Application of Artificial Intelligence to Gastroenterology and Hepatology

Catherine Le Berre, William J. Sandborn, Sabeur Aridhi, Marie-Dominique Devignes, Laure Fournier, Malika Smail-Tabbone, Silvio Danese, Laurent Peyrin-Biroulet

Application of Artificial Intelligence to Gastroenterology and Hepatology

Short title: Artificial intelligence in gastroenterology and hepatology

Catherine Le Berre[1,6], William J. Sandborn[2], Sabeur Aridhi[3], Marie-Dominique Devignes[3], Laure Fournier[4], Malika Smaïl-Tabbone[3], Silvio Danese[5], Laurent Peyrin-Biroulet[6]

[1]: Institut des Maladies de l'Appareil Digestif, Nantes University Hospital, France.

[2]: University of California San Diego, La Jolla, California, United States.

[3]: University of Lorraine, CNRS, Inria, LORIA (Laboratoire lorrain de recherche en informatique et ses applications), F-54000 Nancy, France.

[4]: Université Paris-Descartes, INSERM UMR-S970, Assistance Publique-Hôpitaux de Paris, Paris, France.

[5]: IBD Center and Department of Biomedical Sciences, Humanitas Clinical and Research Center, Humanitas University, Milan, Italy.

[6]: Inserm U954 and Department of Gastroenterology, Nancy University Hospital, University of Lorraine, France.

Grant support: No funding was provided for this study.

Abbreviations:

AI: Artificial intelligence.

ANN: Artificial neural network.

AUROC: Area under the receiver operating characteristic curve.

DL: Deep learning.

DNN: Deep neural network.

IBD: Inflammatory bowel disease.

ML: Machine learning.

TGN: Thioguanine nucleotide.

Corresponding author:

Prof. Laurent Peyrin-Biroulet

Inserm U954 and Department of Gastroenterology

Nancy University Hospital, University of Lorraine

1 Allée du Morvan

54 511 Vandœuvre-lès-Nancy, France

Tel +33 3 83 15 36 61

Fax +33 3 83 15 36 33

peyrinbiroulet@gmail.com

Disclosures: The authors declare that they have no competing interests relevant to this manuscript.

Writing assistance: No writing assistance was provided for this study.

Author contributions: CLB, LPB: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript. SA, MDD, MST: drafting and critical revision of the manuscript. WJS, SD, LF: critical revision of the manuscript for important intellectual content. All the authors read and approved the final version of the manuscript.

Abstract

Since 2010, substantial progress has been made in artificial intelligence (AI) and its application to medicine. AI is explored in gastroenterology for endoscopic analysis of lesions, in detection of cancer, and to facilitate the analysis of inflammatory lesions or gastrointestinal bleeding during wireless capsule endoscopy. AI is also tested to assess liver fibrosis and to differentiate patients with pancreatic cancer from those with pancreatitis. AI might also be used to establish prognoses of patients or predict their response to treatments, based on multiple factors. We review the ways in which AI may help physicians make a diagnosis or establish a prognosis and discuss its limitations, knowing that further randomized controlled studies will be required before the approval of AI techniques by the health authorities.

Keywords: Deep learning; machine learning; neural network; digestive system

There is no single definition of artificial intelligence (AI), but the concept involves computer programs that perform functions that we associate with human intelligence, such as learning and problem solving.[1,2] AI, machine learning (ML), and deep learning (DL) are overlapping disciplines (see **Figure 1**). ML is a vast domain that involves computer science and statistics, in which a machine performs repeated iterations of models progressively improving performance of a specific task. It produces algorithms to analyze data and to learn descriptive and predictive models. Data are mostly in the form of tables with objects or individuals as rows and variables, either numerical or categorical, as columns. ML is roughly divided into supervised and unsupervised methods. Unsupervised learning occurs when the purpose is to identify groups within data according to commonalities, with no a priori knowledge of the number of groups or their significance. Supervised learning occurs when training data contain individuals represented as input–output pairs. Input comprises individual descriptors whereas output comprises outcomes of interest to be predicted—either a class for classification tasks or a numerical value for regression tasks. The supervised ML algorithm then learns predictive models that subsequently allow to map new inputs to outputs.[3]

Artificial neural networks (ANN) are supervised ML models inspired by the neuroanatomy of brain. Each neuron is a computing unit and all neurons are connected to each other to build a network. Signals travel from the first (input), to the last (output) layer, possibly after going through multiple hidden layers (see **Figure 2**)**.** Training an ANN consists of dividing the data into a training set that helps to define the architecture of the network and to find out the various weights between the nodes and then a test set to assess the capability of the ANN to predict the desired output. During training, weights of interneuron connections are adjusted to optimize classification. The competition for more performance has led to a progressive complexity of neural network architectures, resulting in the concept of DL.[4]

Deep neural network (DNN) models are characterized by the application of several consecutive filters which allow the automatic detection of relevant features of input data. For this reason, DNN are considered as capable of learning data representation while including this learning in the global learning of the classification task. A variety of DNN architectures are included in DL-based methods.[5] However, the good performance obtained requires a huge amount of labeled training data. Researchers have addressed this issue by combining DL with reinforcement learning principles.[6]

The limits to these techniques are overfitting and lack of explainability. The models obtained by DL often perform much better than any other at fitting the data, however they are intrinsically dependent on the training dataset. If the training population does not include enough diversity, or contains an unidentified bias, results may not be generalizable to real-life populations leading to problems in model validation. Moreover, DNN, like ANN, provide black-box models lacking explainability. Recent studies are oriented towards improving explainability of DNN models as it is a pre-requisite for their acceptability in many fields, particularly in the biomedical applications.[7,8] There have been reviews on the use of AI in gastroenterology, but mainly focused on AI assisted-endoscopy.[9–11] We provide an overview of important studies assessing the value of AI in helping physicians make a diagnosis or establish a prognosis in the main fields of gastroenterology and hepatology (see **Supplementary Table 1** and **Supplementary Figure 1 and 2**).

Most studies use 1 dataset to train the machine learning process and a second dataset to test its performance. Some studies use common evaluation techniques such as cross-validation and leave 1 out.[8] To increase the size of the dataset, some studies use image-applied data augmentation, by a random resizing and cropping of the frame, followed by a random flipping along either axis. Datasets can include images of negative (normal) results positive (pathologic) results.

Analysis of Malignant and Premalignant Lesions

Fifty-three studies have used AI to detect malignant and premalignant intestinal lesions (**Table 1**). Most of these (48) focused on endoscopy, 3 studies used clinical and biological data extracted from electronic medical records (mainly demographics, cardiovascular comorbidities, concomitant medication, digestive symptoms, complete blood count), 1 study was based on serum tumor markers, and 1 study used data from gut microbiota. Twenty-seven studies were dedicated to improve diagnostic accuracy in case of colorectal polyps or cancer.[12–38] Nineteen studies focused on the diagnosis of premalignant or malignant lesions of the upper gastrointestinal tract,[39–57] only 4 studies were limited to the small bowel,[58–61] and 3 studies assessed the entire digestive tract.[62–64] Twenty-four studies used specific validation techniques—mainly *k*-fold cross-validation. For studies focusing on endoscopy, the size of training and test datasets varied widely across studies. Performance results were also heterogeneous from one study to another, but most of the presented algorithms reached an accuracy of more than 80%.

Two published randomized controlled trials compared the performance of endoscopy with or without assistance of AI-based algorithms. The first study tested the ability of a real-time DL system, WISENSE, to monitor blind spots during esophagogastroduodenoscopy (EGD). A total of 324 patients were randomly assigned to undergo EGD with or without WISENSE. WISENSE monitored blind spots with an accuracy of 90.4%, and the rate of blind spots was significantly lower in WISENSE group than in the control group (5.9% vs 22.5%).[65] The second study investigated the effect of a DL-based automatic polyp detection system during colonoscopy. A total of 1058 patients were randomly assigned to groups that underwent diagnostic colonoscopy with or without this assistance. The AI system significantly increased the rate of adenoma detection to 29.1% from 20.3%, and the mean

number of adenomas per patient, to 0.53 from 0.31.[66] These results indicate that AI systems could be used to improve the diagnostic value of everyday endoscopy for detecting premalignant lesions in the gastrointestinal tract.

Apart from improving diagnostic accuracy, AI might help physicians determine prognoses of patients with digestive cancer. An ANN model was developed from a dataset of 1219 patients with colorectal cancer. It provides more-accurate determinations of survival times and influential factors compared to a conventional Cox regression model,[67] and can be used to determine patients' risk for distant metastases.[68] An ANN model was used to assess 452 patients with gastric cancer, and determined survival times with approximately 90% accuracy.[69] In a study of 117 patients with stage IIA colon cancer after radical surgery, an ANN-based scoring system, based on molecular features of tumors, identified those with high, moderate, and low probability of survival for 10 years. The 10-year overall survival rate and disease free survival rate varied significantly between the three groups.[70] DL identifies patients with a complete response to neo-adjuvant chemoradiation for locally advanced rectal cancer with 80% accuracy. This technology might be used to identify patients most likely to benefit from conservative treatment vs radical resection.[71] A DL-based model was developed to predict survival times at 5 years of 1190 patients with gastric cancer, based on clinical and pathology data and treatment regimens. This system achieved an AUROC of 0.92 and identified associations between molecular features of tumor and optimal adjuvant treatment.[72]

Inflammatory and Other Non-malignant Lesions

AI has been used to identify patients with inflammatory bowel diseases (IBD) (n=6),[73–78] ulcers (n=6),[79–84] celiac disease (n=5),[85–89] lymphangiectasia (n=1),[90] and hookworm (n=1),[91] two studies evaluated endoscopic findings from patients with inflammatory lesions (**Table 2**).[92,93] Two studies used electronic medical records to determine patients' risk of celiac

disease, and 1 study used genetic factors to determine patients' risk of IBD. Two-thirds (14/21) of studies used *k*-fold cross-validation to avoid overfitting of data. Twelve out of twenty-one studies identified patients with approximately 90% accuracy.

