

Artificial intelligence and machine learning technologies in ulcerative colitis

Chiraag Kulkarni, Derek Liu, Touran Fardeen, Eliza Rose Dickson, Hyunsu Jang, Sidhartha R. Sinha* and John Gubatan* 

Ther Adv Gastroenterol

2024, Vol. 17: 1–22

DOI: 10.1177/
17562848241272001

© The Author(s), 2024.
Article reuse guidelines:
[sagepub.com/journals-](https://sagepub.com/journals-permissions)
permissions

Abstract: Interest in artificial intelligence (AI) applications for ulcerative colitis (UC) has grown tremendously in recent years. In the past 5 years, there have been over 80 studies focused on machine learning (ML) tools to address a wide range of clinical problems in UC, including diagnosis, prognosis, identification of new UC biomarkers, monitoring of disease activity, and prediction of complications. AI classifiers such as random forest, support vector machines, neural networks, and logistic regression models have been used to model UC clinical outcomes using molecular (transcriptomic) and clinical (electronic health record and laboratory) datasets with relatively high performance (accuracy, sensitivity, and specificity). Application of ML algorithms such as computer vision, guided image filtering, and convolutional neural networks have also been utilized to analyze large and high-dimensional imaging datasets such as endoscopic, histologic, and radiological images for UC diagnosis and prediction of complications (post-surgical complications, colorectal cancer). Incorporation of these ML tools to guide and optimize UC clinical practice is promising but will require large, high-quality validation studies that overcome the risk of bias as well as consider cost-effectiveness compared to standard of care.

Plain language summary

Artificial intelligence in ulcerative colitis

Ulcerative colitis (UC) is a chronic inflammatory disorder of the colon. The clinical care of patients with UC and research efforts to better understand the disease has inevitably produced a significant quantity of diverse and complex datasets ranging from electronic health records, laboratory values, images (endoscopy, radiology, histology) to gene expression. The size and complexity of datasets derived from UC poses a significant challenge to accurately and effectively predict clinically meaningful endpoints in order to ultimately improve UC outcomes. Artificial intelligence through the application of machine learning tools has the potential to improve the analysis of large, complex, high-dimensional datasets and reveal novel, deeper insights compared to traditional analytical tools. Here, we provide an updated and comprehensive summary of AI applications in UC.

Keywords: artificial intelligence, biomarkers, machine learning, outcomes, prediction, ulcerative colitis

Received: 29 November 2023; revised manuscript accepted: 17 June 2024.

Introduction

Ulcerative colitis (UC) is a chronic inflammatory disorder of the gut without a medical cure that affects nearly 1 million Americans.¹ Inflammatory bowel disease (IBD) is characterized by

intestinal dysbiosis and immune dysregulation. Environmental factors, particularly diet, are thought to play a key role in disease pathogenesis, particularly via impact on the gut microbiome.^{2,3}

Correspondence to:

Sidhartha R. Sinha
John Gubatan

Division of
Gastroenterology and
Hepatology, Stanford
University School of
Medicine, 300 Pasteur
Drive, M211, Stanford, CA
94305, USA
sidsinha@stanford.edu
jgubatan@stanford.edu

Chiraag Kulkarni
Derek Liu
Touran Fardeen
Eliza Rose Dickson
Hyunsu Jang

Division of
Gastroenterology and
Hepatology, Stanford
University, Stanford, CA,
USA

*These authors share
senior co-authorship

Current therapeutics control IBD via broad immunosuppression but do not address the underlying intestinal dysbiosis. Further, despite our best therapies, most patients do not achieve long-term remission highlighting the need for improved disease monitoring and personalized therapeutic interventions.⁴ There is a growing interest in utilizing deep, multi-omics phenotyping in IBD including whole exome sequencing, transcriptomics, proteomics, and metagenomics of the microbiota. Additionally, there is a rapid expansion in the growth of clinical images from endoscopy and pathology samples. The resultant rapid expansion of data has led to interest in the application of artificial intelligence (AI) to IBD.

AI is a multidisciplinary field that seeks to apply computer software to mimic human intelligence. Machine learning (ML) is a subset of AI that uses statistical methods to recognize patterns from datasets and can be done through supervised and unsupervised methods. Supervised learning relies on labeled input to give accurate classification or prediction of the outcome of interest. Examples of supervised learning include regression, K-nearest neighbor, and random forest (RF). Unsupervised learning, in contrast, does not require labeled input and is used to reduce dimensionality and allow for clustering.^{5,6} Deep learning (DL) utilizes artificial neural networks (ANN), which mimic brain logic structures, to perform complex learning tasks by utilizing layers of representation and subsequent transformation to highlight aspects of the input which improves task performance. Examples of DL include virtual assistants and image recognition.⁶

Over the last decade, there has been increased application of AI in IBD.⁷ In particular, computer vision in endoscopy in UC has been a key area of growth.⁸⁻¹¹ The purpose of this review is to provide an updated and comprehensive evaluation of recent advances in AI in UC, with a particular focus on the prediction and diagnosis of new UC, prediction of response to therapy, disease monitoring, and identification of disease complications. We also review challenges to the translation of these novel technologies into the clinic and discuss future directions.

Literature search strategy

We performed a literature review using PubMed (MEDLINE) from inception until July 30, 2023,

of all studies applying AI in UC. Our search strategy consisted of the following combinations: (((((((ulcerative colitis [Title]))) AND (artificial intelligence [Title])) OR (computer-assisted [Title])) OR (computer-aided [Title])) OR (machine learning [Title])) OR (deep learning [Title]). We included studies that used AI in the (1) prediction and diagnosis of UC, (2) prediction of response to therapy in UC, (3) monitoring disease activity in patients with UC, and (4) prediction of complications of UC. We excluded reviews, studies with non-human subjects (animal models), or studies that did not provide objective measures of the efficacy of AI applications (e.g. area under the curve (AUC), sensitivity, specificity, etc.).

Results

Our search strategy yielded 97 studies that applied AI to UC, of which 61 studies met our inclusion criteria. In total, 54 (88.5%) of studies were published in the last 5 years. Eighteen studies focused on the prediction and diagnosis of new UC, 11 studies predicted response to therapy, 15 evaluated disease monitoring, and 14 focused on prediction of UC complications. The AI methods utilized include: linear regression (LR), lasso regression, gradient boosted machine (GBM), principal component analysis (PCA), RF, linear discriminant analysis (LDA), support vector machines (SVM), segment anything model (SAM), ANN.

Prediction and diagnosis of new UC

Identification of biologic pathways in UC

What is already known? While previous research has identified some common genetic, environmental, and microbial risk factors for UC, the associations are neither strong enough nor consistent enough to be clinically useful.¹²⁻¹⁶ The use of AI has enormous potential for assessing risk and identifying biologic pathways enriched in UC compared to the general population.

What do current studies show? Table 1 summarizes studies that applied AI to the diagnosis of new UC. Four studies utilized omics data, including genetic/genomic ($n=3$) and transcriptomic ($n=1$) data sets. While this is a growing area of research interest, only a few studies have specifically focused on prediction of UC from a healthy population.

Table 1. Prediction and diagnosis of new UC.

Author	Study design	Brief description	Main findings
Tang et al., 2023 ¹⁷	Cross-sectional	Used a combination of 3 machine learning algorithms (LASSO, SVM-RFE, RF) trained on microarrays from colon biopsy samples of 298 UC and 55 healthy control patients (validated using samples from 87 active UC and 21 Healthy control (HC) patients) identified differentially expressed cuproptosis-related genes and generated a prediction model of UC.	Seven signature genes to build a nomogram for predicting the risk of UC. Demonstrated outstanding prediction performance based on calibration and ROC curves (AUC 0.982).
Zhang et al., 2022 ¹⁸	Cross-sectional	Identified 2 useful genes for UC diagnosis (<i>OLFM4</i> and <i>C4BPB</i>) using 6 ML methods (SVM, LASSO, RF, GBM, PCA, and neural network) which were trained using 6 microarrays from 201 UC and 106 healthy patients and validated using 4 microarrays (from 186 UC and 33 healthy subjects).	<i>OLFM4</i> had good performance with average AUC of 0.865. Both <i>OLFM4</i> and <i>C4BPB</i> significantly correlated to macrophages (M1 and M2), mast cells (activated and resting), monocytes, and natural killer (NK) cells activated
Li et al., 2020 ¹⁹	Cross-sectional	Used a random forest algorithm (trained on a set of mucosal transcriptomic profiles from rectal biopsies of 206 UC and 20 healthy patients and validated with an independent set from 53 UC and 21 healthy rectal biopsies) to identify 1 downregulated and 29 upregulated DEG's with highest contribution to UC occurrence. ANN calculated DEG weights to UC.	Prediction results agreed with that of an independent data set (AUC=0.951, area under the precision recall curve [AUPRC] =0.975).
Han et al., 2021 ²⁰	Cross-sectional	Applied SVM combined with RFE to construct a disease classifier from 41 genes for the disease diagnosis of UC patients. Trained on a dataset of 108 patients (97 UC, 11 HC), and validated using data from 70 samples (46 UC, 24 HC).	The SVM classifier combined with RFE applied to 41 genes had the highest accuracy of 0.965. In the validation datasets, AUC was 0.832
Kraszewski et al., 2021 ²¹	Prospective cohort	Used RF to create an IBD prediction model based on routinely performed blood, urine, and fecal tests and compare it to diagnosis based on CRP alone. Input data (702 medical records of 372 IBD patients: 319 records from 180 UC patients and 383 records from 192 CD [Crohn's disease] patients) was divided into training and test sets with a 7:3 ratio and compared against a control group of 315 records from 271 patients with noninflammatory and non-malignant bowel diseases.	A majority of robust classifiers from the RF ensemble obtained a mean average precision of 97% for CD and 91% for UC. In comparison, diagnosis based on only CRP demonstrated average precision of only 81% for CD and 61% for UC.
Jiang et al., 2022 ⁸	Cross-sectional	Assessed diagnostic ability of low-dose computed tomography enterography (CTE) based on an improved GIF algorithm from 60 patients suspected to have IBD and compared it to 60 patients who underwent routine computed tomography (CT) examination. Comprehensive diagnosis was used as the standard to assess the diagnostic effect.	For UC, low-dose CTE based on improved GIF had a diagnostic accuracy of 98.33%. Compared to routine CT examination, low-dose CTE from improved GIF had superior classification performance
Dhaliwal et al., 2021 ²²	Cross-sectional	Used clustering with similarity network fusion (SNF) on the top RF features to discriminate between UC and colonic-CD independent of a supervised model. And RF classifier was trained on a dataset of baseline clinical, endoscopic, radiologic, and histologic data from 74 pediatric colonic IBD patients (56 UC, 18 colonic-CD) and validated via leave-one-out approach. A new classifier was constructed from the top features and then tested on 15 previously unused patients.	The classifier accurately distinguished UC from colonic-CD in 97% of patients in the training set and 100% of the patients in an independent set.
Lu et al., 2022 ²³	Cross-sectional	Created a logistic regression model based on 5 genes with potentially significant correlation with UC. Applied the logistic regression model to microarray data from colonic epithelial mucosa biopsy samples from 106 UC patients and 21 healthy patients (randomly assigned into training and test data sets) to identify specific diagnostic signatures to create a diagnostic model for UC.	The logistic regression model had an average AUC of 0.8497 in the training set, and AUC of 0.7208 in another independent verification set.

