

87137_Auto_Edited-check.docx

Name of Journal: *World Journal of Gastroenterology*

Manuscript NO: 87137

Manuscript Type: OPINION REVIEW

**Challenges involved in the application of artificial intelligence in gastroenterology:
the race is on!**

Christou CD *et al.* Challenges of AI in gastroenterology

Abstract

Gastroenterology is a particularly data-rich field, generating vast repositories of data that are fruitful ground for artificial intelligence (AI) and machine learning (ML) applications. In this opinion review, we initially elaborate on the current status of the application of AI/ML-based software in gastroenterology. Currently, AI/ML-based models have been developed in the following applications: Models integrated into the clinical setting following real-time patient data flagging patients at high risk for developing a gastrointestinal disease, models employing non-invasive parameters that provide accurate diagnoses aiming to either replace, minimize or refine the indications of endoscopy, models utilizing genomic data to diagnose various gastrointestinal diseases, computer-aided diagnosis systems facilitating the interpretation of endoscopy images, models to facilitate treatment allocation and predict the response to treatment, and finally, models in prognosis predicting complications, recurrence following treatment, and overall survival. Then, we elaborate on several challenges and how they may negatively impact the widespread application of AI in healthcare and gastroenterology. Specifically, we have elaborated on concerns regarding accuracy, cost-effectiveness, cybersecurity, interpretability, oversight, and liability. While AI is unlikely to replace physicians, it will transform the skillset demanded by future physicians to practice. Thus, physicians are expected to engage with AI to avoid becoming obsolete.

Key Words: Artificial intelligence; Machine learning; Gastroenterology; Cost-effectiveness; Interpretability; Accuracy

Christou CD, Tsoulfas G. Challenges involved in the application of artificial intelligence in gastroenterology: the race is on!. *World J Gastroenterol* 2023; In press

Core Tip: Currently, artificial intelligence (AI) and machine learning (ML) have several applications in the prevention, diagnosis, treatment, and prognosis of various

gastrointestinal diseases, including gastroesophageal reflux disease, esophageal cancer, gastric cancer, gastrointestinal bleeding, inflammatory bowel diseases, polyps, colorectal cancer, and others. Despite their promising results, AI/ML applications in gastroenterology are hindered by several challenges, including accuracy, cost-effectiveness, cybersecurity, interpretability, oversight, and liability concerns. In this opinion review, we elaborate on these challenges and present different ways to overcome them.

INTRODUCTION

Gastroenterology is a particularly data-rich field, with data being omnipresent. Several gastroenterology norms, including patient electronic records, imaging, endoscopy images and videos, capsule endoscopy videos, genomic analyses, microbiome data, and histopathology images, generate vast data repositories^[1-3]. Conforming with big data features, these repositories provide a fruitful ground for artificial intelligence (AI) applications and, more specifically, AI's subfield, machine learning (ML)^[4]. By learning from these data, ML algorithms are able to predict future data points^[5]. AI/ML applications have been described in many studies as the potential solution to long-standing healthcare challenges, including the need for cost reduction, enhancing diagnostic accuracy, facilitating individualized and evidence-based care, and, more importantly, significantly ameliorating the morbidity and mortality associated with various diseases^[6-8].

In this opinion review, we first present the current status of the application of AI in the field of gastroenterology. Then, we pose a series of challenges these applications face and explore ways to overcome them. Specifically, we elaborate on the following challenges: Accuracy concerns, cost-effectiveness, cybersecurity, interpretability, oversight, and liability.

Current status

Physicians use experience, intuition, and quantifiable and non-quantifiable variables for decision-making. ML algorithms, on the other hand, perform a series of precise calculations of all the quantifiable variables for decision-making. Computer systems, in general, and AI/ML tools, particularly, surpass, by far, physicians in their ability to quantify multiple correlations, even in fields where the physicians hold in-depth expertise. Thus, the application of ML tools is particularly useful when the “truth” lies within the data. AI classifiers commonly employed in ML tools in gastroenterology are support vector machines (SVM), artificial neural networks (ANN), and convolutional neural networks (CNNS). SVMs are used in supervised learning. Data points of the instances are mapped in a high-dimensional space, where the hyperplane that separates the instances based on their assigned label, retaining the highest performance, is selected^[9]. SVMs can be used for linear and non-linear problems using kernel functions^[10]. ANNs are inspired by the human brain and consist of an input and output layer with at least one hidden layer in between^[11]. As the architectures of ANNs became more sophisticated with the addition of multiple hidden layers and even more layer connections, the concept of deep neural networks (DNNs) emerged^[12,13]. A particular type of DNN, the CNN, has found profound application in gastroenterology since the demonstrated high performance in image interpretation. CNNs are based on convolution, where the image is processed using a series of filters or kernels to detect patterns within the image, such as edges and textures^[14].