Many studies have evaluated the ability of AI to predict responses to treatments in patients with IBD. Waljee et al developed an ML approach, using age and laboratory values, that is less costly and more accurate than 6-thioguanine nucleotide (6-TGN) metabolite measurement in predicting clinical responses to thiopurines; the ML approach identified patients with a clinical response with an AUROC curve of 0.86, vs 0.60 for 6-TGN levels.[94] This ML model was then improved to predict objective remission of patients receiving thiopurines, based on biomarkers, imaging data, and endoscopy findings. This ML model outperformed measurement of 6-TGN levels, identifying patients in remission with an AUROC curve of 0.79, vs 0.49 for the 6-TGN assay.[95] An ML model was developed to analyze data from a phase 3 trial of vedolizumab in patients with ulcerative colitis. The model predicted which patients would be in corticosteroid-free endoscopic remission at week 52 with an AUROC curve of 0.73, through week 6, vs and AUROC curve of 0.71 for level of fecal calprotectin. This algorithm might be used to select patients for continuation of vedolizumab when the benefits are not apparent in the first 6 weeks.[96] An AI algorithm that incorporates data on the microbiome with clinical data identified patients with IBD who had a clinical response to anti-integrin therapy with an AUROC curve of 0.78.[97] An ANN identified patients with ulcerative colitis who would require surgery after cytoapheresis therapy with 0.96 sensitivity and 0.87 specificity.[98]

AI systems are also being developed to predict onset or progression of IBD. A neural network that analyzes morphometric images of early-stage biopsies from patients with Crohn's disease identified those with disease progression with approximately 83.3% accuracy and a requirement for surgery with 86.0% accuracy.[99] Waljee et al constructed a ML method

to analyze data from electronic medical records that predicted IBD-related hospitalizations and outpatient use of steroids within 6 months with an AUROC curve of 0.87[100]. An ANN predicted the frequency of clinical relapse in patients with IBD with a high level of accuracy.[101]

Gastrointestinal Bleeding

Twelve studies have assessed the use of AI in detection of small bowel bleeding, based on images collected during wireless capsule endoscopy (**Table 3**).[102–111,55,112] Eight of 12 studies used specific validation techniques, mainly *k*-fold cross-validation. Among these studies, 9 identified patients with small bowel bleeding with an accuracy of more than 90%.

For patients with acute upper or lower gastrointestinal bleeding, the cause of hemorrhage can be easily determined by endoscopic examination. However, a significant proportion of patients have recurrent bleeding, which requires repeated endoscopies and treatments. ML models have therefore been developed to identify patients at risk for recurrent bleeding and those most likely to require treatment, and to estimate mortality. These models use clinical and/or biological data and identify these patients with approximately 90% accuracy.[113–117] An ML model, developed in a retrospective analysis of 22,854 patients with peptic ulcer and validated in 1265 patients, was able to identify patients with recurrent ulcer bleeding based on their age, level of hemoglobin, gastric ulcer, gastrointestinal diseases, malignancies, and infections. The model identified patients with recurrent ulcer bleeding within 1 year with an AUROC curve of 0.78 and an accuracy of 84.3%.[117]

Liver and Pancreatobiliary Disorders

Twenty-two studies have tested the ability of AI to aid in identification of patients with pancreatobiliary or liver diseases (**Table 4**). Six studies tested AI in detection of pancreatic

adenocarcinoma, based on endoscopic ultrasound[118–122] or markers in serum samples.[123] These studies identified patients with pancreatic cancer with an AUROC of approximately 90%. Sixteen studies focused on hepatology. Of those studies, 7 aimed to detect fibrosis associated with viral hepatitis,[124–130] 6 developed AI strategies to detect non-alcoholic fatty liver disease,[131–136] 2 were developed to identify patients with esophageal varices[137,138], and 1 to assess patients with chronic liver disease of any cause.[139] Thirteen studies used data from electronic medical records and/or biologic features to build the algorithms and three studies used data from elastography. All except 2 used specific validation techniques, mainly *k*-fold cross-validation. These models identified their target factor with approximately 80% accuracy.

Apart from improving diagnostic accuracy, methods are needed to determine patient prognoses and predict disease progression. Pearce et al developed an ML model to predict severity in patients with acute pancreatitis based on admission values of APACHE II score and levels of C-reactive protein. Their model predicted a severe attack with an AUROC curve of 0.82, 87% sensitivity, and 71% specificity.[140] Hong et al created an ANN to evaluate patients with acute pancreatitis based on their age, hematocrit, serum levels of glucose and calcium, and blood level of urea nitrogen—this model identified patients with persistent organ failure with 96.2% accuracy.[141] Jovanovic et al developed an ANN model to identify patients with suspected choledocholithiasis who require therapeutic endoscopic retrograde cholangiopancreatography based on clinical, laboratory, and transcutaneous ultrasound findings; it did so with an AUROC curve of 0.88.[142]

Banerjee et al developed an ANN based on clinical and laboratory data that identified patients with cirrhosis who would die within 1 year with 90% accuracy. This model can be used to identify the best candidates for liver transplantation.[143] Konerman et al created an ML model based on clinical, laboratory and histologic data that identified patients with chronic

hepatitis C virus (HCV) infection at highest risk for disease progression and liver-related outcomes (liver-related death, hepatic decompensation, hepatocellular carcinoma, liver transplant, or increase in Child-Pugh score to ≥7) with an AUROC curve of 0.78 in a validation cohort of 1007 patients.[144,145] Khosravi et al developed an ANN to predict survival times of 1168 patients undergoing liver transplantation; it estimated survival probability of one to five years with an AUROC curve of 86.4% compared to 80.7% for Cox proportional hazard regression models.[146] Researchers have also used ANN to match liver donors with recipients, which could provide powerful decision-making technology.[147] Moreover, ML models could help predict response to treatments. Takayama et al created an ANN that identified patients with chronic HCV infection who responded to therapy with pegylated interferon alpha-2b plus ribavirin with 82% sensitivity and 88% specificity.[148]

Future Directions

AI will be an important component of methods to determine diagnoses of patients seen by gastroenterologists and hepatologists, select treatments, and predict outcomes. Many methods have been developed with these aims, and found to have varying levels of performance. Differences in performance metrics make it difficult to compare the results from these studies. AI seems particularly valuable for use in endoscopy, where it could increase detection of malignant and premalignant lesions, inflammatory lesions, small-bowel bleeding, and pancreatobiliary disorders. In hepatology, AI techniques could be used to determine patients' risk of liver fibrosis and allow some patients to avoid liver biopsy.

Our review covered only articles listed in PubMed, and might have missed some publications in computer science and medical image analysis journals. Nonetheless, AI has become an important part of gastroenterology and hepatology research in the past 20 years. While this review focused on diagnosis and prognosis assistance, there are other areas where AI is being explored for purposes outside this field, for example the use of ML in assessing quality metrics for gastrointestinal endoscopy (caecal landmarks, ML to assess follow-up recommendations for surveillance colonoscopy), further extending the scope of application of AI in gastroenterology.

Limitations of AI techniques that require caution include the lack of high-quality datasets for ML development. Most evidence used to develop ML algorithms comes from pre-clinical studies, with no applications used in clinical practice at present. Furthermore, DL algorithms are considered to be black-box models, in which it is difficult to understand decision-making processes, preventing physicians from finding potential confounding factors. It is also important to consider ethical challenges; AI is not aware of the patient's preferences or legal liabilities. If an endoscopic misdiagnosis occurs, who is liable—the endoscopist, the programmer, or the manufacturer? Moreover, inherent biases, such as racial discrimination,

can be included in AI algorithms—especially in the field of hepatology, in determining risk of fibrosis related to viral hepatitis. In developing AI models, it is important to consider these factors and validate the models in a range of populations. Medicine always has intrinsic uncertainty, making perfect predictions impossible, and some research gaps related to AI in the field of gastroenterology and hepatology still remain to be investigated (Table 5).

There is no turning back for the development of AI in gastroenterology and hepatology, and future implications are large. The use of AI could expand access to care in undeserved or developing regions, especially in evaluating patients' risk of viral hepatitis or intestinal parasitic diseases. Smartphones can use AI technologies to monitor patients' health remotely—this has already been established with home measurement of fecal calprotectin by patients with IBD.[149] AI can also be used to identify new therapeutic targets, via synthesis of molecular, genetic, and clinical data from large patient datasets. However, AI will not completely replace doctors—computers and health care workers will always have to work together. Although the machine can make accurate predictions, ultimately, health care workers will have to make decisions for their patients based on patient's preferences, environment, and ethics.

