Artificial Intelligence in Gastrointestinal Endoscopy

Artif Intell Gastrointest Endosc 2022 June 28; 3(3): 16-43
MINIREVIEWS

16  Artificial intelligence in endoscopy: More than what meets the eye in screening colonoscopy and endosonographic evaluation of pancreatic lesions
    *Rao B H, Trieu JA, Nair P, Gressel G, Venu M, Venu RP*

31  Artificial intelligence and machine learning in colorectal cancer
    *Awidi M, Bagga A*
ABOUT COVER
Editorial Board Member of Artificial Intelligence in Gastrointestinal Endoscopy, Mohamed Amine Elghali, PhD, Professor, Surgeon, Department of General Surgery, Farhat Hached University Hospital, Sousse 4000, Tunisia. amine.elghali@gmail.com

AIMS AND SCOPE
The primary aim of Artificial Intelligence in Gastrointestinal Endoscopy (AIGE, Artif Intell Gastrointest Endosc) is to provide scholars and readers from various fields of artificial intelligence in gastrointestinal endoscopy with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

AIGE mainly publishes articles reporting research results obtained in the field of artificial intelligence in gastrointestinal endoscopy and covering a wide range of topics, including artificial intelligence in capsule endoscopy, colonoscopy, double-balloon enteroscopy, duodenoscopy, endoscopic retrograde cholangiopancreatography, endosonography, esophagoscopy, gastrointestinal endoscopy, gastroscopy, laparoscopy, natural orifice endoscopic surgery, proctoscopy, and sigmoidoscopy.

INDEXING/ABSTRACTING
The AIGE is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

RESPONSIBLE EDITORS FOR THIS ISSUE
Production Editor: Yi-Xuan Cai; Production Department Director: Xiang Lu; Editorial Office Director: Jin-Lei Wang.
Artificial intelligence in endoscopy: More than what meets the eye in screening colonoscopy and endosonographic evaluation of pancreatic lesions

Harshavardhan Rao B, Judy A Trieu, Priya Nair, Gilad Gressel, Mukund Venu, Rama P Venu

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Abstract

Artificial intelligence (AI)-based tools have ushered in a new era of innovation in the field of gastrointestinal (GI) endoscopy. Despite vast improvements in endoscopic techniques and equipment, diagnostic endoscopy remains heavily operator-dependent, in particular, colonoscopy and endoscopic ultrasound (EUS). Recent reports have shown that as much as 25% of colonic adenomas may be missed at colonoscopy. This can result in an increased incidence of interval colon cancer. Similarly, EUS has been shown to have high inter-observer variability, overlap in diagnoses with a relatively low specificity for pancreatic lesions. Our understanding of Machine-learning (ML) techniques in AI have evolved over the last decade and its application in AI-based tools for endoscopic detection and diagnosis is being actively investigated at several centers. ML is an aspect of AI that is based on neural networks, and is widely used for image classification, object detection, and semantic segmentation which are key functional aspects of AI-related computer aided diagnostic systems. In this review, current status and limitations of ML, specifically for adenoma detection and endosonographic diagnosis of pancreatic lesions, will be summarized from existing literature. This will help to better understand its role as viewed through the prism of real world application in the field of GI endoscopy.

Key Words: Artificial intelligence; Artificial; Machine; Colonoscopy; Polyp; Endosonography; Pancreas
The influence of artificial intelligence (AI) based applications in our everyday practice as endoscopists has been steadily increasing. One of the areas where it has shown promise is in image discrimination and diagnosis, which has many applications in endoscopy. The increasing application and rapid advancement of technology in this area necessitates an understanding of the basics and scope of AI in gastroenterology. In this review, a brief technical basis of AI in image discrimination has been described, followed by an update on the role of AI in the prevention of colorectal cancer and the evaluation of specific pancreatic lesions using endoscopic ultrasound.

INTRODUCTION
Artificial intelligence (AI) has directly impacted the field of endoscopy by nurturing questions directed at the status quo and eventually opened up new paradigms that redefined the boundaries of our abilities as an endoscopist. AI is a broad term that encompasses the development and application of algorithms that can perform tasks that generally necessitate human intelligence[1]. Machine learning (ML), on the other hand, is a subset of AI which refers to a specific algorithm, capable of analyzing features in a dataset, based on raw data, in order to deliver a classification output[2,3]. One of the areas where ML has shown a lot of promise is in image discrimination and diagnosis, which has many applications in the field of gastro-intestinal (GI) endoscopy. The advent of advanced imaging techniques such as high-definition white light endoscopy (HD-WLE) and pre-processing techniques like optical chromo-endoscopy, have paved the way for AI to make a significant impact in diagnostic endoscopy. Currently, AI in GI endoscopy is witnessing a paradigm shift, from mere ‘identification’ to a more composite and clinically relevant ‘interpretation’ of the images[4]. This paradigm shift, in combination with rapid improvement in computing power, has enabled ML algorithms to occupy a central role in the world of endoscopy.

Machine learning has already demonstrated remarkable success in several areas of medicine, such as radiology and pathology[5-9]. More importantly, there has been a deluge of published literature on the utility and potential of ML within the domain of endoscopy in the past decade[10-18]. Deep learning has strengthened the reality that the use of ML in endoscopy is an eventuality that is here to stay[19]. However, we are still in the early stages of understanding its full potential in image differentiation and classification of endoscopic lesions, with many unanswered questions leading to poor acceptance of these technologies.

The relative novelty of ML in the field of endoscopy, coupled with the frequent use of technical terminology around machine learning, has been a major factor that has affected its widespread acceptance among clinicians. Moreover, understanding the progress made in this area and adopting this new tool for clinical practice necessitates a working knowledge of the technical basis and a familiarity of the terminology used. In this review, the common terminology as well as a brief technical basis of image interpretation by AI-based applications will be described. This will be followed by an update on the role of AI in the prevention of colorectal cancer (CRC) and the evaluation of specific pancreatic lesions using EUS.

TECHNICAL BASIS AND COMMON TERMINOLOGY USED
ML in healthcare is a convergence of two diverse and complex areas, namely data science engineering and medicine, each with its unique expertise and jargon, which often results in a relationship that is fraught with misinterpretation and ambiguity. This fosters a disconnect that can be one of the major barriers of progress in this field. In this section, we define the relevant terminology and, in the process, also briefly describe the technical basis of the use of ML in endoscopy.

AI and ML
‘Artificial intelligence’ is a popular term that is commonly used interchangeably with ‘machine
learning’. In essence however, AI encompasses a broader field that includes path finding, logic representation and reasoning[4]. While ML is used to accomplish specific tasks, AI attempts to provide a more generic path for autonomous learning. The field of ML involves the use of existing data to build mathematical models that can predict expected outcomes on new data. There are two broad subtypes of ML models, namely, supervised and unsupervised learning. Supervised learning is achieved on a model with labelled data points (e.g.: Benign vs malignant), following which, the algorithm attempts to predict the labels upon a test set of unseen datapoints. On the other hand, unsupervised learning is used only to find the underlying structure, or a pattern within an unlabelled dataset; in other words, there is access to data but the outcome is not labelled (malignant or benign). Common examples of ML algorithms include deep neural networks (deep learning), support vector machines (SVM), gradient boosted trees and K-nearest neighbours.

**Feature extraction**

Before the generation of a predictive model, the data needs to be transformed into a numerical representation that can be fed to the ML algorithms. This process is called feature extraction and generally requires the input of medical experts in the field. Alternatively, modern ML techniques have automated this process and enabled extraction of features automatically from vision, language and sound datasets.

**Deep learning**

Deep learning (DL) is a type of ML algorithm originally known as Artificial Neural Networks (ANN’s). ANN’s are loosely inspired by the biological process found in a brain. They are comprised of mathematical neurons which “fire” if they are activated, and each neuron is connected to other neurons with “weights”. This connection of neurons and weights makes up what is known as “layers” in the neural network. Deep learning is when you have many layers (10’s to 100’s) connected, with millions of neurons and weights all interconnected. Deep learning models are very promising because they achieve extremely high rates of success in the fields of computer vision, natural language processing, machine translation, and speech recognition. This success is possible because of the enormous amount of data available, modern computing architectures and improved optimization algorithms. The attractiveness of deep learning is that it requires little expert domain knowledge in the form of feature extraction. The algorithm learns directly from the raw data (pixels, sound waves, text) and will automatically learn the correct “weights” which produce the most accurate results.

It has been shown that the lower layers of a deep learning model learn more abstract concepts such as “edges, shapes, lines” and the higher layers of the network learn more specific representations such as “nose, hair, eyes”.

