

# NARRATIVE REVIEW

## Artificial Intelligence and the Future of Gastroenterology and Hepatology



Daniel D. Penrice, Puru Rattan, and Douglas A. Simonetto

*Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota*

**The integration of artificial intelligence (AI) into gastroenterology and hepatology (GI) will inevitably transform the practice of GI in the coming decade. While the application of AI in health care is not new, advancements are occurring rapidly, and the future landscape of AI is beginning to come into focus. From endoscopic assistance via computer vision technology to the predictive capabilities of the vast information contained in the electronic health records, AI promises to optimize and expedite clinical and procedural practice and research in GI. The extensive body of literature already available on AI applications in gastroenterology may seem daunting at first; however, this review aims to provide a breakdown of the key studies conducted thus far and demonstrate the many potential ways this technology may impact the field. This review will also take a look into the future and imagine how GI can be transformed over the coming years, as well as potential limitations and pitfalls that must be overcome to realize this future.**

*Keywords:* Big data, Future of healthcare, Machine learning, Neural networks

### Introduction

The term artificial intelligence (AI) was first introduced in the 1950s and refers to the application of computers to perform complex tasks, such as solving problems and making nuanced decisions, that have traditionally been associated with human intelligence.<sup>1</sup> Interest in AI-driven health care and medicine has been exploding in recent decades, owing largely to the vast increase in data available since the establishment of the electronic health record (EHR). In fact, from 2008 to 2018, over 23,000 papers concerning the application of AI in health care have been published.<sup>2</sup> The current emphasis of AI in medicine is a direct response to the challenges facing health-care systems including increasingly complex patients, higher billing and coding requirements, and a growing population that leaves less time for interactions between a health-care provider and their patients than ever before. The hope is for a future where AI-driven health care can lead to improvements through all aspects of patient care by leveraging the vast quantities of data currently stored in the EHR.<sup>3,4</sup>

Machine learning (ML) lies at the crux of the most modern AI and encompasses algorithms on a wide spectrum

from “supervised” to “unsupervised” learning. These frameworks, as described in the previous issue by Rattan et al., allow researchers to accomplish tasks with accuracy, speed, and precision not previously possible.<sup>5,6</sup> The ability to organize and analyze large amounts of clinical information such as documentation, laboratory, and imaging data makes ML particularly powerful. As the majority of the information included in the EHR is stored as unstructured data, it has previously only been accessible to research via laborious review of individual charts. By accessing these data, ML algorithms can accomplish a wide variety of tasks such as disease identification and outcome prediction with accuracy that outperforms classical statistical methodology.<sup>5</sup>

The application of AI in gastroenterology has shown significant promise, primarily due to the strong body of ML research involving image recognition. These algorithms have already been applied to endoscopy with encouraging results and have subsequently led to improved disease detection and classification with greater accuracy than even the most experienced endoscopists.<sup>7</sup> However, many challenges such as identifying potential biases in algorithms, expanding generalizability, and increasing interpretability still exist.<sup>8</sup> Additionally, for these new advances to truly transform health care, widespread adoption is necessary. A substantial degree of skepticism exists in the medical community regarding the utility of AI in clinical practice, which will need to be addressed in the coming years.<sup>9</sup>

In this article, we highlight important studies utilizing AI in gastroenterology and hepatology to demonstrate the possibilities through which AI can transform clinical practice. We discuss how AI has been applied in various domains of gastroenterology and hepatology (GI) including upper endoscopy, colonoscopy, capsule endoscopy, pathology,

**Abbreviations:** AI, artificial intelligence; ANN, artificial neural network; BE, Barrett’s esophagus; CNN, convolutional neural network; DNN, deep neural network; EHR, electronic health record; SCC, squamous cell carcinoma; SVM, Support Vector Machine; WCE, wireless capsule endoscopy.

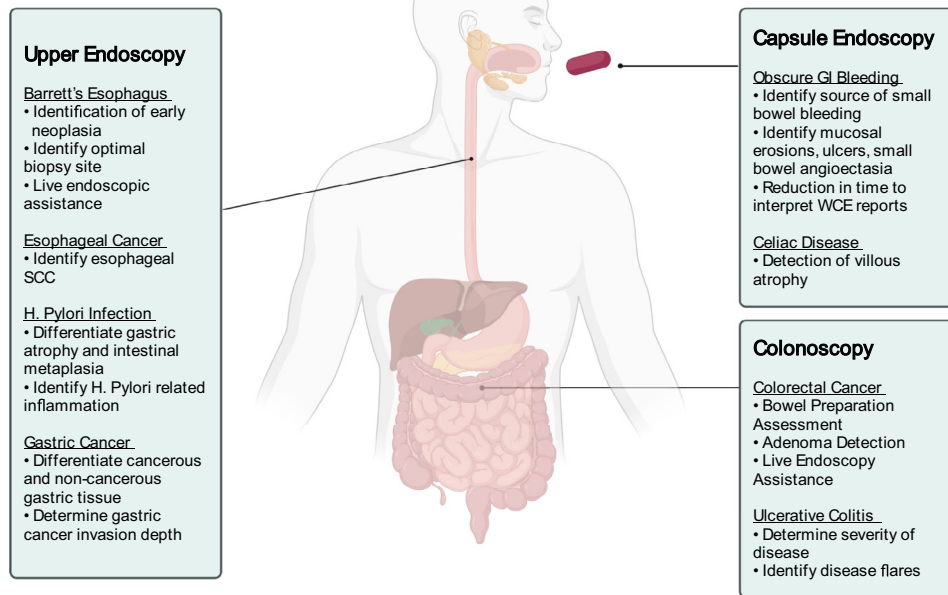
Most current article

Copyright © 2022 The Authors. Published by Elsevier Inc. on behalf of the AGA Institute. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2772-5723

<https://doi.org/10.1016/j.gastha.2022.02.025>

## Machine Learning in GI Endoscopy



**Figure 1.** Potential applications of artificial intelligence in GI endoscopy. SCC, squamous cell cancer; WCE, wireless capsule endoscopy.

radiology, and EHR data while also looking at potential breakthroughs on the horizon that promise to significantly improve patient care while alleviating the administrative burden on health-care providers.

### Computer Vision in Endoscopy

Over the past 10 years, research into the possible applications of novel ML algorithms in upper endoscopy, colonoscopy, and wireless capsule endoscopy (WCE) has been increasing exponentially. Already, major advancements have been made in identifying early stages of diseases, determining potential treatment options, and reducing the overall burden of endoscopists. The extensive work that has been done in understating the ways in which ML can transform endoscopic practice has been outlined in a number of high-quality reviews that focus on specific endoscopic modalities.<sup>7,10,11</sup> We will highlight notable studies in ML-driven endoscopy and the potential of this technology to revolutionize clinical practice over the coming years (Figure 1).

#### Upper Endoscopy

In upper endoscopy, much of the ML research thus far has focused on Barrett's esophagus (BE), esophageal squamous cell cancer (SCC), *Helicobacter pylori* (*H. pylori*) infection, and gastric cancer.

BE is the primary risk factor in the development of esophageal adenocarcinoma, which is associated especially with poor survival.<sup>12</sup> Screening for BE and early esophageal adenocarcinoma is challenging, as evidenced by the fact that only 1 in 10 cases of esophageal adenocarcinoma are diagnosed within a screening program.<sup>13</sup> Thus, an AI-enabled

aid that can assist in the detection of neoplastic changes in BE has been a focus of AI research and has the potential to improve clinical outcomes.<sup>14</sup> In 2016, a Support Vector Machine (SVM) model was created based on endoscopy images to identify early neoplastic changes in patients with BE and was able to do so with a sensitivity and specificity of 0.86 and 0.87, respectively.<sup>15</sup> An algorithm combining multiple ML models has also been developed to detect neoplasia from volumetric laser endomicroscopy images and was able to detect neoplasia with a sensitivity and specificity of 90% and 93%, respectively.<sup>16</sup> In 2020, de Groof et al<sup>17</sup> developed a deep learning model based on 1704 endoscopic images of nondysplastic BE and early stage neoplastic BE. This model was able to classify images as either neoplasia or nondysplastic BE with 90% sensitivity and 88% specificity. Additionally, when compared to general endoscopists, the system identified dysplasia with greater accuracy, 88% vs 73%.<sup>17</sup> This model also proved useful in identifying the optimal site for biopsy in over 92% of cases.<sup>18</sup> Another model was created using a convolutional neural network (CNN) trained on 916 endoscopic images of patients with histology-proven neoplastic BE and 919 control images from patients with nondysplastic BE. This model was able to detect early neoplasia with a sensitivity and specificity of 96.4% and 94.2%, respectively.<sup>19</sup> The first true real-time application of a deep learning model to identify early neoplastic BE was published by Ebigbo et al<sup>20</sup> in 2020 and deployed their previously developed CNN during live endoscopy to accurately identify cases of neoplastic BE with an accuracy of 89.9%, similar to that of expert endoscopists.

