REVIEW

DOI: http://dx.doi.org/10.15403/jgld-4180

A Primer into the Current State of Artificial Intelligence in Gastroenterology

Alexandru Constantin Moldoveanu[1,2], Carmen Fierbințeanu-Braticevici[1,2]

1) Department of Internal Medicine II and Gastroenterology, Carol Davila University of Medicine and Pharmacy Bucharest, Romania
2) Department of Gastroenterology, Emergency University Hospital Bucharest, Romania

Address for correspondence:
Carmen Fierbințeanu-Braticevici
Department of Internal Medicine II and Gastroenterology, Carol Davila University of Medicine and Pharmacy Bucharest, Romania
cfierbinteanu@yahoo.com

Received: 29.12.2021
Accepted: 22.04.2022

ABSTRACT

Artificial intelligence (AI) is not a new idea or field of research. However, recent advancements in computing technology as well as increasing worldwide experience in applying AI to various fields have enabled us to hope that applying it to the medical practice could improve outcomes. The aim of the article was to evaluate potential applications of AI in the clinical practice of gastroenterology. Artificial intelligence algorithms can offer many benefits over the traditional diagnostic, prognostic, therapeutic scores, or non-AI enhanced procedures that we perform today. It can very easily process complex mathematical data, allow using many parameters and highly sophisticated formulas to determine a conclusion, to a degree that would be impractical or impossible for an unaided human to do alone in the clinical practice. This could permit creating significantly more individualized treatments for each patient. As more data becomes available, it can be fed into the algorithm, changing it based on the new data. However, we must be aware that most studies performed today are still early studies, and that their utility still needs to be evaluated in future large-scale studies. Apart from the medical challenges, it must be noted that the large-scale implementation of AI will have challenges related to ethical and financial aspects as well.

Key words: artificial intelligence – machine learning – deep learning – gastroenterology – endoscopy.

Abbreviations: AI: artificial intelligence; ANN: artificial neural network; CNN: convolutional neural network; IBD: inflammatory bowel disease; IBS: irritabile bowel syndrome; NAFLD: nonalcoholic liver disease; NN: neural network; NASH: nonalcoholic steatohepatitis; PNN: probabilistic neural network; RNN: recurrent neural network; SVM: support vector machine.

INTRODUCTION

Artificial intelligence (AI) is not a new idea or field of research. However, recent advancements in computing technology as well as increasing worldwide experience in applying AI to various fields have enabled us to hope that applying it to the medical practice in the field of gastroenterology could improve outcomes.

There is no single definition of an AI (and it is very unlikely that there will be an universally accepted one as it would require defining what constitutes intelligence), but an AI can be mostly described as a machine that imitates complex human thinking [1, 2]. Some of the original concepts of AI were established in the 1950s by Alan Turing, and recent technological advancements have allowed for more complex types of AI [3, 4].

Machine learning uses statistics to establish relationships between the various parameters. It should be noted that, while there are clinical scores in current use that were first established using regression algorithms (most notably the Model of End Stage Liver Disease and other scores based on it [5]), a machine learning algorithm can also be designed as an ever-improving algorithm. Most common types of simple machine learning algorithms include regression algorithms and classification algorithms.

Regression algorithms attempt to predict a mathematical relationship between two or more parameters (e.g., age and incidence of a specific disease), with minimal error. The most commonly used algorithm in this category is linear regression [6]. Classification algorithms attempt to determine thresholds for various categories (e.g., severity category based on blood test results). Algorithms commonly used to classify include random forests [7] and support vector machine (SVM) [8-12], although both can be used for regression analysis as well.

Support vector machine is a simpler type of algorithm, where each observation (e.g. patient) is drawn on a chart, using it's features as distance from the axis (e.g., patients with alcoholic hepatitis might have prothrombin time and bilirubin as features), and then a hyperplane (which is essentially a border between the two categories – e.g., severe and non-severe hepatitis) is drawn in order to separate the two categories of patients. Support vector machine can sometimes use custom features in order to obtain a clear separation between the categories. These custom features are computed from the input ones (e.g., the algorithm might compute the Maddrey score and use it as a feature).

Decision trees are a type of algorithm that attempts to classify a result based on its characteristics by creating a step-by-step process to perform decisions (e.g., if bilirubin > 10 then if also INR > 2: prognosis is poor, otherwise prognosis is good). While this type of algorithm is simple and easy to follow logically, it is very easy for the AI to tailor the step-by-step too closely to the dataset it learns from and be unable to make accurate predictions when presented with new data. The more commonly used version of this algorithm is the random forests algorithm which attempts to build multiple such decision trees, from randomly assigned variables (e.g., if the input variables include coagulation time, bilirubin, and creatinine, the algorithm might build a tree with just coagulation and creatinine, and another tree with bilirubin and creatinine) and randomly assigned subsamples of data (e.g., one tree might be built on the first 40% of the data, while another might be built on the middle 30%). The final output will be determined as follows: each tree "votes" on they think the output should be, and the majority vote will be considered the most likely prediction. While there will be errors (e.g., trees being assigned useless parameters), these errors will be random, and they will most likely cancel each other out.

Neural networks are a more complex type of algorithm, that uses neural nodes. Each node receives input for various parameters (like the dendrites in a natural neural network), processes the data in the neural node (like the core of the neural cell), and outputs a calculated parameter (like the axon in a neural cell). A neural network contains multiple such nodes (hence network), and the algorithm assigns different weights to each node, depending on overall accuracy and importance of each node. This type of network is referred to in articles as either a neural network (NN) or artificial neural network (ANN). These algorithms allow for very complex formulas to be used routinely for establishing diagnosis and predicting treatment outcomes and survival [13, 14].

A deep learning network is essentially a multi-layer neural network, which allows for intermediate conclusions (the conclusions of one layer of the network will be used as input parameters for the next layer). This allows for significantly more nuanced and complex learning algorithms and are generally used in complex applications, such as image and video recognition [15, 16]. The most widely used algorithm type in medical studies is the convolutional neural network (CNN), which is highly regarded for its ability to interpret picture data and recurrent neural network (RNN) which is highly regarded for its ability to interpret video and audio data [17, 18]. Practical applications in gastroenterology mostly relate to AI enhanced endoscopic procedures [19, 20], AI enhanced imaging [21] but also as a potential clinical aid [22].