Table 1 presents a series of studies where AI/ML models have been applied in the field of gastroenterology. These models could be integrated into the clinical setting and follow real-time patient data trends. Then, these models could flag patients at high risk for developing a plethora of gastrointestinal diseases. A notable example is ColonFlag, an ML algorithm used to stratify the risk of developing colorectal cancer. Impressively, the tool has been shown to identify patients with colorectal cancer who would otherwise have avoided screening and to identify patients at the early stages of the disease 18 to 24 mo prior to the usual diagnosis time^[15,16]. Other recent studies focus on the stratification of gastric cancer development by employing electronic health care

records^[17,18]. These models demonstrate how AI/ML models can be used at the primary care level to lessen the burden of a plethora of gastrointestinal diseases significantly.

Currently, endoscopy constitutes the gold standard in the diagnosis of various gastrointestinal diseases. Nevertheless, endoscopy is an unpleasant intervention for the patient, has a significant cost, and is related to a series of potential complications^[19]. AI/ML models utilizing non-invasive parameters that provide comparably accurate diagnoses could either replace, minimize, or refine the indications of endoscopy. Two studies have used data from microbiota and microbiome to identify patients with inflammatory bowel disease (IBD)^[20,21]. Although these models could not replace endoscopy in the diagnosis of IBDs, they can be used to identify patients requiring endoscopic evaluation. However, in the diagnosis of gastroesophageal reflux disease, the employment of ML tools could substitute the use of endoscopy for patients at low risk of developing cancer^[22]. These efforts demonstrate how ML tools could be used to significantly reduce the cost and complications related to endoscopy, avoiding, at the same time, an unpleasant experience for the patient.

The introduction of next-generation sequencing has widened our capabilities regarding detecting host genetic factors related to certain diseases and allowed us to unveil the complex hereditary background of several diseases by clarifying the cascade of gene interactions and the impact of the environment^[23,24]. This progress allowed for the emergence of novel biomarkers, including microRNAs. In gastroenterology, various studies have employed microRNA profiling data to develop tools able to diagnose various gastrointestinal diseases. Specifically, tools have been developed that classify non-tumor from tumor mucosa from gastric samples^[25]. In addition, by employing microRNA profiling, data from serum models have been developed that identify patients with gastric cancer^[26], pancreatic ductal adenocarcinoma^[27], colorectal cancer^[28], and ulcerative colitis^[29]. Finally, a model has been developed using microRNA profiling from colonic samples to predict the response to treatment of patients with active severe ulcerative colitis^[30]. These models again demonstrate how AI/ML tools can provide a non-invasive endoscopy alternative.

Except for replacing endoscopy, AI/ML tools could significantly enhance the endoscopy's clinical outcomes. Such models could be used for developing Computer Aided Detection or Diagnosis systems (CAD systems) to facilitate the interpretation of endoscopy images. Firstly, these CAD systems could aid the endoscopist to navigate in the gastrointestinal tract. Existing models have demonstrated outstanding performance^[31]. These CAD systems could be used as a "third eye" for the endoscopist, an "eye" that does not get tired, is not susceptible to any distractions, and can identify lesions missed by the endoscopist. Following the identification of lesions, these models could be used to classify the lesions, for example, as benign or malignant. In capsule endoscopy, CAD systems could automatically identify and then classify lesions from capsule endoscopy videos and then provide the images to the gastroenterologist for review. Such models could significantly decrease the cost associated with the labor hours required to review the capsule endoscopy videos. Several models have been developed for a plethora of gastrointestinal diseases, including IBDs, gastric cancer, small bowel lesions (ulcerative/hemorrhagic), colon cancer, and others^[6,32]. Another application of CAD systems is proposing an ideal site for biopsy when needed. CAD systems also find a particular application in colonoscopy in classifying benign from malignant polyps^[33]. By avoiding the removal of benign polyps and correctly removing all malignant polyps with the assistance of CAD systems, the cost-efficiency of colonoscopy significantly increases. Notably, regarding polyp classification, the reported negative predictive value in the literature of CAD systems is above the 90% recommended by the American Society of Gastrointestinal Endoscopy for adenoma detection^[34].