**Computer aided detection and computer aided diagnosis**

ML algorithms that are applied to assist in the interpretation of medical images/videos are referred to as computer-aided detection (CADE) and computer-aided diagnosis (CADx). Distinction between CADe and CADx algorithms is important as the former is mainly used to ‘detect’ pathology, while the latter is able to ‘classify’ the pathology. For example, CADe will be used to identify a colonic polyp in a study, while CADx will enable characterization of the polyp as adenomatous or non-adenomatous. This has profound implications in the management of patients undergoing colonoscopy. Therefore, it necessitates a high degree of accuracy, reliability and external validity. Apart from this, ML algorithms can also be applied to guide interventions and is usually referred to as ‘image-guided interventions’; like the use of ML to guide the necessity and site of biopsy using EUS imaging.

**ROLE OF AI IN SCREENING COLONOSCOPY FOR CRC**

CRC is a leading cause of death with a rising incidence especially in younger age-groups, both in western countries as well as many Asian countries in the recent past[20,21]. Most CRC develops from pre-existing adenomas which are pre-cancerous lesions[22]. Resection of adenomas during a screening colonoscopy has been shown to be instrumental in lowering the risk of CRC[23]. Thus, adenoma detection rate (ADR) in particular, apart from withdrawal time, clean colon and caecal intubation rate, is considered to be a vital quality indicator of CRC screening programs. For every 1% increase in adenoma detection rate, there is an associated 3% decrease in interval incidence of colon cancer[23]. Non-visual-ization is a major factor that can lower ADR in most cases. This can mainly be attributed to polyps hidden in poorly accessible areas like the left colon, or behind mucosal folds. Besides hidden polyps, those that are technically in the visual field may still be missed if they are subtle, diminutive, transiently visible, partially obscured by debris, or seen on the edge of the screen[24]. High quality bowel preparation, strict adherence to globally accepted standards for withdrawal time, meticulous mucosal inspection techniques and the use of endoscopes with wider viewing angles can, to a certain extent, address these issues[25]. However, even with the currently performed, careful colonoscopy, rates of missed adenomas can be as high as 26% for adenomatous polyps less than 5 mm in size[26]. Even in the case of advanced adenomas, adenoma missed rates (AMR) has been reported to be as high as 5.4%[27].
An intuitive approach to this problem would be to employ measures that can supplement our capacity for visualisation. To that end, recent studies using full-spectrum colonoscopy (FUSE), which provides 330° of view, have been described to access previously hidden areas during a colonoscopy. However, results have shown sub-optimal with a persistent AMR ranging between 7% to 20.5%[28,29]. Another option explored was the use of second observers (nurse observers/trainees). However, even this approach was not effective in bridging the gap and reducing AMR during screening colonoscopy[30-32]. This indicates that extending the field of vision or supplementing the limits of visualisation with additional human eyes, may not fully overcome the inherent deficiencies of human attention and visualisation, especially in the context of subtle colonic lesions.

In this context, the recent innovation of AI plays a pivotal role in CADe and CADx systems for polyp detection and characterization respectively. They have been pegged as a potentially disruptive technology that can herald a new era in CRC prevention strategies. The success and practical utility of these systems hinges on a low false positive rate and low latency time defined as the time from the first appearance of the polyp to detection in real-time[33]. In other words, these systems have to show high accuracy, fidelity, consistency and enable real-time detection (low latency time) of polyps that are otherwise missed[34]. In this section, we will summarise the current status of ML systems in this area and discuss the future of this technology in the CRC prevention programs.

### Evolution of AI in polyp detection

Initial application of AI in gastroenterology was limited to ‘edge detection’ by identifying sharp changes in image brightness and ‘region growing’ by a group of pixels of similar properties. This was essentially useful in lesions when edges were undetectable in standard endoscopic images[35]. The first polyp detection software CoLD (colorectal lesions detector), was developed in 2003 with an accuracy of 93% [36]. With the advancement of endoscopic imaging quality, subsequent DNN systems could make use of additional features like color, temporal factors and texture of the polyps with a high level of precision[37]. Subsequently, novel deep learning techniques were applied that could take advantage of image processing and vast datasets, to enable complex functions like polyp classification leading to a shift in our approach. Since, then, multiple systems have been developed that have shown improved results and accuracy[16,17,38,39]. Moreover, robust image databases and the use of video-based algorithms have provided an effective training as well as testing platform. This has led to an array of CADe and CADx systems that have become commercially available in the last 5 years[16,38,40].

### Real-time use of CADe systems for polyp detection

CADe systems have been well-validated in real-time colonoscopic examinations. They have demonstrated high accuracy for polyp detection, especially for polyps less than 5mm and those between 5-9 mm. These systems have enabled the identification of lesions that are subtle, obscured by debris, poorly visualised due to specular reflections or lesions at the edge of the screen[41]. Different CADe techniques have demonstrated promising results in polyp detection, especially when combining different DL methodologies. Not surprisingly, larger datasets appear to improve overall measures of performance[17]. Among these, a CADe system developed by Wang et al.[34] was the first one to be validated in a large multi-centric trial. The system was developed on a large dataset of over 1200 patients and was independently validated on two separate datasets, including over 27,000 images and nearly 200 colonoscopy videos, generating 100% specificity and a latency of 76.8 ms. Patients were then randomized to undergo routine diagnostic colonoscopy (n = 536) or real-time CADe assisted colonoscopy (n = 520). The CADe system significantly increased ADR (29.1% vs 20.3%; P < 0.001), mean number of adenomas per patient (0.35 vs 0.31; P < 0.001), and overall polyp detection rate (45% vs 29%; P < 0.001). Not only did the CADe system increase polyp and adenoma detection rates, it identified significantly more flat and sessile polyps, as well as diminutive polyps. There were however, a few false positives in this study (0.075 per colonoscopy) which were attributed to air bubbles, mucosal inflammation and retained fecal matter. The same study group then performed another study of their CADe system tandem colonoscopies were performed for each participant by the same blinded endoscopist, wherein, patients were randomly assigned to groups that received either CADe assisted colonoscopy or routine colonoscopy first, followed immediately by the other procedure. They found that AMR was significantly lower with CADe assisted colonoscopy (13.89%) than with routine colonoscopy (40%)[24].

Real-time CADe during screening colonoscopy, tested on several hours of colonoscopy videos, were also found to have a high accuracy of almost 97%[15,38]. In a study by authors Urban et al[15], deep neural networks (DNN) to detect polyps was developed using a diverse and representative set of 8641 hand labeled images from screening colonoscopies collected from over 2000 patients. This was tested on 20 colonoscopy videos. Gold standards were developed with the help of experts who were asked to identify all polyps in de-identified videos. They found that their CADe system had an accuracy of 96.5% and can detect and localize polyps well within real-time constraints. In a recent publication, Repici et al[42] evaluated the AI system developed by Medtronic based on a convolutional neural network, called GI-Genius™ (Figure 1). In this randomized, controlled study, GI-Genius™ detected significantly more adenomas with an adenoma detection rate of 54.8%, irrespective of withdrawal time[31] (Figure 2). Adenomas detected per colonoscopy were also higher in the GI-Genius™ group (mean 1.07 ± 1.54) than...
in the control group (mean 0.71 ± 1.20) (incidence rate ratio 1.46; 95% CI, 1.15-1.86). This improved ADR was mainly seen in polyps < 5 mm and polyps with 5-9 mm diameter. These findings indicate that CADe systems are clearly an effective strategy to increase ADR and could prove to be indispensable in the future\[42\]. The imperative question however, is not whether it can merely ‘detect’ what was missed by the human eye, but whether it can provide additional information by identifying patterns that are otherwise invisible to the human eye?

**The leap from polyp detection to histological characterization**

The leap from merely detecting a polyp to accurate histological characterization has opened up a new paradigm of screening colonoscopy for CRC prevention. Two alternate strategies have been proposed for the management of diminutive polyps that may have far-reaching consequences in clinical practice and healthcare economics. These two approaches are ‘Resect and discard’ and ‘leave-in-situ’ strategies \[43,44\]. The advanced imaging capabilities achieved through CADx make the above choices a welcome reality. Thus, when an adenomatous diminutive polyp is diagnosed by a CADx system, ‘resect and discard’ approach can be safely undertaken. At the same time, a non-neoplastic diminutive polyp found on colonoscopy can be safely managed with ‘leave-in-situ’ strategy. These alternate strategies have important advantages like cost reduction, avoiding adverse events related to polypectomy with its resultant shorter procedure time\[45\]. Both these strategies are highly dependent on advanced imaging systems that provide a precise, real-time identification of the polyp. However, both strategies have not found good penetration outside of expert centres as current imaging systems do not meet the appropriate thresholds for accuracy\[44,46\]. CADx systems could be the answer in these situations by
improving the diagnostic accuracy of existing imaging systems.[47]

Initial experience with CADx systems showed that they were able to discriminate adenomatous from hyperplastic polyps when using magnification chromoendoscopy or magnification narrow-band imaging (NBI).[18,48,49]. However, these used traditional AI techniques which limited its real-time application as it required manual segmentation of polyp margins and captured images that required magnification techniques that were not widely available. With the development of DNN techniques, newer CADx systems addressed these issues and have shown a lot of promise in preliminary real-time polyp classification. In a prospective single-operator trial of 41 patients, diagnostic accuracy of 93.2% was shown for a real-time CADx system on 118 colorectal lesions evaluated with magnifying NBI before resection. Among the subset of patients with diminutive polyps, exceeding the Preservation and Incorporation of Valuable Endoscopic Innovations (PIVI) initiative threshold of ≥ 90% for the “resect and discard” strategy, 92.7% showed concordance between the CADx diagnosis and the pathological findings.[50]. This highlights the massive impact that CADx systems can potentially have in reducing costs associated with CRC screening programs.