Table 1. Use of AI in Identification of Patients With Intestinal Malignancies or Premalignant Lesions

Lesions	Diagnostic or predictive modality	AI classifier	AI validation methods	Number of images/cases[a] in training dataset (negative/positive)[b]	Number of images/cases[a] in test dataset (negative/positive)[b]	Best average results (%)		Reference
						Accuracy	Sensitivity/ Specificity	
Colon and rectum								
Polyps	high-magnification colonoscopy	regularized discriminant analysis or SVM	LOO	484 (198 non-adenomas/286 adenomas)		96.9	97.2/96.0	12
Polyps	colonoscopy (CE)	CNN	LOO	100 (75/25)[c]	2500 (1,875/625)	93.6	NA	13
Polyps	colonoscopy (WL or NBI)	RF, random subspace, or SVM	LOO	76 videos (15 serrated/21 hyperplastic/40 adenomas)		82.5	72.7/85.9	14
Polyps	colonoscopy	several CNN	not applicable	612 frames + 20 videos (10/10)	192 frames + 18 videos (9/9)	Several methods compared		15
Polyps	colonoscopy (NBI)	CNN	random sub-sampling	60,089[c] (223 videos; 29% type 1 and 53% type 2 based on NBI international colorectal endoscopic;18% no polyp)	125 (51 hyperplastic/74 adenomas	94.0	98.0/83.0	16
Polyps	colonoscopy	CNN	10-fold cross-validation	1200 (600/600)	10	70.0	83.3/50.0	17
Polyps	colonoscopy	SVM	-	100 videos split into training and test datasets		98.7	98.8/98.5	18
Polyps	colonoscopy	CNN	-	196,631 (133,496/63,135)		76.5	90.0/63.3	19
				411 videos	135 videos (85/50)			

				(306/105)				
Polyps	high-magnification colonoscopy (NBI)	CNN	-	2,157 (681 hyperplastic/1476 adenomas)	284 (96 hyperplastic/188 adenomas)	90.1	96.3/78.1	20
Polyps	Colonoscopy (WL or NBI)	CNN	7-fold cross-validation, dropout, early stopping	8641 (4553/4088)[c]	1,330 (658/672)	96.4	96.9/95.0	21
Polyps	colonoscopy (WL or NBI)	CNN	-	788 (205/583): 602 training dataset, 186 test dataset		78.0	92.3/62.5	22
Polyps	colonoscopy	CNN	-	5545 (1911/3634)	27,113 (21,572/5541)	AUROC, 0.98	94.4/95.9	23
Polyps	linked color imaging colonoscopy	Gaussian mixture model	-	208 (69/139) from 112 patients	181 (66/115) from 91 patients	78.4	83.3/70.1	24
Polyps	endocytoscopy (NBI and methylene blue)	SVM	-	61,925	466 (175/287/4 lost)	96.5	93.8/91.0	25
Polyps	endocytoscopy (NBI)	SVM	-	1661 (448 non-neoplasms/1213 neoplasms)	173 (49 non-neoplasms/124 neoplasms)	87.8	94.3/71.4	26
Polyps	WCE (colon)	SVM	-	1000 (800/200)	500 (400/100)	95.0	91.0/95.2	27
Polyps	WCE (colon)	MLP	non-maxima suppression	31,600 (30,000/1600)[c]	30,540 (30,000/540)[c]	80.0	NA	28
Polyps	WCE (colon)	binary	-	18,968 (18,738/230 corresponding to 16 polyps)		NA	81.2/90.2	29
Polyps	WCE (colon) or colonoscopy	CNN	-	7910	1695	96.4	97.1/93.3	30
				from 124 patients without and 131 patients with polyps				
CRC	colonoscopy	CNN	3-fold cross-	9942 (5124	5022 (2604	81.2	67.5/89.0	31

			validation	cTis+cT1a/4818 cT1b)[c]	cTis+cT1a/2418 cT1b)			
CRC	confocal laser endomicroscopy	two-layer NN	Early stopping	1035 (356/679)		84.5	1.17 (cross-entropy)	32
				725	155			
CRC	endocytoscopy	SVM	-	5543 (2,506 non-neoplasms, 2,667 adenomas, 370 cancers)	200 (100 adenomas, 100 cancers)	94.1	89.4/98.9	33
CRC	EMR	classification and regression trees, LR or RF	-	263,879 (262,587/1,292)		AUROC, 0.89	64.2/90.0	34
CRC	EMR	LR	5-fold cross-validation	90,000 (89,412/588)		AUROC, 0.90	68.0/3.5 (precision)	35
CRC	EMR	Gradient boosting model or RF	-	112,584 (112,451/133)		odds ratio, 21.8	17.3/NA	36
CRC	serum markers of tumors	SVM	-	40 (20/20)	166 (66/100)	82.5	85.0/80.0	37
CRC	intestinal microbiota	Bayes net, RF, simple logistic, or logistic model tree	-	141 (93/48)		0.99 (AUROC)	93.5/97.9	38
Upper gastrointestinal tract								
EN-BE	EGD (WL)	SVM	LOO	100 (60/40) from 23 patients without EN-BE and 21 patients with EN-BE		NA	86.0/87.0	39

EN-BE	volumetric laser endomicroscopy	several[d]	LOO	60 (30/30)		0.95 (AUROC)	90.0/93.0	40
BE/ED	EGD (NBI)	SVM	10-fold cross-validation	197 (36/161) from 84 patients		91.8	91.8/92.1	41
ESCC and EAC	EGD (WL or NBI)	CNN	-	8428 from 384 patients (397 ESCC, 32 EAC)	1118 (956/162) from 50 control, 41 ESCC patients, 8 EAC patients	55.7	98.0/16.0	42
ESCC	endocytoscopy (with or without high magnification	CNN	-	4715 (3574/1141) from 114 noncancerous and 126 cancerous patients	1520 from 55 patients (27 ESCC, 28 benign lesions)	90.9	92.6/89.3	43
ESCC	magnifying EGD (NBI)	VGG16 Net	3-fold cross-validation	1,383 from 219 patients (54/165)		89.2	87.0/84.1	44
HP infection	EGD (WL)	CNN	-	32,208 from 1015 patients negative for *H pylori* infection and 753 patients positive	11,481 from 325 patients negative for *H pylori* infection and 72 patients positive	87.7	88.9/87.4	45
HP infection	EGD	CNN	-	596[c] from 74 patients negative for *H pylori* infection and 65 patients positive	30 (15/15)	0.96 (AUROC)	86.7/86.7	46
EGC	EGD (WL, CE, NBI)	CNN	-	13,584 from 2639 lesions	2296 from 77 lesions	NA	92.2/NA	47
EGC	EGD with CE	SVM	-	200 (100/100) from 18 patients	3800 (1900/1900) from 18 patients	0.69 (F1 score)	NA	48
EGC	EGD with CE	several[e]	10-fold cross-	176 (56/120)		87.0	91.0/82.0	49

			validation					
EGC	EGD (WL)	CNN	-	348,943 (176,388/172,555)[c] from 58 EGC patients	9650 (4997/4653)[c] from 58 EGC patients	87.6	80.0/94.8	50
EGC	EGD (M-NBI)	SVM	-	126 (60/66)	81 (20/61)	96.3	96.7/95.0	51
EGC	EGD	CNN	5-fold cross-validation and early stopping	9151 (5981/3170)	200 (100/100)	92.5	94.0/91.0	52
EGC	EGD	inception network, ResNet, or VGGNet	-	717 (180 normal/200 ulcers/337 EGC)	70 (20 normal/20 ulcers/30 EGC)	96.0	NA	53
EGC, BE	EGD (CE, NBI)	SVM	-	426 CE images (132/294) and 672 NBI images (171/501)[c]	426 CE images (132/294) and 672 NBI images (171/501)[c]	83.1 (CE) / 88.4 (NBI)	83.1 (CE) and 87.5 (NBI)	54
EGC, ESCC, and EAC	EGD	joint diagonalization principal component analysis	10-fold cross-validation	800 (520/150 early ESCC or EAC/130 EGC) from 291 patients		90.8 (ESCC or EAC or EGC)	ESCC or EAC, 93.3/89.2 EGC, 90.8/90.7	55
Invasive GC	EGD	CNN	10-fold cross-validation	812 from 344 patients (among 902 images: 448 T1/ 106 T2/149 T3/199 T4)	90	77.2	NA	56
Invasive GC	EGD	ResNet50	Bootstrapping	5056[c] from 790 patients (545/245)	203 from 203 patients (135/68)	89.2	76.5/95.6	57

Small bowel								
Polyps	WCE (Small bowel)	ANN	-	54 videos (46/8)	90 images (58/32)	97.7	93.8/91.4	58
Tumors	WCE (Small bowel)	several[f]	4-fold cross-validation	900 (450/450) from 10 patients	300 (150/150) from 10 patients	90.5	92.3/88.7	59
Tumors	WCE (Small bowel)	SVM	4-fold cross-validation	600 (300/300) from 6 patients	200 (100/100) from 6 patients	93.5	94.0/93.0	60
Tumors	WCE (Small bowel)	SVM	10-fold cross-validation	1800 (900/900) from 90 control and 15 patients with polyps		97.3	97.8/96.7	61
Entire gastrointestinal tract								
Polyps	WCE	MLP	3-fold cross-validation	300 (150/150) from 2 patients		86.1	89.8/82.5	62
Polyps	WCE	Softmax	-	4000 (3,000/1,000) from 35 patients		98.0	95.5/98.5	63
Tumors	WCE	SVM	10-fold cross-validation	1200 (600/600) from 10 patients		92.4	88.6/96.2	64

BE, Barrett's esophagus; CE, chromoendoscopy; CNN, convolutional neural network; CRC, colorectal cancer; EAC, esophageal adenocarcinoma; EGC, early-stage gastric cancer; EMR, electronic medical records; EN-BE, early-stage neoplasia in patients with BE; ESCC, esophageal squamous cell carcinoma; LOO, leave 1 out; LR, logistic regression; MLP, multilayer perceptron network; NA, not available; NBI, narrow-band imaging; RF, random forest; SVM, support vector machine; WCE, wireless capsule endoscopy; WL, white light.

