Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology

Chrysanthos D Christou, Georgios Tsoulfas

ORCID number: Chrysanthos D Christou 0000-0002-5417-8686; Georgios Tsoulfas 0000-0001-5043-7962.

Chrysanthos D Christou, Georgios Tsoulfas, Organ Transplant Unit, Hippokration General Hospital, Aristotle University of Thessaloniki, Thessaloniki 54622, Greece

Corresponding author: Georgios Tsoulfas, FACS, FICS, MD, Associate Professor, Organ Transplant Unit, Hippokration General Hospital, Aristotle University of Thessaloniki, 66 Tsimiski Street, Thessaloniki 54622, Greece. tsoulfasg@gmail.com

Abstract

Artificial intelligence (AI) is an umbrella term used to describe a cluster of interrelated fields. Machine learning (ML) refers to a model that learns from past data to predict future data. Medicine and particularly gastroenterology and hepatology, are data-rich fields with extensive data repositories, and therefore fruitful ground for AI/ML-based software applications. In this study, we comprehensively review the current applications of AI/ML-based models in these fields and the opportunities that arise from their application. Specifically, we refer to the applications of AI/ML-based models in prevention, diagnosis, management, and prognosis of gastrointestinal bleeding, inflammatory bowel diseases, gastrointestinal premalignant and malignant lesions, other nonmalignant gastrointestinal lesions and diseases, hepatitis B and C infection, chronic liver diseases, hepatocellular carcinoma, cholangiocarcinoma, and primary sclerosing cholangitis. At the same time, we identify the major challenges that restrain the widespread use of these models in healthcare in an effort to explore ways to overcome them. Notably, we elaborate on the concerns regarding intrinsic biases, data protection, cybersecurity, intellectual property, liability, ethical challenges, and transparency. Even at a slower pace than anticipated, AI is infiltrating the healthcare industry. AI in healthcare will become a reality, and every physician will have to engage with it by necessity.

Key Words: Artificial intelligence; Machine learning; Gastroenterology; Hepatology; Artificial neural networks; Support vector machine

Core Tip: The opportunities that arise from the application of artificial intelligence/machine learning-based models in gastroenterology and hepatology include the

establishment of targeted screening programs through the identification of patients prone to develop cancer, the development of non-invasive diagnostic tools, the improvement of the diagnostic accuracy, the development of treatment allocation frameworks based on predictions of outcomes for different treatment modalities, the development of models to ensure cost-effective use of resources, the development of triage tools for higher levels of care and decision-making tools for further treatment, based on individualized patient outcome predictions, and finally the development of predictive models of prognosis for patient and family counseling.

INTRODUCTION

Artificial intelligence (AI) is an umbrella term to describe any application where computer systems are used to perform tasks normally associated with human intelligence[1,2]. AI is a cluster of interrelated fields, including machine learning (ML) and probabilistic reasoning, planning and decision making, fuzzy systems, computer vision, natural language processing, knowledge representation, and neural networks (NN)[1]. Despite their differences, these fields are all driven by advancements in computing power and Big Data.

ML could be described as a model that learns from past data in order to predict future data[3]. The application of ML in healthcare was catalyzed by several healthcare trends, such as electronic healthcare records and patient summaries, genomic analyses and biomedical research, routine imaging, and telemedicine, that have transformed the healthcare industry into a data-rich science with data as an omnipresent concept[4-6]. Specifically, while in 2013, the healthcare industry produced 153 exabytes (10^{18} gigabytes) of data, it has been projected to reach 2314 exabytes in 2020[7]. Therefore, the healthcare industry generates an enormous amount of data that conform with the features that define Big Data: Volume, high-velocity, high-variety, and veracity, which cannot be analyzed or managed by traditional software[5,8,9]. AI promises to process and analyze these extensive repositories of data and turn them into meaningful insights.

Several studies have described AI as the potential solution to long-standing healthcare challenges such as increasing diagnostic accuracy, enhancing telemedicine, providing substantial cost reduction, promoting evidence-based medicine, facilitating targeted literature search, and delivering individualized care[10-14]. More importantly, AI could substantially alleviate the burden of diseases by reducing the associated mortality and morbidity through optimizing patient outcomes[15,16].

In gastroenterology and hepatology, physicians handle large amounts of clinical data and an extensive repository of imaging data generated from endoscopy, ultrasound, and computed tomography (CT). Therefore gastroenterology and hepatology are data-rich fields, and thus fruitful ground for AI applications with extensive research conducted in these fields, particularly regarding the prevention, diagnosis, management, and prognosis of diseases[17,18]. We aim to comprehensively review the applications of AI in gastroenterology and hepatology and identify the current challenges of utilizing AI in healthcare in an effort to explore ways to overcome them.

SEARCH STRATEGY

We conducted a comprehensive literature review of the Medline, Cochrane, and Scopus databases using the following algorithm: [(artificial intelligence OR machine learning OR deep learning OR neural networks OR support vector machine OR computer-assisted OR computer-aided) AND (gastroenterology OR hepatology OR

esophageal OR small bowel OR large bowel OR gastric cancer OR capsule endoscopy OR polyps OR colonoscopy OR colorectal cancer OR gastrointestinal bleeding OR inflammatory bowel disease OR Crohn's disease OR celiac disease OR ulcers OR hepatocellular carcinoma OR cholangiocarcinoma OR liver fibrosis OR fatty liver OR chronic liver disease OR cirrhosis)]. Articles were reviewed for eligibility by the two authors independently (CC, GT), and disagreements were solved by a discussion between the two authors. Finally, the reference lists of eligible articles were reviewed to identify further related literature, including articles, books, and other forms of publication. We excluded studies written in a language other than English and publications of abstracts. The review of the literature was completed on January 27, 2021.

AI CLASSIFIERS

Before discussing AI's current applications and challenges in gastroenterology and hepatology, we briefly describe the AI classifiers in ML models that we identified as the most commonly used among the eligible articles. Specifically, we describe the basic features of the support vector machine (SVM), the artificial NN (ANN), and the convolutional NN (CNN). ML is divided into supervised and unsupervised based on whether the training data is labeled or not[19]. In other words, through supervised ML, a new data set is classified for an outcome based on previous data sets that trained the ML model, while in unsupervised ML, there is no outcome but rather an attempt to detect unknown patterns and correlations within the data[20].

SVMs are supervised learning models with associated learning algorithms trained to assign classes to new cases. SVMs require a training set of data where each case is pre-labeled regarding the outcome classification. The data points of the features of each case are treated as points in a high-dimensional space[21]. Based on these mapped points, separating hyperplanes are drawn, aiming to distinguish these points upon their labeled class[22]. The hyperplane with the maximum distance from the mapped data points, known as the maximum functional margin, is selected since it provides the SMV's full potential to classify new examples correctly[21,22]. SVMs typically perform linear classification analysis. For non-linear analysis, kernel function could introduce additional dimensions to the raw data; and thus turning a non-linear problem into a linear problem in a higher-dimensional space[23].

An ANN is an ML model inspired by the human brain's neuronal connections that consist of an input layer, an output layer, and a hidden layer between them[18]. ANNs are applied both in supervised and unsupervised ML[24]. When multiple hidden layers are inserted between the input and output layers, and the network's architecture becomes more complex with multiple interconnections, the concept of deep NN (DNN) emerges[24,25]. During DNN training, the model adjusts the weighted correlations between the nodes of the input layer and the nodes of the multiple hidden layers[18,20]. Within DNN, CNNs are biologically variants inspired by multi-layer perceptrons[26]. CNNs have been extensively applied for medical image analysis[17, 18]. During the CNN model development, the images are preprocessed using multiple filters, and multiple feature maps are created in a process called convolution[26].

APPLICATIONS OF AI IN GASTROENTEROLOGY

AI/ML-based models have been extensively applied in the prevention, diagnostics, management, and prognosis of gastrointestinal (GI) diseases, including GI bleeding (GIB), inflammatory bowel diseases (IBDs), malignant and premalignant lesions, and other nonmalignant lesions/diseases such as gastroesophageal reflux, *Helicobacter pylori* (*H. pylori*) infection, ulcers, celiac disease, and intestinal hookworms.

Prevention

Table 1 summarizes the findings of the identified studies applying AI/ML models in the prevention of gastroenterological diseases. AI/ML-based software could be used in screening programs to identify high-risk patients not currently identified by standard screening guidelines. In a recent study, researchers developed seven different AI/ML models to identify patients at high risk of developing gastric cancer following the eradication of *H. pylori* infection[27]. The extreme gradient boosting (GB) model demonstrated the highest performance[27]. A CNN was developed in a different study to classify patients at low, moderate, and high risk of developing gastric cancer using

Table 1 Artificial intelligence applications in gastroenterology: Prevention

Ref.	Parameters employed	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Leung *et al* [27]	Laboratory results, clinicopathological parameters	Several	64238/25330 patients	Risk of gastric cancer development following *H.pylori* eradication	0.53-0.97[2,6], 59.3-98.1[3,6], 51.5-93.6[4,6]
Nakahira *et al*[28]	Laboratory results, clinicopathological parameters, endoscopic images	CNN	7826/454 patients	Stratify risk of gastric cancer development	---
Taninaga *et al*[29]	Laboratory results, clinicopathological parameters, endoscopic images	CART	1144/287	Prediction of future gastric cancer	63.4-94.8[1,6], 0.736-0.874[2,6]
Goshen *et al*[31]	Laboratory results, clinicopathological parameters	DT, RF, GB	688 flagged patients	High risk of CRC development	----

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. CART: Classification and regression tree; CNN: Convolutional neural network; CRC: Colorectal cancer; DT: Decision tree; GB: Gradient boosting; RF: Random forest.

endoscopic images[28]. A similar study used multiple classification and regression trees (CART) to develop a model that predicts future gastric cancer development[29]. Another notable example is the ColonFlag test, an ML algorithm that uses age, sex, and complete blood count data to identify patients at high risk of developing colorectal cancer[30,31]. In a recent study, among 254 individuals who underwent colonoscopy, 19 cancers (7.5%) and 22 (8.7%) advanced adenomas were identified in a population of patients who would otherwise have avoided screening[31]. These studies demonstrate how AI/ML-based models could be used in a real clinical setting to establish targeted screening programs.

Diagnostics

Table 2 summarizes the findings of the identified studies applying AI/ML models to diagnose gastroenterological diseases. AI/ML-based software is applied in Computed Aided Detection or Diagnosis (CAD) systems used in radiology to augment medical images' interpretation accuracy. A CAD typically includes the stages of preprocessing, extraction of features, and feature selection and classification[32]. CAD systems could also aid endoscopists in navigating through the different anatomical locations of the GI tract. Specifically, a model based on a CNN employed esophagogastroduodenoscopy images to recognize the anatomical locations with outstanding performance[33].

Regarding the diagnosis of gastroesophageal reflux, an ANN model was developed as a non-invasive diagnostic tool by employing only clinical data[34]. For patients with Barrett's esophagus, a deep learning-based (DL) CAD system was developed to differentiate patients with malignant from patients with nondysplastic Barret's esophagus [35]. The CAD system also identified the optimal site to perform a biopsy with an accuracy between 92% and 97%[35]. An SVM model was developed in another study employing white-light endoscopic imaging to identify early neoplastic lesions[36]. In a recent study, images from volumetric laser endomicroscopy were used to develop a CAD system to detect neoplastic lesions for patients with Barret's esophagus[37]. A similar study using volumetric laser endomicroscopy developed an ML model for detecting neoplastic lesions[38]. Notably, the model outperformed the experts in volumetric laser endomicroscopy[38]. Finally, to diagnose esophageal squamous cell carcinoma, a CNN model was developed by employing endocytoscopic images as an alternative to biopsy[39].

For diagnosing patients with *H. pylori* infection, a CNN was developed in a study employing gastroscopic images[40]. A different CNN for *H. pylori* diagnosis was developed in a prospective pilot study analyzing images taken from the stomach's lesser curvature using either white light imaging, blue light imaging, or linked color imaging[41]. Notably, employing blue light and linked color imaging yielded significantly higher performance than white light imaging[41]. In a similar study, a

Table 2 Artificial intelligence applications in gastroenterology: Diagnosis

Ref.	Diagnostic Modality	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Takiyama et al [33]	Esophago-gastro-duodenoscopy imaging	CNN	1750/435[7]	Anatomical classification among larynx, esophagus, stomach, and duodenum	0.99-1.00[2,7]
Pace et al[34]	Laboratory results, clinicopathological parameters	ANN	159 patients	Diagnosis of gastroesophageal reflux disease	67.86-100[1,6]
de Groof et al [35]	Esophageal endoscopic images	DNN	1247/297[6]/80[7]/80[7] patients	Classification of malignant from nondysplastic Barret's esophagus	88.2[1,6], 87.5-88.8[1,7], 87.6[3,6], 90.0-92.5[3,7], 88.6[4,6], 82.5-87.5[4,7]
van der Sommen et al[36]	White-light endoscopic imaging	SVM	44 patients with Barret's esophagus	Diagnosis of early neoplastic lesions	Per image: 62-90[3,6], 65-90[4,6], Per patient: 52-100[3,6], 74-96[4,6]
Struyvenberg et al[37]	Volumetric laser endomicroscopy imaging	Several	29 patients with Barret's esophagus	Diagnosis of neoplastic lesions	0.83-0.94[2,6]
Swager et al[38]	Volumetric laser endomicroscopy imaging	Several	60 images	Diagnosis of neoplastic lesions	0.89-0.95[2,6]
Kumagai et al [39]	Endocytoscopic imaging	CNN	4715/1520[7]	Diagnosis of esophageal squamous cell carcinoma	90.9[1,7], 0.72-0.90[2,7], 39.4-46.4[3,7], 98.2-98.4[4,7]
Zheng et al[40]	Endoscopic images	CNN	1507/452 patients	Diagnosis of H.pylori infection	84.5-93.8[1,6], 0.93-0.97[2,6], 81.4-91.6[3,6], 90.1-98.6[4,6]
Nakashima et al [41]	Endoscopic images	CNN	162/60 patients	Diagnosis of H.pylori infection	0.66-0.96[2,6]
Itoh et al[42]	Endoscopic images	CNN	149/30 images	Diagnosis of H.pylori infection	0.956[2,6], 86.7[3,6], 86.7[4,6]
Shichijo et al[43]	Endoscopic images	CNN	32308/11481[7]	Diagnosis of H.pylori infection	83.1-87.7[1,7], 81.9-88.9[3,7], 83.4-87.4[4,7]
Kanesaka et al [45]	NBI	SVM	126/81 NBI images	Diagnosis of gastric cancer	96.3[1,6], 96.7[3,6], 95.0[4,6]
Hirasawa et al [46]	Endoscopic images	CNN	13584/2296[7]	Diagnosis of gastric cancer	92.2[3,7]
Zhu et al[47]	Laboratory results, clinicopathological parameters, cancer biomarkers	GB/DT	496/213 patients	Diagnosis of gastric cancer	85.9[1,5], 83[1,6], 0.91[2,6], 88[3,5], 87[3,6], 83.4[4,5], 84.1[4,6]
Tenório et al[48]	Laboratory results, clinicopathological parameters	Several	178/38	Diagnosis of celiac disease	71.5-80[1,6], 0.71-0.84[2,6], 69-82[3,6], 67-80[4,6]
Caetano Dos Santos et al[49]	Endomysial autoantibody test for IgA-class antibodies images	SVM	2597 images (training:validation = 7:3)	Diagnosis of celiac disease	96.8-98.85[1,6], 82.84-98.91[3,6], 98.81-99.40[4,6]
Hujoel et al[50]	Laboratory results, clinicopathological parameters	Several	408 undiagnosed patients	Diagnosis of celiac disease	0.49-0.53[2,6]
Manandhar et al [51]	Gut microbiome data	RF	1429 fecal 16S metagenomic data subjects	Diagnosis of IBD	0.80-0.82[2,6]
Wei et al[52]	Single nucleotide polymorphisms data	Several	60828 samples	Classification of CD and UC	0.782-0.866[2,6]
Mossotto et al [53]	Capsule endoscopy, histologic imaging	SVM	239/48[7] pediatric patients	Classification of CD, UC, and unclassified IBD	71-82.7[1,5], 0.78-0.87[2,5], 83.3[1,7], 83-85[3,7]
Xia et al[58]	Capsule endoscopy imaging	CNN	697/100[7] patients, 822590/201365[7], images	Classification among different types of gastric lesions	77.1-86[1,7], 0.80-0.90[2,7], 96.2-100[3,7], 56.5-76.2[4,7]
Seguí et al[59]	Capsule endoscopy imaging	CNN	50 videos	Classification of small bowel mobility events	96[1,6]
Park et al[60]	Capsule endoscopy imaging	CNN	139 videos, 200000 images (training:validation:test = 6:2:2)	Small bowel lesion identification	80.29-98.34[1,6], 0.999[2,5], 0.998[2,6,7]