Advanced imaging techniques such as NBI have come into routine use and supplemented our ability to better characterize colonic polyps. Moreover, emerging techniques of incorporating NBI images, with and without magnification, to create datasets for CADx systems, especially with larger image and video banks, have yielded highly sensitive systems with high negative predictive values.[16,48,51]. The level of performance of these CADx systems in conjunction with NBI imaging have been shown to meet the minimum threshold for a ‘diagnose and leave-in-situ’ strategy (90% NPV) as proposed by the American Society for Gastrointestinal Endoscopy PIVI initiative.[42]. In a very interesting study by Jin et al.[14], CADx improved the overall accuracy of optical polyp diagnosis from 82.5% to 88.5% (P < 0.05). In particular, CADx assistance was most beneficial to novices with limited training in using enhanced imaging techniques for polyp characterization, where accuracy jumped from 73.8% to 85.6% which was comparable to the endoscopy experts. This finding has significant implications on the feasibility of implementation of CADx systems in routine practice.

**Endocytoscopy**

Endocytoscopy is another evolving technology that involves ultramagnification that can detect microscopic changes at the level of the nuclei (abnormal spindle shaped nucleus, loss of polarity).[52]. It is conceivable that innovation in endocytoscopy with CADx systems may one day, replace conventional histopathological examination through tissue acquisition, fixation, staining and microscopic examination. In a study of 791 consecutive patients who underwent colonoscopy with endocytoscopes, CADx was able to characterize diminutive rectosigmoid polyps in real time with an accuracy of 94% and an NPV of 96%, which supports the use of “diagnose and leave in situ strategy” for nonneoplastic polyps.[11].

**Limitations of AI in screening colonoscopy**

Although automatic polyp detection has shown promising results, it is yet to live up to expectations. A number of factors can affect the performance of AI-based systems including camera motion, strong light reflection, poor focus, polyp morphology, presence of bubbles and retained fecal material. When it comes to CADx systems, accuracy of tissue characterisation can be affected by inadequate staining and surface cleaning and inability to obtain a cross sectional view.[53]. Nevertheless, the advent of AI system through improved detection and histological characterisation could lead to increased ADR and reduce missed adenomas, leading to lowered incidence of interval CRC.

**Future of AI in screening colonoscopy**

CADx systems, once validated in real-time use for polyp characterization, could enable the implementation of ‘Resect and discard’ and ‘leave-in-situ’ strategies. These strategies have been shown to reduce the cost of care dramatically. In a study by Mori et al.[54], the use of CADx system for polyp characterisation in order to implement ‘leave in situ’ strategy resulted in a significant cost saving of 10.9%. In addition, these strategies could potentially reduce procedure time and reduce adverse events related to unnecessary polypectomies.

Recent findings have shown promising results with the use of video analysis and its potential advantages. Video-based algorithms have several advantages over image-based algorithms. Since a video is basically a series of images over time, it provides vital spatiotemporal information as in real life, that is not available in still images. When such spatiotemporal information is combined with CAD system, its performance can be significantly improved. This is especially true for colonic polyps since there is marked difference between the polyp and the surrounding mucosa which is easily picked up on a video analysis.[55]. However, video-based algorithms need further validation in controlled settings.

Another aspect where AI could potentially improve colonoscopy performance, in general and screening colonoscopy in particular, is its role in quality control and monitoring.[56,57]. These algorithms can potentially monitor endoscopic quality, by which it can indicate colonic surface missed during withdrawal, need for a slower speed of withdrawal, areas of poor bowel preparation necessitating adequate cleansing before moving on. Although this area has not been investigated
thoroughly, an argument can be made that this might have an equal, if not bigger, impact on clinical outcomes of CRC screening programs than a specific lesion detection tool for a specific pathology.

Several questions remain to be answered in order to fine-tune the role of AI in polyp detection. However, with the advent of advanced systems that combine multiple functions, the time seems appropriate to embrace this technology and troubleshoot issues along the way, rather than delay the adoption of AI in our daily practice in the hope of achieving perfection.

**ROLE OF AI IN THE EVALUATION OF PANCREATIC DUCTAL ADENOCARCINOMA USING EUS**

Pancreatic ductal adenocarcinoma (PDAC) has a dismal prognosis with a five-year survival rate of approximately 6%[88]. PDAC is also associated with significant morbidity and accounts for 3.9% Disability Adjusted Life Years (DALY) related to cancers. Moreover, future estimates indicate that the PDAC burden is likely to double within the next four decades[39]. The incidence of PDAC in the United States is increasing by 0.5% to 1.0% per year, and is expected to be the second-leading cause of cancer-related mortality by 2030[60].

Most patients with PDAC are unresectable at the time of diagnosis owing to locally advanced (30%-35%) or metastatic disease (50%-55%) at presentation[60]. Surgical resection is possible only in around 20% of patients[61]. Despite curative resection, most of these patients will eventually have a recurrence, with a 5 year survival of around 25%-62%. However, cancers < 1 cm in size at the time of diagnosis, have been shown to have an excellent response following resection with a survival rate as high as 84.4%[63]. This highlights the paramount importance of screening and early detection for PDAC. Unfortunately, well-defined pre-malignant conditions and proper guidelines are lacking for pancreatic cancer, as compared to CRC. Moreover, current modalities of screening are inadequate and merit further evaluation before recommending routine clinical use.

Diagnosis of PDAC relies on accurate identification of the tumor by various imaging modalities, followed by a reliable method of tissue acquisition to confirm the histological characteristics of malignancy. Currently available modalities for imaging include transabdominal ultrasonography, computed tomography (CT), magnetic resonance imaging, EUS, and endoscopic retrograde cholangiopancreatography. Of these imaging modalities, EUS enables real-time observation of the pancreas with high spatial resolution, and the sensitivity of detection of PDAC using EUS has been reported to be as high as 94%[64]. Numerous studies indicate that EUS is a highly sensitive modality for the detection of pancreatic tumours and its application is especially useful for lesions less than 2 cm in size which may be missed on contrast enhanced CT studies[65]. Although the sensitivity for tumour detection is high, it is also important to note that it has a very high negative predictive value (NPV) in the background of a normal pancreas[66].

The major drawback of EUS is the fact that it is highly operator dependent and the learning curve to perfect the techniques of EUS imaging can be quite long. The American Society for Gastrointestinal Endoscopy recommends that a trainee should undergo at least two years of standard GI fellowship followed by one year of pancreatic EUS training prior to independently performing EUS[67]. ASGE also recommends that an endosonographer should perform a minimum of 150 supervised EUS procedures, including 75 pancreaticobiliary cases and 50 EUS-guided fine needle aspiration (EUS-FNA) procedures, to achieve competence in this area. In addition, specialised EUS training centres are usually inaccessible hampering the widespread application of standardised protocols for EUS screening of the pancreas[68].

Another major challenge that is faced by endosonographers is inability to correctly identify PDAC in patients with chronic pancreatitis (CP). Several studies have shown that the diagnostic yield of EUS and EUS-guided fine needle aspiration (FNA) are markedly decreased in the presence of CP[69,70]. This can be attributed to the fact that neoplastic lesions and inflammatory masses usually have a similar sonomorphology with very subtle differentiating characteristics. Studies by Fritscher-Ravens et al[71] and Varadarajulu et al[70] found EUS sensitivity to range from 54% to 73.4% respectively, in patients with CP[70,71].

AI could potentially address both these issues. In this section, a brief account of the progress made by AI-based CAD systems in image differentiation among patients with chronic pancreatitis will be presented; followed by the recent developments in the field of AI assisted EUS training systems.

**Evolution of AI in endosonography**

Similar to screening colonoscopy, AI is being actively investigated in the early diagnosis of PDAC. However, its application in this area is still in its infancy with no commercially available CAD systems yet. Initial reports focus on integrating AI with EUS imaging to identify PDAC in the background of CP. Several sonographic features of CP such as calcification and the presence of pseudotumors with intense desmoplasia pose significant challenges to making an accurate diagnosis of PDAC in these patients[72].