In addition to diagnosis, AI/ML-based tools can be employed to facilitate treatment allocation for various gastrointestinal diseases. Data regarding impedance and acid exposure time were employed in a study to develop an ML tool predicting treatment response^[35]. Other studies focusing on IBDs have developed tools that predict treatment response to specific drugs or the need for operation^[35,36]. A different group of studies, which focused on gastrointestinal bleeding management, achieved the development of

models identifying patients in need of a transfusion, emergent endoscopy, emergent surgery, and intensive care unit triage^[37-40]. These studies demonstrate how AI/ML tools could be used to create frameworks for individualized, evidence-based treatment allocations.

AI/ML algorithms have also been employed in the development of predictive tools able to predict complications, recurrence following treatment, and overall survival. Examples of such studies include a study developing an ML tool to predict overall survival in patients with esophageal cancer^[41], a model predicting in-hospital death for patients with gastrointestinal bleeding^[42], and an ML model predicting the chance of recurrence in operated gastric cancer patients^[43]. Such models could be used for family and patient counseling.

Accuracy concerns

The accuracy of the developed AI/ML tool largely depends on the quality of the training data. The notion “garbage in, garbage out” is particularly true in ML models. Flaws and weaknesses of the datasets, such as wrongly labeled instances, low image quality, small dataset size and variability of instances, and discrepancies in the data collection processes, will be inadvertently built into the end model. For example, capsule endoscopy has significantly lower image resolution compared with other types of endoscopy images, and thus, a model trained from capsule endoscopy imaging is expected to have lower classification accuracy. Standardization of data collection methods is essential to acquire datasets of clean, high-quality data that are representative of a diverse patient population. A main challenge encountered in ML development is class imbalance. Most of the time, normal findings represent the majority class, while patients with a disease are usually the minority class. This problem is amplified in multi-class situations where accurate classification of rare classes is particularly challenging. A series of techniques to undersample the majority class and oversample the minority class in medical datasets have been proposed that, to some extent, resolve this issue^[44]. In addition, in a study, researchers developed a model able

to add polyps in the endoscopic images, which can then be used for training ML models^[45]. Such methodologies of creating synthetic instances could again resolve the imbalance challenge. A different challenge in ML model development is underfitting, which can be defined as the incapability of the model to capture the variability of the data^[46]. On the other hand, overfitting can be described as the inability of the model to retain its performance in datasets other than the training dataset^[47]. In other words, the model's performance is significantly superior in the training set and drops in new datasets. CNNs, which are widely used in gastroenterology, are vulnerable to overfitting^[14]. There is a series of proposed methodologies to resolve underfitting and overfitting challenges, including early stopping/dropout, penalty method/regularization, adjusting the learning rate and the optimizer, weight matrix initialization, data augmentation, and batch normalization^[46,48]. Another intriguing point is that ML models are trained from datasets not practically available in the clinical setting. The different training variables are derived at various time points during the treatment of the patient and are not readily available at a given time. These missing parameters could become a shortcoming anchoring the use of ML in the clinical setting. All the above challenges have a common "true" solution. Physicians should be included in the development of ML models from the early stages of the procedure. Effective communication among physicians, the developing team, and the investors could address and resolve any potential shortcomings of these models. The key is to involve physicians in all the stages of the process. Physicians should be involved in problem identification to ensure that the developed tool addresses actual needs and in data collection and annotation to ensure that the data are labelled correctly, which is crucial in supervised learning. Additionally, we believe that the following steps should be followed during model development to avoid potential drawbacks: physicians should provide a detailed description of the problem at hand to the developing team, explain what is requested by the model (the endpoint), describe the features thoroughly, cooperate to identify and engineer features of high predictive value, discuss the sample size needed based on the nature of the task (binary classification, multi-class