Esophageal SCC is another disease in gastroenterology where the possibilities of ML are beginning to be explored.

In 2019, a deep neural network (DNN) was trained on 2428 endoscopy images and was able to identify esophageal SCC with a sensitivity and specificity of 97.8% and 85.4%, respectively. Accuracy in identification was 91.4%, compared with 88.8% in a group of senior endoscopists and 77.2% in junior endoscopists. However, when the endoscopists were given access to the DNN tool to assist with decision-making, the diagnostic ability of all the endoscopists improved considerably.<sup>21</sup> A model was also created to specifically identify dysplasia and early esophageal SCC. This CNN-driven model was based on 6473 narrow-band images and was able to correctly identify dysplasia and early esophageal SCC with 98% sensitivity and 95% specificity.<sup>22</sup> These studies demonstrated the feasibility of deploying a ML-driven model during live endoscopy to assist with the diagnosis of esophageal SCC.

Upper endoscopy can be beneficial in the diagnosis of *H. pylori* gastritis; however, accuracy of diagnosis during endoscopy is quite difficult with a sensitivity of 62% and specificity of 89%.<sup>23</sup> An early ML model from 2004 involved a CNN trained on endoscopy images from 30 patients and was able to predict the presence of gastric atrophy, intestinal metaplasia, and *H. pylori*-related gastric inflammation with a sensitivity of 85% and a specificity of 90%.<sup>24</sup> Further progress was made in 2017 by Shichijo et al who developed a CNN based on over 30,000 endoscopic images that could be employed during live endoscopy to predict the presence of *H. pylori* infection. This model that was able to predict the presence of *H. pylori* infection with a sensitivity of 89% and a specificity of 87% was significantly quicker and more accurate than experienced endoscopists.<sup>25</sup>

Gastric cancer also presents an opportunity for improvements in care with ML, as diagnosing early gastric cancer is challenging, with reports of miss rates as high as 11% during upper endoscopy.<sup>26</sup> In 2012, Kubota et al<sup>27</sup> developed a model able to differentiate T1 through T4 gastric cancer staging with a modest accuracy of 64%. An improved model utilized a CNN trained on 790 endoscopic images that was able to determine depth of tumor invasion with an area under the curve (AUC) of 0.94. This model demonstrated a higher accuracy and specificity in determining tumor invasion than human endoscopists.<sup>28</sup> Another model was developed to identify early gastric cancer, in which a CNN was trained on over 24,000 images and was able to differentiate early gastric cancer with an accuracy of 92.5%, sensitivity of 94%, and specificity of 91%.<sup>29</sup>

## Colonoscopy

The majority of ML research in colonoscopy has focused on colorectal cancer, from improving the detection of colorectal cancer during colonoscopy to ensuring adequate bowel preparation and ultimately in improving adenoma-detection rates. While most attention in colonoscopy-related ML has been focused on improving adenoma-detection rates, ensuring adequate bowel preparation is also essential as poor prep can result in greater than 30%

adenoma miss rate.<sup>30</sup> A reliable and reproducible method to evaluate the effectiveness of bowel preparation has previously been developed in the Boston Bowel Preparation Scale; however, interobserver reliability in application of the scale has been shown to be only 0.74.<sup>31</sup> To address this important problem, a CNN was trained by Zhou et al<sup>32</sup> in 2020 that applied Boston Bowel Preparation Scale to live endoscopy with over 93% accuracy, significantly better than human endoscopists. This is a system that, if applied in the clinical setting, could potentially lead to decreased rate of missed adenomas.<sup>32</sup>

Overall, adenoma miss rate during colonoscopy has been shown to be as high as 25%, with a colorectal cancer miss rate of approximately 5%.<sup>33,34</sup> While there have been over 50 published studies detailing ML-driven algorithms built to detect malignant and premalignant colorectal lesions, essential studies in ML polyp detection demonstrate the greatest potential for advancements in this field, especially those that implement real-time ML-driven polyp-detection systems.<sup>7</sup> These models have shown tremendous promise, for example, a CNN trained on over 8000 colonoscopy images was able to identify polyps with an accuracy of 96.4%. This model was able to improve human endoscopists in detecting polyps, with expert reviewers able to identify over twice as many missed polyps with model assistance.<sup>35</sup>

Wang et al<sup>36</sup> were the first to deploy an ML-driven polyp-detection model to assist with live endoscopy. Patients were randomized to either standard colonoscopy or colonoscopy with a built-in CNN-based polyp-detection system that would alert the endoscopist when a polyp was detected. This system resulted in a 9% increase in adenoma-detection rate and a greater number of adenomas detected per patient.<sup>36</sup> Recently, multiple additional live ML assistant models have been developed which significantly increase adenoma-detection rate, with one model increasing detection rate by over 14% and the number of adenomas detected per procedure increasing from 0.71 to 1.07.<sup>37-39</sup>

Another area of ongoing ML research in colonoscopy is in the determination of ulcerative colitis severity. Severity of inflammation is a key factor in the risk of colorectal cancer development, so accurate detection of more severe disease is crucial.<sup>40</sup> Unfortunately, the determination of disease severity remains somewhat subjective, and consistency varies widely among endoscopists.<sup>41</sup> To address this problem, a CNN model was trained on 16,514 endoscopic images to distinguish moderate-to-severe ulcerative colitis from remission based on endoscopic images. This model was able to distinguish endoscopic remission from moderate-to-severe disease with an area under the receiver operating characteristic curve (AuROC) of 0.966, which is similar to human reviewers.<sup>42</sup> This model demonstrates the potential for computer assistance in detecting subtle changes in inflammation, potentially allowing for future standardization of endoscopic description of inflammation. Another CNN-based model trained on endoscopic images from 29 consecutive patients with ulcerative colitis (UC) was able to identify patients with UC flares requiring treatment

**Table.** Future Directions for ML in Computer Vision Endoscopy

Problem	ML solution
Classification of esophageal varices and risk stratification	Agreement among endoscopists varies widely regarding size of esophageal varices. Automated stratification and risk classification could significantly impact practice and provide endoscopists with a tool to more accurately define varices. <sup>62</sup>
Differentiation of ulcerative colitis vs Crohn's disease	Distinction between ulcerative colitis and Crohn's disease can be challenging endoscopically as well as histologically. <sup>63,64</sup> Numerous scoring tools have been created to assist endoscopists in diagnosis; however, this problem lends itself well to ML assistance in the future.
Assessment of biliary strictures during ERCP	The differentiation between benign and malignant biliary strictures during ERCP is challenging even for the most advanced endoscopists, and it is possible that ML could help identify features associated with benign and malignant disease that could assist the endoscopist. <sup>65</sup>
Predicting bile duct cannulation difficulty	One of the most challenging aspects of ERCP is determining the likelihood of success in achieving bile duct cannulation, and thus presents another opportunity for ML assistance to transform practice. <sup>65</sup>
Quality assessment of mucosal inspection	ML could not only assist with training of endoscopists but also provide real-time feedback regarding the percent of mucosa examined during EGD or colonoscopy to ensure adequate examination is taking place. <sup>66</sup>
Standardized training in endoscopy	ML-developed systems can likely help standardize the training and assessment of fellows in the development of endoscopy skills. <sup>10</sup>

EGD, esophagogastroduodenoscopy; ERCP, endoscopic retrograde cholangiopancreatography.

escalation.<sup>43</sup> Alternatively, a DNN was trained on over 40,000 endoscopic images and was able to identify patients with endoscopic remission with accuracy of 90.1%.<sup>44</sup>

### Wireless Capsule Endoscopy

The use of WCE in gastroenterology has been increasing since its clinical introduction in 2000 and has transformed visualization of the small bowel.<sup>45</sup> WCE is the recommended first-line investigation in patients with obscure GI bleeding but is also used to identify small bowel tumors, inflammatory bowel disease (IBD), celiac disease, and for surveillance of heritable polyposis syndromes.<sup>46</sup> During a single WCE, an average of 12,000 images are generated, which take at least 30 minutes on average for an experienced gastroenterologist to read and interpret. Thus, there has been significant research into the automation of this process through AI.

