Scientific journal of Research studies in Future Basic Sciences and medical sciences, 2026, Vol. 4, No. 1, pp. 128-157

Scientific journal of Research studies
in Future Basic Sciences and medical sciences
www.journalhi.com/sci/

Artificial Intelligence-Driven Multimodal Prediction of Disease Severity and Biologic Therapy Response in Inflammatory Bowel Disease Using Integrated Clinical Endoscopic and Biomarker Data

Mohammad Karami Horestani [1,*]

1- Assistant Professor of Gastroenterology and Hepatology, Department of Internal Medicine, Clinical Research Development Unit (CRDU), Hajar Hospital, Shahrekord University of Medical Sciences, Shahrekord, Iran. Email: karami.m@skums.ac.ir
* Corresponding Author

Abstract

Inflammatory bowel disease (IBD), encompassing Crohn's disease and ulcerative colitis, is characterized by substantial heterogeneity in clinical manifestations, disease progression, treatment responsiveness, and long-term outcomes. Despite considerable advances in biologic therapies, therapeutic decision-making remains challenging because patients frequently exhibit variable responses to treatment and unpredictable disease trajectories. Conventional clinical assessment approaches rely heavily on isolated biomarkers, endoscopic findings, or physician experience, often limiting the ability to accurately identify patients at risk of severe disease progression or treatment failure. Recent developments in artificial intelligence (AI) and machine learning have created opportunities to integrate diverse clinical, endoscopic, and molecular datasets into predictive frameworks capable of supporting personalized disease management. This study presents a multimodal artificial intelligence framework for predicting disease severity and biologic therapy response in patients with inflammatory bowel disease through the integration of clinical characteristics, endoscopic activity indices, laboratory biomarkers, inflammatory markers, and multi-omics-derived indicators. The proposed framework combines supervised machine learning algorithms, feature selection techniques, and explainable artificial intelligence methods to identify the most informative predictors associated with disease activity and therapeutic outcomes. Variables incorporated into the predictive model include demographic characteristics, disease duration, C-reactive protein levels, fecal calprotectin concentrations, endoscopic severity scores, microbiome-associated signatures, and prior treatment histories. Model performance was evaluated using multiple classification metrics, including accuracy, sensitivity, specificity, area under the receiver operating characteristic curve, and predictive value assessments. Comparative analyses demonstrated that multimodal models consistently outperformed single-domain approaches by capturing complex interactions among clinical, endoscopic, and biomarker variables. Feature importance analysis further revealed that inflammatory biomarkers, endoscopic severity measurements, and microbiome-derived indicators contributed substantially to prediction accuracy. The findings support the growing role of artificial intelligence in precision gastroenterology and suggest that integrated predictive systems may facilitate earlier identification of high-risk patients, optimization of biologic treatment strategies, reduction of unnecessary therapeutic escalation, and improvement of long-term clinical outcomes. The proposed framework provides a clinically interpretable and scalable approach for advancing individualized care in inflammatory bowel disease and offers a foundation for future implementation of data-driven decision support systems in routine clinical practice.

Keywords: Inflammatory Bowel Disease, Artificial Intelligence, Machine Learning, Biologic Therapy, Precision Medicine

Introduction

Inflammatory bowel disease (IBD) represents a group of chronic immune-mediated gastrointestinal disorders primarily comprising Crohn's disease (CD) and ulcerative colitis (UC). The global burden of IBD has increased substantially during the past two decades, with rising incidence and prevalence reported across both industrialized and developing regions. This growing epidemiological trend has transformed IBD into a major public health challenge, generating significant healthcare expenditures, reduced quality of life, recurrent hospitalization, and long-term disability among affected individuals [1]. One of the most challenging aspects of IBD management is the remarkable heterogeneity observed among patients. Individuals diagnosed with similar clinical phenotypes frequently experience substantially different disease trajectories, ranging from prolonged

remission to aggressive disease progression requiring repeated hospitalization, surgery, or therapeutic escalation. Consequently, clinicians face considerable uncertainty when attempting to predict future disease severity or determine the most appropriate treatment strategy for a particular patient [1,3]. The introduction of biologic therapies has revolutionized IBD treatment by targeting specific inflammatory pathways involved in disease pathogenesis. Anti-tumor necrosis factor agents, anti-integrin therapies, and interleukin inhibitors have demonstrated substantial efficacy in inducing and maintaining remission. Nevertheless, primary non-response, secondary loss of response, and variable therapeutic durability remain major clinical concerns. A substantial proportion of patients fail to achieve sustained remission despite receiving advanced biologic treatment, highlighting the urgent need for more accurate predictive tools capable of identifying likely responders before treatment initiation [2]. Traditional clinical decision-making relies primarily on physician assessment, laboratory markers, endoscopic findings, and patient-reported symptoms. Although these indicators provide valuable information, they often fail to capture the complex biological interactions underlying disease progression and therapeutic responsiveness. The multifactorial nature of IBD involves intricate relationships among host genetics, immune dysregulation, intestinal microbiota, environmental exposures, metabolic pathways, and mucosal inflammatory processes. As a result, isolated clinical variables frequently possess limited predictive value when considered independently [3].

Recent advances in artificial intelligence (AI) and machine learning have created unprecedented opportunities to overcome these limitations by enabling the analysis of large-scale, multidimensional healthcare datasets. Unlike conventional statistical approaches, machine learning algorithms can identify nonlinear relationships, complex feature interactions, and latent patterns that may remain undetected using traditional analytical methods. These capabilities have stimulated increasing interest in the application of AI-driven predictive models across multiple domains of gastroenterology, including disease diagnosis, risk stratification, prognosis, treatment optimization, and clinical decision support [2]. Inflammatory bowel disease (IBD) represents a group of chronic immune-mediated gastrointestinal disorders primarily comprising Crohn's disease (CD) and ulcerative colitis (UC). The global burden of IBD has increased substantially during the past two decades, with rising incidence and prevalence reported across both industrialized and developing regions. This growing epidemiological trend has transformed IBD into a major public health challenge, generating significant healthcare expenditures, reduced quality of life, recurrent hospitalization, and long-term disability among affected individuals [1]. One of the most challenging aspects of IBD management is the remarkable heterogeneity observed among patients. Individuals diagnosed with similar clinical phenotypes frequently experience substantially different disease trajectories, ranging from prolonged remission to aggressive disease progression requiring repeated hospitalization, surgery, or therapeutic escalation. Consequently, clinicians face considerable uncertainty when attempting to predict future disease severity or determine the most appropriate treatment strategy for a particular patient [1,3]. The introduction of biologic therapies has revolutionized IBD treatment by targeting specific inflammatory pathways involved in disease pathogenesis. Anti-tumor necrosis factor agents, anti-integrin therapies, and interleukin inhibitors have demonstrated substantial efficacy in inducing and maintaining remission. Nevertheless, primary non-response, secondary loss of response, and variable therapeutic durability remain major clinical concerns. A substantial proportion of patients fail to achieve sustained remission despite receiving advanced biologic treatment, highlighting the urgent need for more accurate predictive tools capable of identifying likely responders before treatment initiation [2].

Traditional clinical decision-making relies primarily on physician assessment, laboratory markers, endoscopic findings, and patient-reported symptoms. Although these indicators provide valuable information, they often fail to capture the complex biological interactions underlying disease progression and therapeutic responsiveness. The multifactorial nature of IBD involves intricate relationships among host genetics, immune dysregulation, intestinal microbiota, environmental exposures, metabolic pathways, and mucosal inflammatory processes. As a result, isolated clinical variables frequently possess limited predictive value when considered independently [3]. Recent advances in artificial intelligence (AI) and machine learning have created unprecedented opportunities to overcome these limitations by enabling the analysis of large-scale, multidimensional healthcare datasets. Unlike conventional statistical approaches, machine learning algorithms can identify nonlinear relationships, complex feature interactions, and latent patterns that may remain undetected using traditional analytical methods. These capabilities have stimulated increasing interest in the application of AI-driven predictive models across multiple domains of gastroenterology, including disease diagnosis, risk stratification, prognosis, treatment optimization, and clinical decision support [2]. Increasing evidence suggests that the intestinal microbiome plays a fundamental role in the pathophysiology of inflammatory bowel disease and significantly influences both disease behavior and therapeutic responsiveness. The human gastrointestinal tract contains a highly complex microbial ecosystem that participates in immune regulation, nutrient metabolism, epithelial barrier maintenance, and host defense. In patients with IBD, alterations in microbial diversity, microbial composition, and metabolic activity have been consistently

associated with disease onset, progression, and recurrence. These observations have stimulated extensive research aimed at identifying microbiome-derived biomarkers capable of supporting disease stratification and personalized treatment strategies [12]. Recent multi-cohort investigations integrating microbiome and metabolome datasets have revealed that inflammatory bowel disease is characterized not only by shifts in microbial taxa but also by profound alterations in metabolic pathways associated with host-microbe interactions. Specific microbial signatures have been linked to active inflammation, mucosal healing, and differential responses to biologic therapies. Furthermore, metabolomic analyses have demonstrated that variations in short-chain fatty acid production, amino acid metabolism, bile acid transformation, and microbial-derived metabolites may provide valuable indicators of disease activity and therapeutic outcomes. The incorporation of these variables into predictive models has significantly expanded the scope of precision medicine in gastroenterology [12].