classification, regression), ensure that the patients included in the developing procedure are a representative sample of the targeted group, conduct a rigorous clinical validation before the clinical application of the model, provide input on how the tool can be seamlessly be integrated into the existing clinical workflows without causing disruption, discuss the prospects of real-time learning, where the model continues to learn and improves following clinical application, and provide insights into ethical considerations including ensuring patient privacy and avoiding potential biases. Even following the model's application in the clinical setting is essential to establish a framework of continuous monitoring and feedback from healthcare professionals to address model shortcomings and improve the model's efficiency. Finally, when applying AI/ML-based tools in the clinical setting is crucial to collect data on the impact of the use of such tools on the clinical outcome, the wellbeing of the patient, and the patient-physician relation. Failing to cooperate efficiently and to communicate what exactly is expected from the model could result in an end product that does not meet the expectations of the physicians and jeopardizes the safety of the patients.

Cost-effectiveness

Unfortunately, resources are not infinite. Especially in an era of technological advancement, with other cutting-edge technologies, such as three-dimensional printing and bioprinting, virtual and augmented reality, regenerative medicine, and robotics that also promise to facilitate patient care, AI must prove that it is worth investing in. In reality, AI, which has been at the forefront for many decades, has not yet been able to provide the expected results in healthcare. This is evident during AI winters, where investors demonstrate their frustration by cutting funding^[4]. The expectations are often placed extremely high, and from their perspective, AI overpromises and constantly underdelivers. As mentioned above, effective communication among the physicians, the developing team, and the investors is crucial in agreeing on the expectations and avoiding disappointment in the end product. Nevertheless, even if physicians, developers, and investors are satisfied, the true success of the model should be assessed

in association with the clinical outcome. Thus, conducting a series of cost-effective analyses is crucial to determine whether AI applications' benefits in healthcare are worth the associated cost. The costing of new technologies is generally challenging due to the lack of pre-existing data, and it involves bottom-up costing approaches, which typically require more resources and are time-consuming^[49]. The cost associated with new technologies tends to drop over time; thus, cost estimation studies will need to be updated throughout the lifecycle of the AI application. Despite its importance, the topic of the cost-effectiveness of the application of AI in medicine is poorly investigated, mainly due to the lack of real AI applications in the clinical setting. There are several ways to improve the cost-effectiveness of implementing AI/ML-based technologies in healthcare. First, a series of models has been developed and patented with a high performance in performing several tasks, such as CAD systems that classify benign from malignant polyps. Thus, employing existing or pre-trained models rather than developing new tools from scratch could significantly decrease the cost of adopting these technologies in clinical practice. When, however, healthcare institutions decide to invest in developing AI/ML-based tools from scratch, a plethora of approaches could be adopted to improve the cost-effectiveness, including the utilization of collaborative platforms and open-source tools, run models on existing hardware when possible to reduce the need of specialized hardware, use of existing datasets when available to avoid data acquisition costs, and finally focus on developing tools that allow for the automatization of routine, time-consuming tasks while prioritizing high-incidence conditions with the potential for significant cost savings. One of the few studies addressing this issue is a recent study in the field of gastroenterology, where researchers performed a Markov model microsimulation of AI application in screening colonoscopy and concluded that at the United States population level, AI application could prevent more than 7 thousand colorectal cancer cases more than two thousand related deaths, and provide a yearly saving of 290 million dollars^[50]. These results are very promising, and as soon as AI models are widely applied in the clinical setting, studies proving their cost-effectiveness must be conducted.

Cybersecurity

Cybersecurity is a challenge not exclusive to AI but to the tendency of digitalization in general. Nevertheless, AI introduces further concerns since it requires a wholly electronic tracking of patient records. This raises concerns that a huge amount of sensitive data is vulnerable to massive disclosures^[51]. For example, a transfer of 1.6 million patient records to Google DeepMind was ruled illegal in the United Kingdom^[52]. Additionally, due to the sensitive nature of data (including financial data), healthcare facilities are essentially attractive targets for cyberattacks. The introduction of AI in healthcare allows for the emergence of new vulnerabilities. Imagine an intrusion into an ML, which is now manipulated into making wrong decisions or even introducing malevolent data. For example, in a study, the researchers demonstrated how a DNN was used to add or remove lung tumors from computed tomography scans at a level of accuracy that remained undetected by specialized radiologists^[53]. Imagine a similar intrusion that will introduce colorectal cancer images during real-time colonoscopy. Such incidents could severely undermine any trust in AI in healthcare and have severe consequences on patient outcomes. How could patients and physicians be confident about any AI model suggestions, knowing they can be manipulated? Evidently, the information technology infrastructure will need to be expanded in order to accommodate AI applications and shield these models from potential threats.