[a]: Number of images (frames or videos) for studies analyzing endoscopy. Number of cases (patients) for studies analyzing electronic medical records, microbiota, serum markers.; [b]: The number of negative and positive data is provided if applicable; [c]: After data augmentation; [d]: SVM, discriminant analysis, AdaBoost, RF, K-nearest neighbors, Naïve Bayes, linear regression, logistic regression; [e]: SVM, Naïve Bayes, K-nearest neighbors, linear discriminant analysis, decision tree, and ensemble classifiers; [f]: K-nearest neighbor, MLP and SVM.

-: No specific validation technique was used (or missing data).

Table 2. Use of AI in Identification of Patients With Inflammatory or Other Non-malignant Diseases or Lesions

Diseases/ Lesions	Diagnostic or predictive modality	AI classifier	Validation methods	Number of images/cases[a] in training dataset (negative/positive)[b]	Number of images/cases[a] in test dataset (negative/positive)[b]	Best average results (%)		Reference
						Accuracy	Sensitivity/ Specificity	
CD	WCE (SB)	SVM	-	469 (245/224) from 29 patients with CD, 17 control patients, 1 patient with celiac disease	277 (150/127)	87.0	80.0/93.0	73
CD	WCE (SB)	SVM	3-fold cross-validation	533 (212 normal/213 mild/108 moderate-to-severe) from 47 patients (30 with CD lesions)		80.2	81.1/93.6	74
CD	WCE (SB)	SVM	-	1,828 (1728/100) from 2 patients (1 with CD)		100.0	NA	75
CD	WCE (SB)	SVM	3-fold cross-validation	800 (400 normal/152 mild/248 severe) from 13 patients	102 (66/36)	93.8	95.2/92.4	76
Pediatric IBD	endoscopy / histology	ensemble learners, linear discriminant analysis, or SVM	5-fold cross-validation	210 children (143 with CD and 67 with UC)	48 children (35 with CD and 13 with UC)	83.3	83.0/NA	78
IBD	genetics	SVM or gradient-boosted trees	10-fold cross-validation	53,279 (22,442 controls, 17,379 patients with CD, and 13,458 patients with UC)		AUROC: 0.86 (CD) 0.83 (UC)	71.0/83.0 (CD)	77
Peptic ulcers	WCE (SB)	SVM	-	50 (30/20)		74.0	75.0/73.3	79

Peptic ulcers	WCE (SB)	SVM		250 (184/66)	930 (470/460) from 30 videos	96.3	91.7/99.4	80
Ulcers[c]	WCE (SB)	several[d]	10-fold cross-validation	156 (78 from 78 different lesions)	18 (9 from 9 different lesions)	95.0	96.6/93.5	81
Ulcers[c]	WCE (SB)	SVM	10-fold cross-validation	260 (130 from 130 different lesions)		86.5	84.5/88.6	82
Ulcers[c]	WCE (SB)	SVM	5-fold cross-validation	272 (136 from 136 different lesions) from 20 patients	68 (34 from 34 different lesions) from 20 patients	92.7	94.1/91.2	83
Ulcers[c]	WCE (SB)	SVM or vector-supported convex hull	-	613 (500/113) from 50 videos	200 (100/100)	NA	100.0/100.0	84
Celiac disease	WCE (SB)	Threshold or incremental	-	8600 (4000/4600) from 5 control and 6 celiac patients	10,000 (5,000/5,000) from 5 control and 5 celiac patients	76.7	88.0/80.0	85
Celiac disease	EGD (WL or NBI)	SVM	Cross-validation	2,835 from 215 controls/75 children with celiac disease		99.6	NA	87
Celiac disease	WCE (SB)	GoogLeNet	7-fold cross-validation	8800 (4000/4800) from 5 control and 6 celiac patients	8000 (4000/4000) from 5 control and 5 celiac patients	NA	100.0/100.0	88
Celiac disease	EMR	several[e]	10-fold cross-validation	178 (96 controls and 82 with celiac disease)	38 (24 controls and 14 with celiac disease)	80.0	78.8/80.0	86
Celiac disease	EMR	several[f]	10-fold	816 (408/408)		0.53	NA	89

			cross-validation			(AUROC)		
Lymphangiectasia	WCE (SB)	threshold	-	7218		97.9	48.8/NA	90
Hookworm	WCE (SB)	rusboost	11-fold LOO cross-validation	401,476 (397,087/4389) from 10 patients	40,148 (39,709/439) from 1 patient	78.2	77.2/77.9	91
Several lesions[g]	WCE (SB)	SVM	10-fold cross-validation	1370 from 252 WCE procedures	137 (60/77)	94.0	95.4/82.9	92
Several lesions[h]	WCE/Colonoscopy	KNN	10-fold cross-validation	1250 (800/450)	500 (400/100)	96.9	91.0/97.3	93

CD, Crohn's disease; SB, small bowel.

[a]: Number of images for studies analyzing endoscopy or number of cases for studies analyzing electronic medical records; [b]: The number of negative and positive data is provided if applicable; [c]: Ulcers from Crohn's disease, unexplained ulcerations and ulcerations from NSAID (Nonsteroidal anti-inflammatory drugs); [d]: discriminant analysis-based classifiers, SVM with radial basis function, multilayer neural network; [e]: decision trees, Bayesian inference, K nearest neighbor (KNN), SVM and artificial neural networks; [f]: logistic regression, elastic net, tree-based models, SVM with radial basis functions, a neural network (single layer perceptron), random forest, and linear discriminant analysis; [g]: Angiectasias, intraluminal hemorrhage, aphthae, ulcers, stenoses, villous edema, nodular lymphangiectasias, chylous cysts, polyps; [h]: erythema, blood, polyps, ulcers, erosions.

-: No specific validation technique was used (or missing data).

Table 3. Use of AI in Detection of Small Bowel Bleeding

Diagnostic or predictive modality	AI classifier	Validation methods	Number of images in training dataset (negative/positive)	Number of images in test dataset (negative/positive)	Best average results (%)		Reference
					Accuracy	Sensitivity/ Specificity	
WCE (SB)	color spectrum transformation	Patient adaptive method	4800 (3378/1422) from 12 patients (3 normal, 9 abnormal)		30.8	94.9/96.1	102
WCE (SB)	MLP	4-fold cross-validation	2700 (1350/1350) from 10 patients	900 (450/450) from 10 patients	NA	87.8/88.6	103
WCE (SB)	Probabilistic neural network	-	14,630 (11,458/3,172) from 150 videos		87.4	93.1/85.8	104
WCE (SB)	SVM	5-fold cross-validation	280 (140/140) from 9 videos	280 (140/140) from 9 videos	97.9	97.8/98.0	105
WCE (SB)	SVM	10-fold cross-validation	30,000 pixels (20,000/10,000)	30,000 pixels (20,000/10,000)	94.0	97.0/92.0	106
			5000 (4000/1000) from 20 videos				
WCE (SB)	SVM	LOO cross validation	250 (200/50)	2000 (1600/400)	94.5	93.0/94.9	107
			from 30 videos				
WCE (SB)	MLP	-	100 (50/50)		93.0	96.0/90.0	108
WCE (SB)	SVM	-	1200 (600/600) from 6 control and 6 bleeding patients	1720 (860/860) from 10 control and 10 bleeding patients	99.2	99.4/99.0	109
WCE (SB)	SVM	10-fold cross-	8200 (6,150/2,050)	1800 (1000/800)	99.6 (F1 score)	99.2/99.9 (precision)	110

		validation					
WCE (SB)	SVM	10-fold cross-validation	75,000 pixels (50,000/25,000)	8500 (5500/3,000) from 30 videos	91.8	93.7/90.7	111
WCE (SB)	joint diagonalization principal component analysis	10-fold cross-validation	530 (400/130) from 30 patients		94.3	93.9/94.5	55
WCE (SB)	CNN	-	600 (300/300) from 200 control and 208 abnormal videos	600 (300/300) from 200 control and 208 abnormal videos	98.0	100.0/96.0	112

-: No specific validation technique was used (or missing data).

Table 4. Use of AI in Identification of Patients With Pancreatic or Liver Diseases

Diseases/ Lesions	Diagnostic or predictive modality	AI classifier	Validation methods	Number of images/videos/cases[a] in training dataset (negative/positive)[b]	Number of images/videos/cases[a] in test dataset (negative/positive)[b]	Best average results (%)		Reference
						Accuracy	Sensitivity/ Specificity	
Pancreatobiliary field								
PAAD or CP	EUS	MLP	cross-validation	160	159	0.93 (AUROC)	93.0/92.0	118
				from a total of 319 (110 normal/99 CP/110 PAAD) from 22, 12, 22 patients, respectively				
PAAD or CP	Real-time EUS elastography	MLP	10-fold cross-validation	774 (129 CP/645 PAAD) from 47 and 211 patients, respectively		84.3	87.6/82.9	119
PAAD or CP	EUS	SVM	LOO	194 (131 CP/63 PAAD)	194 (63 CP/131 PAAD)	94.2	91.6/95.0	120
PAAD or CP	Contrast-enhanced EUS	ANN	Early stopping	117 (39/78)	25 (8/17)	94.6	94.6/94.4	121
PAAD	EUS	Age-based MLP	-	260 (100/160)	72 (30/42)	87.5	83.3/93.3	122
				from 40 patients<40, 58 patients aged 40-60, 74 patients>60				
PAAD	Serum tumor markers	ANN	-	658 (195/463)	255 (75/180)	0.91 (AUROC)	NA	123
Hepatology								
HCV-associated fibrosis	EMR, biology	ANN	internal cross-validation	414[c] (319 F0-F1/95 F3-F4) from 123 patients	96[c] (73 F0-F1/23 F3-F4) from 65 patients	0.93 (AUROC)	100.0/79.5	124
HCV-	EMR,	DT, genetic	10-fold	22,690 (19,349 F0-	16,877 (14,200 F0-	84.4	07.0/99.0	125

