Received: 16 May 2024 | Accepted: 15 August 2024

DOI: 10.1002/ueg2.12655

REVIEW ARTICLE

UNITED EUROPEAN GASTROENTEROLOGY ueg journal WILEY

Inflammatory bowel disease genomics, transcriptomics, proteomics and metagenomics meet artificial intelligence

Anna Lucia Cannarozzi[1] | Anna Latiano[1] | Luca Massimino[2] | Fabrizio Bossa[1] | Francesco Giuliani[3] | Matteo Riva[2] | Federica Ungaro[2] | Maria Guerra[1] | Anna Laura Di Brina[1] | Giuseppe Biscaglia[1] | Francesca Tavano[1] | Sonia Carparelli[1] | Gionata Fiorino[4] | Silvio Danese[5] | Francesco Perri[1] | Orazio Palmieri[1]

[1]Division of Gastroenterology and Endoscopy, Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy

[2]Gastroenterology and Digestive Endoscopy Department, IRCCS Ospedale San Raffaele, Milan, Italy

[3]Innovation & Research Unit, Fondazione IRCCS "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Italy

[4]Gastroenterology and Digestive Endoscopy, San Camillo-Forlanini Hospital, Rome, Italy

[5]Faculty of Medicine, Università Vita-Salute San Raffaele, Milan, Italy

Correspondence
Orazio Palmieri.
Email: o.palmieri@operapadrepio.it

Funding information
Italian Next Generation Eu Program, Grant/Award Number: PNRR-MAD-2022-12375729

Abstract

Various extrinsic and intrinsic factors such as drug exposures, antibiotic treatments, smoking, lifestyle, genetics, immune responses, and the gut microbiome characterize ulcerative colitis and Crohn's disease, collectively called inflammatory bowel disease (IBD). All these factors contribute to the complexity and heterogeneity of the disease etiology and pathogenesis leading to major challenges for the scientific community in improving management, medical treatments, genetic risk, and exposome impact. Understanding the interaction(s) among these factors and their effects on the immune system in IBD patients has prompted advances in multi-omics research, the development of new tools as part of system biology, and more recently, artificial intelligence (AI) approaches. These innovative approaches, supported by the availability of big data and large volumes of digital medical datasets, hold promise in better understanding the natural histories, predictors of disease development, severity, complications and treatment outcomes in complex diseases, providing decision support to doctors, and promising to bring us closer to the realization of the "precision medicine" paradigm. This review aims to provide an overview of current IBD omics based on both individual (genomics, transcriptomics, proteomics, metagenomics) and multi-omics levels, highlighting how AI can facilitate the integration of heterogeneous data to summarize our current understanding of the disease and to identify current gaps in knowledge to inform upcoming research in this field.

KEYWORDS
artificial intelligence, Crohn's disease, deep learning, genes, genetics, inflammatory bowel disease, machine learning, omics, pathogenesis, ulcerative colitis

INTRODUCTION

Inflammatory bowel disease (IBD) is a class of chronic inflammatory disorders of the gastrointestinal tract characterized by heterogeneous phenotypes, namely ulcerative colitis (UC)[1] and Crohn's disease (CD),[2] which exhibit consistent differences in clinical presentation and behavior, likely explained by variations in underlying pathogenetic mechanisms.

Over the past few decades, IBD has emerged as a global disease with an increasing incidence worldwide associated with increased morbidity, mortality, and substantial costs to the healthcare system.[3]

Various extrinsic and intrinsic factors characterize IBD, such as drug exposures, antibiotic treatments, smoking, lifestyle, genetics, immune responses and the gut microbiome.[4–11] All these factors contribute to the complexity and heterogeneity of the disease etiology and pathogenesis,[12–15] leading to major challenges for the scientific community in improving management, medical treatments, estimating genetic risk, and exposome impact on the evolution of the disease.

Understanding the interaction(s) among these factors and their effects on the immune system in IBD patients could better explain these complex diseases. This prompted advances in multi-omics research[16] and the development of new tools as part of system biology[17–20] and, more recently, approaches based on artificial intelligence (AI) techniques. Innovations in experimental and computational methodologies along with data generated by consumer-level technologies employed in daily life[21] have ushered in a new era of "big data" research enabling a better understanding of the natural histories, predictors of disease development, severity, complications, and treatment outcomes.[22–24] Promoted by the availability of large volumes of digitized medical datasets and by the advanced computational tools required for complex pattern-recognizing data, AI could be the key in the management of complex multifactorial diseases such as IBD. This consideration is encouraged by the promising results that AI has made it possible to obtain for other diseases, particularly cancers[25] and immune-mediated diseases,[26] where the use has certainly been more extensive than for IBD. The concept of AI revolutionizing medical practice is gradually gaining traction to provide decision support to medical experts, to analyze large and complex datasets, and to boost healthcare research, bringing us closer to the realization of "precision medicine."[23]

The IBD emerges as one of the most studied from various omics perspectives and with different analytical tools. Numerous genomic studies have identified common and rare genetic variants associated with CD and UC, although they have not yet provided definitive clues to the etiology, pathogenesis, localization, diagnosis, or therapy.[10,27–29] Consequently, scientific evidence is still lacking regarding the feasibility of obtaining practical guidelines for diagnosis or therapy solely through genetic evaluation independent of other factors.[30] To address this challenge, there is increasing interest in leveraging AI as an investigative tool in IBD research. This review aims to provide an overview of current multi-omics IBD studies integrated with computational methods, highlighting how AI can facilitate the integration of heterogeneous data to attain greater clinical and therapeutic benefits in the near future.

METHODS

This narrative review was conducted using the free Pubmed database from 2010 to 2024 employing mesh terms related to the disease under examination, such as "inflammatory bowel disease," "ulcerative colitis," and "Crohn's disease," separated by the Boolean operator "OR" and combined with the Boolean operator "AND" with mesh terms related to AI tools, such as "artificial intelligence," "machine learning," "deep learning," "artificial neural networks," "random forest," "decision trees," and "computational methods." This initial research permitted the retrieval of all studies on the understanding and prediction of IBD presence or features using AI strategies. Additional mesh terms, such as "omics," "genomics," "transcriptomics," "proteomics," and "metabolomics," separated by the Boolean operator "OR," were included. Lastly, other terms, such as "multiomic approach," "multiomic integration," and "multiomics strategy" were incorporated to identify works concerning the integrative analysis of various omics. Studies on AI applications not related to omics and multi-omics analyses in the field of IBD were excluded from this search, for example, studies of automatic image analysis for endoscopic, histological, or radiographical evaluations.

Regarding the collected studies, there were no restrictions on the number of subjects, type of AI strategy, and type of omics analyzed.

ARTIFICIAL INTELLIGENCE

AI is a broad and multidisciplinary field that encompasses concepts from computer science, engineering, philosophy, and linguistics,[31] including specific subsets namely machine learning (ML) and deep learning (DL) (Figure 1).

Its goal is to understand and design systems capable of performing tasks and solving problems that typically require human intelligence and decision-making.[32]

AI in medicine has evolved significantly since its inception in the 1950s, with notable advancements in the early 2000s due to DL; its applications in gastroenterology, particularly in endoscopy, began gaining traction around 2020.[33]

AI holds the ambition to revolutionize medical practice, particularly in managing complex diseases influenced by numerous interconnected biological components and events that traditional tools struggle to resolve.[34] Through AI, the analysis of big data can aid expert physicians in decision-making and uncover insights into multifactorial diseases.

For effective utilization, AI tools must logically acquire, store, and organize information, characterizing relationships within the context.[35] To mimic human-comparable understanding, judgment,

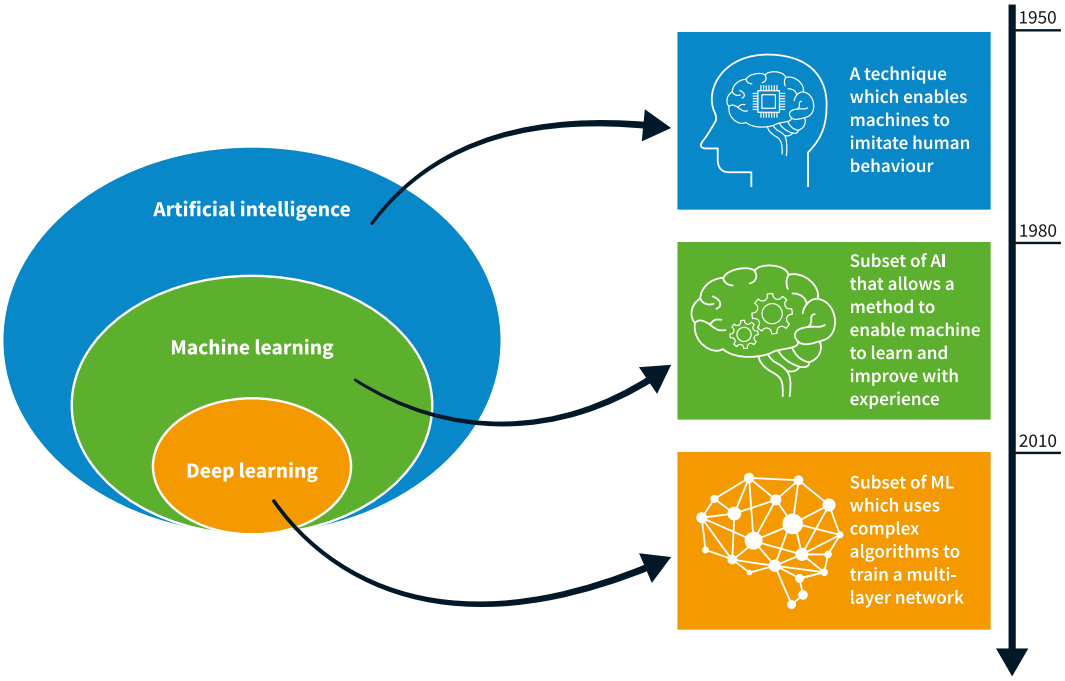

FIGURE 1 Typical graphical representation of primary concepts of artificial intelligence and their historical appearance. AI, artificial intelligence; ML, machine learning.

and prediction, AI relies on electronically stored data and analytical capabilities to discern patterns associated with clinical or pathological traits.

Machine learning

While "human physicians" continually improve their performance based on acquired knowledge and previously made errors that guide them in making clinical decisions, ML allows them to integrate in their mental process the results stemming from models of interaction among variables obtained from large datasets provided as input.[35] Such data or features are annotated with labels or classifications and can include clinical findings, research results, expert measurements conducted by physicians, and even physiological processes.

ML algorithms have the ability, based on previous experiences, to self-improve through training, which can occur principally in three ways: supervised, unsupervised, and reinforcement learning[36] (Figure 2).