The actual implementation of AI can differ slightly for each type of algorithm, but most studies have a few things in common.

The data is usually divided into two datasets. The first set (training set) is used by the AI to learn and form the mathematical relationships, while the second set (testing set) is used to test the accuracy of the newly formed mathematical relationships [23].

Another key characteristic of the AI algorithms is the type of learning performed – supervised learning or unsupervised learning. Supervised learning is done by providing the algorithm in the training phase with both input data (e.g., blood tests results) and output data (e.g., mortality), allowing the algorithm to develop its own relationships between the data. The algorithm is later given in the testing phase just input data, and its ability to correctly predict output data is measured [24]. The datasets for the two phases are different (sometimes allocated at random from a larger dataset). Unsupervised learning is done by providing the algorithm with input data only (in both datasets), to identify new patterns. Previously unknown classification parameters are used to identify clusters of patients with a consistent pattern within the cluster but different patterns between the clusters. Once the algorithm has grouped patients in different clusters, these clusters are analyzed, and their outcomes are determined [25].

In more complex algorithms, a combination of both supervised and unsupervised learning can also be used. One such method implies using a deep learning (multi-layer) network, where the output data of one layer serves as input data for a different layer that is of the opposite learning method (the results of an unsupervised learning algorithm could be used as input for a supervised learning algorithm, or vice versa). For example, an unsupervised clustering algorithm can be run on imaging results, and a classifier is then used to predict which clusters resulting from the first layer correlate to which outcome [26]. Another method involves using the supervised data to help increase the accuracy of the unsupervised learning (essentially the algorithm is designed as an unsupervised learning algorithm but receives some help initially from supervised learning) [27-29].

Artificial intelligence algorithms can offer many benefits over the traditional diagnostic, prognostic, therapeutic scores, or non-AI enhanced procedures that we perform today. Artificial intelligence can very easily process complex mathematical data, accounting for many parameters and using highly sophisticated formulas to determine a conclusion, to a degree that would be impractical or impossible for a human to do alone in the clinical practice. Some neural networks that are being studied today use hundreds of parameters to determine a potential outcome, suggesting the possibility of a significantly more individualized treatment for each patient [30]. Another advantage constitutes the ability of the algorithm to "learn". As more data becomes available, it can be fed into the algorithm, which allows it to change based on the new data and increase its accuracy.

Artificial intelligence algorithms also have many disadvantages. Many of these will probably be overcome as our understanding and application of AI will improve, but at this early stage, it is important to understand the inherent weaknesses of these tools, as they present new challenges to the clinician.

Artificial intelligence algorithms are very dependent on high quality data, much more so than the current practice. For an algorithm to correctly train itself and become proficient, it requires large amounts of data, and the data needs to be accurate and consistent [31]. This presents unique challenges, as the requirement for large amounts of data would suggest multi-center AI as a solution, but issues such as differing testing methodologies or test types between the various centers might impede the algorithms' ability to learn [32].

It should be also noted that these algorithms will require intense human supervision at their early stage, as AI is very susceptible to biases and overfitting. The algorithm might learn very well on the initial training set, but in trying to predict the outcomes extremely accurately in the training set, it fails to correctly predict the actual target population. This can happen due to several reasons, including small number of observations in the training set, inherent bias when selecting the training set (patients of the same center, specifically selected pictures for image training, patient demographics etc.), or algorithm settings. The result is the algorithm either interprets data that is not generalizable (biased data) or creates an overly complex model that is not generalizable (overfitting the prediction model) [33]. Human oversight of this algorithms is especially difficult, as the more complex an algorithm is, the less information the human overseer has regarding the actual decision process of the algorithm [34].

Another significant question that will require a complex answer as AI systems are implemented into the medical practice, is how medical liability would be assigned. A recent article proposes a level system to assess the autonomy of AI systems in performing medical decisions [35], ranging from level 0 (no AI used) to level 5 (AI is fully autonomous in making medical decisions). While current research proposes AIs as a useful tool, it is very likely that some procedures could become partially or completely automated in the future. The question remains, as we progress towards the higher levels of AI autonomy, who is liable when the AI fails [36].

Finally, AI is a processing intensive method, and this will add to the operating costs of every institution that employs them. Most likely, each institution will have to choose between paying for computer hardware and employee (AI programmer/supervisor) salary cost or paying a third party for a software as service package. Some of these costs may be offset by better organizational efficiency [37]. At this early stage of development however, we do not yet know how the implementing AI will change the economics of the medical practice [38]. There are further implications regarding the need for an AI to have large amounts of data. Storage of this data in large centers could have a privacy concern as well as data security concerns as cyberattacks and data leaks have become a common occurrence in other fields with large amounts of data storage and AI implementations [39,40].

The aim of this article was to evaluate potential applications of AI in the clinical practice of gastroenterology. The searching strategy and the Prisma flowcharts are listed in the Supplementary file.

ARTIFICIAL INTELLIGENCE IN GASTROENTEROLOGY

There have been attempts to apply AI systems to almost every problem in gastroenterology. Most studies, however, have very few patients, which would make them very likely to be subject to the above-mentioned disadvantages. We list here some of the more promising results, but it should be noted that most of these are retrospective studies and further research is probably needed in all fields.

Atrophic Gastritis

A study was performed on 350 patients with either atrophic body gastritis (study lot) or without atrophic body gastritis (control lot). Input data used was non-endoscopic – anamnestic, clinical, blood test data and the results were compared to the histopathological examination. Two algorithms were used: ANN and linear discriminant analysis (LDA), a less complex classifier type of algorithm. Overall accuracies were between around 94-98% for most configurations of input data [41].

Upper Gastrointestinal Bleeding

A study performed on 2,380 patients with upper gastrointestinal bleeding, where 30-day mortality was predicted by using an ANN and non-endoscopic input data (a total of 68 input variables). Patients were split randomly between the training and testing sets. The results were compared to a score currently used in clinical practice, the Rockall score. Using ANN yielded much better results, with an overall accuracy of 96.8%, compared to the 52.9% for the Rockall score [42]. Similar studies were performed by other researchers, comparing to the Rockall, the Glasgow-Blatchford and the AIMS65 scores, with superior results for the AI model (AUROC of 0.90 compared to 0.86 for Glasglow-Blatchford, 0.66 for Rockall score, and 0.64 for AIMS65) [43].