The first report of the use of an AI based system for the diagnosis of PDAC was by Norton et al[73] in 2001. In this study, 35 patients were included, of which 21 patients were histologically proven to have PDAC, while 14 patients had focal CP. Representative images with the region of interest were fed into a
CAD system which was then trained to identify subtle differences in the gray scale and overall brightness within the images. These features were then assessed to differentiate between PDAC and focal CP. This early CAD system was found to have an overall diagnostic sensitivity of nearly 89%. In an effort to reduce the chances of missed malignancy, the authors found that even when the sensitivity for malignancy was set to 100%, the overall diagnostic accuracy was still around 80%. This was remarkably close to the 85% accuracy that was observed among blinded, trained endosonographers[73]. Although the technology used in this study was primitive to say the least, it was the first study that demonstrated the feasibility of integrating AI into diagnostic studies using EUS, and formed the foundation to the studies that followed. Since then, many attempts at applying conventional CAD using ANNs or SVMs have been tested, both with traditional grayscale texture features on B-mode imaging as well as on elastography images. The Area under Receiver operating characteristic curve (AUROC) in these studies ranged from 0.8 to 0.94[74-78]. Though these studies showed promising results, the accuracy in the background of CP was still far from ideal.

One of the promises of AI in the field of endoscopy, is the ability of the machine to make a diagnosis in real time imaging and assist the endoscopist in planning the next step in the management of the patient during the procedure itself. However, the multiple intricate post-processing steps that were needed in the studies that assessed the role of CAD system in EUS precluded their use during real time imaging. This was one of the main reasons for the technology remaining dormant for years after the initial proof of concept in 2001. However, encouraged by the benefits of CADe and CADx systems in screening colonoscopy, there has been renewed interest, in recent years, on the application of AI systems in EUS. A sudden surge of publications that have employed novel CAD systems for pancreatic lesions combining EUS-FNA and contrast enhanced EUS studies has opened up new avenues for the role of AI based technology in this area.

**AI and EUS elastography**

EUS elastography (EUS-E) can transform the tissue properties based on elastic coefficients, into visible images composed of color pixels. This can provide vital information regarding the pathological state of the tissue under study and has been shown to be useful in the evaluation of pancreatic lesions. In a seminal study by Săftoiu *et al*[79] real-time EUS-E avoided motion artifacts and color perception errors that arose from individual selection, manipulation bias and static image analysis. Following this, a large multicentric trial was conducted in Europe in which, 744 EUS-E images from 258 patients with pancreatic lesions were studied. A detailed analysis of the color hue histogram data from the dynamic sequence of EUS-E images was performed using a novel neural network in order to distinguish benign from malignant patterns. An overall sensitivity of 87.6%, specificity of 82.9%, and positive predictive value (PPV) of 96.3% indicated that the combination of EUS-E with AI based software, could be beneficial in the real-time evaluation of pancreatic lesions[80].

**Role of AI in contrast EUS and fine needle biopsy**

EUS guided fine needle biopsy (EUS-FNB) has enabled reliable tissue acquisition and accurate histological diagnosis in patients with PDAC. In fact, it is considered to be the cornerstone of management of pancreatic lesions < 3 cms[81]. Multiple studies have documented a high diagnostic accuracy of EUS-FNB for PDAC with a pooled sensitivity of 87% and specificity of 96%[82]. However, these results have been negatively impacted by the presence of chronic pancreatitis. Intense desmoplasia, fibrosis and calcifications seen among patients with CP can decrease diagnostic yield of EUS-FNB because of the higher tissue impedance and inaccessibility of the lesion due to various factors[69,83]. Moreover, Rapid On site examination of the cytology obtained from EUS-FNA which has been shown to be a major factor that impacts diagnostic yield, is not feasible in many centers [84]. ML based algorithms have shown promise in this area by augmenting visual inspection of the histopathology slides. In a study by Inoue *et al*[85], an ML-based automated visual inspection system could reliably highlight areas of abnormal cellularity on the stained smears obtained after an EUS-FNB from solid pancreatic lesions.

Contrast harmonic EUS (C-EUS) uses the enhancement properties of the solid lesions and categorizes them into different patterns[86]. Multiple studies have shown C-EUS to have a pooled sensitivity of around 93% and specificity ranging between 80%-89% for pancreatic lesions[87-89]. Its ability to highlight areas of increased vascularity and to outline areas of reduced vascularity due to necrosis and fibrosis have been used during EUS-FNA, to increase the diagnostic yield[90-94].

In an elegant study by Săftoiu and colleagues, a time intensity curve was made for patients with pancreatic lesions, using dynamic C-EUS examinations. Using a set of 7 features that were extracted from the data using a convolutional neural networks (CNN), sensitivity, specificity, NPV and PPV were 94.6%, 94.4%, 89.4% and 97.2%, respectively, was reported[95]. Since then, multiple studies are underway that highlight a significant ancillary role played by AI-based systems in improving the diagnostic yield of EUS-FNB with C-EUS.

**Future of AI in the field of endosonography**

The immediate clinical application of the results of studies using AI based systems in the field of...
endosonography are unfortunately limited, to say the least. This is in part due to the necessity of pre-
analysis image preparation and post-processing steps that preclude real-time application[96].

A major factor in the development of machine learning models for EUS is the sheer volume of
labelled images required to improve accuracy. ImageNet is one of the most popular datasets used in
machine learning models. This dataset contains as many as 14 million labelled images, which is used by
a majority of image recognition software. This essentially means that it takes millions of labelled images
to train a machine to accurately interpret an image or video. To add to the problem, the concept of,
"Garbage in and garbage out", is another cause for concern. This means that if we feed the machine poor
quality/ poorly labelled images, the output will be inaccurate. So, apart from the quantity of labelled
images, quality is equally, if not more important.

With regard to EUS, trained endosonographers are not widely available. The time and resources
required to have trained endosonographers read, label and edit an adequate number of high quality
videos is impractical to implement. This is why, there has been a recent change in the paradigm of ML
in EUS. Instead of depending on endosonographers alone to edit videos, investigators have begun
training the machine to detect stations which can result in shortened videos focussed on the regions of
interest. This would significantly reduce the time and resources required to create a high quality dataset
of EUS images.

In a study by Zhang and colleagues, a novel CNN was evaluated for the accurate recognition of the
EUS station as well as segment the pancreas for more detailed evaluation. Compared with EUS experts,
the models achieved 90.0% accuracy in classification, which is comparable with that of experts[97]. In
2019, Kuwahara et al[98] evaluated the use of DL based CAD with CNN to achieve two objectives –
accurately determine the station of the EUS probe as well as differentiate between malignant vs benign
intraductal papillary mucinous neoplasms (IPMN) of the pancreas. The area under ROC curve for CAD
systems to diagnose malignant IPMNs was found to be was as high as 0.98 (P < 0.001). The sensitivity,
specificity, and accuracy was found to be 95.7%, 92.6%, and 94.0%, respectively; which was significantly
higher as compared to expert endoscopists in the study.

In addition to accurate classification of lesions, AI based systems could potentially be beneficial by
supplementing EUS training programs. This can eventually result in a uniform, high quality EUS
examinations which are more amenable to the application of CAD systems that can identify and
diagnose pancreatic lesions in real-time. In the study by Zhang et al[97], the developed CAD system was
subsequently validated on trainees, where they found that diagnostic accuracy improved from 67.2% to
a significant 78.4% for the evaluation of solid pancreatic lesions.

In the most recent study by Tonozuka et al[99], a complex CNN based DL method was employed for
the detection of PDAC. They found improved performance of this automated system with AUROC of
0.924 and 0.940 in the validation and test setting, respectively. However, there have been very few head-
to-head comparison studies that have compared the efficacy of CAD systems for the diagnosis of
PDAC and its role in image differentiation merits further clarity.

There are several potential benefits likely to arise from the use of AI based CAD systems in the field
of EUS. Firstly, AI can augment EUS expertise especially by shortening the learning curve. Although,
there is very little data to support this statement, initial results are extremely encouraging and it would
be reasonable to surmise a significant role played by AI-based automated systems in EUS training
programs. Secondly, the recent innovations using CNN based DL algorithms have the potential to
significantly augment the diagnostic accuracy of EUS and could, conceivably overcome the inherent
deficiencies of human error, visualisation, inattention and fatigue. Finally, our rudimentary foray into
this area, coupled with the encouraging results seen in the case of endoscopcy-based CADx systems
for colonic polyps; could pave the way for optical diagnosis of pancreatic lesions in the future. This
could theoretically, expand the role of EUS in the context of solid pancreatic lesions, by enabling the
accurate diagnosis of lesions which are poorly accessible, failed EUS-FNA (high tissue impedance,
intervening vessels) or poor visualisation due to calcifications and fibrosis secondary to CP.

