**EVIDENCE REVIEW**

1847  Pancreatic adenocarcinoma: Beyond first line, where are we?

Cherri S, Noventa S, Zaniboni A

**REVIEW**

1864  Non-alcoholic fatty liver and chronic kidney disease: Retrospect, introspect, and prospect

Heda R, Yazawa M, Shi M, Bhaskaran M, Aloor FZ, Thuluvath PJ, Satapathy SK

1883  Repurposing metformin for the treatment of gastrointestinal cancer

Cunha Junior AD, Bragagnoli AC, Costa FO, Carvalheira JBC

1905  Mesenchymal stromal cell secretome in liver failure: Perspectives on COVID-19 infection treatment


**MINIREVIEWS**

1920  Artificial intelligence applications in inflammatory bowel disease: Emerging technologies and future directions

Gubatan J, Levitte S, Patel A, Balabanis T, Wei MT, Sinha SR

1936  Post-pancreatitis diabetes mellitus and excess intra-pancreatic fat deposition as harbingers of pancreatic cancer

Petrov MS

1943  BRCA mutated pancreatic cancer: A change is coming

Rosen MN, Goodwin RA, Vickers MM

1959  Metabolic and cardiovascular complications after virological cure in hepatitis C: What awaits beyond

Shengir M, Elgara M, Sebastiani G

**ORIGINAL ARTICLE**

**Basic Study**

1973  S100 calcium binding protein A6 and associated long noncoding ribonucleic acids as biomarkers in the diagnosis and staging of primary biliary cholangitis

Dong XH, Dai D, Yang ZD, Yu XO, Li H, Kang H

1993  Long non-coding RNA TP73-AS1 promotes pancreatic cancer growth and metastasis through miRNA-128-3p/GOLM1 axis

Wang B, Sun X, Huang KJ, Zhou LS, Qiu ZJ
Retrospective Study

2015 Prediction of microvascular invasion in solitary hepatocellular carcinoma ≤ 5 cm based on computed tomography radiomics
Liu P, Tan XZ, Zhang T, Gu QB, Mao XH, Li YC, He YQ

2025 Hepatocellular carcinoma progression in hepatitis B virus-related cirrhosis patients receiving nucleoside (acid) analogs therapy: A retrospective cross-sectional study
Yang DH, Wang WP, Zhang Q, Pan HY, Huang YC, Zhang JJ

Observational Study

2039 Hypophosphatemia after high-dose intravenous iron treatment in patients with inflammatory bowel disease: Mechanisms and possible clinical impact
Detlie TE, Lindstrom JC, Jahnsen ME, Finnes E, Zoller H, Moum B, Jahnsen J
ABOUT COVER
Editorial Board Member of World Journal of Gastroenterology, Abhinav Vasudevan, Bachelor of Medicine, MPH, PhD, FRACP, Advanced Inflammatory Bowel Disease Fellow, Mayo Clinic, 200 1st Street SW, Rochester, MN 55902, United States. vasudevan.abhinav@mayo.edu

AIMS AND SCOPE
The primary aim of World Journal of Gastroenterology (WJG, World J Gastroenterol) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online. WJG mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

INDEXING/ABSTRACTING
The WJG is now indexed in Current Contents®/Clinical Medicine, Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports®, Index Medicus, MEDLINE, PubMed, PubMed Central, and Scopus. The 2020 edition of Journal Citation Report® cites the 2019 impact factor (IF) for WJG as 3.665; IF without journal self cites: 3.534; 5-year IF: 4.048; Ranking: 35 among 88 journals in gastroenterology and hepatology; and Quartile category: Q2. The WJG’s CiteScore for 2019 is 7.1 and Scopus CiteScore rank 2019: Gastroenterology is 17/137.

RESPONSIBLE EDITORS FOR THIS ISSUE
Production Editor: Yu-Jie Ma; Production Department Director: Xiang Li; Editorial Office Director: Ze-Mao Gong.

NAME OF JOURNAL
World Journal of Gastroenterology

ISSN
ISSN 1007-9327 (print) ISSN 2219-2840 (online)

LAUNCH DATE
October 1, 1995

FREQUENCY
Weekly

EDITORS-IN-CHIEF
Andrzej S Tarnawski, Subrata Ghosh

EDITORIAL BOARD MEMBERS
http://www.wjgnet.com/1007-9327/editorialboard.htm

PUBLICATION DATE
May 7, 2021

COPYRIGHT
© 2021 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS
https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS
https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH
https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS
https://www.wjgnet.com/bpg/gerinfo/288

PUBLICATION MISCONDUCT
https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE
https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS
https://www.wjgnet.com/bpg/gerinfo/239

ONLINE SUBMISSION
https://www.f6publishing.com

© 2021 Baishideng Publishing Group Inc. All rights reserved. 7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
E-mail: bpgoffice@wjgnet.com https://www.wjgnet.com
Artificial intelligence applications in inflammatory bowel disease: Emerging technologies and future directions

John Gubatan, Steven Levitte, Akshar Patel, Tatiana Balabanis, Mike T Wei, Sidhartha R Sinha

Abstract

Inflammatory bowel disease (IBD) is a complex and multifaceted disorder of the gastrointestinal tract that is increasing in incidence worldwide and associated with significant morbidity. The rapid accumulation of large datasets from electronic health records, high-definition multi-omics (including genomics, proteomics, transcriptomics, and metagenomics), and imaging modalities (endoscopy and endomicroscopy) have provided powerful tools to unravel novel mechanistic insights and help address unmet clinical needs in IBD. Although the application of artificial intelligence (AI) methods has facilitated the analysis, integration, and interpretation of large datasets in IBD, significant heterogeneity in AI methods, datasets, and clinical outcomes and the need for unbiased prospective validations studies are current barriers to incorporation of AI into clinical practice. The purpose of this review is to summarize the most recent advances in the application of AI and machine learning technologies in the diagnosis and risk prediction, assessment of disease severity, and prediction of clinical outcomes in patients with IBD.

Key Words: Artificial intelligence; Machine learning; Inflammatory bowel disease; Crohn’s disease; Ulcerative colitis; Clinical outcomes
INTRODUCTION

Inflammatory bowel disease (IBD), which includes Crohn’s disease (CD) and ulcerative colitis (UC), is a chronic inflammatory disorder of the gastrointestinal tract. IBD has emerged as a global disease with increasing incidence worldwide and associated with significant healthcare utilization[1,2]. The pathogenesis of IBD is complex and is thought to involve an interplay between loss of tolerance to commensal gut bacteria, intestinal epithelial barrier dysfunction, and immune dysregulation[3-7]. The diagnosis of IBD is based on a combination of factors including clinical data (e.g., chronicity of gastrointestinal symptoms), laboratory values (elevated inflammatory markers such as C-reactive protein and fecal calprotectin), imaging, endoscopy, and histology (gastrointestinal inflammation with architectural distortion)[8]. Although treatment algorithms based on clinical trials and experience have been developed to inform clinical management in IBD[9], there is significant heterogeneity among patients with IBD with regards to presentation, response to therapy, and long-term clinical outcomes such development of strictures and need for surgery[10,11]. There is a great need for precision medicine strategies to improve diagnostic and therapeutic approaches in IBD.