Supervised learning entails training ML models using human-labeled input-output pairs to determine outputs for unlabeled inputs. This method often requires a large amount of labeled training data to generate meaningful predictions.[37]

With the aid of unsupervised learning algorithms, computers autonomously recognize associations or patterns in the data without relying on predefined labels. This method is particularly useful for uncovering underlying patterns within the data. Due to the absence of labels, unsupervised methods can be more computationally complex.[38]

Reinforcement learning is a learning process in which decision-making processes are directed toward achieving a given goal through interaction with the environment. In this framework, the machine does not receive explicit instructions on what to do but learns through trial and error, discerning which actions yield the greatest reward based on the context in which it finds itself.[39]

Machine learning process

A standard ML process consists of five main steps: (I) data collection from various sources, (II) data cleaning and feature engineering, (III) model assembly with the right ML algorithm selection, (IV) model evaluation, and (V) model deployment, as shown in Figure 3.

However, several critical aspects need to be carefully evaluated when employing AI. In particular, model selection is an essential aspect because depending on the problem (classification, regression, clustering, etc.) or the dataset characteristics (type and number of features, number of instances, codependency among features, etc.), a specific algorithm may perform better than others. The choice of the algorithm should aim to capture the underlying patterns in the data, avoiding the model to perform too well on the training data and to generalize poorly to new unseen ones (overfitting) or, for the sake of generalizability, to fail to effectively learn the patterns in the data, thus behaving poorly in both the training and unseen inputs (underfitting). Achieving a balance between model complexity and performance with unseen data is crucial for the success of AI applications and the specific predictive model.[40] Examples of different algorithms will be deeply discussed in the next paragraphs.

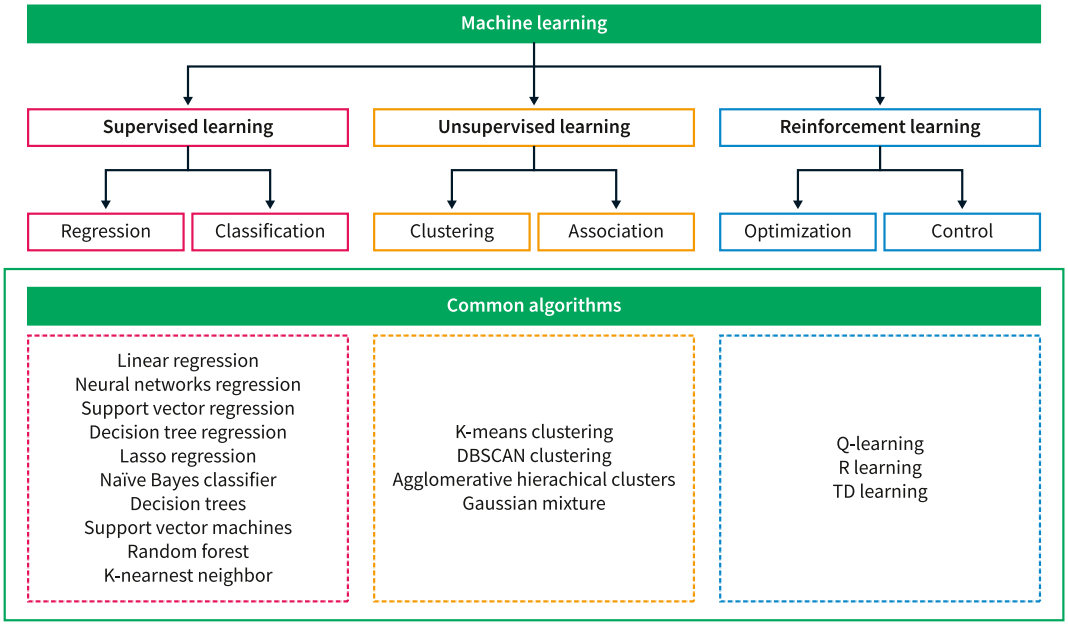

FIGURE 2 Classification of machine learning approaches and their common algorithms.

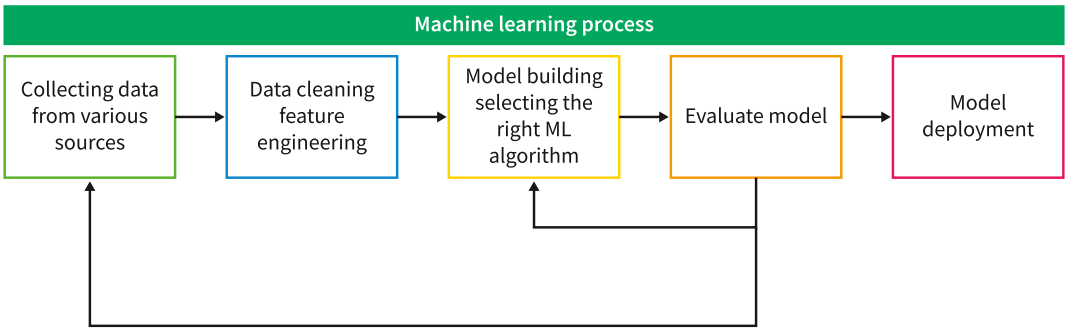

FIGURE 3 The five steps in the ML workflow: data acquisition, data cleaning, model construction and training, model evaluation, and deployment. During the construction of the model, an optimal ML algorithm is selected based on the training dataset and the problem to resolve. This algorithm allows the model to learn and predict behaviors. After the model is trained, its performance is evaluated to test and validate the model itself. If the predictive outcomes are not satisfactory, the model should be further improved by giving feedback in the step "model construction and training", which allows adjustments of some parameters/features. In some cases, it may be required to return to the "data acquisition" step to modify the data entering the ML process. ML, machine learning.

Finally, performance evaluation is key in assessing the effectiveness of AI models and various metrics, such as the area under the receiver operating characteristic curve (AUROC), precision, recall, F1 score, accuracy, sensitivity, specificity, and mean error are commonly used for this purpose.[41] Overall, careful consideration of data type, model selection, and performance evaluation is essential to ensure the effectiveness and reliability of AI-based solutions, being these aspects still subjects of debate in the literature if AI tools are intended to be used in clinical practice.[42]

Deep learning

DL represents both a subset and an evolution of ML, as it emulates neural interactions in the human brain through artificial neural networks (ANNs) and convolutional neural networks (CNNs).[43] These computational structures are composed of multiple perceptrons or "neurons" organized in layers, which autonomously process input data to identify predictive factors for specific outcomes. A key feature of DL is the development of multiple levels of abstraction, which enhances its power. The artificial neurons incorporate nonlinear activation functions to regulate the signals transmitted to subsequent layers, called "hidden layers", and finally give an output. In this case, the entire system is the "model". During the training process, the model parameters are iteratively adjusted to optimize an objective function often linked to prediction confidence.[43] This iterative adjustment involves successive transformations that amplify essential aspects of the input data, enhancing the model discrimination power. As a result, DL can effectively handle more complex and intricate datasets.[31] These advanced computational

models are applied in several fields and tasks from speech recognition and natural language processing to image recognition and object detection, with common use cases including self-driving cars, facial recognition and medical image analysis. Common neural networks that are typically used to perform these specific tasks are ANNs, DNNs (Deep Neural Networks), and CNNs (Convolution Neural Networks). ANN is a generic neural network for modeling nonlinear relationships between inputs and outputs, DNN is an ANN with many hidden layers to capture complex representations. CNN is a DNN that outperforms in image and video recognition using convolutions, while ANN is mostly used for classification and disease prediction.

Common methodologies used for IBD and their applications

In the context of IBD, AI holds significant potential across a spectrum of applications aimed at addressing unmet medical needs. Among these, AI tools can ameliorate the quality of treatment procedures, the classification among IBD, CD, UC, and non-IBD subjects, the evaluation of histological severity and probable remission, disease trajectory, the identification of predictive biomarkers of the pathology and create novel patient-care approaches in IBD.[24,44]

Various ML/DL methods have been used in IBDOmics, in particular: algorithms of classification and/or regression, such as linear algorithms, Support Vector Machine (SVM), k-Nearest Neighbors, decision trees, and ANN,[45–48] or unsupervised learning algorithms such as hierarchical clustering.[49,50]

Linear regression develops a linear model function that best fits the data around a straight line (or hyperplane), while Logistic regression (LR) is used for classification into distinct categories using the sigmoid function that better differentiates among categories.

SVM is a supervised classification method that considers each feature as a dimension of a dataset and finds an optimal hyperplane (boundary) to separate all feature values to best divide a dataset according to a classification of interest.[51]

KNN is another supervised algorithm used for classification or regression that predicts an output considering (k) most similar instances (nearest neighbors).[40]

Classification methods based on decision trees,[52] including techniques like Classification and Regression Trees (CART) and Random Forest (RF), are supervised methods. While CART methods are characterized by a single decision tree, RF methods consist of multiple decision trees to improve accuracy and account for randomness and missing data. Each tree of RF is generated using only a portion of the features providing a vote (or prediction) for the classification. The greatest number of votes determines the final classification of the RF method.

ANNs are particularly used for various IBD-related data types, such as genomics, transcriptomics, proteomics, metagenomics, and clinical data[53] (Figure 4).

Unsupervised approaches include clustering algorithms that group instances based on similarities or distance in feature space in the absence of a predetermined number of clusters. Hierarchical clustering iteratively group instances into larger clusters until they are merged into a single cluster.[40]

The use of AI in IBD will continue to expand. It provides solutions to challenges in other areas such as diagnostic/clinical stratification, endoscopy, histology, intestinal ultrasound, outcome prediction, etc., which are mainly based on AI-based medical imaging analysis, which weren't of primary interest in this review. For imaging analysis in IBD, computer-aided diagnosis (CAD) systems based on CNN and radiomics, a subtype of AI, are most commonly used.[54]

One of the examples using CNN was developed by Iacucci and colleagues[55] who had demonstrated high concordance between CAD

