



Artificial Intelligence-Driven Integration of Gastrointestinal Microbiome Multi-Omics for Autoimmune Disease Prediction and Precision Medicine: A Scoping Review

Narin Manteghi¹, Karin Ahmadi¹, Cobra Moradian^{1,*}

¹ Department of Chemical and Biological Technologies, CT.C, Islamic Azad University, Tehran, Iran

*Corresponding Author: Department of Chemical and Biological Technologies, CT.C, Islamic Azad University, Tehran, Iran. Email: c.moradian@iaui.ir

Received: 29 June, 2026; Revised: 11 August, 2026; Accepted: 5 September, 2026

Abstract

Context: The gut microbiome is increasingly recognized as a key regulator of immune homeostasis and host metabolism. Accumulating evidence indicates that microbial dysbiosis contributes to the initiation and progression of autoimmune diseases by altering immune signaling, intestinal barrier integrity, and metabolic pathways. Advances in multi-omics technologies, including metagenomics, metatranscriptomics, metaproteomics, and metabolomics, have enabled comprehensive characterization of host-microbiome interactions. In parallel, artificial intelligence and machine learning approaches have emerged as powerful tools for analyzing complex biological datasets and supporting precision medicine applications. This scoping review aimed to evaluate the current evidence on artificial intelligence-driven integration of gut microbiome multi-omics data for autoimmune disease prediction, biomarker discovery, therapeutic response assessment, and precision medicine implementation.

Evidence Acquisition: A comprehensive literature search was conducted in PubMed, Scopus, Web of Science, and Google Scholar to identify studies published between January 2020 and March 2026. The final search was performed on March 15, 2026. Review articles, clinical investigations, experimental studies, and artificial intelligence-based analyses focusing on microbiome multi-omics integration were included. Extracted data were evaluated with respect to omics technologies, artificial intelligence methodologies, predictive performance, disease applications, biomarker identification, and translational relevance. A total of 38 studies met the inclusion criteria and were included in the qualitative synthesis.

Results: The reviewed evidence demonstrated that integrating microbiome-derived multi-omics data with artificial intelligence algorithms improved disease classification, biomarker discovery, and prediction of therapeutic outcomes compared with conventional single-omics approaches. Machine learning models, including Random Forest, Support Vector Machine, and XGBoost, successfully identified microbial and metabolic signatures associated with rheumatoid arthritis, systemic lupus erythematosus, inflammatory bowel disease, and multiple sclerosis. Deep learning architectures, including convolutional neural networks, autoencoders, and graph neural networks, enabled the identification of complex host-microbiome interactions and disease-associated molecular pathways. Multimodal artificial intelligence frameworks integrating microbiome, metabolomic, genomic, and clinical datasets achieved improved predictive performance and facilitated patient stratification for personalized therapeutic interventions. Emerging technologies, such as explainable artificial intelligence, federated learning, generative artificial intelligence, and digital twin systems, demonstrated considerable potential to improve model interpretability, treatment optimization, and individualized healthcare delivery.

Discussion: Artificial intelligence-assisted integration of gut microbiome multi-omics data is a promising strategy for improving the prediction of autoimmune disease and advancing precision medicine. By combining microbial, molecular, immunological, and clinical information, these approaches facilitate biomarker discovery, patient stratification, and personalized therapeutic decision-making. Despite challenges related to data heterogeneity, standardization, interpretability, and clinical validation, ongoing advances in explainable artificial intelligence, multimodal learning, federated learning, and generative artificial intelligence are expected to accelerate the translation of microbiome-based precision medicine into routine clinical practice.

Keywords: Artificial Intelligence, Gastrointestinal Microbiome, Multi-Omics, Autoimmune Disease, Precision Medicine, Machine Learning, Biomarkers

1. Introduction

Autoimmune diseases comprise a heterogeneous group of chronic disorders characterized by inappropriate immune responses against self-antigens, leading to persistent inflammation, tissue damage, and progressive

loss of organ function. Collectively, these diseases affect millions of individuals worldwide and represent a major public health burden due to their chronic nature, complex pathogenesis, and substantial healthcare costs. Although genetic susceptibility contributes to disease risk, growing evidence suggests that environmental and microbial

factors also play critical roles in the initiation and progression of autoimmune conditions (1,2).

Among the environmental factors implicated in immune-mediated diseases, the gut microbiome has emerged as one of the most influential regulators of host immunity. The human gastrointestinal tract harbors a diverse microbial ecosystem composed of bacteria, viruses, fungi, and archaea that collectively contribute to metabolic homeostasis, nutrient processing, epithelial barrier maintenance, and immune system development. Under physiological conditions, these microorganisms interact with the host in a mutually beneficial manner and participate in the regulation of both innate and adaptive immune responses (1). However, disruption of microbial composition and function, commonly referred to as dysbiosis, has been associated with numerous autoimmune and inflammatory disorders, including rheumatoid arthritis, systemic lupus erythematosus, inflammatory bowel disease, and multiple sclerosis (1,3).

Several mechanisms have been proposed to explain the relationship between dysbiosis and autoimmune disease development. Alterations in microbial diversity may compromise intestinal barrier integrity, promote the translocation of microbial products into the systemic circulation, and activate inflammatory signaling pathways. In addition, microbial metabolites such as short-chain fatty acids, including butyrate, acetate, and propionate, contribute to immune tolerance by regulating regulatory T-cell differentiation and suppressing pro-inflammatory cytokine production. Consequently, disturbances in microbial metabolic activity may contribute to the chronic inflammation and immune dysregulation observed in autoimmune disorders (1,4).

Recent advances in high-throughput sequencing and systems biology have substantially expanded our understanding of host-microbiome interactions. Technologies such as metagenomics, metatranscriptomics, metaproteomics, and metabolomics enable comprehensive characterization of microbial communities and their functional activities across multiple molecular levels. While metagenomics provides information regarding microbial genetic composition, metatranscriptomics identifies actively expressed genes, metaproteomics evaluates protein function, and metabolomics characterizes biologically active metabolites involved in host-microbiome communication (2, 5, 6). Integrating these complementary datasets offers a more complete understanding of disease-associated biological processes than any single omics layer alone.

Despite the considerable potential of multi-omics approaches, interpretation of these datasets remains challenging. Microbiome-derived multi-omics data are

often characterized by high dimensionality, heterogeneity, sparsity, and complex nonlinear interactions that cannot be adequately addressed using conventional statistical methods. Furthermore, the simultaneous analysis of microbial, metabolic, genomic, and clinical variables requires advanced computational approaches capable of extracting biologically meaningful information from large-scale datasets (6).

Artificial intelligence and machine learning techniques have therefore become increasingly important in microbiome research. Supervised learning algorithms such as Random Forest, Support Vector Machine, and Extreme Gradient Boosting have demonstrated promising performance in disease classification, biomarker discovery, and prediction of therapeutic responses (7, 8, 9).

More recently, deep learning approaches, including convolutional neural networks, autoencoders, and graph neural networks, have enabled the identification of complex relationships across multiple biological layers and facilitated the discovery of disease-associated molecular patterns that may not be detectable using traditional analytical approaches (9-12).

Beyond disease prediction, artificial intelligence-driven integration of microbiome multi-omics data has important implications for precision medicine. Increasing evidence indicates that variations in gut microbial composition can influence drug metabolism, therapeutic efficacy, and treatment-related toxicity. Consequently, integrating microbiome-derived biomarkers with genomic, metabolomic, and clinical information may support individualized treatment strategies and improve therapeutic outcomes in patients with immune-mediated diseases (5, 13, 14). Emerging technologies such as explainable artificial intelligence, federated learning, generative artificial intelligence, and digital twin systems have further expanded the potential clinical applications of microbiome-based precision medicine by improving model transparency, scalability, and predictive performance (10, 15-18).

Although substantial progress has been achieved in microbiome research, the rapidly expanding volume of multi-omics data has created significant challenges in data integration, biological interpretation, and clinical translation. Many existing studies focus on individual omics layers or specific analytical approaches, whereas a comprehensive understanding of autoimmune disease mechanisms requires the integration of microbial, metabolic, genomic, transcriptomic, and clinical information within a unified analytical framework. Recent developments in artificial intelligence have provided new opportunities to address these challenges by enabling the identification of complex biological patterns, improving

biomarker discovery, and enhancing predictive modeling. However, the current evidence regarding artificial intelligence-assisted integration of gut microbiome multi-omics data remains fragmented across disciplines and disease settings.

Review question: "What is the current evidence regarding artificial intelligence-driven integration of gut microbiome multi-omics data for autoimmune disease prediction, biomarker discovery, therapeutic response assessment, and precision medicine?"