However, the current systems possess major drawbacks that hamper the uniform application of AI -
based CAD systems for EUS in clinical practice. One of the major drawbacks is the "black box
phenomenon" where the basis of a decision taken by the machine is not clearly understood by the
programmers and developers. This makes it difficult to course-correct the system in case of sub-optimal
accuracy. Another important drawback is the fact that real-time video and the tactile understanding of
the location of the scope, plays a major role in decisions with regard to EUS-FNA. These data inputs are
currently not factored into the DL algorithms and could significantly hamper its clinical applicability.

CONCLUSION

The tremendous progress witnessed in the field of artificial intelligence and machine learning has
enabled the development of novel and innovative algorithms that can perform specific functions in the
field of endoscopy. Although AI based systems have shown immense promise in the prevention of CRC
by detecting and characterising colonic polyps, the systematic incorporation of these systems in our
everyday practice is still lacking. While it is intuitive to engage our efforts on the implementation of
these systems in our endoscopy practice, there needs to be a clear agreement and consensus as to the specific gaps that can be addressed by AI based systems. This could improve efficiency of implementation and efficacy, thereby enabling the translation from mere ‘promise’ to measurable ‘impact’ on global screening programs. There are encouraging steps taken in that regard, where novel approaches like ‘Leave-in-situ’ and ‘Resect and Discard’, can potentially change the landscape of CRC screening programs. Validated and reliable CADx systems can enable the adoption of these strategies. The most critical and exciting aspect is the potential to implement these strategies at the community level in emerging economies like India, where CRC prevalence have shown alarming upward trends in the past decade, owing to a higher prevalence of metabolic risk factors and changing patterns of diet and lifestyle practices. These strategies can reduce the cost of screening programs significantly by obviating the need for histopathological evaluation of small diminutive polyps. In addition, the reduced requirement of specialised man-power, logistical issues and equipment installation at primary care centres in the community can make CRC screening programs economically viable and a welcome addition to global efforts to reduce the burden of CRC.

AI in the field of EUS, however, is still in its infancy. Given the present lacunae in the diagnosis of early PDAC, there is significant scope for the application of AI-based CADe and CADx systems, which can augment our capabilities to manage patients with solid pancreatic lesions with/without CP in the future. However, there is an acute need to re-examine the available approaches to development of CADe and CADx systems in this area. The specific functions and questions that need the assistance of AI based systems needs to be clarified by expert consensus before we embark further on the development of newer systems.

In conclusion, there is an urgent need, now more than ever before, for future collaborative projects with the ever-expanding world of data science and artificial intelligence, which could pave the way for a brave new world, of man and machine, acting in concert to bring about the technological age of modern medicine.

FOOTNOTES

Author contributions: Rao B H, Trieu JA, and Nair P performed the majority of the writing, prepared the figures and tables; Gressel G provided valuable inputs in the technical aspects of Artificial intelligence and data science; Venu M and Venu RP designed the outline and coordinated the writing of the paper.

Conflict-of-interest statement: The authors declare that they have no conflict of interest.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

Country/Territory of origin: India

ORCID number: Harshavardhan Rao B 0000-0002-6472-4396; Judy A Trieu 0000-0001-6877-2291; Priya Nair 0000-0002-9889-2275; Gilad Gressel 0000-0002-6631-7858; Mukund Venu 0000-0001-9589-4278; Rama P Venu 0000-0003-2251-3942.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

AIGE


Gut Liver 2016; 10: 419-422.


AIGE


Puli SR, Bechtold ML, Buxbaum JL, Eloubeidi MA. How good is endoscopic ultrasound-guided fine-needle aspiration in diagnosing the correct etiology for a solid pancreatic mass? *Pancr'us* 2013; 42: 20-26 [PMID: 23254913 DOI: 10.1097/MPA.0b013e1321546a79]


Inoue H, Ogo K, Tabuchi M, Yamane N, Oka H. An automatic visual inspection method based on supervised machine learning for rapid on-site evaluation in EUS-FNA. *2014 Proc SICE Annu Conf* 2014; 1114-1119 [DOI: 10.1109/sic2014.6935253]


Facciorusso A, Cotsoglou C, Chierici A, Mare R, Crino SF, Muscatelli N. Contrast-Enhanced Harmonic Endoscopic


Artificial intelligence and machine learning in colorectal cancer

Muhammad Awidi, Arindam Bagga

Specialty type: Oncology

Provenance and peer review:
Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report’s scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C, C
Grade D (Fair): 0
Grade E (Poor): E

P-Reviewer: Li C, China; Nazari N, Iran; Yakar M, Turkey

Received: January 17, 2022
Peer-review started: January 17, 2022
First decision: March 8, 2022
Revised: March 24, 2022
Accepted: June 20, 2022
Article in press: June 20, 2022
Published online: June 28, 2022

Abstract

Colorectal cancer (CRC) is a heterogeneous illness characterized by various epigenetic and microenvironmental changes and is the third-highest cause of cancer-related death in the US. Artificial intelligence (AI) with its ability to allow automatic learning and improvement from experiences using statistical methods and Deep learning has made a distinctive contribution to the diagnosis and treatment of several cancer types. This review discusses the uses and application of AI in CRC screening using automated polyp detection assistance technologies capable of accurately detecting polyps during colonoscopy and classifying them. Furthermore, we summarize the current research initiatives geared towards building computer-assisted diagnostic algorithms that aim at improving the diagnostic accuracy of benign from premalignant lesions. Considering the evolving transition to more personalized and tailored treatment strategies for CRC, the review also discusses the development of machine learning algorithms to understand responses to therapies and mechanisms of resistance as well as the future roles that AI applications may play in assisting in the treatment of CRC with the aim to improve disease outcomes. We also discuss the constraints and limitations of the use of AI systems. While the medical profession remains enthusiastic about the future of AI and machine learning, large-scale randomized clinical trials are needed to analyze AI algorithms before they can be used.

Key Words: Artificial intelligence; Machine learning; Colonic polyps; Colorectal neoplasms; Computer-aided diagnosis; Precision oncology

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.
INTRODUCTION

In the United States, the third leading cause of cancer-related deaths is colorectal cancer (CRC) [1]. Since 1980, the number of people diagnosed with colon or rectal cancer has decreased due to improved screening guidelines and lifestyle-related risk factors modification. In addition, treatments for colorectal cancer have improved over the last few decades [2]. CRC is a diverse group of diseases with differences in epidemiology, histology, genomics, and host immune responses [3,4]. Recognizing the diversity of the disease, and the importance of personalized medicine, machine learning models have been utilized to improve detection rates, diagnosis, and treatment of CRC.

Artificial intelligence (AI) is a computer science field dedicated to developing systems capable of performing tasks that typically require human-level intelligence [5]. It is a broad term used to encompass Machine learning (ML), a subset of AI algorithms that allows automatic learning and improvement from experiences using statistical methods and deep learning which imitates higher level human data processing by using multi-layered neural networks for extractions and self-training algorithms [6] (Figure 1).

The increased utilization of this novel technology has made a distinctive contribution to the diagnosis and treatment of several cancer types. From AI models to reduce rates of missed adenomas to novel computer assisted drug delivery techniques and robotic surgery colorectal carcinoma treatment entered a new area rapidly moving towards precision and personalized medicine [7,8].

Our review aims to analyze the AI uses and application in CRC screening, diagnosis, and treatment. In addition, we will discuss potential future directions and limitations for the use of AI systems.

SCREENING

Colorectal screening remains the gold standard for improving patient clinical outcomes, such as avoiding treatment delays and lowering CRC morbidity and mortality [9]. CRC patients are diagnosed at advanced stages of the disease in 60%-70% of cases [9].

It is thought that the alterations from the normal mucosa to malignant state lesion take almost 10 to 20 years [10]. Colonoscopy, flexible sigmoidoscopy, and less invasive capsule endoscopy, computed tomography chrography, blood in stool tests, fecal immune-chemical testing, and multi-target cell DNA testing are just a few of the screening options available for CRC [11,12]. Colonoscopy is the gold standard screening test, though it is not without flaws [13]. It has been reported that around 9% of cases of CRC occurred within three years following a negative colonoscopy [14]. Adenoma detection rates are very variable with reported detection rates of 7% to 50% [15]. The wide range of detection rates is due to different factors, including endoscopic procedural experience, pre-procedure bowel preparation, time of procedure termination, use of sedation, flexure visualization, image enhanced endoscopy, and the presence of flat or diminished polyps [16,17].

The growing interest of AI in CRC yielded automated polyp detection assisted technology to aid in the detection and diagnosis of polyps during colonoscopy [5]. In addition, technologies that use deep learning techniques to improve detection rates and localize premalignant lesions are available and being applied [18].