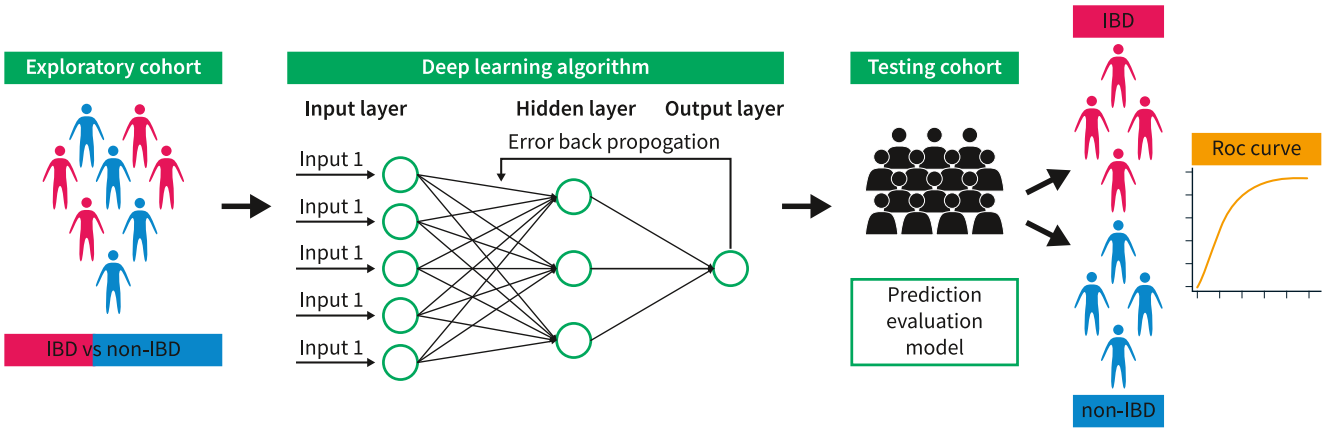

FIGURE 4 A possible structure for an IBD neural network consists of a multi-layer stack: the first layer handles input data from different sources, such as IBD omics elements, endoscopic images, and histological data; one or more hidden layers perform calculations complexes, interactions and combinations of parameters; and finally, an output layer which, after receiving the signals processed by the hidden layers, provides the results. For example, an ANN algorithm could be used to train and instruct an ML model to distinguish IBD patients from non-IBD subjects (exploratory cohort), and subsequently be tested and validated on a separate cohort to evaluate the accuracy of the model in classifying IBD patients from non-IBD subjects. IBD, inflammatory bowel disease.

and human pathologists in stratifying endoscopic assessment and occurrence of risk of flare at 1 year between UC patients with histological activity or histological remission according to the PICaSSO Histological Remission Index.

The same research team used a CNN model to distinguish between endoscopic activity and predicted histological remission and clinical outcome from colonoscopy videos. The AI-PICaSSO model showed a strong association between AI-assessed disease activity and remission and the risk of adverse outcomes, consistent with findings by human endoscopists.[56]

Recently, both radiomics-based analysis and CNN-based models have been applied to IUS for monitoring IBD and distinguishing between normal and endoscopically active inflammation (abnormal) images. The radiomic-based classification model, using the Extreme Gradient Boosting (XGB) classifier, accurately discriminates abnormal from normal images with better performance than CNN.[57]

OMICS IN IBD

AI models in the fields of genomics, transcriptomics, proteomics, and metagenomics have undergone significant evolution, particularly with the advent of DL techniques (Table 1). These models have become more sophisticated, enabling the processing, analysis, and interpretation of complex omics data. The integration of multi-omics and clinical data has been enhanced, leading to breakthroughs in disease diagnosis, drug discovery, and precision medicine.

Genomics

With the advent of genome-wide association studies (GWAS), more than 240 susceptibility loci have been associated with IBD. However, all the associated variants explain only a small portion of the disease risks calculated at about 20%–25%. In these approaches, each variation was analyzed independently from the others, and the signal of single nucleotide polymorphism (SNP) that was most associated was considered for application replications or genotype-phenotype studies.[58]

One of the initial applications that simultaneously considered the effect and possible interactions of 16 SNPs encompassing 11 genes in CD patients was employed by our group using regularized least squares (RLS) classifiers.[59] When all the 16 SNPs were simultaneously employed and with a training set size of $n = 225$ observations, we estimated an accuracy of 62%.

Subsequently, Wei et al.[60] utilized Immunochip data from the International IBD Genetics Consortium's (IIBDGC) Immunochip project employing ML techniques. They developed a predictive model that attained areas under the curve (AUCs) of 0.86 and 0.83 for CD and UC as compared to healthy controls (HC), respectively.

Analyzing more recent data from IIBDGC, Romagnoni et al.[46] employed multiple ML models to predict the probability of developing CD. They observed that the performance of these models could be influenced by factors such as quality control, missing genotype imputation methods, and input data coding strategies.

In 2017 Giollo et al.[61] presented the initial evidence of applying whole exome sequencing (WES) data from Critical Assessment of Genome Interpretation (CAGI) with various computational methods and learning techniques. They utilized three datasets associated with predicting the risk of CD, of which only one demonstrated an AUC of up to 0.7.

Recently, Raimondi et al.[62] developed a novel neural network designed to differentiate CD from HC by utilizing exome data from three datasets sourced from CAGI. The authors utilized feature selection techniques to reduce genetic data complexity and sought to extract a biological interpretation from the model, recognizing that neural networks are among the most challenging DL models to interpret.

Similarly, Wang et al.[63] utilized the CAGI4 dataset as a training set and applied a ML method, namely Analysis of Variation for Association with Disease, to extract signals related to CD and predict CD status. They successfully identified 16% of patients in the validation set of over 3000 individuals with 99% precision with a stringent cut-off, while 58% of the patients (at default cut-off) with 82% precision.

Very recently, Stafford et al.[64] employed a supervised ML classifier with a larger WES dataset to distinguish between CD and UC. They examined three gene panels, including all available genes, an autoimmune gene panel, and an IBD gene panel. Their analysis underscored that the autoimmune gene panel emerged as the top performer for CD and UC classification, achieving an AUROC of approximately 0.7.

Transcriptomics

Investigating the downstream effects of genomic aberrations on the transcriptome and proteome offers additional molecular insights into unraveling IBD pathogenesis. Differential gene expression analysis has proven instrumental in identifying key genes and pathways involved in IBD pathogenesis.[65,66] Several studies have utilized computational approaches to assess whether these additional branches of molecular biology can predict IBD pathogenesis.

Isakov et al.[45] devised an ML-based gene prioritization technique aimed at distinguishing IBD risk genes from non-IBD genes. They trained four distinct ML models to generate gene risk scores. An ROC curve was generated for each of the four models, whose calculated AUC ranged from 0.76 to 0.87 during training, while greater differences between the models emerged in the validation phase to determine the scores between IBD and non-IBD genes, obtaining higher AUC values (0.78–0.83). The model identified 347 genes with high prediction scores for IBD risk, encompassing genes previously recognized to be associated with IBD via GWAS and literature, as well as 67 novel IBD-associated genes.

In a recent study by Khorasani et al.,[50] a predictive model was developed to differentiate UC from HC utilizing multiple colonic

TABLE 1 Summary of key studies exploring AI models applied to genomics, transcriptomic, proteomics, metagenomics and multiomics in IBD.

Studies on AI applied to OMICs	AI classifier	Populations	Primary outcomes/ Clinical results	Performance				
				Precision	Sensitivity	Specificity	Accuracy	AUC
Genomics								
D'Addabbo et al. (2007)	RLS	178 patients with CD and 127 HC	Prediction of CD susceptibility		61%	63%	62% (with a training set size of $n = 225$ examples)	
Wei et al. (2013)	L1-penalized LR	Immunochip data from the International IBD Genetics Consortium's (IIBDGC) immunochip project	Prediction of CD and UC susceptibility					0.86 for CD and 0.83 for UC
Romagnoni et al. (2019)	Comparison of penalized LR, GBT and ANN	Immunochip data from the International IBD Genetics Consortium's (IIBDGC) immunochip project	Prediction of CD susceptibility					0.80 achieved by logistic regression methods
Giollo et al. (2017)	Regularized logistic model, SVM, GBT	Whole exome sequencing (WES) data from critical assessment of genome interpretation (CAGI)	Prediction of CD susceptibility					0.86 on a very large test set using a regularized logistic model
Raimondi et al. (2020)	NN model called Cdkoma	Exome data from three datasets sourced from critical assessment of genome interpretation (CAGI)	Prediction of CD susceptibility	90% for CAGI3 and CAGI2 datasets, 77% for CAGI4	62% for CAGI4, 95% for CAGI3, 98% for CAGI2	72% for CAGI4, 60% for CAGI3, 61% for CAGI2		0.70 for CAGI4, 0.80 for CAGI3, 0.73 for CAGI2
Wang et al. (2019)	Analysis of variation for association with disease (AVA,Dx)	CAGI4 dataset as a training set (train panel included 64 unrelated CD and 47 HC) and testing set (test panel included 51 CD and 15 HC)	Prediction of CD susceptibility	AVA, Dx identified 16% (at strict cutoff) of patients with CD at 99% precision and 58% of the patients (at default cutoff) with 82% precision in over 3000 individuals from separately sequenced panels				

(Continues)

TABLE 1 (Continued)

Studies on AI applied to OMICs	AI classifier	Populations	Primary outcomes/ Clinical results	Performance				
				Precision	Sensitivity	Specificity	Accuracy	AUC
Stafford et al. (2023)	RF	Whole exome sequencing (WES) from pediatric/adult IBD patients (600 with CD and 306 with UC)	Prediction of CD and UC susceptibility	77% for all available genes panel, 82% for the autoimmune gene panel, 79% for the IBD gene panel		51% for all available genes panel, 67% for the autoimmune gene panel, 65% for the IBD gene panel		0.57 for all available genes panel, 0.68 for the autoimmune gene panel, 0.61 for the IBD gene panel
Transcriptomics								
Isakov et al. (2017)	RF, SVM with polynomial kernel (svmPoly), extreme gradient boosting (xgbTree), and GLMNET	Expression data were downloaded from the gene expression omnibus (GEO) database. The samples originated from various intestinal biopsies corresponding to 349 IBD (180 CD and 149 UC), 94 colorectal neoplasms, and 90 normal tissue	Identification IBD-risk genes		For the combined model 63% during the training and 58% during the test	For the combined model 91% during the training and 88%during the test	For the combined model 85% during the training and 81% during the test	For the combined model 0.634 during the training and 0.577 during the test
Khorasani et al. (2020)	SVM	Expression data were downloaded from the gene expression omnibus (GEO) database. The samples originated from subjects with UC (in active and inactive state) and HC	Classification between UC and HC	62% for the classification between inactive UC and HC and 100% for the classification between active UC and HC				
Li et al. (2020)	RF, ANN	Two sets of cases and an HC containing the UC gene expression profile (training set GSE109142 and validation set GSE92415) were downloaded	Prediction model of UC diagnosis					0.95
Yu et al. (2023)	Combination of unsupervised clustering analysis and	Four IBD-associated datasets, including GSE112366,	Prediction of IBD susceptibility	87%			86%	0.96

TABLE 1 (Continued)