(Continued)

Table 1. (Continued)

Author	Study design	Brief description	Main findings
Khorasani et al., 2020 ²⁴	Cross-sectional	A feature selection algorithm combined with an SVM classifier was used to create a model to distinguish between healthy control and UC patients using 32 genes in colon samples from 5 combined datasets. The training set consisted of 52 samples, and the test dataset had 25 samples.	The model perfectly detected all active cases and had an average precision of 0.62 in the inactive cases.
Wang et al., 2023 ²⁵	Cross-sectional	LASSO regression and SVM-RFE identified 5 genes as promising biomarkers from 2 gene expression data sets consisting of a combined 193 UC and 42 healthy control samples.	All 5 genes identified as essential diagnostic genes for UC demonstrated strong discrimination between UC and healthy specimens (average AUC 0.9562). Of these 5 genes, 2 genes (<i>DUOX2</i> , <i>DMBT1</i>) had significantly higher expression levels in UC samples vs healthy control, and 3 genes (<i>CYP2B7P</i> , <i>PITX2</i> , <i>DEFB1</i>) had significantly lower levels of expression in UC samples.
Dutttagupta et al., 2012 ²⁶	Cross-sectional	SAM identified 31 differentially expressed platelet-derived miRNAs from whole genome maps of circulating miRNAs from peripheral blood mononuclear cells (PBMC), microvesicles, and platelets constructed from blood samples of 20 UC and 20 healthy control patients. Used SVM (trained on a cohort of randomly selected 18 case-control subjects, tested on 2 subjects from each enrollment category) to evaluate biomarker performance using non-probabilistic binary linear classification.	SVM classifier measurements revealed a predictive score of 92.8% accuracy, 96.2% specificity, and 89.5% sensitivity in distinguishing ulcerative colitis patients from normal individuals.
Chen et al., 2023 ²⁷	Cross-sectional	Used RF and LASSO regression trained on a dataset from 30 UC and 13 healthy samples and (tested on a dataset from 25 UC and 22 healthy samples). LASSO regression identified 7 genes as potential diagnostic markers of UC.	The LASSO regression model had an AUC of >0.9 in the training set, and most of the genes identified as potential UC-related diagnostic markers had an AUC of 0.65 in a validation set.
Sutton et al., 2022 ¹⁰	Cross-sectional	Deep learning CNNs applied via a weakly supervised approach to distinguish UC from other intestinal disorders and grade endoscopic severity. The diagnostic classification model was trained using 2642 pathological endoscopy images and the diagnostic model for endoscopic grading of UC was trained using 851 UC images with Mayo grades 0–3. 20% of both training sets were set aside for test datasets.	DenseNet121 architecture had superior accuracy (87.50%) and AUC (0.90). Grad-CAM improved visual interpretation of the model. In all model architectures, CNN discrimination between UC and non-UC pathologies had area under the receiver operating characteristic (AUROC) > 0.99.
Chierici et al., 2022 ¹¹	Cross-sectional	Applied a prototype DL framework based on ResNet architectures merged by ensemble learning to 14,226 three-channel RGB endoscopic images of 11,404 IBD (4388 UC, 5949 CD, and 1067 other IBD) and 2822 healthy patients to identify disease patterns and distinguish endoscopic images of UC and CD from non-IBD samples. 20% of the images were reserved for validation. Of the remaining sample size, 90% were used as training dataset and 10% for testing.	In the test dataset, DL model demonstrated strong predictive ability at distinguishing IBD from healthy samples (Matthews correlation coefficient [MCC]=0.940) and specifically distinguishing UC vs healthy patients (MCC=0.931). It had lower but relatively good predictive ability at distinguishing UC vs CD (MCC=0.688).
Gottlieb et al., 2021 ⁹	Prospective cohort	Used a deep learning algorithm (RNN) trained on 795 full-length endoscopy videos from 249 patients, cleaned and abnormality features extracted via CNN, to predict UC severity based on endoscopic Mayo score and UCEIS scores. The full video feature data set was randomly split at the patient level into a training set (80%) and hold-out test set (20%).	The RNN had excellent agreement with human readers with a QWK of 0.844 [95% CI, 0.787–0.901] for Mayo score and 0.855 [95% CI, 0.80–0.91] for UCEIS.

(Continued)

Table 1. (Continued)

Author	Study design	Brief description	Main findings
Mossotto et al., 2017 ²⁸	Cross-sectional	Constructed 3 models (endoscopic data only, histological data only, and combined endoscopic/histological data) by applying supervised and unsupervised ML techniques using data collected at initial diagnosis from 287 pediatric IBD patients (178 CD, 80 UC, 29 IBD). Supervised linear SVM was used to classify CD vs UC samples based on clinical data from 210 patients (143 CD, 67 UC), divided into subsets for discovery (72 CD, 34 UC), training and testing (71 CD, 33 UC), and final reclassification (29 IBD)	All models had fair classification performance for UC with the greatest accuracy from the combined model (82.7%, AUC 0.87). However, in the validation cohort classification performance for CD remained high but only 65% of UC cases were correctly labeled

ANN, artificial neural network; AUC, area under the curve; CI, confidence interval; CNN, convolutional neural networks; CRP, C-reactive protein; DEGs, differentially expressed genes; DL, deep learning; GBM, gradient boosted machine; GIF, guided image filtering; IBD, inflammatory bowel disease; LASSO, least absolute shrinkage and selection operator; MCC, Matthews correlation coefficient; PCA, principal component analysis; QWK, quadratic weighted kappa; RGB, red, green, blue; RF, random forest; RNN, recurrent neural network; SAM, segment anything model; SVM-RFE, support vector machines recursive feature elimination; UC, ulcerative colitis; UCEIS, UC endoscopic index of severity.

In a cross-sectional study of colon biopsy samples from 298 active UC and 76 healthy control patients by Tang et al., a combination of three ML algorithms—including least absolute shrinkage and selection operator (LASSO), SVM recursive feature elimination (SVM-RFE), and RF—identified seven differentially expressed cell death-related genes (average AUC 0.859) to build a prediction model of UC diagnosis. The resulting nomogram had good predictive performance with an AUC of 0.982 in the validation set.¹⁷ In a separate cross-sectional study of 387 UC and 139 healthy patients by Zhang et al., 2 useful genes (*OLFM4* and *C4BPB*) were identified using a combination of 6 ML methods including SVM, LASSO, RF, GBM, PCA, and ANN. *OLFM4* and *C4BPB* were found to be of diagnostic values as determined by an average AUC of 0.865 based on their performance in training, test, and independent validation sets. Notably, both genes were significantly correlated with M1 macrophages, M2 macrophages, activated mast cells, resting mast cells, monocytes, and activated natural killer cells ($p < 0.05$).¹⁸ Another cross-sectional analysis of 259 UC and 41 healthy patients by Li et al. utilized RF to identify differentially expressed genes (DEGs) with highest contribution to UC occurrence from sets of mucosal transcriptomic profiles from rectal biopsies and used an artificial neural net to calculate DEG weights to UC. The algorithms demonstrated excellent prediction performance

of AUC 0.9506, which also agreed with that of an independent data set.¹⁹ In a separate cross-sectional study of 178 patients (143 UC, 35 healthy control), Han et al. constructed a disease classifier from 41 genes using SVM-RFE for diagnosis of UC. The model demonstrated high accuracy of 96.5% and performed excellently in training and validation sets with an AUC of 0.999 and average AUC of 0.832, respectively.²⁰

What could AI add in the future? The prevalence of UC is increasing, and despite this, our understanding of the pathophysiology of UC is still limited. Bench scientists have increasingly applied a systems biology approach to study disease pathogenesis, and there has been a resultant explosion in the volume of scientific data. While the current studies have applied traditional ML methods to these datasets, there is an opportunity to apply DL methods to these data to generate novel insights.

AI in diagnosis of UC. Traditionally, evaluation and diagnosis of UC involve comparing clinical symptoms to relevant laboratory data, radiographic imaging, and endoscopic reports via index colonoscopy.²⁹ In recent years, many studies have begun exploring the potential of AI methodologies to enhance prediction and accuracy of UC diagnosis, improve treatment outcomes through early diagnosis, and discovery of novel pathways associated with UC pathogenesis.

Of the 14 total studies, 3 used AI to assist in the analysis of diagnostic labs and radiographic data, 6 involved the use of AI to aid in the discovery of novel biomarkers for diagnosis, and 5 studies utilized AI or computer vision in index colonoscopy.

AI analysis of laboratory, pathology, and radiographic data

What is already known? Current clinical practice for diagnosis of UC relies on a combination of laboratory testing and radiographic data in combination with endoscopic and histological evaluation. Laboratory testing includes assessment of serum inflammatory markers such as leukocyte count and differential, platelet count, and C-reactive protein (CRP) as well as stool tests such as fecal calprotectin or lactoferrin levels which are stronger indications of activation of immune pathways in the gut.³⁰ Radiographic techniques using magnetic resonance imaging, computed tomography, and abdominal ultrasound can be used to rule out small bowel involvement and distinguish UC from other gastrointestinal pathologies.³¹ Despite the best application of current technology, approximately 5%–10% of patients with IBD are initially diagnosed with indeterminate colitis.

What do current studies show? The overarching goal of research in this area is to use AI techniques to create an objective model for evaluation of clinical labs and radiographic data to improve accuracy and precision of UC diagnosis. Of the studies that employed AI for the analysis of labs and radiographic data, data modalities included electronic health records ($n=2$ studies) and imaging datasets ($n=1$ study).

In a prospective cohort study of 702 medical records belonging to 372 IBD patients (180 UC, 192 CD) conducted by Kraszewski et al., an RF algorithm was used to create an UC diagnostic prediction model based on routine blood, urine, and fecal tests compared to diagnosis based on CRP alone. While the RF ensemble achieved a mean average precision of 91% for UC, the comparison to CRP alone does not represent typical clinical practice; no comparison was made to physician diagnosis.²¹ In a separate cross-sectional study of 74 pediatric colonic IBD (56 UC and 18 colonic-CD) patients, Dhaliwal et al. used an RF classifier that accurately distinguished UC from colonic-CD in 97% of patients in the training set,

and 100% of the patients in the validation set of patients when given a combination of baseline clinical, endoscopic, radiologic, and histologic data.²²

Jiang et al. demonstrated diagnostic ability of a guided image filtering (GIF) algorithm in a cross-sectional study of 60 patients with suspected IBD and 60 non-IBD patients undergoing radiologic examination via CT scan. The improved GIF algorithm accurately diagnosed 98.3% of UC cases. Despite the smaller sample size, the performance characteristics are promising and show the capability of AI to enhance diagnostic accuracy when applied to CT images.⁸

What could AI add in the future? The three studies on the application of AI to the diagnosis of UC have significant limitations. One study did not compare their model to physician diagnosis; all studies focused on differentiating patients who clearly had either UC or CD from each other, which is not a problem a practicing gastroenterologist typically faces. Rather, given the expanding repertoire of IBD medications, some of which are specific to UC, applying AI to accurately classify UC or CD among patients initially diagnosed with indeterminate colitis would be more clinically relevant.