Patients at risk for recurrent upper gastrointestinal bleeding from ulcers were classified into high risk and low risk groups for rebleeding within a year by a model. The AUROC value was not impressive, but the negative predictive value was 91%, suggesting the test could be used to rule out some patients from more intensive follow-up [44].

Gastric Cancer

A meta-analysis of multiple studies that have tried to assess prognosis in gastric cancer, using endpoints such as survival time, recurrence and metastasis shows good accuracy (70-95%). Multiple parameters were used, including demographics, blood tests, imaging results, (computed tomography), tissue samples; most studies used a larger number of patients (hundreds) [45]. Of particular note are studies that tried to predict response to adjuvant chemotherapy [46] and survival [47], with superior results to the currently used tumor – nodules – metastases (TNM) system.

Nonalcoholic Fatty Liver Disease

Most studies involved in identifying nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH) used imaging data from ultrasound or transient elastography. A study used a probabilistic neural network (PNN) type AI (a classifier) to differentiate patients with normal liver vs. patients with NAFLD vs. patients with NAFLD and cirrhosis, based on ultrasound images. The overall accuracy was 97%, with an AUC of 0.98 [48]. Lee et al. [49] have tried to compare results of a CNN type network using ultrasound images from 3,446 patients as input. The test was compared with liver biopsy results or transient elastography results. AUC for determining cirrhosis was 0.857-0.901 [49]. Considering that an experienced operator could also reach similar conclusions, and that a skilled operator is still required to guide data acquisition, it is unclear whether these results, while showing high performance in diagnosis, will lead to a significant improvement in the clinical practice.

Artificial intelligence models were also applied to histopathological examinations, as a possible method of automated examination. These models could recognize the characteristics of NAFLD, with high accuracy (AUROC 95-99%) [50]. Similar accuracies were found in distinguishing between NASH and non-NASH NAFLD in another study [51].

Considering the prevalence of NAFLD, multiple algorithms were developed to try to identify patients at risk, using electronic health records. Many of these have shown significant improvement in detection rates, compared to the simple diagnosis and codifications systems in use today [52-55].

Various types of AI have been used to classify patients with NAFLD and already developed cirrhosis. An ANN using input data based on demographics, personal history, and blood tests managed to correctly classify patients with cirrhosis with an overall accuracy of 86% of cases in a study with nearly 400 patients [56].

Cirrhosis

Artificial intelligence models attempted to determine the risk of readmission and death in patients with cirrhosis. The AI was not able to accurately predict (best AUROC was around 0.78, most being around 0.60-0.65) these parameters and showed minimal improvement over using standard scores (MELD-Na – AUROC of around 0,67 for 1 year mortality) [57, 58]. Relatively few articles explore predictions in cirrhosis, perhaps due to relatively poor significant findings of the AI models attempted so far.

Predictive factors for progression from chronic viral hepatitis to advanced fibrosis and cirrhosis were studied using various types of AI. Different input data were used: clinical data [59], biochemical and imaging data [60], patient genome [61], and virus genome [62] and similar results were reported (AUROCs of around 0.80). Biochemical data was also used to predict progression of patients with primary sclerosing cholangitis, with good results (C-statistic of 0.90), which is significantly better than current scores (C-statistics of 0.72 for MELD, 0.85 for the Mayo primary sclerosing cholangitis risk score) [63].

Liver transplantation has been the „holy grail" of treatment for cirrhosis, but graft rejects and overall poor improvement in survivability has been a significant challenge. Given the scarcity of the donor pieces, a better prediction of which patients benefit most from transplantation would be useful. Lau et al. [64] has attempted to perform a small-scale study, using a random forests type classifier. The results are promising, as the AI had an AUROC of 0.818 in identifying patients that are going to experience graft failure [64].

Inflammatory Bowel Diseases

An automatic histopathological interpretation system was tested on a small number of patients with inflammatory bowel diseases (IBD), with promising results (98% accuracy) compared to expert human interpreters [65]. Histological characteristics were also used to distinguish between Crohn's disease and ulcerative colitis, using an SVM type algorithm, using tissue samples from affected areas of the colon. Input data was based on the proteomic pattern profile of each sample, and the results had a 76.9% accuracy in distinguishing between the two pathologies [66].

Another interesting study has tried to integrate genomic characteristics into a prediction model for Crohn's disease, with improvements from the traditional analysis methods (AUROC of 0.816 compared to 0.786) [67].

While the best results for diagnosis in IBD, even with AIs, still use histopathological, biochemical and imaging data (with accuracies in the range 82-93%, depending on study methodology [68]), it should be noted that the small studies listed above show that the increased computational capacity of an artificial intelligence can allow for novel and complex data to be included into the diagnosis models, which would have been either impractical or impossible in an non-AI clinical practice.

A simple random forest algorithm using biological and demographic data was used to predict Cron's disease response to ustekinumab, as defined by remission at 42 weeks. Artificial intelligence was unable to accurately predict the response before the treatment start (AUROC=0.59) but was reasonably accurate in predicting the outcome once the treatment has started (at 8 weeks, AUROC=0.78) [69].

Another study tried to predict the risk of IBD flares and group patients in "high risk" and "low risk" for flares. Input data included demographics, previous medications, number of previous hospitalizations and biochemical data. The AI type was random forest, with results having an AUROC of 0.85-0.87 [70].

Irritable Bowel Syndrome

The exact mechanisms and triggers or irritable bowel syndrome IBS have been the subject of much debate, with some patients having reported certain foods as triggers for the symptoms. A few small studies have tried to assess the relationships between food composition (by logging foods with smartphone apps) and attempting to correlate diet with symptoms [71, 72]. While these studies are very early-stage, performed with a low number of patients, and it is yet unclear if the AI improves the results, or if they are driven by patient behavior modification (more strict logging of foods), it seems reasonable that data obtained from larger scale studies might allow us to better understand the pathogenesis of the disease, as well as provide better individualized therapy. Analysis of the

gut microbiome through fecal samples and AI implementation may also yield useful information for these patients, but the relationship between the results and IBS is as of yet unclear [73].