Precision medicine efforts in IBD have led to more in-depth phenotyping of patients with IBD using large scale databases from clinical trials and cohort studies, deep immunophenotyping using whole genome gene expression datasets, proteomics, transcriptomics, and metagenomics of gut microbiota, and complex predictive models incorporating computer-assisted analysis of endoscopic images and histology[12-14]. This has inevitably led to vast arrays of high dimensional data that pose significant challenges with traditional statistical and computational methods[15]. Technological advances in artificial intelligence (AI) have revolutionized the ability of clinicians and researchers to process, analyze, and interpret high dimensional data and large datasets.

AI is a broad and multidisciplinary field incorporating concepts from computer science, engineering, philosophy, and linguistics aimed at understanding and designing systems that display or mimic human intelligence. The term was first coined in 1965 by McCarthy [16,17]. Machine learning (ML) is a subdiscipline of AI where computer algorithms apply statistical models to learn associations of predictive power from examples in provided datasets (e.g., Dragon dictation, SPAM, Netflix). ML may be programmed through supervised learning or unsupervised learning. In supervising learning, computer programs are trained to learn associations between inputs and outputs in data through analysis of predefined outputs of interest (by human operator). Once associations have been learned using existing data, supervised ML classifiers could then be used to predict future examples using different datasets. Examples of supervised ML include random forest (RF) and support vector machines (SVM). In unsupervised learning, computer programs learn associations in data without external definitions of associations of interest. This method allows for the identification of previously undiscovered predictors. Deep learning, commonly known as neural networks, includes newer techniques that are based on models with fewer assumptions, rely on multiple layers of representation of the data with successive transformations that amplify aspects of the input which improves discrimination power and thus able to handle more complex data (e.g., Facebook face recognition, credit card fraud)[17]. There has been increased interest in use of AI in IBD in recent years.

DOI: https://dx.doi.org/10.3748/wjg.v27.i17.1920
years with many prior groups applying ML methods to identify meaningful insights in diagnostics and prediction models in IBD. The purpose of this review is to provide a comprehensive summary of advances in the application of AI and ML technologies in the diagnosis and risk prediction, assessment of disease severity, and prediction of clinical outcomes in patients with IBD.

LITERATURE SEARCH

We performed a literature review using PubMed (MEDLINE) from inception to December 15, 2020 of studies applying AI in IBD. Our search strategy included the following combinations: (((((((((inflammatory bowel disease>Title)) OR (ulcerative colitis>Title)) OR (Crohn’s disease>Title)) AND (artificial intelligence>Title)) OR (computer-assisted>Title)) OR (computer-aided>Title)) OR (neural network>Title)) OR (machine learning>Title)) OR (deep learning>Title)). We included studies that used AI in the (1) diagnosis or risk prediction of IBD, (2) assessment of disease severity in IBD, and (3) prediction of therapy response and clinical outcomes in IBD. 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., measures of precision, accuracy, area under the curve (AUC), sensitivity, specificity, etc.).

RESULTS

Our search strategy yielded 98 studies evaluating AI in IBD of which 58 studies[18-74] met inclusion criteria and were included in the final review. About 86.2% (50/58) of studies were published within the past 5 years (2015 and later). There were 23 studies[18-39] that focused on IBD diagnosis and risk prediction, 19 studies[40-58] which evaluated disease activity, and 17 studies[45,59-74] which predicted IBD clinical outcomes (response to therapy, colonic neoplasia, post-surgical complications, quality of life, IBD well-being and emotional content). There were 22 studies with combined IBD cohorts (CD and UC), 16 studies with UC patients only, 18 studies with CD only, and 5 pediatric IBD cohorts. The most common AI classifications used were neural networks (convolutional and deep) at 32.7% (19/58 studies), RF at 29.3% (17/58 studies), and SVM at 29.3% (17/58 studies).

AI in diagnosis and risk prediction of IBD

Table 1 summarizes studies included which applied AI in the diagnosis and risk prediction of IBD. There were 17 studies focused on IBD diagnosis, whereas 5 studies focused on predicting risk of IBD. Data modalities included genetic/genomic datasets (n = 16 studies), imaging and endoscopic datasets (n = 4), and protein expression/proteomics (n = 2 studies). Some groups have used ML to develop IBD risk prediction models based on gene expression datasets. In a cross-sectional study of 180 CD patients, 149 UC patients and 90 healthy controls by Isakov et al[21], RF and SVM used microarray and RNA-seq data sets to classify a list of 16390 genes. Their combined IBD risk prediction model demonstrated an AUC, sensitivity, specificity, and accuracy values of 0.829, 0.577, 0.880, and 0.808, respectively. In another cross-sectional study of 18227 CD patients and 34050 healthy controls, Romagnoni et al[20] used gradient boosted trees and artificial neural networks to analyze gene expression profiles. Using single nucleotide polymorphisms, their final predictive model for CD achieved AUC of 0.80. Likewise, a cross-sectional study of 20 UC patients and 20 healthy controls by Duttagupta et al[33] used SVM to analyze microRNA profiles. Their SVM classifier measurements revealed a predictive score accuracy of 92.8%, specificity of 96.2%, and sensitivity of 89.5% in distinguishing UC patients from normal individuals.