A recent randomized controlled trial studied the effect of computer aided detection deep learning models on polyps and adenoma detection rates. The trial randomized 1058 patients to either conventional colonoscopy \( (n = 536) \) or colonoscopy with computer aided detection system \( (n = 522) \). In the computer aided detection system group there was an increase in both the adenoma detection rates, 29.1% vs 20.3%, \( P < 0.001 \), in addition to the mean number of identified adenomas per patient, 0.53 vs 0.31, \( P < 0.001 \), in comparison to the group assigned standard colonoscopy. This trial, however, did not reveal a significant statistical difference for the detection of large adenomas between the groups (77 vs...
Interestingly, the computer aided detection system arm had more hyperplastic adenomas (114 vs 52, $P < 0.001$) and diminutive polyps (185 vs 102, $P < 0.001$) identified. This study demonstrates the impact of AI-assisted colonoscopy technologies on the detection of small polyps that even highly trained endoscopists may miss [19].

Karkanis et al [20] used color and texture analysis of mucosal surfaces based on color wave covariance features were used to develop a computer-assisted diagnostic algorithm for automatic polyp identification. Rather than a real-time recognition system, the system was able to identify precancerous lesions in static endoscopic images. It accomplished that by examining frame images extracted from 60 colonoscopy video sequences containing small polyps with a sensitivity and specificity of 99.3% and 93.6% respectively.

In a study to evaluate deep learning algorithms for automated polyp detection during colonoscopy using colonoscopy images, colonoscopy videos obtained from four different datasets resulted a significant improvement in real-time colonoscopy video analysis by processing at least 25 frames per second with a latency of 76.8 milliseconds [65].

A recent systematic review and meta-analysis that included 48 studies showed a significant increase in both polyp detection rates [odds ratio (OR) 1.75, 95%CI 1.56-1.96; $P < 0.001$] as well as adenoma detection rates (OR 1.53, 95%CI 1.32-1.77; $P < 0.001$) patients who had a colonoscopy with AI compared to those who did not [21].

Recognizing that colonoscopy is a highly operator-dependent procedure, challenges such as light conditions, morphology of colorectal polyps during colonoscopy, and size could be overcome by AI computer assisted diagnostic systems as they serve as an “extra pair of eyes” and improve adenoma detection rates.

Several alternative screening tools to conventional colonoscopy have been developed. A modified computed tomography (CT) examination known as virtual colonoscopy or computed tomographic colonography (CTC) was first described in 1994 [22]. Its ability to evaluate the entire colorectum, rapid acquisition of imaging, and lack of sedation makes it a valuable alternative for certain patients. The effectiveness of CTC in detecting asymptomatic colorectal lesions is still a point of contention. Several studies reported identification of 90 percent of patients with asymptomatic adenomas or cancers (≥ 10 mm in diameter) using CTC [23,24]. AI-based algorithm concepts have been used to obtain optimal diagnostics standards and image qualities to aid in CRC detection and diagnosis using CTC. Grosu et al [25] developed a machine learning method that had an area under the curve (AUC) of 0.91, a sensitivity of 82%, a specificity of 85% in differentiating between benign and precancerous lesions in average risk asymptomatic patients using CTC. In another study, Song et al [26] developed a virtual pathological model to see if image high-order differentiations (curvature and gradient) could be used to distinguish colorectal lesions (neoplastic and non-neoplastic). The results revealed an improvement of receiver operating characteristic (ROC) curve (AUC) from 0.74 (Using image intensity alone) to 0.85 (Using texture features from high-order differentiations).

In cases of incomplete colonoscopy or when evaluating the small intestines, capsule endoscopy (CE) is used as a minimally invasive technique. It acquires images as it passes through the gastrointestinal tract [27]. Hence, CE can be affected by laxative use. In addition, it requires manual interpretation and analysis of acquired images which is particularly time consuming [28,29]. AI-based systems are being used to automate the reading and examination of the results to reduce the time and the human error inherently present when reading images thereby improving adenoma detection rates [30,31]. Novel
algorithms were developed to match CE and colonoscopy-identified polyps based on their size, morphology and location as well as utilizing deep convolutional neural networks for automatic colorectal polyp detection. When compared to the manual process of polyp detection, localization had a high sensitivity (97.1%), accuracy (96.4%), and specificity (93.3%) for identifying polyps.[30]

Blood-based screening approaches have been developed to detect CRC at early stages. Demographic characteristics and blood test results such as complete blood count (CBC), which may indicate iron deficiency, microcytic anemia, or elevated red cell distribution width are frequently used to evaluate the risk of developing CRC.[32-34] An AI-assisted prediction model (Mescore®, Calgary, Alberta, Canada) was designed to identify people at high risk for CRC using parameters such as age, sex, and CBC data collected 3 to 6 mo prior to cancer diagnosis. A study using this AI-assisted prediction model revealed a 2.1-fold increase in cancer detection rates when the model is used in combination with FOBT.[35] Furthermore, a study using CellMax (CMx®) platform to detect and isolate circulating tumor cells in peripheral blood samples resulted in a sensitivity and specificity of 80%.[36] Table 1 highlights studies focusing on screening.

DIAGNOSIS

A machine learning algorithm can be trained to identify or differentiate polyps in real time in the field of endoscopy. Techniques for analyzing non-magnified endoscopic images and techniques for cellular imaging at a microscopic level have both been investigated (i.e., optical biopsy). The theory behind these methods is that they will improve polyp detection rates, reduce missed adenomas, and thus lower the risk of CRC. However, the increase in polyp detection rates will lead to an increase in financial burdens on health systems, specifically histopathological departments involved in the analysis of resected tissue. Current research initiatives are geared towards building a computer assisted diagnostic algorithm capable of reliably detecting polyps while also characterizing them as hyperplastic or adenomatous during colonoscopy.[37]

The Preservation and Incorporation of Valuable endoscopic Innovations (PIVI) an American Society of Gastrointestinal Endoscopy program set a threshold of negative predictive value (NPV) > 90% for the development of new endoscopic technologies, such as the optical diagnosis of small colorectal polyps[38].

Many AI applications have been developed to assist endoscopist with the aim of adopting a “diagnose and leave” strategy for hyperplastic polyps and a “resect and discard” strategy for diminutive adenomas.[39]. In one study a system was designed to predict the histology of colorectal polyps (adenomatous vs non-adenomatous) by analyzing linked color imaging demonstrated an 83.3% sensitivity, 70.1% specificity, 82.6% positive predictive value (PPV), 71.2% NPV and an accuracy of 78.4% when compared to expert endoscopists.[40]

Magnification Endoscopy with Narrow-Band Imaging (NBI), Endocytoscopy, Magnifying Chromendo-scopy, Confocal Laser Endomicroscopy, Laser-Induced Fluorescence Spectroscopy, Autofluorescence Endoscopy, and White Light Endoscopy are example of advanced endoscopic techniques currently used to aid in the detection and diagnosis of polyps.

Magnification Endoscopy with NBI is a imaging system that allows observation of mucosal surfaces and microvascular patterns.[41] It improves the diagnostic accuracy of benign from premalignant lesions by evaluating depth of submucosal lesions.[42-44] Gross et al.[45] developed a computer-assisted model for polyp classification by analyzing 9 vessel features, including perimeter and brightness from patients who underwent magnifying endoscopy with NBI. The model had a higher sensitivity (95% vs 86%), specificity (90.3% vs 87.8%) and accuracy (93.1% vs 86.8%) when compared to novice endoscopists; however, they are comparable to those of experienced endoscopists (sensitivity, specificity, and accuracy of 93.4%, 91.8% and 92.7%, respectively).

In addition, Chen et al.[46] used magnifying NBI images with 284 diminutive colorectal polyps extracted to create a deep learning model to classify diminutive colorectal polyps When compared to expert endoscopists, the algorithm was able to distinguish between neoplastic and hyperplastic lesions in less time (0.45 vs 1.54 s). It had a sensitivity, specificity, accuracy, PPV, and NPV of 96.3%, 78.1%, 90.1%, 89.6%, and 91.5% respectively.