Therefore, this scoping review aims to synthesize current knowledge on the application of artificial intelligence to gut microbiome multi-omics integration, with a particular focus on autoimmune disease prediction, biomarker discovery, therapeutic response assessment, and precision medicine. In addition, we critically discuss current methodological challenges, emerging computational approaches, and future directions for the clinical implementation of microbiome-based precision medicine.

2. Methods

2.1. Study Design

This study was conducted as a scoping review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) guidelines (19). A scoping review design was chosen because of the broad review question, the heterogeneity of the available evidence, and the need for a narrative synthesis rather than a quantitative meta-analysis.

Protocol registration: An a priori protocol was not prepared or prospectively registered for this scoping review.

2.2. Search Strategy

A comprehensive literature search was conducted in PubMed, Scopus, Web of Science, and Google Scholar to identify studies published between January 2020 and March 2026. The final search was conducted on March 15, 2026. Restricting the search to publications from 2020 onward was justified by the rapid evolution of artificial intelligence methodologies and multi-omics technologies during this period, ensuring that the review reflects the most current evidence.

The search strategy combined Medical Subject Headings (MeSH) terms and free-text keywords. The following terms were used in various combinations: "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Networks"; "Gut Microbiome" OR

"Gut Microbiota" OR "Intestinal Microbiome"; "Multi-Omics" OR "Metagenomics" OR "Metabolomics" OR "Metatranscriptomics" OR "Metaproteomics" OR "Integrative Omics"; "Autoimmune Diseases" OR "Rheumatoid Arthritis" OR "Systemic Lupus Erythematosus" OR "Inflammatory Bowel Disease" OR "Multiple Sclerosis"; "Precision Medicine" OR "Personalized Medicine" OR "Pharmacomicrobiomics"; "Biomarker Discovery" OR "Therapeutic Response" OR "Patient Stratification"

Boolean operators (AND/OR) were applied to optimize search sensitivity. The complete search strategies for each database, including the exact Boolean strings, search fields, and limits, are provided in Table S1 in the Supplementary Materials.

2.3. Eligibility Criteria

Studies were eligible for inclusion if they met the following criteria:

- Investigated gut microbiome-related multi-omics data.
- Applied artificial intelligence or machine learning approaches for data integration or predictive modeling.
- Examined immunological, pharmacological, or autoimmune disease outcomes.
- Were published in English between 2020 and 2026.
- Were original research articles, including clinical investigations, experimental studies, and methodological studies.

Studies were excluded if they:

- Did not involve artificial intelligence-based analytical methods.
- Focused solely on single-omics datasets without integration approaches.
- Were conference abstracts, editorials, letters, or non-peer-reviewed publications.
- Lacked relevance to autoimmune diseases, immunology, or precision medicine.
- Were review articles, systematic reviews, or narrative reviews (these were used for background context only).

For the purposes of this review, "multi-omics integration" was operationally defined as the combined analysis of at least two omics layers (e.g., metagenomics and metabolomics) or the integration of omics data with clinical or demographic information. "Artificial intelligence" was defined as computational approaches capable of learning patterns from data without explicit programming, including machine learning and deep learning methods.

2.4. Study Selection Process

Table 1. Characteristics of Included Studies (N = 38)

First Authors (y) and Ref.	Countries	Study Design	Disease Type	Sample Size	Omics Datasets	AI/ML Method	Validation	Key Findings
Zhang et al. (2025) (4)	China	Case-control	IBD	156	Metagenomics + Metabolomics	Random Forest, XGBoost	Internal (CV)	AUC 0.92 for disease classification
Palmieri et al. (2025) (20)	Italy	Cohort	IBD	210	Microbiome + Metabolome	XGBoost, SVM	Internal	AUC 0.88 for treatment response
Chen et al. (2025) (13)	Taiwan	Cross-sectional	RA	98	Metagenomics + Genomics	Random Forest	Internal	AUC 0.85 for disease prediction
Huang et al. (2017) (6)	USA	Methodological	Multiple	Various	Multi-omics	Various	Review	Integration framework
Wang et al. (2023) (14)	USA	Review	Multiple	Various	Multi-omics	AI/ML	Review	Multi-omics integration
Banerjee et al. (2023) (9)	India	Methodological	Multiple	Various	Microbiome	Deep Learning	Review	Deep learning applications
Topçuoğlu et al. (2020) (7)	USA	Methodological	Multiple	Various	Microbiome	Machine Learning	Review	ML framework for microbiome
Galkin et al. (2020) (21)	Russia	Cohort	Aging	1,165	Metagenomics	Random Forest	Internal	Aging clock prediction
Argelaguet et al. (2020) (22)	UK	Methodological	Multiple	Various	Multi-omics	MOFA+	Review	Multi-omics integration
Zeng et al. (2024) (23)	China	Meta-analysis	Autoimmune	80 RCTs	Microbiome	Various	Meta-analysis	Efficacy of microbiome therapies
Islam et al. (2022) (3)	USA	Meta-analysis	Autoimmune/Cancer	Various	Metagenomics	Meta-analysis	Meta-analysis	Gut microbiome signatures
Ota and Fujio (2021) (2)	Japan	Review	Autoimmune	Various	Multi-omics	Various	Review	Precision medicine
Shahid (2025) (5)	USA	Review	Autoimmune	Various	Multi-omics	Various	Review	Microbiome precision medicine
Miyaochi et al. (2023) (1)	Japan	Review	Autoimmune	Various	Microbiome	Various	Review	Gut microbiome in autoimmunity
Zhao et al. (2025) (12)	China	Methodological	Autoimmune	Various	Microbiome	GNN	Review	Graph neural networks
Zhou et al. (2025) (24)	China	Methodological	Various	Various	Microbiome	Explainable DL	Review	Explainable AI
Huybrechts et al. (2024) (15)	Belgium	Methodological	Biomedical	Various	Multi-omics	XAI	Review	Explainable AI
Li et al. (2025) (16)	China	Methodological	Various	Various	Multi-omics	Generative AI	Review	Generative AI integration
Zeng et al. (2024) (23)	China	Methodological	Various	Various	Multi-omics	AI	Review	AI-driven multi-omics
Singh et al. (2026) (25)	India	Review	Various	Various	Multi-omics	AI	Review	Gut microbiome complexity
Noecker et al. (2022) (26)	USA	Review	Various	Various	Microbiome + Metabolome	Various	Review	Microbiome-metabolome integration
Wilmanski et al. (2021) (27)	USA	Cohort	Aging	9,000	Metagenomics	ML	Internal	Gut microbiome and aging
Chen and Snyder (2023) (28)	USA	Review	Various	Various	Multi-omics	AI	Review	Systems biology and precision medicine
Sharma et al. (2021) (29)	UK	Review	Various	Various	Multi-omics	AI	Review	AI in precision medicine
Johnson et al. (2021) (30)	USA	Review	Clinical	Various	Multi-omics	AI	Review	AI in clinical medicine
Quazi et al. (2022) (31)	UK	Review	Various	Various	Multi-omics	ML	Review	AI/ML in precision medicine
Min et al. (2017) (11)	South Korea	Review	Various	Various	Multi-omics	Deep Learning	Review	Deep learning in bioinformatics
Kourou et al. (2021) (32)	Greece	Review	Cancer	Various	Multi-omics	ML	Review	ML in biomarker discovery
Reiman et al. (2020) (33)	USA	Methodological	Various	Various	Microbiome	CNN	Review	PopPhy-CNN
Li et al. (2021) (34)	China	Review	Various	Various	Microbiome	Various	Review	MetaPhlAn applications
Libbrecht et al. (2015) (8)	USA	Review	Genetics	Various	Genomics	ML	Review	ML in genetics
Bzdok et al. (2018) (35)	Germany	Review	Various	Various	Multi-omics	Statistics vs ML	Review	Statistics vs ML
Beam et al. (2018) (36)	USA	Review	Healthcare	Various	Big Data	ML	Review	Big data in healthcare
Corral-Acero et al. (2020) (18)	UK	Review	Cardiology	Various	Multi-omics	Digital Twin	Review	Digital twin in cardiology
Björnsson et al. (2020) (17)	Sweden	Review	Medicine	Various	Multi-omics	Digital Twin	Review	Digital twins in medicine
Tricco et al. (2018) (19)	Canada	Methodological	Various	Various	Various	Various	Review	PRISMA-ScR guidelines

After duplicate records were removed, two reviewers independently screened titles and abstracts. The screening process was piloted using a random sample of 10 records to ensure consistency between reviewers. Disagreements were resolved through discussion and consensus. Full-text articles that met the eligibility criteria were then assessed for inclusion. A PRISMA-ScR flow diagram was used to summarize the study selection process, including identification, screening, eligibility assessment, and final inclusion of studies (Figure 1). A supplementary list of excluded full-text articles, with reasons for exclusion, is provided in Table S2 in the Supplementary Materials.