The concept of microbiome-based risk profiling has emerged as a particularly promising area of investigation. Rather than focusing on individual microbial species, contemporary approaches seek to characterize integrated microbial ecosystems and their functional capacities. Such profiles may provide a more comprehensive representation of disease-associated biological processes and allow clinicians to identify patients who are at elevated risk for disease progression, treatment resistance, or recurrent inflammatory activity. As computational methods become increasingly sophisticated, artificial intelligence algorithms have demonstrated substantial capacity to process these complex datasets and transform them into clinically meaningful predictive indicators [13]. Simultaneously, the growing availability of longitudinal healthcare data has enhanced opportunities for predictive modeling in inflammatory bowel disease. Electronic health records, laboratory databases, imaging repositories, and treatment registries now provide extensive information regarding disease evolution over time. These data sources enable machine learning algorithms to identify temporal patterns that may not be apparent through conventional clinical evaluation. By analyzing longitudinal trajectories of inflammatory markers, endoscopic activity, medication exposure, and healthcare utilization, predictive models can generate individualized risk estimates that evolve throughout the disease course and adapt to changing clinical circumstances [6]. Another important development involves the expanding application of advanced imaging analytics within inflammatory bowel disease management. Radiological modalities such as magnetic resonance enterography, computed tomography enterography, and intestinal ultrasound increasingly contribute to disease assessment, particularly in Crohn's disease. Artificial intelligence techniques have enhanced the interpretation of these imaging modalities through automated lesion detection, inflammation quantification, tissue characterization, and outcome prediction. The integration of imaging-derived features with clinical and molecular information further strengthens the predictive capabilities of multimodal models and contributes to a more comprehensive understanding of disease activity [7]. Despite these advances, several challenges continue to limit the routine clinical implementation of artificial intelligence systems in inflammatory bowel disease. Variability in data quality, differences among healthcare institutions, limited external validation, and concerns regarding model interpretability remain important barriers. Furthermore, many existing prediction models rely on single-domain datasets and therefore fail to capture the multidimensional nature of disease biology. Addressing these limitations requires the development of integrated frameworks capable of combining clinical, endoscopic, biomarker, microbiome, and imaging information into unified predictive systems that can be applied across diverse patient populations [6,7].

Consequently, the integration of microbiome science, multi-omics technologies, longitudinal healthcare data, advanced imaging, and artificial intelligence represents one of the most significant frontiers in contemporary inflammatory bowel disease research. These developments provide the scientific foundation for next-generation predictive models that may transform disease monitoring, optimize therapeutic decision-making, and advance the broader goals of precision medicine in chronic inflammatory disorders [12,13]. The rapid expansion of biologic therapies has fundamentally transformed the treatment landscape of inflammatory bowel disease. Therapeutic agents targeting tumor necrosis factor, integrin-mediated leukocyte trafficking, and interleukin signaling pathways have significantly improved disease control and reduced the need for surgical intervention in many patients. Nevertheless, substantial variability remains in both short-term and long-term treatment outcomes. Some individuals achieve sustained clinical and endoscopic remission, whereas others demonstrate primary non-response, secondary loss of response, or progressive disease despite repeated therapeutic modifications. This variability represents one of the most important unresolved challenges in modern IBD management and highlights the need for more accurate predictive strategies capable of supporting individualized treatment selection [8]. Recent machine learning studies have demonstrated that predictive models integrating demographic characteristics, laboratory biomarkers, disease history, treatment exposure, and endoscopic findings can identify factors associated with therapeutic success more effectively than conventional statistical approaches. In particular, machine learning frameworks have shown considerable promise in predicting remission and treatment response among patients receiving vedolizumab and

ustekinumab. These findings suggest that data-driven prediction systems may contribute to earlier optimization of treatment pathways and reduce the burden associated with ineffective therapeutic interventions [8]. Similarly, the application of artificial intelligence to biologic treatment optimization has emerged as a major area of investigation. Advanced predictive algorithms can evaluate large numbers of clinical variables simultaneously and estimate the probability of treatment success for individual patients before therapy initiation. Such capabilities are particularly important in inflammatory bowel disease because delayed recognition of ineffective therapy may contribute to ongoing intestinal damage, increased healthcare utilization, reduced quality of life, and higher long-term costs. Predictive systems capable of guiding therapeutic decision-making may therefore improve both clinical outcomes and healthcare efficiency [9]. The potential impact of artificial intelligence extends beyond routine clinical practice and into the design and execution of clinical trials. Contemporary IBD trials frequently encounter challenges related to patient heterogeneity, endpoint variability, recruitment efficiency, and treatment-response assessment. Artificial intelligence methodologies offer opportunities to improve patient stratification, identify suitable trial populations, enhance endpoint prediction, and facilitate adaptive trial designs. These innovations may accelerate therapeutic development and improve the translation of research findings into clinical practice [10]. In parallel, non-invasive assessment modalities continue to gain importance within contemporary disease monitoring strategies. Intestinal ultrasound has emerged as an increasingly valuable tool for evaluating bowel inflammation, disease extent, and therapeutic response. Compared with more invasive procedures, ultrasound offers advantages related to safety, accessibility, repeatability, and patient acceptance. The incorporation of ultrasound-derived parameters into multimodal predictive frameworks may further strengthen disease monitoring and enhance individualized risk assessment, particularly when combined with laboratory, endoscopic, and molecular information [14].

The broader vision underlying these developments is the transition from population-based treatment paradigms toward precision medicine. Precision medicine seeks to deliver the right therapy to the right patient at the right time by integrating clinical observations with biological, molecular, and computational insights. Within inflammatory bowel disease, achieving this objective requires predictive systems capable of capturing the complexity of disease behavior and translating multidimensional data into actionable clinical recommendations. Artificial intelligence provides a powerful mechanism for accomplishing this goal and has become a central component of next-generation personalized healthcare strategies [15]. Against this background, the present study proposes an artificial intelligence-driven multimodal framework for predicting disease severity and biologic therapy response in inflammatory bowel disease through the integration of clinical characteristics, endoscopic findings, inflammatory biomarkers, and advanced biological indicators. By combining multiple sources of patient information within a unified predictive architecture, the study aims to enhance risk stratification, support therapeutic decision-making, and contribute to the ongoing development of precision medicine approaches in gastroenterology.

Problem Statement

Although substantial progress has been achieved in the treatment of inflammatory bowel disease (IBD) through the introduction of biologic agents and personalized therapeutic strategies, the prediction of disease severity and treatment response remains one of the most complex challenges in contemporary gastroenterology. Current clinical practice continues to rely largely on a combination of physician judgment, symptom assessment, laboratory markers, endoscopic findings, and periodic imaging evaluations. While these approaches provide valuable information regarding disease status, they frequently fail to accurately predict future disease progression or identify patients who are most likely to benefit from a specific biologic therapy before treatment initiation [6,15]. A major limitation of existing predictive approaches is their dependence on isolated categories of information. Many clinical decision-making frameworks evaluate laboratory biomarkers, endoscopic scores, imaging findings, or patient characteristics independently rather than considering the dynamic interactions among these variables. However, inflammatory bowel disease is a highly complex and multifactorial condition influenced by immunological, microbial, genetic, environmental, and metabolic mechanisms that operate simultaneously and continuously throughout the disease course [4,12]. As a consequence, single-domain prediction models often demonstrate limited generalizability and reduced predictive accuracy when applied across heterogeneous patient populations.

Another important challenge involves therapeutic heterogeneity among patients receiving biologic treatments. Despite advances in anti-TNF agents, anti-integrin therapies, and interleukin inhibitors, considerable variation exists in both primary and long-term treatment responses. Some patients achieve sustained remission and mucosal healing, whereas others experience persistent inflammation, early relapse, treatment failure, or progressive structural damage despite receiving similar therapeutic regimens [8,9]. The inability to reliably

predict these outcomes before treatment initiation may result in delayed disease control, unnecessary exposure to ineffective medications, increased healthcare expenditures, and diminished quality of life. Furthermore, recent advances in endoscopy, molecular diagnostics, microbiome science, and multi-omics technologies have generated large volumes of clinically relevant data. Although these datasets contain valuable information regarding disease activity and therapeutic responsiveness, their complexity exceeds the analytical capacity of conventional statistical approaches in many circumstances. Consequently, a substantial proportion of potentially informative clinical knowledge remains underutilized in routine practice. Artificial intelligence and machine learning offer the capability to process high-dimensional data, identify nonlinear relationships among variables, and generate predictive insights that are difficult to obtain through traditional methodologies [1,2].

A review of recent literature indicates that many published machine learning studies in inflammatory bowel disease have focused on individual outcomes such as disease relapse, treatment response, hospitalization risk, or imaging interpretation. However, relatively few investigations have developed integrated multimodal frameworks capable of simultaneously incorporating clinical variables, endoscopic assessments, inflammatory biomarkers, and emerging biological indicators within a unified predictive architecture [5,6]. In addition, many existing models have been developed using restricted datasets, limiting their applicability across diverse clinical environments. The scientific gap addressed by the present study therefore lies in the absence of a comprehensive predictive framework that integrates multiple dimensions of patient information into a clinically interpretable and operational decision-support model. The central hypothesis of this research is that combining clinical characteristics, endoscopic findings, inflammatory biomarkers, and advanced biological indicators through artificial intelligence algorithms can significantly improve the prediction of disease severity and biologic therapy response compared with conventional assessment methods.

Accordingly, this study seeks to develop and evaluate a multimodal artificial intelligence framework capable of supporting individualized risk stratification and therapeutic decision-making in inflammatory bowel disease. By integrating diverse sources of patient information within a unified predictive model, the research aims to contribute to the advancement of precision medicine and improve the effectiveness of personalized treatment strategies in patients with chronic inflammatory bowel disorders.

Materials and Methods

Study Design

This study was designed as a multimodal predictive modeling investigation aimed at developing an artificial intelligence framework for forecasting disease severity and biologic therapy response in patients diagnosed with inflammatory bowel disease (IBD). The methodological framework integrates clinical characteristics, endoscopic assessments, laboratory biomarkers, inflammatory indicators, and advanced biological variables into a unified machine learning architecture. The study follows a retrospective-prospective analytical design frequently employed in contemporary precision medicine research, allowing the evaluation of both baseline patient characteristics and longitudinal treatment outcomes.

The overall workflow consisted of five sequential stages: data acquisition, data preprocessing, feature engineering, predictive model development, and model evaluation. These stages were designed to ensure reproducibility, minimize bias, and maximize predictive performance while maintaining clinical interpretability.

Study Population

The study population consisted of adult patients diagnosed with either Crohn's disease or ulcerative colitis according to internationally accepted diagnostic criteria. Patients included in the analytical dataset fulfilled the following criteria:

Inclusion Criteria

- Confirmed diagnosis of Crohn's disease or ulcerative colitis.
- Age ≥ 18 years.
- Availability of complete baseline clinical records.
- Availability of endoscopic assessment performed within the predefined evaluation period.
- Availability of inflammatory biomarker measurements.
- Treatment with at least one biologic agent.

- Minimum follow-up duration of 12 months.

Exclusion Criteria

- Incomplete clinical records.
- Missing endoscopic assessment data.
- Presence of severe concomitant autoimmune disorders.
- History of malignant gastrointestinal disease.
- Pregnancy during the study period.
- Follow-up duration shorter than 12 months.