Colorectal Cancer

Kudo et al. [74] have used an ANN type network with data from 3,134 patients with T1 colorectal cancer to predict lymph nodes metastases. Input data included tumor characteristics, results of histopathological examinations and demographic data. The ANN managed to identify patients with lymph node metastases, with an AUROC of 0.83, compared to the current guidelines which yielded an AUROC of 0.73. This model could have considerable implications in determining which patients require further surgery [74]. Another study has also tried to classify patient response to chemotherapy (5-fluorouracil), using biomarkers with significant differences in survival between the predicted responder and predicted non-responder groups [75].

Automated identification of histopathological tissue samples was attempted, using a CNN type network, with good results; the score proposed by the AI was a significant prognostic factor for patient survival [76]. Microsatellite instability is also easily identified by an AI algorithm with an AUROC of 0.865 vs. expert pathologists with an AUROC of 0.605 [77].

Data from a colorectal cancer registry (a total of 4,299 patients enrolled) was used to evaluate risk factors for colorectal cancer recurrence with promising results, an AUROC of 0.878 for some applications [78].

Computer Aided Endoscopy

A systematic analysis of 16 studies evaluated AI performance in detecting esophageal neoplasms using either a CNN type network (14 studies), or a SVM type algorithm. The results of the AI were very good, with an AUROC of 0.96-0.98. The studies were, however, mostly retrospective, and the analysis concluded that further prospective studies are required [79]. What is more interesting is that one of the studies from the meta-analysis identifies that the AI detection has significantly better accuracy when compared to novice endoscopists, but similar accuracy compared to senior endoscopists [80].

High rates of detection were also obtained for early gastric cancers in a meta-analysis of 17 studies. Most of these studies are retrospective and used a CNN-type network with a training set consisting of endoscopy images. The results had an average AUROC of 0.96 [45, 46, 81, 82].

Computer Aided Colonoscopy

Artificial intelligence-based real-time image detection and categorization permits a new type of aid in clinical practice. Potentially, an AI could be used to either aid the colonoscopy operator, or to perform automatic classification and identification of structures.

Multiple studies were performed in the attempt of improving polyp detection rates and, especially, adenoma detection. Detection rates were consistently improved in most studies, with odds ratio between 1.36–2.33 (for adenoma detection) and between 1.72–2.02 (for polyp detection). More interestingly, most studies report an important improvement in detection rates of all polyps, but especially of small (< 5 mm) polyps, which are more likely to be missed by an un-assisted operator in conventional colonoscopy. In most studies, the colonoscopy videos were examined by a panel of expert reviewers, both with and without AI assistance, and the results were compared. Notably, the examination time was increased by an average of 1-2 minutes for the AI assisted procedures [83-85].

An automatic robotic steering system for endoscopy was compared to conventional steering in both novice and expert human operators. The results showed significant improvement in polyp detection rate for novice operators, but not for expert operators. The cecal intubation time was also significantly better for expert operators with conventional steering, suggesting that, while technically feasible, the model needs significant improvement [86].

Stidham et al. [87] assessed deep learning algorithms (more specifically a CNN type algorithm) for their ability to grade remission in ulcerative colitis. The training set consisted of colonoscopy images and the testing set included colonoscopy images as well as full length colonoscopy video. The study found excellent accuracy and interobserver agreement with human examinators [87]. A similar study was performed by Gottlieb et al. [88] but using a RNN type network and using only full-length colonoscopy video. The algorithm had similar results, with a high degree of accuracy [88].

Video Capsule Endoscopy

Video capsule endoscopy is a time consumptive examination for the operator, as it requires that a human operator examines hours of footage recorded from the gastrointestinal tract. A meta-analysis of 9 studies tried to assess the accuracy of AI (using CNN type networks) as a potential replacement for the human operator. Overall accuracy was around 95% for the detection of ulcers and bleeding [89]. Performance was also found to be good (90.8%) for detecting more superficial lesions, including erosions and ulcerations [90].

Computer Aided Endoscopic Ultrasound

Computer aided endoscopic ultrasonography studies were also performed for patients with gastrointestinal stromal tumors. Results were promising, with an AUROC of 0.965 for lesions larger than 20 mm, which was significantly higher than an AUROC of 0.684 that was obtained by expert operators [91].

Diagnosis of autoimmune pancreatitis and differential diagnosis with pancreatic ductal adenocarcinoma is a difficult proposition. A CNN-type AI was implemented to distinguish between the two diseases as well as chronic pancreatitis and a normal pancreas, with good results (sensitivities of 90-99% and specificities of 71-98%, depending on the exact comparison) in distinguishing between the four entities [92, 93].

Screening Programs

Electronic health records have shown some promise in screening for pancreatic cancer, with current results having an AUROC of 0.68-0.72 [94, 95]. Some studies suggested better performance when applied to high-risk populations [96]. Other suggested parameters to be analyzed by AIs included imaging results and even social media data [97], but further research is needed.

An interesting study was performed by Nakhleh et al. [98], where the researchers studied a potentially novel application of AI. They used an AI pattern analysis algorithm on a specific nano-array that assessed exhaled volatile organic compounds from 1,404 subjects with various diseases, including some gastroenterological diseases like colon cancer, IBS, IBD, but also for non-gastroenterological diseases. The overall accuracy was 86%, and the results suggested that the test evaluated the pathogenic mechanisms behind each disease, as similar diseases seem to have similar volatile organic compounds profiles [98]. Nevertheless, this may have great practical application as a possible screening tool.

LIMITS OF CURRENT STUDIES

Many of the evaluated studies have significant methodology difficulties. Most of them are performed using already existing data (retrospective studies), providing no knowledge of how an AI might be implemented in practice. Most studies also contain low number of patients, which, as already discussed, runs a significant risk of an AI overfitting or otherwise creating a non-representative model for the task it tackles.

Some of the studies with large numbers of input data have issues with data sampling bias and methodology. The most relevant example is training an AI for endoscopy. Some studies use live full-length videos, but most use images, and it is usually not clearly explained how the images are chosen: are they the actual frames of the video (hence representing the entire full-length video) or are they specifically chosen to be easier to interpret (by excluding blurred frames for example)?