Interpretability

Interpretability in ML refers to the degree of explainability and transparency of the model^[54]. In other words, whether the logic behind the model's decisions is understandable or not. Imagine interpretability as a spectrum, where at the "white" end, we find "white-box" models, where their inner workings and decision-making processes are fully explainable and interpretable. At the other end of the spectrum, we find "black-box" models, which are characterized by a highly complex architecture to the extent that we are unable to comprehend their inner mechanisms and decision-

making processes. In simple tasks, for example, when the features and target variable have a linear relation, white box models demonstrate a sufficiently high performance, rendering the use of black box models redundant. However, in complex tasks with high-dimensional data, black boxes exhibit superior performance to white boxes due to their ability to capture intricate patterns and relationships within the data. However, the application of black box models in healthcare introduces a series of challenges, including undermining the trust towards AI, inability to oversight any biases or prejudiced decisions towards minority groups, failure to justify our AI-assisted decision to patients, and violation of fundamental ethical tenets that physicians should hold a basic comprehension of the devices they use. Avoiding using “black-box” models due to the above challenges could significantly handicap the potential of AI applications in healthcare. Thus, we should aim for methodologies that allow for a level of reasoning of “black-box” model predictions. Interpretability can be divided into global and local interpretability. Introducing a level of interpretability in black boxes applied in healthcare is crucial to enhance patient and physician trust in these tools and facilitate clinical decision-making. Interpretability can be divided into global and local. Global interpretability offers a level of explainability of the model as a whole^[55]. Ways to improve global interpretability include feature importance analysis, surrogate models, and interactive visualization tools developed to allow users to explore how the various features influence the model’s predictions^[55-58]. For example, a model stratifying the risk of colorectal cancer development could provide its prediction along with a notification that the prediction is based mainly on the patient’s sex, age, and blood count. On the other hand, local interpretability provides the reasoning behind individual predictions^[55]. Methodologies to enhance local interpretability include surrogate models, Shapley values, saliency regions, and visualization techniques^[55-58]. Shapley values could be used to demonstrate how each value of each feature contributes to the model’s prediction, while saliency maps provide a *post-hoc* visualization to comprehend which parts of the input were used by the model to reach its decision and are particularly helpful for CNNs to visualize which parts of the image were used in the

model's interpretation. For example, a saliency map would highlight exactly which parts of a biopsy the model focused on to reach its diagnosis. Providing a level of explainability for "black-box" models is key to gaining the trust of both physicians and patients and will play a pivotal role in determining which models will dominate the industry.

Oversight and liability concerns

As acknowledged by the Food and Drug Administration (FDA), applications regarding AI/ML-based software marketing in healthcare have increased exponentially^[59,60]. In early 2021, the FDA published the Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan³, while in 2023, it published the Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions²^[59,61]. Similarly, the European Union has announced its first regulation on AI applications, the EU AI Act^[62]. Evidently, governments feel the need for AI oversight and regulation. Regarding law and regulation, the true challenge for the future is to find the golden mean between too little oversight, which may have devastating and severe consequences for the well-being of our patients, and too strict regulation, which will handicap innovation. AI/ML-based tools are unique in the way that they constantly evolve following the initial audit and marketing. Thus, as the model encompasses more data and is retrained, it diverges from the original product to the extent that it can be considered an entirely different product at different time points. Who ensures the safety and credibility of the product at different time points? Evidently, a lifecycle regulatory framework is needed that provides post-marketing oversight, or alternative time-limited authorizations should be granted with periodic audits when several modifications have been made to the initial product^[63].