Studies on AI applied to OMICs	AI classifier	Populations	Primary outcomes/ Clinical results	Performance				
				Precision	Sensitivity	Specificity	Accuracy	AUC
	the XGBoost feature selection method	GSE3365, GSE75214, and the data from the integrative human microbiome project (iHMP) were integrated, and 41,307 features across 705 samples were chosen						
Zhang et al. (2022)	SVM, least absolute shrinkage and selection operator, RF, gradient boosting machine, principal component analysis, and NN	Ten eligible microarrays including 387 patients with UC and 139 HC	Genes discovery for UC susceptibility					Higher than 0.8
Smith et al. (2020)	kNN classifier, RF, and l2-regularized multinomial LR	Expression data from the genotype-tissue expression (GTEx) project, the cancer genome atlas (TCGA) pan-cancer clinical data resource, and the sequence read archive (SRA); 45,000 samples including ~37k samples for training, ~4k samples for validation, and ~4k samples test sets	Prediction of UC susceptibility					0.85
Wacker et al. (2023)	RF	Whole blood samples of 495 patients with UC and 243 HC from Germany, 220 patients with PSC (177 with a concurrent UC diagnosis) and 77 HC from Norway	Classification among UC, PSC and HC					UC versus controls 0.95; UC versus PSC 0.986

(Continues)

TABLE 1 (Continued)

Studies on AI applied to OMICs	AI classifier	Populations	Primary outcomes/ Clinical results	Performance				
				Precision	Sensitivity	Specificity	Accuracy	AUC
Martin et al. (2019)	Hierarchical clustering	Single-cell analysis from 11 patients with ileal CD (iCD)	Prediction of response to anti-TNF therapy in iCD					Patients achieving durable corticosteroid-free remission versus or not achieving 0.69
Abreu et al. (2022)	GLMNET, rpart, randomGLM, RF	37 IBD patients that received vedolizumab therapy	Prediction of response to vedolizumab therapy in IBD				82%	
Proteomics								
Ungaro et al. (2021)	RSF	265 pediatric CD patients	Identification of biomarkers predictors of CD complications					Penetrating: 0.79 (with 5 protein); stricturing: 0.68 (with 4 protein)
Jongsma et al. (2023)	LR	91 pediatric CD patients	Prediction of response to anti-TNF-alpha or conventional therapies in CD pediatric	NA	NA	NA	NA	NA
Microbiomics								
Manandhar et al. (2021)	For the classification of IBD versus non-IBD: RF, DT, EN, SVM with radial kernel, and NN; for the classification of CD versus UC: RF	The metagenomics data were collected from the american gut project: 19,978 stool samples, 934 samples from IBD patients and 19,044 samples from non-IBD; 406 and 179 samples from the participants diagnosed with CD and UC	Classification between IBD patients and non-IBD subjects and between CD and UC	IBD versus non-IBD: 69%; CD versus UC: 90%	IBD versus non-IBD: 80%; CD versus UC: 85%	IBD versus non-IBD: 64%; CD versus UC: 80%	IBD versus non-IBD: 72%; CD versus UC:83%	IBD versus non-IBD: 0.80; CD versus UC: 0.92
Wang et al. (2021)	RF	66 treatment naive IBD pediatric patients and 27 HC as the exploration cohort, 14 early onset IBD patients and 48 IBS patients as the validation cohort	Classification among pediatric IBD patients, HC, and IBS					IBD versus HC: 0.88; IBD versus IBS: 0.84

TABLE 1 (Continued)

Studies on AI applied to OMICs	AI classifier	Populations	Primary outcomes/ Clinical results	Performance				
				Precision	Sensitivity	Specificity	Accuracy	AUC
Vich Vila et al. (2018)	Generalized linear models	Stool samples and fecal calprotectin from 1792 individuals with IBD and IBS compared with HC	Classification between IBD and IBS					IBD versus IBS: 0.93
Zuo et al. (2022)	RF	Feces samples from 19 UC pediatric patients and 23 HC	Prediction of pediatric UC susceptibility					0.90
Ananthakrishnan et al. (2017)	ANN model called vedoNet	85 IBD patients (43 UC and 42 CD)	Prediction of response to vedolizumab therapy in IBD					Responder versus non responder: 0.87
Raygoza Garay et al. (2023)	RSF	3483 healthy first-degree relatives (FDRs) of CD patients	Development of a microbiome risk score (MRS) and prediction of CD					Developed CD within 1.5 years versus healthy relatives: 0.7
Caenepeel et al. (2023)	LR, NN	296 patients with active IBD (203 CD, 93 UC) initiating biological therapy	Prediction of response to biological therapy in IBD		67.5%	67.6%		0.74
Multiomics								
Lloyd-Price et al. (2019)	Linear mixed effect model	2965 stool, biopsy, and blood specimens from 132 IBD patients	Prediction of the disease activity status	NA	NA	NA	NA	NA
Gardiner et al. (2022)	Comparison of RF, XGBoost, SVM, k-NN and adaboost	The data set consisted of 25 patient organoculture (ex vivo) assay data sets, the associated genomics/ transcriptomics and the patient demographic/ clinical information	Prediction of response to therapies	NA	NA	NA	NA	NA
Lee et al. (2021)	ANN	21 participants with profiles of proteomic, metabolomic, and metagenomic data at baseline	Prediction of response to biological therapies					0.96 (multi-omics data)

(Continues)

TABLE 1 (Continued)

Studies on AI applied to OMICs	AI classifier	Populations	Primary outcomes/ Clinical results	Performance				
				Precision	Sensitivity	Specificity	Accuracy	AUC
Arehart et al. (2023)	LR	1785 samples from 130 individuals (103 IBD and 27 HC)	Prediction of IBD diagnosis					0.80 (multi-omics data)
Ning et al. (2023)	RF	Analysis of nine metagenomic cohorts (N = 1363 cases) and four metabolomics cohorts (N = 398 cases) of IBD patients from different countries or regions	Identification of IBD diagnostic marker					0.98 (multi-omics data)
Mishra et al. (2022)	RF	Discovery cohort: 14 IBD patients (10 UC and 4 CD); 17 IBD patients (10 UC and 7 CD), who received first-time therapy with vedolizumab were used as treatment controls	Prediction of response to anti-TNF therapy in IBD		100%	50%	85%	0.88 (multi-omics data)
Gao et al. (2023)	RF	540 tissue samples from 30 CD active patients and HC	Identification of biomarkers predictors of CD					0.96

Note: In Gardiner et al. was evaluated the predictive error using the mean absolute error (MAE) = 4.98%.

Abbreviations: AI, artificial intelligence; ANN, artificial neural network; AUC, Area under the curve; CD, Crohn's disease; DT, decision tree; EN, elastic net; GBT, gradient boosted trees; GLMNET, regularized generalized linear model; HC, healthy controls; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; kNN, k-Nearest Neighbor; LR, logistic regression; NA, not available; NN, neural network; PSC, primary sclerosing cholangitis; RF, random forest; RLS, regularized least squares; RSF, random survival forest; SVM, support vector machines; TNF, tumor necrosis factor; UC, ulcerative colitis.

transcriptome datasets from gene array expression data. They employed an SVM classifier model using a signature of 32 genes, many of which did not have a direct association with the IBD phenotype. The model effectively classified active and inactive UC from HC with an average precision of 1 and 0.62, respectively.

In a subsequent study by Li et al.,[47] utilizing public expression profile datasets, a universal molecular prognostic score (mPS) was established based on the expression data of 30 genes. Through the utilization of an ANN that calculated the weights of these genes with UC and a mPS system-based classification model, they attained an AUC exceeding 0.95 for identifying UC from HC.

Utilizing public datasets of gene expression, Yu et al.[67] employed a combined unsupervised clustering analysis with the XGBoost algorithm for feature selection to differentiate between IBD and HC. The authors identified a 32-gene IBD signature capable of predicting IBD occurrence in new cohorts of IBD subjects, achieving an accuracy of 0.86.

In a subsequent study,[68] 10 eligible UC microarray datasets were analyzed alongside 6 ML methods to identify associated genes potentially useful for diagnosis. Initially, 87 deregulated genes were identified, and through the implementation of ML approaches, two genes, namely OLFM4 and C4BP, achieved an average ROC performance higher than 0.8.

Further studies have applied AI to RNAseq obtained from public databases or conducted experiments based on the whole blood RNA-Seq approach. In their recent study, Smith et al.[69] utilized RNAseq data from various databases and several diseases including UC, autoimmune diseases, and cancers, and employed several ML

techniques. They observed that multivariate predictors surpassed predictors based on single genes, indicating the importance of considering multiple genes simultaneously. Furthermore, they found that larger gene sets and the L2-regularized regression method yielded the most effective choice for predictive analyses using transcriptomic data.

In a more recent study by Wacker et al.[70] transcriptional variances between UC and primary sclerosing cholangitis (PSC), a chronic liver disease regarded as a rare extra-intestinal manifestation of IBD, were dissected. Through RF analysis, the study achieved high accuracy in discriminating PSC, UC, and HC (ROC UC vs. HC: 0.95, PSC vs. HC: 0.88, UC vs. PSC: 0.986), while showing poorer discrimination solely for PSC and PSC/UC (ROC: 0.56).

Single-cell RNA sequencing technology, increasingly utilized in IBD research, enables detailed analysis of various cell type phenotypes. This is particularly crucial in IBD research because inflammatory phenotypes of immune cells are often enriched in inflamed tissues. The detection of such cell phenotypes is associated with disease progression and therapy failure as demonstrated in a study by Martin et al.,[71] which utilized an unsupervised ML approach to differentiate between major cell types. This approach has the potential to distinguish the cellular composition of inflamed and uninflamed tissue, thereby facilitating precise diagnosis, disease localization, and therapeutic decisions.

Very recently, Abreu and colleagues[72] analyzed the transcriptional behavior of T Cells to predict responses to vedolizumab therapy in IBD patients. They sorted T cells from peripheral blood and lamina propria mononuclear cells and subsequently analyzed them using RNA-seq. The analyses revealed that the best computational and ML models for predicting their response to vedolizumab were RF, and random GLM, both achieving an accuracy of 82%.

Proteomics

From a proteomics perspective, numerous efforts[73] have been made although many of them have utilized traditional methods such as regression models. In a case-cohort study utilizing a pediatric CD inception cohort, Ungaro et al.[74] analyzed 92 inflammation-related proteins assayed in baseline plasma of individuals with inflammatory disease (B1) behavior to predict the development of stricturing (B2) or penetrating (B3) complications. The authors employed ML by applying a random survival forest (RSF) to select variables predicting B2 and B3 complications. In particular, the first analysis with 5 protein markers predicted B3 complications with an AUC of 0.79, while using 4 proteins it predicted B2 complications with an AUC of 0.68. These results demonstrate greater reliability compared to serologic markers.

Jongsma et al.[75] evaluated the serum immune profile of 92 proteins to predict the response to anti-TNF-alpha or conventional therapies in pediatric CD patients using ML and statistical approaches. They identified 30 immunoproteins deregulated between patients using anti-TNF-alpha therapy at pre- and after stages. Utilizing ML approaches, the authors categorized patients into two groups: CD-hi (higher clinical disease activity, C-reactive protein, blood neutrophils, and higher immunoproteins at baseline) and CD-lo. The model based on multivariate binary LR analyses highlighted that patients belonging to the cluster CD-lo achieved the best outcomes. Specifically, these patients maintained remission without escalation at week 52 in 58% of IFX-treated patients compared to only 24% of CD-hi patients.