A major challenge in IBD diagnosis is the distinction between CD and UC which is based on clinical features such as the distribution of inflammation along the gastrointestinal tract. The misdiagnosis of IBD subtype is not uncommon[74]. Distinguishing between CD and UC is clinically important as IBD subtype informs clinical management. AI has been employed to analyze molecular data to distinguish between CD and UC. In a cross-sectional study of 59 CD patients, 26 UC patients, and 42 healthy controls applying deep belief networks (DBNs) and SVM to gene expression datasets, Smolander et al[25] explored the diagnosis UC from CD. Using DBN only, the accuracy for diagnosis of UC was 97.06% and CD was 97.07%. Using both DBN and SVM, accuracy for diagnosis of UC was 97.06% and CD was 97.03%.
<table>
<thead>
<tr>
<th>Ref.</th>
<th>AI classifier vs comparator</th>
<th>IBD type</th>
<th>Study design and sample size</th>
<th>Modality</th>
<th>Outcome</th>
<th>Study results/validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mosotto et al [18], 2017</td>
<td>Support vector machines (SVM) vs linear discriminant</td>
<td>CD/UC</td>
<td>Prospective cohort, 287 IBD patients</td>
<td>Endoscopic and histologic inflammation</td>
<td>Diagnosis of IBD</td>
<td>Diagnostic accuracy of 82.7% with an AUC of 0.87 in diagnosing Crohn's disease or ulcerative colitis. Validation cohort included</td>
</tr>
<tr>
<td>Wei et al [19], 2013</td>
<td>SVM with gradient boosted trees (GBT) vs simple log odds method</td>
<td>CD/UC</td>
<td>Cross-sectional, 30,000 IBD patients, 22,000 healthy controls</td>
<td>Genetics, ImmunoChip</td>
<td>Risk of IBD</td>
<td>The SVM demonstrated very comparable performance (AUC 0.862 and 0.826 for CD and UC, respectively), whereas GBT showed inferior performance (AUC 0.802 and 0.782 for CD and UC, respectively. Validation cohort included</td>
</tr>
<tr>
<td>Romagnoni et al [20], 2019</td>
<td>Artificial neural networks (ANNs) vs penalized logistic regression (LR), and GBT</td>
<td>CD</td>
<td>Cross-sectional, 18,227 CD patients, 34,000 healthy controls</td>
<td>Genetics, ImmunoChip</td>
<td>Risk of IBD</td>
<td>Using single nucleotide polymorphisms (SNPs), final predictive model achieved AUC of 0.80. Validation cohort included</td>
</tr>
<tr>
<td>Isakov et al [21], 2017</td>
<td>Random forest (RF), SVM with svmPoly, extreme gradient boosting vs elastic net regularized generalized linear model (glmnet)</td>
<td>CD/UC</td>
<td>Cross-sectional, 180 CD patients, 149 UC patients, 90 healthy controls</td>
<td>Expression data (microarray and RNA-seq)</td>
<td>Risk of IBD</td>
<td>The method was used to classify a list of 16,390 genes. Each gene received a score that was used to prioritize it according to its predicted association to IBD. The combined model demonstrated AUC, sensitivity, specificity, and accuracy values of 0.829, 0.577, 0.88, and 0.808, respectively. Validation cohort included</td>
</tr>
<tr>
<td>Yuan et al [22], 2017</td>
<td>Sequential minimal optimization vs DisGeNET (Version 4.0)</td>
<td>CD/UC</td>
<td>Cross-sectional, 59 CD patients, 26 UC patients, 42 healthy controls</td>
<td>Gene Expression datasets</td>
<td>Risk of IBD</td>
<td>By analyzing the gene expression profiles using minimum redundancy maximum relevance and incremental feature selection, 21 genes were obtained that could effectively distinguish samples from IBD and the non-IBD samples. Highest total prediction accuracy was 97.64% using the 1170th feature set. Validation cohort included</td>
</tr>
<tr>
<td>Hübenthal et al [23], 2015</td>
<td>SVM vs RF</td>
<td>CD/UC</td>
<td>Cross-sectional, 40 CD patients, 36 UC patients, 38 healthy controls</td>
<td>MicroRNAs</td>
<td>Diagnosis of IBD</td>
<td>Measured by the AUC the corresponding median holdout-validated accuracy was estimated as ranging from 0.75 to 1.00 and 0.89 to 0.98, respectively. In combination, the corresponding models provide tools for the distinction of CD and UC as well as CD, UC and healthy control with expected classification error rates of 3.1 and 3.3%, respectively. Validation cohort included</td>
</tr>
<tr>
<td>Tong et al [24], 2020</td>
<td>RF vs convolutional neural network (CNN)</td>
<td>CD/UC</td>
<td>Retrospective Cohort, 875 CD patients, 5128 UC patients</td>
<td>Colonoscopy Endoscopic Images</td>
<td>Diagnosis of IBD</td>
<td>RF sensitivities/specificities of UC/CD were 0.89/0.84, 0.83/0.82, and 0.72/0.77, respectively, while the values for the CNN of CD was 0.90/0.77. The precision/recall of UC-CD when employing RF were 0.97/0.97, 0.65/0.53, respectively, and when employing the CNN were 0.99/0.97 and 0.87/0.83, respectively. Validation cohort included</td>
</tr>
<tr>
<td>Smolander et al [25], 2019</td>
<td>Deep belief networks (DBNs) vs SVM</td>
<td>CD/UC</td>
<td>Cross-sectional, 59 CD patients, 26 UC patients, 42 healthy controls</td>
<td>Gene Expression datasets</td>
<td>Diagnosis of IBD</td>
<td>Using DBN only, accuracy for diagnosis of UC was 97.06% and CD was 97.07%. Using both DBN and SVM, accuracy for diagnosis of UC was 97.06% and CD was 97.03%. Validation cohort included</td>
</tr>
<tr>
<td>Abbas et al [26], 2019</td>
<td>RF vs network-based biomarker discovery</td>
<td>Ped/UC</td>
<td>Cross-sectional, 657 IBD patients, 316 healthy controls</td>
<td>Large dataset of newborn pediatric IBD metagenomics biopsy samples</td>
<td>Diagnosis of IBD</td>
<td>For the diagnosis of IBD, highest AUC attained by top Random Forest classifiers was 0.77. No validation cohort included</td>
</tr>
<tr>
<td>Khorasani et al [27], 2020</td>
<td>SVM vs recently developed feature selection algorithm (robustness-performance tradeoff, RPT)</td>
<td>UC</td>
<td>Cross-sectional, 146 UC patients, 60 healthy controls</td>
<td>Gene Expression dataset</td>
<td>Diagnosis of IBD</td>
<td>Our model perfectly detected all active cases and had an average precision of 0.62 in the inactive cases. Validation cohort included</td>
</tr>
<tr>
<td>Rubin et al [28], 2019</td>
<td>CITRUS supervised machine learning algorithm. No comparator</td>
<td>CD/UC</td>
<td>Cross-sectional, 68 IBD patients</td>
<td>Peripheral blood mononuclear cells and intestinal biopsies mass cytometry</td>
<td>Diagnosis of IBD</td>
<td>An 8-parameter immune signature distinguished Crohn's disease from ulcerative colitis with an AUC = 0.845 (95%CI: 0.742-0.948). No validation cohort included</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Methodology</td>
<td>Subjects</td>
<td>Data</td>
<td>Feature Extraction</td>
<td>Model</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>-------------</td>
<td>---------</td>
<td>------</td>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>Pal et al. [29], 2017</td>
<td>CD</td>
<td>Cross-sectional, 64 CD patients, 47 healthy controls</td>
<td>Genotypes from Exome Sequencing Data</td>
<td>Risk of IBD</td>
<td>The AUC for predicting risk of Crohn’s disease using the SNP model was 0.72. No validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Aoki et al. [30], 2019</td>
<td>CD</td>
<td>Retrospective Cohort, 115 IBD patients</td>
<td>Wireless capsule endoscopy images</td>
<td>Diagnosis of IBD</td>
<td>The AUC for the detection of erosions and ulcerations was 0.958 (95% CI: 0.947-0.968). The sensitivity, specificity, and accuracy of the CNN were 88.2% (95% CI: 84.8-91.0), 90.9% (95% CI: 90.3-91.4), and 90.8% (95% CI: 90.2-91.3), respectively. Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Bielecki et al. [31], 2012</td>
<td>CD/UC</td>
<td>Cross-sectional, 14 CD patients, 13 UC patients, 11 healthy controls</td>
<td>Raman spectroscopic imaging of epithelium cells</td>
<td>Diagnosis of IBD</td>
<td>Raman maps of human colon tissue sections were analyzed by utilizing innovative chemometric approaches. Using SVM, it was possible to separate between healthy control patients, patients with Crohn’s Disease, and patients with ulcerative colitis with an accuracy of 98.90%. No validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Cui et al. [32], 2013</td>
<td>CD/UC</td>
<td>Cross-sectional, 124 IBD patients, 99 healthy controls</td>
<td>16S rRNA gene analysis</td>
<td>Diagnosis of IBD</td>
<td>Selection level of 200 features results in the best leave-one-out cross-validation result (accuracy = 88%, sensitivity = 92%, specificity = 84%). Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Duttagupta et al. [33], 2012</td>
<td>UC</td>
<td>Cross-sectional, 20 UC patients, 20 healthy controls</td>
<td>MicroRNAs</td>
<td>Diagnosis of IBD</td>
<td>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. Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Daneshjou et al. [34], 2017</td>
<td>CD</td>
<td>Cross-sectional, 64 ICD patients, 47 healthy controls</td>
<td>Exome Sequencing</td>
<td>Diagnosis of IBD</td>
<td>In CAGI4, 111 exomes were derived from a mix of 64 Crohn’s disease patients. Top performing methods had an AUC of 0.87. Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Geurts et al. [35], 2005</td>
<td>RF vs SVM</td>
<td>Prospective cohort, 30 CD patients, 30 CD patients</td>
<td>Proteomic Mass Spectrometry</td>
<td>Diagnosis of IBD</td>
<td>Random forest model to diagnosis IBD had a sensitivity of 81.67%, specificity of 81.17%. Support vector machine model to diagnosis IBD had a sensitivity of 87.92%, specificity of 87.87%. Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Li et al. [36], 2020</td>
<td>RF vs ANN</td>
<td>Cross-sectional, 193 UC patients, 21 healthy controls</td>
<td>Gene Expression Profiles</td>
<td>Diagnosis of IBD</td>
<td>The random forest algorithm was introduced to determine 1 downregulated and 29 upregulated differentially expressed genes contributing highest to ulcerative colitis occurrence. ANN was developed to calculate differentially expressed genes weights to ulcerative colitis. Prediction results agreed with that of an independent data set (AUC = 0.9506/PR-AUC = 0.9747). Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Wingfield et al. [37], 2019</td>
<td>RF vs SVM</td>
<td>Cross-sectional, 668 CD patients</td>
<td>Metagenomic Data</td>
<td>Diagnosis of IBD</td>
<td>Highest RPT measure for Crohn’s disease was random forest 0.60 and SVM 0.38. For ulcerative colitis, RPT was random forest 0.70 and SVM 0.48. Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Han et al. [38], 2018</td>
<td>RF vs LR, CORG</td>
<td>Cross-sectional, 24 CD patients, 59 UC patients, 76 healthy controls</td>
<td>Gene Expression Profiles</td>
<td>Diagnosis of IBD</td>
<td>The gene-based feature sets had median AUC on the validation sets ranging from 0.6 to 0.76. Validation cohort included</td>
<td></td>
</tr>
<tr>
<td>Wang et al. [39], 2019</td>
<td>AVADx (Analysis of Variation for Association with Disease) vs two GWAS-based CD evaluation methods</td>
<td>Cross-sectional, 64 CD patients, 47 healthy controls</td>
<td>Whole Exome or Genome Sequencing Data</td>
<td>Diagnosis of IBD</td>
<td>AVADx highlighted known CD genes including NOD2and new potential CD genes. AVADx identified 16% (at strict cutoff) of CD patients at 99% precision and 58% of the patients (at default cutoff) with 82% precision in over 3000 individuals from separately sequenced panels. Validation cohort included</td>
<td></td>
</tr>
</tbody>
</table>

