised learning and multipartite network models: a promising approach for understanding traditional medicine. Front Pharmacol. 2020;11:1319.
20. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202.
21. Cheng Y, Wang X, Xia Y. Supervised t-distributed stochastic neighbor embedding for data visualization and classification. INFORMS J Comput. 2021;33(2):419–835.
22. Liu B, Zhang T, Li Y, Liu Z, Zhang Z. Kernel probabilistic K-means clustering. Sensors (Basel). 2021;21(5):1892.
23. Gollub J, Sherlock G. Clustering microarray data. Methods Enzymol. 2006;411:194–213.
24. Zhao Y, Li X. Spectral clustering with adaptive neighbors for deep learning. IEEE Trans Neural Netw Learn Syst. 2021.
25. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. Behav Ther. 2020;51(5):675–87.
26. Yu Z, Chen H, Liuxs J, You J, Leung H, Han G. Hybrid k-nearest neighbor classifier. IEEE Trans Cybern. 2016;46(6):1263–75.
27. Holodinsky JK, Yu AYX, Kapral MK, Austin PC. Comparing regression modeling strategies for predicting hometime. BMC Med Res Methodol. 2021;21(1):138.
28. Rokem A, Kay K. Fractional ridge regression: a fast, interpretable reparameterization of ridge regression. Gigascience 2020, 9(12).
29. Ji L, Chen S, Gu L, Zhang X. Exploration of potential roles of m6A regulators in colorectal cancer prognosis. Front Oncol. 2020;10:768.
30. Bian D, Liu X, Wang C, Jiang Y, Gu Y, Zhong J, Shi Y. Association between dietary inflammatory index and sarcopenia in Crohn's disease patients. Nutrients. 2022;14(4):901.
31. Zhang Z. Naive Bayes classification in R. Ann Transl Med. 2016;4(12):241.
32. Luckett DJ, Laber EB, El-Kamary SS, Fan C, Jhaveri R, Perou CM, Shebl FM, Kosorok MR. Receiver operating characteristic curves and confidence bands for support vector machines. Biometrics. 2021;77(4):1422–30.
33. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry. 2015;27(2):130–5.
34. Sheikh Hassani M, Green JR. A semi-supervised machine learning framework for microRNA classification. Hum Genomics. 2019;13(Suppl 1):43.
35. van Engelen JE, Hoos HH. A survey on semi-supervised learning. Mach Learn. 2020;109(2):373–440.
36. Matsuo Y, LeCun Y, Sahani M, Precup D, Silver D, Sugiyama M, Uchibe E, Morimoto J. Deep learning, reinforcement learning, and world models. Neural Netw. 2022;152:267–75.
37. Coronato A, Naeem M, De Pietro G, Paragliola G. Reinforcement learning for intelligent healthcare applications: a survey. Artif Intell Med. 2020;109: 101964.
38. Akalin N, Loutfi A. Reinforcement learning approaches in social robotics. Sensors (Basel). 2021;21(4):1292.
39. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. Adv Exp Med Biol. 2020;1213:3–21.
40. Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. Next-generation machine learning for biological networks. Cell. 2018;173(7):1581–92.
41. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24–9.
42. Lin S, Runger GC. GCRNN: group-constrained convolutional recurrent neural network. IEEE Trans Neural Netw Learn Syst. 2018;29(10):4709–18.
43. Phillips H, Soffer S, Klang E. Oncological applications of deep learning generative adversarial networks. JAMA Oncol. 2022;8(5):677–8.
44. Zhou SK, Le HN, Luu K, H VN, Ayache N. Deep reinforcement learning in medical imaging: a literature review.Med Image Anal. 2021; 73:102193.
45. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, Cook G. Introduction to radiomics. J Nucl Med. 2020;61(4):488–95.
46. Levy MA, Freymann JB, Kirby JS, Fedorov A, Fennessy FM, Eschrich SA, Berglund AE, Fenstermacher DA, Tan Y, Guo X, et al. Informatics methods to enable sharing of quantitative imaging research data. Magn Reson Imaging. 2012;30(9):1249–56.
47. Liu B, Liu H, Zhang L, Song Y, Yang S, Zheng Z, Zhao J, Hou F, Zhang J. Value of contrast-enhanced CT based radiomic machine learning algorithm in differentiating gastrointestinal stromal tumors with KIT exon 11 mutation: a two-center study. Diagn Interv Radiol. 2022;28(1):29–38.
48. Kim K, Kim S, Han K, Bae H, Shin J, Lim JS. Diagnostic performance of deep learning-based lesion detection algorithm in CT for detecting hepatic metastasis from colorectal cancer. Korean J Radiol. 2021;22(6):912–21.
49. Ma Z, Fang M, Huang Y, He L, Chen X, Liang C, Huang X, Cheng Z, Dong D, Liang C, et al. CT-based radiomics signature for differentiating Borrmann type IV gastric cancer from primary gastric lymphoma. Eur J Radiol. 2017;91:142–7.