Liability is the main physician-related concern that could undermine the widespread application of AI in healthcare. Even if the research succeeds in providing highly accurate, reliable AI models, for AI to impact the clinical outcome, physicians will need

to accept and adopt these models in their everyday lives. Liability could be the main concern that discourages physicians from doing so. To evade ¹malpractice liability, a physician is expected to provide care at the same level as a competent, same-specialty physician, considering the available resources^[64]. However, the introduction of AI makes liability a far more complicated matter. An intriguing concept to consider is the potential legal binding of AI's decisions in the future. As AI/ML-based software continues to advance and grow, it is reasonable to assume that they may eventually outperform physicians, at least in certain tasks. In such a scenario, how can physicians justify ignoring the decisions presented by AI, especially when AI's decisions are solely data-driven and lack any subjective bias?

Furthermore, who should be held liable if following the AI model's decision leads to patient harm? Currently, no legal precedent clarifies liability in cases where a patient suffers an injury due to an inaccurate output generated by AI/ML-based software^[65]. In a published legal analysis, the authors present a series of scenarios based on the various combinations of whether the physicians follow the AI recommendation, whether the AI recommendation is correct, ¹and whether a patient injury occurs. The authors' conclusion highlights that since existing laws protect physicians from liability when adhering to standard care, it incentivizes them to downplay AI's actual usefulness, potentially turning AI into a mere confirmatory tool instead of a tool to enhance the quality of care^[65]. Until a comprehensive legal framework regarding liability is established, healthcare facilities have valid concerns and may hesitate to adopt AI technologies due to fears and uncertainties about potential liabilities they may expose the facility and its staff to.

CONCLUSION

In this opinion review, we initially elaborate on the current status of the application of AI/ML-based software in gastroenterology. Currently, AI/ML-based models have been developed in the following applications: Models integrated into the clinical setting following real-time patient data flagging patients at high risk for developing a

gastrointestinal disease, models employing non-invasive parameters that provide accurate diagnoses aiming to either replace, minimize or refine the indications of endoscopy, models utilizing genomic data to diagnose various gastrointestinal diseases, CAD systems facilitating the interpretation of endoscopy images, models to facilitate treatment allocation and predict the response to treatment, and finally, models in prognosis predicting complications, recurrence following treatment, and overall survival that can be used for family and patient counseling. Then, we elaborate on several challenges and how they may negatively impact the widespread application of AI in healthcare and gastroenterology, offering possible ways to overcome them. Specifically, we have elaborated on concerns regarding accuracy, cost-effectiveness, cybersecurity, interpretability, oversight, and liability.

It is worth noting that the vast majority of AI research originates from high-income countries^[66]. Despite being described as the solution that could narrow the gap in quality of care between high and low-income countries, AI may become one more brick in the wall of inequality. To avoid that, future AI initiatives should include middle and low-income countries. In any healthcare system where AI/ML-based software will be integrated, it is essential to simultaneously integrate equally sophisticated oversight mechanisms to ensure the safety of patients; otherwise, physicians may encounter unintended ramifications.

As mentioned before, AI/ML models' decision-making is based on quantifiable parameters. However, clinicians rely on non-quantifiable parameters for decision-making. It is crucial to state that overreliance on AI could impair our critical thinking and reasoning. Thus, the ultimate goal is not AI-driven but AI-assisted clinical practice. AI could be the means to augment our intelligence by embracing the inherent complexity of medicine. Envision the hospital of the future where hospitals are fully robotic, AI and robots conduct the majority of procedures, 3D printers are used to print almost everything, from medical equipment to artificial parts, telemedicine allows patients to stay at home, virtual and augmented reality will assist physicians at all procedures, and biosensors are placed at birth diagnosing all diseases at an early stage

and alerting physicians^[67-69]. It is unlikely that AI will replace physicians. Nevertheless, it is also unlikely that the skill set demanded by future physicians will be in any way similar to the present. In this upcoming transformation of healthcare with the integration of cutting-edge technologies, AI is expected to be at the center. Thus, physicians are expected to engage with AI to avoid becoming obsolete.

2%

SIMILARITY INDEX

PRIMARY SOURCES

1

Chrysanthos D Christou, Georgios Tsoulfas.
"Challenges and opportunities in the application of
artificial intelligence in gastroenterology and hepatology",
World Journal of Gastroenterology, 2021
Crossref

60 words — 1%

2

www.lw.com
Internet

21 words — < 1%

3

www.digitalhealthnews.eu
Internet

20 words — < 1%