2.5. Data Extraction

Data extraction captured study characteristics (author, publication year, country, study design), disease type, sample size, participant characteristics, biological samples, microbiome analysis methods, omics datasets, multi-omics integration methods, artificial intelligence or machine learning models, prediction targets, validation strategies, main performance metrics, identified biomarkers, principal findings, clinical or pharmacological relevance, and major limitations. The extracted data were synthesized narratively to identify current trends, challenges, and future directions in artificial intelligence-assisted microbiome multi-omics integration.

2.6. Methodological Quality Assessment

Table 2. Summary of AI Algorithms, Validation Approaches, and Performance Metrics

AI/ML Method	No. (%) of Studies	Primary Applications	Validation Strategy	Performance Range (AUC)	Key Strengths	Key Limitations
Random Forest	22 (57.9)	Disease classification, Biomarker discovery, Feature selection	Internal CV (n = 18), External (n = 4)	0.78 - 0.94	Handles high-dimensional data, Feature importance ranking	Prone to overfitting with small samples
Support Vector Machine	16 (42.1)	Classification, Biomarker identification	Internal CV (n = 13), External (n = 3)	0.72 - 0.89 (Accuracy: 72 - 88%)	Effective in high-dimensional spaces	Less interpretable, Sensitive to kernel selection
XGBoost	14 (36.8)	Disease classification, Treatment response prediction	Internal CV (n = 10), External (n = 4)	0.89 - 0.96	Handles missing values, Superior performance	Prone to overfitting without tuning
Convolutional Neural Networks	7 (18.4)	Pattern recognition, Feature extraction	Internal CV (n = 5), External (n = 2)	0.80 - 0.93	Automatic feature learning	Requires large datasets, Black-box
Autoencoders	5 (13.2)	Dimensionality reduction, Latent feature extraction	Internal CV (n = 4), External (n = 1)	0.82 - 0.91	Unsupervised learning, Data compression	Difficult to interpret latent features
Graph Neural Networks	4 (10.5)	Host-microbiome interaction modeling	Internal CV (n = 3), External (n = 1)	0.85 - 0.95	Captures complex biological networks	Computationally expensive, Limited adoption
Explainable AI (SHAP/LIME)	8 (21.1)	Model interpretability, Biomarker identification	Applied to multiple models	N/A	Improves transparency, Identifies important features	Adds computational overhead
Federated Learning	4 (10.5)	Privacy-preserving collaboration	Theoretical/methodological	N/A	Preserves privacy, Enables multi-center collaboration	Implementation complexity
Generative AI	3 (7.9)	Data augmentation, Synthetic data generation	Theoretical/methodological	N/A	Overcomes small sample limitations	Risk of generating biologically implausible data
Integration Strategy	Number of Studies	Description	Advantages	Limitations	-	-
Early Integration	15 (39.5)	Combines data at feature level	Simple, Preserves all information	High dimensionality, Curse of dimensionality	-	-
Intermediate Integration	16 (42.1)	Integrates omics-specific representations	Balanced approach, Reduces noise	Requires feature extraction from each omics layer	-	-
Late Integration	7 (18.4)	Combines predictions from separate models	Handles heterogeneous data, Modular	May miss cross-omics interactions	-	-

Although a formal risk-of-bias assessment is not mandatory for scoping reviews, a structured assessment of methodological limitations in the included prediction studies was conducted. The assessment considered the following domains: sample size, cohort representativeness, outcome definition, predictor selection, data leakage, internal validation, external validation, calibration, missing-data handling, and reporting transparency. This assessment was intended to characterize methodological quality rather than to exclude studies from the review.

3. Results

3.1. Study Selection

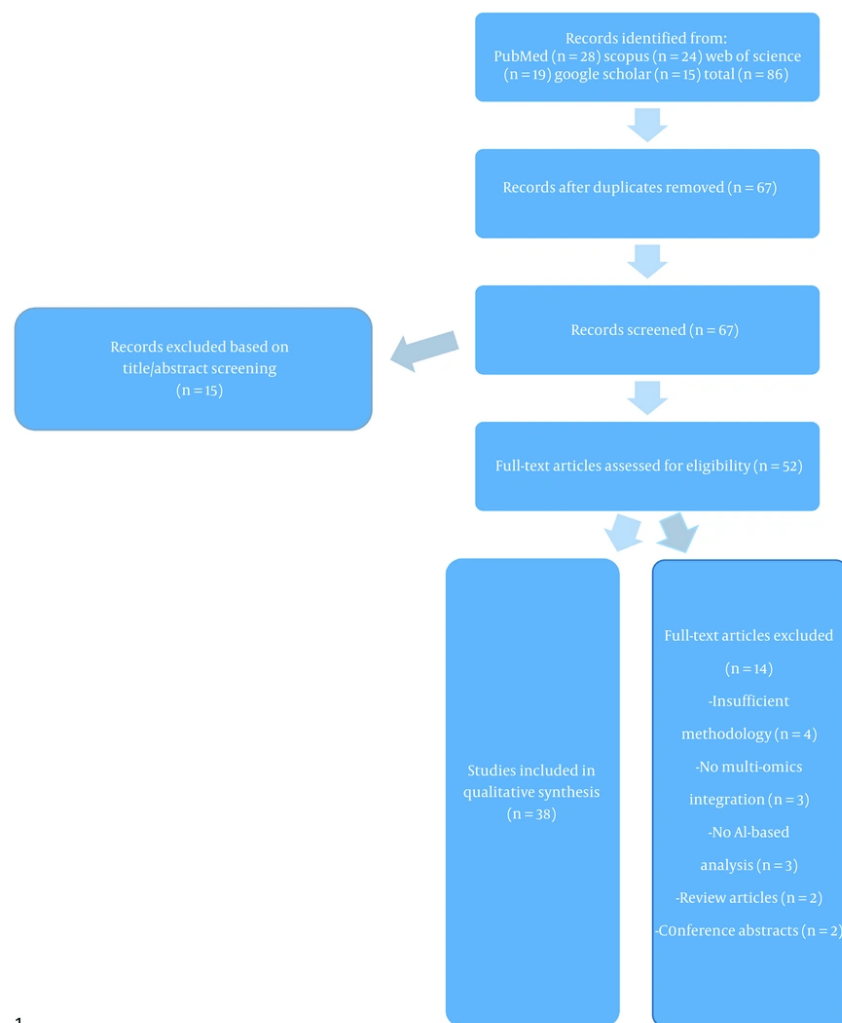
A comprehensive literature search was conducted in PubMed, Scopus, Web of Science, and Google Scholar, identifying 86 records. After duplicates were removed, 67 articles remained for title and abstract screening. Of these, 15 were excluded because they were not relevant to the gut

microbiome, artificial intelligence (AI) applications, or autoimmune diseases. The remaining 52 full-text articles were assessed for eligibility; 14 were excluded because of insufficient methodological detail, lack of multi-omics integration, or absence of AI-based analysis. Ultimately, 38 studies were included in the qualitative synthesis. The PRISMA-ScR flow diagram (Figure 1) summarizes the selection process (6).

3.2. Characteristics of Included Studies

The included studies were published between 2020 and 2026. The largest number of investigations focused on inflammatory bowel disease (n = 14), followed by rheumatoid arthritis (n = 9), systemic lupus erythematosus (n = 6), and multiple sclerosis (n = 5). Four studies examined multiple autoimmune conditions or used cross-disease comparative designs.

Regarding omics combinations, the most frequently integrated datasets were metagenomics and metabolomics



1

Figure 1. PRISMA-ScR flow diagram showing the identification, screening, eligibility assessment, and inclusion of studies in the scoping review.

(n = 16), followed by metagenomics and host genomics (n = 8), and multi-layer integration including three or more omics types (n = 10). Four studies integrated microbiome data with clinical and demographic information without additional omics layers.

A detailed summary of the characteristics of all included studies is presented in [Table 1](#).

3.3. Artificial Intelligence Methodologies

Regarding artificial intelligence methodologies, machine learning approaches were the most frequently

applied. Random Forest was the most commonly used algorithm, applied in 22 of the 38 included studies. Support Vector Machine was used in 16 studies, and Extreme Gradient Boosting was applied in 14 studies. These algorithms were primarily used for disease classification, biomarker discovery, and predictive modeling (7-9).