50. Mori Y, Kudo SE, Berzin TM, Misawa M, Takeda K. Computer-aided diagnosis for colonoscopy. Endoscopy. 2017;49(8):813–9.
51. Yasar A, Saritas I, Korkmaz H. Computer-aided diagnosis system for detection of stomach cancer with image processing techniques. J Med Syst. 2019;43(4):99.

52. Li X, Liang D, Meng J, Zhou J, Chen Z, Huang S, Lu B, Qiu Y, Baker ME, Ye Z, et al. Development and validation of a novel computed-tomography enterography radiomic approach for characterization of intestinal fibrosis in Crohn's disease. Gastroenterology. 2021;160(7):2303-2316e2311.
53. Yuan Z, Xu T, Cai J, Zhao Y, Cao W, Fichera A, Liu X, Yao J, Wang H. Development and validation of an image-based deep learning algorithm for detection of synchronous peritoneal carcinomatosis in colorectal cancer. Ann Surg. 2022;275(4):e645–51.
54. Wu M, Yan C, Wang X, Liu Q, Liu Z, Song T. Automatic classification of hepatic cystic echinococcosis using ultrasound images and deep learning. J Ultrasound Med. 2022;41(1):163–74.
55. Kundu AK, Fattah SA, Wahid KA. Multiple linear discriminant models for extracting salient characteristic patterns in capsule endoscopy images for multi-disease detection. IEEE J Transl Eng Health Med. 2020;8:3300111.
56. Klang E, Barash Y, Margalit RY, Soffer S, Shimon O, Albshesh A, Ben-Horin S, Amitai MM, Eliakim R, Kopylov U. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. Gastrointest Endosc. 2020;91(3):606-613e602.
57. Dmitriev K, Marino J, Baker K, Kaufman AE. Visual analytics of a computer-aided diagnosis system for pancreatic lesions. IEEE Trans Vis Comput Graph. 2021;27(3):2174–85.
58. Meng J, Luo Z, Chen Z, Zhou J, Chen Z, Lu B, Zhang M, Wang Y, Yuan C, Shen X, et al. Intestinal fibrosis classification in patients with Crohn's disease using CT enterography-based deep learning: comparisons with radiomics and radiologists. Eur Radiol. 2022;32(12):8692–705.
59. Wang J, Kang B, Sun C, Du F, Lin J, Ding F, Dai Z, Zhang Y, Yang C, Shang L, et al. CT-based radiomics nomogram for differentiating gastric hepatoid adenocarcinoma from gastric adenocarcinoma: a multicentre study. Expert Rev Gastroenterol Hepatol. 2023;17(2):205–14.
60. Shi YJ, Zhu HT, Yan S, Li XT, Zhang XY, Sun YS. A CT-based radiomics nomogram model for differentiating primary malignant melanoma of the esophagus from esophageal squamous cell carcinoma. Biomed Res Int. 2023;2023:6057196.
61. Zhou Z, Xiong Z, Cheng R, Luo Q, Li Y, Xie Q, Xiao P, Hu D, Hu X, Shen Y, Li Z. Volumetric visceral fat machine learning phenotype on CT for differential diagnosis of inflammatory bowel disease. Eur Radiol. 2023;33(3):1862–72.
62. Sun ZQ, Hu SD, Li J, Wang T, Duan SF, Wang J. Radiomics study for differentiating gastric cancer from gastric stromal tumor based on contrast-enhanced CT images. J Xray Sci Technol. 2019;27(6):1021–31.
63. Lonseko ZM, Du W, Adjei PE, Luo C, Hu D, Gan T, Zhu L, Rao N. Semi-supervised segmentation framework for gastrointestinal lesion diagnosis in endoscopic images. J Pers Med. 2023;13(1):118.
64. Jia X, Wan L, Chen X, Ji W, Huang S, Qi Y, Cui J, Wei S, Cheng J, Chai F, et al. Risk stratification for 1- to 2-cm gastric gastrointestinal stromal tumors: visual assessment of CT and EUS high-risk features versus CT radiomics analysis. Eur Radiol. 2023;33(4):2768–78.
65. Guo J, Cao W, Nie B, Qin Q. Unsupervised learning composite network to reduce training cost of deep learning model for colorectal cancer diagnosis. IEEE J Transl Eng Health Med. 2023;11:54–9.
66. Du H, Dong Z, Wu L, Li Y, Liu J, Luo C, Zeng X, Deng Y, Cheng D, Diao W, et al. A deep-learning based system using multi-modal data for diagnosing gastric neoplasms in real-time (with video). Gastric Cancer. 2023;26(2):275–85.
67. Tang S, Yu X, Cheang CF, Liang Y, Zhao P, Yu HH, Choi IC. Transformer-based multi-task learning for classification and segmentation of gastrointestinal tract endoscopic images. Comput Biol Med. 2023;157: 106723.