Despite the generation of a large number of images, WCE is currently only able to locate the source of bleeding in approximately 60% of cases.<sup>47</sup> An SVM model was developed in 2014 and trained on WCE images from 50 subjects with obscure GI bleeding and 200 subjects with normal mucosa. This model had a sensitivity of 93% and a specificity of 95% for identifying the source of bleeding.<sup>48</sup> Another study carried out a few years later improved on this model and was able to detect small bowel bleeding based on WCE images with sensitivity and specificity >99%.<sup>49</sup> Additional models were also created using CNNs that were able to identify mucosal erosions, ulcers, and small bowel angioectasia with high accuracy.<sup>50,51</sup> The success of these models led to a study evaluating the effectiveness of a CNN trained to identify small-bowel mucosal

breaks in clinical practice. This study reported that the time to read WCE reports was reduced by as much as 75% with no significant change in detection rate of mucosal breaks.<sup>52</sup>

In celiac disease, WCE has proven to be a useful alternative to upper endoscopy, with WCE demonstrating good sensitivity and excellent specificity for the detection of villous atrophy.<sup>53</sup> Since the adoption of WCE as an acceptable means of diagnosing celiac disease, a number of studies have evaluated the utility of incorporating ML to streamline and improve the detection of disease. The first notable study was published by Ciaccio et al<sup>54</sup> in 2010 and was able to detect the presence of celiac disease with 80% sensitivity and 96% specificity using still images from WCE. An SVM model trained on WCE images from 13 control and 13 celiac patients was able to identify celiac disease with a sensitivity of 88% and a specificity of 87%.<sup>55</sup> Another SVM model improved upon these results and was able to detect disease with sensitivity and specificity of 97% and 96%, respectively.<sup>56</sup>

WCE also plays an important role in IBD, in particular, Crohn's disease as it allows for assessment of the entire small bowel. WCE is currently used to identify disease, assist in determining disease severity, and help identify response to therapy. In 2020, a CNN was trained on over 17,000 WCE images from 49 patients and was found to be able to differentiate normal vs diseased mucosa in subjects with known Crohn's disease with an AUC of 0.99.<sup>57</sup> Another study attempted to decrease the time to read WCE reports to assess disease severity in IBD and trained an ML model to identify 100 images from the WCE video that are most likely to contain abnormalities. They had 2 sets of endoscopists review WCE reports and assess inflammatory activity; however, one group only had access to the ML-identified 100 images. There was strong agreement in the degree of

inflammation between the 2 reading groups; however, those only reviewing the 100-image set identified by the model were able to come to a conclusion up to 30 minutes quicker.<sup>58</sup> Another CNN-based algorithm was developed which demonstrated an accuracy of 91% in classifying ulcers in Crohn's disease.<sup>59</sup>

### The Future

In the coming years, ML-driven innovation in endoscopy is likely to impact practice in numerous ways. Endoscopists' diagnostic capabilities during procedures are likely to be enhanced through the use of this technology, with the computer able to assist in challenging diagnosis, such as differentiating cancer vs adenoma in the duodenum.<sup>60</sup> In the near future, computer vision will serve as an augmentation tool for endoscopists, improving accuracy and decreasing procedural time. These new systems will also likely serve as training tools for learners.<sup>10</sup> This future is rapidly approaching, with the FDA recently approving Breakthrough Device Designation for a real-time endoscopic AI system<sup>61</sup> (Table).

Significant changes will likely be seen in clinical practice over the next decade as a result of these innovations, and endoscopy is likely to be impacted more in the near term than any other area of gastroenterology. For example, endoscopists can expect computer vision systems to be implemented in the near future to assist in the diagnosis of early esophageal adenocarcinoma and polyp detection and classification. The reading of WCE images is also likely to be heavily impacted over the coming years, with ML algorithms able to significantly reduce reading time and improve accuracy. This will likely greatly expand the use of capsule endoscopy.

However, there is a vital need for clinical trials that can ascertain long-term benefits, risks, and the cost-effectiveness of increased early disease detection.<sup>67</sup> We must determine whether these models are generalizable across different populations and clinical settings. After high-quality clinical trials take place, it is not hard to imagine a future in which ML-driven endoscopy improves diagnostic capability of all endoscopists, reduces the burden on over-worked endoscopists, and also serves a vital role in the learning process for young endoscopists.

## AI and ML in GI Radiology/"Radiomics"

At the beginning of the ML revolution in medicine, the majority of the early research focused on the automation of radiology interpretation, including ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI). The potential applications of ML in radiology are broad, from decreasing radiologist burden to improving diagnostic accuracy. In gastroenterology, ML-assisted radiology has shown significant promise in hepatology and pancreatology (Figure 2).

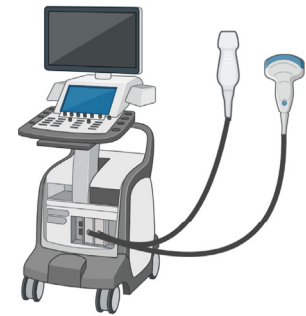
## Machine Learning in GI Radiology

### Hepatology

- Identification of liver disease
- Determine fibrosis stage
- Identify Liver Lesions
- Differentiate benign vs malignant lesions

### Pancreatology

- Identification of intraductal papillary mucinous neoplasms
- Identification and staging of pancreatic cancer
- Diagnosis of pancreatic ductal adenocarcinoma
- Quantify degree of intrapancreatic fat



**Figure 2.** Potential applications of artificial intelligence in GI radiology.

### Hepatology

ML has been applied extensively to radiology in the field of hepatology to identify and assess various conditions. Ahn et al<sup>68</sup> published a recent review on the use of ML in hepatology, specifically highlighting recent advancements in ML-assisted radiology.

Ultrasound is often the first test those with suspected liver disease undergo for diagnostic purposes; thus, much attention has been focused on automating the interpretation of hepatic ultrasound and improving diagnostic accuracy via ML. In 2016, Gatos et al<sup>69</sup> developed an SVM model that was able to distinguish subjects with liver disease from controls using ultrasound shear wave elastography with a sensitivity of 83.3%, specificity of 89.1%, and an accuracy of 87%. Another model was able to predict fibrosis stage in subjects with hepatitis B using real-time elastography with an accuracy of 83%.<sup>70</sup> A recent multicenter study was published by Wang et al<sup>71</sup> in 2019 in which an ML model to predict liver fibrosis using shear wave elastography 2D images was developed. This model was able to achieve an AUC of 0.97 for cirrhosis, 0.98 for advanced fibrosis, and 0.85 for identifying significant fibrosis.<sup>71</sup> The identification of liver lesions via ML-guided ultrasound has also been demonstrated, with a DL model developed to detect and characterize benign and focal liver lesions based on ultrasound images. This model had an AUC of 0.935 for identifying liver lesions and 0.916 for classifying these lesions as either benign or malignant.<sup>72</sup>

CT and MRI are also heavily used modalities in diagnosing and monitoring various liver diseases. Nayak et al<sup>73</sup> developed an radial basis function kernel algorithm based on CT images to identify cirrhosis and hepatocellular carcinoma (HCC). This model was able to accurately identify cirrhosis and HCC with an accuracy of 80%.<sup>73</sup> A model was also built to identify hepatic lesions, including HCC,

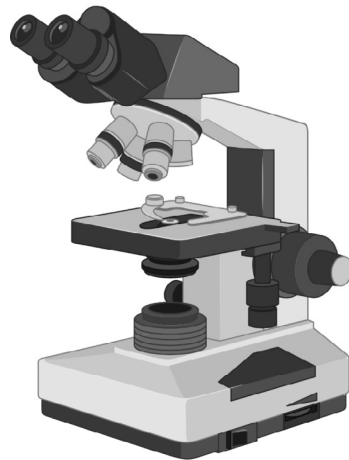
## Machine Learning in GI Pathology

### Luminal GI

- Diagnosis of Celiac Disease
- Classify Colorectal Polyps
- Diagnosis of Barrett's Esophagus