Microbiomics

Several series of predominantly case-control studies involving patients with established IBD documented alterations in gut microbiome composition.[76] However, these studies often fail to distinguish whether the altered gut microbiome composition is associated with the initiation of IBD or is a consequence of inflammation or drug treatment. Recent advancements in sequencing technologies and data analysis techniques have facilitated a more comprehensive characterization of microbiota communities for IBD patients. Moreover, researchers have begun leveraging ML techniques to better analyze metatranscriptomic data alongside clinical characteristics, disease behavior, and responses to therapies.

Utilizing data from the American Gut Project, Manandhar et al.[77] employed five different ML algorithms to discriminate between IBD patients and non-IBD subjects as well as between CD and UC. They utilized 50 differential bacterial taxa between the IBD and non-IBD groups achieving an AUC of approximately 0.80. Similar results were obtained when 500 operational taxonomic units (OTUs), representative of groups of closely related individuals, were used instead of bacterial taxa. Additionally, they tested whether supervised ML could differentiate between CD and UC. One hundred and seven differential bacterial taxa or 500 OTUs were identified between the diseases. By employing either taxa or OTUs and applying ML approaches, the research group was able to discern between CD and UC with an AUC exceeding 0.9.

Subsequently, an ML strategy was employed in pediatric IBD to differentiate between IBD patients, controls, and irritable bowel syndrome (IBS) subjects.[78] The researchers utilized the top 11 OTUs obtained from the RF model comparing IBD and HC, resulting in a model performance with an AUC of 0.88. Furthermore, using a validation cohort consisting of 14 early onset IBD patients and 48 IBS patients, the model achieved a performance of 0.84.

In a further work, Vich Vila and colleagues[79] characterized the gut microbiota composition of both disorders using shotgun metagenomic sequencing of stool samples from 1792 individuals with IBD and IBS compared with controls achieving similar results. When fecal calprotectin and biometrical data were incorporated into the model, it achieved the highest accuracy with AUC of 0.93.

Comparable results were also observed in pediatric UC patients[80] using both 16S and shotgun data, employing ML based on RF with a large number of trees. When 500 trees were utilized, the study was able to distinguish UC from HC with an ROC of 0.90.

Furthermore, the inclusion of age and gender in the model did not significantly affect this value.

To assess whether the gut microbiome could predict responses to IBD therapy, Ananthakrishnan et al.[81] conducted a prospective study involving IBD patients initiating anti-integrin therapy. They utilized an ANN model incorporating microbiome-related data, including abundance and functional profiles, along with clinical data. Their best-performing model for predicting clinical remission achieved an AUC of 0.87 when incorporating clinical data, pathway relative abundance, and a manually curated list of 40 taxa.

Some of the most intriguing papers that focused on the composition of the intestinal microbiota and employed ML techniques demonstrated the ability to anticipate changes in the gut microbiota composition up to 5 years before the onset of CD,[82] and the capability to predict the response to therapy with biological drugs in patients with IBD.[83]

Raygoza Garay et al.[82] developed a microbiome risk score (MRS) utilizing an RSF model that incorporated microbial composition and clinical variables from healthy first-degree relatives of patients with CD. Their study demonstrated that the gut microbiome was associated with the future onset of CD, suggesting its critical role in CD pathogenesis. The subjects were monitored for several years from baseline and 73/3483 developed CD with a median time from enrollment to CD diagnosis of about 3 years. The cohort was divided into training and validation groups. In the discovery group, the MRS yielded a hazard ratio (HR) of 1.58, while in the validation cohort, patients with the highest MRS had more than twice the risk of developing CD. Subsequently, the method was applied to predict the risk of developing CD within 1.5, 3, and 5 years after the baseline stool sample was collected. Similar AUCs of 0.7 were obtained at all three time points analyzed.

Caenepeel et al.[83] utilized quantitative microbiome profiles of pre-and post-treatment fecal samples to predict treatment response to biological therapies. Using baseline data, they identified two enterotypes of which BACT2 was found in 65.9% of dysbiotic individuals, with a significantly higher prevalence among patients with ileal involvement (76.8%). To predict treatment outcomes, logistic regressions with various ML approaches were employed and after training the model, they achieved an AUC of 0.7. The best performance was attained when the quantitative microbiome profiling abundance was excluded from the variables, resulting in treatment outcomes of different biologics with 73.9% accuracy. Subsequently, the model was applied to predict the efficacy of alternative therapies in case of nonresponse to the initially prescribed biologic. Among the non-responders to biologics, 27.63% were predicted to respond to second-line therapy with anti-TNF, 10.53% with vedulizumab, and 8.55% with any alternative therapy.

AI-enabled MultiOMIC

While numerous studies on the utilization of multi-omics data analyzed into computational algorithms are documented in the literature, many of them are characterized by small sample sizes. Currently, tangible and satisfactory results are primarily focused on specific areas, including predicting therapy response and anticipating remission and relapse. Some intriguing initial applications of multi-omics data integration, although not employing stringent ML approaches, are in characterizing dysregulated interactions between host and environmental factors in IBD[5,84] as well as in understanding the response to medical therapy.[85]

For instance, Lloyd-Price et al.[5] conducted extensive multi-omics molecular profiling on 132 IBD patients. By applying a model based on a linear mixed effect, they observed significant alterations in microbiota composition and function based on disease activity states.

A further attempt was made by Gardiner L. J. et al.,[85] who employed a ML workflow integrating multi-omics data (genomics and transcriptomics), demographic information, medicinal records, and pharmacology data derived from a preclinical human fresh tissue assay to predict patient drug responses. They achieved the best results with a mean absolute error of 4.98, when the model was used to predict inflammatory drug response from a combination of integrated demographic, medicinal and SNPs on unseen patients.

Ananthakrishnan's group conducted a further study[86] involving IBD patients initiating anti-integrin therapy. They utilized an ANN model incorporating microbiome, metabolome, and proteome data along with clinical information to predict responses to biological therapies. Building separate models with metagenomic, metabolomic, or proteomic features individually, they achieved predictive values with AUCs of 0.85, 0.77, and 0.81, respectively. Despite having only 21 participants with profiles of proteomic, metabolomic, and metagenomic data at baseline, they challenged an ML model with a combination of clinical, metagenomic, metabolomic, and proteomic features. The AUC of the model yielded the best result of 0.96.

Recently, Arehart et al.[87] utilized data from the Human Microbiome Project 2 IBD multi-omic database and employed ML approaches to develop a polygenic risk score framework across multiple omic data types to predict IBD diagnosis. They identified several species, pathways, and metabolites known to be associated with IBD risk and built a model incorporating demographic covariates. During the training phase, the model achieved an AUC of 0.80 using multi-omics data. However, individually, metabolomics (AUC = 0.82) and viromic (AUC = 0.83) scores were more predictive than metagenomics (AUC = 0.73) or metatranscriptomics (AUC = 0.73) ones.

Recently Ning and colleagues[88] identified multi-omics biomarkers for IBD diagnosis by analyzing 9 metagenomic and 4 metabolomics cohorts of IBD from different populations. They first selected 31 bacterial species, 25 Kegg Ontology genes and 13 metabolites, and then combined these features using ML approaches to identify IBD patients from non-IBD subjects. They further discovered heterogeneous results with a reduced AUC score in both intra- and inter - cohorts, attributing such variations to geographical diversity. However, applying the multi-omics model approach, they achieved the best results with an AUC >0.90 in all the datasets having both metagenomics and metabolites. In a further study,[89] both RNA-seq

and DNA methylation profiling were analyzed in IBD patients receiving infliximab therapy to identify dynamic molecular signatures associated with clinical remission or non-remission at weeks 14. An integrative ML approach was used to identify and build several models composed of distinct features. The best score was achieved in predicting the response to CD patients when the model comprised 31 features from transcriptomics and epigenomics data, yielding an AUC of 1. When using 87 or 14 features, AUCs of 0.93 and 0.9 were obtained in IBD and UC, respectively. However, in the CD validation cohort, the ROC curve of the prediction model with 31 selected features achieved a lower AUC of 0.88. Recently, Gao and colleagues[90] revealed possible host-microbe interactions in CD patients by analyzing spatial omics (RNA-seq at both microbiota and transcriptomics) levels, sera proteomics and fecal microbiota. They identified several candidate interactions between host proteins and microbes associated with CD gut inflammation. They constructed a classification model, based on an RF classifier, for diagnosing patients with active CD using the 5 overlapped differential proteins and 19 overlapped differential genera. The model built using the combination of protein and bacterial biomarkers achieved an accuracy (AUC = 0.96), displaying high diagnostic capability in distinguishing active CD patients from HC; this value was higher than that obtained by the signature of only proteins (AUC = 0.91) and the key genera of microbiota (AUC = 0.91).

AI LIMITATIONS AND ROADMAPS FOR CLINICAL IMPLEMENTATION

Unfortunately, the works outlined above have certain limitations. For instance, single-center studies, while advantageous in terms of data uniformity regarding recruitment, may have limitations related to geographical and racial aspects, and the sample size analyzed could be a major issue. Conversely, cross-cohort studies, while analyzing several subjects, may encounter difficulties in eliminating biases related to cohort selection, sample collection, and methodological differences.

Recent studies have contributed to bioinformatics by developing AI-based tools that improve the reproducibility and interpretability of models applied to omics data. For instance, new algorithms for the integrative analysis of omics and health data, including biomedical images, have been introduced. Moreover, AI has been instrumental in advancing protein structure prediction and gene function prediction, which are crucial for understanding biological processes and disease mechanisms.

The improvements in AI models are evident in their increased accuracy, reliability, and efficiency. They are now better equipped to handle the vast amounts of data generated by omics technologies, making them invaluable for bioinformatics research. The recent contributions of AI in bioinformatics include novel machine-learning methods for examining molecular structures and classifying biological data, thereby pushing the boundaries of what is possible in the field.

In addition, critical aspects may include the lack of generalization capabilities of the AI models when applied to independent validation cohorts, which currently represent a major limitation of ML approaches applied to multifaceted or multifactorial diseases such as IBD. The primary limitation of the clinical applicability of AI lies precisely in the high heterogeneity of the disease and its variations over time, leading to inadequate reproducibility and generalizability of predictive results and a possible overestimation of prediction accuracy.

However, some challenges remain that require further improvements and a multi-faceted approach. The roadmap for implementation in clinical practice necessitates standardization, clear study design, simplicity, performance auditing, cost-effectiveness, the creation of international AI-working groups, and acceptance by patients and healthcare providers. In the next few years, we hope that international data-sharing consortia initiatives, including not only genomics, transcriptomics, proteomics, and metagenomics but most importantly endo-histo-OMIC[44,91] and radio-OMIC[54,57] approaches in IBD, will help train AI tools on large, unbiased data sets that reflect the heterogeneity and complexity of IBD patient characteristics. At the same time, our hope is for the full acceptance of AI by both physicians and patients, so that the patients can take advantage of the benefits of AI for their well-being and better control these complex and chronic conditions characterized by fluctuating periods of active inflammation and clinical remission.