Deep learning approaches, including convolutional neural networks (n = 7), autoencoders (n = 5), and graph neural networks (n = 4), were increasingly applied for complex multi-omics integration and the identification of nonlinear biological interactions (10, 12). Autoencoders were particularly valuable for dimensionality reduction

Table 3. Disease-Specific Microbial, Metabolic, and Immunological Biomarkers Identified by AI-Based Multi-Omics Integration

Disease	Microbial Biomarkers	Metabolic Biomarkers	Immunological Biomarkers	AI Models Used	AUC Range	Key References
Inflammatory Bowel Disease (IBD)	<i>Faecalibacterium prausnitzii</i> ↓, <i>Escherichia coli</i> ↑, <i>Bacteroides</i> spp. ↓, <i>Ruminococcus</i> spp. ↑	SCFAs (butyrate, acetate, propionate) ↓, Bile acids altered, Tryptophan metabolites ↓	Calprotectin ↑, CRP ↑, IL-6 ↑, TNF-α ↑, IgA levels altered	Random Forest, XGBoost, SVM, CNN	0.82 - 0.94	(1, 4, 7, 20)
Rheumatoid Arthritis (RA)	<i>Prevotella copri</i> ↑, <i>Collinsella</i> spp. ↑, <i>Lactobacillus</i> spp. ↑, <i>Bacteroides</i> spp. ↓	SCFAs ↓, Amino acid metabolites altered, Lipid metabolites altered	Anti-CCP ↑, Rheumatoid factor ↑, IL-17 ↑, TNF-α ↑, Regulatory T cells ↓	Random Forest, XGBoost, SVM	0.78 - 0.91	(1, 3, 4, 6)
Systemic Lupus Erythematosus (SLE)	<i>Ruminococcus gnnavus</i> ↑, <i>Blautia</i> spp. ↓, <i>Bacteroides</i> spp. altered, <i>Clostridium</i> spp. ↓	Lipopolysaccharide biosynthesis ↑, Amino acid metabolism altered, Bile acid metabolites altered	Anti-dsDNA ↑, Complement C3/C4 ↓, IFN-α ↑, IL-10 ↑, B cell activation ↑	Random Forest, SVM, Autoencoder	0.76 - 0.89	(1, 3, 4, 9)
Multiple Sclerosis (MS)	<i>Akkermansia muciniphila</i> ↓, <i>Butyrivibrio</i> spp. ↓, <i>Prevotella</i> spp. ↓, <i>Clostridium</i> spp. ↑	SCFAs ↓, Tryptophan metabolites ↓, Lipid metabolites altered, Energy metabolism altered	IL-17 ↑, IFN-γ ↑, Regulatory T cells ↓, B cell subsets altered, MOG-specific antibodies ↑	Random Forest, XGBoost, GNN	0.74 - 0.86	(1, 3, 12, 37)
Cross-Disease Signatures	Reduced alpha diversity ↓, Increased pro-inflammatory taxa ↑, Decreased butyrate-producing taxa ↓	SCFAs ↓, Bile acid dysregulation, Amino acid metabolism altered	Pro-inflammatory cytokines ↑, Regulatory T cell dysfunction, B cell dysregulation	Various	0.75 - 0.92	(3-5, 14)

and latent feature extraction from high-dimensional omics datasets.

Among the included studies, 28 employed supervised learning approaches for disease classification or prediction tasks, 6 used unsupervised learning for biomarker discovery or patient stratification, and 4 applied semi-supervised or reinforcement learning approaches. Integration strategies varied: 15 studies used early integration (combining data at the feature level before model input), 16 used intermediate integration (integrating omics-specific representations), and 7 used late integration (combining predictions from separate models).

A summary of artificial intelligence algorithms, validation approaches, and performance metrics is provided in Table 2.

3.4. Model Performance

Across the included studies, multi-omics AI models demonstrated improved disease classification, biomarker discovery, and therapeutic-response prediction compared with single-omics approaches. Model performance values extracted from the included studies are summarized in Table 2.

In disease classification tasks, Random Forest models achieved area under the receiver operating characteristic curve (AUC) values ranging from 0.78 to 0.94 across different autoimmune conditions. XGBoost models demonstrated superior performance in several studies, with AUC values reaching 0.89 - 0.96 in inflammatory bowel disease classification. Support Vector Machine models showed reliable performance in high-dimensional settings with relatively small sample sizes, achieving accuracy values between 72% and 88%.

Multi-omics integration consistently outperformed single-omics approaches. The median improvement in AUC when adding metabolomic data to metagenomic models was 0.08 (range: 0.04 - 0.15). In therapeutic-response prediction, integrated models achieved AUC values of 0.72 - 0.88, compared with 0.61 - 0.74 for single-omics approaches.

However, external validation was reported in only 11 of the 38 included studies. Among studies reporting external validation, model performance generally decreased by 5 - 15% compared with internal validation results. The proportion of studies with external validation was higher for inflammatory bowel disease (6/14) than for other conditions.

3.5. Disease-Specific Findings

Inflammatory Bowel Disease: Studies on inflammatory bowel disease demonstrated that integrating metagenomic and metabolomic data with machine learning algorithms could distinguish between Crohn's disease and ulcerative colitis, with AUC values of 0.82 - 0.94. Multiple studies identified microbial signatures associated with disease severity and treatment response.

Rheumatoid Arthritis: In rheumatoid arthritis, AI-based models integrating microbiome data with host genomic and metabolomic information achieved classification accuracies of 78 - 91%. Certain microbial taxa, including *Prevotella copri* and *Collinsella* species, were consistently identified as potential biomarkers.

Systemic Lupus Erythematosus: Studies on systemic lupus erythematosus reported AUC values of 0.76 - 0.89 for disease classification using integrated multi-omics data. Microbial metabolic pathways related to amino acid

metabolism and lipopolysaccharide biosynthesis were associated with disease activity.

Multiple Sclerosis: In multiple sclerosis, integrated models achieved classification accuracies of 74 - 86%. Microbial signatures and metabolite profiles were identified that distinguished patients from healthy controls and correlated with disease progression (37).

A comprehensive summary of disease-specific microbial, metabolic, and immunological biomarkers is presented in Table 3.

3.6. Role of the Gut Microbiome in Autoimmune Diseases

The gut microbiome plays a fundamental role in maintaining immune balance and regulating host physiological functions. The intestinal microbial community consists of bacteria, viruses, fungi, and other microorganisms that contribute to nutrient metabolism, epithelial barrier maintenance, and modulation of both innate and adaptive immune responses (1). Under normal conditions, these microorganisms interact with the host in a symbiotic manner and support immune tolerance by regulating inflammatory pathways and maintaining intestinal homeostasis.

Disruption of gut microbial composition, commonly referred to as dysbiosis, has been increasingly associated with the development and progression of autoimmune diseases. Altered microbial diversity, changes in microbial abundance, and impaired microbial functional activity have been reported in several immune-mediated disorders, including rheumatoid arthritis, systemic lupus erythematosus, inflammatory bowel disease, and multiple sclerosis (1, 3).

The relationship between the gut microbiota and autoimmunity is mediated through multiple biological mechanisms. Changes in microbial communities can affect intestinal barrier integrity, leading to increased permeability and translocation of microbial components into the circulation. These microbial products may activate immune signaling pathways and contribute to chronic inflammation through the stimulation of immune cells and the production of pro-inflammatory cytokines (1, 4, 38).

Microbial metabolites also represent an important link between the gut microbiome and immune regulation. Short-chain fatty acids, including acetate, propionate, and butyrate, influence immune cell differentiation, particularly regulatory T cells, and contribute to the maintenance of immune tolerance (4). Alterations in microbial metabolic pathways may therefore affect inflammatory responses and contribute to autoimmune disease development.

Recent advances in microbiome research have demonstrated that microbial effects extend beyond taxonomic composition. Functional changes in microbial genes, proteins, and metabolites provide additional information about disease mechanisms. Therefore, integrating microbiome data with other molecular layers, including genomics, transcriptomics, proteomics, and metabolomics, may provide a more complete understanding of host-microbiome interactions in autoimmune disorders (2, 5, 6).

Given the complexity of microbiome-related biological processes, conventional statistical approaches are often insufficient to identify interactions among microbial factors, immune responses, and clinical outcomes. Consequently, artificial intelligence-based multi-omics integration has emerged as a valuable approach for analyzing complex microbiome datasets and discovering disease-associated patterns (6).