AI-assisted discovery of novel biomarkers for diagnosis

What is already known? Leading biomarkers for UC include serum CRP and fecal calprotectin.^{32,33} While both of these biomarkers indicate inflammatory states at a systemic and gastrointestinal level, respectively, elevated levels of these markers are not sufficient to diagnose UC without additional more invasive testing through endoscopic and histological evaluation.³³ There is ongoing and increasing interest in the identification of novel biomarkers that are specific and sensitive enough to have strong clinical relevance to UC diagnosis. The lack of specific diagnostic signatures for UC has been noted as a potential barrier to early detection.³⁴ Of the six studies which focused on using AI to assist in the discovery of novel biomarkers for UC diagnosis, data modalities included genetic/genomics ($n=2$ studies), transcriptomics ($n=3$ studies), and proteomics ($n=1$ study).

What do current studies show? In a cross-sectional study of blood samples from 20 UC

and 20 healthy patients, Duttgupta *et al.* applied a SAM to identify 31 differentially expressed platelet-derived miRNAs from whole genome maps of circulating miRNAs from PBMC, micro-vesicles, and platelets. They then used SVM to evaluate biomarker performance using non-probabilistic binary linear classification, which revealed predictive scores with 92.8% accuracy and specificity and sensitivity of 96.2% and 89.5%, respectively. Candidate biomarkers independently validated by qPCR assays run on pooled patient and control samples and demonstrated 88% success.²⁵

Lu *et al.* created a logistic regression model based on five genes (*REG3A*, *REG1A*, *DEFA6*, *REG1B*, and *DEFA5*) determined to be strongly associated with UC occurrence based on analysis of a microarray of colonic biopsies from 106 UC and 21 healthy patients. The logistic model demonstrated strong performance at predicting UC with average AUC of 0.850, and AUC of 0.721 when evaluated in an independent set of 137 unseen samples.²³ Khorasani *et al.* took a similar approach, using a SVM classifier to distinguish between healthy controls and patients with UC by gene expression, but the model had poor precision for identifying inactive UC.²⁴ In a separate cross-sectional study involving microarray expression data of 193 UC and 42 healthy control patients, Wang *et al.* identified 64 upregulated and 38 downregulated genes then used LASSO regression and SVM-RFE to identify 5 diagnostic genes with strong ability to distinguish UC cases from normal samples. They found UC samples had significantly higher expression levels of *DUOX2* and *DMBT1* (AUC 0.985 and 0.896, respectively) and lower expression of *CYP2B7P*, *PITX2*, and *DEFB1* (AUC 0.966, 0.968, and 0.966, respectively) compared to samples from healthy patients. These genes were also found to be associated with infiltration of regulatory T cells, CD8 T cells, activated and resting memory CD4 T cells, activated natural killer cells, neutrophils, activated and resting mast cells, activated and resting dendritic cells, and M0, M1, and M2 macrophages.²⁵ Using RF to identify 54 feature genes from expression profiles of 55 UC and 35 healthy patients, Chen *et al.* constructed a LASSO regression model to screen for diagnostic markers of UC. The model performed well in the training set but when validated in an external

data set, model performance was not found to be clinically useful (AUC = 0.650).²⁷

What could AI add in the future? Current studies have sought to identify novel biomarkers for the diagnosis of UC. Patients with UC typically do not experience significant diagnostic delay, and there is no meaningful clinical action a gastroenterologist could take even if a high-risk patient was identified prior to developing overt symptoms. Therefore, the current approach has limited clinical utility and is more likely to have an impact by aiding our understanding of disease pathogenesis.

Computer vision in index colonoscopy

What is already known? Endoscopic and histological evaluation via index colonoscopy is the gold standard for confirming UC diagnosis, and it is frequently analyzed in collaboration with clinical symptoms, laboratory, and radiological findings.³⁵ However, endoscopic scoring is inherently subjective despite attempts to create consistent scoring systems, leading to observed high rates of inter- and intra-observer variability and general lack of widespread use among endoscopists.^{10,35} AI techniques such as ML and computer vision are promising in the creation of an objective approach to analyzing endoscopic and histological data for early and accurate diagnosis of UC at index colonoscopy.¹⁰ The studies included in this section have different applications. Some studies focused on automated scoring of disease severity and others focused on distinguishing UC and CD.

For included studies exploring computer vision and ML for evaluation of endoscopic data, data modalities included imaging and endoscopic datasets ($n=4$ studies), combined endoscopic and histological datasets ($n=1$ study), electronic health data ($n=1$ study), metagenomics ($n=1$ study), and metabolomics ($n=1$ study).

What do current studies show? Using a CNN to clean and extract abnormality features, Gottlieb *et al.* trained a recurrent neural network to predict UC severity in a prospective cohort study using 795 full-length endoscopy videos (19.5 million image frames) from 249 patients enrolled in a phase II trial of mirikizumab. The model's predictions agreed strongly with

endoscopic scoring by centralized human readers demonstrated by quadratic weighted kappa score of 0.844 (95% confidence interval (CI): 0.787–0.901) for endoscopic Mayo score and 0.855 (95% CI: 0.80–0.91) for the UC endoscopic index of severity (UCEIS). Notably, the performance metrics met or exceeded those previously published for endoscopic Mayo score and the UCEIS scores.⁹

In one cross-sectional study, Sutton et al. used DL CNNs to discriminate between UC and non-UC pathologies with high accuracy when compared against review by consensus labeled data from a single gastroenterologist and three medical trainees. The initial diagnostic classification model based on 2643 pathological endoscopy images was only able to make predictions of majority class with 72.02% accuracy, compared to the final diagnostic model for grading endoscopic severity of UC, which had prediction accuracy of 87.50%. The final model was based on 851 images from diagnostic colonoscopies with endoscopic Mayo scores of 0–3 and had stronger overall performance with AUC of 0.90.¹⁰ In a separate cross-sectional study, Chierici et al. applied a prototype DL framework based on ResNet architectures merged by ensemble learning to 14,226 three-channel RGB (red, green, blue) endoscopic images of 11,404 IBD (4388 UC, 5949 CD, and 1067 other IBD) and 2822 healthy patients to identify disease patterns and distinguish endoscopic images of UC (Matthews Correlation Coefficient=0.931) from healthy patients.¹¹

In a cross-sectional study of endoscopic and histological data from 287 pediatric patients (178 CD, 80 UC, 29 IBD unclassified) at time of diagnosis, Mossotto et al. identified four new subgroups of disease based on colonic disease using unsupervised PCA and multidimensional scaling. They then applied supervised linear SVM with RFE fivefold cross-validation to construct a model to discriminate UC from CD with 82.7% accuracy (AUC 0.87) based on a combination of histological and endoscopic data when

compared against physicians. While the model performed well overall, this still falls below the requirement needed for clinical application. Notably, this combined model outperformed models that relied on either endoscopic or histological data alone in terms of accuracy (71% and 76.9%, respectively) and AUC (0.78 and 0.82, respectively). However, even the optimized model was able to identify Crohn's disease more precisely versus UC.²⁸

What could AI add in the future? There is considerable inter- and intra-observer variability in endoscopic scoring in UC, and rates of agreement with agreement for the endoscopic Mayo score and UCEIS reported to be as low as 0.58. Computer vision offers a promising avenue for recording disease severity, allowing for standardized scoring between different providers and institutions. It may also serve as an alternative to centralized reading for IBD clinical trials, potentially allowing for significant cost savings.

Predicting response to medical therapy

What is already known? Rational selection of therapy in UC is an area of great promise and interest. Investigators have applied ML techniques to omics and clinical data in order to develop models that can accurately predict response to therapy a priori with varying degrees of success. Prediction of response to medical therapy in UC has focused on prediction of response to anti-tumor necrosis factor (TNF) therapy ($n=7$), though some studies exist for thiopurines ($n=1$) and anti-integrin therapy ($n=3$). Overall, these efforts have been limited by small datasets and study quality is variable. A key limitation of all the studies discussed is that they do not definitively show that the models predict response to a specific drug; for example, a model predicting response to anti-integrin may not in fact identify features that specifically are associated with response to anti-integrin therapy, and rather as associated with response to any form of IBD therapy. The studies are summarized in Table 2.

Table 2. Predicting response to medical therapy.

Reference	Study design	Brief description	Main findings
Waljee et al., 2017 ³⁶	Retrospective	Using a cohort of 1080 patients with IBD on thiopurines, an RF model was developed to predict remission. The model was compared to current gold standard, 6-thioguanine nucleotide levels	AUROC of the RF model was 0.79 vs 0.49 for 6-TGN. Patients with predicted remission had fewer steroid prescriptions, hospitalizations, and surgeries
Mishra et al., 2022 ³⁷	Prospective	In an IBD cohort of 14 patients (10 with UC) on infliximab, samples were collected at up to 7 time points from baseline to 14 weeks after induction. RNA sequencing and DNA methylation data were analyzed to predict clinical remission by pMayo using an RF model	Downregulation of nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) and TNF pathways at week 2 were associated with remission but no baseline factors were. RF model using week 0 and 2 to predict had 85% accuracy.
Feng et al., 2021 ³⁸	Prospective	Gene expression data at baseline in patients with UC who were treated with infliximab (IFX) were used to predict 14-week endoscopic remission. An RF model was applied to select differentially expressed genes, and an ANN was used to assign weights.	28 downregulated genes and 2 upregulated genes were identified. In the validation cohort, AUROC was 0.81
Popa et al., 2020 ⁴⁷	Retrospective	In 50 patients with UC treated with anti-TNF therapy, authors aimed to predict 52-week endoscopic remission. Using ANOVA 4 variables were selected, and after application of SMOTE, a neural net was applied.	In the internal validation cohort, neural net had an AUC of 0.92 for prediction of 52-week endoscopic remission
Obraztsov et al., 2018 ³⁹	Prospective	49 patients with UC-prescribed IFX were followed prospectively. Most patients were male and had pancolitis. Using baseline measurements of 17 serum cytokines, the authors predicted clinical remission at 12 weeks using LDA.	Baseline levels of TNF- α , IL-12, IL-8, IL-2, IL-5, IL1- β , and IFN- γ predict 12-week clinical remission. Confusion matrix shows a sensitivity of 84.2% and specificity of 93.3%
Chen et al., 2021 ⁴⁰	Retrospective	Three GEO datasets were mined for discovery. 44 patients with UC-prescribed IFX 5mg/kg were included in the study. Endoscopic remission between weeks 6 and 8 was predicted by pre-IFX gene expression. An independent cohort was used for validation	The ANN showed that <i>CDX2</i> , <i>CHP2</i> , <i>HSD11B2</i> , <i>RANK</i> , <i>NOX4</i> , and <i>VDR</i> levels predicted endoscopic remission. AUROC declines from 0.850 and degrades to 0.759 in the validation cohort
Miyoshi et al., 2021 ⁴¹	Retrospective	The discovery cohort consisted of 34 patients with UC-prescribed vedolizumab (VDZ), and the validation cohort had 35. Using baseline clinical features steroid-free, the authors predicted 22-week remission by Lichtiger index. An RF model was used for feature selection, and missing values were imputed. LR was applied	RF selected pMayo score, mean corpuscular hemoglobin (MCH), body mass index (BMI), blood urea nitrogen (BUN), concomitant azathioprine (AZA) (+/-), lymphocyte count, height, CRP, total cholesterol, and neutrophil count. There was 100% accuracy in cohort 1 suggesting overfitting, and 68.6% in cohort 2. However, the negative predictive value (NPV) of 92.3% suggested it may be able to rule out non-response.
Chen et al., 2021 ⁴⁰	Prospective	Data from VARSITY ($n=160$) and VISIBLE 1 ($n=383$) were mined. 429 patients on VDZ were included. 52-week steroid-free clinical remission was the outcome of interest. Baseline clinical features were used in elastic net regression and RF with and without SMOTE.	Pop: Varsity (160) + Visible 1 (383), 543 patients \rightarrow 429 without missing baseline information Definition of Outcome: The model showed that baseline steroid use, serum albumin, endoscopic Mayo score, prior anti-TNF use, immunomodulator (IM) use, and complete Mayo score had predictive value. Complete Mayo and endoscopic Mayo are not independent variables. AUC was only 0.614 in the training dataset, and 0.811 in the test, raising concerns about possible data leak
Waljee et al., 2018 ⁴²	Prospective	Data from a phase III clinical trial for vedolizumab was mined. Using baseline and week 6 data from 491 patients on VDZ, the authors predicted 52-week corticosteroid-free endoscopic remission using an RF model. Delta calprotectin and VDZ levels at week 6 were calculated	Using baseline data, AUROC was 0.62 baseline. With incorporation of week 6 data, AUROC modestly rose to 0.73.