AI: Artificial intelligence; IBD: Inflammatory bowel disease; CD: Crohn’s disease; UC: Ulcerative colitis; AUC: Area under the curve.
a cross-sectional study of 68 IBD patients using a CITRUS supervised ML algorithm to analyze single cell immunophenotyping of peripheral blood mononuclear cells by mass cytometry, Rubin et al[28] demonstrated that an 8-parameter immune signature distinguished CD from UC with an AUC = 0.845 [95% confidence interval (CI): 0.742-0.948]. ML algorithms have also been applied to analyze large arrays of endoscopic images to differentiate between UC and CD. In a recent retrospective cohort study of 875 CD patients and 5128 UC patients by Tong et al[24] using RF and convolutional neural networks (CNNs) on endoscopic images, the precision of diagnosing UC/CD with RF and CNNs were 0.97/0.65 and 0.99/0.87, respectively. Taken together, these studies suggest that AI classifiers have high performance in diagnosing or predicting risk of IBD but have some variability with type of AI classifier and modality of data (molecular vs endoscopic).

**AI in assessment of disease severity in IBD**

The assessment of disease activity and grading of severity in IBD could be accomplished using validated clinical symptom scores (e.g., Harvey Bradshaw Index for CD, Mayo Score for UC)[75,76], biomarkers of inflammation (e.g., C-reactive protein, fecal calprotectin)[77,78], endoscopic inflammation indices (e.g., Mayo endoscopic score, simple endoscopic score)[79,80], and histologic scoring systems (e.g., Geboes Score, Robarts Histopathology Index)[81,82]. However, these systems may be subject to recall bias, heterogeneity in patient clinical presentation, and intraobserver and interobserver variability[83]. AI has been applied to these existing systems to improve precision and accuracy of quantifying disease severity in IBD.