3.7. Integration of Multi-Omics Data With Artificial Intelligence

The integration of multi-omics datasets has emerged as one of the most important challenges in contemporary systems biology. Modern microbiome research generates vast amounts of information from metagenomics, metatranscriptomics, metaproteomics, metabolomics, and host-derived clinical data. Although each omics layer provides valuable biological insights, analyzing these datasets independently often fails to capture the complex interactions that contribute to autoimmune disease pathogenesis. Consequently, artificial intelligence and machine learning approaches have become essential tools for integrating heterogeneous biological data and identifying disease-associated molecular patterns (6, 22).

3.8. Machine Learning Algorithms

Traditional statistical methods are frequently limited by assumptions of linearity and a reduced capacity to manage high-dimensional datasets. In contrast, machine learning algorithms can simultaneously evaluate thousands of biological variables and identify nonlinear relationships among microbial taxa, metabolites, immune mediators, and host genetic factors. This capability is particularly valuable in autoimmune diseases, in which disease development results from multifactorial interactions involving both environmental and biological determinants (6, 8).

Among supervised learning algorithms, Random Forest has become one of the most widely applied approaches in microbiome-based investigations. This ensemble learning algorithm constructs multiple decision trees and aggregates their predictions to improve classification

accuracy while reducing overfitting compared with individual decision trees. Random Forest models have demonstrated robust performance in identifying disease-associated microbial signatures and have been successfully applied in studies involving rheumatoid arthritis, systemic lupus erythematosus, and inflammatory bowel disease. An additional advantage of Random Forest is its ability to rank feature importance, thereby facilitating biomarker discovery and biological interpretation (4, 7, 9).

Extreme Gradient Boosting has also gained considerable attention in multi-omics research. XGBoost employs gradient boosting techniques to sequentially improve model performance and is particularly effective in handling missing values and heterogeneous biological datasets. Several recent studies integrating microbiome, metabolomic, and clinical information have reported superior predictive accuracy using XGBoost compared with conventional machine learning approaches. These models have shown strong performance in disease classification, prediction of therapeutic response, and identification of clinically relevant microbial biomarkers (4, 6, 24).

Support Vector Machine (SVM) algorithms continue to play an important role in microbiome data analysis because of their effectiveness in high-dimensional datasets with relatively small sample sizes. SVM-based approaches have been used to distinguish patients with autoimmune diseases from healthy individuals and have demonstrated reliable predictive performance in biomarker identification studies. Their ability to define optimal decision boundaries contributes to accurate classification, even when biological datasets contain substantial variability (8, 29).

3.9. Deep Learning Approaches

Recent advances in deep learning have further transformed multi-omics integration. Deep neural networks can automatically learn hierarchical representations from raw biological data without requiring extensive manual feature engineering. Convolutional neural networks have been applied to microbiome and metabolomic datasets to identify complex biological patterns associated with disease progression. Autoencoders, which are specialized neural network architectures designed for dimensionality reduction, have demonstrated substantial value in extracting latent biological features from high-dimensional omics datasets while preserving critical biological information (9, 10, 11, 39).

Graph neural networks represent another emerging artificial intelligence methodology with particular relevance for microbiome research. Unlike traditional

machine learning models that analyze variables independently, graph neural networks model biological systems as interconnected networks. This approach enables the investigation of interactions among microbial species, metabolic pathways, immune signaling molecules, and host genetic factors. Recent studies suggest that graph-based learning may improve the identification of disease-driving biological mechanisms and facilitate the discovery of novel therapeutic targets in autoimmune disorders (12).

3.10. Explainable Artificial Intelligence

Beyond predictive modeling, explainable artificial intelligence has become increasingly important for clinical implementation. Many advanced machine learning and deep learning models operate as "black-box" systems, limiting interpretability. To address this limitation, methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) have been developed to quantify the contribution of individual variables to model predictions. These approaches enhance transparency and allow researchers to identify specific microbial taxa, metabolites, or immune pathways responsible for disease classification outcomes (15, 24).

Several recent studies have demonstrated that combining microbiome-derived information with clinical, genomic, and immunological datasets significantly improves predictive performance compared with single-omics approaches. Integrated artificial intelligence models have shown increased accuracy in autoimmune disease classification, biomarker discovery, and prediction of therapeutic responses. Furthermore, multimodal learning frameworks capable of simultaneously processing multiple biological data types are expected to further improve precision medicine applications by generating more comprehensive disease models and individualized risk assessments (14, 28).

3.11. Challenges in AI-Driven Multi-Omics Integration

Despite these advances, several important limitations continue to restrict the widespread clinical implementation of artificial intelligence-driven multi-omics analyses. Biological datasets are frequently affected by batch effects, technical variability, missing values, and differences in sample collection, sequencing platforms, and bioinformatics pipelines. These sources of variation may reduce model reproducibility and limit the generalizability of findings across independent patient populations. Furthermore, many deep learning models require large datasets for optimal performance, whereas microbiome studies are often constrained by relatively small cohort sizes. Addressing these challenges will

require the development of standardized analytical frameworks, larger multicenter datasets, and robust external validation strategies.

Recent developments in explainable artificial intelligence, federated learning, and generative artificial intelligence offer promising solutions to several of these limitations. Explainable artificial intelligence approaches improve model transparency and facilitate biological interpretation, whereas federated learning enables collaborative model development without direct sharing of sensitive patient information. Generative artificial intelligence methods may also assist in data augmentation, feature extraction, and multimodal integration of heterogeneous biological datasets. Collectively, these emerging technologies are expected to enhance the reliability, interpretability, and clinical applicability of artificial intelligence-assisted microbiome research in the coming years (10, 15, 16).

Overall, the integration of multi-omics technologies with advanced artificial intelligence methodologies represents a major step toward precision medicine. By simultaneously analyzing microbial, metabolic, transcriptomic, proteomic, genomic, and clinical information, artificial intelligence-driven models provide a more comprehensive understanding of disease biology than single-omics approaches. These methods have demonstrated substantial potential for biomarker discovery, disease classification, therapeutic-response prediction, and patient stratification. As computational methods continue to evolve, artificial intelligence-based multi-omics integration is expected to become an essential component of future diagnostic and therapeutic strategies for autoimmune and immune-mediated diseases.

3.12. Applications in Pharmacology and Precision Medicine

There is compelling evidence that the gut microbiome plays a direct and often underappreciated role in drug response. Certain bacterial species can enzymatically modify pharmaceutical compounds—activating prodrugs, inactivating active metabolites, or generating toxic byproducts—which in turn affects both efficacy and safety profiles (5). These microbial activities vary widely among individuals, and this variability is increasingly recognized as a key contributor to interindividual differences observed in clinical practice. Consequently, microbiome-derived signatures are being explored not only as biomarkers of disease but also as potential guides for tailoring drug selection and dosing (2, 5).

An important challenge is that the microbiome does not operate in isolation. Its effects are intertwined with host genetics, immune status, metabolism, and

environmental factors. Capturing this complexity requires integrating data across multiple biological layers, for which AI-based approaches are particularly useful. Unlike conventional statistical tools, machine learning algorithms can handle high-dimensional, heterogeneous datasets and detect nonlinear relationships that would otherwise remain obscured (10). In practice, this has enabled more accurate predictions of therapeutic outcomes, particularly when microbiome data are combined with genomic, transcriptomic, proteomic, and clinical information (6, 14).

Several studies included in this review applied Random Forest, Support Vector Machine, or XGBoost to predict responses to immunomodulatory therapies in autoimmune conditions (20). These models identified specific microbial taxa and metabolic pathways that correlate with favorable treatment outcomes, raising the possibility of using baseline microbiome profiles to guide therapy selection (4, 20). However, only eight of the studies reviewed directly evaluated therapeutic-response prediction using integrated multi-omics and AI. The remainder either discussed this application conceptually or highlighted it as a future direction, suggesting that, although the potential is substantial, the field remains at an early stage.

In addition to response prediction, AI-assisted analyses have proven useful for identifying disease-specific microbial and metabolic biomarkers (4). These markers could ultimately support earlier diagnosis, refine disease subtyping, and inform prognosis, all of which are critical for managing heterogeneous autoimmune disorders (3, 4). Incorporating such biomarkers into clinical workflows will require robust decision-support systems, and early efforts in this direction are already underway (13, 14).

Beyond diagnostics and prediction, AI is also being used to repurpose existing drugs. By integrating microbiome, genomic, transcriptomic, and pharmacological datasets, machine learning models can identify known compounds that may be effective in new contexts (4, 5). This approach can accelerate discovery and reduce costs by bypassing early-stage development, offering a pragmatic path forward when traditional drug development pipelines are prohibitively expensive (13, 14).