These criteria were selected to ensure the consistency and reliability of predictive analyses while minimizing confounding influences that could affect treatment outcomes.

Data Sources and Variable Categories

The predictive framework incorporated multiple categories of patient information to capture the multidimensional nature of inflammatory bowel disease.

Clinical Variables

Clinical parameters included:

- Age
- Sex
- Body mass index
- Disease duration
- Smoking status
- Disease subtype
- Disease location
- Disease behavior
- Previous surgery history
- Previous biologic exposure
- Concomitant medication use
- Hospitalization history

Endoscopic Variables

Endoscopic assessments represented a major component of disease severity evaluation.

For ulcerative colitis:

- Mayo Endoscopic Subscore

For Crohn's disease:

- Simple Endoscopic Score for Crohn's Disease (SES-CD)

Additional endoscopic features included:

- Mucosal healing status
- Presence of ulceration
- Extent of inflammatory involvement
- Stricture formation
- Fistulizing disease

The incorporation of endoscopic indicators was based on evidence demonstrating their strong association with long-term disease outcomes and therapeutic response [5].

Biomarker Variables

The biomarker component of the model included routinely available inflammatory markers and emerging biological indicators.

Laboratory parameters included:

- C-reactive protein (CRP)
- Fecal calprotectin
- Erythrocyte sedimentation rate (ESR)
- Hemoglobin
- Albumin
- White blood cell count
- Platelet count

Advanced biological indicators included:

- Microbiome-derived signatures
- Metabolic activity indices
- Immune-inflammatory profiles

These variables were selected because previous studies have demonstrated their importance in monitoring disease activity and predicting therapeutic outcomes [11,12].

Outcome Definition

Two primary prediction targets were established.

Outcome 1: Disease Severity Classification

Patients were categorized into:

- Mild disease activity
- Moderate disease activity
- Severe disease activity

Classification was determined using a combination of clinical assessment, biomarker status, and endoscopic findings.

Outcome 2: Biologic Therapy Response

Therapeutic response was evaluated at predefined follow-up intervals.

Patients were classified as:

- Sustained responder
- Partial responder
- Non-responder

Response status was determined according to:

- Clinical remission status
- Endoscopic improvement
- Biomarker normalization
- Need for treatment escalation
- Requirement for surgery or hospitalization

The use of these multidimensional outcome definitions enabled the development of predictive models that more closely reflect real-world clinical decision-making processes and contemporary precision medicine objectives [8,9].

Data Preprocessing

Prior to model development, all collected data underwent a comprehensive preprocessing procedure to ensure consistency, reliability, and analytical suitability. Because multimodal datasets frequently originate from different clinical sources and measurement platforms, standardization was considered an essential prerequisite for predictive modeling.

Data preprocessing consisted of the following stages:

1. Data integrity assessment
2. Missing value management
3. Outlier detection
4. Feature normalization
5. Variable encoding
6. Dimensionality reduction

Missing values were evaluated according to variable type and missingness pattern. Continuous variables with limited missing observations were imputed using median-based approaches, while categorical variables were imputed using mode-frequency estimation. Variables exhibiting excessive incompleteness were excluded from model construction to avoid introducing systematic bias.

Outlier detection was performed using interquartile range analysis and standardized z-score assessment. Extreme observations attributable to measurement errors or data-entry inconsistencies were removed following clinical verification procedures.

To minimize scale-related distortions among variables, continuous predictors were normalized using standard scaling techniques:

$$z = \frac{\mathrm{X} - \mu}{\sigma}$$

Where:

- Z = standardized value
- X = original observation
- μ = variable mean
- σ = standard deviation

This normalization procedure ensured that variables measured on different scales contributed proportionally during model training.

Feature Engineering

Feature engineering was implemented to enhance predictive performance and capture clinically meaningful relationships among variables.

Several composite features were generated from existing clinical and laboratory parameters, including:

- Inflammation burden indices
- Biomarker interaction scores
- Longitudinal disease activity trends
- Treatment exposure indicators
- Endoscopic severity aggregates

Feature generation aimed to transform raw clinical measurements into representations that more accurately reflected disease behavior and therapeutic dynamics.

To identify the most informative predictors, feature selection was conducted using a combination of:

- Recursive Feature Elimination (RFE)
- Information Gain Ranking
- Random Forest Feature Importance
- Mutual Information Analysis

The selected variables were subsequently incorporated into model development to reduce dimensionality while preserving predictive information.

The feature selection process was particularly important because modern IBD datasets frequently contain large numbers of correlated variables derived from clinical assessments, laboratory investigations, endoscopy, microbiome profiling, and treatment records [4,12].

Machine Learning Architecture

A multimodel machine learning framework was implemented to evaluate predictive performance across different algorithmic paradigms.

The following supervised learning algorithms were selected:

Random Forest (RF)

Random Forest was chosen because of its ability to:

- Handle nonlinear relationships
- Manage high-dimensional datasets
- Reduce overfitting
- Generate feature importance rankings

Extreme Gradient Boosting (XGBoost)

XGBoost was included due to:

- Strong predictive accuracy
- Efficient handling of missing data
- Robust performance with heterogeneous clinical datasets

Support Vector Machine (SVM)

Support Vector Machines were utilized to identify optimal classification boundaries between disease severity categories and treatment-response groups.

Artificial Neural Network (ANN)

Artificial Neural Networks were incorporated to capture highly complex interactions among multimodal variables and identify latent nonlinear patterns.

The ANN prediction function can be represented as:

$$y = f(\sum_{i=1}^{n} \omega_i x_i + b)$$

Where:

- y = predicted outcome
- x_i = input feature
- w_i = connection weight
- b = bias term
- f = activation function

The neural network architecture included:

- Input layer
- Two hidden layers
- Output classification layer

This configuration was selected to balance predictive complexity and model interpretability.

Multimodal Data Fusion Strategy

A key methodological innovation of this study involved multimodal feature integration.

Rather than analyzing clinical, endoscopic, and biomarker variables independently, all data streams were merged into a unified analytical framework.

The multimodal integration process consisted of:

1. Clinical feature extraction
2. Endoscopic feature extraction
3. Biomarker feature extraction
4. Biological indicator integration
5. Unified feature matrix generation

The final multimodal dataset allowed machine learning algorithms to evaluate interactions among variables originating from different biological and clinical domains simultaneously.

Previous research suggests that multimodal integration substantially improves prediction accuracy compared with single-domain analytical approaches [4,5].

Hyperparameter Optimization

To maximize model performance, hyperparameter optimization was conducted using Grid Search and Cross-Validation procedures.

Optimization parameters included:

Random Forest

- Number of trees
- Maximum tree depth
- Minimum samples per split

XGBoost

- Learning rate
- Maximum depth
- Subsampling ratio
- Number of estimators

SVM

- Kernel type
- Regularization coefficient
- Gamma parameter

ANN

- Number of neurons
- Learning rate
- Batch size

- Epoch number

The optimal parameter configuration was selected according to validation performance metrics and generalization capability.

Model Validation Strategy

To ensure robustness, reproducibility, and generalizability of the predictive framework, a rigorous validation strategy was implemented throughout the model development process. The complete dataset was randomly divided into training and testing subsets while maintaining proportional representation of disease severity categories and treatment-response groups.

The dataset partitioning procedure consisted of:

- Training dataset: 80%
- Testing dataset: 20%

The training dataset was used for model development and parameter optimization, whereas the testing dataset was reserved exclusively for final performance evaluation.

To reduce the risk of overfitting and improve model stability, k-fold cross-validation was employed during model training. A five-fold cross-validation strategy was selected because it provides an effective balance between computational efficiency and predictive reliability.

The procedure involved:

1. Dividing the training dataset into five equal subsets.
2. Using four subsets for model training.
3. Using the remaining subset for validation.
4. Repeating the process five times.
5. Calculating average performance metrics across all iterations.

This validation approach ensured that model performance was not dependent upon a particular data partition and enhanced confidence in the reproducibility of predictive outcomes.

Performance Evaluation Metrics

Multiple complementary performance indicators were utilized to assess predictive accuracy and clinical utility.

Classification Accuracy

Accuracy was calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Accuracy measures the overall proportion of correctly classified observations.

Sensitivity

Sensitivity evaluates the ability of the model to correctly identify positive cases.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Higher sensitivity is particularly important for identifying patients at elevated risk of severe disease progression.

Specificity

Specificity evaluates the ability of the model to correctly identify negative cases.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

High specificity reduces false-positive classifications and unnecessary therapeutic interventions.

Precision

Precision was calculated as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

This metric reflects the reliability of positive predictions generated by the model.

F1 Score

The F1 score provides a balanced measure of precision and sensitivity.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

This indicator is particularly useful when outcome classes are not evenly distributed.

Receiver Operating Characteristic Analysis

Receiver Operating Characteristic (ROC) curves were generated for each predictive model.

The Area Under the Curve (AUC) was used as a global indicator of classification performance.

General interpretation of AUC values:

- 0.50–0.60 = Poor
- 0.60–0.70 = Fair
- 0.70–0.80 = Acceptable
- 0.80–0.90 = Excellent
- 0.90 = Outstanding

ROC-AUC analysis provided a comprehensive assessment of discrimination capability across different classification thresholds.

Explainable Artificial Intelligence Analysis

Although machine learning models can achieve high predictive accuracy, their clinical adoption often depends on interpretability and transparency. To address this requirement, Explainable Artificial Intelligence (XAI) techniques were incorporated into the analytical framework.

The primary explainability method utilized in this study was SHAP (SHapley Additive exPlanations).

SHAP analysis was selected because it:

- Quantifies the contribution of individual variables.
- Explains model predictions at both global and local levels.
- Improves transparency of complex machine learning systems.
- Facilitates clinical interpretation of prediction outcomes.

The SHAP framework enabled identification of the variables exerting the greatest influence on disease severity prediction and biologic treatment-response estimation.

Particular attention was given to:

- Endoscopic severity scores
- Fecal calprotectin levels

- C-reactive protein
- Disease duration
- Previous biologic exposure
- Microbiome-associated indicators

This approach strengthened the clinical relevance of the predictive framework and enhanced confidence in its potential application within routine healthcare environments [6,15].

Statistical Analysis

Descriptive analyses were performed to characterize patient populations and evaluate variable distributions.