Hwang *et al*[61]	Capsule endoscopy imaging	CNN	7556/5760[7] images	Classification of hemorrhagic and ulcerative lesions	96.62-96.83[1,7], 95.07-97.61[3,7], 96.04-98.18[4,7]
Otani *et al*[62]	Capsule endoscopy imaging	DNN	167/40[7] patients	Classification among different types of small bowel lesions	0.950-0.996[2,6], 0.884-0.928[2,7]
Yuan *et al*[63]	Capsule endoscopy imaging	SVM	20 patients, 340 images (training:validation = 8:2)	Diagnosis of peptic ulcers	92.65[1,6], 94.12[3,6], 91.18[4,6]
Karargyris *et al*[64]	Capsule endoscopy imaging	SVM	80 frames	Diagnosis of peptic ulcers	75[3,6], 73.3[4,6]
He *et al*[65]	Capsule endoscopy imaging	CNN	11 patients, 440000 images	Diagnosis of intestinal hookworms	88.5[1,6], 0.895[2,6], 84.6[3,6], 88.6[4,6]
Leenhardt *et al*[66]	Capsule endoscopy imaging	CNN	600/600 images	Diagnosis of gastrointestinal angiectasia	100[3,6], 96[4,6]
Zhou *et al*[67]	Capsule endoscopy imaging	CNN	21 videos	Diagnosis of celiac disease	100[3,6], 100[4,6]
Yamada *et al*[68]	Colon capsule endoscopy imaging	CNN	15933/4784[7]	Diagnosis of colorectal neoplasias	83.9[7], 0.902[2,7], 79[3,7], 87[4,7]
Wang *et al*[69]	Colonoscopy imaging	CNN	5545 images/27113[7] images/612[7] images/138[7] videos/54[7] videos	Identification of colorectal polyps	0.984[2,7], 88.24-100[3,7], 95.40-95.92[2,7]
Misawa *et al*[70]	Colonoscopy imaging	CNN	411/35 short videos	Identification of colorectal polyps	76.5[1,6], 0.87[2,6], 90[3,6], 63.3[4,6]
Urban G *et al*[71]	Colonoscopy imaging	CNN	8641 images/20[7] videos	Identification of colorectal polyps	96.4[1,7], 0.991[2,7]
Ozawa *et al*[72]	Colonoscopy imaging	CNN	20431/7077[7] images	Identification of colorectal polyps, Classification of colorectal polyps	90-97[3,7], 47-98[1,7]
Mori *et al*[73]	NBI and methylene blue staining images	SVM	466 diminutive polyps	Classification of diminutive rectosigmoid adenomas	NPV(%): 93.7-96.5
Tischendorf *et al*[74]	NBI	SVM	209 colorectal polyps	Classification of colorectal polyps	90[3,6], 70.2[4,6]
Gross *et al*[75]	NBI	SVM	434 colorectal polyps	Classification of small colorectal polyps	93.1[1,6], 95.0[3,6], 90.3[4,6]
Kominami *et al*[76]	NBI	SVM	118 colorectal polyps	Classification of colorectal polyps	93.2[1,6], 93.0[3,6], 93.3[4,6]
Misawa *et al*[77]	NBI endocytoscopy	SVM	979/100 endocytoscopy, images	Classification of colorectal polyps	90[1,6], 84.5[3,6], 97.6[4,6]
Takeda *et al*[78]	NBI endocytoscopy	SVM	5543/200 endocytoscopy, images	Diagnosis of invasive CRC	94.1[1,6], 89.4[3,6], 98.9[4,6]
Chen *et al*[79]	NBI	CNN	2157/284[7]	Classification neoplastic from hyperplastic polyps	96.3[3,7], 78.1[4,7], NPV(%): 91.5[7]
Komeda *et al*[80]	NBI	CNN	1200/600 images	Classification of adenomatous from non-adenomatous polyps	75.1[1,6]
Byrne *et al*[81]	NBI	CNN	223/40[7] videos	Classification of adenomas from hyperplastic polyps	94[1,7], 0.95[2,7], 98[3,7], 83[4,7], NPV(%): 97[7]

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. ANN: Artificial neural network; CD: Chron's disease; CNN: Convolutional neural network; CRC: Colorectal cancer; DNN: Deep neural network; DT: Decision tree; GB: Gradient boosting; IBD: Inflammatory bowel disease; NBI: Narrow-band imaging; NPV: Negative predictive value; RF: Random forest; SVM: Support vector machine; UC: Ulcerative colitis.

CNN model employed images from upper GI endoscopy to detect *H. pylori* infection

[42]. Finally, Shichijo *et al*[43] developed two different CNN models to diagnose *H. pylori* infection. The first, a 22-layered deep CNN, performed comparably with endoscopists, while the second, which classified images based on their location in the stomach, performed comparably with endoscopists in terms of sensitivity and specificity but demonstrated a significantly higher accuracy[43]. The time needed to analyze all the images was 3 min and 18 s and 3 min and 14 s for the two CNNs, while for endoscopists, the average time needed was 230.1 min[43]. A recent meta-analysis evaluating the performance of CNNs to diagnose infection with *H. pylori* concluded that it is currently equivalent to physicians[44].

Regarding the diagnosis of gastric cancer, an SVM-based CAD model was developed to identify early gastric cancer features in narrow-band imaging (NBI) gastroscopy[45]. In a different retrospective study, a CNN was developed to detect gastric cancer from endoscopic images. Even though the model managed to correctly classify 71 of the 77 gastric cancer lesions (92.2% sensitivity), it also falsely identified 161 non-cancerous lesions as gastric cancer, yielding a low positive predictive value of 30.6%[46]. Approximately half of these lesions were gastritis with an irregular mucosal surface or changes in color tone[46]. Finally, in a recent study, a non-invasive GB/ decision tree (DT) model employing only nonendoscopic parameters was developed to diagnose gastric cancer[47].

Regarding the diagnosis of celiac disease, five different AI/ML-based models were developed in a study using clinical data as a base for non-invasively diagnosing celiac disease[48]. Interestingly, the models were tested using 13 different algorithms and different input variables variations, resulting in 270 different tested models[48]. Among them, a model based on the Bayesian classifier demonstrated the highest performance[48]. A different study developed an SVM that employed images from the endomysial autoantibody test for IgA-class antibodies to classify patients with celiac disease[49]. Finally, in a different study, the authors aimed to develop several AI/ML-based models, which employ clinical data, to predict celiac disease in a group of patients who remained undiagnosed[50]. The models were unsuccessful, with only two models slightly outperforming random chance in predicting celiac disease[50].

In the diagnosis of IBDs, a recent study developed five different AI/ML-based models using data from the gut microbiome to classify patients with IBD[51]. The Random Forest (RF) model demonstrated the highest performance with AUROCs of 0.80 when bacterial taxa were used and 0.82 when operational taxonomic features were used[51]. Finally, the models were tested in distinguishing between Chron's disease (CD) and ulcerative colitis (UC), with the RF model demonstrating an AUROC > 0.9[51]. A multicentered, genome-wide association study used data from single-nucleotide polymorphisms to develop several AI/ML-based models to classify patients with CD and UC[52]. For pediatric IBDs, an SVM-based model employed data from histologic and endoscopic images to classify pediatric patients with CD, UC, or unclassified IBD[53]. Other studies have employed images from wireless capsule endoscopy to build SVM models that identify patients with CD with reported accuracies between 80.2% and 100%[54-57].

Focusing on capsule endoscopy, in a recent study, magnetically controlled capsule endoscopy imaging was employed to develop a CNN that provides an automatic detection and classification system for gastric lesions[58]. These lesions included erosions, polyps, ulcers, submucosal tumors, normal mucosa, and xanthomas[58]. A different study used images generated by a wireless capsule to develop a CNN able to classify the small bowel motility among six distinct intestinal motility events[59]. In a recent study, an AI-assisted capsule endoscopy reading model was developed to assist lesion identification[60]. Notably, the model significantly shortened the reading time of images by trainees[60].

To distinguish between hemorrhagic and ulcerative lesions, a CNN was developed using images from small bowel capsule endoscopy[61]. In a similar, recent study, a DNN model used capsule endoscopy images to classify different small bowel lesions (erosions, ulcers, tumors, and vascular lesions)[62]. SVM-based models were developed in two other studies employing images from capsule endoscopy to identify peptic ulcers[63,64]. In three different studies, CNNs have been developed, employing images from capsule endoscopy to diagnose intestinal hookworms, GI angiectasia, and celiac disease[65-67]. Finally, a CNN was developed in a recent study to detect colorectal neoplasias in images from colon capsule endoscopy[68].

Regarding the identification of colorectal polyps in coloscopy, a CNN model was designed to analyze colonoscopy images and videos and identify colorectal polyps [69]. Interestingly, the model's sensitivity to identify flat isochromatic, less than 0.5 cm polyps, which are typically associated with a higher missing rate, was estimated at 91.65% per image[69]. A different study, employing short videos from colonoscopies,

developed a CNN model that identifies colorectal polyps[70]. In another study, a CNN was developed employing videos and images from coloscopy to detect colorectal polyps in real-time[71]. Interestingly, when experts reviewed the colonoscopy videos, they managed to identify additional non-removed polyps with the assistance of the CNN model[71]. In a recent study, an automated polyp detection model based on CNN was developed to identify and then classify colorectal polyps into adenomas, hyperplastic, sessile serrated adenomas, and cancer[72]. The model managed to process the images at a speed of 20 m per frame[72].

Specifically for colorectal polyps classification, in a prospective study, an SVM-based model was developed using NBI and methylene blue staining images to classify diminutive rectosigmoid adenomas in real-time[73]. A prospective pilot study developed an SVM model that employed NBI images to detect and classify colorectal polyps based on vascularization features as neoplastic and non-neoplastic[74]. In a comparative study, an SVM model demonstrated comparable performance at classifying the neoplastic nature of diminutive polyps (< 10 mm) to that of experts in NBI colonoscopy but surpassed the performance of non-experts[75]. Two other studies developed SVM models that employed NBI images to classify colorectal lesions (neoplastic or non-neoplastic)[76,77]. Finally, an SVM model was developed in a retrospective study to diagnose invasive colorectal cancer based on NBI endocyt-oscopy images[78].

Except for SVMs, several studies employed CNNs for colorectal polyps classi-fication. One comparative study developed a CNN employing NBI colonoscopy images to classify neoplastic polyps[79]. Notably, the model's performance was found comparable to that of experts but superior to the performance of non-experts[79]. In another study focusing on NBI colonoscopy, a CNN model was developed to classify adenomatous from non-adenomatous polyps[80]. A different study developed a CNN model to classify adenomas from hyperplastic polyps during NBI colonoscopy[81].

The performances reported by the majority of the AI/ML-based models surpass both the NPV threshold recommended by the American Society of Gastrointestinal Endoscopy (90%) for adenoma detection and the estimated pooled NPV reported in a meta-analysis conducted by the society (91%)[82,83]. Finally, we should mention that currently, the majority of the CAD systems that we reported have the shortcoming of manual segmentations of lesions. The endoscopists should identify the areas of interest before the model could analyze and attempt to classify. This weakness has been acknowledged by the European Society of Gastrointestinal Endoscopy[84]. Other obstacles to developing CAD systems constitute the lack of large datasets and the lack of variability in images. A recent study aimed to resolve this by developing a CNN that "adds" polyps to the images to increase the repository of images for training and advance the development of automated polyp detection models[85].

Management

Table 3 summarizes the findings of the identified studies applying AI/ML models for the management of gastroenterological diseases. In a recent study, data from baseline impedance, nocturnal baseline impedance, and acid exposure time were used as a base for a DT model to predict the treatment response with proton pump inhibitors for patients with gastroesophageal reflux disease[86]. The aim was to establish a decision-making framework for treatment allocation[86].

AI/ML-based models have been used in the management of malignant GI lesions. Regarding gastric cancer, a CNN has been developed to predict whether the early gastric cancer has invaded the mucosa and submucosa layers of the stomach and act as a decision tool for endoscopic resection[87]. Interestingly, the CNN model outper-formed endoscopists[87]. In the same concept, a DNN was developed to classify gastric cancer based on invasion depth as a basis for treatment allocation[88]. Specifically, the model demonstrated accuracy for predicting T1, T1a, T1b T2, T3, T4, and an overall accuracy of 77.2%, 68.9%, 63.6%, 49.1%, 51.0%, 55.3%, and 64.7%, respectively[88]. In colorectal cancer, the identification of microsatellite instability significantly impacts the treatment allocation process. A DL model was developed to identify microsatellite instability directly from hematoxylin and eosin-stained whole slide image[89]. Notably, the model outperformed a group of pathologists[89]. Finally, an SVM model that predicted the lymph node metastasis status of patients with colorectal cancer was designed as a tool to identify patients who would benefit from additional treatment following the endoscopic resection of T1 tumors[90]. Notably, the model outperformed the staging systems endorsed by current guidelines[90].

Regarding the management of GIB, in a recent study, an ML model was developed using data from patients admitted to the intensive care unit (ICU) following a GIB to predict the need for transfusion[91]. The authors suggested that it could potentially be

Table 3 Artificial intelligence applications in gastroenterology: Treatment

Ref.	Parameters employed	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Rogers *et al* [86]	Data from baseline impedance, nocturnal baseline impedance, and acid exposure time	DT	335 patients	Prediction of treatment response with proton pump inhibitors for patients with gastroesophageal reflux disease	0.31-0.938[2,6]
Zhu *et al* [87]	Endoscopic images	CNN	790/203[7] images	Invasion of gastric cancer at the mucosa and submucosa layers of the stomach	89.16[1,7], 0.94[2,7], 76.47[3,7], 95.56[4,7]
Kubota *et al* [88]	Endoscopic images	DNN	800/90 images	Invasion depth of gastric cancer	64.7[1,6]
Yamashita *et al*[89]	Hematoxylin and eosin-stained WSI	DNN	100/15[6]/484[7]	Identificication of CRC microsatellite instability	0.931[2,6], 0.779[2,7], 76[3,7], 66.6[4,7]
Ichimasa *et al*[90]	Laboratory results, clinicopathological parameters	SVM	590/100[7]	Prediction of lymph node metastasis status	69[1,7], 0.821[2,7], 100[3,7], 66[4,7]
Levi *et al* [91]	Laboratory results, clinicopathological parameters	RFE	14620 patients	Prediction of the need for transfusion following GIB	50.21-74.88[1,6], 0.7858-0.8141[2,6], 69.17-92.77[3,6], 35.02-79.82[4,6]
Chu *et al* [92]	Laboratory results, clinicopathological parameters	Several	122/67 patients	Prediction of the source of GIB	69.7-94.3[1,6], 0.658-0.999[2,6], 90.1-98.0[3,6], 89-100[4,6]
				Prediction of the need for blood resuscitatio	64.7-94.1[1,6], 0.381-0.993[2,6], 90.3-93.9[3,6], 18.4-95.5[4,6]
				Prediction of the need for emergent endoscopy	62.7-83.3[1,6], 0.404-0.913[2,6], 80.1-89.1[3,6], 13.8-85.7[4,6]
				Prediction of disposition	58.4-89.7[1,6], 0.324-0.972[2,6], 81.9-92.9[3,6], 18.4-90.9[4,6]
Das *et al*[93]	Laboratory results, clinicopathological parameters	ANN	194/193[6]/200[7] patients	Prediction of major stigmata of recent hemorrhage	89[1,3,4,6], 77[1,7], 96[3,7], 63[4,7]
				Prediction of the need for emergent endoscopy	81[1,3,6], 61[1,7], 94[3,6], 82[4,6], 48[4,7]
Augustin *et al*[94]	Laboratory results, clinicopathological parameters	CART	164/103[7] patients	Stratification of risk of rebleeding and mortality following acute variceal hemorrhage	0.81-0.83[2,7]
Loftus *et al* [95]	Laboratory results, clinicopathological parameters	ANN	103/44 patients	Prediction of severe lower GIB	0.979[2]
				Prediction of the need for surgical intervention	0.954[2,6]
Ayaru *et al* [96]	Laboratory results, clinicopathological parameters	GB	170/130[7]	Prediction of severe lower GIB	78[1,6], 83[1,7]
				Prediction of recurrent bleeding	88[1,6], 88[1,7]
				Prediction of the need for intervention	88[1,6], 91[1,7]

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. ANN: Artificial neural network; CART: Classification and regression tree; CNN: Convolutional neural network; CRC: Colorectal cancer; DT: Decision tree; DNN: Deep neural network; GB: Gradient boosting; GIB: Gastrointestinal bleeding; RFE: Recursive feature elimination; WSI: Whole-slide image.

used as a decision-making tool to triage patients to the ICU or the ward[91]. A different study developed several AI/ML models for patients with acute GIB to predict four different outputs: The source of bleeding, the need for urgent blood transfusion, the need for urgent endoscopy, and the disposition[92]. The study aimed to establish a decision-making framework for the efficient management of GIB. The RF model outperformed all seven other models for all four outputs[92].

Focusing on the upper GIB, an ANN model was developed as a non-invasive triage tool for patients with upper GIB[93]. In the external cohort, the ANN performed similarly to the complete Rockall Score (includes endoscopic variables) in predicting stigmata of recent hemorrhage[93]. A different study developed a CART model regarding acute variceal hemorrhage to predict rebleeding and mortality and achieved the discrimination of three distinct prognostic groups of low, intermediate, and high risk[94]. Its performance was significantly superior to Child-Pugh and Model for End-Stage Disease (MELD) but comparable to a conventional logistic regression (LR) model [94].

Focusing on the lower GIB, two different ANN models were developed in a study aiming to predict severe acute lower GIB and the need for surgical intervention[95]. The first ANN significantly outperformed the Strate prediction rule for predicting severe bleeding (AUROCs: 0.98 *vs* 0.66)[95]. A different study employed nonendoscopic variables to develop a model based on a GB classifier to predict severe lower GIB, recurrent bleeding, and the need for clinical intervention[96]. On external validation, the model was found equally accurate to a conventional LR model for recurrent bleeding and the need for clinical intervention but superior in predicting severe lower GIB[96].

Prognosis

Table 4 summarizes the findings of the identified studies applying AI/ML models regarding the prognosis of gastroenterological diseases. Several SVM-based nomograms were developed in a study to predict distant metastasis for operated patients with oesophageal squamous cell carcinoma[97]. A different study employed clinicopathological data to develop an ANN model to predict the survival of patients with esophageal cancer operated with curative intends[98]. The model surpassed the Tumor–Node–Metastasis (TNM) model[98]. Regarding gastric cancer, a recent study developed five different AI/ML-based models to predict recurrence in operated patients[99]. Among the models, the RF demonstrated the best performance[99].

An ANN model was developed for patients with IBD, which used meteorological data to predict seasonal variations of onset and relapse in patients with CD and UC [100]. In the validation cohort, the model predicted the onset frequency and the frequency of relapse of the IBD with a mean absolute percentage error of 37.58% and 17.1%, respectively[100]. A study focusing on CD developed an ANN model to predict mucosal remission for patients treated with azathioprine 16 wk following treatment [101]. In a study focusing on UC, an ANN was developed employing clinical data to predict the patients with UC treated with cytoapheresis, who will eventually require operation[102].

Regarding the prognosis of GIB, a CART model was developed to predict in-hospital mortality of cirrhotic patients presenting with upper GIB[103]. In a multi-centered study, an ANN model was developed employing pre-endoscopic variables to predict 30-d mortality in patients with non-variceal upper GIB[104]. Similarly, a prospective, multicentered study employed pre-endoscopic variables to develop an ANN model that predicts 30-d mortality in patients with non-variceal upper GIB[105]. The ANN significantly outperformed the Rockall scoring system (AUROCs: 0.95 *vs* 0.67)[105].