CURRENT AND FUTURE PERSPECTIVES

Inflammatory bowel disease is an unpredictable, relapsing-remitting, and fluctuating course. Despite the significant advances in molecular biology, metagenomics, proteomics, and metabolomics, they still represent critical challenges for researchers involved in IBD studies, primarily due to the biological complexity and heterogeneity of the data, and the lack of standard pipelines dedicated to individual omics or their integrations (Figure 5).

A wide range of artificial intelligence applications offer promising solutions and could change our perspective on some of the unresolved issues, in particular, to improve personalized medicine or the discovery of non-invasive biomarkers applied to these complex multifactorial diseases as well as to predictors of cancer-associated colitis.[92] Currently, integrating data from individual omics relevant to IBD is seen as an approach that could significantly enhance precision medicine (Figure 6).

This review aims to focus on recent publications in IBD from 2012 to 2023 conducted both individual omic and multi-omics studies, where artificial intelligence has been applied.

Our work highlights critical aspects of the applicability of AI in clinical practice. In the first reports, AI was applied to genomic databases including immunochip and WES datasets. Data from genomic studies and AI models underscore how genetic variants can distinguish between CD and UC. Models using GWAS data[46,59,60] have shown better performance compared to WES data,[61–64] likely due to

the larger sample size in GWAS. However, many AI algorithms have been used to develop the best-performing model by incorporating statistically associated features such as SNPs and clinical data. Nevertheless, in most studies, when the trained models were subsequently applied to the replication cohort, their performance was reduced.

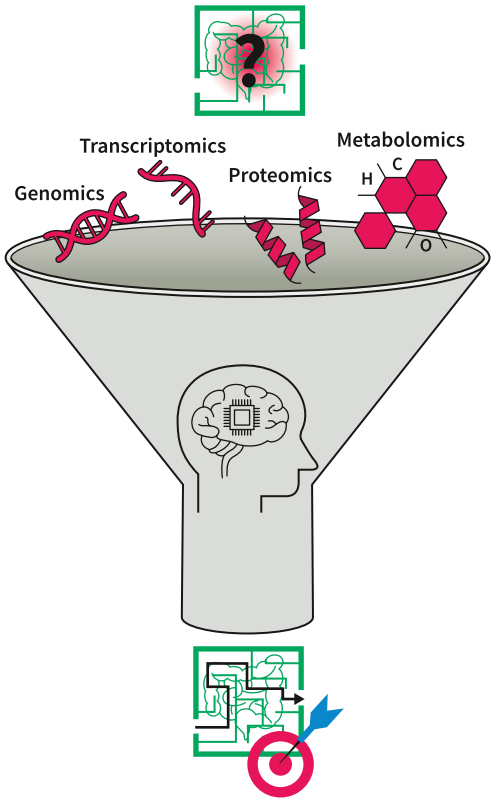

FIGURE 5 Data integration at individual OMICs (genomics, transcriptomics, proteomics, and metabolomics) with artificial intelligence tools could provide multi-omics analysis to decipher the complex labyrinth of inflammatory bowel disease.

The transcriptomics studies focused on the implementation of gene expression datasets resulting from bulk and single-cell RNA-seq technology from mucosal gut tissues or blood. In most of these studies, various models were built, trained, and optimized using deregulated genes (i.e. features) that were differently associated using traditional statistical strategies or tools based on ML-driven regression algorithms. It was observed that the best performances were achieved using gene expression data from microarrays, with a relatively small number of genes (<100), to discern between UC and HC,[50,67,68] and between IBD and non-IBD[47] also in this case likely due to the use of bigger sample sizes.

However, when RNA-seq data were analyzed, the most effective model was characterized by using a larger number of genes to discern among UC and other diseases or malignancies,[69] such as UC and PBC.[70]

Single-cell RNA-seq data were also applied with AI to identify inflammatory phenotypes. In the first study, the AI model was able to differentiate between inflamed and uninflamed tissue,[71] while subsequently, the transcriptional behavior of T cells was used to predict responses to vedolizumab therapy in IBD patients.[72] Despite the numerous studies and various gene expression approaches used, none of them succeeded in identifying unique molecular signatures associated with IBD or response to clinical therapy.

The state of the art in proteomics is unfortunately less advanced, with most studies relying on more traditional statistical-based approaches rather than AI. One study attempted to use ML strategies to predict future disease complications; however, it did not achieve satisfactory accuracy.[74]

Studies interested in evaluating the influences of the composition of the intestinal microbiota in IBD using AI techniques mainly focus on the integration of metatranscriptomic data with clinical and demographic parameters as well as on predicting disease progression or response to therapies. Several studies have attempted to characterize the viral communities of the mucosa using statistical and ML techniques, resulting in different clusters. However, the results obtained from each study, while characterized by good accuracy, are not directly comparable due to differences in cohorts (e.g., UC vs. HC

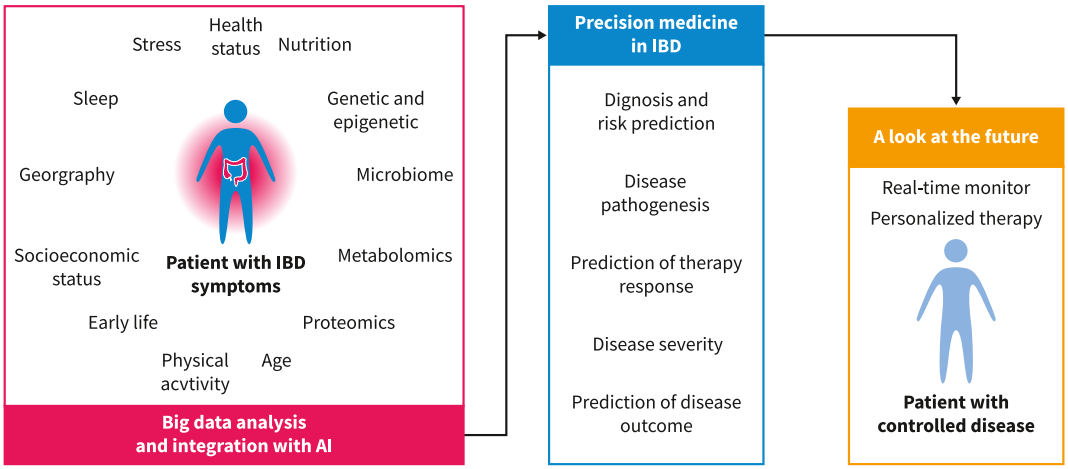

FIGURE 6 Data-integration and analysis with artificial intelligence in inflammatory bowel disease: a combination of heterogeneous information, including omics data, clinical variables, exposomes, and biometric data.

subjects,[80] pediatric IBD vs. HC vs. IBS subjects[78]) or variations in feature combinations[77,80] (e.g., OTUs, bacterial taxa, etc.) and therefore were conducted using non-standardized methods.

In this field, two papers emerged as intriguing: the first was able to anticipate changes in the gut microbiota composition up to 5 years before the onset of CD in a cohort of healthy family members of CD,[82] and the other one focused on predicting the response to therapy with biological drugs in patients with IBD.[83]

Although several studies on the utilization of multi-omics data incorporated into computational algorithms are documented in the literature, tangible and satisfactory results mainly focus on specific areas, including the prediction of therapeutic response and the anticipation of remission and relapse.

From the literature analysis emerges with clarity that, to date, the application of AI for a more efficient multi-omics investigation of IBD is still in the emerging and trial phase. While some initial studies achieved promising predictive outcomes, others relied on more traditional computational techniques.[5] The utilization of AI has facilitated achieving good accuracy in predicting therapy response in Gradinier's study,[85] which involved efficient multi-omics integration. Additionally, further studies have emerged aiming to predict treatment response,[86] disease risk[87] or the identification of biomarkers to distinguish disease activity.[90]

These studies demonstrated good performance through both individual models based on single omics and integrative models.

Despite its significant potential, there remains a notable lack of real clinical use of AI in IBD. After a thorough review of the current literature, it seems overly optimistic to expect AI-based multi-omics studies to produce revolutionary advances in IBD, at least for now. Satisfactory results in the application of AI in the field of IBD have not currently been achieved, even though sophisticated analysis tools have been adopted in many studies. However, these tools are typically based on biological samples obtained from single subjects at a single point in time, reflecting only the biology at that specific moment of the patient's life rather than capturing the complete biology of ever-changing chronic conditions such as CD or UC. In the near future, the AI approaches are expected to be especially valuable in classifying already diagnosed patients into disease sub-phenotypes, predicting disease progression, and evaluating response to treatment.

AUTHOR CONTRIBUTIONS

Anna Lucia Cannarozzi, Orazio Palmieri, Francesco Perri, Silvio Danese, Federica Ungaro Study design and data analysis and final approval of article; Anna Lucia Cannarozzi, Orazio Palmieri, Fabrizio Bossa, Giuseppe Biscaglia, Luca Massimino, Anna Latiano, Sonia Carparelli, Gionata Fiorino, Anna Laura Di Brina, Francesca Tavano, Maria Guerra, Francesco Giuliani and Matteo Riva, data collection and writing up of the first draft of the paper, and final approval of the article.

ACKNOWLEDGMENTS

The work was supported by the "Italian Next Generation Eu Program (PNRR)" PNRR-MAD-2022-12375729.

Open access funding provided by BIBLIOSAN.