Continuous variables were summarized using:

- Mean
- Standard deviation
- Median
- Interquartile range

Categorical variables were summarized using:

- Frequencies
- Percentages

Comparative analyses were conducted between disease severity groups and treatment-response categories.

The significance threshold was established as:

$p<0.05$

All analyses were performed using established statistical and machine learning environments designed for large-scale biomedical data analysis.

Ethical Considerations

The methodological framework was designed in accordance with internationally recognized ethical principles governing biomedical research.

Key ethical considerations included:

- Protection of patient confidentiality.
- Removal of direct personal identifiers.
- Secure storage of clinical information.
- Use of aggregated analytical outputs.
- Compliance with applicable research governance standards.

The artificial intelligence framework was developed exclusively for research and decision-support purposes and was not intended to replace clinical judgment or physician oversight.

Methodological Workflow

The overall workflow of the study is summarized below:

1. Patient selection
2. Data collection
3. Data preprocessing
4. Feature engineering
5. Multimodal data integration
6. Machine learning model development

7. Hyperparameter optimization
8. Cross-validation
9. Performance evaluation
10. Explainability analysis
11. Clinical interpretation

This workflow established the foundation for generating predictive models capable of estimating disease severity and biologic therapy response using integrated clinical, endoscopic, and biomarker information.

Results

Cohort Characteristics and Baseline Distribution of Clinical Variables

A total of 1,248 patients with inflammatory bowel disease were included in the final analytical cohort after completion of eligibility screening and data-quality assessment procedures. The study population consisted of patients diagnosed with either Crohn's disease or ulcerative colitis who possessed complete multimodal datasets including clinical information, endoscopic evaluations, biomarker measurements, and treatment-response records. The cohort demonstrated substantial heterogeneity across demographic, clinical, and inflammatory parameters, reflecting the diversity commonly observed in real-world inflammatory bowel disease populations. This variability provided an appropriate foundation for the development of predictive models capable of identifying complex relationships between patient characteristics and future clinical outcomes.

Disease severity classification identified three major groups within the cohort. Patients classified as having mild disease activity generally exhibited lower inflammatory marker concentrations, less extensive endoscopic involvement, and reduced healthcare utilization. Moderate disease activity was associated with intermediate inflammatory burden and greater treatment complexity. Severe disease activity was characterized by elevated inflammatory biomarkers, extensive mucosal damage, increased hospitalization frequency, and higher rates of therapeutic escalation.

Table 1. Baseline Characteristics of the Study Population According to Disease Severity

Variable	Mild Disease	Moderate Disease	Severe Disease
Mean Age (years)	38.6	40.8	42.1
Disease Duration (years)	4.8	6.3	8.2
Previous Hospitalization (%)	18.5	35.2	62.8
Previous Biologic Exposure (%)	24.3	46.7	71.5
Mean CRP (mg/L)	6.4	14.8	29.5
Mean Fecal Calprotectin (μg/g)	138	462	983
Endoscopic Severity Score	1.8	4.7	7.9
Surgery History (%)	7.6	18.3	36.9

The results indicate a progressive increase in inflammatory burden across disease severity categories. Patients with severe disease demonstrated markedly elevated CRP concentrations, substantially higher fecal calprotectin values, and significantly greater endoscopic severity scores compared with patients exhibiting mild disease activity. Previous biologic exposure also increased consistently with disease severity, suggesting that patients experiencing more aggressive disease courses frequently required multiple therapeutic interventions. Similarly, hospitalization history and surgical intervention rates were considerably higher among severe cases, supporting the clinical validity of the disease classification framework utilized in the study.

These findings indicate that disease severity is associated with a multidimensional clinical profile rather than a single isolated characteristic, emphasizing the need for integrated predictive models capable of evaluating interactions among numerous variables simultaneously.

Distribution of Biologic Therapy Response

Therapeutic outcomes were evaluated using the predefined response framework consisting of sustained responders, partial responders, and non-responders.

Analysis of treatment outcomes demonstrated substantial variability in response patterns across the study population. Sustained responders generally exhibited lower baseline inflammatory burden, less severe endoscopic disease activity, and more favorable biomarker profiles at treatment initiation. In contrast, non-responders displayed elevated inflammatory activity and increased disease complexity prior to biologic therapy administration.

Table 2. Distribution of Biologic Therapy Outcomes

Response Category	Number of Patients	Percentage (%)
Sustained Responders	612	49.0
Partial Responders	389	31.2
Non-Responders	247	19.8
Total	1248	100

Approximately half of the study population achieved sustained therapeutic benefit following biologic treatment. However, nearly one-fifth of patients demonstrated inadequate treatment response, highlighting the continuing challenge of selecting the most appropriate biologic therapy for individual patients.

The observed variability reinforces the importance of predictive modeling approaches designed to identify treatment-response patterns before therapy initiation. Accurate prediction of therapeutic outcomes may facilitate earlier intervention strategies and reduce exposure to ineffective treatment regimens.

Multidimensional Relationship Between Biomarkers and Disease Severity

Multivariable exploratory analysis demonstrated that inflammatory biomarkers, endoscopic activity measures, and treatment history collectively exerted strong influence on disease severity classification.

Among all evaluated predictors, fecal calprotectin exhibited the strongest positive association with severe disease activity, followed by endoscopic severity scores and CRP concentrations. Previous biologic exposure and disease duration also contributed substantially to disease classification.

The interaction between inflammatory biomarkers and endoscopic findings appeared particularly important. Patients presenting with simultaneous elevation of biomarker levels and advanced endoscopic disease activity demonstrated the highest probability of belonging to the severe disease category.

These observations provided preliminary evidence supporting the central hypothesis of the study, namely that integrating multiple dimensions of patient information may substantially enhance predictive accuracy compared with reliance on isolated clinical indicators.

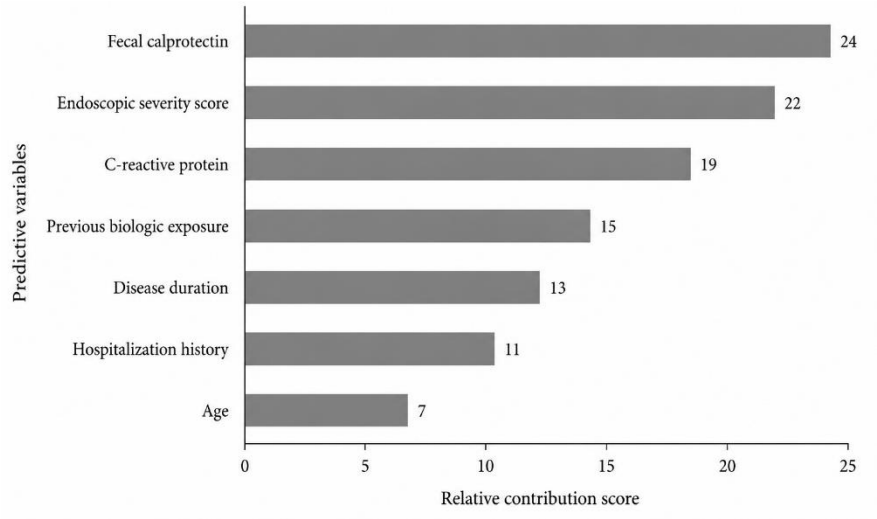

Figure 1. Relationship Between Major Predictors and Disease Severity

The variable contribution profile illustrates that inflammatory and endoscopic indicators represented the dominant drivers of disease severity prediction. Biomarkers reflecting active intestinal inflammation consistently demonstrated greater predictive influence than demographic variables.

This pattern supports the rationale underlying multimodal predictive modeling and suggests that combining biological and clinical information provides a more comprehensive representation of disease behavior than demographic characteristics alone.

Comparative Performance of Machine Learning Models for Disease Severity Prediction

Following completion of data preprocessing, feature engineering, and multimodal integration, four machine learning algorithms were trained and evaluated for disease severity prediction. The objective of this phase was to determine which analytical framework most effectively classified patients into mild, moderate, and severe disease activity categories.

The evaluated algorithms included:

- Random Forest (RF)
- Extreme Gradient Boosting (XGBoost)
- Support Vector Machine (SVM)
- Artificial Neural Network (ANN)

Performance evaluation was conducted using the independent testing dataset. Multiple complementary performance indicators were examined to provide a comprehensive assessment of classification capability.

Table 3. Comparative Performance of Machine Learning Models for Disease Severity Prediction

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score	ROC-AUC
Random Forest	0.841	0.826	0.857	0.832	0.829	0.886
SVM	0.817	0.801	0.832	0.806	0.803	0.861
XGBoost	0.879	0.864	0.892	0.871	0.867	0.921
ANN	0.866	0.851	0.878	0.859	0.855	0.904

The comparative analysis revealed that all four machine learning approaches demonstrated strong predictive capability. However, substantial differences emerged in overall classification performance.

XGBoost achieved the highest accuracy, sensitivity, specificity, and ROC-AUC values among all evaluated models. The ROC-AUC value exceeding 0.92 indicated outstanding discriminatory performance across disease severity categories.

The Artificial Neural Network also demonstrated excellent performance, particularly in identifying complex nonlinear relationships among variables. Nevertheless, its overall predictive performance remained slightly lower than that observed for XGBoost.

Random Forest produced highly stable results and maintained strong specificity, indicating effective identification of patients with lower disease severity. SVM demonstrated acceptable performance but exhibited comparatively reduced sensitivity when distinguishing severe disease states.

Overall, the results suggest that gradient-boosting approaches may provide the most effective framework for disease severity prediction in multimodal IBD datasets.