ANN, artificial neural networks; AUC, area under the curve; CRP, C-reactive protein; GEO, gene expression omnibus; IBD, inflammatory bowel disease; LDA, linear discriminant analysis; LR, linear regression; partial Mayo, pMayo; RF, random forest; SVM, support vector machines; TGN, thioguaninenucleotides; UC, ulcerative colitis.

What do current studies show?

Thiopurines. Thiopurines are antimetabolite drugs that function as immunomodulators. 6-Mercaptopurine and its prodrug azathioprine, are enzymatically metabolized to 6-thioguanine-nucleotides (6-TGN) which reduces gut inflammation. Thiopurines have a narrow therapeutic window, and traditionally therapeutic drug monitoring (TDM) has been utilized.⁴³ However, two small, albeit underpowered, randomized controlled trials failed to show a clinical difference between TDM-guided and weight-based dosing regimens.^{44,45} Up to half of patients on thiopurines discontinue treatment within 2 years of initiation due to either adverse event or failure of therapy.⁴³ To address these shortcomings a ML algorithm that could predict clinical remission was developed. A RF model was trained using approximately 1000 patients, and the authors showed that the model was superior to 6-TGN in predicting remission (AUROC 0.79 vs 0.49); patients with algorithm-predicted remission had lower rates of steroid prescription, hospitalization, and surgery.³⁶

Anti-tumor necrosis alpha therapy. Most studies that applied ML techniques to UC have focused on omics data ($n=5$), rather than clinical data. These studies are promising and provide key insights into the underlying mechanisms of disease but are far from being clinically applicable. Only two studies utilized clinical data that is readily available to the practicing clinician. The studies are summarized in Table 2.

In a study by Mishra et al., the authors aimed to predict clinical remission by partial Mayo scores at week 14 using whole blood samples to obtain RNA sequencing and DNA methylation data. Data was obtained from a discovery cohort of 14 patients; all but 1 patient was prescribed infliximab. They applied an RF model using data obtained at baseline and 2 weeks after induction. Downregulation of NF- κ B and TLF signaling at week 2 predicted response to therapy (accuracy 85%), but no baseline findings accurately predicted response.³⁷ Feng et al. used colonic mucosal gene expression, from gene expression omnibus (GEO) datasets, at baseline to predict endoscopic remission at week 14. They utilized RF for feature selection and then applied an ANN to assign weights to the DEGs. They tested in a separate cohort with an AUC of 0.81. The datasets were small and may have contributed to large

uncertainty in the test set.³⁸ Obraztsov et al. evaluated 49 patients with UC treated with IFX; most patients were male and had pancolitis. They used a pre-specified 17 cytokine panel to predict clinical remission at 12 weeks by using baseline data. Applying LDA, they showed that TNF- α , IL-12, IL-8, IL-2, IL-5, IL1- β , and IFN- γ levels predicted remission. The confusion matrix showed a sensitivity of 84.2% and a specificity of 93.3%.³⁹ Finally, Chen et al. mined three GEO datasets for discovery, utilizing only patients receiving 5 mg/kg of infliximab. Baseline mucosal gene expression prior to IFX infusion was used to predict 8-week endoscopic remission. Given the small datasets, synthetic bootstrapping was used and then an ANN was applied. A model incorporating *CDX2*, *CHP2*, *HSD11B2*, *RANK*, *NOX4*, and *VDR* levels was shown to have an AUC of 0.850, but AUC declined to 0.759 in an independent cohort.⁴⁰

Two studies aimed to predict response to anti-TNF therapy with clinical variables. Xiojun et al. used a heterogeneous group of 420 patients with UC on a variety of therapies including aminosalicylates, thiopurines, and biologics. They used demographics, laboratory measurements, and medicines to predict endoscopic remission, but the time points for the input data were not clearly defined. They used inferential analysis for feature selection, and then applied multiple models including LR, RF, and SVM to the data; to address under-sampling of patients in remission and those with mild disease SMOTE was utilized. The final model had an AUC of 0.80.⁴⁶ A second study using clinical data by Popa et al. used baseline data from a cohort of 50 patients with UC to predict endoscopic remission at 52 weeks. Feature selection was done with ANOVA and the final model incorporated four variables: neutrophil count, platelet distribution width, CRP, and alpha-1-globulins. SMOTE and cross-validation were performed to reduce overfitting and imbalance data with a small dataset as much as possible. The model had an AUC of 0.92 in a validation dataset of five patients from the same center.⁴⁷

Overall, there are multiple promising models for assessing response to anti-TNF therapy, but overall any clinical application is currently limited either by use of data that is not widely available or derivation from small samples that have unclear external validity.

Anti-integrin therapy. Studies predicting response to anti-integrin therapy are less numerous than those predicting response to anti-TNF therapy. A study by Miyoshi *et al.* utilized baseline demographic, IBD, laboratory, and prescription data to predict 22-week remission by Lichtiger index. Data was trained on 34 patients at a single hospital and tested on 35 patients at a different institution. Missing data was imputed, carrying a risk of bias given the small sample size and retrospective data. RF was used for feature selection, and ultimately eight features were included in the LR model (MCH, BMI, BUN, Concomitant AZA use, lymphocyte count, height, CRP, total cholesterol, and neutrophil count). The model only had 68.6% accuracy in the validation cohort suggesting overfitting but had an NPV of 92.3%. It may be valuable to rule out non-response.⁴¹ Chen *et al.* combined data from VARSITY and VISIBLE 1, resulting in a dataset of 429 patients. Fifty-two-week steroid-free remission by Mayo score was the outcome of interest, and baseline clinical features were used as the predictors. Elastic net regression was compared with RF, with and without SMOTE on a 75:25 split dataset. Baseline steroid use, albumin, endoscopic Mayo score, prior anti-TNF use, IM use, and complete Mayo score were included; notably complete Mayo is not an independent predictor from endoscopic Mayo. AUC was 0.614 in the training set, and 0.811 in the test set raising the possibility

of data leak given the unexpected increase in performance.⁴⁸ Finally, Waljee *et al.* used baseline and week 6 clinical data from a phase III clinical trial to predict 52-week corticosteroid-free endoscopic remission. An RF model was used with a 70:30 split of data. AUC using only baseline data was 0.62, and AUC with addition of week 6 data was 0.73.⁴² Overall, no model had adequate predictive characteristics to inform clinical decision making at this time.

What could AI add in the future? Rational selection of IBD therapy is both an area of research interest and significant clinical need. Current models suffer from significant limitations as noted above. AI, likely in conjunction with fundamental advances in basic science, has the potential to bring the era of precision medicine to patients with IBD by providing true class-specific predictions on the likelihood of response to medications.

Monitoring disease activity in patients with established UC. Endoscopy, histologic assessment, and laboratory testing play important roles in the surveillance of UC. There have been numerous AI and ML applications aimed toward evaluating endoscopic lesions, predicting histological indices to grade severity of UC activity, and identifying biomarkers of active disease. These studies are summarized in Table 3.

Table 3. Monitoring disease activity in patients with established ulcerative colitis.

Author	Study design	Brief description	Main findings
Ozawa <i>et al.</i> , 2019 ⁴⁹	Retrospective	CNN-based CAD trained using 26,304 colonoscopy images from 841 UC patients, then tested on predicting MES in independent set of 3981 images from 114 patients	Accurately differentiated between Mayo 0 (AUC 0.86) and Mayo 0-1 (AUC 0.98) states compared with more inflamed disease states
Stidham <i>et al.</i> , 2019 ⁵⁰	Retrospective	CNN was trained and tested on 16,514 images from 3082 patients with UC to categorize patients into an endoscopic remission group (Mayo 0 or 1) vs a moderate-to-severe disease group (Mayo 2 or 3)	Distinguished endoscopic remission state from moderate-to-severe disease state with high accuracy (AUC 0.966)
Yao <i>et al.</i> , 2021 ⁵¹	Retrospective	CNN model was trained using 51 high-resolution UC endoscopic videos and tested on 264 endoscopic videos from multicenter clinical trials to predict whole-video MES	Automated MES scoring predicted MES correctly in 78% of high-resolution endoscopic videos; in external clinical trial videos, automated and central reviewer scoring agreement occurred in 57.1% of videos, which improved to 69.5% when accounting for inter-reviewer disagreement
Gottlieb <i>et al.</i> , 2021 ⁹	Retrospective	Using 795 full-length endoscopic videos from a transnational multicenter phase II trial of mirikizumab, a recurrent neural network (NN) model was trained to predict MES and UCEIS in individual full-length videos	Quadratic weighted kappa metric to assess agreement between automated and central reviewer scoring was 0.844 for MES and 0.855 for UCEIS

(Continued)

Table 3. (Continued)