68. Gong EJ, Bang CS, Lee JJ, Baik GH, Lim H, Jeong JH, Choi SW, Cho J, Kim DY, Lee KB, et al. Deep learning-based clinical decision support system for gastric neoplasms in real-time endoscopy: development and validation study. Endoscopy. 2023;55:701.
69. Yang J, Ou Y, Chen Z, Liao J, Sun W, Luo Y, Luo C. A benchmark dataset of endoscopic images and novel deep learning method to detect intestinal metaplasia and gastritis atrophy. IEEE J Biomed Health Inform. 2023;27(1):7–16.
70. Ding Z, Shi H, Zhang H, Zhang H, Tian S, Zhang K, Cai S, Ming F, Xie X, Liu J, Lin R. Artificial intelligence-based diagnosis of abnormalities in small-bowel capsule endoscopy. Endoscopy. 2023;55(1):44–51.
71. Muniz FB, Baffa MFO, Garcia SB, Bachmann L, Felipe JC. Histopathological diagnosis of colon cancer using micro-FTIR hyperspectral imaging and deep learning. Comput Methods Programs Biomed. 2023;231: 107388.
72. Du W, Rao N, Yong J, Adjei PE, Hu X, Wang X, Gan T, Zhu L, Zeng B, Liu M, Xu Y. Early gastric cancer segmentation in gastroscopic images using a co-spatial attention and channel attention based triple-branch ResUnet. Comput Methods Programs Biomed. 2023;231: 107397.
73. Yuan L, Yang L, Zhang S, Xu Z, Qin J, Shi Y, Yu P, Wang Y, Bao Z, Xia Y, et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. EClinicalMedicine. 2023;57: 101834.
74. Faust O, De Michele S, Koh JE, Jahmunah V, Lih OS, Kamath AP, Barua PD, Ciaccio EJ, Lewis SK, Green PH, et al. Automated analysis of small intestinal lamina propria to distinguish normal, celiac disease, and non-celiac duodenitis biopsy images. Comput Methods Programs Biomed. 2023;230: 107320.
75. Kim J, Tomita N, Suriawinata AA, Hassanpour S. Detection of colorectal adenocarcinoma and grading dysplasia on histopathologic slides using deep learning. Am J Pathol. 2023;193(3):332–40.
76. Abdelrahim M, Saiko M, Maeda N, Hossain E, Alkandari A, Subramaniam S, Parra-Blanco A, Sanchez-Yague A, Coron E, Repici A, Bhandari P. Development and validation of artificial neural networks model for detection of Barrett's neoplasia: a multicenter pragmatic nonrandomized trial (with video). Gastrointest Endosc. 2023;97(3):422–34.
77. Fockens KN, Jukema JB, Boers T, Jong MR, van der Putten JA, Pouw RE, Weusten B, Alvarez Herrero L, Houben M, Nagengast WB, et al. Towards a robust and compact deep learning system for primary detection of early Barrett's neoplasia: initial image-based results of training on a multi-center retrospectively collected data set. United Eur Gastroenterol J. 2023;11(4):324–36.
78. Zhang L, Lu Z, Yao L, Dong Z, Zhou W, He C, Luo R, Zhang M, Wang J, Li Y, et al. Effect of a deep learning-based automatic upper gastrointestinal endoscopic reporting system: a randomized crossover study. Gastrointest Endosc. 2023;98:181.

79. Zhou H, Liu Z, Li T, Chen Y, Huang W, Zhang Z. Classification of precancerous lesions based on fusion of multiple hierarchical features. Comput Methods Programs Biomed. 2023;229: 107301.
80. Fan Y, Mu R, Xu H, Xie C, Zhang Y, Liu L, Wang L, Shi H, Hu Y, Ren J, et al. Novel deep learning-based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. Gastrointest Endosc. 2023;97(2):335–46.
81. Faghani S, Codipilly DC, David V, Moassefi M, Rouzrokh P, Khosravi B, Agarwal S, Dhaliwal L, Katzka DA, Hagen C, et al. Development of a deep learning model for the histologic diagnosis of dysplasia in Barrett's esophagus. Gastrointest Endosc. 2022;96(6):918-925.e913.
82. Yang H, Wu Y, Yang B, Wu M, Zhou J, Liu Q, Lin Y, Li S, Li X, Zhang J, et al. Identification of upper GI diseases during screening gastroscopy using a deep convolutional neural network algorithm. Gastrointest Endosc. 2022;96(5):787-795.e786.
83. Yuan XL, Liu W, Liu Y, Zeng XH, Mou Y, Wu CC, Ye LS, Zhang YH, He L, Feng J, et al. Artificial intelligence for diagnosing microvessels of precancerous lesions and superficial esophageal squamous cell carcinomas: a multicenter study. Surg Endosc. 2022;36(11):8651–62.
84. Luo J, Cao S, Ding N, Liao X, Peng L, Xu C. A deep learning method to assist with chronic atrophic gastritis diagnosis using white light images. Dig Liver Dis. 2022;54(11):1513–9.