Regarding colorectal cancer, a recent study developed several ML models to predict the RAS and BRAF mutation status for patients with advanced colorectal cancer[106]. Notably, the ANN demonstrated the best performance[106]. In a different study employing clinicopathologic variables and data generated from immunochemistry, a least absolute shrinkage and selection operator regression model was developed to predict the lymph node metastasis status in a cohort of operated patients for T1 colorectal cancer[107].

Opportunities of AI application in gastroenterology

Therefore, the opportunities that arise from applying AI/ML-based software in gastroenterology include:

AI/ML models could be developed and integrated into the clinical setting to employ routinely collected data directly from the patient's electronic health records and flag patients at high risk of developing certain GI diseases in real-time. Current efforts include the prevention of gastric[27-29] and colorectal cancer[30,31]. Such ML models could become the basis for tailoring targeted screening programs.

Endoscopy is the gold standard for the diagnosis of a plethora of GI diseases. AI/ML models employing non-invasive parameters that provide reliably accurate diagnosis could substitute endoscopy or significantly minimize its use, thus ameliorating the impact of endoscopy-related complications, significantly decreasing

Table 4 Artificial intelligence applications in gastroenterology: Prognosis

Ref.	Parameters employed	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Yang *et al*[97]	Laboratory results, immunomarkers, clinicopathological parameters	SVM	319/164 patients	Distant metastasis of oesophageal squamous cell carcinoma following surgery	69.5-80.1[1,6], 44.7-67.2[3,6], 81.6-97.7[4,6]
Sato *et al*[98]	Laboratory results, clinicopathological parameters, tumor characteristics	ANN	395 patients (training:validation:test = 53:27:20)	1-year and 5-year survival of patients with esophageal cancer following surgery	0.883-0.884[2,7], 78.1-80.7[3,7], 84.7-86.5[4,7]
Zhou *et al*[99]	Laboratory results, clinicopathological parameters, tumor characteristics	Several	2012 patients (training:validation = 8:2)	Recurrence of gastric cancer following surgery	0.790-0.962[2,5], 0.771-0.795[2,6]
Peng *et al*[100]	Meteorological data	ANN	901 patients	Variations of onset and relapse of IBDs	----
Hardalaç *et al* [101]	Clinicopathological parameters, treatment data	ANN	129 patients (training:validation:test = 80:10:10)	Prediction of mucosal remission for CD patients treated with azathioprine	58.1-79.1[1,6], 0.527-0.883[2,6]
Takayama *et al*[102]	Clinicopathological parameters, treatment data	ANN	54/36 patients	Prediction of the need for operation for UC patients treated with cytoapheresis	96[3,6], 97[4,6]
Lyles *et al* [103]	Laboratory results, clinicopathological parameters	CART	884 patients	Prediction of in-hospital mortality of upper GIB in cirrhotic patients	----
Grossi *et al* [104]	Laboratory results, clinicopathological parameters	ANN	807 patients	30-d mortality of patients with non-variceal upper GIB	81.2-89.0[1,6], 0.87[2,6], 81.5-93.3[3,6], 80.9-84.7[4,6]
Rotondano *et al*[105]	Laboratory results, clinicopathological parameters	ANN	2380 patients	30-d mortality of patients with non-variceal upper GIB	96.8[1,6], 0.95[2,6], 83.8[3,6], 97.5[4,6]
Shi *et al*[106]	CT radiomics	Several	124/35 patients	Prediction of the presence of RAS and BRAF mutations in CRC	ANN: 87[1,5], 71[1,6], 0.90-0.95[2,5], 0.79[2,6]
Kang *et al* [107]	Laboratory results, immunomarkers, clinicopathological parameters, tumor characteristics	LASSO	221/95 patients	Prediction of lymph node metastasis status in operated patients for T1 CRC	0.795[2,5], 0.765[2,6]

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. ANN: Artificial neural network; CART: Classification and regression tree; CD: Chron's disease; CRC: Colorectal cancer; CT: Computed tomography; GIB: Gastrointestinal bleeding; IBD: Inflammatory bowel disease; LASSO: Least absolute shrinkage and selection operator; SVM: Support vector machine; UC: Ulcerative colitis.

the cost for diagnosis, and providing an alternative to an unpleasant intervention for the patient. Current efforts include the diagnosis of gastroesophageal reflux disease [34], gastric cancer[47], celiac disease[48,50], and IBDs[51].

Expect from replacing endoscopy; AI/ML models could also improve its efficacy. CAD systems could facilitate navigating the GI tract and serve as the second "observer" for the endoscopist, an "observer" non-susceptible to distraction, which identifies lesions missed by the endoscopist. Particularly for capsule endoscopy, a CAD system that automatically detects and classifies lesions could significantly decrease the time required to evaluate the images by endoscopists while increasing diagnostic accuracy. Current efforts include the diagnosis of IBDs[53-57], gastric lesions[58], small bowel mobility disorders[59], small bowel lesions such as hemorrhagic and ulcerative lesions[60-64], intestinal hookworms[65], GI angiectasia[66], celiac disease[67], and colorectal cancer[68].

Physicians could use CADs to augment the accuracy of classifying polyps based on their neoplastic nature. As a result, the morbidity and mortality associated with failing to remove a neoplastic polyp could be lessened. At the same time, the complications related to removing a non-neoplastic polyp could be avoided. Therefore, CADs could significantly increase the cost-effectiveness of polyp management. Current efforts include the development of SVMs[73-76] and CNNs[79-81] for the classification of the neoplastic nature of colorectal polyps during NBI colonoscopy.

AI/ML models that accurately predict the response to different treatments could be used as a basis for individualized treatment allocation. Current efforts include the response of treatment with proton pump inhibitors for patients with gastroesophageal reflux disease[86], the mucosal remission for CD patients treated with azathioprine [101], and the need for operation for UC patients treated with cytoapheresis[102].

Particularly for GIB management, AI/ML models could be used to identify the source of bleeding[92] and as frameworks for decision-making, including the need to transfuse patients[92], perform emergent endoscopy[92,93] or emergent surgery[95, 96], and finally, triage patients with severe GIB to the ICU[95,96].

Finally, AI/ML models could be used as predictive tools that stratify the risk of complications and predict overall survival and recurrence following treatment. Such models could be used to tailor individualized follow-up schedules and for patient and family counseling. Current efforts include the prediction of overall survival for patients with esophageal cancer[98], the risk of recurrence of operated patients with gastric cancer[99], and 30-d mortality of patients with non-variceal GIB[104,105].

APPLICATIONS OF AI IN HEPATOLOGY

Prevention

Table 5 summarizes the findings of the identified studies applying AI/ML models for the prevention of disease in the field of hepatology. In a recent study, a prospective cohort of apparently healthy volunteers was enrolled in a study, which developed a DT-based model to classify patients based on their risk of developing non-alcoholic fatty liver disease (NAFLD) and liver fibrosis[108]. A similar study used routinely collected laboratory and clinical parameters to develop several AI/ML-based models to identify patients with NAFLD in the general population[109]. In a different study, several AI/ML models were developed as potential tools for targeted screening for NAFLD[110]. The Bayesian network model demonstrated the highest accuracy, followed by an SVM model[110]. Notably, an LR model outperformed all the models. These studies are examples of how AI/ML-based models could be used in primary care as tools for detecting chronic liver diseases. In a different study, multiple AI/ML models were developed as a basis for a non-invasive tool for assessing the level of fibrosis progression in NAFLD patients[111]. Notably, the RF model demonstrated the highest performance[111]. A different study managed to develop an ML model for patients at high risk of developing NAFLD to classify patients with non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH)[112]. Several AI/ML models were developed in a different study to classify healthy individuals from patients with NAFL and NASH[113]. Among them, a GB tree model demonstrated the highest performance, followed by an RF model[113]. In another study, four different AI/ML-based models were developed utilizing data from biochemical and enzyme-linked immunosorbent assays to classify patients with NAFLD and alcoholic liver disease with or without cirrhosis[114].

AI/ML models that utilize data directly from the electronic health records from patients infected with hepatitis B (HBV) and hepatitis C (HCV) could be used as the basis for targeted screening tools for liver carcinomas. A GB model was developed in a study using only serum markers aiming to predict fibrosis and classify the stage of fibrosis in two cohorts of patients with HBV and HCV[115]. An ANN model was developed in another study that employed only routine clinical data to predict significant fibrosis for patients with HBV[116]. In a similar study, an ANN model was developed by employing only routine laboratory data from HBV patients to identify patients with cirrhotic liver[117]. A different study developed an ANN model employing only non-invasive data to predict advanced liver fibrosis in HCV patients [118]. A similar study developed several AI/ML-based models to develop a non-invasive tool for identifying HCV patients with advanced fibrosis[119]. Among the developed models, a model based on alternating DTs demonstrated the highest performance[119]. These studies demonstrate how AI/ML models using routine clinical data could be used to identify patients who will benefit from a sustained

Table 5 Artificial intelligence applications in hepatology: Prevention

Ref.	Parameters employed	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Goldman *et al*[108]	National database of routine annual health check-ups	DT-based	12019 patients	Risk of NAFLD and cirrhosis	84.50-85.73[1,6], 0.7740-0.8486[2,6]
Yip *et al*[109]	Laboratory results, clinicopathological parameters	Several	500/422	Identify patients with NAFLD	0.87-0.90[2,5], 0.78-0.88[2,6], 55.48-94.52[3,5], 51.69-92.37[3,6], 58.47-91.53[4,5], 50.99-90.46[4,6]
Ma *et al*[110]	Laboratory results, clinicopathological parameters	Several	10508 patients (training:validation = 9:1)	Identify patients with NAFLD	49.47-82.92[1,6], 20.2-68.0[3,6], 54.4-94.6[4,6]
Sowa *et al* [111]	Laboratory results, clinicopathological parameters	Several	126 morbidly obese patients (training:validation = 9:1)	Fibrosis in NAFLD patients	79[1,6], 30.8-60.0[3,6], 77.0-92.2[4,6]
Canbay *et al* [112]	Laboratory results, clinicopathological parameters	EFS	164/122 obese patients	Classification of NAFLD and NASH	0.7339[2,5], 0.7028[2,6]
Fialoke *et al* [113]	National database of routine annual health check-ups	Several	108139 patients (training:validation = 4:1)	Classification among healthy, NAFLD, and NASH	77.2-79.7[1,6], 0.842-0.876[2,6], 74.5-77.4[3,6]
Sowa *et al* [114]	Data from biochemical and enzyme-linked immunosorbent assays	Several	133 patients (training:validation = 9:1)	Classification of NAFLD and ALD	DT: 89.02-95.1[1,6], 74.19-94.12[3,6], 96.08-98.04[4,6], RF: 0.8932-0.9846[2,6], SVM: 0.9058-0.9118[2,6]
Wei *et al* [115]	Laboratory results, clinicopathological parameters	GB	576 HBV patients, (training:validation = 7:3), 368[7] HCV patients	Classification of fibrosis/cirrhosis in HBV patients	0.904-0.974[2,5], 0.871-0.918[2,6], 79-88[3,5], 78-84[3,6], 86-92[4,5], 85[4,6]
				Classification of fibrosis/cirrhosis in HCV patients	0.797-0.849[2,7]
Wang *et al* [116]	Laboratory results, clinicopathological parameters	ANN	226/113[6]/116[7] HBV patients	Classification of significant fibrosis	0.883[2,5], 0.884[2,6], 0.920[2,7]
Raoufy *et al* [117]	Laboratory results, clinicopathological parameters	ANN	86/58 HBV patients	Classification of liver cirrhosis	91.38[1,6], 0.898[2,6], 87.5[3,6], 92[4,6]
Piscaglia *et al*[118]	Laboratory results, clinicopathological parameters	ANN	414/96 HCV patients	Classification of significant fibrosis	45.8-86.5[1,6], 0.87[2,5], 0.93[2,6], 30.4-100[3,6], 30.1-98.6[4,6]
Hashem *et al* [119]	Laboratory results, clinicopathological parameters	Several	22690/16877 HCV patients	Classification of significant fibrosis	66.3-84.4[1,6], 0.73-0.76[2,6]
Ioannou *et al* [120]	Clinical/laboratory data extracted directly from electronic health records	DNN	48151 patients with HCV-related cirrhosis (training:validation = 9:1)	HCC development in HCV cirrhosis	0.759-0.806[2,6]
Emu *et al* [121]	Laboratory results, clinicopathological parameters	Several	1385 patients HCV (training:validation = 4:1)	Stage of liver cirrhosis	97.228-97.831[1,6]

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. ALD: Alcoholic liver disease; ANN: Artificial neural network; DNN: Deep neural network; DT: Decision tree; EFS: Ensemble feature selection; GB: Gradient boosting; HBV: Hepatitis B; HCC: Hepatocellular carcinoma; HCV: Hepatitis C; NAFLD: Non-alcoholic fatty liver disease; NASH: Non-alcoholic steatohepatitis RF: Random forest; SVM: Support vector machine.

virological response (SVR) to delay chronic liver disease progression.

Regarding the development of HCC, a recent study investigated if a DNN could outperform conventional LR models in predicting HCC development in patients with HCV[120]. A different study using variables including demographic, laboratory results, and clinical findings, developed RF, DNN, and LR models to reliably predict

the stage of liver cirrhosis in patients infected with HCV[121]. Accurately predicting patients with HCV prone to develop HCC could help identify patients who would be benefited from a targeted screening and are in a greater need of antiviral treatment to achieve an SVR.

Diagnostics

Table 6 summarizes the findings of the identified studies applying AI/ML models for diagnosis in the field of hepatology. Regarding chronic liver disease, a study used CT imaging data of patients with a confirmed liver fibrosis diagnosis to develop a CNN model for the staging of liver fibrosis[122]. Notably, the model outperformed the radiologists' interpretation[122]. A different study employed ultrasound imaging of patients with fatty liver disease to develop an SVM model and an extreme learning machine model (a type of ANN) for diagnosis and risk stratification[123]. A different study that employed ultrasound shear wave elastography features developed an SVM model to identify patients with chronic liver disease[124]. Another study used images from real-time tissue elastography to develop four different AI/ML-based models to classify liver fibrosis[125]. Notably, the RF model outperformed the rest, followed by the KNN and the SVM models[125].

In a study, CT imaging was employed to develop an ANN model that differentiates between HCC, intrahepatic peripheral cholangiocarcinoma (CCA), hemangioma, and metastasis[126]. Interestingly, when radiologists evaluated the images, their performance significantly increased when they considered the ANN's output, from an AUROC of 0.888 to one of 0.934[126]. Focusing on HCC diagnosis, a recent retrospective study developed a CNN that employed MRI images of patients with HCC. The model was trained with a combination of images that met the Liver Imaging Reporting and Data System (typical) and with images that did not (atypical)[127]. In a multicenter, retrospective study, a CNN was developed that employed MRI scans to identify HCC lesions[128]. Notably, the model surpassed the performance of less experienced radiologists in the diagnosis of small HCC lesions[128]. The model was able to analyze 100 images in just 3.4 s[128]. Finally, a different study aimed to develop a non-invasive ANN model that predicts the presence of microscopic vascular invasion and the tumor grade of HCC[129].

Regarding diagnosis of CCA and pancreatic adenocarcinoma, a study developed an ANN model using data generated by metabolomic and proteomic analyses of bile from patients undergoing endoscopic retrograde cholangiopancreatography aiming to classify patients with and without cancer[130]. A different study used data from the plasma levels of bile acids to develop six different AI/ML-based models to classify patients as having CCA or a benign biliary disease[131]. Among the six developed models, a model based on the Naive Bayes classifier demonstrated the highest performance[131]. Finally, in another study, an ANN model was designed to analyze images from magnetic resonance cholangiopancreatography to diagnose CCA with a reported accuracy of 92.8%[132].

Management

Table 7 summarizes the findings of the identified studies applying AI/ML models to manage hepatic diseases. With the guidance of AI/ML-based software trained by data generated from gene mutation biomarkers, serum markers, imaging, and the clinical setting, high-quality, evidence-based, and individualized treatments could be employed regarding chemotherapy, radiotherapy, and immunotherapy. In HBV management, a recent study developed several AI/ML-based models that use soluble immune markers to predict early virological relapse after discontinuation of nucleoside analogs treatment[133]. The model could be used to exclude patients at high risk of virological relapse from treatment cessation. In HCV management, a study utilized data from full-length HCV genome sequencing of variants of HCV to develop multiple AI/ML-based models that classify HCV anti-viral resistance variants[134]. Notably, the SVM model demonstrated the highest performance[134].

In HCC management, a recent study used data from DNA methylation profiling to develop an RF-based model that predicts 6-mo progression-free survival. Such models could be used to personalize patient surveillance[135]. In an international, multi-institutional study, a CART model was developed that aimed to create a framework beyond the Barcelona-Clinic-Liver-Cancer (BCLC) staging system, which is currently endorsed by guidelines for treatment allocation[136,137]. The model defined six distinct prognostic groups of patients based on predictive factors of overall survival that could be used as a framework for treatment allocation[137]. Notably, the radiologic tumor burden score, which is not integrated into the BCLC staging system, was found as the best predictor of long-term outcome for BCLC stage B patients[137].