associated fibrosis	biology	analysis, particle swarm optimization, or MLR	cross-validation	F2/3,341 F3-F4)	F2/2,677 F3-F4)			
HBV-associated fibrosis	EMR, biology	Bayesian ANN	early stopping	226 (166 F0-F1/60 F2-F4)	116 (80 F0-F1/36 F2-F4)	0.92 (AUROC)	83.3/85.0	126
HBV-associated fibrosis	real-time tissue elastography	SVM/NB/RF/KNN	4-fold cross-validation	257 (60 stage 0, 82 stage 1, 44 stage 2, 36 stage 3, 35 stage 4)	256 (59 stage 0, 82 stage 1, 44 stage 2, 36 stage 3, 35 stage 4)	91.3	72.9/99.2	127
HBV-associated fibrosis	Shear wave elastography	CNN	-	1,330 from 266 patients	660 from 132 patients	0.98 (AUROC)	90.4/98.3	128
HBV- or HCV-associated fibrosis	EMR, biology	DT, RF, or gradient boosting	10-fold cross-validation	343	147	0.92 (AUROC)	84.0/85.0	129
HBV-associated cirrhosis	EMR, biology	ANN/LR	random subsampling	86 (75/11)	58 (50/8)	91.4	87.5/92.0	130
Chronic liver disease	Shear wave elastography	SVM	LOO	126 (56/70) from 126 patients		87.3	93.5/81.2	139
Esophageal varices	EMR, biology	MLP	holdout	110	30	87.8	93.8/71.7	137
				from 197 (53/144) patients				
Esophageal varices	EMR, biology	RF	bootstrapping	238 (129 negative, 54 varices not needing treatment and 55 that	109 (45 varices not needing treatment and 34 that needed	0.82 (AUROC)	100.0/49.3	138

				needed treatment)	treatment)			
NAFLD	biology	LR/KNN/ SVM/DT/ RF	10-fold LOO cross-validation	126 (101 fibrosis stage 2 and 25 fibrosis stage 1)		79.0	60.0/77.0	131
NAFLD or alcoholic liver disease	biology	LR/DT/SVM /RF	10-fold LOO cross-validation	133 (31 with NAFLD and 102 with alcoholic liver disease)		89.0	74.2/98.0	132
NAFLD	EMR, biology	several[d]	-	500 (354/146)	422 (304/118)	0.88 (AUROC)	92.4/90.5	133
NAFLD	EMR, biology, ultrasonography	several[e]	10-fold cross-validation	10,508 (7,986/2,522)		82.9	67.5/87.8	134
NAFLD	EMR, biology	RF, naïve Bayes, ANN, or LR	10-fold cross-validation	519 (180/339)	58 (20/38)	86.5	87.2/85.9	135
Non-alcoholic steato-hepatitis	EMR	LR/DT/RF/XG-Boost	Random subsampling and 5-fold cross-validation	108,139 (17,590 controls/73,190 NAFLD/17,359 NASH)		79.7	77.4/80.8 (precision)	136

CP, chronic pancreatitis; DT, decision tree; HBV, hepatitis B virus; NAFLD: Non-alcoholic fatty liver disease; PAAD, pancreatic adenocarcinoma

[a]: Number of images for studies analyzing EUS and shear wave elastography. / Number of videos for studies analyzing EUS elastography. / Number of cases for studies analyzing electronic medical records, biology and CEH-EUS; [b]: The number of negative and positive data is provided if applicable; [c]: Biopsies from liver transplant recipients; [d]: LR, ridge regression, AdaBoost and DT models; [e]: KNN, SVM, LR, NB, Bayesian network, DT, adaptive boosting, bootstrap aggregating, RF, hidden naïve Bayes and aggregating one-dependence estimators.

-: No specific validation technique was used (or missing data).

Table 5. Main research gaps still to be investigated for AI in the field of gastroenterology and hepatology.

Main research gaps
Variations in performance levels of AI due to: - Lack of high-quality datasets for ML development and great heterogeneity in the size of training and testing datasets - Wide variety of performance metrics (accuracy, sensitivity, specificity, precision, F1 score, AUROC) - Lack of validation techniques in multiple studies
Lack of randomized controlled trials comparing AI-assisted approaches to current non-AI-based approaches: - Only two published RCTs up to date, most evidence used to develop ML algorithms coming from pre-clinical studies
Limitations of AI techniques that require further investigation: - "Black-box models" preventing physicians from finding potential confounding factors - Ethical challenges - Inherent biases
Multiple other areas outside the field of diagnosis and prognosis assistance are still under-investigated: - Quality metrics for gastrointestinal endoscopy - Bedside computer vision especially in intensive care units - Therapeutic advances particularly in immunotherapy

References

1. Russell S, Norvig P. Artificial Intelligence: A Modern Approach, Global Edition. 3rd ed. Pearson; 2016.
2. Colom R, Karama S, Jung RE, et al. Human intelligence and brain networks. Dialogues Clin Neurosci 2010;12:489–501.
3. Shalev-Shwartz S, Ben-David S. Understanding Machine Learning: From Theory to Algorithms-. New York, NY, USA: Cambridge University Press; 2014.
4. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–444.
5. Goodfellow I, Bengio Y, Courville A. Deep Learning. The MIT Press; 2016.
6. Mahmud M, Kaiser MS, Hussain A, et al. Applications of Deep Learning and Reinforcement Learning to Biological Data. IEEE Transactions on Neural Networks and Learning Systems 2018;29:2063–2079.
7. Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. Digital Signal Processing 2018;73:1-15.
8. Nathalie Japkowicz and Mohak Shah. Evaluating Learning Algorithms: A Classification Perspective. Cambridge University Press 2011, New York, NY, USA.
9. Ahmad OF, Soares AS, Mazomenos E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. The Lancet Gastroenterology & Hepatology 2019;4:71–80.
10. Ruffle JK, Farmer AD, Aziz Q. Artificial Intelligence-Assisted Gastroenterology—Promises and Pitfalls: The American Journal of Gastroenterology 2019;114:422–428.
11. Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. Nature Reviews Gastroenterology & Hepatology 2015;12:172–186.

12. Häfner M, Brunauer L, Payer H, et al. Computer-aided classification of zoom-endoscopical images using Fourier filters. IEEE Trans Inf Technol Biomed 2010;14:958–970.

13. Ribeiro E, Uhl A, Wimmer G, et al. Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification. Comput Math Methods Med 2016;2016. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5101370/ [Accessed September 14, 2018].

14. Mesejo P, Pizarro D, Abergel A, et al. Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy. IEEE Trans Med Imaging 2016;35:2051–2063.

15. Bernal J, Tajkbaksh N, Sanchez FJ, et al. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge. IEEE Transactions on Medical Imaging 2017;36:1231–1249.

16. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2017.

17. Komeda Y, Handa H, Watanabe T, et al. Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. Oncology 2017;93 Suppl 1:30–34.

18. Billah M, Waheed S, Rahman MM. An Automatic Gastrointestinal Polyp Detection System in Video Endoscopy Using Fusion of Color Wavelet and Convolutional Neural Network Features. Int J Biomed Imaging 2017;2017:9545920.

19. Misawa M, Kudo S-E, Mori Y, et al. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. Gastroenterology 2018;154:2027-2029.e3.

20. Chen P-J, Lin M-C, Lai M-J, et al. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. Gastroenterology 2018;154:568–575.

21. Urban G, Tripathi P, Alkayali T, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology 2018.

22. Renner J, Phlipsen H, Haller B, et al. Optical classification of neoplastic colorectal polyps - a computer-assisted approach (the COACH study). Scand J Gastroenterol 2018;53:1100–1106.

23. Wang P, Xiao X, Brown JRG, et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nature Biomedical Engineering 2018;2:741.

24. Min M, Su S, He W, et al. Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology. Sci Rep 2019;9:2881.

25. Mori Y, Kudo S-E, Misawa M, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Ann Intern Med 2018;169:357–366.

26. Misawa M, Kudo S, Mori Y, et al. Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for diagnosing colorectal lesions: comparison with experts. International Journal of Computer Assisted Radiology and Surgery 2017;12:757–766.

27. Romain O, Histace A, Silva J, et al. Towards a multimodal wireless video capsule for detection of colonic polyps as prevention of colorectal cancer. 2013:1–6.

28. David E, Boia R, Malaescu A, et al. Automatic colon polyp detection in endoscopic capsule images. 2013:1–4.

29. Mamonov AV, Figueiredo IN, Figueiredo PN, et al. Automated Polyp Detection in Colon Capsule Endoscopy. IEEE Transactions on Medical Imaging 2014;33:1488–1502.

30. Blanes-Vidal V, Baatrup G, Nadimi ES. Addressing priority challenges in the detection and assessment of colorectal polyps from capsule endoscopy and colonoscopy in colorectal cancer screening using machine learning. Acta Oncol 2019;58:S29–S36.

31. Ito N, Kawahira H, Nakashima H, et al. Endoscopic Diagnostic Support System for cT1b Colorectal Cancer Using Deep Learning. Oncology 2018:1–7.

32. Ştefănescu D, Streba C, Cârţână ET, et al. Computer Aided Diagnosis for Confocal Laser Endomicroscopy in Advanced Colorectal Adenocarcinoma. PLoS ONE 2016;11:e0154863.

33. Takeda K, Kudo S-E, Mori Y, et al. Accuracy of diagnosing invasive colorectal cancer using computer-aided endocytoscopy. Endoscopy 2017;49:798–802.