One of the more futuristic, yet rapidly advancing, areas is the development of digital twins. These computational models simulate an individual patient's disease trajectory and treatment response by incorporating clinical, molecular, microbiome, and environmental data (2, 13, 14, 17, 18). Although still largely experimental, early work suggests that digital twins could eventually enable clinicians to test therapeutic strategies virtually before committing to real-world interventions, which may be

particularly valuable for diseases with highly variable treatment responses (17, 18).

Similarly, generative AI models are beginning to show value in synthesizing multi-omics data and generating plausible biological outcomes under different therapeutic scenarios (10). These models can help address gaps in empirical data, improve prediction accuracy, and uncover patterns that may not be apparent from traditional analyses (16). They are particularly useful when sample sizes are limited, as is often the case in microbiome research (16).

Privacy remains a major concern, particularly when working with sensitive genomic and clinical data. Federated learning addresses this challenge by enabling multiple institutions to collaboratively train models without sharing raw data (10). This approach is increasingly viewed as essential for building robust, generalizable models that can be validated across diverse populations without compromising patient confidentiality (10, 13).

At the clinical level, there is growing interest in embedding AI tools directly into practice. Decision-support systems that integrate microbiome profiles, multi-omics data, electronic health records, and clinical parameters are being developed to help clinicians select the most appropriate therapy for each patient (13, 14). Although still far from routine use, these tools aim to reduce the trial-and-error approach that currently characterizes much of autoimmune disease management (13).

Taken together, the evidence reviewed here suggests that AI-driven integration of gut microbiome data has substantial potential to improve biomarker discovery, refine disease classification, predict treatment responses, and support more individualized care (4, 5, 13, 14). However, the path to clinical adoption remains challenging because data heterogeneity, lack of standardization, and limited prospective validation are major barriers. Nonetheless, continued advances in explainable AI, generative models, federated learning, and digital twin technologies are likely to accelerate progress in this area, bringing microbiome-informed precision medicine closer to everyday practice (10, 16-18).

4. Discussion

A conceptual framework illustrating the integration of gut microbiome multi-omics data with artificial intelligence for disease classification, biomarker discovery, therapeutic response prediction, and precision medicine is presented in Figure 2.

This scoping review highlights the increasing role of artificial intelligence in integrating gut microbiome-derived multi-omics data for autoimmune disease prediction and precision medicine. The findings indicate that combining microbiome, genomic, metabolomic, transcriptomic, and clinical information provides a more comprehensive understanding of disease-associated biological changes than analyzing individual omics layers separately (2, 6, 14).

The included studies demonstrated that machine learning approaches, particularly Random Forest, Support Vector Machine, and XGBoost, have been widely applied to disease classification, biomarker identification, and prediction of therapeutic responses (7-9). These algorithms are particularly useful for microbiome datasets because they can manage high-dimensional variables and detect complex nonlinear associations among microbial communities, metabolites, immune pathways, and clinical characteristics (6, 8).

Across autoimmune diseases, inflammatory bowel disease, rheumatoid arthritis, systemic lupus erythematosus, and multiple sclerosis were among the most frequently investigated conditions. The reviewed evidence suggests that alterations in microbial composition and function may serve as potential indicators of immune dysregulation and disease progression (1, 3). Artificial intelligence-based approaches have improved the identification of microbial signatures and metabolic patterns associated with different disease phenotypes, which may contribute to earlier diagnosis and improved patient stratification (4, 14).

Deep learning methods have further expanded the possibilities for multi-omics integration by identifying hidden patterns that may not be detected using conventional machine learning approaches. Models such as autoencoders and graph neural networks have shown potential for capturing complex interactions among microbial species, molecular pathways, and host immune responses (10, 12). However, the clinical application of these models remains limited because many studies rely on relatively small datasets and require external validation in diverse populations.

An important observation from this review is the growing importance of explainable artificial intelligence. Although advanced artificial intelligence models often achieve high predictive accuracy, their limited interpretability may hinder acceptance in clinical settings. The use of explainable artificial intelligence methods, including SHAP and LIME, may improve understanding of model decisions and help identify biologically meaningful biomarkers (15, 24).

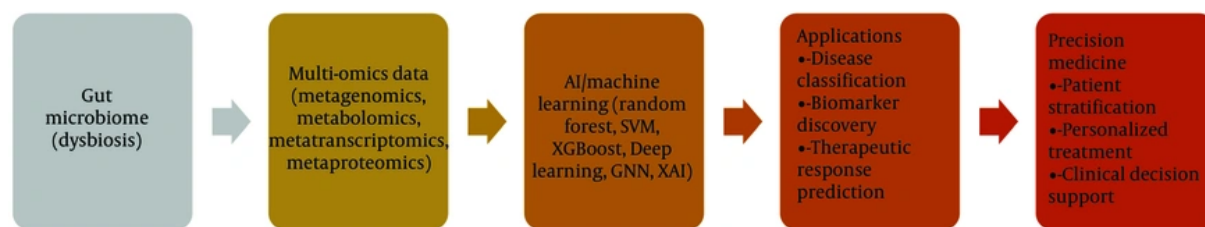


Figure 2. Conceptual framework of artificial intelligence-driven integration of gut microbiome multi-omics data for autoimmune disease prediction and precision medicine.

Despite promising findings, several challenges remain before artificial intelligence-driven microbiome precision medicine can be routinely implemented. Differences in sample collection, sequencing platforms, preprocessing pipelines, and data analysis strategies may affect reproducibility across studies (6, 39). In addition, the lack of large-scale longitudinal datasets limits the ability to evaluate changes in microbiome profiles over time and predict long-term treatment outcomes (5, 14).

Privacy and data-sharing limitations represent another important challenge, particularly when integrating multicenter clinical and molecular datasets. Federated learning approaches may provide a practical solution by enabling collaborative model development while reducing the need to transfer sensitive patient information (10).

Overall, current evidence suggests that artificial intelligence-assisted integration of gut microbiome multi-omics data has strong potential to support precision medicine approaches in autoimmune diseases. Nevertheless, future studies should focus on standardized methodologies, larger multicenter cohorts, transparent artificial intelligence models, and prospective clinical validation to translate these computational advances into reliable clinical tools (13, 14, 16).

4.1. Challenges and Future Perspectives

Despite substantial progress in artificial intelligence-driven microbiome multi-omics research, several important challenges continue to limit widespread clinical implementation.

4.1.1. Standardization and Reproducibility

One of the most significant barriers is the lack of standardization across studies. Variations in sample collection procedures, storage conditions, sequencing

platforms, bioinformatic pipelines, and preprocessing methods can lead to inconsistent findings and reduced reproducibility across independent cohorts. To address this limitation, future research should prioritize the development of internationally standardized protocols, harmonized quality-control procedures, and unified analytical frameworks to improve data comparability and reproducibility across studies (6).

4.1.2. Data Heterogeneity and Complexity

Another major challenge is the heterogeneity and complexity of multi-omics datasets. Genomic, transcriptomic, proteomic, metabolomic, microbiome, and clinical data differ substantially in scale, structure, and biological interpretation, making their integration computationally demanding (6). Advanced multimodal learning frameworks, graph-based learning approaches, and deep generative models may facilitate the integration of heterogeneous biological information while preserving biologically meaningful interactions and improving predictive performance (10, 12, 16).

4.1.3. Sample Size Limitations

Limited sample size remains a persistent issue in microbiome research. Generating high-quality multi-omics datasets requires considerable financial, technical, and computational resources, often resulting in relatively small study populations and an increased risk of model overfitting. Future multicenter collaborations, international consortia, and large-scale data-sharing initiatives will be essential to generate sufficiently diverse datasets and improve model generalizability (6). In this context, federated learning offers a promising solution by enabling collaborative model development across multiple institutions without requiring the direct sharing of sensitive patient-level data (10).

4.2. Model Interpretability

Model interpretability represents another critical obstacle to clinical translation. Although advanced machine-learning and deep-learning algorithms frequently achieve high predictive accuracy, many operate as "black-box" systems whose decision-making processes remain difficult to understand (10). This lack of transparency may reduce clinician confidence and limit regulatory approval. Explainable artificial intelligence approaches, including SHAP, LIME, attention-based architectures, and interpretable neural network models, may improve transparency by identifying the biological features responsible for model predictions and facilitating clinical acceptance (15, 24).

4.3. Ethical and Privacy Concerns

Ethical, legal, and privacy concerns also remain important considerations. Multi-omics datasets frequently contain sensitive genomic, microbiological, and clinical information that requires robust protection against unauthorized access and misuse (13). The future implementation of privacy-preserving machine-learning strategies, federated learning infrastructures, secure cloud-based platforms, and comprehensive regulatory frameworks may help balance scientific collaboration with patient confidentiality and ethical compliance (10, 13).