85. Wang YC, Wu Y, Choi J, Allington G, Zhao S, Khanfar M, Yang K, Fu PY, Wrubel M, Yu X, et al. Computational genomics in the era of precision medicine: applications to variant analysis and gene therapy. J Pers Med. 2022;12(2):175.
86. Biagioni A, Skalamera I, Peri S, Schiavone N, Cianchi F, Giommoni E, Magnelli L, Papucci L. Update on gastric cancer treatments and gene therapies. Cancer Metastasis Rev. 2019;38(3):537–48.
87. Chia NY, Tan P. Molecular classification of gastric cancer. Ann Oncol. 2016;27(5):763–9.
88. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202–9.
89. Ichikawa H, Nagahashi M, Shimada Y, Hanyu T, Ishikawa T, Kameyama H, Kobayashi T, Sakata J, Yabusaki H, Nakagawa S, et al. Actionable gene-based classification toward precision medicine in gastric cancer. Genome Med. 2017;9(1):93.
90. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. 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.
91. Wang W, Xie Y, Chen F, Liu X, Zhong LL, Wang HQ, Li QC. LncRNA MEG3 acts a biomarker and regulates cell functions by targeting ADAR1 in colorectal cancer. World J Gastroenterol. 2019;25(29):3972–84.
92. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med. 2014;370(14):1287–97.
93. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, Wang W, Sheng H, Pu H, Mo H, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med. 2020;12(524):eaax7533.
94. Romagnoni A, Jegou S, Van Steen K, Wainrib G, Hugot JP, International Inflammatory Bowel Disease Genetics C. Comparative performances of machine learning methods for classifying Crohn Disease patients using genome-wide genotyping data. Sci Rep. 2019;9(1):10351.
95. Chung J, Negm L, Bianchi V, Stengs L, Das A, Liu ZA, Sudhaman S, Aronson M, Brunga L, Edwards M, et al. Genomic microsatellite signatures identify germline mismatch repair deficiency and risk of cancer onset. J Clin Oncol. 2023;41(4):766–77.
96. Zuo Y, Zhong J, Bai H, Xu B, Wang Z, Li W, Chen Y, Jin S, Wang S, Wang X, et al. Genomic and epigenomic profiles distinguish pulmonary enteric adenocarcinoma from lung metastatic colorectal cancer. EBioMedicine. 2022;82: 104165.
97. Wan N, Weinberg D, Liu TY, Niehaus K, Ariazi EA, Delubac D, Kannan A, White B, Bailey M, Bertin M, et al. Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. BMC Cancer. 2019;19(1):832.
98. Cakmak A, Ayaz H, Arıkan S, Ibrahimzada AR, Demirkol Ş, Sönmez D, Hakan MT, Sürmen ST, Horozoğlu C, Doğan MB, et al. Predicting the predisposition to colorectal cancer based on SNP profiles of immune phenotypes using supervised learning models. Med Biol Eng Comput. 2023;61(1):243–58.
99. Guo C, Xie B, Liu Q. Weighted gene co-expression network analysis combined with machine learning validation to identify key hub biomarkers in colorectal cancer. Funct Integr Genomics. 2022;23(1):24.
100. Killcoyne S, Gregson E, Wedge DC, Woodcock DJ, Eldridge MD, de la Rue R, Miremadi A, Abbas S, Blasko A, Kosmidou C, et al. Genomic copy number predicts esophageal cancer years before transformation. Nat Med. 2020;26(11):1726–32.
101. Jiang Z, Zhou X, Li R, Michal JJ, Zhang S, Dodson MV, Zhang Z, Harland RM. Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. Cell Mol Life Sci. 2015;72(18):3425–39.
102. Romanov V, Davidoff SN, Miles AR, Grainger DW, Gale BK, Brooks BD. A critical comparison of protein microarray fabrication technologies. Analyst. 2014;139(6):1303–26.
103. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
104. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12(2):87–98.
105. Xu L, Li X, Li X, Wang X, Ma Q, She D, Lu X, Zhang J, Yang Q, Lei S, et al. RNA profiling of blood platelets noninvasively differentiates colorectal cancer from healthy donors and noncancerous intestinal diseases: a retrospective cohort study. Genome Med. 2022;14(1):26.
106. Zhao X, Wu S, Jing J. Identifying diagnostic and prognostic biomarkers and candidate therapeutic drugs of gastric cancer based on transcriptomics and single-cell sequencing. Pathol Oncol Res. 2021;27:1609955.
107. Liu XN, Cui DN, Li YF, Liu YH, Liu G, Liu L. Multiple "Omics" data-based biomarker screening for hepatocellular carcinoma diagnosis. World J Gastroenterol. 2019;25(30):4199–212.

108. Kaur H, Dhall A, Kumar R, Raghava GPS. Identification of platform-independent diagnostic biomarker panel for hepatocellular carcinoma using large-scale transcriptomics data. Front Genet. 2019;10:1306.