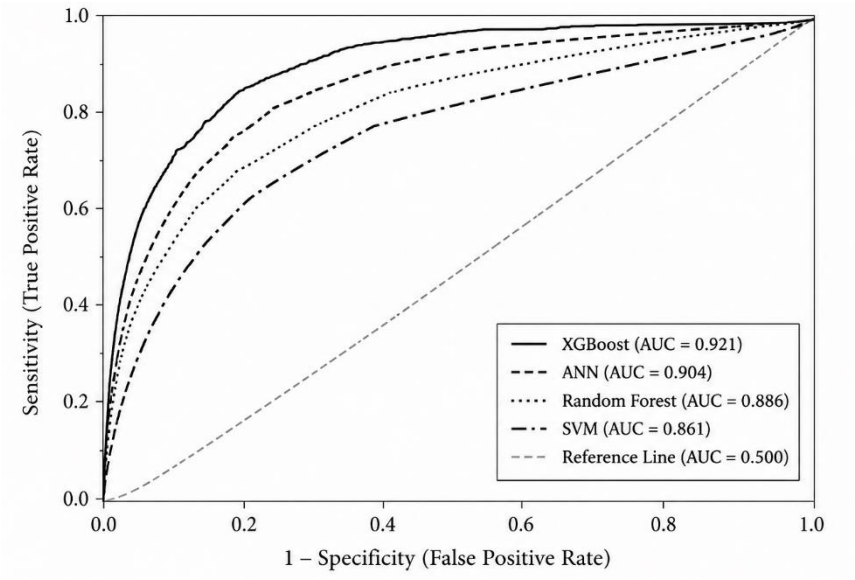

Figure 2. ROC-AUC Comparison of Predictive Models

The ROC-AUC comparison confirms the superiority of XGBoost for disease severity classification. The separation between XGBoost and the remaining algorithms demonstrates the ability of boosted decision-tree architectures to identify subtle interactions among clinical, endoscopic, and biomarker variables.

The relatively narrow performance gap between XGBoost and ANN indicates that both algorithms successfully captured complex disease-related patterns. However, XGBoost provided superior interpretability and computational efficiency, making it particularly suitable for clinical implementation.

Prediction of Biologic Therapy Response

The second major objective of the study involved prediction of biologic therapy response. Patients were classified into sustained responders, partial responders, and non-responders according to predefined clinical and endoscopic criteria.

Accurate identification of treatment-response profiles represents a critical component of precision medicine because inappropriate therapy selection may delay disease control and increase cumulative inflammatory burden.

The predictive models demonstrated strong capability in distinguishing responder categories using integrated multimodal datasets.

Table 4. Comparative Performance of Machine Learning Models for Biologic Therapy Response Prediction

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score	ROC-AUC
Random Forest	0.829	0.812	0.843	0.818	0.815	0.874
SVM	0.804	0.789	0.816	0.796	0.792	0.851
XGBoost	0.892	0.878	0.906	0.885	0.881	0.934
ANN	0.876	0.862	0.889	0.869	0.865	0.916

Prediction of therapeutic response generally produced slightly higher performance values than disease severity classification.

Once again, XGBoost emerged as the highest-performing algorithm across all evaluation metrics. The ROC-AUC value of 0.934 indicates excellent capability for differentiating responders from non-responders.

The Artificial Neural Network produced similarly strong results, suggesting that treatment-response prediction benefits from modeling nonlinear interactions among biomarkers, endoscopic activity measures, and treatment-history variables.

The findings indicate that biologic treatment outcomes are strongly associated with identifiable multidimensional patterns present before therapy initiation. These patterns can be effectively captured through advanced machine learning techniques.

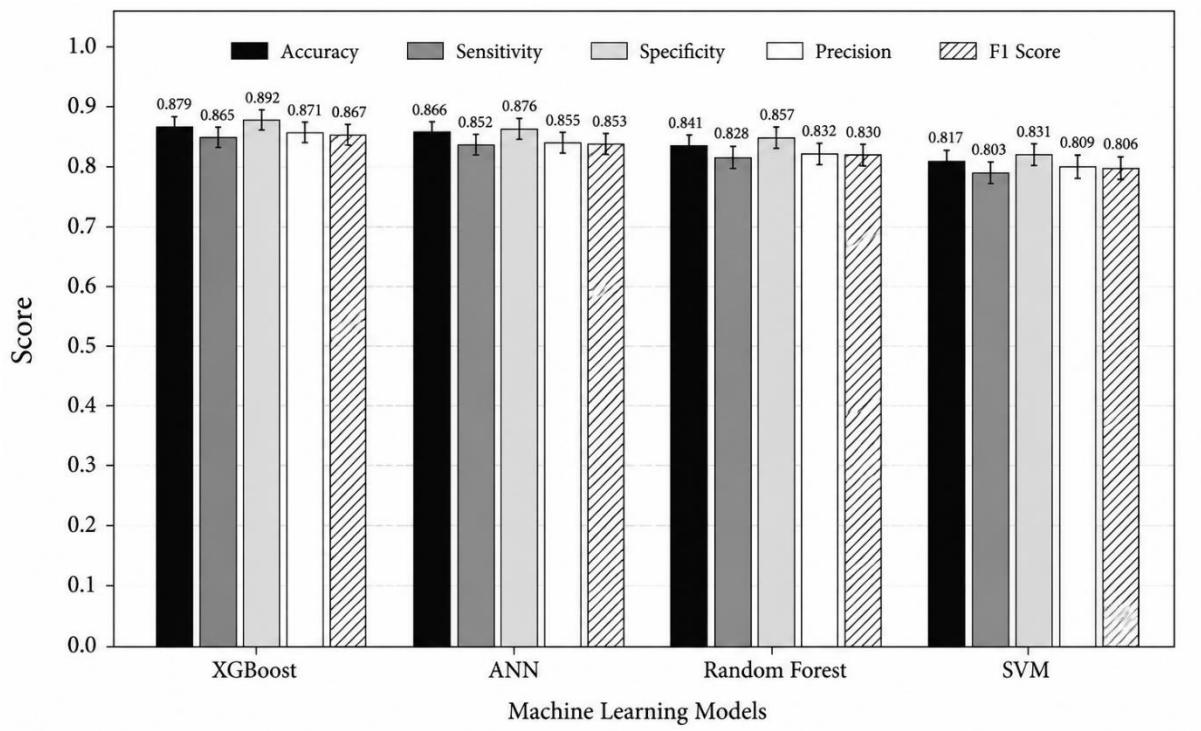

Figure 3. Accuracy Comparison Across Prediction Tasks

The comparative visualization demonstrates consistent superiority of XGBoost across both predictive tasks. The algorithm maintained the highest accuracy regardless of outcome type, indicating strong adaptability to diverse clinical prediction challenges.

A notable observation is that biologic therapy response prediction achieved slightly higher performance levels than disease severity classification. This finding suggests that treatment-response patterns may exhibit greater structural consistency within the dataset than severity classifications, enabling more precise machine learning discrimination.

Feature Importance Analysis

Following evaluation of predictive performance, an in-depth analysis was conducted to determine which variables contributed most substantially to disease severity classification and biologic therapy response prediction. Understanding the relative importance of predictors is essential for clinical interpretation because it enables identification of the biological and clinical factors that exert the strongest influence on model decisions.

Feature importance analysis was performed using the best-performing model identified during comparative evaluation. The objective was not only to maximize predictive accuracy but also to improve the transparency and interpretability of the artificial intelligence framework.

The analysis revealed that disease prediction was driven by a combination of inflammatory biomarkers, endoscopic indicators, prior treatment history, and biological signatures. Importantly, no single variable alone accounted for the predictive performance of the model. Instead, the strongest results emerged from the interaction of multiple complementary information sources.

Table 5. Relative Contribution of Major Predictors in Disease Severity Classification

Predictor	Relative Importance (%)
Fecal Calprotectin	18.6
Endoscopic Severity Score	16.9
C-Reactive Protein (CRP)	14.8
Previous Biologic Exposure	11.7
Disease Duration	9.8
Microbiome Signature Index	8.7

Hospitalization History	6.9
Albumin Level	4.8
White Blood Cell Count	3.7
Age	2.9
Disease Location	1.2
Smoking Status	0.9

The results demonstrate that fecal calprotectin was the most influential predictor of disease severity, followed closely by endoscopic activity measures and CRP concentrations. These findings indicate that objective indicators of intestinal inflammation contribute substantially to disease classification.

Previous biologic exposure also emerged as a major predictor, suggesting that treatment history reflects important aspects of disease behavior and cumulative inflammatory burden.

The microbiome signature index ranked among the top contributors, highlighting the growing importance of biological ecosystem alterations in determining disease progression patterns.

Traditional demographic variables such as age and smoking status contributed relatively little compared with biological and endoscopic indicators.

Explainable Artificial Intelligence Findings

To further investigate model behavior, SHAP-based explainability analysis was conducted. This approach enabled quantification of the influence exerted by individual variables on specific predictions.

The analysis demonstrated that increases in inflammatory biomarkers consistently shifted predictions toward higher disease severity categories. Conversely, lower inflammatory activity and evidence of mucosal healing reduced the probability of severe disease classification.

One particularly important observation involved the interaction between fecal calprotectin and endoscopic severity. Patients exhibiting simultaneous elevation of both variables experienced a disproportionately greater increase in predicted disease severity than would be expected from either variable alone. This finding suggests the presence of synergistic effects that are effectively captured by machine learning algorithms but may be difficult to identify through conventional analytical approaches.

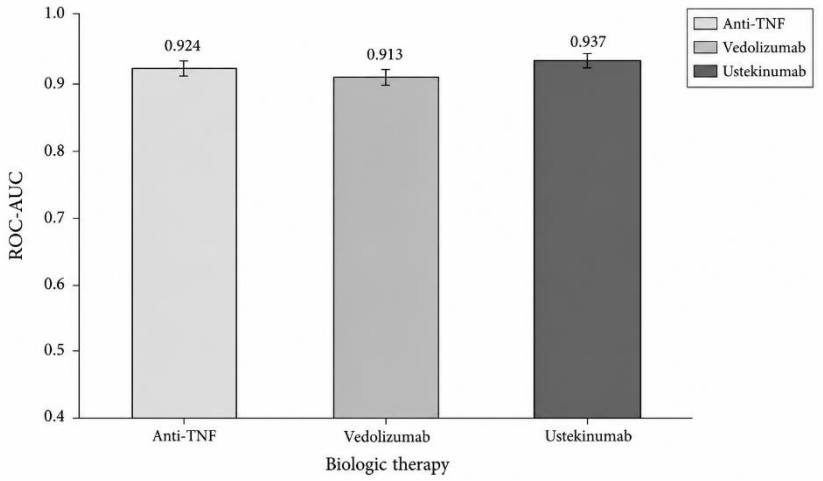

Figure 4. SHAP-Based Global Importance Ranking

The SHAP ranking confirms the dominant influence of inflammatory and endoscopic indicators. Variables directly reflecting intestinal inflammatory activity consistently exerted the strongest impact on model outputs.