Another issue of methodology is that not all studies completely explain how the AI was built and tested – size of training set, size of testing set, how the sets were chosen, or any other parameter related to the specific type of AI implemented (e.g., setting a maximum number of decision nodes in a decision tree). We should recognize that, while seeking to fine tune an AI to perform well, the human programing the AI is also unwillingly or unknowingly susceptible to bias. A quality hallmark of a good scientific study is the complete reproductivity of the study based on the methodology description, however, due to the significantly increased computational complexity, AI related research includes new obstacles to maintaining this important objective. One possible solution could be to post the implementation code in an online open repository, similar to how datasets are currently being shared for traditional research [99]. This will however introduce new problems: it will either require that most researchers be able to understand and evaluate the code (which will perhaps require that AI programming and implementation be taught in the University), or that frequent collaboration is performed with specialized engineers (current implementation, but to a much larger scale, especially if AIs go from being a largely academic and research tool to becoming a clinical practice tool).

In some fields (e.g., non-endoscopic evaluation of atrophic gastritis) there are very few studies performed so far, to the point that it is very difficult to form a clear picture whether the AI is useful or accurate. In other fields there are many studies, but the implementation is significantly different between the studies, which, while very useful for comparing different implementations of AI for a specific task, makes it more difficult to draw a conclusion. Some study results, while most likely being not random, might not perform well-enough to have a practical application or improvement in the clinical practice.

We suspect that the most difficult implementation of AI would probably be in rare diseases (due to the small sample sizes and making multi-center AI mandatory) and in invasive maneuvers (especially if automation is desired, due to the incredibly varied and large amounts of input data).

Finally, we should mention that, for quality assessment tools it is difficult to determine if they are correctly applied to AI studies (e.g., assessing bias or sample size of the study).

POTENTIAL NEW RESEARCH DIRECTIONS

Artificial intelligence models could potentially be used as training tools, for either students or doctors at the start of their carriers.

Several of the studies performed here showed that the AI could be as good as an experienced human investigator, but significantly better than a novice investigator. While the impact on the clinical practice might be low (depending on the overall level required to be classified as "experienced" for that task), the AI might be used to train the novice more quickly.

Artificial intelligence algorithms could also be used to train students. As some of the commercially available AI driven products can suggest (e.g., voice assistants), the AI can be tailored to the individual student and suggest potential learning improvements. An overall analysis on batches of students can also be performed, which may find flaws within a course or teaching tool (unclear explanation, insufficient context etc.). Finally, like a voice assistant, a potential "patient simulator" could be created, that will more easily allow students to practice and train on specific skills (e.g., anamnesis), in a more realistic environment.

CONCLUSIONS

Artificial intelligence has a great potential to improve healthcare, and its implementation can yield improvements in multiple areas related to gastroenterology. However, we must be aware that most studies included in this paper have significant methodological issues, either by being retrospective, containing a low number of patients, or potentially having sampling biases make drawing conclusions difficult. The increased computational complexity also makes these studies very difficult to reproduce for validation. We fully expect that, with more research and better methodology, the results and conclusion of this paper may significantly change. Apart from the medical challenges, it must be noted that the impact of large-scale implementation of artificial intelligence will have challenges related to ethical and financial aspects as well.

Conflicts of interest: None to declare.

Authors' contribution: A.C.M. and C.F.B conceived the study. A.C.M. designed the study, collected data, interpreted the results and drafted

the manuscript. C.F.B.: supervised data collection and analysis and revised the paper. All the authors approved the final version of the manuscript.

Supplementary material: To access the supplementary material visit the online version of the *J Gastrointestin Liver Dis* at http://dx.doi.org/10.15403/jgld-4180