109. Sallis BF, Erkert L, Moñino-Romero S, Acar U, Wu R, Konnikova L, Lexmond WS, Hamilton MJ, Dunn WA, Szepfalusi Z, et al. An algorithm for the classification of mRNA patterns in eosinophilic esophagitis: integration of machine learning. J Allergy Clin Immunol. 2018;141(4):1354-1364.e1359.
110. Samadi P, Soleimani M, Nouri F, Rahbarizadeh F, Najafi R, Jalali A. An integrative transcriptome analysis reveals potential predictive, prognostic biomarkers and therapeutic targets in colorectal cancer. BMC Cancer. 2022;22(1):835.
111. Maurya NS, Kushwaha S, Chawade A, Mani A. Transcriptome profiling by combined machine learning and statistical R analysis identifies TMEM236 as a potential novel diagnostic biomarker for colorectal cancer. Sci Rep. 2021;11(1):14304.
112. Long NP, Park S, Anh NH, Nghi TD, Yoon SJ, Park JH, Lim J, Kwon SW. High-throughput omics and statistical learning integration for the discovery and validation of novel diagnostic signatures in colorectal cancer. Int J Mol Sci. 2019;20(2):296.
113. Sallis BF, Acar U, Hawthorne K, Babcock SJ, Kanagaratham C, Goldsmith JD, Rosen R, Vanderhoof JA, Nurko S, Fiebiger E. A distinct esophageal mRNA pattern identifies eosinophilic esophagitis patients with food impactions. Front Immunol. 2018;9:2059.
114. Su Y, Tian X, Gao R, Guo W, Chen C, Chen C, Jia D, Li H, Lv X. Colon cancer diagnosis and staging classification based on machine learning and bioinformatics analysis. Comput Biol Med. 2022;145: 105409.
115. Lu J, Wang Z, Maimaiti M, Hui W, Abudourexiti A, Gao F. Identification of diagnostic signatures in ulcerative colitis patients via bioinformatic analysis integrated with machine learning. Hum Cell. 2022;35(1):179–88.
116. He B, Huang Z, Huang C, Nice EC. Clinical applications of plasma proteomics and peptidomics: towards precision medicine. Proteomics Clin Appl. 2022;16:e2100097.
117. Islam Khan MZ, Tam SY, Law HKW. Advances in high throughput proteomics profiling in establishing potential biomarkers for gastrointestinal cancer. Cells. 2022;11(6):973.
118. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
119. Liu W, Xie L, He YH, Wu ZY, Liu LX, Bai XF, Deng DX, Xu XE, Liao LD, Lin W, et al. Large-scale and high-resolution mass spectrometry-based proteomics profiling defines molecular subtypes of esophageal cancer for therapeutic targeting. Nat Commun. 2021;12(1):4961.
120. Komor MA, Bosch LJ, Coupe VM, Rausch C, Pham TV, Piersma SR, Mongera S, Mulder CJ, Dekker E, Kuipers EJ, et al. Proteins in stool as biomarkers for non-invasive detection of colorectal adenomas with high risk of progression. J Pathol. 2020;250(3):288–98.
121. Bhardwaj M, Weigl K, Tikk K, Benner A, Schrotz-King P, Brenner H. Multiplex screening of 275 plasma protein biomarkers to identify a signature for early detection of colorectal cancer. Mol Oncol. 2020;14(1):8–21.
122. Kalla R, Adams AT, Bergemalm D, Vatn S, Kennedy NA, Ricanek P, Lindstrom J, Ocklind A, Hjelm F, Ventham NT, et al. Serum proteomic profiling at diagnosis predicts clinical course, and need for intensification of treatment in inflammatory bowel disease. J Crohns Colitis. 2021;15(5):699–708.
123. Demirhan DB, Yılmaz H, Erol H, Kayili HM, Salih B. Prediction of gastric cancer by machine learning integrated with mass spectrometry-based N-glycomics. Analyst. 2023;148(9):2073–80.
124. Fan H, Li X, Li ZW, Zheng NR, Cao LH, Liu ZC, Liu MW, Li K, Wu WH, Li ZX, et al. Urine proteomic signatures predicting the progression from premalignancy to malignant gastric cancer. EBioMedicine. 2022;86: 104340.
125. Bergemalm D, Andersson E, Hultdin J, Eriksson C, Rush ST, Kalla R, Adams AT, Keita ÅV, D'Amato M, Gomollon F, et al. Systemic inflammation in preclinical ulcerative colitis. Gastroenterology. 2021;161(5):1526-1539.e1529.
126. Zhao Y, Yang L, Sun C, Li Y, He Y, Zhang L, Shi T, Wang G, Men X, Sun W, et al. Discovery of urinary proteomic signature for differential diagnosis of acute appendicitis. Biomed Res Int. 2020;2020:3896263.