34. Kop R, Hoogendoorn M, Teije AT, et al. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. Comput Biol Med 2016;76:30–38.

35. Hoogendoorn M, Szolovits P, Moons LMG, et al. Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. Artif Intell Med 2016;69:53–61.

36. Kinar Y, Akiva P, Choman E, et al. Performance analysis of a machine learning flagging system used to identify a group of individuals at a high risk for colorectal cancer. PLoS ONE 2017;12:e0171759.

37. Zhang B, Liang XL, Gao HY, et al. Models of logistic regression analysis, support vector machine, and back-propagation neural network based on serum tumor markers in colorectal cancer diagnosis. Genet Mol Res 2016;15.

38. Ai L, Tian H, Chen Z, et al. Systematic evaluation of supervised classifiers for fecal microbiota-based prediction of colorectal cancer. Oncotarget 2017;8:9546–9556.

39. Sommen F van der, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. Endoscopy 2016;48:617–624.

40. Swager A-F, Sommen F van der, Klomp SR, et al. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. Gastrointest Endosc 2017;86:839–846.

41. Riaz F, Ribeiro M-D, Pimentel-Nunes P, et al. Integral scale histogram local binary patterns for classification of narrow-band gastroenterology images. Conf Proc IEEE Eng Med Biol Soc 2013;2013:3714–3717.

42. Horie Y, Yoshio T, Aoyama K, et al. The diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. Gastrointest Endosc 2018.

43. Kumagai Y, Takubo K, Kawada K, et al. Diagnosis using deep-learning artificial intelligence based on the endocytoscopic observation of the esophagus. Esophagus 2018.

44. Zhao Y-Y, Xue D-X, Wang Y-L, et al. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy. Endoscopy 2019;51:333–341.

45. Shichijo S, Nomura S, Aoyama K, et al. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. EBioMedicine 2017;25:106–111.

46. Itoh T, Kawahira H, Nakashima H, et al. Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images. Endosc Int Open 2018;6:E139–E144.

47. Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018;21:653–660.

48. Ogawa R, Nishikawa J, Hideura E, et al. Objective Assessment of the Utility of Chromoendoscopy with a Support Vector Machine. J Gastrointest Cancer 2018.

49. Ali H, Yasmin M, Sharif M, et al. Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images. Comput Methods Programs Biomed 2018;157:39–47.

50. Sakai Y, Takemoto S, Hori K, et al. Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. Conf Proc IEEE Eng Med Biol Soc 2018;2018:4138–4141.

51. Kanesaka T, Lee T-C, Uedo N, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. Gastrointest Endosc 2018;87:1339–1344.

52. Wu L, Zhou W, Wan X, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. Endoscopy 2019. Available at: http://www.thieme-connect.de/DOI/DOI?10.1055/a-0855-3532 [Accessed April 30, 2019].

53. Lee JH, Kim YJ, Kim YW, et al. Spotting malignancies from gastric endoscopic images using deep learning. Surg Endosc 2019.

54. Riaz F, Silva FB, Ribeiro MD, et al. Invariant Gabor texture descriptors for classification of gastroenterology images. IEEE Trans Biomed Eng 2012;59:2893–2904.

55. Liu D-Y, Gan T, Rao N-N, et al. Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process. Medical Image Analysis 2016;32:281–294.

56. Kubota K, Kuroda J, Yoshida M, et al. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. Surg Endosc 2012;26:1485–1489.

57. Zhu Y, Wang Q-C, Xu M-D, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. Gastrointest Endosc 2019;89:806-815.e1.

58. Constantinescu AF, Ionescu M, Iov V-F, et al. A computer-aided diagnostic system for intestinal polyps identified by wireless capsule endoscopy. :Rom J Morphol Embryol 2016, 57(3):979–984.

59. Li B, Meng MQ-H, Lau JYW. Computer-aided small bowel tumor detection for capsule endoscopy. Artif Intell Med 2011;52:11–16.

60. Faghih Dinevari V, Karimian Khosroshahi G, Zolfy Lighvan M. Singular Value Decomposition Based Features for Automatic Tumor Detection in Wireless Capsule Endoscopy Images. Applied Bionics and Biomechanics 2016;2016:1–8.

61. Liu G, Yan G, Kuang S, et al. Detection of small bowel tumor based on multi-scale curvelet analysis and fractal technology in capsule endoscopy. Computers in Biology and Medicine 2016;70:131–138.

62. Li B, Meng MQH, Xu L. A comparative study of shape features for polyp detection in wireless capsule endoscopy images. Conf Proc IEEE Eng Med Biol Soc 2009;2009:3731–3734.

63. Yuan Y, Meng MQ-H. Deep learning for polyp recognition in wireless capsule endoscopy images. Medical Physics 2017;44:1379–1389.

64. Baopu Li, Meng MQ-H. Tumor Recognition in Wireless Capsule Endoscopy Images Using Textural Features and SVM-Based Feature Selection. IEEE Transactions on Information Technology in Biomedicine 2012;16:323–329.

65. Wu L, Zhang J, Zhou W, et al. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019.

66. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019.

67. Gohari MR, Biglarian A, Bakhshi E, et al. Use of an artificial neural network to determine prognostic factors in colorectal cancer patients. Asian Pac J Cancer Prev 2011;12:1469–1472.

68. Biglarian A, Bakhshi E, Gohari MR, et al. Artificial neural network for prediction of distant metastasis in colorectal cancer. Asian Pac J Cancer Prev 2012;13:927–930.

69. Nilsaz-Dezfouli H, Abu-Bakar MR, Arasan J, et al. Improving Gastric Cancer Outcome Prediction Using Single Time-Point Artificial Neural Network Models. Cancer Inform 2017;16:1176935116686062.

70. Peng J-H, Fang Y-J, Li C-X, et al. A scoring system based on artificial neural network for predicting 10-year survival in stage II A colon cancer patients after radical surgery. Oncotarget 2016;7:22939–22947.

71. Bibault J-E, Giraud P, Durdux C, et al. Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. Sci Rep 2018;8:12611.

72. Lee J, An JY, Choi MG, et al. Deep Learning–Based Survival Analysis Identified Associations Between Molecular Subtype and Optimal Adjuvant Treatment of Patients With Gastric Cancer. JCO Clinical Cancer Informatics 2018:1–14.

73. Girgis HZ, Mitchell BR, Dassopoulos T, et al. An intelligent system to detect Crohn's disease inflammation in Wireless Capsule Endoscopy videos. 2010:1373–1376.

74. Kumar R, Qian Zhao, Seshamani S, et al. Assessment of Crohn's Disease Lesions in Wireless Capsule Endoscopy Images. IEEE Transactions on Biomedical Engineering 2012;59:355–362.

75. Jebarani WSL, Daisy VJ. Assessment of Crohn's disease lesions in Wireless Capsule Endoscopy images using SVM based classification. 2013:303–307.

76. Charisis VS, Hadjileontiadis LJ. Potential of hybrid adaptive filtering in inflammatory lesion detection from capsule endoscopy images. World J Gastroenterol 2016;22:8641–8657.

77. Wei Z, Wang W, Bradfield J, et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. Am J Hum Genet 2013;92:1008–1012.

78. Mossotto E, Ashton JJ, Coelho T, et al. Classification of Paediatric Inflammatory Bowel Disease using Machine Learning. Sci Rep 2017;7:2427.

79. Karargyris A, Bourbakis N. Identification of ulcers in Wireless Capsule Endoscopy videos. 2009:554–557.

80. Chen Y, Lee J. Ulcer detection in wireless capsule endoscopy video. In: Proceedings of the 20th ACM international conference on Multimedia - MM '12. Nara, Japan: ACM Press; 2012:1181. Available at: http://dl.acm.org/citation.cfm?doid=2393347.2396413 [Accessed September 12, 2018].

81. Charisis VS, Hadjileontiadis LJ, Liatsos CN, et al. Capsule endoscopy image analysis using texture information from various colour models. Computer Methods and Programs in Biomedicine 2012;107:61–74.

82. Eid A, Charisis VS, Hadjileontiadis LJ, et al. A curvelet-based lacunarity approach for ulcer detection from Wireless Capsule Endoscopy images. 2013:273–278.

83. Yuan Y, Wang J, Li B, et al. Saliency Based Ulcer Detection for Wireless Capsule Endoscopy Diagnosis. IEEE Transactions on Medical Imaging 2015;34:2046–2057.

84. Szczypiński P, Klepaczko A, Pazurek M, et al. Texture and color based image segmentation and pathology detection in capsule endoscopy videos. Computer Methods and Programs in Biomedicine 2014;113:396–411.

85. Ciaccio EJ, Tennyson CA, Bhagat G, et al. Classification of videocapsule endoscopy image patterns: comparative analysis between patients with celiac disease and normal individuals. BioMedical Engineering OnLine 2010;9:44.

86. Tenório JM, Hummel AD, Cohrs FM, et al. Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease. Int J Med Inform 2011;80:793–802.

87. Gadermayr M, Kogler H, Karla M, et al. Computer-aided texture analysis combined with experts' knowledge: Improving endoscopic celiac disease diagnosis. World J Gastroenterol 2016;22:7124–7134.

88. Zhou T, Han G, Li BN, et al. Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method. Computers in Biology and Medicine 2017;85:1–6.

89. Hujoel IA, Murphree DH, Van Dyke CT, et al. Machine Learning in Detection of Undiagnosed Celiac Disease. Clin Gastroenterol Hepatol 2018;16:1354-1355.e1.

90. Lei Cui, Chao Hu, Yuexian Zou, et al. Detection of lymphangiectasia disease from wireless capsule endoscopy images with adaptive threshold. In: 2010 8th World Congress on Intelligent Control and Automation. Jinan, China: IEEE; 2010:3088–3093. Available at: http://ieeexplore.ieee.org/document/5554005/ [Accessed September 17, 2018].