Table 6 Artificial intelligence applications in hepatology: Diagnosis

Ref.	Diagnostic Modality	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Choi *et al*[122]	CT imaging	CNN	7461/421[6]/298[7]/172[7] patients	Liver fibrosis staging (F0-F4)	83.1[1,5], 80.8[1,6], 74.4-80.2[1,7]
				Classification among significant fibrosis, advanced fibrosis, and cirrhosis	92.1-95.0[1,6,7], 0.95-0.97[2,6,7], 84.6-95.5[3,6,7], 89.9-96.6[4,6,7]
Kuppili *et al* [123]	US imaging	ELM, SVM	63 patients	Diagnosis of FLD	ELM: 81.7-92.4[1,6], 0.81-0.92[2,6], 85.10-91.30[3,6], 78.52-92.10[4,6], SVM: 76.14-86.42[1,6], 0.74-0.86[2,6], 76.80-88.20[3,6], 74.52-86.30[4,6]
Gatos *et al*[124]	US shear wave elastography imaging	SVM	126 patients	Classification of chronic liver disease from healthy patients	87.3[1,6], 0.87[2,6], 93.5[3,6], 81.2[4,6]
Chen *et al*[125]	Real-time tissue elastography imaging, age, sex	Several	513 patient (training:validation = 3:1)	Classification of liver fibrosis	80.44-82.87[1,6], 79.67-92.97[3,6], 46.25-82.50[4,6]
Matake *et al* [126]	Clinicopathological parameters, CT imaging	ANN	120 patients	Classification among four types of focal liver lesions	0.961[2,6]
Oestmann *et al* [127]	Multiphasic MRI scans	CNN	150/10 patients	Classification of HCC and non-HCC lesions	94.1[1,5], 87.3[1,6], 0.912[2,6]
					For HCC: 92.7[3,6], 82.0[4,6]
					For non-HCC: 82.0[3,6], 92.7[4,6]
Kim *et al*[128]	MRI scans	CNN	455[5,6]/54[7] patients	HCC detection	0.97[2,6], 94[3,6], 99[4,6], 0.90[2,7], 87[3,7], 93[4,7]
Cucchetti *et al* [129]	Laboratory results, clinicopathological parameters, radiological data, histological data	ANN	175/75 patients	MVI	0.92[2,5], 91.0[1,6]
				Histopathological Grade	0.94[2,5], 93.3[1,6]
Urman *et al* [130]	Metabolomic and proteomic analyses of bile	Several	139 patients	Classification of CCA and pancreatic adenocarcinoma	0.98-1.00[2,6], 88-94.1[3,6], 92.3-100[4,6]
Negrini *et al* [131]	Plasma bile acids profiles	Several	112 patients (training:validation = 4:1)	Classification of CCA and benign biliary disease	68.2-86.4[1,6], 0.77-0.95[2,6], 64-79[3,6], 63-100[4,6]
Logeswaran [132]	MRCP	MLP	55/593[7] images	CCA diagnosis	92.8-96.3[1,6], 83.64-90.14[1,7]

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. ANN: Artificial neural network; CCA: Cholangiocarcinoma; CNN: Convolutional neural network; CT: Computed tomography; ELM: Extreme learning machine; FLD: Fatty liver disease; HCC: Hepatocellular carcinoma; MLP: Multi-layer perceptron; MRCP: Magnetic resonance cholangiopancreatography MRI: Magnetic resonance imaging; MVI: Microvascular invasion; SVM: Support vector machine; US: Ultrasound.

These findings could help us reevaluate HCC management into a multidisciplinary, individualized approach that goes beyond the BCLC criteria[138].

In the management of patients with CCA, in an international study, a CART model was developed with solely preoperative variables to identify patients who would be more likely to benefit from surgery[139]. The model managed to isolate four distinct prognostic groups of patients with similar patient outcomes[139]. The authors concluded that this model could be used to inform presurgical decisions-for example, the use of neoadjuvant therapy for patients with poor prognoses[139]. In a different study, the researchers developed a DNN model to establish an AI framework through which specific prognostic groups could be used to identify which patients were more likely to benefit from different treatment modalities such as neoadjuvant chemo-therapy or transarterial chemoembolization[140]. The framework was found to be

Table 7 Artificial intelligence applications in hepatology: Treatment

Ref.	Parameters employed	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Wübbolding *et al*[133]	Analyze soluble immune markers	Several	28/49[7] HBV patients	Prediction of early virological relapse	0.73-0.89[2,6], 0.59-0.67[2,7]
Haga *et al*[134]	WGS of HCV	Several	86/87 HCV patients	Classification of HCV variants resistant to antiviral drugs	0.5-0.937[2,5], 0.597-0.954[2,6]
Bedon *et al*[135]	DNA methylation profiling	RF-based	300/74 HCC specimens	6-mo progression-free survival	67.1-80.6[1,5], 64.8-80.2[1,7]
Tsilimigras *et al* [137]	Laboratory results, clinicopathological parameters, tumor characteristics	CART	976 HCC patients	Determining factors of prognostic weigh preoperatively within the BCLC staging system	---
Tsilimigras *et al* [139]	Laboratory results, clinicopathological parameters, tumor characteristics	CART	1146 CCA patients	Determining factors of prognostic weigh preoperatively	---
Jeong *et al*[140]	Laboratory results, clinicopathological parameters	DNN	1421/234[7]	Intrahepatic CCA susceptible to adjuvant therapy following resection	0.84[2,5], 0.78[2,7]
Shao *et al*[141]	Clinicopathological parameters	ANN	288 CCA patients (training:validation = 8:2)	Predict early occlusion following bilateral plastic stent placement	0.9648[2,5], 0.9544[2,6]

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. ANN: Artificial neural network; BCLC: Barcelona clinic liver cancer; CART: Classification and regression tree; CCA: Cholangiocarcinoma; DNN: Deep neural network; HCC: Hepatocellular carcinoma; HCV: Hepatitis C; RF: Random forest; WGS: Whole-genome sequencing.

significantly more accurate than the current guidelines of the American Joint Committee of Cancer[140]. Finally, a study developed an ANN to predict which patients with inoperable hilar CCA will develop early occlusion following a bilateral plastic stent placement[141].

Prognosis

Table 8 summarizes the findings of the identified studies applying AI/ML models regarding prognosis in the field of hepatology. An ANN model was developed in a study to identify patients with HBV cirrhosis at a high risk of developing esophageal varices[142]. In a different study, two different RF models were created using clinical data, the first to identify esophageal varices and the second to classify patients with esophageal varices that require treatment[143].

Regarding HCC, several studies have focused on developing AI/ML-based models that would reliably predict patient outcomes (survival and recurrence). A retrospective study compared the performance among an ANN, an LR model, and a DT to predict the 1-, 3-, and 5-year disease-free survival in patients with HCC following resection [144]. A similar study used a nationwide database to compare an ANN and an LR model in predicting the 5-year survival of patients with HCC following hepatic resection and concluded that the ANN model surpassed the performance of the LR model[145]. In a different study by the same department, an ANN model and an LR model were compared, but instead, the outcome was in-hospital mortality, and the ANN was found superior to the LR model[146]. Interestingly, the study reported that the surgeon volume was the best single predictor of in-hospital mortality[146]. Regarding predictors, a different study comparing ANN and LR models reported that the ANN model identified a greater number of significant predictors than the LR model, except for outperforming it in survival predictions[147]. In a prospective study that aimed to compare an ANN model's performance to the performance of traditional staging systems in predicting survival for patients with early HCC, the ANN model outperformed the Hepato-Pancreato-Biliary Association's, the TNM 6th, and the BCLC staging systems with higher reported AUROCs in all training, internal validation, and external validation cohorts[148]. In a recent study, a GB survival classifier-based

Table 8 Artificial intelligence applications in hepatology: Prognosis

Ref.	Parameters employed	AI classifier	Sizes of the training/validation sets	Outcomes	Performance
Hong *et al* [142]	Laboratory results, clinicopathological parameters	ANN	197 HBV patients (training:validation = 4:1)	Development of esophageal varices in HBV cirrhosis	87.82[1,6], 93.75[3,6], 71.70[4,6]
Dong *et al* [143]	Laboratory results, clinicopathological parameters	RF	238/109[7]	Identification of esophageal varices	0.84[2,5], 0.82[2,7]
				Classification of esophageal varices requiring treatment	0.74[2,5], 0.75[2,7]
Ho *et al*[144]	Laboratory results, clinicopathological parameters, surgery parameters	ANN, DT	427, 354, and 297 patients for 1-, 3-, and 5-year survival (training:validation = 8:2)	1-, 3-, and 5-year disease-free survival Following surgical resection	ANN: 0.963-0.989[2,5], 93.5-96.3[3,5], 91.6-97.9[4,5], 0.774-0.864[2,6], 70.0-78.7[3,6], 54.2-92.7[4,6] DT: 0.675-0.825[2,5], 19.6-94.8[3,5], 45.8-97.9[4,5], 0.561-0.718[2,6], 0-88.5[3,6], 37.5-96.4[4,6]
Shi *et al*[145]	Laboratory results, clinicopathological parameters, tumor characteristics	ANN	22926 patients	5-year survival following surgical resection	96.57[1,6], 0.885[2,6], 97.43[1,7], 0.871[2,7], 74.23[3,7],
Shi *et al*[146]	Laboratory results, clinicopathological parameters, surgery parameters	ANN	22926 hepatectomies	In-hospital mortality following surgical resection	97.28[1,6], 0.84[2,6], 95.93[1,7], 0.82[2,7], 78.40[3,7], 94.57[4,7]
Chiu *et al* [147]	Laboratory results, clinicopathological parameters, tumor characteristics	ANN	434, 341, and 264 patients for 1-, 3-, and 5-year survival, (training:validation = 8:2)	1-, 3-, and 5-year overall survival, following surgical resection	98.5-99.5[1,5], 0.980-0.993[2,5], 99.7-100[3,5], 96.2-99.2[4,5], 72.1-85.1[1,6], 0.798-0.875[2,6], 71.4-88.6[3,6], 50.0-82.1[4,6]
Qiao *et al* [148]	Laboratory results, clinicopathological parameters, tumor characteristics	ANN	362/181[6]/104[7] patients	Survival following surgical resection	0.855[2,5], 80.00[3,5], 73.40[4,5], 0.832[2,6], 78.67[3,6], 75.70[4,6], 0.829[2,7], 77.42[3,7], 78.08[4,7]
Liu *et al*[149]	Laboratory results, data from immunochemistry of peripheral blood mononuclear cells, tumor characteristics	GB survival classifier	136/56[6]/105[7]	Risk of HCC-related death	0.844[2,5], 0.827[2,6], 0.806[2,7]
Zhong *et al* [150]	ALBI/CTP stage	ANN	319 / 61[7] / 124[7]	Survival of patients treated with chemoembolization and sorafenib	ALBI-based: 0.716[2,7], 0.823[2,7] CTP-based: 0.779[2,7], 0.693[2,7]
Divya and Radha[152]	Laboratory results, clinicopathological parameters, tumor characteristics	APO, SVM, RF	152 patients	Recurrence following RFA	95.5[1,6], 95.1[3,6], 95.8[4,6]
Yamashita *et al*[153]	Hematoxylin and eosin-stained WSI	CNN	299 / 53[6]/198[7] WSIs	Recurrence following Surgical Resection	0.724[2,6], 0.683[2,7]
Liang *et al* [154]	Laboratory results, clinicopathological parameters	SVM	83 patients	Recurrence following RFA	73-82[1,6], 0.60-0.69[2,6], 77-86[3,6], 73-82[4,6]
Eaton *et al* [155]	Laboratory results, clinicopathological parameters	GB-based	509/278 patients with primary sclerosing cholangitis	Classify risk of primary sclerosing cholangitis-related complications	0.96[2,6], 0.90[2,7]
Andres *et al* [156]	Laboratory results, clinicopathological parameters, donor characteristics	PSSP system	2769 patients	Survival following transplantation for primary sclerosing cholangitis	----
Rodriguez-Luna *et al* [157]	Genotyping data from microsatellite mutations/deletions	ANN	19 transplated patients	Post-transplant HCC recurrence	89.5[1,6]
Lau *et al*[158]	Laboratory results, clinicopathological parameters, donor characteristics	ANN, RF	90/90 transplants	Graft failure/primary nonfunction	ANN: 0.734-0.835[2,6] RF: 0.787-0.818[2,6]
				3-mo graft failure	ANN: 0.559[2,6], R6: 0.715[2,6]
Briceño *et al* [160]	Laboratory results, clinicopathological parameters,	ANN	1003 liver transplants	3-mo graft failure	0.806-0.821[2,6]

	surgical parameters, donor characteristics

[1]Accuracy (%).

[2]Area under the receiver operating curve or c-index.

[3]Sensitivity (%).

[4]Specificity (%).

[5]Training.

[6]Internal validation.

[7]External validation/testing. ALBI: Albumin-bilirubin; ANN: Artificial neural network; APO: Artificial plant optimization; CNN: Convolutional neural network; CTP: Child–Turcotte–Pugh; DT: Decision tree; GB: Gradient boosting; HBV: Hepatitis B; HCC: Hepatocellular carcinoma; PSSP: Patient-specific survival prediction; RF: Random forest; RFA: Radiofrequency ablation; SVM: Support vector machine; WSI: Whole-slide image.

model was developed to stratify the risk of an HCC-related death into three distinct categories[149]. Finally, in another recent study, an ANN identified albumin-bilirubin grade as the most important prognostic factor for the survival of patients with HCC treated with the combination of transarterial chemoembolization and sorafenib as initial treatment[150]. A systematic review aimed to compare the performance of AI/ML-based software and that of traditional linear prediction models in predicting survival for patients with HCC concluded that AI/ML models provided enhanced accuracy[151].

Except for survival, other studies have developed AI/ML models to predict the recurrence of HCC following therapeutical treatment. An interesting study employed several AI/ML methods, including Artificial Plant Optimization, SVM, and RF to predict HCC recurrence following radiofrequency ablation[152]. A recent study developed a CNN employing histopathologic images to predict recurrence in operated HCC patients[153]. A different study developed an SVM model to predict the recurrence in a group of patients who underwent radiofrequency ablation[154].

Regarding primary sclerosing cholangitis, a team of researchers derived and validated a risk estimate tool based on GB algorithms to predict the outcomes of the disease[155]. In a different study, an ML model was developed to predict survival curves for patients with primary sclerosing cholangitis following liver transplantation [156]. The *P* value of the χ^2 test of the distributional calibration was 1, indicating excellent calibration of the model[156].

Focusing on liver transplantation, a team developed an ANN model combined with genotyping for microsatellite mutations/deletion to predict HCC recurrence in a cohort of patients receiving a liver transplant[157]. Several other studies have focused on the survival of individual liver grafts following transplantation[158-160]. In a multi-centered study, the authors developed an ANN model to predict the 3-mo graft survival and loss[160]. Interestingly, their model outperformed all extensively validated scores, including the MELD, the donor risk index (DRI), the survival outcome following liver transplantation, and the balance of risk, with their performance significantly lower (AUROCs range: 0.42-0.67)[160]. In a different study, an RF model was developed to predict the 30-d failure graft. Notably, this model too outperformed MELD and DRI[158]. These findings could help reevaluate our thinking regarding the current models of recipient-donor matching.

Opportunities of AI application in hepatology
Therefore, the opportunities that arise from applying AI/ML-based software in hepatology include:

Regarding primary care, ML models could be integrated into the clinical setting and flag individuals in the general population at high risk of developing chronic liver disease in real-time by employing routinely collected data from the electronic health records. Current efforts include models that identify patients at high risk to develop NAFLD, NASH, fibrosis, and cirrhosis[108-114]. These models could be used to design targeted screening programs.

Particularly for patients with chronic HBV and HCV infection, AI/ML models could be used to stratify each individual's risk to progress through the several stages of cirrhosis and develop HCC. Multiple studies have been conducted in this regard for both HBV-related[115-117] and HCV-related[115,118-121] cirrhosis. Such models could be used to identify patients in greater need of SVR and tailor individualized follow-up schedules.

In diagnosis, AI/ML models provide the opportunity for increased diagnostic accuracy of various hepatic diseases and in multiple diagnostic modalities (such as

US/CT/MRI imaging and histologic images). Current efforts include CAD models for the diagnosis of chronic liver disease, HCC[127-129], and CCA[130-132].

AI/ML models that accurately and reliably predict the response to specific treatments could be used to tailor evidence-based frameworks for individualized treatment allocation. Current efforts include the development of frameworks for the management of HCC[136,137] and CCA[139,140].

Regarding prognosis, AI/ML models could be used to reliably predict complications, in-hospital mortality, overall survival, and recurrence following treatment. Current efforts include the prediction of in-hospital and HCC-related death[146,149], of overall survival of HCC patients following surgery[144,145,147,148], and HCC recurrence following surgery or RFA[152-154]. Such models could be used to tailor individualized follow-up schedules and for patient and family counseling.

Particularly for liver transplantation, AI/ML models could be used to predict graft failure or cancer recurrence for patients transplanted for cancer. Such models could be used to reevaluate and optimize our current practices regarding recipient-donor matching for graft allocation[157-160].

CURRENT CHALLENGES OF AI APPLICATIONS

Intrinsic bias and accuracy

The level of accuracy of AI systems is mostly dependent on the quality of the training dataset. Existing biases and prejudices in the training data set will inadvertently be built into the algorithms limiting the AI/ML-based software's accuracy[161]. Discrepancies in the data collection process, imperfections of standardization, and incorrectly labeled cases become part of the algorithms' training and are thus integrated into the end product. Two of the most common biases found in these models are: (1) Spectrum bias; and (2) Overfitting. Spectrum bias occurs when the patients (whose generated data were used during the training and internal validation of these models) do not constitute a representative sample of the target population[162]. On the other hand, overfitting refers to the tendency of models to be customized for the training data [162]. The performance of the model is thus exaggerated for the training dataset but is significantly inferior in new datasets[162]. CNNs, which are extensively used in gastroenterology and hepatology, are particularly vulnerable to overfitting[26]. Another substantial bias source is that physicians often misrecord data in the electronic health records, sometimes even the chief complaint[163]. These biases significantly impact the performance of AI/ML-based models and undermine their applications. A measure to alleviate the impact of biases is the standardization of data collection methods while establishing evaluation systems that scrutinize underlying biases and check data accuracy[15,164]. Another crucial step to tackle these biases is the external validation of models, in a clinical setting, from data generated from patients prospectively enrolled in the study[165]. Initial implementation of AI/ML-based models in the clinical setting should occur on a small scale, similar to phase I and phase II of the clinical trials[166]. Existing tools, such as the PROBAST tool, could be used during the development of AI/ML-based models to comprehensively assess the risk of bias[164,167]. The American Medical Association has recently acknowledged the need to identify and address bias in data when testing or deploying AI/ML-based software to avoid introducing or exacerbating health care disparities [168]. Therefore, the real challenge is to develop an AI/ML-based software that taps into the true potential hidden by the data without picking up the biases.

Several other limitations affect the performance of these models. An example is the limited resolution of capsule endoscopy images compared to other types of endoscopy [18]. A different limitation is the in silico nature of most of the currently developed models, which significantly impacts the expectations of similar performances in a real clinical setting. Another limitation of these algorithms is that they utilize a series of variables that in the real clinical setting are derived in a series of careful decisions made by physicians and are not readily available as a complete set of data. Therefore, these variables' presence or absence could become a significant shortcoming of these models in a real clinical setting, making them impractical to use. Finally, there is a lack of consistency in the metrics used to assess performance (sensitivity, specificity, AUROC, accuracy, *etc.*), limiting the ability to draw meaningful comparisons between the models[17].