### Hepatology

- Diagnosis of NAFLD and NASH
- Identification of NAFLD
- Identification of specific histologic features such as macrosteatosis
- Determine degree of fibrosis
- Predict treatment response in NASH



### GI Malignancy

- Survival prediction in colorectal cancer
- Identification of microsatellite instability
- Identification of Esophageal Adenocarcinoma
- Differentiation between HCC and Cholangiocarcinoma
- Predict prognosis and survival in HCC

**Figure 3.** Potential applications of artificial intelligence in GI pathology. NAFLD, nonalcoholic fatty liver disease.

identified on multiphasic MRI and achieved an accuracy of 92%, a sensitivity of 92%, and specificity of 98%. Notably, computation time was 5.6 ms per lesion, highlighting the significant time that can be potentially saved by these models. This group further identified exact features that its model was utilizing to make decisions, improving interpretability of these predictions and potentially allowing for implementation into the radiologist workflow.<sup>74,75</sup> Multiple recent models have demonstrated that adipose tissue, muscle mass, and bone density can be determined by applying deep learning methods to CT images, ultimately allowing for clinicians to identify trends in patients' nutritional status that could signify future clinical decline. These models have been trialed in patients with liver disease and have been shown to accurately predict future mortality.<sup>76,77</sup>

### Pancreas

Diagnosis of diseases of the pancreas also relies heavily on imaging. Limited research in ML-based radiology in diseases of the pancreas has taken place; however, these few studies have shown promise for the future of this field. In 2019, Kuwahara et al<sup>78</sup> developed a DNN based on endoscopic ultrasound images to predict intraductal papillary mucinous neoplasms, a precursor lesion of pancreatic adenocarcinoma. This model had an accuracy, a sensitivity, and specificity of 92.6%, 94.0%, and 95.7%, respectively, for identifying intraductal papillary mucinous neoplasms, which was much better than the accuracy of 56% for radiologists.<sup>78</sup> Liu et al<sup>79</sup> trained a CNN based on CT images to identify pancreatic cancer. They utilized 370 CT scans from subjects with pancreatic cancer and 320 controls and found that the developed algorithm had a sensitivity of 98%, which was significantly higher than the 93% among experienced radiologists.<sup>79</sup> In 2021, Janssens et al<sup>80</sup> trained a CNN on 469 CT scans to identify intrapancreatic fat and

pancreatic ductal adenocarcinoma and were able to identify markers that could potentially be used to identify pancreatic adenocarcinoma in the prediagnostic phase.

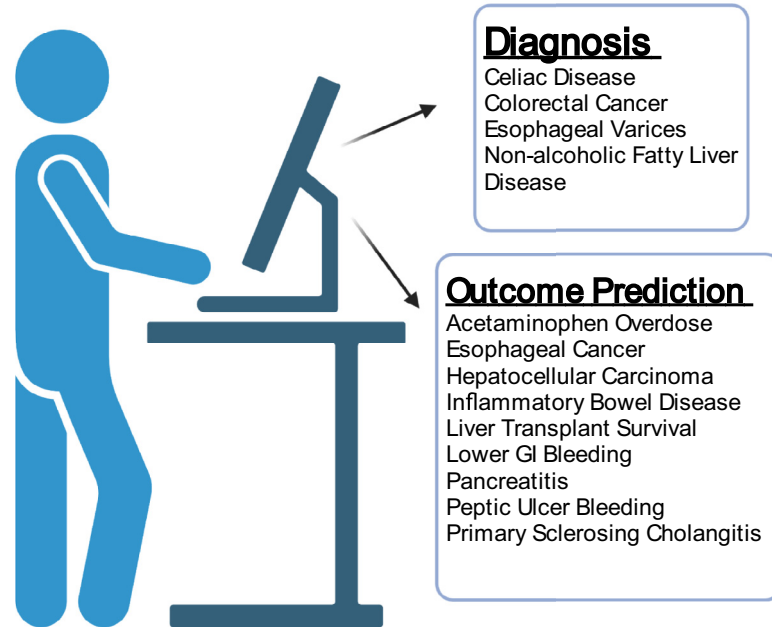
### The Future

In the near future, the application of ML algorithms in radiology will likely improve radiology workflow, improve diagnostic accuracy, and decrease time to diagnosis. For example, ML models that can not only analyze the current radiology image but also interpret this image in the context of all the previous imaging a certain patient has had and the interval changes that have occurred can be developed. Additionally, one could also envision a future in which ML-driven mobile-based ultrasound is available for patient use at home for tasks such as volume status evaluation, determination of degree of ascites, and bladder scans. ML-driven radiology is a crucial component of a future in which individualized medicine is at the forefront of gastroenterology.

### AI and ML in GI Pathology

While there has been significant focus on the application of ML in endoscopy, ML-driven histopathology has the potential to significantly impact clinical practice. Histopathology has a unique advantage over imaging modalities in that models can quantify precise cellular features such as shape, size, and color from hematoxylin- or immunohistochemistry-stained slides that can be used to create ML models.<sup>81</sup> Recent technological advances have made it possible to digitally archive traditional glass slides at full resolution, leading to the possibility of standardizing histopathological analysis through the use of ML.<sup>82,83</sup> These models have the potential to reduce the clinical burden on pathologists and improve diagnostic accuracy. Kobayashi et al<sup>81</sup> recently published a comprehensive review on the state of ML in GI pathology. We

## Machine Learning in GI Electronic Health Record Data



**Figure 4.** Potential applications of artificial intelligence via utilization of the electronic health record.

will review high-impact studies in this field that demonstrate the ways in which ML can transform clinical practice (Figure 3).

### Luminal GI

In 2019, Wei et al<sup>84</sup> developed a CNN with the goal of creating a deep learning system that can assist pathologists with the diagnosis of celiac disease. The CNN was trained on hematoxylin and eosin-stained duodenal tissue and was able to identify and differentiate between normal tissue and celiac disease with an accuracy of 91% and 95%, respectively.<sup>84</sup> A similar approach was taken by Martin et al<sup>85</sup> who trained a CNN on 300 hematoxylin-stained slides (100 normal, 100 *H. pylori*, 100 reactive gastropathy) and found the model could differentiate between the 3 with sensitivity and specificity greater than 70%. Neural networks have also shown promise in the classification of colorectal polyps. Wei et al<sup>86</sup> build a DNN that could classify colorectal polyps as either tubular adenoma, villous adenoma, hyperplastic polyp, and sessile serrated adenoma with an accuracy of 87%. The authors found this rate of correct classification to be equivalent to that of pathologist review and suggested that this model could be deployed to assist pathologists in improving efficiency and diagnostic accuracy in classifying colorectal polyps.<sup>86</sup> A recent publication by Gehring et al<sup>87</sup> found that pathologic diagnosis of BE via a nonendoscopic Cytosponge-TFF3 test could be feasible with an ML-driven process and that review of these slides by pathologists could be decreased by 57% with a similar diagnostic accuracy.