127. Song Y, Wang J, Sun J, Chen X, Shi J, Wu Z, Yu D, Zhang F, Wang Z. Screening of potential biomarkers for gastric cancer with diagnostic value using label-free global proteome analysis. Genomics Proteomics Bioinformatics. 2020;18(6):679–95.
128. Shen Q, Polom K, Williams C, de Oliveira FMS, Guergova-Kuras M, Lisacek F, Karlsson NG, Roviello F, Kamali-Moghaddam M. A targeted proteomics approach reveals a serum protein signature as diagnostic biomarker for resectable gastric cancer. EBioMedicine. 2019;44:322–33.
129. Chatziioannou AC, Wolters JC, Sarafidis K, Thomaidou A, Agakidis C, Govorukhina N, Kuivenhoven JA, Bischoff R, Theodoridis G. Targeted LC–MS/MS for the evaluation of proteomics biomarkers in the blood of neonates with necrotizing enterocolitis and late-onset sepsis. Anal Bioanal Chem. 2018;410(27):7163–75.
130. Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. Mass Spectrom Rev. 2019;38(3):221–38.
131. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17(7):451–9.
132. Miao YD, Mu LJ, Mi DH. Metabolism-associated genes in occurrence and development of gastrointestinal cancer: latest progress and future prospect. World J Gastrointest Oncol. 2021;13(8):758–71.
133. Wishart DS. Metabolomics for investigating physiological and pathophysiological processes. Physiol Rev. 2019;99(4):1819–75.
134. Jimenez B, Mirnezami R, Kinross J, Cloarec O, Keun HC, Holmes E, Goldin RD, Ziprin P, Darzi A, Nicholson JK. 1H HR-MAS NMR spectroscopy of tumor-induced local metabolic "field-effects" enables colorectal cancer staging and prognostication. J Proteome Res. 2013;12(2):959–68.
135. Yuan Y, Zhao Z, Xue L, Wang G, Song H, Pang R, Zhou J, Luo J, Song Y, Yin Y. Identification of diagnostic markers and lipid dysregulation in oesophageal squamous cell carcinoma through lipidomic analysis and machine learning. Br J Cancer. 2021;125(3):351–7.

136. Takis PG, Taddei A, Pini R, Grifoni S, Tarantini F, Bechi P, Luchinat C. Fingerprinting acute digestive diseases by untargeted NMR based metabolomics. Int J Mol Sci. 2018;19(11):3288.
137. Wang H, Yin Y, Zhu ZJ. Encoding LC–MS-based untargeted metabolomics data into images toward AI-based clinical diagnosis. Anal Chem. 2023;95(16):6533–41.
138. Yang C, Zhou S, Zhu J, Sheng H, Mao W, Fu Z, Chen Z. Plasma lipid-based machine learning models provides a potential diagnostic tool for colorectal cancer patients. Clin Chim Acta. 2022;536:191–9.
139. Huang S, Guo Y, Li ZW, Shui G, Tian H, Li BW, Kadeerhan G, Li ZX, Li X, Zhang Y, et al. Identification and validation of plasma metabolomic signatures in precancerous gastric lesions that progress to cancer. JAMA Netw Open. 2021;4(6): e2114186.
140. Yu J, Zhao J, Yang T, Feng R, Liu L. Metabolomics reveals novel serum metabolic signatures in gastric cancer by a mass spectrometry platform. J Proteome Res. 2023;22(3):706–17.
141. Matsumoto T, Fukuzawa M, Itoi T, Sugimoto M, Aizawa Y, Sunamura M, Kawai T, Nemoto D, Shinohara H, Muramatsu T, et al. Targeted metabolomic profiling of plasma samples in gastric cancer by liquid chromatography-mass spectrometry. Digestion. 2023;104(2):97–108.
142. Pan C, Deng D, Wei T, Wu Z, Zhang B, Yuan Q, Liang G, Liu Y, Yin P. Metabolomics study identified bile acids as potential biomarkers for gastric cancer: a case control study. Front Endocrinol (Lausanne). 2022;13:1039786.
143. Zhao J, Zhao X, Yu J, Gao S, Zhang M, Yang T, Liu L. A multi-platform metabolomics reveals possible biomarkers for the early-stage esophageal squamous cell carcinoma. Anal Chim Acta. 2022;1220: 340038.
144. Heintz-Buschart A, Wilmes P. Human gut microbiome: function matters. Trends Microbiol. 2018;26(7):563–74.
145. Gao B, Chi L, Zhu Y, Shi X, Tu P, Li B, Yin J, Gao N, Shen W, Schnabl B. An introduction to next generation sequencing bioinformatic analysis in gut microbiome studies. Biomolecules. 2021;11(4):530.
146. Escobar-Zepeda A, Vera-PoncedeLeon A, Sanchez-Flores A. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. Front Genet. 2015;6:348.
147. Wensel CR, Pluznick JL, Salzberg SL, Sears CL. Next-generation sequencing: insights to advance clinical investigations of the microbiome. J Clin Invest. 2022;132(7):1549544.