Table 2 summarizes studies included which applied AI in the assessment of disease severity in IBD. There were 2 studies that assessed clinical disease activity, 2 studies that assessed disease activity by biomarker (C-reactive protein), 13 studies that focused on endoscopic inflammation, and 3 studies that focused on histologic inflammation. Data modalities included electronic health records (n = 2), molecular datasets (n = 3), endoscopic datasets (n = 11 studies), and histologic datasets via endomicroscopy/endoctyscopy (n = 2). Using RF to integrate and analyze clinical and laboratory data from publicly available clinical trials (UNITI-I, UNITI-2, and IM-UNITI) data consisting of 401 CD patients, Waljee et al[42] constructed a CD remission prediction model using the week 6 albumin to C-reactive protein ratio with an AUC of 0.76 (95%CI: 0.71-0.82). Reddy et al[44] applied gradient boosting machines to electronic health records and predicted inflammation severity in a retrospective cohort of 3335 CD patients with a very high accuracy (AUC) = 92.82%. In a CNN analysis of colonoscopy images from a retrospective cohort of 841 UC patients by Ozawa et al[55], the CNN-based computer aided diagnostic system showed a high level of performance with AUC of 0.86 and 0.98 to identify Mayo 0 and 0-1, respectively. The performance of the CNN was better for the rectum than for the right side and left side of the colon when identifying Mayo 0 (AUC = 0.92, 0.83, and 0.83, respectively). Likewise, in an ordinal CNN analysis of wireless capsule endoscopy images in a retrospective cohort of 49 CD patients by Barash et al[50], the classification accuracy of the algorithm was 0.91 for grade 1 vs grade 3 ulcers, 0.78 for grade 2 vs grade 3, and 0.624 for grade 1 vs grade 2. The role of AI in grading severity of histologic inflammation in IBD has also been explored. For example, in a retrospective cohort study of 187 UC patients by Maeda et al[46], application of SVM to data derived from endocytoscopy to assess histologic inflammation provided diagnostic sensitivity, specificity, and accuracy of 74% (95%CI: 65-81), 97% (95%CI: 95-99), and 91% (95%CI: 83-95), respectively. These examples highlight the clinical utility, versatility, and performance of AI classifiers in grading the disease activity of IBD patients at the clinical, endoscopic, and histologic level. AI performance may be affected by location of inflammation and may be limited by ability to discriminate between subtle differences.

**AI in prediction of therapy response and clinical outcomes in IBD**

The armamentarium of therapies in IBD have expanded significantly in recent years with diverse mechanisms of action ranging from biologics that inhibit proinflammatory cytokines (anti-tumor necrosis factor-α, anti-interleukin-12/23) and leukocyte trafficking to the gut (anti-α4β7) to small molecule inhibitors of the JAK-STAT signaling pathway[84-86]. Despite several IBD treatment options available to clinicians, there are no effective biomarkers or tools to predict response to therapy or to guide selection of alternative therapies after a failed response. Likewise, there is also an unmet clinical need to predict long term clinical outcomes in IBD such as colon cancer. To address these challenges, several groups have applied AI and ML algorithms to existing clinical and molecular datasets.
Table 2 Artificial Intelligence in assessment of disease severity in inflammatory bowel disease