Endocytoscopy is an endoscopic imaging modality, that allows in vivo microscopic imaging and real-time diagnosis of cellular structures at high magnifications (400× magnification power in endoscope-based to 1400× magnification in probe-based endocytoscopy) during colonoscopy.[47] A computer-aided algorithm was designed to histologically differentiate colorectal lesions in vivo using endocytoscopy.[48] Initially, this model used nuclear features (area, standard deviation of area, circularity, circularity of the 20 largest nuclei, shortest and longest diameter) after nuclear segmentation from the endocytoscopic images with a 92% sensitivity and 89.2% accuracy in establishing a histological diagnosis. This model was later improved by extracting features from texture analysis and utilizing SVM to classify benign, adenomatous lesions or invasive carcinoma.[49,50]. Another model looked at the role of a computer-aided endocytoscopy system in the diagnosis of invasive colorectal carcinoma, and found that it had 89.4% sensitivity, 98.9% specificity, 98.8% positive predictive value, 90.1 percent
**Table 1 Overview of screening studies**

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Objective</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al [19], 2019</td>
<td>Effect of computer aided detection deep learning models on polyps and adenoma detection rates</td>
<td>Increase in adenoma detection rates [29.1% vs 20.3%, ( P &lt; 0.001 )] and mean number of identified adenomas per patient [0.53 vs 0.31, ( P &lt; 0.001 )]; More hyperplastic adenomas (114 vs 52, ( P &lt; 0.001 )) and diminutive polyps (185 vs 102, ( P &lt; 0.001 )) identified</td>
</tr>
<tr>
<td>Nazarian et al [20], 2021</td>
<td>Detection rates of polyp and adenoma with AI vs without AI</td>
<td>Increase in both polyp detection rates (odds ratio [OR] 1.75, 95%CI 1.56-1.96; ( P &lt; 0.001 )) as well as adenoma detection rates (OR 1.53, 95%CI 1.32-1.77; ( P &lt; 0.001 ))</td>
</tr>
<tr>
<td>Johnson et al [23], 2008; Pickhardt et al [24], 2005</td>
<td>Degree to which CTC is effective in detecting asymptomatic colorectal lesions</td>
<td>Reported identification of 90% of patients with asymptomatic adenomas or cancers (≥10 mm in diameter) using CT colonography</td>
</tr>
<tr>
<td>Grosu et al [25], 2021</td>
<td>Development of machine learning method differentiating between benign and precancerous lesions in average risk asymptomatic patients using CTC</td>
<td>Sensitivity of 82%, specificity of 85% and AUC of 0.91</td>
</tr>
<tr>
<td>Song et al [26], 2015</td>
<td>Development of virtual pathological model to assess the suitability of using image high-order differentiations to distinguish colorectal lesions</td>
<td>Improvement of ROC curve (AUC) from 0.74 to 0.85</td>
</tr>
<tr>
<td>Blanes-Vidal et al [30], 2019</td>
<td>Algorithms developed to match CE and colonoscopy-identified polyps based on their estimated size, morphology and location as well as utilizing deep convolutional neural networks for automatic colorectal polyp detection</td>
<td>Localization resulted in high sensitivity (97.1%), specificity (93.3%), and accuracy (96.4%) for identifying polyps when compared to the manual process of polyp detection</td>
</tr>
<tr>
<td>Kinar et al [35], 2017</td>
<td>AI-assisted prediction model (MeScore®, Calgary, Alberta, Canada) was designed to identify people at high risk for CRC</td>
<td>Revealed a 2.1-fold increase in cancer detection rates when the model is used in combination with FOBT</td>
</tr>
<tr>
<td>Gupta et al [36], 2019</td>
<td>Using CellMax (CMx®) platform to detect and isolate circulating tumor cells in peripheral blood samples</td>
<td>A sensitivity and specificity of 80%</td>
</tr>
</tbody>
</table>

AI: Artificial intelligence; AUC: Area under the curve; CTC: Computed tomographic colonography; CT: Computed tomography; CE: Capsule endoscopy; ROC: Receiver operating characteristic.

negative predictive value, and 94.1 percent accuracy[51].

Magnifying Chromoendoscopy is a technique that uses dye to inspect and analyze the pit patterns of the polyp surfaces resulting in high diagnostic performance (97.8% sensitivity, 91.4% specificity and 97.1% accuracy) when performed by expert endoscopists[52]. Takemura et al [53] created a software model to automatically quantify and classify pit patterns. They used texture and quantitative analysis (area, perimeter, and circularity) to classify pit patterns. Using this model type I and II pit patterns were in complete agreement with the endoscopic diagnosis on discriminant analysis. Type III was found in 29 of the 30 cases (96.7%), while type IV was found in one. Type IV pit pattern was found in 29 of the 30 cases (96.7%). The computerized recognition system’s overall accuracy was 132 out of 134 (98.5%).

Confocal Laser Endomicroscopy is a microscopic imaging modality that allows in vivo examination of cellular and subcellular structures at 1000× magnification power [54]. Andrét al [55] used an automated polyp characterization system to distinguish between benign and malignant lesions using the k-nearest neighbor classification with an accuracy of 89.6%. A neural network analysis algorithm had an accuracy of 84.5% in differentiating advanced colorectal adenocarcinomas from normal mucosa [56]. Algorithms using Confocal Laser Endomicroscopy are yet to be validated in randomized clinical trials.

Autofluorescence imaging endoscope characterizes colorectal polyps by analyzing different color emissions of tissue after exposure to a light source. It has shown promising results in differentiating non-neoplastic from neoplastic lesions during colonoscopy [57,58].

White light endoscopy and laser-induced fluorescence spectroscopy technologies have been tested as potential models to discriminate between neoplastic and non-neoplastic lesions with results that were inferior to NBI or chromoendoscopy with or without magnification [59,60]. Table 2 summarized relevant diagnostic research.

**TREATMENT SELECTION, TREATMENT RESPONSE, TOXICITY, AND PROGNOSIS**

Colorectal cancer is a heterogenic disease with numerous epigenetic and microenvironment alterations that affects drug response, aggressiveness, and prognosis[61,62]. The shift to a more personalized and tailored treatment tactic considering the various alternations is evolving to improve disease outcomes [63].
### Table 2 Overview of diagnosis studies

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Objective</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min et al[40], 2019</td>
<td>System designed to predict the histology of colorectal polyps by analyzing linked color imaging</td>
<td>83.3% sensitivity, 70.1% specificity, 82.6% PPV, 71.2% NPV and an accuracy of 78.4% when compared to expert endoscopists</td>
</tr>
<tr>
<td>Gross et al[45], 2011</td>
<td>Development of computer-assisted model for polyp classification by analyzing 9 vessel features, from patients who underwent magnifying endoscopy with NBI</td>
<td>Higher sensitivity (95% vs 86%), specificity (90.3% vs 87.8%) and accuracy (93.1% vs 86.8%) when compared to novice endoscopists but comparable to those of expert endoscopists (sensitivity, specificity, and accuracy of 93.4%, 91.8% and 92.7%, respectively)</td>
</tr>
<tr>
<td>Chen et al[46], 2018</td>
<td>Designed a deep learning model to classify diminutive colorectal polyps using magnifying NBI images with 284 diminutive colorectal polyps extracted</td>
<td>Able to distinguish between neoplastic and hyperplastic lesions in a shorter period compared to expert endoscopists (0.45 vs 1.54 seconds) and had a sensitivity, specificity, accuracy, PPV, and NPV of 96.3%, 78.1%, 90.1%, 89.6% and 91.5% respectively</td>
</tr>
<tr>
<td>Mori et al[48], 2015</td>
<td>Computer-aided algorithm designed to histologically differentiate colorectal lesions in vivo using endocytoscopy</td>
<td>92% sensitivity and 89.2% accuracy in establishing a histological diagnosis.</td>
</tr>
<tr>
<td>Takeda et al [51], 2017</td>
<td>Model investigated the role of a computer-aided endocytoscopy system on the diagnosis of invasive colorectal carcinoma</td>
<td>89.4% sensitivity, 98.9% specificity, 98.8% PPV, 90.1% NPV and 94.1% accuracy</td>
</tr>
<tr>
<td>Takemura et al[53], 2010</td>
<td>Software model to automatically quantify and classify pit patterns. Used texture and quantitative analysis to classify pit patterns</td>
<td>Type I and II pit patterns were in complete agreement with the endoscopic diagnosis on discriminant analysis. Type III was diagnosed in 29 of 30 cases (96.7%) and type IV was diagnosed in one case. Twenty-nine of 30 cases (96.7%) were diagnosed as type IV pit pattern. The overall accuracy of the computerized recognition system was 132 of 134 (98.5%)</td>
</tr>
<tr>
<td>André et al[55], 2012</td>
<td>Automated polyp characterization system to distinguish between benign and malignant lesions using the k-nearest neighbor classification</td>
<td>Accuracy of 89.6%</td>
</tr>
<tr>
<td>Ştefănescu et al[56], 2016</td>
<td>A neural network analysis algorithm differentiating advanced colorectal adenocarcinomas from the normal mucosa</td>
<td>Accuracy of 84.5%</td>
</tr>
</tbody>
</table>

PPV: Positive predictive value; NPV: Negative predictive value.

### Treatment selection

AI is being integrated in treatment selection to provide a true individualized treatment strategy. A MATCH system was developed to integrate clinical and genetic sequence data using data from hospitals, pharmaceutical laboratories, and research centers. The MATCH system aided in correlating between medical features and genetic data, giving the oncologist the opportunity to understand patient’s individual situation[64].