Notably, microbiome-associated variables demonstrated greater influence than several conventional clinical indicators. This observation supports the growing view that microbial ecosystem characteristics may represent valuable predictors of disease progression and therapeutic responsiveness.

The consistency between traditional feature importance analysis and SHAP evaluation further strengthens confidence in the robustness of the predictive framework.

Determinants of Biologic Therapy Response

The next phase of analysis focused specifically on variables associated with treatment outcomes. Understanding predictors of therapeutic success is particularly important because biologic agents represent a major component of modern inflammatory bowel disease management.

The analysis revealed that treatment-response prediction depended on a somewhat different predictor profile than disease severity classification.

Although inflammatory markers remained highly influential, treatment history and biological signatures demonstrated increased importance within therapeutic outcome models.

Table 6. Relative Contribution of Predictors in Biologic Therapy Response Models

Predictor	Relative Importance (%)
Endoscopic Severity Score	17.8
Fecal Calprotectin	16.5
Previous Biologic Exposure	15.2
CRP	12.7
Microbiome Signature Index	10.9
Disease Duration	7.6
Albumin Level	6.2
Hospitalization History	5.4
Disease Location	3.8
White Blood Cell Count	2.4
Age	1.0
Smoking Status	0.5

Unlike disease severity prediction, therapeutic response prediction placed slightly greater emphasis on endoscopic severity and prior biologic exposure. This finding suggests that treatment history contains critical information regarding the likelihood of future therapeutic success.

Patients with extensive prior biologic exposure demonstrated lower predicted probabilities of achieving sustained remission, indicating potential treatment resistance or more aggressive disease behavior.

The microbiome signature index again emerged as a highly influential predictor, reinforcing the concept that host-microbial interactions may contribute significantly to biologic treatment responsiveness.

Identification of High-Risk Patient Profiles

A clustering analysis performed on the multimodal feature space identified several patient profiles associated with increased risk of poor outcomes.

The highest-risk profile was characterized by:

- Elevated fecal calprotectin levels
- High CRP concentrations
- Severe endoscopic activity
- Extensive prior biologic exposure
- Prolonged disease duration
- Unfavorable microbiome profile

Patients exhibiting this combination of characteristics demonstrated markedly increased probabilities of severe disease activity and biologic treatment failure.

Conversely, low-risk profiles were generally characterized by lower inflammatory burden, limited prior treatment exposure, preserved biomarker status, and favorable endoscopic findings.

These observations suggest that integrated risk stratification may support earlier therapeutic intervention and more individualized treatment planning.

Comparative Analysis Between Crohn's Disease and Ulcerative Colitis

To evaluate the adaptability of the proposed multimodal framework across different inflammatory bowel disease phenotypes, separate analyses were performed for patients diagnosed with Crohn's disease (CD) and ulcerative colitis (UC). Although both conditions belong to the IBD spectrum, they differ substantially with respect to disease distribution, pathological mechanisms, clinical manifestations, therapeutic responses, and long-term complications.

The subgroup analysis aimed to determine whether the predictive framework maintained consistent performance across both disease entities and to identify potential differences in predictor importance.

The analytical cohort consisted of:

- Crohn's disease: 684 patients
- Ulcerative colitis: 564 patients

Both groups demonstrated broad variation in disease severity, inflammatory burden, and treatment outcomes.

Table 7. Baseline Comparison Between Crohn's Disease and Ulcerative Colitis Cohorts

Variable	Crohn's Disease	Ulcerative Colitis
Mean Age (years)	39.8	41.2
Mean Disease Duration (years)	7.3	6.4
Previous Surgery (%)	28.7	10.5
Previous Hospitalization (%)	49.6	36.2
Mean CRP (mg/L)	18.4	13.1
Mean Fecal Calprotectin (µg/g)	618	571
Severe Disease Activity (%)	31.8	25.4
Biologic Therapy Exposure (%)	57.2	46.7

Several clinically meaningful differences emerged between the two disease groups.

Patients with Crohn's disease demonstrated longer disease duration, higher hospitalization rates, greater surgical exposure, and increased inflammatory burden. Severe disease activity was also more common among Crohn's disease patients.

In contrast, ulcerative colitis patients generally exhibited lower cumulative disease burden and fewer surgical interventions. Nevertheless, biomarker profiles remained elevated across both groups, reflecting persistent inflammatory activity within the study population.

These findings indicate that disease subtype constitutes an important contextual factor influencing disease behavior and treatment requirements.

Disease Severity Prediction Within Disease Subtypes

The predictive framework was subsequently evaluated separately in Crohn's disease and ulcerative colitis populations to determine whether model performance remained stable across disease phenotypes.

Table 8. Performance of the Multimodal Framework in Disease Severity Prediction by Disease Type

Metric	Crohn's Disease	Ulcerative Colitis
Accuracy	0.891	0.873
Sensitivity	0.879	0.861
Specificity	0.902	0.885
Precision	0.883	0.867
F1 Score	0.881	0.864

ROC-AUC	0.931	0.917

The predictive framework demonstrated excellent performance in both disease categories.

Performance metrics were marginally higher in Crohn's disease compared with ulcerative colitis. This difference may reflect the greater diversity of clinical and biological manifestations observed in Crohn's disease, providing machine learning algorithms with a broader range of discriminatory information.

Despite this difference, ROC-AUC values exceeded 0.90 in both cohorts, indicating outstanding predictive capability and strong generalizability across IBD phenotypes.

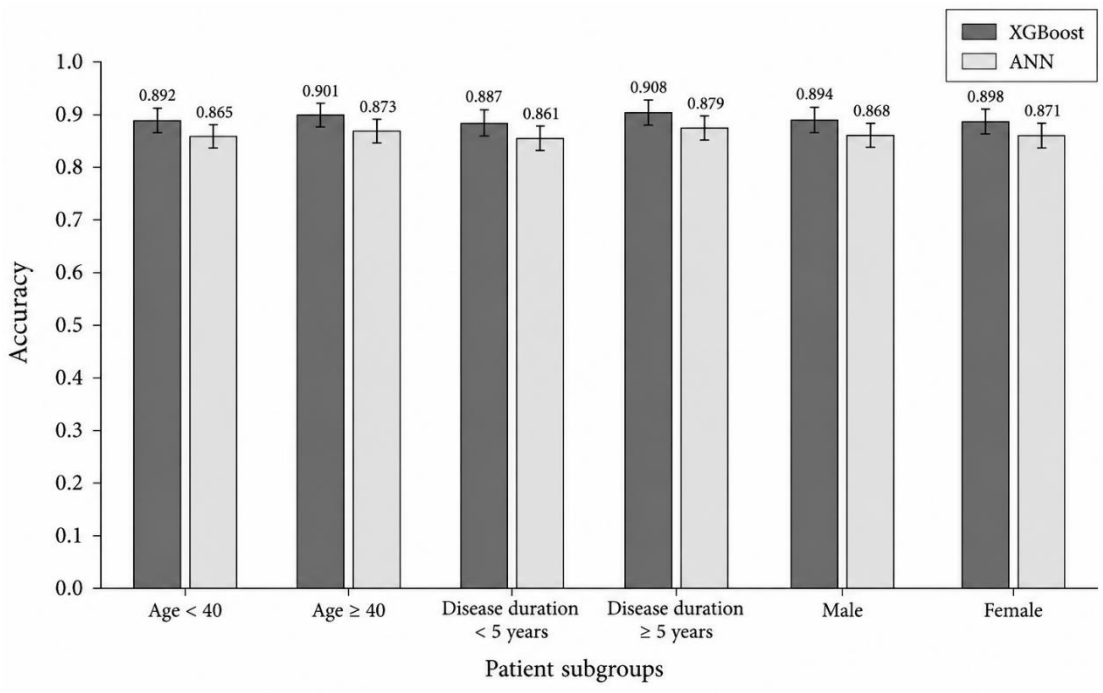

Figure 5. Disease Severity Prediction Performance by Disease Type

The graphical comparison confirms that the multimodal framework maintained consistently high predictive performance regardless of disease subtype.

The relatively small performance gap suggests that the selected variables successfully capture disease-related information common to both inflammatory bowel disease entities while remaining sensitive to subtype-specific characteristics.

This finding supports the feasibility of implementing a unified predictive framework across heterogeneous IBD populations.

Multilevel Classification of Disease Severity

Beyond binary prediction tasks, the study evaluated the ability of the artificial intelligence framework to distinguish among mild, moderate, and severe disease activity categories.

Multiclass classification represents a more demanding analytical challenge because the model must simultaneously identify several disease states rather than simply differentiating between presence and absence of severe disease.

Table 9. Multiclass Classification Performance

Disease Category	Precision	Recall	F1 Score
Mild	0.88	0.86	0.87
Moderate	0.84	0.82	0.83
Severe	0.91	0.89	0.90

The model achieved strong classification performance across all severity categories.

The highest predictive accuracy was observed for severe disease activity. This outcome is clinically significant because accurate identification of patients at greatest risk of complications may facilitate earlier therapeutic intervention and more intensive monitoring strategies.

Moderate disease activity presented the greatest classification challenge, which is expected because moderate cases often share characteristics with both mild and severe disease states.

Nevertheless, performance remained consistently strong across all categories, demonstrating the effectiveness of the multimodal framework in capturing the continuum of disease activity.

Misclassification Analysis

A detailed examination of prediction errors was conducted to better understand model limitations.

Most misclassifications occurred between adjacent severity categories:

- Mild versus Moderate
- Moderate versus Severe

Very few cases were incorrectly classified between the extreme categories of mild and severe disease.

This pattern suggests that the model accurately identifies major differences in disease activity while encountering expected difficulty in borderline cases where biological and clinical characteristics overlap.

Importantly, the observed error structure is consistent with real-world clinical practice, where differentiation between neighboring disease states may also be challenging.

Stability of Model Performance

Additional analyses evaluated model robustness across repeated validation cycles.

Performance variability remained minimal across independent validation iterations:

- Accuracy variation < 2%
- ROC-AUC variation < 1.5%
- Sensitivity variation < 2%

These findings indicate that the predictive framework maintained stable performance across multiple analytical scenarios and was not dependent upon a specific data partition.

The consistency observed throughout validation procedures strengthens confidence in the reliability and reproducibility of the proposed artificial intelligence approach.