Author	Study design	Brief description	Main findings
Gutierrez Becker et al., 2021 ⁵²	Retrospective	Developed deep learning-based system using 1672 endoscopic videos from a multicenter clinical trial to predict binary ratings of the MES	The automated system was able to grade endoscopic videos with accuracy and robustness (MES ≥ 1 , AUC 0.84; MES ≥ 2 , AUC 0.85; MES ≥ 3 , AUC 0.85)
Fan et al., 2023 ⁵³	Retrospective	Trained CNN using 5875 endoscopic images and 20 full-length videos from 332 patients with UC who underwent colonoscopy between 2017 and 2021, which was then used for full-length video scoring and to generate a visualization of full-length intestinal inflammatory activity	CNN model achieved 86.54% accuracy in the MES-related task, had a kappa coefficient of 0.813 compared with endoscopist scoring, and was able to display the distribution of intestinal inflammation using a 2-dimensional visualization
Maeda et al., 2022 ⁵⁴	Prospective	145 UC patients in clinical remission underwent AI-assisted colonoscopy that classified patients into active vs healing groups to assess the role of AI in stratifying patients by risk for clinical relapse of UC	Clinical UC relapse rate was found to be significantly higher in the AI-active group vs the AI-healing group (28.4% vs 4.9%, $p < 0.001$)
Takenaka et al., 2020 ⁵⁵	Prospective	Constructed a deep neural network using 40,578 colonoscopy images and 6885 biopsy results from 2012 UC patients who underwent colonoscopy from 2014 to 2018 at a single center in Japan. The algorithm was evaluated in an independent cohort of 875 patients with UC who underwent colonoscopy in 2018–2019, predicting endoscopic and histologic remission in 4187 endoscopic images and 4104 biopsy specimens.	In the validation cohort, the ML algorithm was able to predict endoscopic remission with 90.1% accuracy and a kappa coefficient of 0.798 and was able to predict histologic remission with 92.9% accuracy and a kappa coefficient of 0.859.
Takenaka et al., 2020 ⁵⁵	Prospective	The algorithm developed in Takenaka et al., 2020 (above) was applied to a prospective cohort of 875 patients to predict mucosal healing based on endoscopic and histologic remission. Mucosal healing was then correlated with clinical endpoints of worse prognosis, including hospitalization, colectomy, steroid use, and clinical relapse.	Mucosal healing as predicted by the ML algorithm was associated with significantly lower risk of hospitalization, colectomy, steroid use, and clinical relapse ($p < 0.001$). Compared to expert review, the ML algorithm had a high sensitivity (92.0%) and specificity (91.3%) for evaluating mucosal healing.
Maeda et al., 2019 ⁵⁶	Retrospective	Developed a CAD system to predict histologic inflammation using endocytoscopy using a training data set from 187 UC patients who underwent endocytoscopy with corresponding biopsies. Diagnostic ability of the CAD to predict histologic inflammation was then tested in a validation dataset of 100 patients.	Performance of the CAD system yielded diagnostic sensitivity of 74%, specificity of 97%, and accuracy of 91%.
Bossuyt et al., 2020 ⁵⁷	Prospective	Constructed computer algorithm based on RD to predict measures of endoscopic and histologic inflammation. Compared results to blinded central readers.	In the construction cohort, RD correlated with rhi ($r = 0.74$, $p < 0.0001$), Mayo endoscopic sub-scores ($r = 0.76$, $p < 0.0001$), and endoscopic index of severity scores ($r = 0.74$, $p < 0.0001$). The RD sensitivity to change had a standardized effect size of 1.16. In the validation set, RD correlated with rhi ($r = 0.65$, $p = 0.00002$). Validation cohort included
Iacucci et al., 2023 ⁵⁸	Retrospective	Using 1090 WLE and VCE endoscopic videos from 283 patients, a CNN was trained to distinguish endoscopic remission vs activity, predict histologic remission vs activity, and predict risk of flare. CNN performance was compared to expert scoring.	In WLE videos, the CNN detected endoscopic remission with 72% sensitivity, 87% specificity, and AUC of 0.85. In VCE videos, the CNN detected endoscopic remission with 79% sensitivity, 95% specificity, and AUC of 0.94. Histologic remission was predicted with 80%–85% accuracy in WLE and VCE videos.
Vande et al., 2022 ⁵⁹	Retrospective	Developed a deep learning algorithm to identify eosinophils in colonic biopsies, which was then applied to sigmoid colon biopsies from a cohort of 88 UC patients with histologically active disease. Analyzed associations between eosinophil density, histologic activity, and clinical features.	The eosinophil deep learning algorithm demonstrated a high degree of agreement with manual eosinophil counts determined by expert pathologists [correlation coefficients 0.805–0.917]. Eosinophil density was not significantly correlated with histologic activity as measured by Roberts Histopathology Index but was associated with disease extent (as measured by Montreal classification) and corticosteroid use.

(Continued)

Table 3. (Continued)

Author	Study design	Brief description	Main findings
Gui et al., 2022 ⁶⁰	Prospective	Using a dataset of 614 biopsies from 307 patients with UC, developed the PHRI. Constructed a CAD using CNN algorithm to detect neutrophils, calculate PHRI, and identify active from quiescent disease.	CAD system incorporating PHRI was able to differentiate active from quiescent UC with 78% sensitivity, 91.7% specificity, and 86% accuracy.
Peyrin-Biroulet et al., 2022 ⁶¹	Retrospective	A novel AI system was developed using image processing and ML algorithms to characterize histological images and measure Nancy index. Results were compared with manual annotation and Nancy index scoring by 3 independent histopathologists.	The AI system scoring was found to be highly correlated with histopathologists with an average ICC of 87.20. The average ICC among the histopathologists was 89.33.
Morilla et al., 2019 ⁶²	Retrospective	Performed microarray analysis of mRNA expression profiles on 47 patients with ASUC before and within 3 days of treatment with steroids, cyclosporine, or infliximab. Deep neural network-based classifier was used to propose biomarkers for discriminating responders from non-responders for each treatment, and classification was tested on an independent cohort of 29 patients.	The deep neural network-based classifier identified 9 miRNA and 5 clinical factors, and the classification algorithm was able to discriminate responders from non-responders accurately (steroid-treated cohort, 93% accuracy, AUC 0.91; cyclosporine-treated cohort, 80% accuracy, AUC 0.79; infliximab treated cohort, 84% accuracy, AUC 0.82)
He et al., 2022 ⁶³	Retrospective with prospective animal model	Performed feature selection of differentially expressed mRNAs on 2 microarray data sets (19 normal samples, 31 UC samples) as well as immune infiltrate and gene set enrichment analysis. Further validated candidate biomarker mRNA expression levels in UC cell cultures and mouse models.	Identified 8 candidate mRNA biomarkers, 3 of which were then validated in a separate dataset and correlated with immune infiltrate analysis and gene set variation analysis; these biomarkers' expression levels in UC were then validated in cell culture models via qrtPCR as well as in UC mouse models colon tissue expression levels
Biasci et al., 2019 ⁶⁴	Prospective	Patients with active IBD underwent transcriptomic analyses on CD8 T cells and/or whole blood prior to treatment. ML was used to identify differentially expressed genes to stratify patients into higher and lower risk groups. The classifier was tested in an independent patient cohort and correlated with clinical phenotypes, such as earlier need for treatment escalation and number of escalations over time.	Developed a blood-based, 17-gene quantitative PCR (qPCR) classifier that stratified IBD patients into low-risk group and high-risk group who experienced significantly more aggressive disease (earlier need for treatment escalation, UC patients HR=3.12; multiple escalations within 18 months, UC patients sensitivity 100% and NPV 100%)
Lai et al., 2021 ⁶⁵	Retrospective	Analyzed UC datasets from GEO database (2 gene expression datasets, 1 microRNA dataset), performed differential expression analysis, and applied weighted co-expression network analysis of UC-related genes to characterize 4 UC subtypes, which were correlated with clinical features, such as Mayo scores, and baseline calprotectin levels.	Through differential gene expression analysis and network analysis, proposed 4 subtypes and 6 genes as biomarkers for UC classification
Li et al., 2023 ⁶⁶	Retrospective	In a cohort of 65 UC patients, used ML techniques to identify Vitamin D, albumin, prealbumin, and fibrinogen levels as blood-based biomarkers to predict moderate-to-severe endoscopic activity of UC.	Constructed a dynamic nomogram prediction model using blood levels of Vitamin D, albumin, prealbumin, and fibrinogen to predict moderate-to-severe endoscopic activity of UC with a concordance index of 0.860 using UCEIS scoring and a concordance index of 0.891 with MES scoring.
Gazouli et al., 2019 ⁶⁷	Retrospective	In a cohort of 573 Greek IBD patients (209 UC) and 445 controls, whole genome association analysis was performed to detect disease-associated single nucleotide polymorphisms (SNPs), which were then used for pathway analysis to identify proteins with high significance and potential biomarkers of disease.	Analyses revealed several novel and well-known pathways associated with IBD, and IBD sub-phenotypes were found to have distinct genetic and functional profiles that may be used for classification of disease
Bakir-Gungor et al., 2022 ⁶⁸	Retrospective	Developed a classification model to aid IBD diagnosis and discover IBD-associated biomarkers using metagenomics data from microbiome DNA sequencing data of 148 IBD patients and 234 control samples.	Tested multiple different types of classifiers and found that a random forest classifier using 10 selected taxonomic biomarkers resulted in the best performance measures for classification of IBD (validation cohort accuracy 86%, AUC 0.9)

(Continued)

Table 3. (Continued)

Author	Study design	Brief description	Main findings
Stidham et al., 2023 ⁶⁹	Retrospective	Developed an MES classifier layered with a motion-detection algorithm to calculate CDS among patients in the UNIFI clinical trial	The authors show that CDS is correlated with the MES, but required 50% less patients than MES to detect a statistically significant difference in endoscopic outcome
Najdawi et al., 2023 ⁷⁰	Retrospective	Developed a CNN to quantitatively measure histologic features in stained whole image slides. Current methods are only semi-quantitative	There was high correlation with Nancy histologic index as determined by a pathologist ($r=0.89$), and it was highly accurate in predicting remission (97%)
Iacucci et al., 2023 ⁷¹	Retrospective	A CNN was compared again to multiple histologic scoring indices. The model was also used to predict endoscopic activity at 12 months.	The CNN was able to distinguish between disease activity and remission with accuracies between 80% and 90%.

AI, artificial intelligence; ASUC, acute severe ulcerative colitis; AUC, area under the curve; CAD, computer-assisted diagnosis; CDS, cumulative disease score; CNN, convolutional neural networks; GEO, gene expression omnibus; IBD, inflammatory bowel disease; ICC, intraclass correlation coefficient; MES, Mayo endoscopic score; ML, machine learning; PHRI, PICaSSO Histologic Remission Index; RD, red density; UC, ulcerative colitis; UCEIS, UC endoscopic index of severity; VCE, virtual chromoendoscopy; WLE, white-light endoscopy.

AI applications in endoscopic monitoring

What is already known? Endoscopy is a cornerstone of assessing UC disease activity, and several endoscopic scores have been developed to define disease activity. The most commonly used and extensively validated endoscopic scores include the Mayo endoscopic score (MES), the UCEIS, and the UC colonoscopic index of severity.⁷² However, these endoscopic scores are limited by their qualitative nature, subjectivity, and corresponding interobserver variability. Further, these scores typically report the maximum severity observed and fail to capture the heterogeneity of disease and total disease burden. Thus, AI applications in UC endoscopy have focused on identifying signs of inflammation on endoscopy as well as standardizing the interpretability of endoscopic findings in UC disease surveillance.