REFERENCES

1. Wang P. On Defining Artificial Intelligence. J Artif Gen Intell 2019;10:1–37. doi:10.2478/jagi-2019-0002
2. Legg S, Hutter M. Universal Intelligence: A Definition of Machine Intelligence. Minds Mach 2007;17:391–444. doi:10.1007/s11023-007-9079-x
3. Ramesh AN, Kambhampati C, Monson JR, Drew P. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004;86:334–338. doi:10.1308/147870804290
4. Oka A, Ishimura N, Ishihara S. A New Dawn for the Use of Artificial Intelligence in Gastroenterology, Hepatology and Pancreatology. Diagnostics (Basel) 2021;11:1719. doi:10.3390/diagnostics11091719
5. Kartoun U, Corey KE, Simon TG, et al. The MELD-Plus: A generalizable prediction risk score in cirrhosis. PLoS One 2017;12:e0186301. doi:10.1371/journal.pone.0186301
6. Emmert-Streib F, Dehmer M. High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection. Mach Learn Knowl Extr 2019;1:359–383. doi:10.3390/make1010021
7. Breiman L. Random Forests. Mach Learn 2001;45:5–32. doi:10.1023/A:1010933404324
8. Deo RC. Machine Learning in Medicine. Circulation 2015;132:1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593
9. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 2019;1:281. doi:10.1186/s12911-019-1004-8
10. Vanderpool CPB. Artificial Intelligence and Gastroenterology: Expanding Technology. J Pediatr Gastroenterol Nutr 2021;72:183. doi:10.1097/MPG.0000000000002986
11. Bishop CM. Pattern Recognition and Machine Learning. Springer New York, NY: 2016.
12. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, NY: Springer; 2009.
13. Drew PJ, Monson JR. Artificial neural networks. Surgery 2000;127:3–11. doi:10.1067/msy.2000.102173
14. Wesolowski M, Suchacz B. Artificial Neural Networks: Theoretical Background and Pharmaceutical Applications: A Review. J AOAC Int 2012;95:652–668. doi:10.5740/jaoacint.sge_wesolowski_ann
15. Lee JG, Jun S, Cho YW, et al. Deep Learning in Medical Imaging: General Overview. Korean J Radiol 2017;18:570-584. doi:10.3348/kjr.2017.18.4.570
16. Kather JN, Krause J, Luedde T. Künstliche Intelligenz in der Gastroenterologie. Dtsch Med Wochenschr 2020;145:1450–1454. doi:10.1055/a-1013-6593
17. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. J Med Imaging Radiat Sci 2019;50:477–487. doi:10.1016/j.jmir.2019.09.005
18. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60:84–90. doi:10.1145/3065386
19. Min JK, Kwak MS, Cha JM. Overview of Deep Learning in Gastrointestinal Endoscopy. Gut Liver 2019;13:388–393. doi:10.5009/gnl18384
20. Jia X, Meng MQH. A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Orlando, FL, USA. 2016:639–642. doi:10.1109/EMBC.2016.7590783
21. Hassan TM, Elmogy M, Sallam ES. Diagnosis of Focal Liver Diseases Based on Deep Learning Technique for Ultrasound Images. Arab J Sci Eng 2017;42:3127–3140. doi:10.1007/s13369-016-2387-9
22. Forte C, Voinea A, Chichirau M, et al. Deep Learning for Identification of Acute Illness and Facial Cues of Illness. Front Med 2021;8:661309. doi:10.3389/fmed.2021.661309
23. 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. doi:10.1148/radiol.2017171920
24. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med 2018;284:603–619. doi:10.1111/joim.12822
25. Vranas KC, Jopling JK, Sweeney TE, et al. Identifying Distinct Subgroups of ICU Patients: A Machine Learning Approach. Crit Care Med 2017;45:1607–1615. doi:10.1097/CCM.0000000000002548
26. Batmanghelich KN, Ye DH, Pohl KM, Taskar B, Davatzikos C, Adni. Disease classification and prediction via semi-supervised dimensionality reduction. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Chicago, IL, USA. 2011:1086–1090. doi:10.1109/ISBI.2011.5872590
27. Chapelle O, Scholkopf B, Zien A. Eds. Semi-Supervised Learning (Chapelle O, et al. Eds.; 2006) [Book reviews]. In: IEEE Trans Neural Netw 2009;20:542–542. doi:10.1109/TNN.2009.2015974
28. Timsina P, Liu J, El-Gayar O, Shang Y. Using Semi-Supervised Learning for the Creation of Medical Systematic Review: An Exploratory Analysis. 2016 49th Hawaii International Conference on System Sciences (HICSS). Koloa, HI, USA. 2016:1195–1203. doi:10.1109/HICSS.2016.151
29. Azmi R, Norozi N, Anbiaee R, Salehi L, Amirzadi A. IMPST: A New Interactive Self-Training Approach to Segmentation Suspicious Lesions in Breast MRI. J Med Signals Sens 2011;1:138–148.
30. Ruffle JK, Farmer AD, Aziz Q. Artificial Intelligence-Assisted Gastroenterology-Promises and Pitfalls. Am J Gastroenterol 2019;114:422–428. doi:10.1038/s41395-018-0268-4
31. Chishti S, Jaggi KR, Saini A, Agarwal G, Ranjan A. Artificial Intelligence-Based Differential Diagnosis: Development and Validation of a Probabilistic Model to Address Lack of Large-Scale Clinical Datasets. J Med Internet Res 2020;22:e17550. doi:10.2196/17550
32. Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. JAMA 2019;322:1806-1816. doi:10.1001/jama.2019.16489
33. Handelman GS, Kok HK, Chandra RV, et al. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. AJR Am J Roentgenol 2019;212:38–43. doi:10.2214/AJR.18.20224
34. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They