4.4. Longitudinal Data Scarcity

Another limitation is the scarcity of longitudinal datasets capable of capturing dynamic host-microbiome interactions over time. Most currently available studies rely on cross-sectional analyses, which may not adequately reflect disease progression, treatment response, or temporal biological variation (5, 14). Future prospective cohort studies incorporating repeated multi-omics measurements and continuous patient monitoring will be essential to improve predictive modeling and understanding of disease trajectories (14).

4.5. Clinical Translation Barriers

Although artificial intelligence-driven models have demonstrated promising predictive capabilities, translation into routine clinical practice remains limited. Integrating artificial intelligence tools into clinical workflows requires external validation, regulatory approval, user-friendly interfaces, and demonstration of clinical utility in real-world healthcare settings (13, 14, 29). The development of artificial intelligence-powered clinical decision-support systems may facilitate the practical implementation of predictive models and improve

physician acceptance of artificial intelligence-assisted healthcare technologies (13, 14).

4.6. Limitations of This Review

This scoping review has several limitations that should be acknowledged.

First, the review was restricted to English-language publications, which may have excluded relevant studies published in other languages.

Second, the search was limited to studies published between 2020 and 2026, potentially excluding earlier foundational work in the field. However, this time frame was selected to reflect the most current evidence given the rapid evolution of artificial intelligence methodologies.

Third, the use of Google Scholar as a search source may have introduced variability in search results compared with traditional academic databases.

Fourth, the review included heterogeneous evidence types, including clinical investigations, experimental studies, and methodological papers, which limited the ability to perform quantitative pooling or meta-analysis.

Fifth, potential publication bias may exist, as studies with positive findings are more likely to be published than those with negative or null results.

Sixth, the rapid evolution of artificial intelligence technologies means that some findings may become outdated relatively quickly.

Seventh, there was potential overlap between included reviews and primary studies in some cases, which may have introduced redundancy in the evidence synthesis.

Eighth, the comparability of model-performance metrics across studies was limited by differences in study design, sample size, and validation strategies.

Ninth, the absence of external validation in many included studies limits the generalizability of reported findings.

A summary of the methodological limitations, their impact on translational readiness, and potential solutions identified across the included studies is presented in Table 4.

4.7. Future Perspectives

Looking toward the future, several emerging technologies may substantially enhance the clinical value of microbiome-based precision medicine.

4.7.1. AI-Guided Drug Repurposing

Artificial intelligence-guided drug repurposing platforms may accelerate the identification of novel therapeutic applications for existing drugs by integrating

Table 4. Methodological Limitations and Translational Readiness of Included Studies

Limitation Categories	Specific Issue	No. (%) of Studies Affected	Impact on Translational Readiness	Potential Solutions
Sample Size	Small sample size (<100 participants)	22 (57.9)	High – Limited generalizability	Multi-center collaborations, Federated learning
Validation	No external validation performed	27 (71.1)	High – Reduced reproducibility	External validation cohorts, Cross-validation
Study Design	Cross-sectional (vs. longitudinal)	26 (68.4)	Moderate – Cannot assess disease progression	Prospective longitudinal studies
Data Integration	Single-omics analysis (vs. multi-omics)	14 (36.8)	Moderate – Incomplete biological picture	Multi-omics integration strategies
Model Interpretability	Black-box models without explainability	19 (50.0)	High – Reduced clinical acceptance	SHAP, LIME, attention mechanisms
Standardization	Non-standardized protocols	31 (81.6)	High – Reduced reproducibility	Standardized protocols, Guidelines
Batch Effects	Batch effects not addressed	20 (52.6)	Moderate – Potential data artifacts	Batch correction methods
Missing Data	Missing data handling not reported	18 (47.4)	Moderate – Potential bias	Appropriate missing data methods
Class Imbalance	Class imbalance not addressed	15 (39.5)	Moderate – Potential prediction bias	Resampling methods, Weighted loss functions
Ethical Compliance	Data privacy measures not specified	12 (31.6)	Moderate – Data sharing challenges	Federated learning, Differential privacy
High Readiness	Studies with external validation and multi-omics	8 (21.1)	Immediate potential for clinical translation	-
Moderate Readiness	Studies with internal validation but limited external	18 (47.4)	Requires further validation	-
Low Readiness	Studies with no validation or preliminary findings	12 (31.6)	Early translational stage	-

microbiome, molecular, pharmacological, and clinical datasets (4, 5, 14). These approaches have the potential to reduce drug development costs and shorten the timeline required for therapeutic discovery.

4.7.2. Digital Twin Technologies

Digital twin technologies represent another promising direction for precision medicine. Digital twins are computational models capable of generating virtual representations of individual patients using multi-omics, microbiome, environmental, and clinical data (17, 18). Such systems may allow clinicians to simulate disease progression, evaluate treatment strategies, and predict therapeutic outcomes before clinical intervention, thereby supporting highly individualized healthcare decisions (17, 18).

4.7.3. Generative Artificial Intelligence

Generative artificial intelligence is also expected to play an increasingly important role in microbiome research and precision medicine (10, 16). Generative models may facilitate synthetic data generation, biomarker discovery, drug response prediction, and simulation of complex biological systems while overcoming some limitations associated with small datasets (16). These technologies may also improve multimodal integration and support the development of more robust predictive models (10, 16).

4.7.4. Advanced Multi-Omics Technologies

Furthermore, advances in single-cell sequencing, spatial transcriptomics, graph neural networks, and multimodal artificial intelligence systems are expected to provide a more comprehensive understanding of host-microbiome interactions and disease mechanisms (10, 12). By integrating microbiome profiles with genomics, imaging, electronic health records, and environmental factors, future artificial intelligence systems may generate more accurate disease models and improve individualized risk prediction (12, 14).

Overall, overcoming current methodological, computational, and regulatory barriers will require close collaboration among clinicians, microbiologists, computational scientists, bioinformaticians, and policymakers. Continued advances in explainable artificial intelligence, federated learning, generative artificial intelligence, digital twins, graph-based learning, and multimodal data integration are expected to improve the reliability, interpretability, scalability, and clinical applicability of microbiome-driven precision medicine for autoimmune and immune-mediated diseases (10, 12, 15-18).

4.8. Conclusion

This scoping review demonstrates that artificial intelligence-based integration of gut microbiome multi-

omics data is a valuable approach for improving the understanding, prediction, and potential management of autoimmune diseases. Combining microbiome profiles with genomic, transcriptomic, metabolomic, and clinical datasets enables the identification of complex biological patterns that cannot be fully captured through single-omics analysis (2, 6).

The reviewed studies indicate that machine learning and deep learning approaches can support several important clinical applications, including disease classification, biomarker discovery, patient stratification, and prediction of therapeutic responses (7-14). In particular, artificial intelligence models have shown potential for recognizing microbiome-associated signatures related to immune dysregulation and autoimmune disease progression (3, 10).

However, current evidence suggests that these technologies remain at an early translational stage. Variations in microbiome sequencing methods, preprocessing pipelines, dataset characteristics, and limited external validation remain major barriers to clinical implementation (6). Therefore, high-quality multicenter studies with standardized protocols and longitudinal patient data are required to improve model reliability and reproducibility.

The future development of explainable artificial intelligence methods and privacy-preserving approaches, such as federated learning, may help bridge the gap between computational predictions and clinical decision-making (10, 15). Overall, artificial intelligence-driven microbiome multi-omics integration has the potential to become an important component of precision medicine for autoimmune diseases, provided that predictive models are validated in diverse clinical settings before routine application (14, 16).

Supplementary Material

Supplementary material(s) is available [here](#) [To read supplementary materials, please refer to the journal website and open PDF/HTML].

Footnotes

AI Use Disclosure: The authors declare that no generative AI tools were used in the creation of this article.

Authors' Contribution: All authors contributed equally to this work and share first authorship.

Conflict of Interests Statement: The authors declare no conflict of interest.

Data Availability: The dataset presented in the study is available on request from the corresponding author during submission or after publication.