<table>
<thead>
<tr>
<th>Ref.</th>
<th>AI classifier vs comparator</th>
<th>IBD type</th>
<th>Study design and sample size</th>
<th>Modality</th>
<th>Outcomes</th>
<th>Study results/validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar et al [40], 2012</td>
<td>Support vector machines (SVM) vs human observers</td>
<td>CD</td>
<td>Cross-sectional, 5000 images (number of patients not given)</td>
<td>Small bowel capsule endoscopy</td>
<td>Endoscopic Inflammation</td>
<td>Database of 47 studies including 50000 capsule endoscopy images evaluating severity of small bowel lesions. Method had good precision (&gt;90% for lesion detection) and recall (&gt;90%) for lesions of varying severity. Validation cohort included</td>
</tr>
<tr>
<td>Biasci et al [41], 2019</td>
<td>Logistic regression with an adaptive Elastic-Net penalty. No comparator</td>
<td>CD/UC</td>
<td>Prospective cohort, 118 IBD patients</td>
<td>Transcriptomics from purified CD8 T cells and/or whole blood</td>
<td>Disease severity, medication escalation</td>
<td>A 17-gene qPCR-based classifier stratified patients into two distinct subgroups. IBDhi patients experienced significantly more aggressive disease than IBDlo patients (analogous to IBD2), with earlier need for treatment escalation [HR 2.65 (CD), 3.12 (UC)] and more escalations over time [for multiple escalations within 18 months: sensitivity=72.7% (CD), 100% (UC); negative predictive value = 90.9% (CD), 100% (UC)]. Validation cohort included</td>
</tr>
<tr>
<td>Waljee et al [42], 2019</td>
<td>RF. No comparator</td>
<td>CD</td>
<td>Post-hoc analysis of prospective clinical trials, 401 CD patients</td>
<td>Clinical and laboratory data from publicly available clinical trials (UNITI-1, UNITI-2, and IM-UNITI)</td>
<td>Crohn’s disease remission, C-reactive protein &lt; 5 mg/L</td>
<td>A prediction model using the week-6 albumin to C-reactive protein ratio had an AUC of 0.76 [95% confidence interval (CI): 0.71-0.82]. Validation cohort included</td>
</tr>
<tr>
<td>Mahapatra et al [43], 2016</td>
<td>RF. No comparator</td>
<td>CD</td>
<td>Cross-sectional, 35 CD patients</td>
<td>Abdominal magnetic resonance imaging</td>
<td>Segmentation of diseased colon (intestinal inflammation)</td>
<td>Model segmentation accuracy ranged from 82.7% to 92.2%. Validation cohort included</td>
</tr>
<tr>
<td>Reddy et al [44], 2019</td>
<td>Gradient boosting machines vs logistic regression</td>
<td>CD</td>
<td>Retrospective, 3335 CD patients</td>
<td>Electronic medical record</td>
<td>Severity of intestinal inflammation (by C-reactive protein)</td>
<td>Machine-learning-based analytic methods such as gradient boosting machines can predict the inflammation severity with a very high accuracy (AUC = 92.82%). Validation cohort included</td>
</tr>
<tr>
<td>Douglas et al [45], 2018</td>
<td>RF. No comparator</td>
<td>Peds CD</td>
<td>Cross-sectional, 20 CD patients, 20 healthy controls</td>
<td>Shotgun metagenomics (MGS), 16S rRNA gene sequencing</td>
<td>Disease State (Relapse/Remission)</td>
<td>MGS modules significantly classified samples by disease state (accuracy = 68.4%, P = 0.043 and accuracy = 65.8%, P = 0.03, respectively), 165 datasets had a maximum accuracy of 68.4% and P = 0.016 based on strain level for disease state. Validation cohort included</td>
</tr>
<tr>
<td>Maeda et al [46], 2019</td>
<td>SVM vs human reader</td>
<td>UC</td>
<td>Retrospective cohort, 187 UC patients</td>
<td>Endocytoscopy</td>
<td>Histologic inflammation</td>
<td>Computer aided diagnosis (CAD) of histologic inflammation provided diagnostic sensitivity, specificity, and accuracy as follows: 74% (95%CI: 65-81), 97% (95%CI 95-99), and 91% (95%CI 83-95), respectively. Its reproducibility was perfect (k = 1). Validation cohort included</td>
</tr>
<tr>
<td>Charisis et al [47], 2016</td>
<td>SVM vs human reader</td>
<td>CD</td>
<td>Retrospective cohort, 13 CD patients</td>
<td>Wireless capsule endoscopy (WCE) images</td>
<td>Endoscopic Inflammation</td>
<td>Experimental results, along with comparison with other related efforts, have shown that the hybrid adaptive filtering [HAF-Differential Lacunarity (DLac) analysis (HAF-DLac)] via SVM approach evidently outperforms them in the field of WCE image analysis for automated lesion detection, providing higher classification results, up to 93.8% (accuracy), 95.2% (sensitivity), 92.4% (specificity) and 92.6% (precision). Validation cohort included</td>
</tr>
<tr>
<td>Klang et al [48], 2020</td>
<td>Convolutional neural network (CNN) vs human reader</td>
<td>CD</td>
<td>Retrospective cohort, 49 CD patients</td>
<td>WCE images</td>
<td>Endoscopic Inflammation</td>
<td>Dataset included 17640 CE images from 49 patients: 7391 images with mucosal ulcers and 10249 images of normal mucosa. For randomly split images results, AUC was 0.99 with accuracies ranging from 95.4% to 96.7%. For individual patient-level experiments, the AUCs were 0.94-0.99. Validation cohort included</td>
</tr>
<tr>
<td>Ungaro et al [49], 2021</td>
<td>Random survival forest. No comparator</td>
<td>Peds CD</td>
<td>Retrospective case-control, 265 peds CD patients</td>
<td>Protein biomarkers using a proximity extension assay (Olink Proteomics)</td>
<td>Penetrating and structuring complications</td>
<td>A model with 5 protein markers predicted penetrating complications with an AUC of 0.79 (95%CI: 0.76-0.82) compared to 0.69 (95%CI: 0.66-0.72) for serologies and 0.74 (95%CI: 0.71-0.77) for clinical variables. A model with 4 protein markers predicted structuring complications with an AUC of 0.68 (95%CI: 0.65-0.71) compared to 0.62 (95%CI: 0.59-0.65) for serologies and 0.52 (95%CI: 0.50-0.55) for clinical variables. Validation cohort included</td>
</tr>
</tbody>
</table>
CNN exhibited Dice similarity coefficient of 75% ± 18%, 81% ± 8%, and 97% ± 2% for the lumen, wall, and background, respectively. The extracted markers of wall thickness at the location of min radius ($P = 0.0013$) and the median value of relative contrast enhancement ($P = 0.0033$) could differentiate active and nonactive disease segments. Other extracted markers could differentiate between segments with strictures and segments without strictures ($P < 0.05$). Validation cohort included

The CNN-based CAD system showed a high level of performance with AUC of 0.86 and 0.98 to 1927

### Table 3

<table>
<thead>
<tr>
<th>Study</th>
<th>Methodology</th>
<th>Disease</th>
<th>Modality</th>
<th>Marker</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barash et al.[50], 2021</td>
<td>Ordinal CNN. No comparator</td>
<td>CD</td>
<td>Retrospective cohort, 49 CD patients</td>
<td>WCE images</td>
<td>Ulcer Severity Grading</td>
</tr>
<tr>
<td>Lamash et al.[51], 2019</td>
<td>CNN vs semi-supervised and active learning models</td>
<td>CD</td>
<td>Retrospective cohort, 23 CD patients</td>
<td>Magnetic resonance imaging</td>
<td>Active Crohn’s Disease</td>
</tr>
<tr>
<td>Takenaka et al.[52], 2020</td>
<td>Deep neural networks vs human reader (endoscopist)</td>
<td>UC</td>
<td>Prospective cohort, 2012 UC patients</td>
<td>Colonoscopy images</td>
<td>Endoscopic inflammation</td>
</tr>
<tr>
<td>Bossuyt et al.[53], 2020</td>
<td>Computer algorithm based on red density (RD) vs blinded central readers</td>
<td>UC</td>
<td>Prospective cohort, 29 UC patients, 6 healthy controls</td>
<td>Colonoscopy Images</td>
<td>Endoscopic and histologic inflammation</td>
</tr>
<tr>
<td>Bhambhvani et al.[54], 2021</td>
<td>CNN vs human reader (endoscopist)</td>
<td>UC</td>
<td>Retrospective cohort, 777 UC patients</td>
<td>Colonoscopy images</td>
<td>Mayo Endoscopic Scores (MES)</td>
</tr>
<tr>
<td>Ozawa et al.[55], 2019</td>
<td>CNN vs human reader (endoscopist)</td>
<td>UC</td>
<td>Retrospective cohort, 841 UC patients</td>
<td>Colonoscopy images</td>
<td>MES</td>
</tr>
<tr>
<td>Bossuyt et al.[56], 2021</td>
<td>Automated CAD Algorithm vs human reader</td>
<td>UC</td>
<td>Prospective cohort, 48 UC patients</td>
<td>Colonoscopy images with confocal laser endomicroscopy</td>
<td>Histologic Remission</td>
</tr>
<tr>
<td>Stidham et al.[57], 2019</td>
<td>CNN vs human reader</td>
<td>UC</td>
<td>Retrospective cohort, 3082 UC patients</td>
<td>Colonoscopy images</td>
<td>Endoscopy severity</td>
</tr>
<tr>
<td>Gottlieb et al.[58], 2021</td>
<td>Neural network vs human central reader</td>
<td>UC</td>
<td>Prospective cohort, 249 UC patients</td>
<td>Colonoscopy images</td>
<td>Endoscopy severity</td>
</tr>
</tbody>
</table>