148. Mathieu A, Leclercq M, Sanabria M, Perin O, Droit A. Machine learning and deep learning applications in metagenomic taxonomy and functional annotation. Front Microbiol. 2022;13: 811495.
149. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol. 2022;23(1):40–55.
150. Calle ML. Statistical analysis of metagenomics data. Genomics Inform. 2019;17(1): e6.
151. Zhong H, Ren H, Lu Y, Fang C, Hou G, Yang Z, Chen B, Yang F, Zhao Y, Shi Z, et al. Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naive type 2 diabetics. EBioMedicine. 2019;47:373–83.
152. Sheka AC, Adeyi O, Thompson J, Hameed B, Crawford PA, Ikramuddin S. Nonalcoholic steatohepatitis: a review. JAMA. 2020;323(12):1175–83.
153. Loomba R, Seguritan V, Li W, Long T, Klitgord N, Bhatt A, Dulai PS, Caussy C, Bettencourt R, Highlander SK, et al. Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. Cell Metab. 2017;25(5):1054–10621055.
154. Yang J, Li D, Yang Z, Dai W, Feng X, Liu Y, Jiang Y, Li P, Li Y, Tang B, et al. Establishing high-accuracy biomarkers for colorectal cancer by comparing fecal microbiomes in patients with healthy families. Gut Microbes. 2020;11(4):918–29.
155. Bang S, Yoo D, Kim SJ, Jhang S, Cho S, Kim H. Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. Sci Rep. 2019;9(1):10189.
156. Dai Z, Coker OO, Nakatsu G, Wu WKK, Zhao L, Chen Z, Chan FKL, Kristiansen K, Sung JJY, Wong SH, Yu J. Multicohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. Microbiome. 2018;6(1):70.
157. Abbas M, Matta J, Le T, Bensmail H, Obafemi-Ajayi T, Honavar V, El-Manzalawy Y. Biomarker discovery in inflammatory bowel diseases using network-based feature selection. PLoS ONE. 2019;14(11): e0225382.
158. Syama K, Jothi JAA, Khanna N. Automatic disease prediction from human gut metagenomic data using boosting GraphSAGE. BMC Bioinformatics. 2023;24(1):126.
159. Lee SJ, Rho M. Multimodal deep learning applied to classify healthy and disease states of human microbiome. Sci Rep. 2022;12(1):824.
160. Forbes JD, Chen CY, Knox NC, Marrie RA, El-Gabalawy H, de Kievit T, Alfa M, Bernstein CN, Van Domselaar G. A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiosis exist? Microbiome. 2018;6(1):221.
161. Liang JQ, Li T, Nakatsu G, Chen YX, Yau TO, Chu E, Wong S, Szeto CH, Ng SC, Chan FKL, et al. A novel faecal Lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut. 2020;69(7):1248–57.
162. Hollister EB, Oezguen N, Chumpitazi BP, Luna RA, Weidler EM, Rubio-Gonzales M, Dahdouli M, Cope JL, Mistretta TA, Raza S, et al. Leveraging human microbiome features to diagnose and stratify children with irritable bowel syndrome. J Mol Diagn. 2019;21(3):449–61.
163. Li H, Lin J, Xiao Y, Zheng W, Zhao L, Yang X, Zhong M, Liu H. Colorectal cancer detected by machine learning models using conventional laboratory test data. Technol Cancer Res Treat. 2021;20:15330338211058352.
164. Hu B, Wang C, Jiang K, Shen Z, Yang X, Yin M, Liang B, Xie Q, Ye Y, Gao Z. Development and validation of a novel diagnostic model for initially clinical diagnosed gastrointestinal stromal tumors using an extreme gradient-boosting machine. BMC Gastroenterol. 2021;21(1):481.
165. Shung DL, Au B, Taylor RA, Tay JK, Laursen SB, Stanley AJ, Dalton HR, Ngu J, Schultz M, Laine L. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. Gastroenterology. 2020;158(1):160–7.
166. Wang Z, Hou M, Yan L, Dai Y, Yin Y, Liu X. Deep learning for tracing esophageal motility function over time. Comput Methods Programs Biomed. 2021;207: 106212.
167. Zhu SL, Dong J, Zhang C, Huang YB, Pan W. Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics. PLoS ONE. 2020;15(12): e0244869.

168. Phan-Mai TA, Thai TT, Mai TQ, Vu KA, Mai CC, Nguyen DA. Validity of machine learning in detecting complicated appendicitis in a resource-limited setting: findings from Vietnam. Biomed Res Int. 2023;2023:5013812.
169. Nemlander E, Ewing M, Abedi E, Hasselström J, Sjövall A, Carlsson AC, Rosenblad A. A machine learning tool for identifying non-metastatic colorectal cancer in primary care. Eur J Cancer. 2023;182:100–6.