Machine learning techniques are also being used to predict protein-protein interactions of a potential therapeutic target protein (S100A9) with different drugs[65]. Several other models are being developed to identify molecular biomarkers and targets by integrating transcriptomics, proteomics data, and RNA-sequencing data[66,67].

### Treatment response

Chemotherapy, neoadjuvant chemoradiotherapy (nCRT) and other approaches are treatment options for CRC. Studies have applied AI technology to CRC treatment to help clinicians choose the appropriate treatment option and improve efficacy and limit potential toxicities.

In a study based on an unsupervised machine learning algorithm comparing pharmacological response relationships between cancer therapies, distinct intrinsic subpopulation sensitivity to one drug but resistance to others was identified. They also identified genetic alterations that could be used as biomarkers for those subpopulations[68].

In another study, artificial neural network K-nearest neighbors, support vector machine, naïve Bayesian classifier, mixed logistic regression models were used to predict response demonstrated an accuracy of 0.88, AUC of 0.86 and sensitivity of 0.94[69].

Ferrari et al[70] used AI models to assess response to therapy in locally advanced rectal cancer. The AI model was able to identify patients who will have complete response at the end of the treatment and those who will not respond to therapy at an early stage of the treatment with an AUC of 0.83.

Shayesteh et al[71] used MRI based ensemble learning methods to predict the response to nCRT with AUC of 95% and accuracy of 90%.

Other algorithms to identify pathological complete responders (CR) and non-responders (NR) patients after neoadjuvant chemoradiotherapy (CRT) in locally advanced rectal cancer showed an AUC
of 0.86 and 0.83 for pathological CRs and NRs respectively by analyzing textural features of T2-weighted magnetic resonance images [70]. Shi et al. [72] created a model to predict the neoadjuvant CRT response by using pre-treatment and early-treatment MRI imaging. They reported that using deep learning achieved a higher accuracy of prediction.

**Toxicity**

Oyaga-Iriarte et al. [73] used algorithms in metastatic CRC patients to predict Irinotecan toxicity with an accuracy of 76%, 75%, and 91% for predicting leukopenia, neutropenia, and diarrhea respectively. Abraham et al. [74] used machine learning to predict the efficacy of bevacizumab combined with oxaliplatin based chemotherapies in patients with metastatic colorectal cancers.

AI technology is also being incorporated in drug research. Drug delivery models using nanoparticles are being developed [75, 76]. Cruz et al. [77] created a model using molecular and nuclear magnetic resonance to detect the half-maximal inhibitory concentration of a drug against HCT116 cell line with predicted accuracy of over 63% for both training and test sets.

**Prognosis**

Traditional mathematical and statistical analysis does not provide accurate predictions on patient’s progress. However, AI can process and analyze many features based on previous data to potentially predict prognosis.

Weiser et al. [78], developed a nomogram to predict recurrence of CRC after curative resection to identify patients who may benefit from adjuvant therapy and early follow-up.

In addition, long term prediction models using independent prognostic factors such as tumor size, high mitotic count, non-gastric location, and sex are established and accurately predict patients who may be cured by surgery alone [79].

The prognosis in CRC is highly dependent on pathology. Kather et al. [80] used CNN to automatically extract prognostic factors from HE-stained CRC tissues. They used 420 digitalized HE-stained samples to predict the 5-year survival with an AUC of 0.69 consistent with “expect level” accuracy.

Sailer et al. [81] compared ten data mining algorithm’s to predict the 5-year survival based on seven attributes and reported an accuracy of 67.7% compared to clinical judgment of 59%. Table 3 summarizes relevant treatment, toxicity, and prognosis studies.

### LIMITATIONS

Artificial intelligence and deep learning algorithms assist physicians in detecting and diagnosing CRC. They are also used to develop and identify treatment strategies to personalize CRC treatment. Until now, AI tools have been able to detect and diagnose CRC in a manner that is comparable to, if not superior to, that of humans (Figure 2).

Despite the significant advance in AI applications, AI-based technologies have several limitations. Machine training is a complex task and requires integrating the technology into clinical practice to
Figure 2  Stages in designing and implementing an artificial intelligence model.

provide high quality large volume training data to train the AI systems and obtain the best results. This process requires robust computational infrastructure.

The variability between patients’ clinical presentation could lead to a deviation from the training model environment which could result in the unpredictable performance of an algorithm\cite{82}. Furthermore, the input and output data of an algorithm is known, there is limited information on the exact working and process in-between, frequently referred to as the “black box” problem in machine learning. As a result of this limited visibility, factors used by a deep learning algorithm to reach a particular decision could be missed potentially leading to significant confounders in output data\cite{82}.

Additionally, there is a lack of evidence-based standards in AI development. The data used to train algorithms vary in size, number, and quality. This results in inconsistencies in validating machine learning systems deterring their implementation on a wide scale clinical setting. Limited research on the
application of AI in CRC treatment is currently present. Most of the existing studies assessed AI algorithm’s ability to predict response after nCRT and chemotherapy. However, they have small sample sizes and therefore lack generalization\[83\]. In addition, current AI algorithms linking clinical features to prognostic status are promising. However, there is a significant difference between sensitivities, specificities, and accuracies of different AI applications.

Machine learning systems can unintentionally exacerbate health disparities by magnifying existing biases used in their training datasets\[84\].

Machine learning and artificial intelligence is evolving, though the medical community remains highly optimistic about the future of AI, wide scale randomized clinical trials are needed to evaluate and validate AI algorithms prior to wide scale clinical implementation. Additionally, these systems should provide a high-quality standard with robust ethical and legal frameworks prior to integration in health systems.

**FUTURE DIRECTIVES**

With the rapid expansion in AI research and technology we believe that AI algorithms will improve and personalize patient care.

Initially, AI algorithms integrate clinical data such as age, health status, disease history and other comorbidities to stratify patients. Though the current gold standard for CRC screening and diagnosis is endoscopy and pathological biopsy\[12\], it carries a significant risk in a subset of patients. We believe that future research directives will focus on less invasive technologies in certain patient groups for diagnosis instead on colonoscopy. Any model must maintain or even exceed the diagnostic accuracy offered by conventional diagnostic modalities. Furthermore, incorporating AI in screen colonoscopy may improve the diagnosis of precancerous lesions.

Moreover, AI technologies could assist in a establishing a more accurate staging system that incorporates not only the classical TNM stages but also proteomics, metabolomics, and genetic data to account for the heterogeneous presentation of CRC. This algorithm would potentially identify patients who would benefit from neoadjuvant therapy.

As more datasets are made available, a sufficiently large dataset could support the prediction of the prognosis of AI technology. This can help identify factors with the greatest impact on prognosis and establish future prognostic and intervention research.

**CONCLUSION**

Artificial intelligence and deep learning are becoming an integral part of modern-day medicine. Though the research advances in the field is an exciting new venture, it currently remains in the infant stage. Colorectal cancer screening, diagnosis and treatment will be distinctly enhanced by the incorporation of artificial intelligence technologies. AI has showed promise in therapeutic recommendations and prediction of treatment toxicity and responses this will hopefully result in a better and more personalized treatments for those in need.

**FOOTNOTES**

**Author contributions:** Awidi M and Bagga A contributed equally to the work; All authors have read and approve the final manuscript.

**Conflict-of-interest statement:** There is no conflict of interest associated with any of the authors who contributed their efforts to this manuscript.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/Licenses/by-nc/4.0/

**Country/Territory of origin:** United States

**ORCID number:** Muhammad Awidi 0000-0002-4205-8417; Arindam Bagga 0000-0003-3096-177X.

**S-Editor:** Liu JH

**L-Editor:** A
REFERENCES

1 Cancer Facts & Figures 2021. Atlanta, Ga: American Cancer Society, 2021
5 Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms: Cambridge university press, 2014 [DOI: 10.1017/cbo97811070298019]
21 Nazarian S, Glover B, Ashrafian H, Darzi A, Teare J. Diagnostic Accuracy of Artificial Intelligence and Computer-Aided
Gross S

Johnson CD


Gross S


Gross S

Gross S


Gross S

Gross S

Gross S

Gross S

Gross S

Gross S

Gross S

Gross S


Gross S

Gross S

Gross S

Gross S

Gross S

Gross S

Gross S

Gross S

Gross S

Gross S


Cruz S, Gomes SE, Borralho PM, Rodrigues CMP, Gaudêncio SP, Pereira F. In Silico HCT116 Human Colon Cancer Cell-Based Models En Route to the Discovery of Lead-Like Anticancer Drugs. *Biomolecules* 2018; 8 [PMID: 30018273 DOI: 10.3390/biom8030056]