AI: Artificial intelligence; IBD: Inflammatory bowel disease; CD: Crohn’s disease; UC: Ulcerative colitis; AUC: Area under the curve; NPV: Negative predictive value; PPV: Positive predictive value; qPCR: Quantitative real-time polymerase chain reaction; HR: Hazard ratio.
**Table 3 Artificial intelligence in prediction of therapy response and clinical outcomes in inflammatory bowel disease**

<table>
<thead>
<tr>
<th>Ref.</th>
<th>AI classifier vs comparator</th>
<th>IBD type</th>
<th>Study design and sample size</th>
<th>Modality</th>
<th>Outcomes</th>
<th>Study results/validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waljee et al [59], 2018</td>
<td>Random forest (RF). No comparator</td>
<td>CD/UC</td>
<td>Post-hoc analysis of prospective clinical trial, 594 CD patients</td>
<td>Veteran’s Health Administration Electronic Health Record (EHR)</td>
<td>Outpatient corticosteroids prescribed for IBD and inept patient hospitalizations associated with a diagnosis of IBD</td>
<td>AUC for the RF longitudinal model was 0.85 [95% CI: 0.84–0.85]. AUC for the RF longitudinal model using previous hospitalization or steroid use was 0.87 [95% CI: 0.87–0.88]. Validation cohort included</td>
</tr>
<tr>
<td>Uttam et al [60], 2019</td>
<td>Support vector machines (SVM) vs nanoscale nuclear architecture mapping (NanoNAM)</td>
<td>CD/UC</td>
<td>Prospective cohort, 103 IBD patients</td>
<td>3-dimensional NanoNAM of normal-appearing rectal biopsies</td>
<td>Colonic neoplasia</td>
<td>NanoNAM detects colonic neoplasia with an AUC of 0.87 ± 0.04, sensitivity of 0.81 ± 0.09, and specificity of 0.82 ± 0.07 in the independent validation set. Validation cohort included</td>
</tr>
<tr>
<td>Waljee et al [1], 2017</td>
<td>RF. No comparator</td>
<td>CD/UC</td>
<td>Retrospective cohort, 1080 IBD patients</td>
<td>EHR, lab values</td>
<td>Remission and clinical outcomes with thiopurines</td>
<td>AUC for algorithm-predicted remission in the validation set was 0.79 vs 0.49 for 6-TGN. The mean number of clinical events per year in patients with sustained algorithm-predicted remission (APR) was 1.08 vs 3.95 in those that did not have sustained APR (P &lt; 1 x 10^(-5)). Validation cohort included</td>
</tr>
<tr>
<td>Popa et al [62], 2020</td>
<td>Neural network model. No comparator</td>
<td>UC</td>
<td>Prospective cohort, 55 UC patients</td>
<td>Clinical and biological parameters and the endoscopic Mayo score</td>
<td>Disease activity after one year of anti-TNF treatment</td>
<td>The classifier achieved an excellent performance predicting the disease activity at one year with an accuracy of 90% and AUC 0.92 on the test set and an accuracy of 100% and an AUC of 1 on the validation set. Validation cohort included</td>
</tr>
<tr>
<td>Douglas et al [13], 2018</td>
<td>RF. No comparator</td>
<td>Peds CD</td>
<td>Cross-sectional, 20 CD patients, 20 healthy controls</td>
<td>Shotgun metagenomics (MGS), 16S rRNA gene sequencing</td>
<td>Response to induction therapy</td>
<td>16S genera were again the top dataset (accuracy = 77.8%; P = 0.008) for predicting response to therapy. MGS-strain (P = 0.029), genus (P = 0.013), and KEGG pathway (P = 0.018) datasets could also classify patients according to therapy response with accuracy = 72.2% for all three. Validation cohort included</td>
</tr>
<tr>
<td>Waljee et al [63], 2010</td>
<td>RF vs boosted trees, RuleFit</td>
<td>CD/UC</td>
<td>Cross-sectional, 774 IBD patients</td>
<td>EHR, lab values (tiopurine metabolites)</td>
<td>Response to thiopurine therapy</td>
<td>A RF algorithm using laboratory values and patient age differentiated clinical response from nonresponse in the model validation data set with an AUC of 0.856 (95% CI: 0.793–0.919). Validation cohort included</td>
</tr>
<tr>
<td>Menti et al [64], 2016</td>
<td>Naïve bayes vs Bayesian additive regression trees vs Bayesian networks</td>
<td>CD/UC</td>
<td>Retrospective cohort, 152 CD patients</td>
<td>Genomic DNA, genetic polymorphism</td>
<td>Presence of extra-intestinal manifestations in IBD patients</td>
<td>Bayesian networks achieved accuracy of 82% when considering only clinical factors and 89% when considering also genetic information, outperforming the other techniques. Validation cohort included</td>
</tr>
<tr>
<td>Waljee et al [65], 2017</td>
<td>RF vs baseline regression model</td>
<td>CD/UC</td>
<td>Retrospective cohort, 20368 IBD patients</td>
<td>EHR, lab values</td>
<td>Corticosteroid-free biologic remission with vedolizumab</td>
<td>The AUC for corticosteroid-free biologic remission at week 52 using baseline data was only 0.65 (95% CI: 0.53–0.77), but was 0.75 (95% CI: 0.64–0.86) with data through week 6 of vedolizumab. Validation cohort included</td>
</tr>
<tr>
<td>Morilla et al [66], 2019</td>
<td>Deep neural networks. No comparator</td>
<td>UC</td>
<td>Retrospective cohort, 47 UC patients</td>
<td>Colonic microRNA profiles</td>
<td>Responses to therapy</td>
<td>A deep neural network-based classifier identified 9 microRNAs plus 5 clinical factors, routinely recorded at time of hospital admission, that were associated with responses of patients to treatment. This panel discriminated responders to steroids from non-responders with 93% accuracy (AUC, 0.91). Three algorithms, based on microRNA levels, identified responders to infliximab vs non-responders (84% accuracy, AUC 0.82) and responders to cyclosporine vs non-responders (80% accuracy, AUC 0.79). Validation cohort included</td>
</tr>
<tr>
<td>Wang et al [67], 2020</td>
<td>Back-propagation neural network (BPNN), SVM vs logistic regression</td>
<td>CD</td>
<td>Cross-sectional, 446 CD patients</td>
<td>EHR</td>
<td>Medication nonadherence to maintenance therapy</td>
<td>The average classification accuracy and AUC of the three models were 85.9% and 0.912 for BPNN, and 87.7% and 0.930 for SVM, respectively. Validation cohort included</td>
</tr>
<tr>
<td>Bottiglia et al</td>
<td>Bayesian machine</td>
<td>CD/UC</td>
<td>Retrospective cohort, 55 UC patients</td>
<td>EHR, genetic</td>
<td>Presence of extra-intestinal manifestations</td>
<td>BMLTs had an AUC of 0.50 for classifying the presence of extra-intestinal manifestations. Validation cohort included</td>
</tr>
</tbody>
</table>
datasets (\(n = 4\)), and histologic data (\(n = 1\)). Waljee et al[59,65] and Popa et al[62] have previously applied RF classifiers to clinical data from electronic health records and laboratory values to predict response to various IBD therapies. In one study using data from a prospective clinical trial consisting of 594 CD patients[59], the AUC for a RF longitudinal model for predicting inpatient hospitalizations in IBD patients prescribed outpatient corticosteroids was 0.85 (95%CI: 0.84-0.85). Using a similar RF approach for predicting remission with thiopurine therapy in a prospective cohort of 55 UC patients yielded an AUC of 0.79[62]. Applying RF to data from a retrospective cohort of 20368 IBD patients with vedolizumab use yielded an AUC of 0.65 (95%CI: 0.53-0.77) for corticosteroid-free vedolizumab remission at week 52 using baseline data and an AUC of 0.75 (95%CI: 0.64-0.86) with data through week 6 of vedolizumab[65]. Molecular datasets have also been used to differentiate between responders and non-responders to various IBD therapies. For example, Morilla et al[66] used a deep neural network classifier to construct a predictive panel of colonic microRNAs for IBD therapies in a retrospective cohort of 47 UC patients. Their panel discriminated responders to steroids from non-responders with 93% accuracy (AUC, 0.91). In addition, three
algorithms, based on microRNA levels, identified responders to infliximab vs non-responders (84% accuracy, AUC 0.82) and responders to cyclosporine vs non-responders (80% accuracy, AUC 0.79). A more recent prospective study of 55 UC patients by Popa et al. [62] integrated clinical, laboratory, and endoscopic (Mayo scores) datasets using a neural network classifier to predict disease activity after one year of anti-tumor necrosis factor therapy in patients with UC. This classifier achieved an AUC of 0.92 for predicting the disease activity at one year on the test set and an AUC of 1.00 on the validation set. These studies suggest that AI classifiers may play a role in predicting clinical outcomes and response to specific therapies in patients with IBD. However, future clinical trials are needed to compare the efficacy of AI applications in IBD clinical management vs standard of care before incorporation into real life clinical practice.