170. Popa SL, Surdea-Blaga T, Dumitrascu DL, Chiarioni G, Savarino E, David L, Ismaiel A, Leucuta DC, Zsigmond I, Sebestyen G, et al. Automatic diagnosis of high-resolution esophageal manometry using artificial intelligence. J Gastrointestin Liver Dis. 2022;31(4):383–9.
171. Fan Z, Guo Y, Gu X, Huang R, Miao W. Development and validation of an artificial neural network model for non-invasive gastric cancer screening and diagnosis. Sci Rep. 2022;12(1):21795.
172. Kou W, Carlson DA, Baumann AJ, Donnan EN, Schauer JM, Etemadi M, Pandolfino JE. A multi-stage machine learning model for diagnosis of esophageal manometry. Artif Intell Med. 2022;124: 102233.
173. Ho KMA, Rosenfeld A, Hogan Á, McBain H, Duku M, Wolfson PB, Wilson A, Cheung SM, Hennelly L, Macabodbod L, et al. Development and validation of a multivariable risk factor questionnaire to detect oesophageal cancer in 2-week wait patients. Clin Res Hepatol Gastroenterol. 2023;47(3): 102087.
174. Pavel AB, Sonkin D, Reddy A. Integrative modeling of multi-omics data to identify cancer drivers and infer patient-specific gene activity. BMC Syst Biol. 2016;10:16.
175. Liu G, Dong C, Liu L. Integrated multiple "-omics" data reveal subtypes of hepatocellular carcinoma. PLoS ONE. 2016;11(11): e0165457.
176. Al-Harazi O, Kaya IH, El Allali A, Colak D. A network-based methodology to identify subnetwork markers for diagnosis and prognosis of colorectal cancer. Front Genet. 2021;12: 721949.
177. Hoshino I, Yokota H, Iwatate Y, Mori Y, Kuwayama N, Ishige F, Itami M, Uno T, Nakamura Y, Tatsumi Y, et al. Prediction of the differences in tumor mutation burden between primary and metastatic lesions by radiogenomics. Cancer Sci. 2022;113(1):229–39.
178. Gawel DR, Lee EJ, Li X, Lilja S, Matussek A, Schäfer S, Olsen RS, Stenmarker M, Zhang H, Benson M. An algorithm-based meta-analysis of genome- and proteome-wide data identifies a combination of potential plasma biomarkers for colorectal cancer. Sci Rep. 2019;9(1):15575.
179. Gai X, Qian P, Guo B, Zheng Y, Fu Z, Yang D, Zhu C, Cao Y, Niu J, Ling J, et al. Heptadecanoic acid and pentadecanoic acid crosstalk with fecal-derived gut microbiota are potential non-invasive biomarkers for chronic atrophic gastritis. Front Cell Infect Microbiol. 2022;12:1064737.
180. Huang H, Cao W, Long Z, Kuang L, Li X, Feng Y, Wu Y, Zhao Y, Chen Y, Sun P, et al. DNA methylation-based patterns for early diagnostic prediction and prognostic evaluation in colorectal cancer patients with high tumor mutation burden. Front Oncol. 2022;12:1030335.
181. Cao R, Yang F, Ma SC, Liu L, Zhao Y, Li Y, Wu DH, Wang T, Lu WJ, Cai WJ, et al. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. Theranostics. 2020;10(24):11080–91.
182. Gonzalez CG, Mills RH, Zhu Q, Sauceda C, Knight R, Dulai PS, Gonzalez DJ. Location-specific signatures of Crohn's disease at a multi-omics scale. Microbiome. 2022;10(1):133.
183. Adel-Patient K, Campeotto F, Grauso M, Guillon B, Moroldo M, Venot E, Dietrich C, Machavoine F, Castelli FA, Fenaille F, et al. Assessment of local and systemic signature of eosinophilic esophagitis (EoE) in children through multi-omics approaches. Front Immunol. 2023;14:1108895.
184. Xing F, Zheng R, Liu B, Huang K, Wang D, Su R, Feng S. A new strategy for searching determinants in colorectal cancer progression through whole-part relationship combined with multi-omics. Talanta. 2023;259: 124543.
185. Kel A, Boyarskikh U, Stegmaier P, Leskov LS, Sokolov AV, Yevshin I, Mandrik N, Stelmashenko D, Koschmann J, Kel-Margoulis O, et al. Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer. BMC Bioinformatics. 2019;20(Suppl 4):119.
186. Ding D, Han S, Zhang H, He Y, Li Y. Predictive biomarkers of colorectal cancer. Comput Biol Chem. 2019;83: 107106.
187. Wallace MB, Sharma P, Bhandari P, East J, Antonelli G, Lorenzetti R, Vieth M, Speranza I, Spadaccini M, Desai M, et al. Impact of artificial intelligence on miss rate of colorectal neoplasia. Gastroenterology. 2022;163(1):295-304.e295.
188. Luo H, Xu G, Li C, He L, Luo L, Wang Z, Jing B, Deng Y, Jin Y, Li Y, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol. 2019;20(12):1645–54.
189. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317–8.

Publisher's Note