What do current studies show? An initial study by Ozawa et al. constructed a computer-assisted diagnosis (CAD) system using a CNN that accurately identified endoscopic disease remission from still images.⁴⁹ A subsequent study by Stidham et al.⁵⁰ employed a CNN constructed as a DL model to differentiate endoscopic remission (MES 0 or 1) from moderate-to-severe disease (MES 2 or 3) from still images (AUC 0.966). Both studies are significantly limited due to their applications to still images, which does not represent typical clinical practice in which decisions are made based off of the entire colonoscopy. However, neural networks have since been applied to analysis of full-length endoscopic video data and have been demonstrated to reliably predict

endoscopic disease severity with reasonably high rates of agreement with expert reviewers.^{9,51–53} In one prospective study, AI-assisted colonoscopy was able to stratify patients with UC in clinical remission into higher and lower risk groups for clinical relapse of UC, evidencing the prognostic potential of AI-assisted endoscopy to predict clinical outcomes and accordingly influence disease management.⁵⁴ A recent publication by Stidham et al. showed that computer vision could be used to calculate a cumulative disease score (CDS) by assigning MES to all frames of adequate quality for a given colonoscopy that were mapped to an estimated location; CDS was defined as the sum of MES-squared values. The authors showed that CDS was more sensitive than MES for detecting change; CDS required 50% fewer participants to demonstrate a difference in the endoscopic outcome, a finding with clear cost implications for clinical trials.⁶⁹

What could AI add in the future? Computer vision may augment the capabilities of general gastroenterologists, allowing them to perform at a similar level as IBD specialists. Future studies will be needed to evaluate if AI assessment will obviate the need for virtual or dye-based chromoendoscopy. Use of tools such as the CDS may be able to increase power and decrease cost for clinical trials in IBD.

AI applications in histology assessment

What is already known? In addition to assessing endoscopic outcomes in UC, AI has also been applied to histological evaluation. Histological

signs of inflammation, even in the absence of endoscopic inflammation, have been associated with adverse clinical outcomes in UC, making histologic remission an important adjunct goal of UC treatment.^{73–75} To this end, there have been widespread efforts to apply AI and ML techniques to predict histologic activity in patients with UC.

What do current studies show? In an initial study, Takenaka *et al.*⁵⁵ constructed a deep neural network using colonoscopy images and biopsy results from a cohort of patients with UC, which was then able to predict endoscopic remission with 90.1% accuracy as well as histologic remission with 92.9% accuracy in the validation cohort. In a follow-up study using a prospective cohort of 875 patients, mucosal healing predicted by the deep neural network algorithm based on endoscopic and histologic remission was shown to be correlated with significantly lower risk of hospitalization, colectomy, steroid use, and clinical relapse.⁷⁶ In another study, Maeda *et al.*⁵⁶ developed a CAD system to predict histologic inflammation using endocytoscopy with a sensitivity of 74%, specificity of 97%, and accuracy of 91% when compared to pathologist interpretation of corresponding biopsies. Najdawi *et al.* developed a CNN which was compared against the Nancy index. The CNN showed strong correlation with pathologist-determined Nancy index ($r=0.89$) and was highly accurate at determining histologic remission (accuracy 97%). Notably, the CNN was only designed to assess disease activity and was unable to identify other clinically important features including signs of infection or dysplasia.⁷⁰ Iacucci *et al.* also developed a CNN which was compared against Nancy index, PICaSSO Histologic Remission Index, and Robarts. The CNN when compared against these three indices had sensitivities ranging from 89% to 94% and specificities ranging from 76% to 85%. This algorithm was also unable to identify other clinically important features including infection and dysplasia.⁷¹

Other studies have incorporated additional endoscopy features, such as red density lighting and virtual chromoendoscopy into ML algorithms to predict measures of endoscopic and histologic inflammation.^{57,58} Models to assess histologic inflammation have also been developed from direct analysis of biopsies themselves, using image processing techniques to detect eosinophils, neutrophils, and other histologic

features.^{59–61} These models have consistently demonstrated a high degree of agreement with scoring by independent experts.

What could AI add in the future? There are a variety of scoring systems for pathology in IBD, but these scoring methods are not standardized across institutions and there are issues with interobserver variability which may be addressed with AI. Further, with continued improvement, these tools may expand access to specialist care by enabling general pathologists to evaluate IBD specimens at a similar level to GI pathologists at referral centers.

AI applications in novel biomarker discovery

What is already known? ML techniques have also been applied toward multi-omics data sets including genetic, transcriptional, and microbiome data to identify novel biomarkers of UC disease activity. Most of these datasets have a very high number of predictors (omics output) derived from a small cohort of patients; this problem, called “big-p, little-n,” causes significant issues for prediction, and require specialized data preparation and proper algorithms to properly handle in the input.

What do the current studies show? In a study by Morilla *et al.*,⁶² microarray data was utilized to build a deep neural network-based classifier consisting of nine miRNAs and five clinical factors that accurately discriminated patients with acute severe UC as responders versus non-responders to steroids (accuracy 93%, AUC=0.91), infliximab (accuracy 84%, AUC=0.82), and cyclosporine (accuracy 80%, AUC=0.79). In another study by He *et al.*,⁶³ ML algorithms were used to identify differentially expressed mRNAs to serve as diagnostic biomarkers for UC, which were then further validated in cell lines and mouse models of colitis. Whole blood transcriptomic data has been used to develop a qPCR-based classifier that stratified patients into high and low-risk groups associated with earlier need for treatment escalation (hazard ratio 3.12) and more escalations over time in UC patients.⁶⁴ Notably, endoscopic severity at baseline did not predict need for treatment escalation in this cohort, highlighting the ability of ML algorithms to impact treatment decisions in ways that would be undetectable by conventional methods of disease surveillance. Transcriptomic data has also been leveraged to characterize different subtypes of UC patients, which were

associated with various relevant clinical features including Mayo scores, calprotectin levels, and histological severity scores.⁶⁵ Other studies have used ML techniques to investigate blood-based, genome-wide association studies, and microbiome biomarkers related to UC diagnosis, phenotypes, and disease severity.^{66–68} Overall, the application of ML techniques to biomarker discovery in UC has revealed promising biomarkers related to diagnosis, disease subtyping, and prognostication. Further testing is required to determine the clinical translatability of these biomarkers.

What could AI add in the future? Even though calprotectin is a reliable biomarker, between 5% and 10% of patients have discordant results when compared with colonoscopy. Application of AI

for the identification of biomarkers for more reliable non-invasive clinical monitoring would be extremely clinically valuable. Further, while any individual biomarker is unlikely to rival the diagnostic accuracy of colonoscopy, an AI tool that can combine multiple biomarkers may be able to provide similar accuracy.

Predicting complications of UC. The literature regarding applications of AI to complications of UC is limited. There were four broad areas that the literature focused on: predicting the need for colectomy ($n=2$), predicting postoperative complications ($n=2$), predicting colorectal cancer (CRC) ($n=2$), and prediction of COVID-19 outcomes ($n=1$). These studies are summarized in Table 4.

Table 4. Predicting complications of ulcerative colitis.

Author	Study design	Brief description	Main findings
Noguchi et al., 2022 ⁷⁷	Retrospective	46 H&E stained slides and corresponding p53 stained slides were prepared from samples of 12 patients with colitis-associated neoplasia who underwent total colectomy. All glands were annotated and grouped into 3 classes: p53 positive, p53 negative, and p53 null. Ten patients were used for training a CNN, 1 patient for validation, and 1 patient for final testing.	The trained CNN was able to predict p53 immunohistochemical staining with accuracy of 0.86–0.91 with the limitations of single-person validation and testing cohort.
Yu et al., 2022 ⁷⁸	Retrospective	129 patients treated with IVCS for acute severe ulcerative colitis; 102 (79.1%) responded, and 27 failed (20.9%). Classification models, including logistic regression, decision tree, random forest, and extreme-gradient boosting models were used to analyze prediction rates of IVCS resistance.	The LR model had the best classification performance of 0.873, falling to 0.703 in the validation cohort
Takayama et al., 2015 ⁷⁹	Retrospective	Used artificial neural network (ANN) as a prediction model for the need for surgery post-cytapheresis (CAP) therapy. The sample population consisted of 90 UC patients who had undergone CAP therapy.	ANN showed high predictive accuracy, with a sensitivity of 0.96 and a specificity of 0.97. However, the nature of prior operations used as a predictor in the model is unclear (prior colorectal surgery vs any surgery) and may strongly impact the validity of the model.
Sofa et al., 2020 ⁸⁰	Retrospective	Clinical data from 32 UC patients who had undergone total abdominal colectomy were used for the ML algorithm.	The algorithm predicted minor infectious complications with high strike rate (84.3%) but was unable to predict any serious complications
Mizuno et al., 2022 ⁸¹	Retrospective	In 43 patients with UC who underwent IPAA, the ability of pre-ileostomy closure mPDAI and a CNN to predict pouchitis was compared	CNN produced 20% greater pouchitis prediction rate than mPDAI (62% vs 84%)
Roy et al., 2021 ⁸²	Prospective	Data from SECURE-IBD with 20,000 IBD patients was inputted into various supervised machine learning algorithms to generate clinical COVID-19 outcomes (outpatient management, hospitalized and recovered, and hospitalized and deceased)	A variety of supervised machine learnings were evaluated, and all had poor classification performance

ANN, artificial neural networks; CAP, colectomy after cytopheresis; CNN, convolutional neural networks; H&E, hematoxylin and eosin; IPAA, ileal pouch–anal anastomosis; IVCS, IV corticosteroid; LR, linear regression; ML, machine learning; mPDAI, modified pouchitis disease activity index; UC, ulcerative colitis.

Predicting the need for colectomy

What is already known? Colectomy is used to treat medically refractory acute severe UC. We know that approximately 10%–15% of patients with UC will undergo colectomy during their lifetime. While traditional epidemiologic methods have found risk factors associated with colectomy, these models are unable to predict risk for a given patient.

What do current studies show? Two studies aimed to construct and validate models that could predict post-treatment complications that require follow-up treatment. In one study by Yu *et al.*, traditional LR models and ML models were compared as predictive models for IV corticosteroid (IVCS) resistance in patients with acute severe ulcerative colitis (ASUC). UCEIS and CRP level at day 3 of IVCS therapy were independent predictors of IVCS response. No ML method was able to outperform traditional LR (AUC of 0.703 in the validation cohort). The study was limited by small sample size from a single patient population and the relatively poor classification performance of the algorithms.⁷⁸ In a second study by Takayama *et al.*, an ANN was utilized to predict the need for colectomy after cytopheresis (CAP) therapy based on 13 input factors using a training data set ($n=54$) and validation data set ($n=36$). The prediction model identified four key factors: history of prior admissions, prior operations, use of immunomodulators, and response to CAP therapy. The model had a sensitivity of 0.96 and specificity of 0.97. The nature of prior operations used as a predictor in the model are unclear (prior colorectal surgery vs any surgery) and may strongly impact the validity of the model.⁷⁹

What could AI add in the future? Predicting which patients with ASUC will require colectomy is an area of clinical need. Patients who fail IVCSs are often given rescue therapy, typically infliximab or cyclosporine. By identifying patients who are not likely to respond to medical therapy, algorithms may help clinicians avoid unnecessary immunosuppression prior to surgery. A significant barrier to AI for this application is the relative rarity of ASUC, and the lack of large databases for training models for this end use.

Prediction of postoperative complications

What is already known? Patients with acute severe UC or treatment-refractory UC often undergo surgery. These patients have a high risk

of post-surgical complications, and there is a clear clinical benefit of being able to predict surgical outcomes. In particular, pouchitis is a vexing complication that can lead to persistence of symptoms and impaired quality of life after colectomy. Two studies applied AI methodology to this clinical problem.