Data protection and cybersecurity

Similarly, with all patient data, those utilized in AI/ML-based software should

conform with the seven principles described in Article 5 of the General Data Protection Regulation. Specifically, the personal data are required to: (1) Be processed fairly, lawfully, and transparently; (2) Be relevant, adequate, and limited to the intended purpose; (3) Be collected for explicit, specific, and lawful purposes; (4) Be accurate and up to date; (5) Permit identification for only as long as necessary; (6) Ensure appropriate security; and (7) Demonstrate compliance and accountability[169]. A particular challenge for AI is the resistance to the concept of utterly electronic tracking of healthcare records due to the belief that it exposes the vast amount of stored sensitive health record data to massive disclosures[170]. These concerns are not entirely unjustifiable if one considers examples such as the transfer of data from 1.6 million patient records from the Royal Free Nathional Health Service Foundation Trust to Google DeepMind, which was later ruled illegal[171]. Nevertheless, any effort or policy towards paper-based data due to data protection concerns could substantially undermine the application of AI modes in healthcare.

The healthcare industry is a particularly attractive target for cyberattacks as it contains sensitive personal data and financial information. Several physical and technical safeguards have been implemented under the Health Insurance Portability and Accountability Act to protect against the breach of sensitive patient data[172]. However, from the application of AI technology in the healthcare industry, new vulnerabilities and dangers emerge except traditional cybersecurity concerns. If ML models can learn from data, they could also be fooled by data. Data could be introduced malevolently in the algorithms to manipulate the developed AI/ML models into making wrong decisions with currently unknown ramifications to patient outcomes[173]. In a recent study, the authors demonstrated how attackers could use DL to add or remove lung cancer tumors in CT scans[174]. This study demonstrated how both a group of radiologists and a state-of-the-art deep AI model were particularly susceptible to the attack[174]. How could we, therefore, be confident that the AI/ML model has not been compromised? Evidently, healthcare facilities will have to develop additional information technology infrastructure to shield the healthcare system from these new threats.

Intellectual property

The Food and Drug Administration (FDA) recently acknowledged that it receives a high volume of submissions regarding AI/ML-based software marketing, with the list of already approved algorithms increasing rapidly[175,176]. A challenging point will be determining the law framework and regulatory standards that AI/ML-based software should follow. The key is classifying AI/ML software as "medical device", "service", or as "product", and achieving this requires a careful evaluation of the intentive use of AI/ML-based software. AI/ML-based software that aims to assist physicians in diagnosing, interpreting, and treatment decisions could be classified as medical devices and fall under the respective regulations[12]. In 2019, the FDA announced its aim to review AI/ML-based software regulation and has recently published the AI/ML– Based Software as a Medical Device Action Plan[175,177]. The action plan published by the FDA focuses on five pillars aimed to facilitate innovation and advance AI/ML-based software that are classified as medical devices. These include: (1) A tailored regulatory framework; (2) Good ML practice that could be achieved by consensus standards efforts; (3) A patient-centered approach incorporating transparency to users that takes into account usability, trust, equity, and accountability; (4) Regulatory science methods related to algorithm bias and robustness, and finally; and (5) Real-world performance[175]. When it comes to law and regulation, the real challenge is finding the golden snitch between too much regulation that strangles innovation and creation and too little regulation, which could have unexpected and devastating consequences for healthcare, and by extension, the well-being of patients.

Another interesting point regarding intellectual property and safety is the substantial divergence of the AI/ML-based software from the original product years after its approval and distribution[178]. What are the rights of developers on the product following its purchase? Since the original product constantly changes by learning from the clinical setting's data, the deviated model years after the purchase and the original software could be seen as two entirely different products. The first is protected under copyright law. However, the latter has now produced intellectual property on its own since it encompasses data generated by the healthcare facility [173]. Who could therefore claim legal rights over the final product? Also, who ensures the credibility and safety of the model as it deviates from the original product? Clearly, there is a need for lifecycle regulation of AI/ML-based software that ensures postapproval guardrails such as built-in audits[165]. Alternatively, time-limited authoriz-

ations could be employed to allow the FDA to perform periodic audits to review the accumulated modifications to the initial product[165].

Liability

Except for intellectual property, another legal challenge that arises from applying AI/ML-based software in the clinical setting is liability. A quite intriguing concept is that AI's findings and decisions could become legally binding in the future. As AI/ML-based models evolve and become more sophisticated, it is fair to assume that they will eventually surpass physicians, at least in specific tasks. How could then, the physicians justify ignoring the decision presented by AI? Especially when their decisions are made solemnly based on data and lack any sense of subjectivity. And who is liable when the followed decision made by the model causes injury? Currently, there is no legal precedent that assigns liability in a case where the injury was inflicted on a patient due to an erroneous output generated by AI/ML-based software[179].

To avoid malpractice liability, a physician is required to provide medical care at the same level as a competent physician of the same specialty while considering the available resources[180]. However, when an AI/ML-based software recommendation is involved, the concept of liability becomes more complicated. In an insightful recent legal analysis, the authors provide eight scenarios based on the combinations of whether the AI recommendation follows standard care and/or is accurate, the physician follows or rejects that recommendation, and whether a medical injury occurs [179]. The authors conclude that since current law shields physicians from liability when the standard care is followed, it also incentivizes physicians to minimize AI's actual usefulness, transforming AI into a confirmatory tool rather than a tool to augment the level of care[179]. Until a comprehensive legal framework regarding liability is developed, healthcare facilities would justifiably hesitate to adopt AI technologies due to the fear and unawareness of how they will expose the facility and its staff to liability[181].

Ethical challenges and transparency

AI could become the third participant in the physician-patient relationship and potentially undermine the trust between them. First, the idea that data are shared with third parties for AI model development could lead patients to withhold information from physicians and become less transparent[182]. Second, AI/ML-based models cannot act like compassionate human beings, which is an integral part of a physician's clinical life. Therefore, AI-driven decision-making neither encompasses an understanding of the patient's needs nor respects the patient's wishes nor demonstrates empathy, nor realizes when a patient feels discomfort or requires some rest or a hand to hold on to. Therefore, retaining the trust in the physician-patient relationship could be proven challenging in the AI era.

Another challenge is patient consent in AI applications. Clearly, it is practically impossible to acquire informed consent from each and every patient whose data are used during the development and validation of AI/ML-based software. In any case, we cannot predict how the algorithms would use specific data points during the training of the model and whether data of a specific patient have any significant impact on the model as a whole[183]. However, when AI/ML-based models are used in the clinical setting, especially for patient recording (in computer vision), patients should be adequately informed, and explicit consent should be acquired[184].

Achieving equilibrium between ensuring a high level of care and avoiding privacy encroachment could be challenging. For example, with computer vision advancement, monitoring could be used to detect any deviations from the optimal bedside practices such as patient mobilization and hand hygiene practices, which leaves patients vulnerable to identification[185,186]. Practices such as data minimization (collecting the least data needed) could address these challenges[185].

Changing the physicians' and patients' stance towards AI technologies could be proven a herculean task. Currently, around 2/3 of the population feels uncomfortable using their data to improve care quality and are unfavorable to AI/ML-based software performing tasks typically performed by physicians[183]. There are justifiable concerns that certain biases included in the training data due to racism, sexism, and socioeconomic inequality would be integrated into the AI/ML-based software. A notorious example is the COMPAS algorithm, which was found to falsely flag black people as likely to re-offend[187]. This was aggravated when developers made the argument that the algorithm was protected under intellectual property law and thus is not open to scrutiny[187]. These examples could erroneously undermine the trust towards AI altogether. A trustworthy AI/ML-based software should be built around the principles of transparency, credibility, auditability, reliability, and recoverability

[188].

The last but not least challenge in applying AI/ML-based software in healthcare is the lack of transparency, which is currently a major concern. The lack of transparency is demonstrated by the non-interpretability associated with these algorithms, described as a "black-box", where the inner logic is hidden[18]. This violates a fundamental tenet of medical ethics that physicians should comprehend at least the basic features of the devices they use and could undermine the trustworthiness of these technologies[189]. Thus, those who develop AI/ML-based models and physicians should collaborate to reach at least a degree of explainability on how AI/ML-based algorithms reach decisions. Attention maps and saliency region are examples of methods that could lessen the lack of interpretability of these models[190]. A slightly different and intriguing point of view is that withholding the widespread application of AI/ML-based software, due to their opacity, when their application could significantly benefit patients is, in fact, unethical[183]. With most of the population being prejudiced against AI, developing trust for these technologies would require several steps, including addressing bias, increasing transparency, communicating with the patients their role in provided care, protecting data privacy, and developing a robust regulatory framework.

CONCLUSION

Even though multiple efforts for AI's integration into healthcare have been made, they all originate from high-income countries[191]. Understandably, since all AI/ML-based software is data-driven, the opportunities to participate in the AI ecosystem for low- and middle-income countries, where the healthcare system does not generate extensive electronic, standardized data, are significantly limited. Not being included in the development of these models would introduce a significant spectrum bias, which could undermine these models' application in low- and middle-income countries. Thus, despite being described as the key to healthcare equity, AI could become another brick in the wall of inequality. A challenge for the future would be to include low- and middle-income countries in AI model development initiatives.

Even though these AI/ML-based software's central idea is that they could surpass physicians in performing certain tasks, the existing literature comparing AI/ML-based models and physicians' performance is limited[192]. A recent meta-analysis, which included multidisciplinary comparative studies, identified only 25 studies meeting their inclusion criteria from all medical specialties[193]. The authors concluded that currently, DL models' diagnostic performance in detecting diseases from medical imaging is equivalent to that of healthcare professionals[193]. Clearly, there is a contrast between the abundance of studies developing and validating ML algorithms and the lack of studies comparing these models' performance to physicians' performance at the investigated task. Thus, there is a crystalline need for further comparative studies in the clinical setting.

Another challenge for the future would be to combine AI with other emerging fields, such as 3D printing and bioprinting, augmented reality, novel biomarkers such as microRNAs, and robotics[194-198]. Envision a world where all these technologies and their applications are unified and integrated into everyday clinical practice. AI could be the means through which this new sophisticated, complex clinical setting is handled. The future AI clinician rejects oversimplifying an inherently complex field but instead embraces the complexity.

AI can outperform physicians in the ability to precisely quantify correlations even in domains where physicians possess in-depth knowledge[173]. However, AI models are simply tools like any other and should be treated as such. Tools with their accuracy, sensitivity, and specificity in performing certain tasks, tools that carry biases, and whose findings should be evaluated in conjunction with other clinical and paraclinical findings. Reliance on AI should not exclude non-quantifiable information from decision-making[183]. Clinical reasoning and critical thinking should not be subsumed by AI at the altar of technological advancement. AI's integration into healthcare should not replace the physicians' intelligence but rather augment it. Thus, we should aim for AI-assisted and not AI-driven clinical practice[14]. Finally, AI systems should be applied in healthcare along with equally advanced evaluation systems, which properly assess their ramifications within the clinical practice environment, investigate their unintended consequences, and, most notably, evaluate their impact on patient outcomes[163]. An example of an evaluative initiative is the Digital Health Innovation Action Plan launched by the FDA that aims to assess medical software under

development based on five excellence criteria: product quality, patient safety, clinical responsibility, cybersecurity responsibility, and proactive culture[199].

Despite being on the front line for several decades, AI still comes short of delivering the presumable solutions to long-standing healthcare problems. AI winters, where AI funding is significantly reduced, demonstrate the frustration that AI is not evolving at a pace investors would feel comfortable sustaining the funding[1]. AI may seem like a field that constantly overpromises but usually underdelivers, given its current impact on healthcare. In this review, we have highlighted the current challenges that we believe restrain the extensive application of AI in healthcare in an attempt to explore ways to overcome them.

Nevertheless, even at a slower pace, AI will eventually "infiltrate" not only the hospital setting but rather the healthcare industry as a whole, from central healthcare facilities to private practice and telemedicine, from medical schools and teaching hospitals to pharmaceutical companies, and even healthcare policy-makers in government. AI will become a reality; everyone will have to conform with, to avoid becoming obsolete. Therefore, sooner or later, physicians will have to engage with the field of AI by necessity. The role of physicians in this upcoming "revolution", however, remains to be seen. Will we be shaping this "revolution", or will we be mere observers?

Limitations

Finally, our review has few limitations. First, our study is a narrative review, and despite our best effort to follow a carefully designed search strategy to provide a comprehensive review of the current literature, our study is prone to selection bias due to its nature. Second, each study described in our study carries its own risk of bias and faces several limitations described in each study. We did not use any bias risk assessment tool to systematically evaluate if a study was worth inclusion rather than the followed search strategy and our judgment. Again this makes our review prone to selection bias. In addition, the majority of our included studies were performed in silico, and the performances reported could deviate substantially when the models are applied in a real clinical setting. Finally, with few exceptions, the majority of results report a superior or at least equivalent performance of AI/ML-based algorithms compared to the performance of conventional statistic models, widely used staging systems, extensively investigated and validated scores, and physician knowledge and reasoning, which have dominated decision-making in healthcare for decades. Thus, at least in the concept of publication bias, our findings should be interpreted with caution.

REFERENCES

1   **Jayabalasingham B**.   ArtificiaI Intelligence: How knowledge is created, transferred, and used. Elsevier, 2019

2   **Russell SJ**.   Artificial intelligence: a modern approach. Harlow: Pearson, 2016

3   **Cruz JA**, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007; **2**: 59-77 [PMID: 19458758]

4   **Raghupathi W**, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014; **2**: 3 [PMID: 25825667 DOI: 10.1186/2047-2501-2-3]

5   **Pastorino R**, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *Eur J Public Health* 2019; **29**: 23-27 [PMID: 31738444 DOI: 10.1093/eurpub/ckz168]

6   **Dash S**, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data* 2019; **6**: 54 [DOI: 10.1186/s40537-019-0217-0]

7   **Statista**.   Healthcare data volume globally 2020 forecast. [cited 20 December 2020]. Available from: https://www.statista.com/statistics/1037970/global-healthcare-data-volume/

8   **McAfee A**, Brynjolfsson E. Big data: the management revolution. *Harv Bus Rev* 2012; **90**: 60-66, 68, 128 [PMID: 23074865]

9   **Beyer MA**, Laney D. The importance of 'big data: a definition. *CT Gart* 2012; 2014

10   Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. Stead WW, Lin HS, editors. Washington (DC): National Academies Press (US), 2009

11   **Matheny ME**, Whicher D, Thadaney Israni S. Artificial Intelligence in Health Care: A Report From the National Academy of Medicine. *JAMA* 2020; **323**: 509-510 [PMID: 31845963 DOI: 10.1001/jama.2019.21579]

12   **Sugawara E**, Nikaido H. Properties of AdeABC and AdeIJK efflux systems of Acinetobacter baumannii compared with those of the AcrAB-TolC system of Escherichia coli. *Antimicrob Agents Chemother* 2014; **58**: 7250-7257 [PMID: 25246403 DOI: 10.1128/AAC.03728-14]

13   **Lai-Hung Wong G**, Sung JJ. The emerging role of big data in gastroenterology and hepatology. *J*

Gastroenterol Hepatol 2019; **34**: 307-308 [PMID: 30698322 DOI: 10.1111/jgh.14606]

14   **Ruffle JK**, Farmer AD, Aziz Q. Artificial Intelligence-Assisted Gastroenterology- Promises and Pitfalls. *Am J Gastroenterol* 2019; **114**: 422-428 [PMID: 30315284 DOI: 10.1038/s41395-018-0268-4]

15   **Loftus TJ**, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, Rashidi P, Upchurch GR Jr, Bihorac A. Artificial Intelligence and Surgical Decision-making. *JAMA Surg* 2020; **155**: 148-158 [PMID: 31825465 DOI: 10.1001/jamasurg.2019.4917]

16   **Rumsfeld JS**, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol* 2016; **13**: 350-359 [PMID: 27009423 DOI: 10.1038/nrcardio.2016.42]

17   **Le Berre C**, Sandborn WJ, Aridhi S, Devignes MD, Fournier L, Smaïl-Tabbone M, Danese S, Peyrin-Biroulet L. Application of Artificial Intelligence to Gastroenterology and Hepatology. *Gastroenterology* 2020; **158**: 76-94 [PMID: 31593701 DOI: 10.1053/j.gastro.2019.08.058]

18   **Yang YJ**, Bang CS. Application of artificial intelligence in gastroenterology. *World J Gastroenterol* 2019; **25**: 1666-1683 [PMID: 31011253 DOI: 10.3748/wjg.v25.i14.1666]

19   **Huang S**, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics* 2018; **15**: 41-51 [PMID: 29275361 DOI: 10.21873/cgp.20063]

20   **Deo RC**. Machine Learning in Medicine. *Circulation* 2015; **132**: 1920-1930 [PMID: 26572668 DOI: 10.1161/CIRCULATIONAHA.115.001593]

21   **Noble WS**. What is a support vector machine? *Nat Biotechnol* 2006; **24**: 1565-1567 [PMID: 17160063 DOI: 10.1038/nbt1206-1565]

22   **Guo CY**, Chou YC. A novel machine learning strategy for model selections - Stepwise Support Vector Machine (StepSVM). *PLoS One* 2020; **15**: e0238384 [PMID: 32853243 DOI: 10.1371/journal.pone.0238384]

23   **Aizerman M**, Braverman E, Rozonoer L. Theoretical foundations of the potential function method in pattern recognition learning. *Autom Remote Control* 1964; **25**: 821-837

24   **Kriegeskorte N**, Golan T. Neural network models and deep learning. *Curr Biol* 2019; **29**: R231-R236 [PMID: 30939301 DOI: 10.1016/j.cub.2019.02.034]

25   **LeCun Y**, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436-444 [PMID: 26017442 DOI: 10.1038/nature14539]

26   **Anwar SM**, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical Image Analysis using Convolutional Neural Networks: A Review. *J Med Syst* 2018; **42**: 226 [PMID: 30298337 DOI: 10.1007/s10916-018-1088-1]

27   **Leung WK**, Cheung KS, Li B, Law SYK, Lui TKL. Applications of machine learning models in the prediction of gastric cancer risk in patients after Helicobacter pylori eradication. *Aliment Pharmacol Ther* 2021; **53**: 864-872 [PMID: 33486805 DOI: 10.1111/apt.16272]