Impact Health Care? J Arthroplasty 2018;33:2358–2361. doi:10.1016/j.arth.2018.02.067
35. Yang GZ, Cambias J, Cleary K, et al. Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy. Sci Robot 2017;2:eaam8638. doi:10.1126/scirobotics.aam8638
36. Poon NC, Sung JJ. Self-driving cars and AI-assisted endoscopy: Who should take the responsibility when things go wrong? J Gastroenterol Hepatol 2019;34:625–626. doi:10.1111/jgh.14641
37. Abonamah AA, Tariq MU, Shilbayeh S. On the Commoditization of Artificial Intelligence. Front Psychol 2021;12:696346. doi:10.3389/fpsyg.2021.696346
38. Wolff J, Pauling J, Keck A, Baumbach J. The Economic Impact of Artificial Intelligence in Health Care: Systematic Review. J Med Internet Res 2020;22:e16866. doi:10.2196/16866
39. Park CW, Seo SW, Kang N, et al. Artificial Intelligence in Health Care: Current Applications and Issues. J Korean Med Sci 2020;35:e379. doi:10.3346/jkms.2020.35.e379
40. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial Intelligence Transforms the Future of Health Care. Am J Med 2019;132:795–801. doi:10.1016/j.amjmed.2019.01.017
41. Lahner E, Grossi E, Intraligi M, et al. Possible contribution of advanced statistical methods (artificial neural networks and linear discriminant analysis) in recognition of patients with suspected atrophic body gastritis. World J Gastroenterol 2005;11:5867-5873. doi:10.3748/wjg.v11.i37.5867
42. Rotondano G, Cipolletta L, Grossi E, et al. Artificial neural networks accurately predict mortality in patients with nonvariceal upper GI bleeding. Gastrointest Endosc 2011;73:218-226.e2. doi:10.1016/j.gie.2010.10.006
43. Shung DL, Au B, Taylor RA, et al. Validation of a Machine Learning Model That Outperforms Clinical Risk Scoring Systems for Upper Gastrointestinal Bleeding. Gastroenterology 2020;158:160–167. doi:10.1053/j.gastro.2019.09.009
44. Wong GL, Ma AJ, Deng H, et al. Machine learning model to predict recurrent ulcer bleeding in patients with history of idiopathic gastroduodenal ulcer bleeding. Aliment Pharmacol Ther 2019;49:912–918. doi:10.1111/apt.15145
45. Niu PH, Zhao LL, Wu HL, Zhao DB, Chen YT. Artificial intelligence in gastric cancer: Application and future perspectives. World J Gastroenterol 2020;26:5408–5419. doi:10.3748/wjg.v26.i36.5408
46. Jiang Y, Xie J, Han Z, et al. Immunomarker Support Vector Machine Classifier for Prediction of Gastric Cancer Survival and Adjuvant Chemotherapeutic Benefit. Clin Cancer Res 2018;24:5574–5584. doi:10.1158/1078-0432.CCR-18-0848
47. Oh SE, Seo SW, Choi MG, Sohn TS, Bae JM, Kim S. Prediction of Overall Survival and Novel Classification of Patients with Gastric Cancer Using the Survival Recurrent Network. Ann Surg Oncol 2018;25:1153–1159. doi:10.1245/s10434-018-6343-7
48. Acharya UR, Raghavendra U, Fujita H, et al. Automated characterization of fatty liver disease and cirrhosis using curvelet transform and entropy features extracted from ultrasound images. Comput Biol Med 2016;79:250–258. doi:10.1016/j.compbiomed.2016.10.022
49. Lee JH, Joo I, Kang TW, et al. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. Eur Radiol 2020;30:1264–1273. doi:10.1007/s00330-019-06407-1
50. Forlano R, Mullish BH, Giannakeas N, et al. High-Throughput, Machine Learning–Based Quantification of Steatosis, Inflammation, Ballooning, and Fibrosis in Biopsies From Patients With Nonalcoholic Fatty Liver Disease. Clin Gastroenterol Hepatol 2020;18:2081-2090.e9. doi:10.1016/j.cgh.2019.12.025
51. Teramoto T, Shinohara T, Takiyama A. Computer-aided classification of hepatocellular ballooning in liver biopsies from patients with NASH using persistent homology. Comput Methods Programs Biomed 2020;195:105614. doi:10.1016/j.cmpb.2020.105614
52. Gunter TD, Terry NP. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. J Med Internet Res 2005;7:e3. doi:10.2196/jmir.7.1.e3
53. Van Vleck TT, Chan L, Coca SG, et al. Augmented intelligence with natural language processing applied to electronic health records for identifying patients with non-alcoholic fatty liver disease at risk for disease progression. Int J Med Inf 2019;129:334–341. doi:10.1016/j.ijmedinf.2019.06.028
54. Corey KE, Kartoun U, Zheng H, Shaw SY. Development and Validation of an Algorithm to Identify Nonalcoholic Fatty Liver Disease in the Electronic Medical Record. Dig Dis Sci 2016;61:913–919. doi:10.1007/s10620-015-3952-x
55. Popa SL, Ismaiel A, Cristina P, et al. Non-Alcoholic Fatty Liver Disease: Implementing Complete Automated Diagnosis and Staging. A Systematic Review. Diagnostics 2021;11:1078. doi:10.3390/diagnostics11061078
56. Pournik O, Dorri S, Zabolinezhad H, Alavian SM, Eslami S. A diagnostic model for cirrhosis in patients with non-alcoholic fatty liver disease: an artificial neural network approach. Med J Islam Repub Iran 2014;28:116.
57. Hu C, Anjur V, Saboo K, et al. Low Predictability of Readmissions and Death Using Machine Learning in Cirrhosis. Am J Gastroenterol 2021;116:336–346. doi:10.14309/ajg.0000000000000971
58. Kanwal F, Taylor TJ, Kramer JR, et al. Development, Validation, and Evaluation of a Simple Machine Learning Model to Predict Cirrhosis Mortality. JAMA Netw Open 2020;3:e2023780. doi:10.1001/jamanetworkopen.2020.23780
59. Wei R, Wang J, Wang X, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. EBioMedicine 2018;35:124–132. doi:10.1016/j.ebiom.2018.07.041
60. Konerman MA, Lu D, Zhang Y, et al. Assessing risk of fibrosis progression and liver-related clinical outcomes among patients with both early stage and advanced chronic hepatitis C. PLoS One 2017;12:e0187344. doi:10.1371/journal.pone.0187344
61. Shousha HI, Awad AH, Omran DA, Elnegouly MM, Mabrouk M. Data Mining and Machine Learning Algorithms Using *IL28B* Genotype and Biochemical Markers Best Predicted Advanced Liver Fibrosis in Chronic Hepatitis C. Jpn J Infect Dis 2018;71:51–57. doi:10.7883/yoken.JJID.2017.089
62. Lara J, López-Labrador FX, González-Candelas F, Berenguer M, Khudyakov YE. Computational models of liver fibrosis progression for hepatitis C virus chronic infection. BMC Bioinformatics 2014;15(Suppl 8):S5. doi:10.1186/1471-2105-15-S8-S5
63. Eaton JE, Vesterhus M, McCauley BM, et al. 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. doi:10.1002/hep.30085
64. Lau L, Kankanige Y, Rubinstein B, et al. Machine-Learning Algorithms Predict Graft Failure After Liver Transplantation. Transplantation 2017;101:e125–e132. doi:10.1097/TP.0000000000001600
65. Matalka II, Al-Omari FA, Salama RM, Mohtaseb AH. A novel approach for quantitative assessment of mucosal damage in inflammatory bowel disease. Diagn Pathol 2013;8:156. doi:10.1186/1746-1596-8-156