CONFLICT OF INTEREST STATEMENT

S Danese has served as a speaker, consultant, and advisory board member for Schering-Plough, AbbVie, Actelion, Alphawasserman, AstraZeneca, Cellerix, Cosmo Pharmaceuticals, Ferring, Genentech, Grunenthal, Johnson and Johnson, Millenium Takeda, MSD, Nikkiso Europe GmbH, Novo Nordisk, Nycomed, Pfizer, Pharmacosmos, UCB Pharma and Vifor.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Anna Latiano https://orcid.org/0000-0003-3719-2061
Matteo Riva https://orcid.org/0009-0002-7054-1913
Federica Ungaro https://orcid.org/0000-0001-5395-7795
Sonia Carparelli https://orcid.org/0000-0003-2054-8754
Gionata Fiorino https://orcid.org/0000-0001-5623-2968
Orazio Palmieri https://orcid.org/0000-0002-0019-7929

REFERENCES

1. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. 2017;389(10080):1756–70. https://doi.org/10.1016/s0140-6736(16)32126-2
2. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. Lancet. 2017;389(10080):1741–55. https://doi.org/10.1016/s0140-6736(16)31711-1
3. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. 2017;390(10114):2769–78. https://doi.org/10.1016/s0140-6736(17)32448-0
4. Di Narzo AF, Peters LA, Argmann C, Stojmirovic A, Perrigoue J, Li K, et al. Blood and intestine eQTLs from an anti-TNF-resistant Crohn's disease cohort inform IBD genetic association loci. Clin Transl Gastroenterol. 2016;7(6):e177. https://doi.org/10.1038/ctg.2016.34
5. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655–62. https://doi.org/10.1038/s41586-019-1237-9
6. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol. 2017;2(5):17004. https://doi.org/10.1038/nmicrobiol.2017.4
7. Vieira-Silva S, Sabino J, Valles-Colomer M, Falony G, Kathagen G, Caenepeel C, et al. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nat Microbiol. 2019;4(11):1826–31. https://doi.org/10.1038/s41564-019-0483-9
8. Pascal V, Pozuelo M, Borruel N, Casellas F, Campos D, Santiago A, et al. A microbial signature for Crohn's disease. Gut. 2017;66(5):813–22. https://doi.org/10.1136/gutjnl-2016-313235
9. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014;15(3):382–92. https://doi.org/10.1016/j.chom.2014.02.005
10. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119–24. https://doi.org/10.1038/nature11582

11. Lee JC, Biasci D, Roberts R, Gearry RB, Mansfield JC, Ahmad T, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. Nat Genet. 2017;49(2):262–8. https://doi.org/10.1038/ng.3755
12. Ananthakrishnan AN. Epidemiology and risk factors for IBD. Nat Rev Gastroenterol Hepatol. 2015;12(4):205–17. https://doi.org/10.1038/nrgastro.2015.34
13. Van Kruiningen HJ, Joossens M, Vermeire S, Joossens S, Debeugny S, Gower-Rousseau C, et al. Environmental factors in familial Crohn's disease in Belgium. Inflamm Bowel Dis. 2005;11(4):360–5. https://doi.org/10.1097/01.mib.0000158536.31557.90
14. Cleynen I, Vermeire S. The genetic architecture of inflammatory bowel disease: past, present and future. Curr Opin Gastroenterol. 2015;31:456–63. https://doi.org/10.1097/mog.0000000000000215
15. Verstockt B, Bressler B, Martinez-Lozano H, McGovern D, Silverberg MS. Time to revisit disease classification in inflammatory bowel disease: is the current classification of inflammatory bowel disease good enough for optimal clinical management? Gastroenterology. 2022;162(5):1370–82. https://doi.org/10.1053/j.gastro.2021.12.246
16. Huang H, Vangay P, McKinlay CE, Knights D. Multi-omics analysis of inflammatory bowel disease. Immunol Lett. 2014;162(2):62–8. https://doi.org/10.1016/j.imlet.2014.07.014
17. Polytarchou C, Koukos G, Iliopoulos D. Systems biology in inflammatory bowel diseases: ready for prime time. Curr Opin Gastroenterol. 2014;30(4):339–46. https://doi.org/10.1097/mog.0000000000000081
18. Moco S, Candela M, Chuang E, Draper C, Cominetti O, Montoliu I, et al. Systems biology approaches for inflammatory bowel disease: emphasis on gut microbial metabolism. Inflamm Bowel Dis. 2014;20(11):2104–14. https://doi.org/10.1097/mib.0000000000000116
19. Suzuki S, Takai-Igarashi T, Fukuoka Y, Wall DP, Tanaka H, Tonellato PJ. Systems analysis of inflammatory bowel disease based on comprehensive gene information. BMC Med Genet. 2012;13(1):25. https://doi.org/10.1186/1471-2350-13-25
20. Dovrolis N, Filidou E, Kolios G. Systems biology in inflammatory bowel diseases: on the way to precision medicine. Ann Gastroenterol. 2019;32:233–46.
21. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28(9):1773–84. https://doi.org/10.1038/s41591-022-01981-2
22. Fiocchi C. Omics and multi-omics in IBD: No integration, No breakthroughs. Int J Mol Sci. 2023;24(19):14912. https://doi.org/10.3390/ijms241914912
23. Javaid A, Shahab O, Adorno W, Fernandes P, May E, Syed S. Machine learning predictive outcomes modeling in inflammatory bowel diseases. Inflamm Bowel Dis. 2022;28(6):819–29. https://doi.org/10.1093/ibd/izab187
24. Seyed Tabib NS, Madgwick M, Sudhakar P, Verstockt B, Korcsmaros T, Vermeire S. Big data in IBD: big progress for clinical practice. Gut. 2020;69(8):1520–32. https://doi.org/10.1136/gutjnl-2019-320065
25. Sufyan M, Shokat Z, Ashfaq UA. Artificial intelligence in cancer diagnosis and therapy: current status and future perspective. Comput Biol Med. 2023;165:107356. https://doi.org/10.1016/j.compbiomed.2023.107356
26. Peng J, Jury EC, Dönnes P, Ciurtin C. Machine learning techniques for personalised medicine approaches in immune-mediated chronic inflammatory diseases: applications and challenges. Front Pharmacol. 2021;12:720694. https://doi.org/10.3389/fphar.2021.720694
27. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008;40(8):955–62. https://doi.org/10.1038/ng.175
28. Sazonovs A, Stevens CR, Venkataraman GR, Yuan K, Avila B, Abreu MT, et al. Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. Nat Genet. 2022;54(9):1275–83. https://doi.org/10.1038/s41588-022-01156-2
29. Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. Lancet. 2016;387(10014):156–67. https://doi.org/10.1016/s0140-6736(15)00465-1
30. McCauley JL, Abreu MT. Genetics in diagnosing and managing inflammatory bowel disease. Gastroenterol Clin North Am. 2012;41(2):513–22. https://doi.org/10.1016/j.gtc.2012.01.002
31. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health. 2018;8(2):020303. https://doi.org/10.7189/jogh.08.020303
32. Sung JJ, Stewart CL, Freedman B. Artificial intelligence in health care: preparing for the fifth Industrial Revolution. Med J Aust. 2020;213(6):253–5.e1. https://doi.org/10.5694/mja2.50755
33. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc. 2020;92(4):807–12. https://doi.org/10.1016/j.gie.2020.06.040
34. Iyengar R. Complex diseases require complex therapies. EMBO Rep. 2013;14(12):1039–42. https://doi.org/10.1038/embor.2013.177
35. Stidham RW, Takenaka K. Artificial intelligence for disease assessment in inflammatory bowel disease: how will it change our practice? Gastroenterology. 2022;162(5):1493–506. https://doi.org/10.1053/j.gastro.2021.12.238
36. Scott IA, Cook D, Coiera EW, Richards B. Machine learning in clinical practice: prospects and pitfalls. Med J Aust. 2019;211(5):203–5.e1. https://doi.org/10.5694/mja2.50294
37. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–58. https://doi.org/10.1056/nejmra1814259
38. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J. 2018;16:34–42. https://doi.org/10.1016/j.csbj.2018.01.001
39. Sutton RS, Barto AG. Reinforcement learning: an introduction. MIT press; 2018.
40. Stankovic B, Kotur N, Nikcevic G, Gasic V, Zukic B, Pavlovic S. Machine learning modeling from omics data as prospective tool for improvement of inflammatory bowel disease diagnosis and clinical classifications. Genes. 2021;12(9):1438. https://doi.org/10.3390/genes12091438
41. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 2013;4:627–35.
42. Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical AI studies. Int J Med Inform. 2021;153:104510. https://doi.org/10.1016/j.ijmedinf.2021.104510
43. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539
44. Iacucci M, Santacroce G, Zammarchi I, Maeda Y, Del Amor R, Meseguer P, et al. Artificial intelligence and endo-histo-omics: new dimensions of precision endoscopy and histology in inflammatory bowel disease. Lancet Gastroenterol Hepatol. 2024;S2468-1253(24). 00053-0.
45. 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(9):1516–23. https://doi.org/10.1097/mib.0000000000001222
46. Romagnoni A, Jégou S, Van Steen K, Wainrib G, Hugot JP, Peyrin-Biroulet L, et al. International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). Comparative performances of machine learning methods for classifying Crohn Disease patients using genome-wide genotyping data. Sci Rep. 2019;9(1):10351. https://doi.org/10.1038/s41598-019-46649-z