28   **Nakahira H**, Ishihara R, Aoyama K, Kono M, Fukuda H, Shimamoto Y, Nakagawa K, Ohmori M, Iwatsubo T, Iwagami H, Matsuno K, Inoue S, Matsuura N, Shichijo S, Maekawa A, Kanesaka T, Yamamoto S, Takeuchi Y, Higashino K, Uedo N, Matsunaga T, Tada T. Stratification of gastric cancer risk using a deep neural network. *JGH Open* 2020; **4**: 466-471 [PMID: 32514455 DOI: 10.1002/jgh3.12281]

29   **Taninaga J**, Nishiyama Y, Fujibayashi K, Gunji T, Sasabe N, Iijima K, Naito T. Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study. *Sci Rep* 2019; **9**: 12384 [PMID: 31455831 DOI: 10.1038/s41598-019-48769-y]

30   **Kinar Y**, Kalkstein N, Akiva P, Levin B, Half EE, Goldshtein I, Chodick G, Shalev V. Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *J Am Med Inform Assoc* 2016; **23**: 879-890 [PMID: 26911814 DOI: 10.1093/jamia/ocv195]

31   **Goshen R**, Choman E, Ran A, Muller E, Kariv R, Chodick G, Ash N, Narod S, Shalev V. Computer-Assisted Flagging of Individuals at High Risk of Colorectal Cancer in a Large Health Maintenance Organization Using the ColonFlag Test. *JCO Clin Cancer Inform* 2018; **2**: 1-8 [PMID: 30652563 DOI: 10.1200/CCI.17.00130]

32   **Mosquera-Lopez C**, Agaian S, Velez-Hoyos A, Thompson I. Computer-Aided Prostate Cancer Diagnosis From Digitized Histopathology: A Review on Texture-Based Systems. *IEEE Rev Biomed Eng* 2015; **8**: 98-113 [PMID: 25055385 DOI: 10.1109/RBME.2014.2340401]

33   **Takiyama H**, Ozawa T, Ishihara S, Fujishiro M, Shichijo S, Nomura S, Miura M, Tada T. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. *Sci Rep* 2018; **8**: 7497 [PMID: 29760397 DOI: 10.1038/s41598-018-25842-6]

34   **Pace F**, Buscema M, Dominici P, Intraligi M, Baldi F, Cestari R, Passaretti S, Bianchi Porro G, Grossi E. Artificial neural networks are able to recognize gastro-oesophageal reflux disease patients solely on the basis of clinical data. *Eur J Gastroenterol Hepatol* 2005; **17**: 605-610 [PMID: 15879721 DOI: 10.1097/00042737-200506000-00003]

35   **de Groof AJ**, Struyvenberg MR, van der Putten J, van der Sommen F, Fockens KN, Curvers WL, Zinger S, Pouw RE, Coron E, Baldaque-Silva F, Pech O, Weusten B, Meining A, Neuhaus H, Bisschops R, Dent J, Schoon EJ, de With PH, Bergman JJ. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology* 2020; **158**: 915-929.e4

[PMID: 31759929 DOI: 10.1053/j.gastro.2019.11.030]

36   **van der Sommen F**, Zinger S, Curvers WL, Bisschops R, Pech O, Weusten BL, Bergman JJ, de With PH, Schoon EJ. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy* 2016; **48**: 617-624 [PMID: 27100718 DOI: 10.1055/s-0042-105284]

37   **Struyvenberg MR**, van der Sommen F, Swager AF, de Groof AJ, Rikos A, Schoon EJ, Bergman JJ, de With PHN, Curvers WL. Improved Barrett's neoplasia detection using computer-assisted multiframe analysis of volumetric laser endomicroscopy. *Dis Esophagus* 2020; **33**: doz065 [PMID: 31364700 DOI: 10.1093/dote/doz065]

38   **Swager AF**, van der Sommen F, Klomp SR, Zinger S, Meijer SL, Schoon EJ, Bergman JJGHM, de With PH, Curvers WL. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017; **86**: 839-846 [PMID: 28322771 DOI: 10.1016/j.gie.2017.03.011]

39   **Kumagai Y**, Takubo K, Kawada K, Aoyama K, Endo Y, Ozawa T, Hirasawa T, Yoshio T, Ishihara S, Fujishiro M, Tamaru JI, Mochiki E, Ishida H, Tada T. Diagnosis using deep-learning artificial intelligence based on the endocytoscopic observation of the esophagus. *Esophagus* 2019; **16**: 180-187 [PMID: 30547352 DOI: 10.1007/s10388-018-0651-7]

40   **Zheng W**, Zhang X, Kim JJ, Zhu X, Ye G, Ye B, Wang J, Luo S, Li J, Yu T, Liu J, Hu W, Si J. High Accuracy of Convolutional Neural Network for Evaluation of Helicobacter pylori Infection Based on Endoscopic Images: Preliminary Experience. *Clin Transl Gastroenterol* 2019; **10**: e00109 [PMID: 31833862 DOI: 10.14309/ctg.0000000000000109]

41   **Nakashima H**, Kawahira H, Kawachi H, Sakaki N. Artificial intelligence diagnosis of *Helicobacter pylori* infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. *Ann Gastroenterol* 2018; **31**: 462-468 [PMID: 29991891 DOI: 10.20524/aog.2018.0269]

42   **Itoh T**, Kawahira H, Nakashima H, Yata N. Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images. *Endosc Int Open* 2018; **6**: E139-E144 [PMID: 29399610 DOI: 10.1055/s-0043-120830]

43   **Shichijo S**, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, Takiyama H, Tanimoto T, Ishihara S, Matsuo K, Tada T. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. *EBioMedicine* 2017; **25**: 106-111 [PMID: 29056541 DOI: 10.1016/j.ebiom.2017.10.014]

44   **Mohan BP**, Khan SR, Kassab LL, Ponnada S, Mohy-Ud-Din N, Chandan S, Dulai PS, Kochhar GS. Convolutional neural networks in the computer-aided diagnosis of *Helicobacter pylori* infection and non-causal comparison to physician endoscopists: a systematic review with meta-analysis. *Ann Gastroenterol* 2021; **34**: 20-25 [PMID: 33414617 DOI: 10.20524/aog.2020.0542]

45   **Kanesaka T**, Lee TC, Uedo N, Lin KP, Chen HZ, Lee JY, Wang HP, Chang HT. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. *Gastrointest Endosc* 2018; **87**: 1339-1344 [PMID: 29225083 DOI: 10.1016/j.gie.2017.11.029]

46   **Hirasawa T**, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, Ohnishi T, Fujishiro M, Matsuo K, Fujisaki J, Tada T. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018; **21**: 653-660 [PMID: 29335825 DOI: 10.1007/s10120-018-0793-2]

47   **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**: e0244869 [PMID: 33382829 DOI: 10.1371/journal.pone.0244869]

48   **Tenório JM**, Hummel AD, Cohrs FM, Sdepanian VL, Pisa IT, de Fátima Marin H. Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease. *Int J Med Inform* 2011; **80**: 793-802 [PMID: 21917512 DOI: 10.1016/j.ijmedinf.2011.08.001]

49   **Caetano Dos Santos FL**, Michalek IM, Laurila K, Kaukinen K, Hyttinen J, Lindfors K. Automatic classification of IgA endomysial antibody test for celiac disease: a new method deploying machine learning. *Sci Rep* 2019; **9**: 9217 [PMID: 31239486 DOI: 10.1038/s41598-019-45679-x]

50   **Hujoel IA**, Murphree DH Jr, Van Dyke CT, Choung RS, Sharma A, Murray JA, Rubio-Tapia A. Machine Learning in Detection of Undiagnosed Celiac Disease. *Clin Gastroenterol Hepatol* 2018; **16**: 1354-1355.e1 [PMID: 29253540 DOI: 10.1016/j.cgh.2017.12.022]

51   **Manandhar I**, Alimadadi A, Aryal S, Munroe PB, Joe B, Cheng X. Gut microbiome-based supervised machine learning for clinical diagnosis of inflammatory bowel diseases. *Am J Physiol Gastrointest Liver Physiol* 2021 epub ahead of print [PMID: 33439104 DOI: 10.1152/ajpgi.00360.2020]

52   **Wei Z**, Wang W, Bradfield J, Li J, Cardinale C, Frackelton E, Kim C, Mentch F, Van Steen K, Visscher PM, Baldassano RN, Hakonarson H; International IBD Genetics Consortium. 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 [PMID: 23731541 DOI: 10.1016/j.ajhg.2013.05.002]

53   **Mossotto E**, Ashton JJ, Coelho T, Beattie RM, MacArthur BD, Ennis S. Classification of Paediatric Inflammatory Bowel Disease using Machine Learning. *Sci Rep* 2017; **7**: 2427 [PMID: 28546534 DOI: 10.1038/s41598-017-02606-2]

54   **Charisis VS**, Hadjileontiadis LJ. Potential of hybrid adaptive filtering in inflammatory lesion detection from capsule endoscopy images. *World J Gastroenterol* 2016; **22**: 8641-8657 [PMID: 27818583 DOI: 10.3748/wjg.v22.i39.8641]

55   **Jebarani W**, Daisy VJ. Assessment of Crohn's disease lesions in Wireless Capsule Endoscopy images using SVM based classification. Signal Processing Image Processing & Pattern Recognition (ICSIPR); Semantic Scholar, 2013: 303-307 [DOI: 10.1109/ICSIPR.2013.6497945]

56   **Girgis HZ**, Mitchell BR, Dassopoulos T, Mullin G, Hager G. An intelligent system to detect Crohn's disease inflammation in Wireless Capsule Endoscopy videos. 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2010 Apr 14-17; Rotterdam, Netherlands. IEEE, 2010 [DOI: 10.1109/ISBI.2010.5490253]

57   **Kumar R**, Zhao Q, Seshamani S, Mullin G, Hager G, Dassopoulos T. Assessment of Crohn's disease lesions in wireless capsule endoscopy images. *IEEE Trans Biomed Eng* 2012; **59**: 355-362 [PMID: 22020661 DOI: 10.1109/TBME.2011.2172438]

58   **Xia J**, Xia T, Pan J, Gao F, Wang S, Qian YY, Wang H, Zhao J, Jiang X, Zou WB, Wang YC, Zhou W, Li ZS, Liao Z. Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy. *Gastrointest Endosc* 2021; **93**: 133-139.e4 [PMID: 32470426 DOI: 10.1016/j.gie.2020.05.027]

59   **Seguí S**, Drozdzal M, Pascual G, Radeva P, Malagelada C, Azpiroz F, Vitrià J. Generic feature learning for wireless capsule endoscopy analysis. *Comput Biol Med* 2016; **79**: 163-172 [PMID: 27810622 DOI: 10.1016/j.compbiomed.2016.10.011]

60   **Park J**, Hwang Y, Nam JH, Oh DJ, Kim KB, Song HJ, Kim SH, Kang SH, Jung MK, Jeong Lim Y. Artificial intelligence that determines the clinical significance of capsule endoscopy images can increase the efficiency of reading. *PLoS One* 2020; **15**: e0241474 [PMID: 33119718 DOI: 10.1371/journal.pone.0241474]

61   **Hwang Y**, Lee HH, Park C, Tama BA, Kim JS, Cheung DY, Chung WC, Cho YS, Lee KM, Choi MG, Lee S, Lee BI. Improved classification and localization approach to small bowel capsule endoscopy using convolutional neural network. *Dig Endosc* 2021; **33**: 598-607 [PMID: 32640059 DOI: 10.1111/den.13787]

62   **Otani K**, Nakada A, Kurose Y, Niikura R, Yamada A, Aoki T, Nakanishi H, Doyama H, Hasatani K, Sumiyoshi T, Kitsuregawa M, Harada T, Koike K. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. *Endoscopy* 2020; **52**: 786-791 [PMID: 32557474 DOI: 10.1055/a-1167-8157]

63   **Yuan Y**, Wang J, Li B, Meng MQ. Saliency based ulcer detection for wireless capsule endoscopy diagnosis. *IEEE Trans Med Imaging* 2015; **34**: 2046-2057 [PMID: 25850085 DOI: 10.1109/TMI.2015.2418534]

64   **Karargyris A**, Bourbakis N. Identification of ulcers in Wireless Capsule Endoscopy videos. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. 2009 Jun 28-Jul 1; Boston, MA, USA. IEEE, 2009 [DOI: 10.1109/ISBI.2009.5193107]

65   **He JY**, Wu X, Jiang YG, Peng Q, Jain R. Hookworm Detection in Wireless Capsule Endoscopy Images With Deep Learning. *IEEE Trans Image Process* 2018; **27**: 2379-2392 [PMID: 29470172 DOI: 10.1109/TIP.2018.2801119]

66   **Leenhardt R**, Vasseur P, Li C, Saurin JC, Rahmi G, Cholet F, Becq A, Marteau P, Histace A, Dray X; CAD-CAP Database Working Group. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; **89**: 189-194 [PMID: 30017868 DOI: 10.1016/j.gie.2018.06.036]

67   **Zhou T**, Han G, Li BN, Lin Z, Ciaccio EJ, Green PH, Qin J. Quantitative analysis of patients with celiac disease by video capsule endoscopy: A deep learning method. *Comput Biol Med* 2017; **85**: 1-6 [PMID: 28412572 DOI: 10.1016/j.compbiomed.2017.03.031]

68   **Yamada A**, Niikura R, Otani K, Aoki T, Koike K. Automatic detection of colorectal neoplasia in wireless colon capsule endoscopic images using a deep convolutional neural network. *Endoscopy* 2021; **53**: 832-836 [PMID: 32947623 DOI: 10.1055/a-1266-1066]

69   **Wang P**, Xiao X, Glissen Brown JR, Berzin TM, Tu M, Xiong F, Hu X, Liu P, Song Y, Zhang D, Yang X, Li L, He J, Yi X, Liu J, Liu X. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat Biomed Eng* 2018; **2**: 741-748 [PMID: 31015647 DOI: 10.1038/s41551-018-0301-3]

70   **Misawa M**, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, Ogawa Y, Maeda Y, Takeda K, Ichimasa K, Nakamura H, Yagawa Y, Toyoshima N, Ogata N, Kudo T, Hisayuki T, Hayashi T, Wakamura K, Baba T, Ishida F, Itoh H, Roth H, Oda M, Mori K. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. *Gastroenterology* 2018; **154**: 2027-2029.e3 [PMID: 29653147 DOI: 10.1053/j.gastro.2018.04.003]

71   **Urban G**, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, Baldi P. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. *Gastroenterology* 2018; **155**: 1069-1078.e8 [PMID: 29928897 DOI: 10.1053/j.gastro.2018.06.037]

72   **Ozawa T**, Ishihara S, Fujishiro M, Kumagai Y, Shichijo S, Tada T. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. *Therap Adv Gastroenterol* 2020; **13**: 1756284820910659 [PMID: 32231710 DOI: 10.1177/1756284820910659]

73   **Mori Y**, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, Ohtsuka K, Urushibara F, Kataoka S, Ogawa Y, Maeda Y, Takeda K, Nakamura H, Ichimasa K, Kudo T, Hayashi T, Wakamura K, Ishida F, Inoue H, Itoh H, Oda M, Mori K. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. *Ann Intern Med* 2018; **169**: 357-366 [PMID: 30105375 DOI: 10.7326/M18-0249]

74   **Tischendorf JJ**, Gross S, Winograd R, Hecker H, Auer R, Behrens A, Trautwein C, Aach T, Stehle

T. Computer-aided classification of colorectal polyps based on vascular patterns: a pilot study. *Endoscopy* 2010; **42**: 203-207 [PMID: 20101564 DOI: 10.1055/s-0029-1243861]

75   **Gross S**, Trautwein C, Behrens A, Winograd R, Palm S, Lutz HH, Schirin-Sokhan R, Hecker H, Aach T, Tischendorf JJ. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. *Gastrointest Endosc* 2011; **74**: 1354-1359 [PMID: 22000791 DOI: 10.1016/j.gie.2011.08.001]

76   **Kominami Y**, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, Tamaki T, Koide T, Kaneda K, Chayama K. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc* 2016; **83**: 643-649 [PMID: 26264431 DOI: 10.1016/j.gie.2015.08.004]

77   **Misawa M**, Kudo SE, Mori Y, Nakamura H, Kataoka S, Maeda Y, Kudo T, Hayashi T, Wakamura K, Miyachi H, Katagiri A, Baba T, Ishida F, Inoue H, Nimura Y, Mori K. Characterization of Colorectal Lesions Using a Computer-Aided Diagnostic System for Narrow-Band Imaging Endocytoscopy. *Gastroenterology* 2016; **150**: 1531-1532.e3 [PMID: 27072671 DOI: 10.1053/j.gastro.2016.04.004]

78   **Takeda K**, Kudo SE, Mori Y, Misawa M, Kudo T, Wakamura K, Katagiri A, Baba T, Hidaka E, Ishida F, Inoue H, Oda M, Mori K. Accuracy of diagnosing invasive colorectal cancer using computer-aided endocytoscopy. *Endoscopy* 2017; **49**: 798-802 [PMID: 28472832 DOI: 10.1055/s-0043-105486]

79   **Chen PJ**, Lin MC, Lai MJ, Lin JC, Lu HH, Tseng VS. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. *Gastroenterology* 2018; **154**: 568-575 [PMID: 29042219 DOI: 10.1053/j.gastro.2017.10.010]

80   **Komeda Y**, Handa H, Watanabe T, Nomura T, Kitahashi M, Sakurai T, Okamoto A, Minami T, Kono M, Arizumi T, Takenaka M, Hagiwara S, Matsui S, Nishida N, Kashida H, Kudo M. Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. *Oncology* 2017; **93** Suppl 1: 30-34 [PMID: 29258081 DOI: 10.1159/000481227]

81   **Byrne MF**, Chapados N, Soudan F, Oertel C, Linares Pérez M, Kelly R, Iqbal N, Chandelier F, Rex DK. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019; **68**: 94-100 [PMID: 29066576 DOI: 10.1136/gutjnl-2017-314547]

82   **ASGE Technology Committee**, Abu Dayyeh BK, Thosani N, Konda V, Wallace MB, Rex DK, Chauhan SS, Hwang JH, Komanduri S, Manfredi M, Maple JT, Murad FM, Siddiqui UD, Banerjee S. ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015; **81**: 502.e1-502.e16 [PMID: 25597420 DOI: 10.1016/j.gie.2014.12.022]