91. Wu X, Chen H, Gan T, et al. Automatic Hookworm Detection in Wireless Capsule Endoscopy Images. IEEE Transactions on Medical Imaging 2016;35:1741–1752.

92. Iakovidis DK, Koulaouzidis A. Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. Gastrointestinal Endoscopy 2014;80:877–883.

93. Nawarathna R, Oh J, Muthukudage J, et al. Abnormal image detection in endoscopy videos using a filter bank and local binary patterns. Neurocomputing 2014;144:70–91.

94. Waljee AK, Joyce JC, Wang S, et al. Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. Clin Gastroenterol Hepatol 2010;8:143–150.

95. Waljee AK, Sauder K, Patel A, et al. Machine Learning Algorithms for Objective Remission and Clinical Outcomes with Thiopurines. J Crohns Colitis 2017;11:801–810.

96. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. Aliment Pharmacol Ther 2018;47:763–772.

97. Ananthakrishnan AN, Luo C, Yajnik V, et al. Gut Microbiome Function Predicts Response to Anti-integrin Biologic Therapy in Inflammatory Bowel Diseases. Cell Host Microbe 2017;21:603-610.e3.

98. Takayama T, Okamoto S, Hisamatsu T, et al. Computer-Aided Prediction of Long-Term Prognosis of Patients with Ulcerative Colitis after Cytoapheresis Therapy. PLoS ONE 2015;10:e0131197.

99. Klein A, Mazor Y, Karban A, et al. Early histological findings may predict the clinical phenotype in Crohn's colitis. United European Gastroenterol J 2017;5:694–701.

100. Waljee AK, Lipson R, Wiitala WL, et al. Predicting Hospitalization and Outpatient Corticosteroid Use in Inflammatory Bowel Disease Patients Using Machine Learning. Inflamm Bowel Dis 2017;24:45–53.

101. Peng JC, Ran ZH, Shen J. Seasonal variation in onset and relapse of IBD and a model to predict the frequency of onset, relapse, and severity of IBD based on artificial neural network. Int J Colorectal Dis 2015;30:1267–1273.

102. Jung YS, Kim YH, Lee DH, et al. Automatic patient-adaptive bleeding detection in a capsule endoscopy. In: Karssemeijer N, Giger ML, eds. Lake Buena Vista, FL; 2009:72603T. Available at: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.813793 [Accessed September 17, 2018].

103. Li B, Meng MQ-H. Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. Computers in Biology and Medicine 2009;39:141–147.

104. Pan G, Yan G, Qiu X, et al. Bleeding Detection in Wireless Capsule Endoscopy Based on Probabilistic Neural Network. Journal of Medical Systems 2011;35:1477–1484.

105. Guolan Lv, Guozheng Yan, Zhiwu Wang. Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines. In: 2011 Annual International Conference of the IEEE Engineering in

Medicine and Biology Society. Boston, MA: IEEE; 2011:6643–6646. Available at: http://ieeexplore.ieee.org/document/6091638/ [Accessed September 17, 2018].

106. Fu Y, Zhang W, Mandal M, et al. Computer-Aided Bleeding Detection in WCE Video. IEEE Journal of Biomedical and Health Informatics 2014;18:636–642.

107. Ghosh T, Fattah SA, Shahnaz C, et al. An automatic bleeding detection scheme in wireless capsule endoscopy based on histogram of an RGB-indexed image. 2014:4683–4686.

108. Sainju S, Bui FM, Wahid KA. Automated Bleeding Detection in Capsule Endoscopy Videos Using Statistical Features and Region Growing. Journal of Medical Systems 2014;38. Available at: http://link.springer.com/10.1007/s10916-014-0025-1 [Accessed September 12, 2018].

109. Hassan AR, Haque MA. Computer-aided gastrointestinal hemorrhage detection in wireless capsule endoscopy videos. Comput Methods Programs Biomed 2015;122:341–353.

110. Jia X, Meng MQ-H. A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images. 2016:639–642.

111. Usman MA, Satrya GB, Usman MR, et al. Detection of small colon bleeding in wireless capsule endoscopy videos. Computerized Medical Imaging and Graphics 2016;54:16–26.

112. Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointest Endosc 2018.

113. Das A, Ben-Menachem T, Cooper GS, et al. Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model. Lancet 2003;362:1261–1266.

114. Das A, Ben-Menachem T, Farooq FT, et al. Artificial neural network as a predictive instrument in patients with acute nonvariceal upper gastrointestinal hemorrhage. Gastroenterology 2008;134:65–74.

115. Ayaru L, Ypsilantis P-P, Nanapragasam A, et al. Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting. PLoS ONE 2015;10:e0132485.

116. Sengupta N, Tapper EB. Derivation and Internal Validation of a Clinical Prediction Tool for 30-Day Mortality in Lower Gastrointestinal Bleeding. Am J Med 2017;130:601.e1-601.e8.

117. Wong GL-H, Ma AJ, Deng H, et al. Machine learning model to predict recurrent ulcer bleeding in patients with history of idiopathic gastroduodenal ulcer bleeding. Aliment Pharmacol Ther 2019;49:912–918.

118. Das A, Nguyen CC, Li F, et al. Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. Gastrointest Endosc 2008;67:861–867.

119. Săftoiu A, Vilmann P, Gorunescu F, et al. Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. Clin Gastroenterol Hepatol 2012;10:84-90.e1.

120. Zhu M, Xu C, Yu J, et al. Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images: a diagnostic test. PLoS ONE 2013;8:e63820.

121. Săftoiu A, Vilmann P, Dietrich CF, et al. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). Gastrointest Endosc 2015;82:59–69.

122. Ozkan M, Cakiroglu M, Kocaman O, et al. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. Endosc Ultrasound 2016;5:101–107.

123. Yang Y, Chen H, Wang D, et al. Diagnosis of pancreatic carcinoma based on combined measurement of multiple serum tumor markers using artificial neural network analysis. Chin Med J 2014;127:1891–1896.

124. Piscaglia F, Cucchetti A, Benlloch S, et al. Prediction of significant fibrosis in hepatitis C virus infected liver transplant recipients by artificial neural network analysis of clinical factors. Eur J Gastroenterol Hepatol 2006;18:1255–1261.

125. Hashem S, Esmat G, Elakel W, et al. Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2018;15:861–868.

126. Wang D, Wang Q, Shan F, et al. Identification of the risk for liver fibrosis on CHB patients using an artificial neural network based on routine and serum markers. BMC Infect Dis 2010;10:251.

127. Chen Y, Luo Y, Huang W, et al. Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. Comput Biol Med 2017;89:18–23.

128. Wang K, Lu X, Zhou H, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. Gut 2019;68:729–741.

129. Wei R, Wang J, Wang X, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. EBioMedicine 2018;35:124–132.

130. Raoufy MR, Vahdani P, Alavian SM, et al. A novel method for diagnosing cirrhosis in patients with chronic hepatitis B: artificial neural network approach. J Med Syst 2011;35:121–126.

131. Sowa J-P, Heider D, Bechmann LP, et al. Novel algorithm for non-invasive assessment of fibrosis in NAFLD. PLoS ONE 2013;8:e62439.

132. Sowa J-P, Atmaca Ö, Kahraman A, et al. Non-invasive separation of alcoholic and non-alcoholic liver disease with predictive modeling. PLoS ONE 2014;9:e101444.

133. Yip TC-F, Ma AJ, Wong VW-S, et al. Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. Aliment Pharmacol Ther 2017;46:447–456.

134. Ma H, Xu C-F, Shen Z, et al. Application of Machine Learning Techniques for Clinical Predictive Modeling: A Cross-Sectional Study on Nonalcoholic Fatty Liver Disease in China. Biomed Res Int 2018;2018:4304376.

135. Wu C-C, Yeh W-C, Hsu W-D, et al. Prediction of fatty liver disease using machine learning algorithms. Comput Methods Programs Biomed 2019;170:23–29.

136. Fialoke S, Malarstig A, Miller MR, et al. Application of Machine Learning Methods to Predict Non-Alcoholic Steatohepatitis (NASH) in Non-Alcoholic Fatty Liver (NAFL) Patients. AMIA Annu Symp Proc 2018;2018:430–439.

137. Hong W-D, Ji Y-F, Wang D, et al. Use of artificial neural network to predict esophageal varices in patients with HBV related cirrhosis. Hepat Mon 2011;11:544–547.

138. Dong T, Kalani A, Aby E, et al. Machine Learning-based Development and Validation of a Scoring System for Screening High Risk Esophageal Varices. Clin Gastroenterol Hepatol 2019.

139. Gatos I, Tsantis S, Spiliopoulos S, et al. A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography. Ultrasound Med Biol 2017;43:1797–1810.

140. Pearce CB, Gunn SR, Ahmed A, et al. Machine learning can improve prediction of severity in acute pancreatitis using admission values of APACHE II score and C-reactive protein. Pancreatology 2006;6:123–131.

141. Hong W, Chen X, Jin S, et al. Use of an artificial neural network to predict persistent organ failure in patients with acute pancreatitis. Clinics (Sao Paulo) 2013;68:27–31.

142. Jovanovic P, Salkic NN, Zerem E. Artificial neural network predicts the need for therapeutic ERCP in patients with suspected choledocholithiasis. Gastrointest Endosc 2014;80:260–268.

143. Banerjee R, Das A, Ghoshal UC, et al. Predicting mortality in patients with cirrhosis of liver with application of neural network technology. J Gastroenterol Hepatol 2003;18:1054–1060.