47. Li H, Lai L, 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):20471–82. https://doi.org/10.18632/aging.103861
48. Pal LR, Kundu K, Yin Y, Moult J. CAGI4 Crohn's exome challenge: marker SNP versus exome variant models for assigning risk of Crohn disease. Hum Mutat. 2017;38(9):1225–34. https://doi.org/10.1002/humu.23256
49. Cushing KC, Mclean R, McDonald KG, Gustafsson JK, Knoop KA, Kulkarni DH, et al. Predicting risk of postoperative disease recurrence in Crohn's disease: patients with indolent Crohn's disease have distinct whole transcriptome profiles at the time of first surgery. Inflamm Bowel Dis. 2019;25(1):180–93. https://doi.org/10.1093/ibd/izy228
50. Khorasani HM, Usefi H, Peña-Castillo L. Detecting ulcerative colitis from colon samples using efficient feature selection and machine learning. Sci Rep. 2020;10(1):13744. https://doi.org/10.1038/s41598-020-70583-0
51. Medicine PLOS, editors Advancing the beneficial use of machine learning in health care and medicine: toward a community understanding. PLoS Med. 2018;15:e1002708.
52. Che D, Liu Q, Rasheed K, Tao X. Decision tree and ensemble learning algorithms with their applications in bioinformatics. Adv Exp Med Biol. 2011;696:191–9. https://doi.org/10.1007/978-1-4419-7046-6_19
53. de Bruijne M. Machine learning approaches in medical image analysis: from detection to diagnosis. Med Image Anal. 2016;33:94–7. https://doi.org/10.1016/j.media.2016.06.032
54. Gu P, Mendonca O, Carter D, Dube S, Wang P, Huang X, et al. AIluminating artificial intelligence in inflammatory bowel diseases: a narrative review on the role of AI in endoscopy, histology, and imaging for IBD. Inflamm Bowel Dis. 2024:izae030. https://doi.org/10.1093/ibd/izae030
55. Iacucci M, Parigi TL, Del Amor R, Meseguer P, Mandelli G, Bozzola A, et al. Artificial intelligence enabled histological prediction of remission or activity and clinical outcomes in ulcerative colitis. Gastroenterology. 2023;164(7):1180–8.e2. https://doi.org/10.1053/j.gastro.2023.02.031
56. Iacucci M, Cannatelli R, Parigi TL, Nardone OM, Tontini GE, Labarile N, 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(04):332–41. https://doi.org/10.1055/a-1960-3645
57. Gu P, Chang JH, Carter D, McGovern DPB, Moore J, Wang P, et al. Radiomics-based analysis of intestinal ultrasound images for inflammatory bowel disease: a feasibility study. Crohns Colitis 360. 2024;6(2):otae034. https://doi.org/10.1093/crocol/otae034
58. Jans D, Cleynen I. The genetics of non-monogenic IBD. Hum Genet. 2023;142(5):669–82. https://doi.org/10.1007/s00439-023-02521-9
59. D'Addabbo A, Latiano A, Palmieri O, Maglietta R, Annese V, Ancona N. Regularized least squares classifiers may predict Crohn's disease from profiles of single nucleotide polymorphisms. Ann Hum Genet. 2007;71:537–49.
60. Wei Z, Wang W, Bradfield J, Li J, Cardinale C, Frackelton E, et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. Am J Hum Genet. 2013;92(6):1008–12. https://doi.org/10.1016/j.ajhg.2013.05.002
61. Giollo M, Jones DT, Carraro M, Leonardi E, Ferrari C, Tosatto SCE. Crohn disease risk prediction-Best practices and pitfalls with exome data. Hum Mutat. 2017;38(9):1193–200. https://doi.org/10.1002/humu.23177
62. Raimondi D, Simm J, Arany A, Fariselli P, Cleynen I, Moreau Y. An interpretable low-complexity machine learning framework for robust exome-based in-silico diagnosis of Crohn's disease patients. NAR Genom Bioinform. 2020;2(1):lqaa011. https://doi.org/10.1093/nargab/lqaa011
63. Wang Y, Miller M, Astrakhan Y, Petersen BS, Schreiber S, Franke A, et al. Identifying Crohn's disease signal from variome analysis. Genome Med. 2019;11(1):59. https://doi.org/10.1186/s13073-019-0670-6
64. Stafford IS, Ashton JJ, Mossotto E, Cheng G, Mark Beattie R, Ennis S. Supervised machine learning classifies inflammatory bowel disease patients by subtype using whole exome sequencing data. J Crohns Colitis. 2023;17(10):1672–80. https://doi.org/10.1093/ecco-jcc/jjad084
65. Palmieri O, Creanza TM, Bossa F, Palumbo O, Maglietta R, Ancona N, et al. Genome-wide pathway analysis using gene expression data of colonic mucosa in patients with inflammatory bowel disease. Inflamm Bowel Dis. 2015;21:1260–8. https://doi.org/10.1097/mib.0000000000000370
66. Massimino L, Lamparelli LA, Houshyar Y, D'Alessio S, Peyrin-Biroulet L, Vetrano S, et al. The inflammatory bowel disease transcriptome and metatranscriptome meta-analysis (IBD TaMMA) framework. Nat Comput Sci. 2021;1(8):511–5. https://doi.org/10.1038/s43588-021-00114-y
67. Yu S, Zhang M, Ye Z, Wang Y, Wang X, Chen YG. Development of a 32-gene signature using machine learning for accurate prediction of inflammatory bowel disease. Cell Regen. 2023;12(1):8. https://doi.org/10.1186/s13619-022-00143-6
68. Zhang L, Mao R, Lau CT, Chung WC, Chan JCP, Liang F, et al. Identification of useful genes from multiple microarrays for ulcerative colitis diagnosis based on machine learning methods. Sci Rep. 2022;12(1):9962. https://doi.org/10.1038/s41598-022-14048-6
69. Smith AM, Walsh JR, Long J, Davis CB, Henstock P, Hodge MR, et al. Standard machine learning approaches outperform deep representation learning on phenotype prediction from transcriptomics data. BMC Bioinf. 2020;21(1):119. https://doi.org/10.1186/s12859-020-3427-8
70. Wacker EM, Uellendahl-Werth F, Bej S, Wolkenhauer O, Vesterhus M, Lieb W, et al. Whole blood RNA sequencing identifies transcriptional differences between primary sclerosing cholangitis and ulcerative colitis. JHEP Rep. 2023;6(2):100988. https://doi.org/10.1016/j.jhepr.2023.100988
71. Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, et al. Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. Cell. 2019;178(6):1493–508.e20. https://doi.org/10.1016/j.cell.2019.08.008
72. Abreu MT, Davies JM, Quintero MA, Delmas A, Diaz S, Martinez CD, et al. Transcriptional behavior of regulatory T cells predicts IBD patient responses to vedolizumab therapy. Inflamm Bowel Dis. 2022;28(12):1800–12. https://doi.org/10.1093/ibd/izac151
73. Gisbert JP, Chaparro M. Clinical usefulness of proteomics in inflammatory bowel disease: a comprehensive review. J Crohns Colitis. 2019;13(3):374–84. https://doi.org/10.1093/ecco-jcc/jjy158
74. Ungaro RC, Hu L, Ji J, Nayar S, Kugathasan S, Denson LA, et al. Machine learning identifies novel blood protein predictors of penetrating and stricturing complications in newly diagnosed paediatric Crohn's disease. Aliment Pharmacol Ther. 2021;53(2):281–90. https://doi.org/10.1111/apt.16136
75. Jongsma MME, Costes LMM, Tindemans I, Cozijnsen MA, Raatgreep RHC, van Pieterson M, et al. Serum immune profiling in paediatric Crohn's disease demonstrates stronger immune modulation with first-line infliximab than conventional therapy and pre-treatment profiles predict clinical response to both treatments. J Crohns Colitis. 2023;17(8):1262–77. https://doi.org/10.1093/ecco-jcc/jjad049
76. Meade S, Liu Chen Kiow J, Massaro C, Kaur G, Squirell E, Bressler B, et al. Gut microbiome-associated predictors as biomarkers of response to advanced therapies in inflammatory bowel disease: a

systematic review. Gut Microb. 2023;15(2):2287073. https://doi.org/10.1080/19490976.2023.2287073
77. Manandhar I, Alimadadi A, Aryal S, Munroe PB, Joe B, Cheng X. Gut microbiome-based supervised machine learning for clinical diagnosis of inflammatory bowel diseases. Am J Physiol Gastrointest Liver Physiol. 2021;320(3):G328–37. https://doi.org/10.1152/ajpgi.00360.2020
78. Wang X, Xiao Y, Xu X, Guo L, Yu Y, Li N, et al. Characteristics of fecal microbiota and machine learning strategy for fecal invasive biomarkers in pediatric inflammatory bowel disease. Front Cell Infect Microbiol. 2021;11:711884. https://doi.org/10.3389/fcimb.2021.711884
79. Vich Vila A, Imhann F, Collij V, Jankipersadsing SA, Gurry T, Mujagic Z, et al. Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. Sci Transl Med. 2018;10(472):eaap8914. https://doi.org/10.1126/scitranslmed.aap8914
80. Zuo W, Wang B, Bai X, Luan Y, Fan Y, Michail S, et al. 16S rRNA and metagenomic shotgun sequencing data revealed consistent patterns of gut microbiome signature in pediatric ulcerative colitis. Sci Rep. 2022;12(1):6421. https://doi.org/10.1038/s41598-022-07995-7
81. Ananthakrishnan AN, Luo C, Yajnik V, Khalili H, Garber JJ, Stevens BW, et al. Gut microbiome function predicts response to anti-integrin biologic therapy in inflammatory bowel diseases. Cell Host Microbe. 2017;21(5):603–10.e3. https://doi.org/10.1016/j.chom.2017.04.010
82. Raygoza Garay JA, Turpin W, Lee SH, Smith MI, Goethel A, Griffiths AM, et al. Gut microbiome composition is associated with future onset of Crohn's disease in healthy first-degree relatives. Gastroenterology. 2023;165(3):670–81. https://doi.org/10.1053/j.gastro.2023.05.032
83. Caenepeel C, Falony G, Machiels K, Verstockt B, Goncalves PJ, Ferrante M, et al. Dysbiosis and associated stool features improve prediction of response to biological therapy in inflammatory bowel disease. Gastroenterology. 2024;166:483–95. https://doi.org/10.1053/j.gastro.2023.11.304
84. Borren NZ, Plichta D, Joshi AD, Bonilla G, Sadreyev R, Vlamakis H, et al. Multi-"-Omics" profiling in patients with quiescent inflammatory bowel disease identifies biomarkers predicting relapse. Inflamm Bowel Dis. 2020;26(10):1524–32. https://doi.org/10.1093/ibd/izaa183
85. Gardiner LJ, Carrieri AP, Bingham K, Macluskie G, Bunton D, McNeil M, et al. Combining explainable machine learning, demographic and multi-omic data to inform precision medicine strategies for inflammatory bowel disease. PLoS One. 2022;17(2):e0263248. https://doi.org/10.1371/journal.pone.0263248
86. Lee JWJ, Plichta D, Hogstrom L, Borren NZ, Lau H, Gregory SM, et al. Multi-omics reveal microbial determinants impacting responses to biologic therapies in inflammatory bowel disease. Cell Host Microbe. 2021;29(8):1294–304.e4. https://doi.org/10.1016/j.chom.2021.06.019
87. Arehart CH, Sterrett JD, Garris RL, Quispe-Pilco RE, Gignoux CR, Evans LM, et al. Poly-omic risk scores predict inflammatory bowel disease diagnosis. mSystems. 2023;9(1):e0067723. https://doi.org/10.1128/msystems.00677-23
88. Ning L, Zhou YL, Sun H, Zhang Y, Shen C, Wang Z, et al. Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. Nat Commun. 2023;14(1):7135. https://doi.org/10.1038/s41467-023-42788-0
89. Mishra N, Aden K, Blase JI, Baran N, Bordoni D, Tran F, 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. https://doi.org/10.1186/s13073-022-01112-z
90. Gao X, Sun R, Jiao N, Liang X, Li G, Gao H, et al. Integrative multi-omics deciphers the spatial characteristics of host-gut microbiota interactions in Crohn's disease. Cell Rep Med. 2023;4(6):101083. https://doi.org/10.1016/j.xcrm.2023.101050
91. Iacucci M, Jeffery L, Acharjee A, Grisan E, Buda A, Nardone OM, et al. Computer-aided imaging analysis of probe-based confocal laser endomicroscopy with molecular labeling and gene expression identifies markers of response to biological therapy in IBD patients: the endo-omics study. Inflamm Bowel Dis. 2023;29(9):1409–20. https://doi.org/10.1093/ibd/izac233
92. Nardone OM, Zammarchi I, Santacroce G, Ghosh S, Iacucci M. Inflammation-driven colorectal cancer associated with colitis: from pathogenesis to changing therapy. Cancers (Basel). 2023;15(8):2389. https://doi.org/10.3390/cancers15082389

How to cite this article: Cannarozzi AL, Latiano A, Massimino L, Bossa F, Giuliani F, Riva M, et al. Inflammatory bowel disease genomics, transcriptomics, proteomics and metagenomics meet artificial intelligence. United European Gastroenterol J. 2024;12(10):1461–80. https://doi.org/10.1002/ueg2.12655