83   **Rex DK**, Kahi C, O'Brien M, Levin TR, Pohl H, Rastogi A, Burgart L, Imperiale T, Ladabaum U, Cohen J, Lieberman DA. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2011; **73**: 419-422 [PMID: 21353837 DOI: 10.1016/j.gie.2011.01.023]

84   **East JE**, Vleugels JL, Roelandt P, Bhandari P, Bisschops R, Dekker E, Hassan C, Horgan G, Kiesslich R, Longcroft-Wheaton G, Wilson A, Dumonceau JM. Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) Technology Review. *Endoscopy* 2016; **48**: 1029-1045 [PMID: 27711949 DOI: 10.1055/s-0042-118087]

85   **de Almeida Thomaz V**, Sierra-Franco CA, Raposo AB. Training data enhancements for improving colonic polyp detection using deep convolutional neural networks. *Artif Intell Med* 2021; **111**: 101988 [PMID: 33461694 DOI: 10.1016/j.artmed.2020.101988]

86   **Rogers B**, Samanta S, Ghobadi K, Patel A, Savarino E, Roman S, Sifrim D, Gyawali CP. Artificial intelligence automates and augments baseline impedance measurements from pH-impedance studies in gastroesophageal reflux disease. *J Gastroenterol* 2021; **56**: 34-41 [PMID: 33151406 DOI: 10.1007/s00535-020-01743-2]

87   **Zhu Y**, Wang QC, Xu MD, Zhang Z, Cheng J, Zhong YS, Zhang YQ, Chen WF, Yao LQ, Zhou PH, Li QL. 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 [PMID: 30452913 DOI: 10.1016/j.gie.2018.11.011]

88   **Kubota K**, Kuroda J, Yoshida M, Ohta K, Kitajima M. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. *Surg Endosc* 2012; **26**: 1485-1489 [PMID: 22083334 DOI: 10.1007/s00464-011-2036-z]

89   **Yamashita R**, Long J, Longacre T, Peng L, Berry G, Martin B, Higgins J, Rubin DL, Shen J. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol* 2021; **22**: 132-141 [PMID: 33387492 DOI: 10.1016/S1470-2045(20)30535-0]

90   **Ichimasa K**, Kudo SE, Mori Y, Misawa M, Matsudaira S, Kouyama Y, Baba T, Hidaka E, Wakamura K, Hayashi T, Kudo T, Ishigaki T, Yagawa Y, Nakamura H, Takeda K, Haji A, Hamatani S, Mori K, Ishida F, Miyachi H. Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer. *Endoscopy* 2018; **50**: 230-240 [PMID: 29272905 DOI: 10.1055/s-0043-122385]

91   **Levi R**, Carli F, Arévalo AR, Altinel Y, Stein DJ, Naldini MM, Grassi F, Zanoni A, Finkelstein S,

Vieira SM, Sousa J, Barbieri R, Celi LA. Artificial intelligence-based prediction of transfusion in the intensive care unit in patients with gastrointestinal bleeding. *BMJ Health Care Inform* 2021; **28**: e100245 [PMID: 33455913 DOI: 10.1136/bmjhci-2020-100245]

92   **Chu A**, Ahn H, Halwan B, Kalmin B, Artifon EL, Barkun A, Lagoudakis MG, Kumar A. A decision support system to facilitate management of patients with acute gastrointestinal bleeding. *Artif Intell Med* 2008; **42**: 247-259 [PMID: 18063351 DOI: 10.1016/j.artmed.2007.10.003]

93   **Das A**, Ben-Menachem T, Farooq FT, Cooper GS, Chak A, Sivak MV Jr, Wong RC. Artificial neural network as a predictive instrument in patients with acute nonvariceal upper gastrointestinal hemorrhage. *Gastroenterology* 2008; **134**: 65-74 [PMID: 18061180 DOI: 10.1053/j.gastro.2007.10.037]

94   **Augustin S**, Muntaner L, Altamirano JT, González A, Saperas E, Dot J, Abu-Suboh M, Armengol JR, Malagelada JR, Esteban R, Guardia J, Genescà J. Predicting early mortality after acute variceal hemorrhage based on classification and regression tree analysis. *Clin Gastroenterol Hepatol* 2009; **7**: 1347-1354 [PMID: 19699816 DOI: 10.1016/j.cgh.2009.08.011]

95   **Loftus TJ**, Brakenridge SC, Croft CA, Smith RS, Efron PA, Moore FA, Mohr AM, Jordan JR. Neural network prediction of severe lower intestinal bleeding and the need for surgical intervention. *J Surg Res* 2017; **212**: 42-47 [PMID: 28550920 DOI: 10.1016/j.jss.2016.12.032]

96   **Ayaru L**, Ypsilantis PP, Nanapragasam A, Choi RC, Thillanathan A, Min-Ho L, Montana G. Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting. *PLoS One* 2015; **10**: e0132485 [PMID: 26172121 DOI: 10.1371/journal.pone.0132485]

97   **Yang HX**, Feng W, Wei JC, Zeng TS, Li ZD, Zhang LJ, Lin P, Luo RZ, He JH, Fu JH. Support vector machine-based nomogram predicts postoperative distant metastasis for patients with oesophageal squamous cell carcinoma. *Br J Cancer* 2013; **109**: 1109-1116 [PMID: 23942069 DOI: 10.1038/bjc.2013.379]

98   **Sato F**, Shimada Y, Selaru FM, Shibata D, Maeda M, Watanabe G, Mori Y, Stass SA, Imamura M, Meltzer SJ. Prediction of survival in patients with esophageal carcinoma using artificial neural networks. *Cancer* 2005; **103**: 1596-1605 [PMID: 15751017 DOI: 10.1002/cncr.20938]

99   **Zhou C**, Hu J, Wang Y, Ji MH, Tong J, Yang JJ, Xia H. A machine learning-based predictor for the identification of the recurrence of patients with gastric cancer after operation. *Sci Rep* 2021; **11**: 1571 [PMID: 33452440 DOI: 10.1038/s41598-021-81188-6]

100   **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 [PMID: 25976931 DOI: 10.1007/s00384-015-2250-6]

101   **Hardalaç F**, Başaranoğlu M, Yüksel M, Kutbay U, Kaplan M, Özderin Özin Y, Kılıç ZM, Demirbağ AE, Coşkun O, Aksoy A, Gangarapu V, Örmeci N, Kayaçetin E. The rate of mucosal healing by azathioprine therapy and prediction by artificial systems. *Turk J Gastroenterol* 2015; **26**: 315-321 [PMID: 26039001 DOI: 10.5152/tjg.2015.0199]

102   **Takayama T**, Okamoto S, Hisamatsu T, Naganuma M, Matsuoka K, Mizuno S, Bessho R, Hibi T, Kanai T. Computer-Aided Prediction of Long-Term Prognosis of Patients with Ulcerative Colitis after Cytoapheresis Therapy. *PLoS One* 2015; **10**: e0131197 [PMID: 26111148 DOI: 10.1371/journal.pone.0131197]

103   **Lyles T**, Elliott A, Rockey DC. A risk scoring system to predict in-hospital mortality in patients with cirrhosis presenting with upper gastrointestinal bleeding. *J Clin Gastroenterol* 2014; **48**: 712-720 [PMID: 24172184 DOI: 10.1097/MCG.0000000000000014]

104   **Grossi E**, Marmo R, Intraligi M, Buscema M. Artificial Neural Networks for Early Prediction of Mortality in Patients with Non Variceal Upper GI Bleeding (UGIB). *Biomed Inform Insights* 2008; **1**: 7-19 [PMID: 27429551 DOI: 10.4137/bii.s814]

105   **Rotondano G**, Cipolletta L, Grossi E, Koch M, Intraligi M, Buscema M, Marmo R; Italian Registry on Upper Gastrointestinal Bleeding (Progetto Nazionale Emorragie Digestive). Artificial neural networks accurately predict mortality in patients with nonvariceal upper GI bleeding. *Gastrointest Endosc* 2011; **73**: 218-226 [PMID: 21295635 DOI: 10.1016/j.gie.2010.10.006]

106   **Shi R**, Chen W, Yang B, Qu J, Cheng Y, Zhu Z, Gao Y, Wang Q, Liu Y, Li Z, Qu X. Prediction of KRAS, NRAS and BRAF status in colorectal cancer patients with liver metastasis using a deep artificial neural network based on radiomics and semantic features. *Am J Cancer Res* 2020; **10**: 4513-4526 [PMID: 33415015]

107   **Kang J**, Choi YJ, Kim IK, Lee HS, Kim H, Baik SH, Kim NK, Lee KY. LASSO-Based Machine Learning Algorithm for Prediction of Lymph Node Metastasis in T1 Colorectal Cancer. *Cancer Res Treat* 2021; **53**: 773-783 [PMID: 33421980 DOI: 10.4143/crt.2020.974]

108   **Goldman O**, Ben-Assuli O, Rogowski O, Zeltser D, Shapira I, Berliner S, Zelber-Sagi S, Shenhar-Tsarfaty S. Non-alcoholic Fatty Liver and Liver Fibrosis Predictive Analytics: Risk Prediction and Machine Learning Techniques for Improved Preventive Medicine. *J Med Syst* 2021; **45**: 22 [PMID: 33426569 DOI: 10.1007/s10916-020-01693-5]

109   **Yip TC**, Ma AJ, Wong VW, Tse YK, Chan HL, Yuen PC, Wong GL. 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 [PMID: 28585725 DOI: 10.1111/apt.14172]

110   **Ma H**, Xu CF, Shen Z, Yu CH, Li YM. 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 [PMID: 30402478 DOI: 10.1155/2018/4304376]

111   **Sowa JP**, Heider D, Bechmann LP, Gerken G, Hoffmann D, Canbay A. Novel algorithm for non-

invasive assessment of fibrosis in NAFLD. *PLoS One* 2013; **8**: e62439 [PMID: 23638085 DOI: 10.1371/journal.pone.0062439]

112   **Canbay A**, Kälsch J, Neumann U, Rau M, Hohenester S, Baba HA, Rust C, Geier A, Heider D, Sowa JP. Non-invasive assessment of NAFLD as systemic disease-A machine learning perspective. *PLoS One* 2019; **14**: e0214436 [PMID: 30913263 DOI: 10.1371/journal.pone.0214436]

113   **Fialoke S**, Malarstig A, Miller MR, Dumitriu A. 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 [PMID: 30815083]

114   **Sowa JP**, Atmaca Ö, Kahraman A, Schlattjan M, Lindner M, Sydor S, Scherbaum N, Lackner K, Gerken G, Heider D, Arteel GE, Erim Y, Canbay A. Non-invasive separation of alcoholic and non-alcoholic liver disease with predictive modeling. *PLoS One* 2014; **9**: e101444 [PMID: 24988316 DOI: 10.1371/journal.pone.0101444]

115   **Wei R**, Wang J, Wang X, Xie G, Wang Y, Zhang H, Peng CY, Rajani C, Kwee S, Liu P, Jia W. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. *EBioMedicine* 2018; **35**: 124-132 [PMID: 30100397 DOI: 10.1016/j.ebiom.2018.07.041]

116   **Wang D**, Wang Q, Shan F, Liu B, Lu C. 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 [PMID: 20735842 DOI: 10.1186/1471-2334-10-251]

117   **Raoufy MR**, Vahdani P, Alavian SM, Fekri S, Eftekhari P, Gharibzadeh S. A novel method for diagnosing cirrhosis in patients with chronic hepatitis B: artificial neural network approach. *J Med Syst* 2011; **35**: 121-126 [PMID: 20703578 DOI: 10.1007/s10916-009-9348-8]

118   **Piscaglia F**, Cucchetti A, Benlloch S, Vivarelli M, Berenguer J, Bolondi L, Pinna AD, Berenguer M. 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 [PMID: 17099373 DOI: 10.1097/01.meg.0000243885.55562.7e]

119   **Hashem S**, Esmat G, Elakel W, Habashy S, Raouf SA, Elhefnawi M, Eladawy M, ElHefnawi M. Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients. *IEEE/ACM Trans Comput Biol Bioinform* 2018; **15**: 861-868 [PMID: 28391204 DOI: 10.1109/TCBB.2017.2690848]

120   **Ioannou GN**, Tang W, Beste LA, Tincopa MA, Su GL, Van T, Tapper EB, Singal AG, Zhu J, Waljee AK. Assessment of a Deep Learning Model to Predict Hepatocellular Carcinoma in Patients With Hepatitis C Cirrhosis. *JAMA Netw Open* 2020; **3**: e2015626 [PMID: 32870314 DOI: 10.1001/jamanetworkopen.2020.15626]

121   **Emu M**, Kamal FB, Choudhury S, Alves de Oliveira TE. Assisting the Non-invasive Diagnosis of Liver Fibrosis Stages using Machine Learning Methods. *Annu Int Conf IEEE Eng Med Biol Soc* 2020; **2020**: 5382-5387 [PMID: 33019198 DOI: 10.1109/EMBC44109.2020.9176542]

122   **Choi KJ**, Jang JK, Lee SS, Sung YS, Shim WH, Kim HS, Yun J, Choi JY, Lee Y, Kang BK, Kim JH, Kim SY, Yu ES. Development and Validation of a Deep Learning System for Staging Liver Fibrosis by Using Contrast Agent-enhanced CT Images in the Liver. *Radiology* 2018; **289**: 688-697 [PMID: 30179104 DOI: 10.1148/radiol.2018180763]

123   **Kuppili V**, Biswas M, Sreekumar A, Suri HS, Saba L, Edla DR, Marinho RT, Sanches JM, Suri JS. Extreme Learning Machine Framework for Risk Stratification of Fatty Liver Disease Using Ultrasound Tissue Characterization. *J Med Syst* 2017; **41**: 152 [PMID: 28836045 DOI: 10.1007/s10916-017-0797-1]

124   **Gatos I**, Tsantis S, Spiliopoulos S, Karnabatidis D, Theotokas I, Zoumpoulis P, Loupas T, Hazle JD, Kagadis GC. A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography. *Ultrasound Med Biol* 2017; **43**: 1797-1810 [PMID: 28634041 DOI: 10.1016/j.ultrasmedbio.2017.05.002]

125   **Chen Y**, Luo Y, Huang W, Hu D, Zheng RQ, Cong SZ, Meng FK, Yang H, Lin HJ, Sun Y, Wang XY, Wu T, Ren J, Pei SF, Zheng Y, He Y, Hu Y, Yang N, Yan H. 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 [PMID: 28779596 DOI: 10.1016/j.compbiomed.2017.07.012]

126   **Matake K**, Yoshimitsu K, Kumazawa S, Higashida Y, Irie H, Asayama Y, Nakayama T, Kakihara D, Katsuragawa S, Doi K, Honda H. Usefulness of artificial neural network for differential diagnosis of hepatic masses on CT images. *Acad Radiol* 2006; **13**: 951-962 [PMID: 16843847 DOI: 10.1016/j.acra.2006.04.009]

127   **Oestmann PM**, Wang CJ, Savic LJ, Hamm CA, Stark S, Schobert I, Gebauer B, Schlachter T, Lin M, Weinreb JC, Batra R, Mulligan D, Zhang X, Duncan JS, Chapiro J. Deep learning-assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (HCC) *vs* non-HCC on contrast-enhanced MRI of the liver. *Eur Radiol* 2021; **31**: 4981-4990 [PMID: 33409782 DOI: 10.1007/s00330-020-07559-1]

128   **Kim J**, Min JH, Kim SK, Shin SY, Lee MW. Detection of Hepatocellular Carcinoma in Contrast-Enhanced Magnetic Resonance Imaging Using Deep Learning Classifier: A Multi-Center Retrospective Study. *Sci Rep* 2020; **10**: 9458 [PMID: 32527998 DOI: 10.1038/s41598-020-65875-4]

129   **Cucchetti A**, Piscaglia F, Grigioni AD, Ravaioli M, Cescon M, Zanello M, Grazi GL, Golfieri R, Grigioni WF, Pinna AD. Preoperative prediction of hepatocellular carcinoma tumour grade and micro-vascular invasion by means of artificial neural network: a pilot study. *J Hepatol* 2010; **52**: 880-888 [PMID: 20409605 DOI: 10.1016/j.jhep.2009.12.037]

130   **Urman JM**, Herranz JM, Uriarte I, Rullán M, Oyón D, González B, Fernandez-Urién I, Carrascosa

J, Bolado F, Zabalza L, Arechederra M, Alvarez-Sola G, Colyn L, Latasa MU, Puchades-Carrasco L, Pineda-Lucena A, Iraburu MJ, Iruarrizaga-Lejarreta M, Alonso C, Sangro B, Purroy A, Gil I, Carmona L, Cubero FJ, Martínez-Chantar ML, Banales JM, Romero MR, Macias RIR, Monte MJ, Marín JJG, Vila JJ, Corrales FJ, Berasain C, Fernández-Barrena MG, Avila MA. Pilot Multi-Omic Analysis of Human Bile from Benign and Malignant Biliary Strictures: A Machine-Learning Approach. *Cancers (Basel)* 2020; **12**: 1644 [PMID: 32575903 DOI: 10.3390/cancers12061644]

131   **Negrini D**, Zecchin P, Ruzzenente A, Bagante F, De Nitto S, Gelati M, Salvagno GL, Danese E, Lippi G. Machine Learning Model Comparison in the Screening of Cholangiocarcinoma Using Plasma Bile Acids Profiles. *Diagnostics (Basel)* 2020; **10**: 551 [PMID: 32748848 DOI: 10.3390/diagnostics10080551]

132   **Logeswaran R**. Cholangiocarcinoma--an automated preliminary detection system using MLP. *J Med Syst* 2009; **33**: 413-421 [PMID: 20052894 DOI: 10.1007/s10916-008-9203-3]

133   **Wübbolding M**, Lopez Alfonso JC, Lin CY, Binder S, Falk C, Debarry J, Gineste P, Kraft ARM, Chien RN, Maasoumy B, Wedemeyer H, Jeng WJ, Meyer Hermann M, Cornberg M, Höner Zu Siederdissen C. Pilot Study Using Machine Learning to Identify Immune Profiles for the Prediction of Early Virological Relapse After Stopping Nucleos(t)ide Analogues in HBeAg-Negative CHB. *Hepatol Commun* 2021; **5**: 97-111 [PMID: 33437904 DOI: 10.1002/hep4.1626]