144. Konerman MA, Zhang Y, Zhu J, et al. Improvement of predictive models of risk of disease progression in chronic hepatitis C by incorporating longitudinal data. Hepatology 2015;61:1832–1841.

145. Konerman MA, Lu D, Zhang Y, et al. Assessing risk of fibrosis progression and liver-related clinical outcomes among patients with both early stage and advanced chronic hepatitis C. PLoS ONE 2017;12:e0187344.

146. Khosravi B, Pourahmad S, Bahreini A, et al. Five Years Survival of Patients After Liver Transplantation and Its Effective Factors by Neural Network and Cox Poroportional Hazard Regression Models. Hepat Mon 2015;15:e25164.

147. Briceño J, Cruz-Ramírez M, Prieto M, et al. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. J Hepatol 2014;61:1020–1028.

148. Takayama T, Ebinuma H, Tada S, et al. Prediction of effect of pegylated interferon alpha-2b plus ribavirin combination therapy in patients with chronic hepatitis C infection. PLoS ONE 2011;6:e27223.

149. Heida A, Knol M, Kobold AM, et al. Agreement Between Home-Based Measurement of Stool Calprotectin and ELISA Results for Monitoring Inflammatory Bowel Disease Activity. Clin Gastroenterol Hepatol 2017;15:1742-1749.e2.

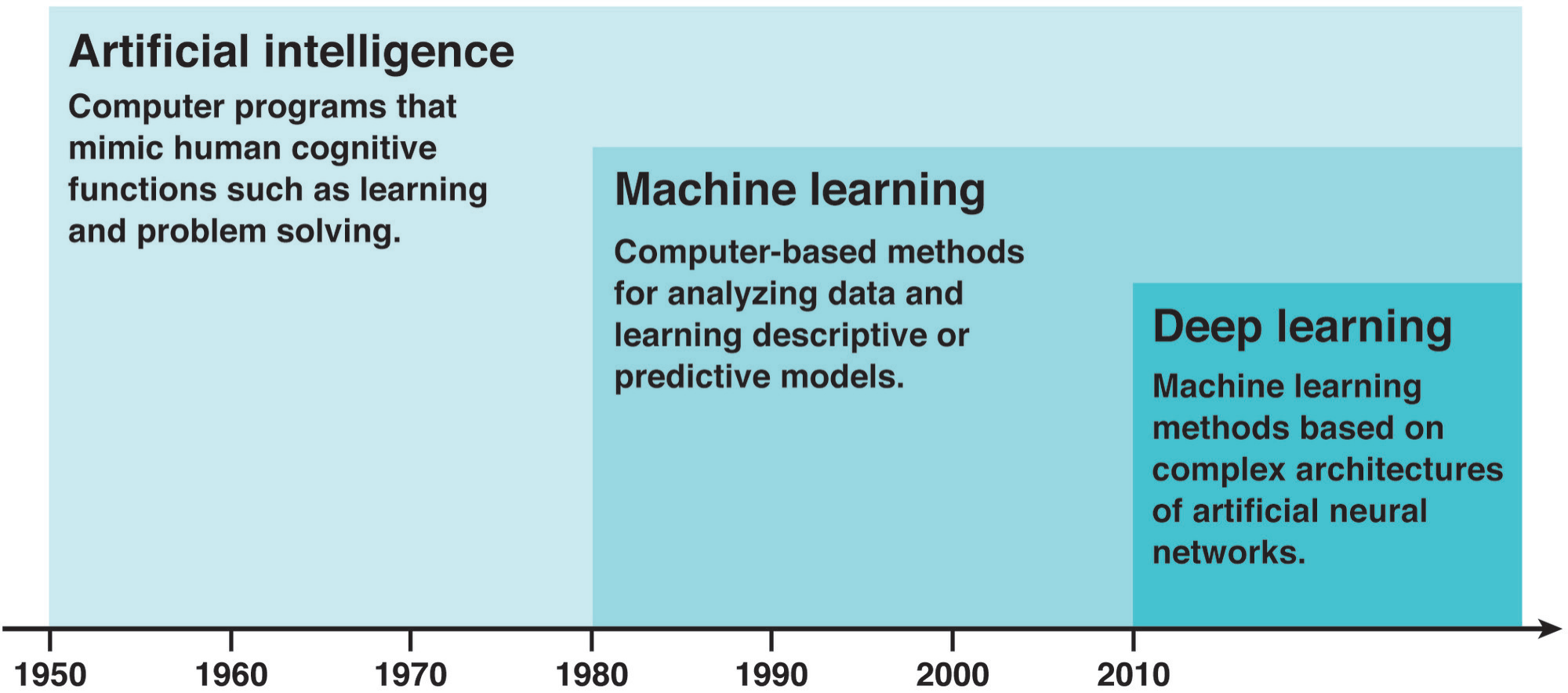
Figure legends

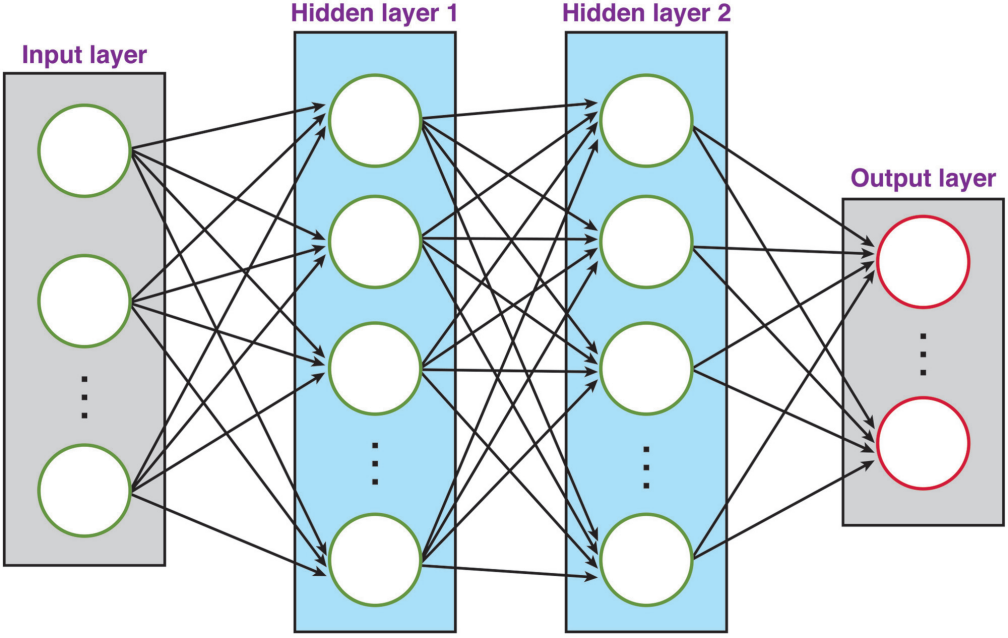
Figure 1. Timeline of Artificial Intelligence main concepts.

Figure 2. Overview of an artificial neural network (ANN) with one input layer, two hidden layers and one output layer. During training on a dataset of input-output pairs, weights of inter-neuron connections are adjusted to optimize classification. Once trained, such ANNs allow to map any new input (represented in the input layer) to a given output (represented in the output layer).

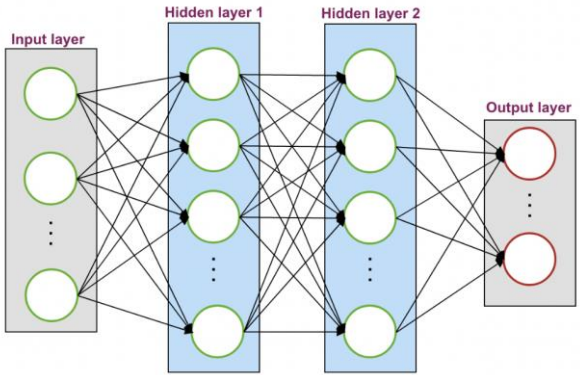
Supplementary Figure 1. Flowchart of study selection and inclusion.

Supplementary Figure 2. Results from most studies were quantified in terms of accuracy (Equation 1), and/or sensitivity (or recall) (Equation 2) and specificity (Equation 3), according to the formulas described below. The area under the receiver operating characteristic curve (AUROC) was adopted in some studies instead of accuracy. Precision (Equation 4) and F1 score (Equation 5) were also used in some studies.


Artificial intelligence
Computer programs that mimic human cognitive functions such as learning and problem solving.
Machine learning
Computer-based methods for analyzing data and learning descriptive or predictive models.
Deep learning
Machine learning methods based on complex architectures of artificial neural networks.
1950
1960
1970
1980
1990
2000
2010

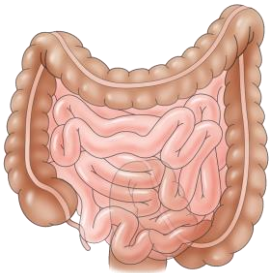

Input layer
Hidden layer 1
Hidden layer 2
Output layer

Artificial intelligence
Neural networks



Literature search
2000-2019
n=138 studies

Gastroenterology (n=107)

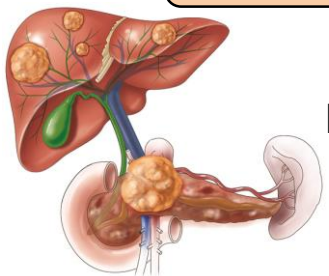


(Pre-)malignant lesions

Inflammatory lesions

Gastrointestinal bleeding

Hepatology (n=31)



Liver / pancreatobiliary disorders

Improving diagnostic accuracy

Establishing a prognosis

Predicting response to treatments

Gastroenterology