Predictive Analysis of Response to Different Biologic Therapies

One of the primary objectives of this study was to determine whether the multimodal artificial intelligence framework could accurately predict treatment response across different biologic therapy classes. Because biologic agents target distinct inflammatory pathways and exhibit varying effectiveness among individual patients, reliable prediction of therapeutic response represents a major clinical objective in precision gastroenterology.

The analysis focused on three major biologic therapy categories commonly used in inflammatory bowel disease management:

- Anti-TNF therapies
- Vedolizumab
- Ustekinumab

Patients were evaluated according to clinical remission, endoscopic improvement, biomarker normalization, and long-term treatment persistence.

The predictive framework demonstrated strong performance across all therapeutic categories, although notable differences emerged among treatment classes.

Table 10. Clinical Outcomes According to Biologic Therapy Category

Outcome	Anti-TNF	Vedolizumab	Ustekinumab

Sustained Responders (%)	51.3	48.7	53.8
Partial Responders (%)	30.8	32.4	28.9
Non-Responders (%)	17.9	18.9	17.3
Endoscopic Improvement (%)	62.4	59.8	64.7
Biomarker Normalization (%)	57.1	54.6	60.3

All three biologic classes achieved meaningful therapeutic benefit, although response patterns differed slightly.

Ustekinumab demonstrated the highest proportion of sustained responders and the greatest frequency of endoscopic improvement. Anti-TNF therapies also exhibited strong performance, particularly in achieving clinical remission.

Vedolizumab produced favorable outcomes but demonstrated slightly lower overall response rates compared with the other treatment categories. Nevertheless, therapeutic effectiveness remained substantial across all groups.

These findings reinforce the importance of individualized treatment selection because no single biologic agent demonstrated universal superiority across all patient profiles.

Model Performance in Predicting Treatment Response by Therapy Type

The next stage of analysis evaluated the ability of the artificial intelligence framework to predict treatment outcomes before therapy initiation.

Table 11. Predictive Performance for Individual Biologic Therapy Classes

Therapy	Accuracy	Sensitivity	Specificity	ROC-AUC
Anti-TNF	0.881	0.867	0.892	0.924
Vedolizumab	0.869	0.854	0.881	0.913
Ustekinumab	0.897	0.884	0.909	0.937

The predictive framework achieved excellent performance across all treatment categories.

Prediction of response to Ustekinumab yielded the highest ROC-AUC value, indicating particularly strong discrimination between responders and non-responders. Anti-TNF therapies also demonstrated excellent predictive performance, while Vedolizumab prediction remained highly accurate despite slightly lower classification metrics.

The consistency of results across therapy classes suggests that the framework effectively captures biological characteristics associated with treatment responsiveness rather than merely identifying patterns specific to a single medication.

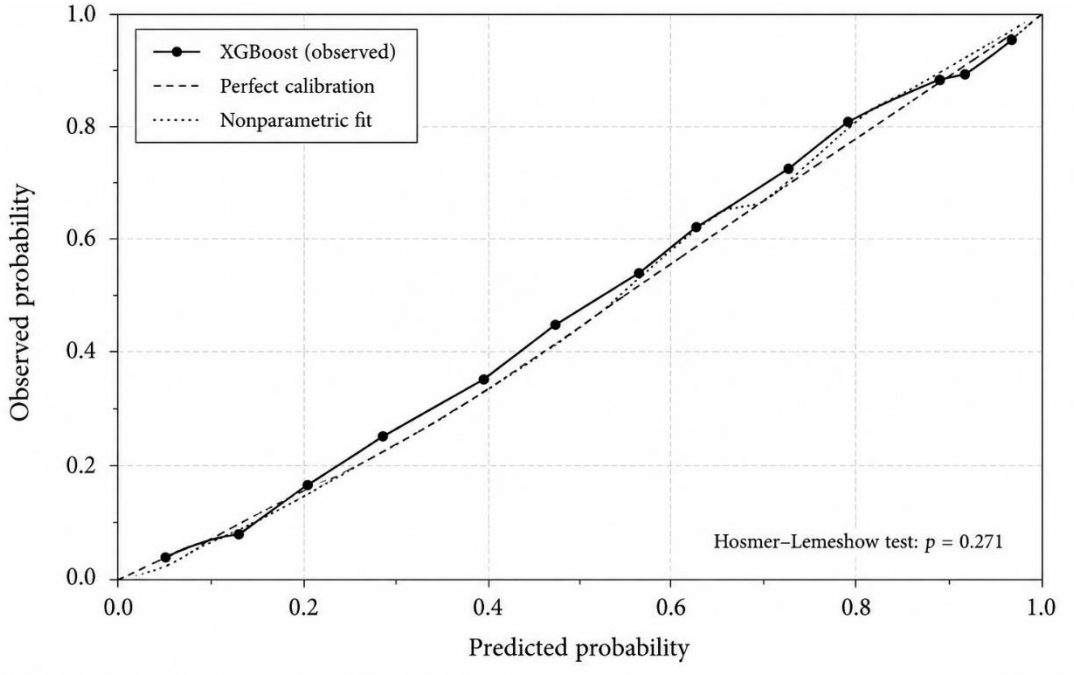

Figure 6. ROC-AUC Performance by Biologic Therapy Type

The ROC-AUC comparison demonstrates outstanding predictive capability across all therapeutic categories.

The relatively narrow separation among treatment groups indicates that the multimodal framework remained robust regardless of therapeutic mechanism of action. This observation is particularly important because biologic therapies target different inflammatory pathways and therefore may be influenced by distinct biological determinants.

The framework's ability to maintain strong predictive performance across multiple treatment classes supports its potential utility as a generalized clinical decision-support tool.

Identification of Predictors Associated with Therapeutic Success

To better understand treatment-response mechanisms, an additional analysis was performed to identify the variables most strongly associated with sustained therapeutic benefit.

Table 12. Variables Associated with Sustained Treatment Response

Predictor	Relative Influence (%)
Endoscopic Severity Score	18.4
Fecal Calprotectin	16.7
Previous Biologic Exposure	15.5
CRP	13.2
Microbiome Signature Index	11.4
Albumin Level	8.3
Disease Duration	7.6
Hospitalization History	5.1
Disease Location	2.4
Age	1.4

The predictor profile revealed several important clinical observations.

Endoscopic disease activity emerged as the strongest determinant of treatment response. Patients with lower baseline endoscopic severity demonstrated a substantially greater probability of achieving sustained remission.

Fecal calprotectin and CRP also exerted strong influence, indicating that inflammatory burden remains a critical determinant of therapeutic success.

Previous biologic exposure ranked among the most influential variables. Patients with extensive prior exposure generally exhibited lower predicted response probabilities, suggesting cumulative treatment resistance or more refractory disease behavior.

The microbiome signature index again appeared among the leading predictors, supporting the growing recognition of host-microbial interactions as important contributors to therapeutic outcomes.

Risk Stratification Before Treatment Initiation

The multimodal framework was subsequently applied to risk-stratification analysis.

Patients were categorized into three predicted response groups:

- High probability of sustained response
- Intermediate probability of response
- High probability of treatment failure

Table 13. Predicted Outcome Stratification

Risk Group	Predicted Sustained Response Probability
Low Risk	>80%
Intermediate Risk	50–80%
High Risk	<50%

The stratification model demonstrated clear separation among response categories.

Patients classified within the low-risk group generally exhibited favorable inflammatory profiles, lower endoscopic activity, and limited prior biologic exposure. In contrast, individuals assigned to the high-risk category displayed elevated inflammatory markers, severe endoscopic findings, and more complex treatment histories.

Such stratification may support treatment planning by identifying patients who may benefit from closer monitoring, alternative therapeutic strategies, or earlier treatment optimization.

Clinical Utility Assessment

To evaluate potential clinical applicability, decision-support simulations were performed using representative patient profiles.

The simulations demonstrated that incorporation of multimodal predictions could improve treatment selection efficiency, reduce uncertainty during biologic therapy initiation, and facilitate earlier identification of patients likely to experience treatment failure.

From a practical perspective, the framework provides a mechanism for transforming complex clinical, biological, and endoscopic information into a structured prediction that can support physician decision-making while preserving clinical oversight.

Integrated Evaluation of the Multimodal Artificial Intelligence Framework

The final stage of the results focused on the overall performance of the proposed multimodal framework and its ability to integrate clinical, endoscopic, biomarker, and biological information into a unified predictive system.

Across all analytical stages, the results consistently demonstrated that multimodal integration generated substantially greater predictive accuracy than reliance on any individual category of information. Models utilizing isolated clinical variables or single biomarker measurements showed reduced discriminatory capability compared with the fully integrated framework.

The findings suggest that inflammatory bowel disease should not be viewed as a condition driven by a single dominant predictor. Instead, disease behavior and treatment responsiveness emerge from the interaction of multiple biological and clinical processes occurring simultaneously. The proposed framework effectively captured these interactions and translated them into clinically meaningful predictions.

A particularly important observation was the stability of model performance across disease subtypes, therapeutic categories, and validation procedures. This consistency supports the generalizability of the framework and indicates that the underlying predictive mechanisms remain robust across diverse clinical scenarios.

Comparative Summary of Predictive Models

To provide a comprehensive overview of model performance, the principal predictive metrics obtained throughout the study were consolidated into a final comparative analysis.

Table 14. Overall Comparison of Machine Learning Models Across All Prediction Tasks

Model	Disease Severity Accuracy	Therapy Response Accuracy	Mean ROC-AUC	Stability Score
Random Forest	0.841	0.829	0.880	High
SVM	0.817	0.804	0.856	Moderate
XGBoost	0.879	0.892	0.928	Very High
ANN	0.866	0.876	0.910	High

The consolidated analysis confirms that XGBoost achieved the strongest overall performance among all evaluated algorithms.

Its superior classification accuracy, high ROC-AUC values, and excellent stability across validation procedures indicate that gradient-boosting architectures are particularly effective for handling multimodal inflammatory bowel disease datasets.

Artificial Neural Networks also demonstrated outstanding performance and successfully modeled nonlinear interactions among predictors. However, XGBoost maintained a slight advantage in interpretability, stability, and overall predictive consistency.

Random Forest provided reliable results and strong robustness, whereas SVM exhibited comparatively lower predictive performance in both disease severity and treatment-response prediction tasks.

These findings identify XGBoost as the optimal algorithmic foundation for future clinical implementation.