What do the current studies show? In one study by Mizuno *et al.*, the researchers aimed to determine whether a CNN could accurately predict pouchitis development after ileal pouch–anal anastomosis (IPAA) in UC patients. Modified pouchitis disease activity index (mPDAI) before ileostomy closure was compared with a CNN model based on the endoscopic imaging of a retrospective cohort of 43 patients with 5-fold cross-validation. The predictive rates for pouchitis of mPDAI prior to ileostomy closure and the CNN model were estimated by ROC. mPDAI had an accuracy of 62% and the CNN had an accuracy of 84%. Limitations include a small number of images and variation in image scoring due to use of different endoscopists, which could be overcome in the future with a multicenter design and standardization of imaging methods. Nevertheless, the findings suggest that CNN models may predict pouchitis, and allow for early intervention.⁸¹ In a second study, Sofo *et al.* looked at a cohort of high-risk UC patients who had undergone a total colectomy and aimed to predict various types of postoperative complications using data available before surgery. This simulated a prospective study and was able to predict minor infectious complications accurately, but major infection and non-infectious complications were not predicted as accurately, greatly limiting the clinical utility of this study.⁸⁰

What could AI add in the future? Pouchitis is a common complication after IPAA. While many patients respond to a single course of antibiotics, a subset develops chronic pouchitis, a devastating complication that affects quality of life after a theoretically curative colectomy. Being able to predict complications like pouchitis may help surgical planning.

Colorectal cancer surveillance

What is already known? Although patients with long-standing UC have a higher risk of developing CRC and there is significant literature regarding CAD in general, there is limited data on the application of AI to CRC risk in UC patients.⁸³

What do the current studies show? One study by Uttam et al. aimed to aid early detection by applying three-dimensional nanoscale nuclear architecture mapping to detect advanced dysplasia or neoplasia in normal-appearing rectal biopsies of patients with both UC and Crohn's disease prior to detection by conventional history. They applied SVM as a binary classifier and the final model had an AUROC of 0.870.⁸⁴ Noguchi et al.⁷⁷ used a CNN to predict p53 immunohistochemical staining from hematoxylin and eosin stained slides without dedicated p53 stains. The trained CNN was able to predict p53 immunohistochemical staining with accuracy of 86%–91%. Although the results are promising, the study did not validate the CNN in an external dataset, and the sample size was small with only 12 patients, with strong risk of overfitting.

What could AI add in the future? Further studies should incorporate external validation and larger sample sizes in order to develop strong predictive models for colitis-associated dysplasia, and biomarkers aside from p53 should be investigated.⁷⁷ Surveillance in patients with dense pseudopolypoidosis is technically challenging and represents an area in which computer vision may prove to be useful.

COVID-19 outcomes. UC is often treated with immunosuppressants, which may lead to higher risk of infection. The outcome of COVID-19 in UC patients is of significant interest and there have been numerous studies which have applied traditional epidemiologic methods. However, there is a paucity of studies which have applied AI methods to this patient population. A single study by Roy et al.⁸² addresses this issue by using the SECURE-IBD database. They applied a variety of and supervised learning methods, but the best performing model only had an accuracy of only 70%.

Conclusion

AI shows great promise in UC, and there has been burgeoning interest in the field. ML and DL techniques have been applied to a wide range of meaningful clinical problems in UC, including the identification of new UC, personalized therapy, monitoring of disease activity, and prediction of complications. Despite the considerable promise of AI in UC, there are also key limitations; many studies have small sample sizes and biases

that risk overfitting. There have been limited validation of studies in truly independent external datasets. On the whole, rarely have the developed models had adequate performance characteristics to justify potential clinical deployment. Given the current status of the field of AI in UC, future research should include: (1) robust, large scale external validation of models to overcome the many limitations and bias that come with using small internal training datasets, (2) studies that predict clinically meaningful outcomes that are in line with current standard of care, such as endoscopic remission rather than clinical remission, (3) studies that evaluate the cost-effectiveness of model-guided therapy compared to the current standard of care, (4) head-to-head studies of models which predict the same outcomes to guide clinical implementation, and (5) randomized controlled trials of AI models to determine if they meaningfully impact clinical outcomes.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Author contributions

Chiraag Kulkarni: Data curation; Formal analysis; Writing – original draft; Writing – review & editing.

Derek Liu: Data curation; Formal analysis; Writing – original draft; Writing – review & editing.

Touran Fardeen: Data curation; Formal analysis; Writing – original draft; Writing – review & editing.

Eliza Rose Dickson: Data curation; Formal analysis; Writing – original draft; Writing – review & editing.

Hyunsu Jang: Data curation; Writing – original draft; Writing – review & editing.

Sidhartha R. Sinha: Supervision; Writing – review & editing.

John Gubatan: Conceptualization; Funding acquisition; Investigation; Methodology; Resources; Supervision; Visualization; Writing – review & editing.

Acknowledgements

None.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: J.G. was in part supported by a Chan Zuckerberg Biohub Physician Scientist Scholar Award, NIH NIDDK LRP Award (L30 DK126220), and a Doris Duke Physician Scientist Fellowship Award (Grant #2021091), Stanford MCHRI Pediatric IBD, and Celiac Disease Research Award. S.R.S. was in part supported by the Paul G. Allen Frontiers Group and The Leona M. and Harry B. Helmsley Charitable Trust (Grant 2007-04026).

Competing interests

The authors declare that there is no conflict of interest.

Availability of data and materials

Not applicable.

ORCID iD

John Gubatan  <https://orcid.org/0000-0001-6037-2883>

References

- Shivashankar R, Tremaine WJ, Harmsen WS, et al. Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmstead County, Minnesota From 1970 Through 2010. *Clin Gastroenterol Hepatol* 2017; 15(6): 857–863.
- Lee D, Albenberg L, Compher C, et al. Diet in the pathogenesis and treatment of inflammatory bowel diseases. *Gastroenterology* 2015; 148(6): 1087–1106.
- Lemons JMS, Conrad M, Tanes C, et al. Enterobacteriaceae growth promotion by intestinal acylcarnitines, a biomarker of dysbiosis in inflammatory bowel disease. *Cell Mol Gastroenterol Hepatol* 2023; 17(1): 131–148.
- White JR, Phillips F, Monaghan T, et al. Review article: novel oral-targeted therapies in inflammatory bowel disease. *Aliment Pharmacol Ther* 2018; 47(12): 1610–1622.
- Wang W and Siau K. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: a review and research agenda. In: Management Association, Information Resources (ed.) *Research anthology on machine learning techniques, methods, and applications*. IGI Global, 2022, pp. 1460–1481.
- Panch T, Szolovits P and Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health* 2018; 8(2): 020303.
- Gubatan J, Mitsuhashi S, Zenlea T, et al. Low serum Vitamin D during remission increases risk of clinical relapse in patients with ulcerative colitis. *Clin Gastroenterol Hepatol* 2017; 15(2): 240–246.e1.
- Jiang F, Fu X, Kuang K, et al. Artificial intelligence algorithm-based differential diagnosis of Crohn's disease and ulcerative colitis by CT image. *Comput Math Methods Med* 2022; 2022: 3871994.
- Gottlieb K, Requa J, Karnes W, et al. Central reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology* 2021; 160(3): 710–719.e2.
- Sutton RT, Zaiane OR, Goebel R, et al. Artificial intelligence enabled automated diagnosis and grading of ulcerative colitis endoscopy images. *Sci Rep* 2022; 12: 2748.
- Chierici M, Puica N, Pozzi M, et al. Automatically detecting Crohn's disease and ulcerative colitis from endoscopic imaging. *BMC Med Inform Decis Mak* 2022; 22(Suppl. 6): 300.
- Sarlos P, Kovesdi E, Magyari L, et al. Genetic update on inflammatory factors in ulcerative colitis: review of the current literature. *World J Gastrointest Pathophysiol* 2014; 5(3): 304–321.
- Loftus EV. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* 2004; 126(6): 1504–1517.
- Ye Y, Pang Z, Chen W, et al. The epidemiology and risk factors of inflammatory bowel disease. *Int J Clin Exp Med* 2015; 8(12): 22529–22542.
- Popov J, Caputi V, Nandeeshan N, et al. Microbiota-immune interactions in ulcerative colitis and colitis associated cancer and emerging microbiota-based therapies. *Int J Mol Sci* 2021; 22(21): 11365.
- Rogler G, Zeitz J and Biedermann L. The search for causative environmental factors in inflammatory bowel disease. *Dig Dis* 2016; 34(Suppl. 1): 48–55.
- Tang D, Pu B, Liu S, et al. Identification of cuproptosis-associated subtypes and signature genes for diagnosis and risk prediction of ulcerative colitis based on machine learning. *Front Immunol* 2023; 14: 1142215.

18. Zhang L, Mao R, Lau CT, et al. Identification of useful genes from multiple microarrays for ulcerative colitis diagnosis based on machine learning methods. *Sci Rep* 2022; 12: 9962.
19. Li H, Lai L and Shen J. Development of a susceptibility gene based novel predictive model for the diagnosis of ulcerative colitis using random forest and artificial neural network. *Aging (Albany NY)* 2020; 12(20): 20471–20482.
20. Han Y, Liu X, Dong H, et al. Screening of characteristic genes in ulcerative colitis by integrating gene expression profiles. *BMC Gastroenterol* 2021; 21: 415.
21. Kraszewski S, Szczurek W, Szymczak J, et al. Machine learning prediction model for inflammatory bowel disease based on laboratory markers. Working model in a discovery cohort study. *J Clin Med* 2021; 10(20): 4745.
22. Dhaliwal J, Erdman L, Drysdale E, et al. Accurate classification of pediatric colonic inflammatory bowel disease subtype using a random forest machine learning classifier. *J Pediatr Gastroenterol Nutr* 2021; 72(2): 262–269.
23. Lu J, Wang Z, Maimaiti M, et al. Identification of diagnostic signatures in ulcerative colitis patients via bioinformatic analysis integrated with machine learning. *Hum Cell* 2022; 35(1): 179–188.
24. Khorasani HM, Usefi H and Peña-Castillo L. Detecting ulcerative colitis from colon samples using efficient feature selection and machine learning. *Sci Rep* 2020; 10. <https://doi.org/10.1038/s41598-020-70583-0>
25. Wang Y, Huang J, Zhang J, et al. Identifying biomarkers associated with the diagnosis of ulcerative colitis via bioinformatics and machine learning. *Math Biosci Eng* 2023; 20(6): 10741–10756.
26. Duttagupta R, DiRienzo S, Jiang R, et al. Genome-wide maps of circulating miRNA biomarkers for ulcerative colitis. *PLoS One* 2012; 7(2): e31241.
27. Chen Q, Bei S, Zhang Z, et al. Identification of diagnostic biomarkers and immune cell infiltration in ulcerative colitis. *Sci Rep* 2023; 13(1): 6081.
28. Mossotto E, Ashton JJ, Coelho T, et al. Classification of paediatric inflammatory bowel disease using machine learning. *Sci Rep* 2017; 7: 2427.
29. da Silva BC, Lyra AC, Rocha R, et al. Epidemiology, demographic characteristics and prognostic predictors of ulcerative colitis. *World J Gastroenterol* 2014; 20(28): 9458–9467.
30. Kucharzik T, Koletzko S, Kannengiesser K, et al. Ulcerative colitis—diagnostic and therapeutic algorithms. *Dtsch Arztebl Int* 2020; 117(33–34): 564–574.
31. Panes J, Bouhnik Y, Reinisch W, et al. Imaging techniques for assessment of inflammatory bowel disease: joint ECCO and ESGAR evidence-based consensus guidelines. *J Crohns Colitis* 2013; 7(7): 556–585.
32. Wagatsuma K, Yokoyama Y and Nakase H. Role of biomarkers in the diagnosis and treatment of inflammatory bowel disease. *Life (Basel)* 2021; 11(12): 1375.
33. Dulai PS, Peyrin-Biroulet L, Danese S, et al. Approaches to integrating biomarkers into clinical trials and care pathways as targets for the treatment of inflammatory bowel diseases. *Gastroenterology* 2019; 157(4): 1032–1043.e1.
34. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019; 13: 144–164K.
35. Sturm A, Maaser C, Calabrese E, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD Part 2: IBD scores and general principles and technical aspects. *J Crohns Colitis* 2019; 13(3): 273–284.
36. Waljee AK, Sauder K, Patel A, et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. *J Crohns Colitis* 2017; 11(7): 801–810.
37. Mishra N, Aden K, Blase JI, et al. Longitudinal multi-omics analysis identifies early blood-based predictors of anti-TNF therapy response in inflammatory bowel disease. *Genome Med* 2022; 14(1): 110.
38. Feng J, Chen Y, Feng Q, et al. Novel gene signatures predicting primary non-response to infliximab in ulcerative colitis: development and validation combining random forest with artificial neural network. *Front Med* 2021; 8: 678424.
39. Obratsov IV, Shirokikh KE, Obratsova OI, et al. Multiple cytokine profiling: a new model to predict response to tumor necrosis factor antagonists in ulcerative colitis patients. *Inflamm Bowel Dis* 2019; 25(3): 524–531.
40. Chen X, Jiang L, Han W, et al. Artificial neural network analysis-based immune-related signatures of primary non-response to infliximab in patients with ulcerative colitis. *Front Immunol* 2021; 12: 742080.
41. Miyoshi J, Maeda T, Matsuoka K, et al. Machine learning using clinical data at baseline predicts