66. Seeley EH, Washington MK, Caprioli RM, M'Koma AE. Proteomic patterns of colonic mucosal tissues delineate Crohn's colitis and ulcerative colitis. Proteomics Clin Appl 2013;7:541–549. doi:10.1002/prca.201200107
67. Jeong CS, Kim D. Inferring Crohn's disease association from exome sequences by integrating biological knowledge. BMC Med Genomics 2016;9(Suppl 1):35. doi:10.1186/s12920-016-0189-2
68. Chen G, Shen J. Artificial Intelligence Enhances Studies on Inflammatory Bowel Disease. Front Bioeng Biotechnol 2021;9:635764. doi:10.3389/fbioe.2021.635764
69. Waljee AK, Wallace BI, Cohen-Mekelburg S, et al. Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease. JAMA Netw Open 2019;2:e193721. doi:10.1001/jamanetworkopen.2019.3721
70. Waljee AK, Lipson R, Wiitala WL, et al. Predicting Hospitalization and Outpatient Corticosteroid Use in Inflammatory Bowel Disease Patients Using Machine Learning. Inflamm Bowel Dis 2018;24:45–53. doi:10.1093/ibd/izx007
71. Chung CF, Wang Q, Schroeder J, et al. Identifying and Planning for Individualized Change: Patient-Provider Collaboration Using Lightweight Food Diaries in Healthy Eating and Irritable Bowel Syndrome. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019;3:27. doi:10.1145/3314394
72. Zia J, Schroeder J, Munson S, et al. Feasibility and Usability Pilot Study of a Novel Irritable Bowel Syndrome Food and Gastrointestinal Symptom Journal Smartphone App. Clin Transl Gastroenterol 2016;7:e147. doi:10.1038/ctg.2016.9
73. Fukui H, Nishida A, Matsuda S, et al. Usefulness of Machine Learning-Based Gut Microbiome Analysis for Identifying Patients with Irritable Bowels Syndrome. J Clin Med 2020;9:2403. doi:10.3390/jcm9082403
74. Kudo SE, Ichimasa K, Villard B, et al. Artificial Intelligence System to Determine Risk of T1 Colorectal Cancer Metastasis to Lymph Node. Gastroenterology 2021;160:1075-1084.e2. doi:10.1053/j.gastro.2020.09.027
75. Kong J, Lee H, Kim D, et al. Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. Nat Commun 2020;11:5485. doi:10.1038/s41467-020-19313-8
76. Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. Butte AJ, editor. PLoS Med 2019;16:e1002730. doi:10.1371/journal.pmed.1002730
77. Yamashita R, Long J, Longacre T, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol 2021;22:132–141. doi:10.1016/S1470-2045(20)30535-0
78. Ting WC, Lu YCA, Ho WC, Cheewakriangkrai C, Chang HR, Lin CL. Machine Learning in Prediction of Second Primary Cancer and Recurrence in Colorectal Cancer. Int J Med Sci 2020;17:280–291. doi:10.7150/ijms.37134
79. Zhang SM, Wang YJ, Zhang ST. Accuracy of artificial intelligence-assisted detection of esophageal cancer and neoplasms on endoscopic images: A systematic review and meta-analysis. J Dig Dis 2021;22:318–328. doi:10.1111/1751-2980.12992
80. Cai SL, Li B, Tan WM, et al. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). Gastrointest Endosc 2019;90:745-753.e2. doi:10.1016/j.gie.2019.06.044
81. Jin P, Ji X, Kang W, et al. Artificial intelligence in gastric cancer: a systematic review. J Cancer Res Clin Oncol 2020;146:2339–2350. doi:10.1007/s00432-020-03304-9
82. Wu L, Zhou W, Wan X, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. Endoscopy 2019;51:522–531. doi:10.1055/a-0855-3532
83. Ashat M, Klair JS, Singh D, Murali AR, Krishnamoorthi R. Impact of real-time use of artificial intelligence in improving adenoma detection during colonoscopy: A systematic review and meta-analysis. Endosc Int Open 2021;9:E513–E521. doi:10.1055/a-1341-0457
84. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–1819. doi:10.1136/gutjnl-2018-317500
85. Wang P, Liu P, Glissen Brown JR, et al. Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study. Gastroenterology 2020;159:1252-1261.e5. doi:10.1053/j.gastro.2020.06.023
86. Pullens HJ, van der Stap N, Rozeboom ED, et al. Colonoscopy with robotic steering and automated lumen centralization: a feasibility study in a colon model. Endoscopy. 2015;48:286–290. doi:10.1055/s-0034-1392550
87. Stidham RW, Liu W, Bishu S, et al. Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis. JAMA Netw Open 2019;2:e193963. doi:10.1001/jamanetworkopen.2019.3963
88. Gottlieb K, Requa J, Karnes W, et al. Central Reading of Ulcerative Colitis Clinical Trial Videos Using Neural Networks. Gastroenterology 2021;160:710-719.e2. doi:10.1053/j.gastro.2020.10.024
89. Mohan BP, Khan SR, Kassab LL, et al. High pooled performance of convolutional neural networks in computer-aided diagnosis of GI ulcers and/or hemorrhage on wireless capsule endoscopy images: a systematic review and meta-analysis. Gastrointest Endosc 2021;93:356-364.e4. doi:10.1016/j.gie.2020.07.038
90. Aoki T, Yamada A, Aoyama K, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest Endosc.2019;89:357-363.e2. doi:10.1016/j.gie.2018.10.027
91. Minoda Y, Ihara E, Komori K, et al. Efficacy of endoscopic ultrasound with artificial intelligence for the diagnosis of gastrointestinal stromal tumors. J Gastroenterol 2020;55:1119–1126. doi:10.1007/s00535-020-01725-4
92. Marya NB, Powers PD, Chari ST, et al. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. Gut 2021;70:1335–1344. doi:10.1136/gutjnl-2020-322821
93. Tonozuka R, Itoi T, Nagata N, et al. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: a pilot study. J Hepatobiliary Pancreat Sci 2021;28:95–104. doi:10.1002/jhbp.825
94. Baecker A, Kim S, Risch HA, et al. Do changes in health reveal the possibility of undiagnosed pancreatic cancer? Development of a risk-prediction model based on healthcare claims data. Real FX, editor. PLoS One 2019;14:e0218580. doi:10.1371/journal.pone.0218580
95. Kenner BJ, Abrams ND, Chari ST, et al. Early Detection of Pancreatic Cancer: Applying Artificial Intelligence to Electronic Health Records. Pancreas 2021;50:916–922. doi:10.1097/MPA.0000000000001882
96. Sharma A, Kandlakunta H, Nagpal SJS, et al. Model to Determine Risk of Pancreatic Cancer in Patients With New-Onset Diabetes. Gastroenterology 2018;155:730-739.e3. doi:10.1053/j.gastro.2018.05.023

97. Pereira SP, Oldfield L, Ney A, et al. Early detection of pancreatic cancer. Lancet Gastroenterol Hepatol 2020;5:698–710. doi:10.1016/S2468-1253(19)30416-9
98. Nakhleh MK, Amal H, Jeries R, et al. Diagnosis and Classification of 17 Diseases from 1404 Subjects *via* Pattern Analysis of Exhaled Molecules. ACS Nano 2017;11:112–125. doi:10.1021/acsnano.6b04930
99. Peng RD. Reproducible research in computational science. Science 2011;334:1226–1227. doi:10.1126/science.1213847