### GI Malignancy

In colorectal cancer, ML-driven histopathology has shown the potential to impact prognosis and determine appropriate treatment. For example, in 2019, a DNN was found to be able to predict survival in colorectal cancer based on only cancer histopathological images with an hazard ratio of 1.63 (95% CI 1.14–2.33).<sup>88</sup> Another study found that a DL model could successfully identify microsatellite instability and mismatch-repair deficiency in hematoxylin and eosin slides obtained from colorectal tumors with an AuROC of 0.92.<sup>89</sup> In the evaluation and screening of esophageal cancer, Tomita et al<sup>90</sup> created a CNN-based model that was able to differentiate esophageal tissue based on histopathological slides as either normal, non-dysplastic Barrett's, dysplastic Barrett's, and adenocarcinoma with an accuracy greater than 85%. Similar techniques have also been used for the detection of gastric cancer, with a CNN trained on whole-slide images from gastric biopsies able to identify gastric adenocarcinoma with an AUC up to 0.97.<sup>91</sup> A DL model has also been developed to identify and distinguish between the 2 primary types of liver cancer, hepatocellular carcinoma and cholangiocarcinoma. This model developed by Kiani et al<sup>92</sup> in 2020 was trained on hematoxylin and eosin-stained whole-slide images and was able to identify and distinguish between these 2 types of hepatic malignancies with an accuracy of 86%. Predicting prognosis in HCC has also been shown to be possible via pathology-based ML, with a DL model shown to have a c-statistic of greater than 0.75 in predicting prognosis after a histopathological diagnosis of

HCC, performing better than previous prediction models based on alpha-fetoprotein level, vascular invasion, or disease stage.<sup>93</sup>

## Hepatology

The current standard for diagnosis of nonalcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH) is manual histologic assessment by an expert pathologist, a time-consuming process that results in significant interpreter variability. Vanderbeck et al<sup>94</sup> created a model to assess histological features of NAFLD, with a goal of creating a scoring system that could provide continuous measurement of disease features. They created this model using an SVM that was able to identify the presence of NAFLD with 89% accuracy and was able to identify specific histopathological features such as macrosteatosis, bile ducts, portal veins, and sinusoids with high accuracy.<sup>94</sup> Models have also been developed to evaluate degree of fibrosis in NAFLD, such as an SVM-based model that had an ROC of greater than 90% for detecting normal fibrosis, bridging fibrosis, and presence of nodules or cirrhosis.<sup>95</sup> Another CNN model was built to characterize disease severity and treatment response in NASH and found a Cohen's kappa coefficient of 0.8 for staging NASH.<sup>96</sup>

## The Future

The future of ML in transforming the practice of GI pathology is incredibly promising, and we expect a number of changes to clinical practice in the coming years. For example, these models will lead to accurate and standardized grading of disease states such as esophageal dysplasia and NAFLD/NASH that significantly decrease intraobserver variability. We also expect more precise interpretations of histopathology, where previously specific components would be estimated in broad categories. With ML, we will be able to quantify these components with greater granularity, for example, in defining degree of fibrosis, percent dysplastic tissue in GI malignancy, degree of inflammation in eosinophilic esophagitis, and many more. Similar to endoscopy, these models will serve as a second set of eyes for pathologists in reducing cognitive burden and improving standardization, while also serving as an invaluable teaching tool.

## AI and ML in GI—EHR Data

Clinicians consider a vast variety of information when making a diagnosis, including clinical history, laboratory values, and vital signs, all of which are reviewed in the EHR. When many clinicians think of AI in medicine, they imagine a future in which the vast amounts of information included in the EHR can be utilized to identify particular diseases and predict outcomes. A significant amount of research has already taken place in making this future a reality, and we will highlight the advances made in HER-based ML

algorithms in the identification of disease and the prediction of outcomes (Figure 4).

## Disease Identification

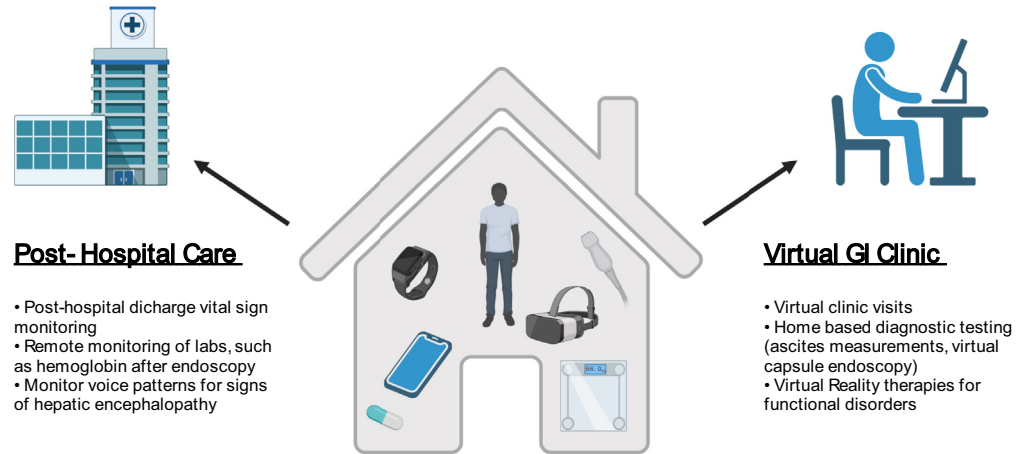
ML has shown significant promise in identifying disease utilizing data present in the EHR. For example, a clinical decision-support system was developed in 2011 to assist in the diagnosis of celiac disease using only patient-reported symptoms and physical exam findings extracted from the EHR, with a Bayesian classifier model having an AUC of 0.84 in the identification of celiac disease.<sup>97</sup> A random forest (RF) model was created in 2016 that was able to diagnose colorectal cancer prior to clinical diagnosis with an AUC of 0.891 using only raw data included in the EHR.<sup>98</sup> Another disease-identification model was developed by Wu et al in 2019 that attempted to identify NAFLD utilizing only clinical data. The authors created an RF model and included variables such as age, gender, vital signs, and labs and were able to identify NAFLD with an AUC of 0.925.<sup>99</sup> Another RF-based model was created to identify patients with esophageal varices and those with esophageal varices requiring treatment. Screening for esophageal varices requires endoscopy, and the authors hypothesized that a noninvasive strategy could be appropriate in screening for varices. The model was developed using clinical data available in the EHR including age, gender, vital signs, labs, and reported complications of cirrhosis and was found to have an AUC of 0.82 for identifying esophageal varices and 0.74 for identifying varices requiring treatment.<sup>100</sup>

## Outcome Prediction

The potential of ML-driven outcome prediction in gastroenterology was first demonstrated in 2003 when Das et al<sup>101</sup> published in the *Lancet* a model for predicting outcomes following lower GI bleeding. They built an artificial neural network (ANN) using clinical variables available in the EHR at initial presentation of lower GI bleed and were able to predict mortality, recurrent bleeding, and need for endoscopic intervention with accuracy over 90%.<sup>101</sup> Another ML model utilized 6 parameters extracted from the EHR (age, baseline hemoglobin, presence of gastric ulcer, gastrointestinal diseases, malignancies, and infections) and was able to predict recurrent peptic ulcer bleeding in 1 year with an AuROC of 0.775% and with 84% accuracy.<sup>102</sup> Sato et al<sup>103</sup> developed an ANN-based model to predict survival in patients diagnosed with esophageal cancer who underwent endoscopic resection with curative intent and were able to predict 5-year survival with an AUC of 0.884. Another ANN model was able to predict progression to severe pancreatitis, organ failure, and mortality in those hospitalized with pancreatitis with an AUC greater than 0.8.<sup>104</sup> A novel risk prediction model was created by Eaton et al<sup>105</sup> in 2020 to predict outcomes in those with primary sclerosing cholangitis. This model utilized 9 clinical variables (bilirubin, albumin, alkaline phosphatase, platelets,



## The GI Practice of the Future



**Figure 5.** Future impact of artificial intelligence on the practice of gastroenterology and hepatology.

aspartate aminotransferase, hemoglobin, sodium, patient age, and number of years since primary sclerosing cholangitis was diagnosed) to build a gradient boosted model that was able to predict hepatic decompensation with c-statistic of 0.85.<sup>105</sup>

Waljee et al<sup>106</sup> have published multiple ML studies predicting outcomes in IBD. In 2017, this group developed an RF-based model that was able to predict future IBD-related hospitalization with an AuROC of 0.87. The main independent risk factors identified in this model included age, albumin concentration, immunosuppressive medication use, and platelet count.<sup>106</sup> They next used an RF algorithm to predict response to vedolizumab among those with UC. AuROC for this algorithm was 0.73 when including data through the first 6 weeks of therapy, potentially giving clinicians further insight into the individualized utility of this expensive medication.<sup>107</sup> Finally, this group again employed an RF model to predict response to ustekinumab in patients with Crohn's disease and found their model had an AuROC of 0.76.<sup>108</sup>

Significant work has also been carried out in predicting outcomes in a variety of liver diseases. For example, one study following up patients with early-stage cirrhosis until liver transplant or death developed an RF model to predict development of HCC. This model was created utilizing EHR data and was able to predict HCC with 80.7% sensitivity and 46.8% specificity.<sup>109</sup> A Classification and Regression Tree model was developed to improve prognostication in acetaminophen overdose that incorporated demographics, labs, medications administered, and other clinical variables extracted from the EHR and was able to predict mortality with an AUC of 0.79.<sup>110</sup> Predicting survival after liver transplant was also shown to be possible, with a recurrent neural network developed that was able to predict 5-year survival after liver transplantation with an AUC of 0.864.<sup>111</sup>

