

75143\_Auto\_Edited.docx



**Name of Journal:** *Artificial Intelligence in Gastrointestinal Endoscopy*

**Manuscript NO:** 75143

**Manuscript Type:** MINIREVIEWS

**Artificial intelligence and machine learning in colorectal cancer; A review**

Muhammad Awidi, Arindam Bagga

**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 to the development of computer-assisted diagnostic algorithms 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.



## **INTRODUCTION**

In the United States, the third leading cause of cancer-related deaths is colorectal cancer(CRC) <sup>[1]</sup>. 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 <sup>[2]</sup>. CRC is a diverse group of diseases with differences in epidemiology, histology, genomics, and host immune responses <sup>[3, 4]</sup>. 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 <sup>[5]</sup>. 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 (DL) which imitates higher level human data processing by using multi-layered neural networks for extractions and self-training algorithms <sup>[6]</sup>. [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 <sup>[7, 8]</sup>.

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 <sup>[9]</sup>. CRC patients are diagnosed at advanced stages of the disease in 60–70% of cases <sup>[9]</sup>.



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 chorography, 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* 58,  $p = 0.075$ ). 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 (CWC) 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 [89].

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;  $p < .001$ ) as well as adenoma detection rates (OR 1.53, 95%CI 1.32-1.77;  $p < .001$ ) 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 ( $\geq 10$  mm in diameter) using CT colonography [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



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 (RDW) 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 Chromoendoscopy, Confocal Laser Endomicroscopy, Laser-Induced Fluorescence Spectroscopy (LIFS), Autofluorescence Endoscopy (AFE), and White Light



Endoscopy (WLE) are example of advanced endoscopic techniques currently used to aid in the detection and diagnosis of polyps.

Magnification Endoscopy with Narrow-Band Imaging (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 (400x magnification power in endoscope-based to 1400x 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 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 1000x magnification power [54]. Andre *et 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 (AFI) 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].

### **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 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 [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 [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 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.