- the efficacy of vedolizumab at week 22 in patients with ulcerative colitis. *Sci Rep* 2021; 11(1): 16440.
42. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther* 2018; 47(6): 763–772.
 43. de Boer NKH, Peyrin-Biroulet L, Jharap B, et al. Thiopurines in inflammatory bowel disease: new findings and perspectives. *J Crohns Colitis* 2018; 12(5): 610–620.
 44. Dassopoulos T, Dubinsky MC, Bentsen JL, et al. Randomised clinical trial: individualized versus weight-based dosing of azathioprine in crohn's disease. *Aliment Pharmacol Ther* 2014; 39(2): 163–175.
 45. Osterman MT, Kundu R, Lichtenstein GR, et al. Association of 6-thioguanine nucleotide levels and inflammatory bowel disease activity: a meta-analysis. *Gastroenterology* 2006; 130(4): 1047–1053.
 46. Li X, Yan L, Wang X, et al. Predictive models for endoscopic disease activity in patients with ulcerative colitis: practical machine learning-based modeling and interpretation. *Front Med* 2022; 9: 1043412.
 47. Popa IV, Burlacu A, Mihai C, et al. A machine learning model accurately predicts ulcerative colitis activity at one year in patients treated with anti-tumour necrosis factor α agents. *Medicina (Kaunas)* 2020; 56(11): 628.
 48. Chen J, Girard M, Wang S, et al. Using supervised machine learning approach to predict treatment outcomes of vedolizumab in ulcerative colitis patients. *J Biopharm Stat* 2022; 32(2): 330–345.
 49. Ozawa T, Ishihara S, Fujishiro M, et al. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019; 89(2): 416–421.e1.
 50. Stidham RW, Liu W, Bishu S, et al. Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Netw Open* 2019; 2(5): e193963.
 51. Yao H, Najarian K, Gryak J, et al. Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc* 2021; 93(3): 728–736.e1.
 52. Gutierrez Becker B, Arcadu F, Thalhammer A, et al. Training and deploying a deep learning model for endoscopic severity grading in ulcerative colitis using multicenter clinical trial data. *Ther Adv Gastrointest Endosc* 2021; 14: 2631774521990623.
 53. Fan Y, Mu R, Xu H, et al. Novel deep learning-based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis. *Gastrointest Endosc* 2023; 97(2): 335–346.
 54. Maeda Y, Kudo S-E, Ogata N, et al. Evaluation in real-time use of artificial intelligence during colonoscopy to predict relapse of ulcerative colitis: a prospective study. *Gastrointest Endosc* 2022; 95(4): 747–756.e2.
 55. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology* 2020; 158(8): 2150–2157.
 56. Maeda Y, Kudo SE, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019; 89(2): 408–415.
 57. Bossuyt P, Vermeire S and Bisschops R. Scoring endoscopic disease activity in IBD: artificial intelligence sees more and better than we do. *Gut* 2020; 69(4): 788–789.
 58. Iacucci M, Cannatelli R, Parigi TL, et al. A virtual chromoendoscopy artificial intelligence system to detect endoscopic and histologic activity/remission and predict clinical outcomes in ulcerative colitis. *Endoscopy* 2023; 55(4): 332–341.
 59. Vande Castele N, Leighton JA, Pasha SF, et al. Utilizing deep learning to analyze whole slide images of colonic biopsies for associations between eosinophil density and clinicopathologic features in active ulcerative colitis. *Inflamm Bowel Dis* 2022; 28(4): 539–546.
 60. Gui X, Bazarova A, Del Amor R, et al. PICaSSO Histologic Remission Index (PHRI) in ulcerative colitis: development of a novel simplified histological score for monitoring mucosal healing and predicting clinical outcomes and its applicability in an artificial intelligence system. *Gut* 2022; 71(5): 889–898.
 61. Peyrin-Biroulet L, Adsul S, Dehmeshki J, et al. DOP58 an artificial intelligence-driven scoring system to measure histological disease activity in Ulcerative Colitis. *J Crohns Colitis* 2022; 16(Suppl_1): i105.
 62. Morilla I, Uzzan M, Laharie D, et al. Colonic microRNA profiles, identified by a deep learning

- algorithm, that predict responses to therapy of patients with acute severe ulcerative colitis. *Clin Gastroenterol Hepatol* 2019; 17(5): 905–913.
63. He T, Wang K, Zhao P, et al. Integrative computational approach identifies immune-relevant biomarkers in ulcerative colitis. *FEBS Open Bio* 2022; 12(2): 500–515.
 64. Biasci D, Lee JC, Noor NM, et al. A blood-based prognostic biomarker in IBD. *Gut* 2019; 68(8): 1386–1395.
 65. Lai L, Li H, Feng Q, et al. Multi-factor mediated functional modules identify novel classification of ulcerative colitis and functional gene panel. *Sci Rep* 2021; 11(1): 5669.
 66. Li X, Tang Z, Liu Y, et al. Risk prediction model based on blood biomarkers for predicting moderate to severe endoscopic activity in patients with ulcerative colitis. *Front Med (Lausanne)* 2023; 10: 1101237.
 67. Gazouli M, Dovrolis N, Franke A, et al. Differential genetic and functional background in inflammatory bowel disease phenotypes of a Greek population: a systems bioinformatics approach. *Gut Pathog* 2019; 11: 31.
 68. Bakir-Gungor B, Hacilar H, Jabeer A, et al. Inflammatory bowel disease biomarkers of human gut microbiota selected via different feature selection methods. *PeerJ* 2022; 10: e13205.
 69. Stidham R, Cai L, Cheng S, et al. Using computer vision to improve endoscopic disease quantification in therapeutic clinical trials of ulcerative colitis. *Gastroenterology* 2023; 166(1): 155–167.
 70. Najdawi F, Sucipto K, Mistry P, et al. Artificial intelligence enables quantitative assessment of ulcerative colitis histology. *Mod Pathol* 2023; 36(6): 100–124.
 71. Iacucci M, Parigi TL, Del Amor R, et al. Artificial intelligence enabled histologic prediction of remission or activity and clinical outcomes in ulcerative colitis. *Gastroenterology* 2023; 164(7): 1190–1188.
 72. Mohammed Vashist N, Samaan M, Mosli MH, et al. Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis. *Cochrane Database Syst Rev* 2018; 1(1): CD011450.
 73. Bryant RV, Burger DC, Delo J, et al. Beyond endoscopic mucosal healing in UC: histological remission better predicts corticosteroid use and hospitalisation over 6 years of follow-up. *Gut* 2016; 65(3): 408–414.
 74. Kaneshiro M, Takenaka K, Suzuki K, et al. Pancolonic endoscopic and histologic evaluation for relapse prediction in patients with ulcerative colitis in clinical remission. *Aliment Pharmacol Ther* 2021; 53(8): 900–907.
 75. Cushing KC, Tan W, Alpers DH, et al. Complete histologic normalization is associated with reduced risk of relapse among patients with ulcerative colitis in complete endoscopic remission. *Aliment Pharmacol Ther* 2020; 51(3): 347–355.
 76. Takenaka K, Ohtsuka K, Fujii T, et al. Deep neural network accurately predicts prognosis of ulcerative colitis using endoscopic images. *Gastroenterology* 2021; 160(6): 2175–2177.e3.
 77. Noguchi T, Ando T, Emoto S, et al. Artificial intelligence program to predict p53 mutations in ulcerative colitis-associated cancer or dysplasia. *Inflamm Bowel Dis* 2022; 28(7): 1072–1080.
 78. Yu S, Li H, Li Y, et al. Development and validation of novel models for the prediction of intravenous corticosteroid resistance in acute severe ulcerative colitis using logistic regression and machine learning. *Gastroenterol Rep* 2022; 10: goac053.
 79. Takayama T, Okamoto S, Hisamatsu T, et al. Computer-aided prediction of long-term prognosis of patients with ulcerative colitis after cytoapheresis therapy. *PLOS One* 2015; 10(6): e0131197.
 80. Sofo L, Caprino P, Schena CA, et al. New perspectives in the prediction of postoperative complications for high-risk ulcerative colitis patients: machine learning preliminary approach. *Eur Rev Med Pharmacol Sci* 2020; 24(24): 12781–12787.
 81. Mizuno S, Okabayashi K, Ikebata A, et al. Prediction of pouchitis after ileal pouch–anal anastomosis in patients with ulcerative colitis using artificial intelligence and deep learning. *Tech Coloproctol* 2022; 26(6): 471–478.
 82. Roy S, Sheikh SZ and Furey TS. A machine learning approach identifies 5-ASA and ulcerative colitis as being linked with higher COVID-19 mortality in patients with IBD. *Sci Rep* 2021; 11(1): 16522.
 83. Eaden JA, Abrams KR and Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 2001; 48(4): 526–535.
 84. Uttam S, Hashash JG, LaFace J, et al. Three-dimensional nanoscale nuclear architecture mapping of rectal biopsies detects colorectal neoplasia in patients with inflammatory bowel disease. *Cancer Prev Res* 2019; 12(8): 527–538.