In 2017, Konerman et al<sup>112</sup> developed an RF model that utilized clinical and laboratory variables extracted from the EHR to predict transplant-free survival in hepatitis C-related cirrhosis with an AuROC of 0.85. Multiple studies have developed models to predict hospital readmission among those with cirrhosis, a source of significant morbidity and mortality. However, the most promising model was built using a traditional LR model that predicted readmission with an AUC of 0.67.<sup>113</sup> Another model utilized multiple ML models to predict readmission and mortality in those hospitalized with cirrhosis; the best model was an SVM that achieved an AUC of 0.62.<sup>114</sup> The limited predicted capabilities of these models are likely due to the difficulties with considering multiple data points in relation to each other over time when building a model. However, if just attempting to predict mortality in those with cirrhosis, multiple models have been developed using deep learning techniques and have been able to predict 1-year mortality with greater accuracy than the model for end-stage liver disease-sodium score.<sup>115,116</sup>

### The Future

The overall goal when initially creating the EHR was to improve health-care quality. As described, ML is showing significant promise in utilizing these vast data to individualize medicine. However, it is still very challenging to determine if guideline-driven quality health care is being provided, especially considering situations when there are reasonable explanations for deviations in standard of care. Utilizing only structured data in the EHR, it is likely impossible to achieve these goals. However, it is evident that these data are present as it can be determined on individual chart review. But how do we evaluate this data on large scales? The answer may be in the use of natural language processing (NLP) to create models based on unstructured data.<sup>117</sup>

## AI and ML in NLP

The current frontier in ML research is NLP. The majority of the previously discussed ML advances in gastroenterology involve structured data, or data that can easily be stored and organized into tables. However, significant amounts of clinical data for individual patients exist in an unstructured format, in places such as clinical notes and pathology and radiology reports. NLP aims to automate the extraction of this unstructured clinical data so that it can be used to develop clinical models.

Studies involving NLP in gastroenterology have primarily been published in the last 5 years, beginning with a study in 2016 that utilized colonoscopy reports to determine quality of colonoscopy and adherence to proper surveillance intervals. This NLP-based algorithm was able to report adenoma-detection rate with 100% accuracy and determine proper colonoscopy surveillance intervals across 13 medical centers.<sup>118,119</sup> Another NLP-based study attempted to improve upon the identification and risk stratification of patients with cirrhosis. Previous International Classification of Diseases (ICD) code-based studies to identify cirrhosis performed well, with a positive predictive value of 90% and an negative predictive value of 87%.<sup>120</sup> However, an NLP model based on radiology reports was able to identify cirrhosis with a positive predictive value of 92% and an negative predictive value of 97%.<sup>121</sup> Similar models have been successfully built using NLP in Crohn's disease, where an NLP-based model was shown to improve retrospective disease identification from 83% to 95%.<sup>122,123</sup> NLP has also been used to identify cases of hepatorenal syndrome, again performing better than ICD-based identification models with an AUC of 0.93.<sup>124</sup> The ability of NLP in retrospective identification of disease will greatly increase the ease and feasibility of conducting large population-based cohort studies.

Another type of unstructured data that is becoming increasingly prevalent in medicine is patient messages through the EHR. The possibilities for this type of data have already started to be explored with NLP. For example, an NLP-based algorithm has shown promise in detecting early hepatic encephalopathy based on sentence and word length in patient messages.<sup>125</sup>

NLP also has the potential to greatly reduce physician administrative burden through the automatic generation of ICD codes based on clinical notes, automated prior authorization requests, and the ability to automatically generate sufficient documentation for billing based on unstructured notes.<sup>126</sup> This has the potential to allow clinicians to return to a system where clinical notes are simple, readable, and oriented toward providing important information to patient care as opposed to a document designed to maximize billing potential. NLP also has the potential to serve as a valuable clinician assistance tool, automatically extracting clinical information from the chart and presenting applicable scoring tools and clinical evidence for treatment. This again highlights the potential for NLP to ease physician burden while increasing evidence driven care.<sup>126</sup>

Preliminary studies have even shown promise in developing automated chat-bots that can collect patient symptoms via messages in the EHR and provide simple medical advice or recognize when higher level of care is needed.<sup>127</sup>

## Conclusion—Where Do We Go From Here?

The application of AI and ML in gastroenterology is not new; however, advancements are occurring rapidly, and the future landscape of AI-driven gastroenterology is beginning to come into focus. Before this future is fully realized, a few key developments that are on the horizon need to take place. First, the medical community needs to fully recognize that AI-driven algorithms can and should be treated as a medical device, and thus regulated by the FDA.<sup>128</sup> Naturally, there remains some hesitancy in integrating ML-based solutions into clinical practice, and increased awareness and understanding of these methods by the gastroenterology community is a crucial next step. Next, the medical community must determine who is responsible for the accuracy of AI-based algorithms and who holds responsible when errors inevitably result as a consequence of these models.<sup>5</sup> There is also an urgent need for randomized trials and extensive external validation of developed algorithms in diverse populations to mitigate inherent biases. Although many of the above-described trials exhibit outstanding performance, it is expected that the real-world performance of these models will decrease considerably. Finally, the pathway to developing AI solutions needs to be simplified so that physicians without computer engineering skills can develop models and deploy them into clinical practice.<sup>129</sup> The practice of gastroenterology and hepatology will significantly change over the coming years as advances in AI and ML continue to accelerate (Figure 5). Thus, it is important that the current state of this technology and the future possibilities are understood so that they can be incorporated into future practice.

## References

1. Turing AM. Computing machinery and intelligence. *Comput Mach Intell Mind* 1950;49:433–460.
2. Tran BX, Vu GT, Ha GH, et al. Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. *J Clin Med* 2019;8:360.
3. Topol EJ. *Deep medicine: how artificial intelligence can make healthcare human again*. Basic Books, 2019:378.
4. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. *World J Gastroenterol* 2019; 25:1666–1683.
5. Stidham RW. Artificial intelligence for understanding imaging, text, and data in gastroenterology. *Gastroenterol Hepatol (N Y)* 2020;16:341.
6. Rattan P, Penrice DD, Simonetto DA. Artificial intelligence and machine learning: what you always wanted to

- know but were afraid to ask. *Gastro Hep Adv* 2021; 1:P70–P78.
7. Le Berre C, Sandborn WJ, Aridhi S, et al. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020;158:76–94.e2.
  8. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:1–9.
  9. Shaw J, Rudzicz F, Jamieson T, et al. Artificial intelligence and the implementation challenge. *J Med Internet Res* 2019;21:e13659.
  10. Mori Y, Kudo S, Mohamed HEN, et al. Artificial intelligence and upper gastrointestinal endoscopy: current status and future perspective. *Dig Endosc* 2019;31:378–388.
  11. Mascarenhas M, Afonso J, Andrade P, et al. Artificial intelligence and capsule endoscopy: unravelling the future. *Ann Gastroenterol* 2021;34:300.
  12. Coleman HG, Xie S-H, Lagergren J. The epidemiology of esophageal adenocarcinoma. *Gastroenterology* 2018; 154:390–405.
  13. Bhat SK, McManus DT, Coleman HG, et al. Oesophageal adenocarcinoma and prior diagnosis of Barrett's oesophagus: a population-based study. *Gut* 2015; 64:20–25.
  14. Bell MG, Iyer PG. Innovations in screening tools for Barrett's esophagus and esophageal adenocarcinoma. *Curr Gastroenterol Rep* 2021;23:1–7.
  15. van der Sommen F, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy* 2016;48:617–624.
  16. Swager A-F, van der Sommen F, Klomp SR, et al. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017;86:839–846.
  17. de Groof AJ, Struyvenberg MR, Fockens KN, et al. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). *Gastrointest Endosc* 2020; 91:1242–1250.
  18. de Groof AJ, Struyvenberg MR, van der Putten J, et al. 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.
  19. Hashimoto R, Requa J, Dao T, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointest Endosc* 2020; 91:1264–1271.e1.
  20. Ebigbo A, Mendel R, Probst A, et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut* 2020;69:615.
  21. Cai S-L, Li B, Tan W-M, 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.
  22. Guo L, Xiao X, Wu C, et al. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest Endosc* 2020;91:41–51.
  23. Watanabe K, Nagata N, Shimbo T, et al. Accuracy of endoscopic diagnosis of *Helicobacter pylori* infection according to level of endoscopic experience and the effect of training. *BMC Gastroenterol* 2013;13:1–7.
  24. Huang C-R, Sheu B-S, Chung P-C, et al. Computerized diagnosis of *Helicobacter pylori* infection and associated gastric inflammation from endoscopic images by refined feature selection using a neural network. *Endoscopy* 2004;36:601–608.
  25. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of *Helicobacter pylori* infection based on endoscopic images. *EBioMedicine* 2017;25:106–111.
  26. Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis. *Endosc Int Open* 2014;2:E46.
  27. Kubota K, Kuroda J, Yoshida M, et al. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. *Surg Endosc* 2011; 26:1485–1489.
  28. Zhu Y, Wang Q-C, Xu M-D, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc* 2019; 89:806–815.e1.
  29. 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.
  30. Belsey J, Epstein O, Heresbach D. Systematic review: oral bowel preparation for colonoscopy. *Aliment Pharmacol Ther* 2007;25:373–384.
  31. Lai EJ, Calderwood AH, Doros G, et al. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc* 2009;69:620–625.
  32. Zhou J, Wu L, Wan X, et al. A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc* 2020;91:428–435.e2.
  33. Zhao S, Wang S, Pan P, et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. *Gastroenterology* 2019;156:1661–1674.e11.
  34. Than M, Witherspoon J, Shami J, et al. Diagnostic miss rate for colorectal cancer: an audit. *Ann Gastroenterol* 2015;28:94–98.
  35. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018;155:1069.
  36. Wang P, Berzin TM, Brown JRG, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019;68:1813–1819.
  37. Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020;5:352–361.
  38. Su J-R, Li Z, Shao X-J, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized

- controlled study (with videos). *Gastrointest Endosc* 2020; 91:415–424.e4.
39. Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* 2020;159:512–520.e7.
  40. Rutter M, Saunders B, Wilkinson K, et al. Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology* 2004;126:451–459.
  41. Osada T, Ohkusa T, Yokoyama T, et al. Comparison of several activity indices for the evaluation of endoscopic activity in UC: inter- and intraobserver consistency. *Inflamm Bowel Dis* 2010;16:192–197.
  42. 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.
  43. Bossuyt P, Nakase H, Vermeire S, et al. Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density. *Gut* 2020; 69:1778–1786.
  44. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology* 2020;158:2150–2157.
  45. Soffer S, Klang E, Shimon O, et al. Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. *Gastrointest Endosc* 2020;92:831–839.e8.
  46. Rondonotti E, Spada C, Adler S, et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) technical review. *Endoscopy* 2018;50:423–446.
  47. Brunk T, Schmidt A, Hochberger J, et al. Telemetric capsule-based upper gastrointestinal tract – blood detection – first multicentric experience. *Minim Invasive Ther Allied Technol* 2021;1–8.
  48. Ghosh T, Fattah SA, Shahnaz C, et al. An automatic bleeding detection scheme in wireless capsule endoscopy based on histogram of an RGB-indexed image. *Annu Int Conf IEEE Eng Med Biol Soc* 2014; 2014:4683–4686.
  49. Jia X, Meng MQH. Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and CNN features. *Annu Int Conf IEEE Eng Med Biol Soc* 2017;2017:3154–3157.
  50. Fan S, Xu L, Fan Y, et al. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Phys Med Biol* 2018;63:165001.
  51. Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; 89:189–194.
  52. Aoki T, Yamada A, Aoyama K, et al. Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. *Dig Endosc* 2020;32:585–591.
  53. Rondonotti E, Spada C, Cave D, et al. Video capsule enteroscopy in the diagnosis of celiac disease: a multicenter study. *Am J Gastroenterol* 2007; 102:1624–1631.
  54. Ciaccio EJ, Tennyson CA, Bhagat G, et al. Classification of videocapsule endoscopy image patterns: comparative analysis between patients with celiac disease and normal individuals. *Biomed Eng Online* 2010;9:44.
  55. Koh JEW, Hagiwara Y, Oh SL, et al. Automated diagnosis of celiac disease using DWT and nonlinear features with video capsule endoscopy images. *Futur Gener Comput Syst* 2019;90:86–93.
  56. Wang X, Qian H, Ciaccio EJ, et al. Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction. *Comput Methods Programs Biomed* 2020;187:105236.
  57. Klang E, Barash Y, Margalit RY, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020;91:606–613.e2.
  58. Freitas M, Arieira C, Carvalho PB, et al. Simplify to improve in capsule endoscopy – TOP 100 is a swift and reliable evaluation tool for the small bowel inflammatory activity in Crohn's disease. *Scand J Gastroenterol* 2020; 55:408–413.
  59. Barash Y, Azaria L, Soffer S, et al. Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. *Gastrointest Endosc* 2021;93:187–192.
  60. Kikuchi D, Hoteya S, Iizuka T, et al. Diagnostic algorithm of magnifying endoscopy with narrow band imaging for superficial non-ampullary duodenal epithelial tumors. *Dig Endosc* 2014;26:16–22.
  61. AI Medical Service Inc. announces FDA breakthrough device designation for endoscopic AI system. Accessed October 27, 2021. <https://www.prnewswire.com/news-releases/ai-medical-service-inc-announces-fda-breakthrough-device-designation-for-endoscopic-ai-system-300953301.html>.
  62. Perri RE, Chiorean MV, Fidler JL, et al. A prospective evaluation of computerized tomographic (CT) scanning as a screening modality for esophageal varices. *Hepatology* 2008;47:1587–1594.
  63. Gottlieb K, Daperno M, Usiskin K, et al. Endoscopy and central reading in inflammatory bowel disease clinical trials: achievements, challenges and future developments. *Gut* 2021;70:418–426.
  64. Leoncini G, Donato F, Reggiani-Bonetti L, et al. Diagnostic interobserver variability in Crohn's disease- and ulcerative colitis-associated dysplasia: a multicenter digital survey from the IG-IBD, Pathologists Group. *Tech Coloproctol* 2020;25:101–108.
  65. Ahmad OF, Stassen P, Webster GJ. Artificial intelligence in biliopancreatic endoscopy: is there any role? *Best Pract Res Clin Gastroenterol* 2021; 52–53:101724.
  66. Ahmad OF, Soares AS, Mazomenos E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. *Lancet Gastroenterol Hepatol* 2019;4:71–80.