Finally, AI algorithms have been previously applied to enhance the detection of colonic polyps [87] and distinguish among subtypes of neoplastic colorectal lesions [88] in the general population. Although patients with IBD who have extensive colitis have a significantly greater risk of colorectal cancer compared to the general population [89,90], there have been limited studies applying AI technologies to improve colorectal cancer surveillance or develop prediction risk models in patients with IBD. Most studies evaluating polyp detection have excluded IBD patients [91-93]. Our literature search yielded only one study applying AI for the detection of colonic neoplasia in IBD. Uttam et al. [60] employed support SVM to analyze 3-dimensional nanoscale nuclear architecture mapping (NanoNAM) of normal-appearing rectal biopsies in a prospective cohort of 103 IBD patients. In their study, NanoNAM detected colonic neoplasia with an AUC of 0.87 ± 0.04, sensitivity of 0.81 ± 0.09, and specificity of 0.82 ± 0.07 in the independent validation set. Further studies should focus on determining the clinical utility of incorporating AI methods to enhance standard of care surveillance in patients with IBD such as chromoendoscopy [94] and to develop predictive models for risks of colorectal malignancy in IBD patient populations.

CONCLUSION

In conclusion, our literature review has revealed that the applications of AI in IBD have significantly increased in recent years. Our review also highlighted that various AI classifiers may be applied to analyze and integrate large datasets ranging from clinical data from electronic health records, molecular data including gene expression and protein-based studies to a wide array of datasets consisting of endoscopic and histologic images. The application of AI has the potential to improve the accuracy and precision of predicting risk and diagnosis of IBD, assessing disease severity, and predicting outcomes with various IBD therapies. Currently, the application of AI methods in IBD has been limited to the research setting and has not yet been adopted in real life clinical practice. Furthermore, studies applying AI in the context of colorectal cancer surveillance or prediction in IBD are much needed. Given the current status of the field of AI in IBD, future directions should include: (1) Prospective validation of AI applications in IBD in independent cohorts as there is a risk of bias from internal training cohorts and potential limitations with generalizability; (2) Standardization of AI methods and comparative studies evaluating effect of heterogeneity from using different types of datasets on outcomes of interest; (3) Randomized controlled trials to determine whether application of AI in the clinical management of IBD improves clinical outcomes and could be translated into clinical practice; and (4) Randomized controlled trials to determine whether application of AI leads to greater clinical efficacy and cost-effectiveness compared to standard of care before incorporation into real life clinical practice.


6 van der Soot KWJ, Amini M, Peters V, Dijkstra G, Alizadeh BZ. Inflammatory Bowel Diseases: Review of Known Environmental Protective and Risk Factors Involv...[PMID: 28777999 DOI: 10.1097/MIB.0000000000001217]


14 Syed S, Siddham RW. Potential for Standardization and Automation for Pathology and Endoscopy in Inflammatory Bowel Disease. Inflamm Bowel Dis 2020, 26: 1490-1497 [PMID: 32869844 DOI: 10.1093/ibd/iaab211]


21 Isakov O, Dotan I, Ben-Shachar S. Machine Learning-Based Gene Prioritization Identifies Novel Candidate Risk Genes for Inflammatory Bowel Disease. Inflamm Bowel Dis 2017; 23: 1516-1523 [PMID: 28795970 DOI: 10.1097/MIB.0000000000001222]


Reddy BK, Delen D, Agrawal RK. Predicting and explaining inflammation in Crohn's disease


**Popa IV**, Burlacu A, Mihai C, Prelipcean CC. 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: 33235514 DOI: 10.3390/medicina56110628

Gubatan J et al. Artificial intelligence in IBD


Best WR. Predicting the Crohn's disease activity index from the Harvey-Bradshaw Index. Inflamm Bowel Dis 2006; 12: 304-310 [PMID: 16633052 DOI: 10.1097/01.MIB.0000215091.77492.2a]


85 Neurath MF. Current and emerging therapeutic targets for IBD. *Nat Rev Gastroenterol Hepatol* 2017; 14: 269-278 [PMID: 28144028 DOI: 10.1038/nrgastro.2016.209]