134   **Haga H**, Sato H, Koseki A, Saito T, Okumoto K, Hoshikawa K, Katsumi T, Mizuno K, Nishina T, Ueno Y. A machine learning-based treatment prediction model using whole genome variants of hepatitis C virus. *PLoS One* 2020; **15**: e0242028 [PMID: 33152046 DOI: 10.1371/journal.pone.0242028]

135   **Bedon L**, Dal Bo M, Mossenta M, Busato D, Toffoli G, Polano M. A Novel Epigenetic Machine Learning Model to Define Risk of Progression for Hepatocellular Carcinoma Patients. *Int J Mol Sci* 2021; **22**: 1075 [PMID: 33499054 DOI: 10.3390/ijms22031075]

136   **European Association for the Study of the Liver**. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2018; **69**: 182-236 [PMID: 29628281 DOI: 10.1016/j.jhep.2018.03.019]

137   **Tsilimigras DI**, Mehta R, Moris D, Sahara K, Bagante F, Paredes AZ, Farooq A, Ratti F, Marques HP, Silva S, Soubrane O, Lam V, Poultsides GA, Popescu I, Grigorie R, Alexandrescu S, Martel G, Workneh A, Guglielmi A, Hugh T, Aldrighetti L, Endo I, Pawlik TM. Utilizing Machine Learning for Pre- and Postoperative Assessment of Patients Undergoing Resection for BCLC-0, A and B Hepatocellular Carcinoma: Implications for Resection Beyond the BCLC Guidelines. *Ann Surg Oncol* 2020; **27**: 866-874 [PMID: 31696396 DOI: 10.1245/s10434-019-08025-z]

138   **Christou CD**, Tooulias A, Tsolakidis A, Papayiannis V, Pianetcki-Tsiantzi B, Tsoulfas G, Papadopoulos VN. Management of Hepatocellular Carcinoma in the Era of Individualized Therapy: The Experience of a Greek Tertiary Center. *Ochsner J* 2020; **20**: 272-278 [PMID: 33071659 DOI: 10.31486/toj.19.0092]

139   **Tsilimigras DI**, Mehta R, Moris D, Sahara K, Bagante F, Paredes AZ, Moro A, Guglielmi A, Aldrighetti L, Weiss M, Bauer TW, Alexandrescu S, Poultsides GA, Maithel SK, Marques HP, Martel G, Pulitano C, Shen F, Soubrane O, Koerkamp BG, Endo I, Pawlik TM. A Machine-Based Approach to Preoperatively Identify Patients with the Most and Least Benefit Associated with Resection for Intrahepatic Cholangiocarcinoma: An International Multi-institutional Analysis of 1146 Patients. *Ann Surg Oncol* 2020; **27**: 1110-1119 [PMID: 31728792 DOI: 10.1245/s10434-019-08067-3]

140   **Jeong S**, Ge Y, Chen J, Gao Q, Luo G, Zheng B, Sha M, Shen F, Cheng Q, Sui C, Liu J, Wang H, Xia Q, Chen L. Latent Risk Intrahepatic Cholangiocarcinoma Susceptible to Adjuvant Treatment After Resection: A Clinical Deep Learning Approach. *Front Oncol* 2020; **10**: 143 [PMID: 32140448 DOI: 10.3389/fonc.2020.00143]

141   **Shao F**, Huang Q, Wang C, Qiu L, Hu YG, Zha SY. Artificial Neural Networking Model for the Prediction of Early Occlusion of Bilateral Plastic Stent Placement for Inoperable Hilar Cholangiocarcinoma. *Surg Laparosc Endosc Percutan Tech* 2018; **28**: e54-e58 [PMID: 29252936 DOI: 10.1097/SLE.0000000000000502]

142   **Hong WD**, Ji YF, Wang D, Chen TZ, Zhu QH. Use of artificial neural network to predict esophageal varices in patients with HBV related cirrhosis. *Hepat Mon* 2011; **11**: 544-547 [PMID: 22087192]

143   **Dong TS**, Kalani A, Aby ES, Le L, Luu K, Hauer M, Kamath R, Lindor KD, Tabibian JH. Machine Learning-based Development and Validation of a Scoring System for Screening High-Risk Esophageal Varices. *Clin Gastroenterol Hepatol* 2019; **17**: 1894-1901.e1 [PMID: 30708109 DOI: 10.1016/j.cgh.2019.01.025]

144   **Ho WH**, Lee KT, Chen HY, Ho TW, Chiu HC. Disease-free survival after hepatic resection in hepatocellular carcinoma patients: a prediction approach using artificial neural network. *PLoS One* 2012; **7**: e29179 [PMID: 22235270 DOI: 10.1371/journal.pone.0029179]

145   **Shi HY**, Lee KT, Wang JJ, Sun DP, Lee HH, Chiu CC. Artificial neural network model for predicting 5-year mortality after surgery for hepatocellular carcinoma: a nationwide study. *J Gastrointest Surg* 2012; **16**: 2126-2131 [PMID: 22878787 DOI: 10.1007/s11605-012-1986-3]

146   **Shi HY**, Lee KT, Lee HH, Ho WH, Sun DP, Wang JJ, Chiu CC. Comparison of artificial neural network and logistic regression models for predicting in-hospital mortality after primary liver cancer surgery. *PLoS One* 2012; **7**: e35781 [PMID: 22563399 DOI: 10.1371/journal.pone.0035781]

147   **Chiu HC**, Ho TW, Lee KT, Chen HY, Ho WH. Mortality predicted accuracy for hepatocellular

carcinoma patients with hepatic resection using artificial neural network. *Scientific World Journal* 2013; **2013**: 201976 [PMID: 23737707 DOI: 10.1155/2013/201976]

148   **Qiao G**, Li J, Huang A, Yan Z, Lau WY, Shen F. Artificial neural networking model for the prediction of post-hepatectomy survival of patients with early hepatocellular carcinoma. *J Gastroenterol Hepatol* 2014; **29**: 2014-2020 [PMID: 24989634 DOI: 10.1111/jgh.12672]

149   **Liu X**, Lu J, Zhang G, Han J, Zhou W, Chen H, Zhang H, Yang Z. A Machine Learning Approach Yields a Multiparameter Prognostic Marker in Liver Cancer. *Cancer Immunol Res* 2021; **9**: 337-347 [PMID: 33431375 DOI: 10.1158/2326-6066.CIR-20-0616]

150   **Zhong BY**, Yan ZP, Sun JH, Zhang L, Hou ZH, Yang MJ, Zhou GH, Wang WS, Li Z, Huang P, Zhang S, Zhu XL, Ni CF. Prognostic Performance of Albumin-Bilirubin Grade With Artificial Intelligence for Hepatocellular Carcinoma Treated With Transarterial Chemoembolization Combined With Sorafenib. *Front Oncol* 2020; **10**: 525461 [PMID: 33392064 DOI: 10.3389/fonc.2020.525461]

151   **Lai Q**, Spoletini G, Mennini G, Laureiro ZL, Tsilimigras DI, Pawlik TM, Rossi M. Prognostic role of artificial intelligence among patients with hepatocellular cancer: A systematic review. *World J Gastroenterol* 2020; **26**: 6679-6688 [PMID: 33268955 DOI: 10.3748/wjg.v26.i42.6679]

152   **R D**, P R. An Optimized HCC Recurrence Prediction Using APO Algorithm Multiple Time Series Clinical Liver Cancer Dataset. *J Med Syst* 2019; **43**: 193 [PMID: 31115780 DOI: 10.1007/s10916-019-1265-x]

153   **Yamashita R**, Long J, Saleem A, Rubin DL, Shen J. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. *Sci Rep* 2021; **11**: 2047 [PMID: 33479370 DOI: 10.1038/s41598-021-81506-y]

154   **Liang JD**, Ping XO, Tseng YJ, Huang GT, Lai F, Yang PM. Recurrence predictive models for patients with hepatocellular carcinoma after radiofrequency ablation using support vector machines with feature selection methods. *Comput Methods Programs Biomed* 2014; **117**: 425-434 [PMID: 25278224 DOI: 10.1016/j.cmpb.2014.09.001]

155   **Eaton JE**, Vesterhus M, McCauley BM, Atkinson EJ, Schlicht EM, Juran BD, Gossard AA, LaRusso NF, Gores GJ, Karlsen TH, Lazaridis KN. Primary Sclerosing Cholangitis Risk Estimate Tool (PREsTo) Predicts Outcomes of the Disease: A Derivation and Validation Study Using Machine Learning. *Hepatology* 2020; **71**: 214-224 [PMID: 29742811 DOI: 10.1002/hep.30085]

156   **Andres A**, Montano-Loza A, Greiner R, Uhlich M, Jin P, Hoehn B, Bigam D, Shapiro JAM, Kneteman NM. A novel learning algorithm to predict individual survival after liver transplantation for primary sclerosing cholangitis. *PLoS One* 2018; **13**: e0193523 [PMID: 29543895 DOI: 10.1371/journal.pone.0193523]

157   **Rodriguez-Luna H**, Vargas HE, Byrne T, Rakela J. Artificial neural network and tissue genotyping of hepatocellular carcinoma in liver-transplant recipients: prediction of recurrence. *Transplantation* 2005; **79**: 1737-1740 [PMID: 15973178 DOI: 10.1097/01.tp.0000161794.32007.d1]

158   **Lau L**, Kankanige Y, Rubinstein B, Jones R, Christophi C, Muralidharan V, Bailey J. Machine-Learning Algorithms Predict Graft Failure After Liver Transplantation. *Transplantation* 2017; **101**: e125-e132 [PMID: 27941428 DOI: 10.1097/TP.0000000000001600]

159   **Cruz-Ramírez M**, Hervás-Martínez C, Fernández JC, Briceño J, de la Mata M. Multi-objective evolutionary algorithm for donor–recipient decision system in liver transplants. *Eur J Oper Res* 2012; 317-327 [DOI: 10.1016/j.ejor.2012.05.013]

160   **Briceño J**, Cruz-Ramírez M, Prieto M, Navasa M, Ortiz de Urbina J, Orti R, Gómez-Bravo MÁ, Otero A, Varo E, Tomé S, Clemente G, Bañares R, Bárcena R, Cuervas-Mons V, Solórzano G, Vinaixa C, Rubín A, Colmenero J, Valdivieso A, Ciria R, Hervás-Martínez C, de la Mata M. 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 [PMID: 24905493 DOI: 10.1016/j.jhep.2014.05.039]

161   **Crawford K**, Calo R. There is a blind spot in AI research. *Nature* 2016; **538**: 311-313 [PMID: 27762391 DOI: 10.1038/538311a]

162   **Park SH**, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology* 2018; **286**: 800-809 [PMID: 29309734 DOI: 10.1148/radiol.2017171920]

163   **Patel VL**, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, Abu-Hanna A. The coming of age of artificial intelligence in medicine. *Artif Intell Med* 2009; **46**: 5-17 [PMID: 18790621 DOI: 10.1016/j.artmed.2008.07.017]

164   **Parikh RB**, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. *JAMA* 2019; **322**: 2377-2378 [PMID: 31755905 DOI: 10.1001/jama.2019.18058]

165   **Hwang TJ**, Kesselheim AS, Vokinger KN. Lifecycle Regulation of Artificial Intelligence- and Machine Learning-Based Software Devices in Medicine. *JAMA* 2019; **322**: 2285-2286 [PMID: 31755907 DOI: 10.1001/jama.2019.16842]

166   **Shortliffe EH**, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. *JAMA* 2018; **320**: 2199-2200 [PMID: 30398550 DOI: 10.1001/jama.2018.17163]

167   **Wolff RF**, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S; PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019; **170**: 51-58 [PMID: 30596875 DOI: 10.7326/M18-1376]

168   **American Medical Association**.   AMA passes first policy recommendations on augmented

intelligence. 2018 Jun 14 [cited 20 December 2020]. Available from: https://www.ama-assn.org/press-center/press-releases/ama-passes-first-policy-recommendations-augmented-intelligence

169 **Chico V**. The impact of the General Data Protection Regulation on health research. *Br Med Bull* 2018; **128**: 109-118 [PMID: 30445448 DOI: 10.1093/bmb/ldy038]

170 **Council NR**. For the Record: Protecting Electronic Health Information. Washington, DC: The National Academies Press, 1997

171 **The Guardian**. Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind. [cited 20 December 2020]. Available from: https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act

172 **U.S. Department of Labor**. Health Insurance Portability and Accountability Act (HIPAA). 2004 [cited 20 December 2020]. Available from: https://www.dol.gov/agencies/ebsa/laws-and-regulations/laws/hipaa

173 **National Academies of Sciences, Engineering, and Medicine**. Artificial Intelligence and Machine Learning to Accelerate Translational Research: Proceedings of a Workshop-in Brief. Washington (DC): National Academies Press, 2018 [PMID: 30048078 DOI: 10.17226/25197]

174 **Mirsky Y**, Mahler T, Shelef I, Elovici Y. CT-GAN: Malicious tampering of 3D medical imagery using deep learning. Proceedings of the 28th USENIX Security Symposium (USENIX Security 2019). arXiv, 2019: 461–478

175 **U.S. Food and Drug**. Software as a Medical Device (SaMD) Action Plan. 2021 [cited 2021 January 23]. Available from: Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

176 **Topol EJ**. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; **25**: 44-56 [PMID: 30617339 DOI: 10.1038/s41591-018-0300-7]

177 **U.S. Food and Drug**. Proposed Regulatory Framework for Modifications to Artificial Intelligence / Machine Learning (AI / ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback. 2019 [cited 20 January 23]. Available from: Available from: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf

178 **Institute of Medicine**. Medical devices and the public's health: the FDA(k) 510 clearance process at 35 years. Washington: the National Academies Press, 2011

179 **Price WN 2nd**, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. *JAMA* 2019; **322**: 1765-1766 [PMID: 31584609 DOI: 10.1001/jama.2019.15064]

180 **Froomkin AM**, Kerr I, Pineau J. When ais outperform doctors: confronting the challenges of a tort-induced over-reliance on machine learning. *Ariz L Rev* 2018; **61**

181 **Cohen IG**, Lynch HF, Vayena E, Gasser U. Big data, health law, and bioethics. New York, NY: Cambridge University Press, 2018

182 **Nundy S**, Montgomery T, Wachter RM. Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence. *JAMA* 2019; **322**: 497-498 [PMID: 31305873 DOI: 10.1001/jama.2018.20563]

183 **Future Advocacy**. Ethical, Social, and political challenges of artificial intelligence in health. [cited 20 December 2020]. Available from: https://wellcome.org/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf

184 **Gerke S**, Yeung S, Cohen IG. Ethical and Legal Aspects of Ambient Intelligence in Hospitals. *JAMA* 2020; **323**: 601-602 [PMID: 31977033 DOI: 10.1001/jama.2019.21699]

185 **Yeung S**, Rinaldo F, Jopling J, Liu B, Mehra R, Downing NL, Guo M, Bianconi GM, Alahi A, Lee J, Campbell B, Deru K, Beninati W, Fei-Fei L, Milstein A. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. *NPJ Digit Med* 2019; **2**: 11 [PMID: 31304360 DOI: 10.1038/s41746-019-0087-z]

186 **Yeung S**, Downing NL, Fei-Fei L, Milstein A. Bedside Computer Vision - Moving Artificial Intelligence from Driver Assistance to Patient Safety. *N Engl J Med* 2018; **378**: 1271-1273 [PMID: 29617592 DOI: 10.1056/NEJMp1716891]

187 **ProPublica**. Machine Bias. [cited 2021 January 23]. Available from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

188 **Al'Aref SJ**, Singh G, Baskaran L, Metaxas D. Machine Learning in Cardiovascular Medicine. 2020 [cited 2021 January 23]. Available from: http://public.eblib.com/choice/PublicFullRecord.aspx?p=6404325

189 **Vayena E**, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. *PLoS Med* 2018; **15**: e1002689 [PMID: 30399149 DOI: 10.1371/journal.pmed.1002689]

190 **Park SH**. Artificial Intelligence in Medicine: Beginners Guide. *DOAJ* 2018; **78**: 301-308 [DOI: 10.3348/jksr.2018.78.5.301]

191 **Zhang L**, Wang H, Li Q, Zhao MH, Zhan QM. Big data and medical research in China. *BMJ* 2018; **360**: j5910 [PMID: 29437562 DOI: 10.1136/bmj.j5910]

192 **Cook TS**. Human *vs* machine in medicine: can scientific literature answer the question? *Lancet Digit Health* 2019; **1**: e246-e247 [PMID: 33323246 DOI: 10.1016/S2589-7500(19)30124-4]

193 **Liu X**, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK.

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019; **1**: e271-e297 [PMID: 33323251 DOI: 10.1016/S2589-7500(19)30123-2]

194   **Christou CD**, Tsoulfas G. The Role of microRNA in Hepatic Ischemia/Reperfusion Injury. *Microrna* 2020; **9**: 248-254 [PMID: 31995027 DOI: 10.2174/2211536609666200129162531]

195   **Panesar S**, Cagle Y, Chander D, Morey J, Fernandez-Miranda J, Kliot M. Artificial Intelligence and the Future of Surgical Robotics. *Ann Surg* 2019; **270**: 223-226 [PMID: 30907754 DOI: 10.1097/SLA.0000000000003262]

196   **Kassahun Y**, Yu B, Tibebu AT, Stoyanov D, Giannarou S, Metzen JH, Vander Poorten E. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *Int J Comput Assist Radiol Surg* 2016; **11**: 553-568 [PMID: 26450107 DOI: 10.1007/s11548-015-1305-z]

197   **Wang DD**, Qian Z, Vukicevic M, Engelhardt S, Kheradvar A, Zhang C, Little SH, Verjans J, Comaniciu D, O'Neill WW, Vannan MA. 3D Printing, Computational Modeling, and Artificial Intelligence for Structural Heart Disease. *JACC Cardiovasc Imaging* 2021; **14**: 41-60 [PMID: 32861647 DOI: 10.1016/j.jcmg.2019.12.022]

198   **Volonté F**, Pugin F, Bucher P, Sugimoto M, Ratib O, Morel P. Augmented reality and image overlay navigation with OsiriX in laparoscopic and robotic surgery: not only a matter of fashion. *J Hepatobiliary Pancreat Sci* 2011; **18**: 506-509 [PMID: 21487758 DOI: 10.1007/s00534-011-0385-6]

199   **U.S. Food and Drug**. Developing a Software Precertification Program: A Working Model. 2018 [cited 2021 January 23]. Available from: https://www.fda.gov/media/113802/download