67. Mori Y, Bretthauer M, Kalager M. Hopes and hypes for artificial intelligence in colorectal cancer screening. *Gastroenterology* 2021;161:774–777.
68. Ahn JC, Connell A, Simonetto DA, et al. Application of artificial intelligence for the diagnosis and treatment of liver diseases. *Hepatology* 2021;73:2546–2563.
69. Gatos I, Tsantis S, Spiliopoulos S, et al. A new computer aided diagnosis system for evaluation of chronic liver disease with ultrasound shear wave elastography imaging. *Med Phys* 2016;43:1428–1436.
70. Chen Y, Luo Y, Huang W, et al. Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patients with chronic hepatitis B. *Comput Biol Med* 2017;89:18–23.
71. Wang K, Lu X, Zhou H, et al. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut* 2019;68:729–741.
72. Schmauch B, Herent P, Jehanno P, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. *Diagn Interv Imaging* 2019;100:227–233.
73. Nayak A, Baidya Kayal E, Arya M, et al. Computer-aided diagnosis of cirrhosis and hepatocellular carcinoma using multi-phase abdomen CT. *Int J Comput Assist Radiol Surg* 2019;14:1341–1352.
74. Hamm CA, Wang CJ, Savic LJ, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol* 2019;29:3338–3347.
75. Wang CJ, Hamm CA, Savic LJ, et al. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. *Eur Radiol* 2019;29:3348–3357.
76. Zou WY, Enchakalody BE, Zhang P, et al. Automated measurements of body composition in abdominal CT scans using artificial intelligence can predict mortality in patients with cirrhosis. *Hepatol Commun* 2021;5:1901.
77. Wang NC, Zhang P, Tapper EB, et al. Automated measurements of muscle mass using deep learning can predict clinical outcomes in patients with liver disease. *Am J Gastroenterol* 2020;115:1210.
78. Kuwahara T, Hara K, Mizuno N, et al. Usefulness of deep learning analysis for the diagnosis of malignancy in intraductal papillary mucinous neoplasms of the pancreas. *Clin Transl Gastroenterol* 2019;10:1–8.
79. Liu K-L, Wu T, Chen P-T, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. *Lancet Digit Health* 2020;2:e303–e313.
80. Janssens LP, Weston AD, Singh D, et al. Determining age and sex-specific distribution of pancreatic whole-gland CT attenuation using artificial intelligence aided image segmentation: associations with body composition and pancreatic cancer risk. *Pancreatol* 2021;21:1524–1530.
81. Kobayashi S, Saltz JH, Yang VW. State of machine and deep learning in histopathological applications in digestive diseases. *World J Gastroenterol* 2021;27:2545.
82. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology* 2017;70:134–145.
83. Kather JN, Calderaro J. Development of AI-based pathology biomarkers in gastrointestinal and liver cancer. *Nat Rev Gastroenterol Hepatol* 2020;17:591–592.
84. Wei JW, Wei JW, Jackson CR, et al. Automated detection of celiac disease on duodenal biopsy slides: a deep learning approach. *J Pathol Inform* 2019;10:7.
85. Martin DR, Hanson JA, Gullapalli RR, et al. A deep learning convolutional neural network can recognize common patterns of injury in gastric pathology. *Arch Pathol Lab Med* 2020;144:370–378.
86. Wei JW, Suriawinata AA, Vaickus LJ, et al. Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides. *JAMA Netw Open* 2020;3:e203398.
87. Gehrung M, Crispin-Ortiz M, Berman AG, et al. Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. *Nat Med* 2021;27:833–841.
88. Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med* 2019;16:e1002730.
89. Echle A, Grabsch HI, Quirke P, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. *Gastroenterology* 2020;159:1406–1416.e11.
90. Tomita N, Abdollahi B, Wei J, et al. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. *JAMA Netw Open* 2019;2:e1914645.
91. Iizuka O, Kanavati F, Kato K, et al. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci Rep* 2020;10:1504.
92. Kiani A, Uyumazturk B, Rajpurkar P, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *NPJ Digit Med* 2020;3:1–8.
93. Saillard C, Schmauch B, Laifa O, et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. *Hepatology* 2020;72:2000–2013.
94. Vanderbeck S, Bockhorst J, Komorowski R, et al. Automatic classification of white regions in liver biopsies by supervised machine learning. *Hum Pathol* 2014;45:785–792.
95. Gawrieh S, Sethunath D, Cummings OW, et al. Automated quantification and architectural pattern detection of hepatic fibrosis in NAFLD. *Ann Diagn Pathol* 2020;47:151518.
96. Taylor-Weiner A, Pokkalla H, Han L, et al. A machine learning approach enables quantitative measurement of liver histology and disease monitoring in NASH. *Hepatology* 2021;74:133–147.
97. Tenório JM, Hummel AD, Cohrs FM, et al. Artificial intelligence techniques applied to the development of a

- decision-support system for diagnosing celiac disease. *Int J Med Inform* 2011;80:793–802.
98. Kop R, Hoogendoorn M, Ten Teije A, et al. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. *Comput Biol Med* 2016;76:30–38.
99. Wu CC, Yeh WC, Hsu WD, et al. Prediction of fatty liver disease using machine learning algorithms. *Comput Methods Programs Biomed* 2019;170:23–29.
100. Dong TS, Kalani A, Aby ES, et al. Machine learning-based development and validation of a scoring system for screening high-risk esophageal varices. *Clin Gastroenterol Hepatol* 2019;17:1894–1901.e1.
101. Das A, Ben-Menachem T, Cooper GS, et al. Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model. *Lancet* 2003;362:1261–1266.
102. Wong GL-H, Ma AJ, Deng H, et al. Machine learning model to predict recurrent ulcer bleeding in patients with history of idiopathic gastroduodenal ulcer bleeding. *Aliment Pharmacol Ther* 2019;49:912–918.
103. Sato F, Shimada Y, Selaru FM, et al. Prediction of survival in patients with esophageal carcinoma using artificial neural networks. *Cancer* 2005;103:1596–1605.
104. Mofidi R, Duff MD, Madhavan KK, et al. Identification of severe acute pancreatitis using an artificial neural network. *Surgery* 2007;141:59–66.
105. 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.
106. 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.
107. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther* 2018;47:763.
108. 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.
109. Singal AG, Mukherjee A, Joseph Elmunzer B, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol* 2013;108:1723–1730.
110. Speiser JL, Lee WM, Karvellas CJ, et al. Predicting outcome on admission and post-admission for acetaminophen-induced acute liver failure using classification and regression tree models. *PLoS One* 2015;10:e0122929.
111. Khosravi B, Pourahmad S, Bahreini A, et al. Five years survival of patients after liver transplantation and its effective factors by neural network and Cox proportional hazard regression models. *Hepat Mon* 2015;15:25164.
112. 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.
113. Koola JD, Ho SB, Cao A, et al. Predicting 30-day hospital readmission risk in a national cohort of patients with cirrhosis. *Dig Dis Sci* 2020;65:1003–1031.
114. 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.
115. 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.
116. Cucchetti A, Vivarelli M, Heaton ND, et al. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. *Gut* 2007;56:253.
117. Jha AK. The promise of electronic records: around the corner or down the road? *JAMA* 2011;306:880–881.
118. Imler TD, Morea J, Kahi C, et al. Multi-center colonoscopy quality measurement utilizing natural language processing. *Am J Gastroenterol* 2015;110:543–552.
119. Naylor J, Borges LF, Goryachev S, et al. Natural language processing accurately calculates adenoma and sessile serrated polyp detection rates. *Dig Dis Sci* 2018;63:1794–1800.
120. Kramer JR, Davila JA, Miller ED, et al. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs Administrative Databases. *Aliment Pharmacol Ther* 2008;27:274–282.
121. Chang EK, Yu CY, Clarke R, et al. Defining a patient population with cirrhosis. *J Clin Gastroenterol* 2016;50:889–894.
122. Hou J, Tan M, Stidham R, et al. Accuracy of diagnostic codes for identifying patients with ulcerative colitis and Crohn's disease in the Veterans Affairs Health Care System. *Dig Dis Sci* 2014;59:2406.
123. Ananthakrishnan AN, Cai T, Savova G, et al. Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis* 2013;19:1411.
124. Koola JD, Davis SE, Al-Nimri O, et al. Development of an automated phenotyping algorithm for hepatorenal syndrome. *J Biomed Inform* 2018;80:87–95.
125. Dickerson LK, Rouhizadeh M, Korotkaya Y, et al. Language impairment in adults with end-stage liver disease: application of natural language processing towards patient-generated health records. *NPJ Digit Med* 2019;2:1–7.
126. Nehme F, Feldman K. Evolving role and future directions of natural language processing in gastroenterology. *Dig Dis Sci* 2020;66:29–40.
127. Safi Z, Abd-Alrazaq A, Khalifa M, et al. Technical aspects of developing chatbots for medical applications: scoping review. *J Med Internet Res* 2020;22:e19127.
128. Software as a medical device (SAMd): clinical evaluation - guidance for industry and Food and drug administration staff. 2017. Accessed July 28, 2021. <https://www.fda.gov/MedicalDevices/InternationalPrograms/IMDRF/default.htm>.
129. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by

health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health* 2019;1:e232–e242.

---

Received December 13, 2021. Accepted February 22, 2022.

**Correspondence:**

Address correspondence to: Douglas A. Simonetto, MD, Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905. e-mail: [Simonetto.Douglas@mayo.edu](mailto:Simonetto.Douglas@mayo.edu).

**Acknowledgments:**

Figures created with [BioRender.com](https://BioRender.com).

**Authors' Contributions:**

Daniel D. Penrice contributed to manuscript formulation, writing, and revision. Douglas A. Simonetto contributed to manuscript design and critical revision.

**Conflicts of Interest:**

The authors disclose no conflicts.

**Funding:**

Dr Simonetto's research is funded by NIAAA U01AA026886-03 and NIH 1U01DK130181.

**Ethical Statement:**

The study did not require the approval of an institutional review board.