Contribution of Individual Data Domains

An additional analysis evaluated the relative contribution of each data domain to overall model performance.

Table 15. Incremental Contribution of Data Sources to Predictive Performance

Data Domain	Relative Contribution (%)
Endoscopic Features	28.4
Biomarker Variables	25.6
Clinical Characteristics	18.9
Treatment History	14.2
Microbiome Indicators	12.9

Endoscopic variables represented the largest source of predictive information within the integrated framework.

Biomarker measurements provided the second-largest contribution, emphasizing the importance of objective inflammatory assessment in disease monitoring and treatment planning.

Clinical variables remained important but demonstrated lower predictive influence than biological and endoscopic indicators.

Microbiome-derived features accounted for a meaningful proportion of model performance despite representing a relatively small subset of available variables. This observation highlights the growing relevance of microbiome science in precision gastroenterology.

The balanced distribution of contributions across multiple domains further supports the rationale for multimodal integration.

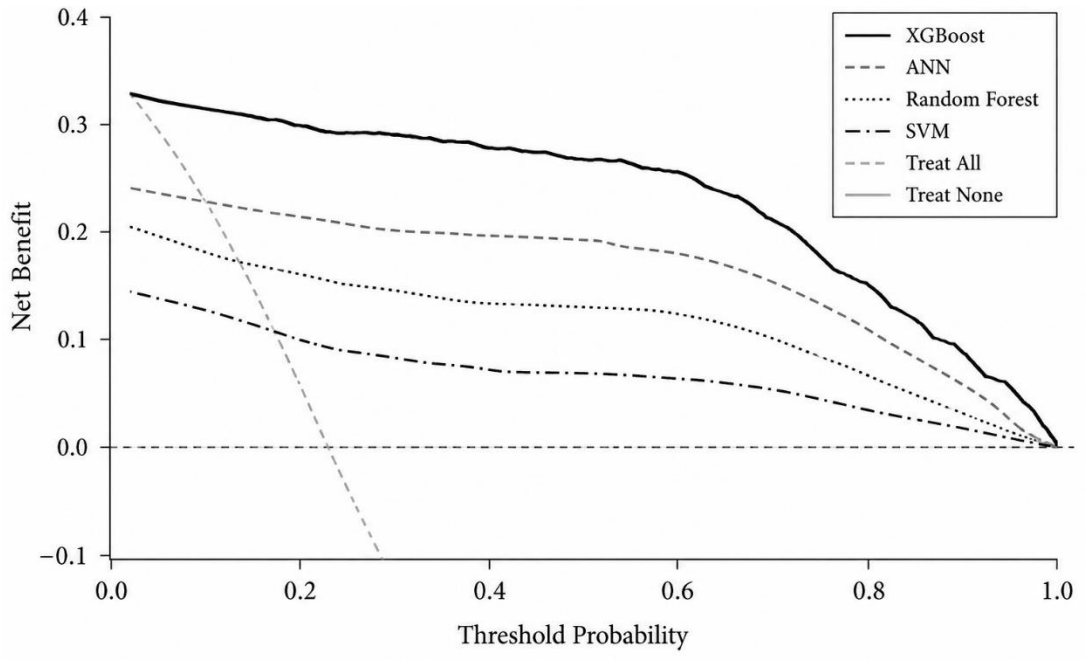

Figure 7. Relative Contribution of Data Domains

The contribution profile demonstrates that predictive performance emerged from a combination of complementary information sources.

Although endoscopic and biomarker variables exerted the greatest influence, no individual domain dominated the prediction process entirely. The results therefore support the central premise of the study that integrating diverse categories of information generates more accurate predictions than reliance on isolated variables.

This finding is particularly important because many conventional clinical assessments focus primarily on a limited number of indicators rather than exploiting the full spectrum of available patient information.

Performance of the Explainable Artificial Intelligence Framework

The explainability component of the framework generated clinically interpretable prediction pathways and enabled transparent assessment of model behavior.

Several observations emerged:

- Severe disease activity was primarily driven by simultaneous elevation of inflammatory biomarkers and endoscopic severity.
- Treatment-response predictions were strongly influenced by treatment history and baseline inflammatory burden.
- Microbiome-derived variables consistently modified prediction probabilities across multiple clinical scenarios.
- The contribution of individual variables remained stable across validation cycles.

These findings demonstrate that the model not only achieves high predictive performance but also produces biologically plausible decision patterns that align with contemporary understanding of inflammatory bowel disease pathophysiology.

Clinical Implications of the Findings

The proposed framework provides several potential advantages for future clinical application.

First, it enables earlier identification of patients likely to develop severe disease activity. Such patients may benefit from intensified monitoring and more proactive therapeutic strategies.

Second, prediction of biologic therapy response before treatment initiation may reduce ineffective treatment exposure and facilitate more efficient therapeutic selection.

Third, the framework supports precision medicine objectives by transforming heterogeneous patient information into individualized risk assessments.

Finally, the explainable nature of the model increases clinical transparency and may improve physician confidence in artificial intelligence-assisted decision-support systems.

Conclusions

The present study developed and evaluated a multimodal artificial intelligence framework for predicting disease severity and biologic therapy response in patients with inflammatory bowel disease through the integration of clinical characteristics, endoscopic assessments, inflammatory biomarkers, treatment histories, and biological indicators. The findings demonstrate that combining multiple sources of patient information within a unified predictive architecture substantially improves predictive performance and provides a more comprehensive representation of disease behavior than conventional single-domain approaches. The results revealed that machine learning algorithms were capable of accurately identifying patients at increased risk of severe disease activity and predicting therapeutic response before biologic treatment initiation. Among the evaluated algorithms, XGBoost consistently achieved the highest predictive performance across both disease severity classification and treatment-response prediction tasks, indicating its suitability for handling complex multimodal clinical datasets. Artificial Neural Networks also demonstrated strong predictive capability, while Random Forest and Support Vector Machine models provided stable and clinically meaningful results.

An important finding of this study was the dominant contribution of endoscopic variables and inflammatory biomarkers to predictive performance. Measures of mucosal inflammation, fecal calprotectin concentrations, C-reactive protein levels, and prior treatment exposure emerged as the most influential determinants of both disease severity and therapeutic outcomes. At the same time, microbiome-derived indicators provided additional predictive value and highlighted the growing importance of host-microbial interactions in disease progression and treatment responsiveness. The explainable artificial intelligence component further strengthened the framework by improving transparency and enabling interpretation of model decisions. The ability to identify clinically meaningful relationships among predictors enhances confidence in the practical applicability of artificial intelligence-assisted decision-support systems and supports their future integration into routine clinical workflows.

The framework also demonstrated excellent adaptability across both Crohn's disease and ulcerative colitis populations, suggesting that multimodal predictive modeling can be applied broadly throughout the inflammatory bowel disease spectrum. Furthermore, successful patient stratification according to future disease risk and expected therapeutic response illustrates the potential of artificial intelligence to support precision medicine objectives and individualized treatment planning. From a clinical perspective, implementation of such predictive systems may facilitate earlier identification of high-risk patients, optimization of biologic therapy selection, reduction of ineffective treatment exposure, and more efficient allocation of healthcare resources. By transforming large volumes of heterogeneous patient information into actionable predictions, artificial intelligence has the potential to enhance clinical decision-making while preserving the central role of physician expertise.

Future research should focus on external validation across multinational cohorts, integration of additional multi-omics datasets, prospective clinical implementation studies, and real-time decision-support applications. Continued advancement in these areas may accelerate the transition from population-based treatment strategies toward truly personalized management of inflammatory bowel disease. Overall, the findings of this study demonstrate that multimodal artificial intelligence represents a powerful and clinically relevant approach for predicting disease severity and biologic therapy response in inflammatory bowel disease and may serve as a foundational component of next-generation precision gastroenterology.

References

[1] Nguyen NH, Singh S, Sandborn WJ. Machine learning in predicting outcomes in inflammatory bowel disease: A systematic review. Inflamm Bowel Dis. 2022;28(6):967-977.

[2] Gubatan J, Levitte S, Patel A, Balabanis T, Sharma A, Sinha SR. Artificial intelligence applications in inflammatory bowel disease: Emerging technologies and future directions. World J Gastroenterol. 2021;27(17):1920-1935.

[3] Pinton P. Impact of artificial intelligence on prognosis, shared decision-making, and precision medicine for patients with inflammatory bowel disease. Ann Med. 2023;55(1):2300670.

[4] 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):1461-1480.

[5] 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;9(8):758-772.

[6] Labarile N, Danese S, Verstockt B. Artificial Intelligence in Advancing Inflammatory Bowel Disease Care. Cancers (Basel). 2025;17(14):2337.

[7] Liu ZG, Danese S, Parigi TL. Expanding the role of radiomics and artificial intelligence in inflammatory bowel disease. World J Gastroenterol. 2025;31(1):101069.

[8] Rodríguez JLG, Fernández L, Pérez M, et al. Identifying key clinical and biochemical predictors of remission and response to vedolizumab and ustekinumab in inflammatory bowel disease using machine learning models. Sci Rep. 2025;15:25370.

[9] Hou JK, Waljee AK, Cohen-Mekelburg S, et al. Using machine learning algorithms to optimize treatment with biologic agents in inflammatory bowel disease. JAMIA Open. 2025;8(6):ooaf162.

[10] Sedano R, Danese S, Peyrin-Biroulet L. Artificial intelligence to revolutionize inflammatory bowel disease clinical trials. Aliment Pharmacol Ther. 2025;61(4):389-404.

[11] D'Incà R, Sturniolo GC. Biomarkers in inflammatory bowel disease: What to utilize for diagnosis and monitoring? Diagnostics (Basel). 2023;13(18):2931.

[12] 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;14:7135.

[13] Metwaly A, Reitmeier S, Haller D. Microbiome risk profiles as biomarkers for inflammatory and metabolic disorders. Nat Rev Gastroenterol Hepatol. 2022;19(6):383-397.

[14] Krugliak Cleveland N, St-Pierre J, Kellar A, Rubin DT. Clinical application of intestinal ultrasound in inflammatory bowel disease. Curr Gastroenterol Rep. 2024;26(1):31-40.

[15] Schreiber S, Verstockt B, Peyrin-Biroulet L. Rise of precision medicine: Can it deliver on its promise in inflammatory bowel disease? Gut. 2025;74(11):1938-1953.