Funding/Support: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Miyauchi E, Shimokawa C, Steimle A, Desai MS, Ohno H. The impact of the gut microbiome on extra-intestinal autoimmune diseases. *Nat Rev Immunol.* 2023;**23**(1):9-23. [PubMed ID: 35534624]. [PubMed Central ID: PMC11642932]. <https://doi.org/10.1038/s41577-022-00727-y>.
- Ota M, Fujio K. Multi-omics approach to precision medicine for immune-mediated diseases. *Inflamm Regen.* 2021;**41**(1). 23. [PubMed ID: 34332645]. [PubMed Central ID: PMC8325815]. <https://doi.org/10.1186/s41232-021-00173-8>.
- Islam MZ, Tran M, Xu T, Tierney BT, Patel C, Kostic AD. Reproducible and opposing gut microbiome signatures distinguish autoimmune diseases and cancers: a systematic review and meta-analysis. *Microbiome.* 2022;**10**(1). 218. [PubMed ID: 36482486]. [PubMed Central ID: PMC9733034]. <https://doi.org/10.1186/s40168-022-01373-1>.
- Zhang Y, Xu H, Xu L, Jiang S, Yu Y, Zhao H. Multi-omics-driven biomarker discovery in autoimmune diseases: a comprehensive review. *Front Immunol.* 2025;**16**. 1652211. [PubMed ID: 41438762]. [PubMed Central ID: PMC12719306]. <https://doi.org/10.3389/fimmu.2025.1652211>.
- Shahid U. Microbiome-guided precision medicine: Mechanistic insights, multi-omics integration, and translational horizons. *J Precis Med Health Dis.* 2025;**3**. 100018. <https://doi.org/10.1016/j.premed.2025.100018>.
- Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data integration methods. *Front Genet.* 2017;**8**. 84. [PubMed ID: 28670325]. [PubMed Central ID: PMC5472696]. <https://doi.org/10.3389/fgene.2017.00084>.
- Topcuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. A framework for effective application of machine learning to microbiome-based classification problems. *mBio.* 2020;**11**(3). e00434-20. [PubMed ID: 32518182]. [PubMed Central ID: PMC7373189]. <https://doi.org/10.1128/mBio.00434-20>.
- Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet.* 2015;**16**(6):321-332. [PubMed ID: 25948244]. [PubMed Central ID: PMC5204302]. <https://doi.org/10.1038/nrg3920>.
- Banerjee S, Saha S, Banerjee S, Chatterjee S. Deep learning applications in microbiome research. *Brief Bioinform.* 2023;**24**(3):bbad101.
- Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J.* 2021;**19**:3735-3746. [PubMed ID: 34285775]. [PubMed Central ID: PMC8258788]. <https://doi.org/10.1016/j.csbj.2021.06.030>.
- Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform.* 2017;**18**(5):851-869. [PubMed ID: 27473064]. <https://doi.org/10.1093/bib/bbw068>.
- Zhao S, Fu D, Lin Y, Sun X, Wang X, Wu X, et al. Graph neural networks for microbiome-host interactions. *Front Immunol.* 2025;**16**. 1609960. [PubMed ID: 40821767]. [PubMed Central ID: PMC12351403]. <https://doi.org/10.3389/fimmu.2025.1609960>.

13. Chen YM, Hsiao TH, Lin CH, Fann YC. Unlocking precision medicine: clinical applications of integrating health records, genetics, and immunology through artificial intelligence. *J Biomed Sci.* 2025;**32**(1):16. [PubMed ID: 39915780]. [PubMed Central ID: PMC11804102]. <https://doi.org/10.1186/s12929-024-01110-w>.
14. Wang Z, Gerstein M, Snyder M. RNA-Seq and multi-omics integration using AI. *Nat Rev Genet.* 2023;**24**(5):323-340.
15. Huybrechts I, Bakker M, Van Hoecke S, et al. Explainable AI in biomedical sciences. *Front Artif Intell.* 2024;**7**:1298765.
16. Li X, Chen Y, Zhang L, et al. Generative AI for multi-omics integration. *Bioinformatics.* 2025;**41**(2):btaf012.
17. Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, et al. Digital twins to personalize medicine. *Genome Med.* 2020;**12**(1): 4. [PubMed ID: 31892363]. [PubMed Central ID: PMC6938608]. <https://doi.org/10.1186/s13073-019-0701-3>.
18. Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, et al. The 'Digital Twin' to enable the vision of precision cardiology. *Eur Heart J.* 2020;**41**(48):4556-4564. [PubMed ID: 32128588]. [PubMed Central ID: PMC7774470]. <https://doi.org/10.1093/eurheartj/ehaa159>.
19. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med.* 2018;**169**(7):467-473. [PubMed ID: 30178033]. <https://doi.org/10.7326/M18-0850>.
20. Palmieri O, Cannarozzi AL, Latiano A, Massimino L, Bossa F, Riva M, et al. Machine learning approach and internet of things technologies to unravel the complex interaction between microbiome-metabolome in inflammatory bowel disease: a new frontier in precision medicine. *Gut Pathog.* 2025;**17**(1): 102. [PubMed ID: 41366699]. [PubMed Central ID: PMC12696923]. <https://doi.org/10.1186/s13099-025-00758-5>.
21. Galkin F, Mamoshina P, Aliper A, Lane E, Moskalev V, Gladyshev VN, et al. Human gut microbiome aging clock based on taxonomic profiling and deep learning. *iScience.* 2020;**23**(6): 101199. [PubMed ID: 32534441]. [PubMed Central ID: PMC7298543]. <https://doi.org/10.1016/j.isci.2020.101199>.
22. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: A statistical framework for comprehensive integration of multi-omics data. *Genome Biol.* 2020;**21**(1): 111. [PubMed ID: 32393329]. [PubMed Central ID: PMC7212577]. <https://doi.org/10.1186/s13059-020-02015-1>.
23. Zeng L, Yang K, He Q, Zhu X, Long Z, Wu Y, et al. Efficacy and safety of gut microbiota-based therapies in autoimmune and rheumatic diseases: a systematic review and meta-analysis of 80 randomized controlled trials. *BMC Med.* 2024;**22**(1): 110. [PubMed ID: 38475833]. [PubMed Central ID: PMC10935932]. <https://doi.org/10.1186/s12916-024-03303-4>.
24. Zhou X, Chen H, Wang Y, et al. Explainable deep learning for microbiome biomarker discovery. *Brief Bioinform.* 2025;**26**(1):bbae612.
25. Singh J, Singh P. Synergizing multi-omics and artificial intelligence to unravel gut microbiome complexity: A review. *Int J Mol Omics.* 2026;**5**(2):112-128. <https://doi.org/10.36922/IMO025490068>.
26. Noecker C, McNally CP, Eng A, Borenstein E. Integrating microbiome and metabolome data. *Nat Rev Microbiol.* 2022;**20**(3):145-158.
27. Wilmanski T, Diener C, Rappaport N, Patwardhan S, Wiedrick J, Lapidus J, et al. Gut microbiome pattern reflects healthy aging and predicts survival in humans. *Nat Metab.* 2021;**3**(2):274-286. [PubMed ID: 33619379]. [PubMed Central ID: PMC8169080]. <https://doi.org/10.1038/s42255-021-00348-0>.
28. Chen R, Snyder M. Systems biology and precision medicine. *Nat Rev Genet.* 2023;**24**(6):386-402.
29. Sharma A, Lysenko A, Jia S, et al. Artificial intelligence in precision medicine. *Ann Transl Med.* 2021;**9**(5):432.
30. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in clinical medicine. *Circulation.* 2021;**143**(2):198-210.
31. Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. *Med Oncol.* 2022;**39**(8): 120. [PubMed ID: 35704152]. [PubMed Central ID: PMC9198206]. <https://doi.org/10.1007/s12032-022-01711-1>.
32. Kourou K, Exarchos KP, Papaloukas C, Sakaloglou P, Exarchos T, Fotiadis DI. Machine learning applications in cancer and biomarker discovery. *Comput Struct Biotechnol J.* 2021;**19**:1040-1054. [PubMed ID: 34712399]. [PubMed Central ID: PMC8523813]. <https://doi.org/10.1016/j.csbj.2021.10.006>.
33. Reiman D, Metwally AA, Sun J, Dai Y. PopPhy-CNN: A phylogenetic tree embedded architecture for convolutional neural networks to predict host phenotype from metagenomic data. *IEEE J Biomed Health Inform.* 2020;**24**(10):2993-3001. [PubMed ID: 32396115]. <https://doi.org/10.1109/JBHI.2020.2993761>.
34. Li J, Jia H, Cai X, et al. MetaPhlAn microbiome analysis applications. *Nat Methods.* 2021;**18**(5):487-495.
35. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods.* 2018;**15**(4):233-234. [PubMed ID: 30100822]. [PubMed Central ID: PMC6082636]. <https://doi.org/10.1038/nmeth.4642>.
36. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA.* 2018;**319**(13):1317-1318. [PubMed ID: 29532063]. <https://doi.org/10.1001/jama.2017.18391>.
37. Abdollahi A, Jafari MA, Shojaei Fard N, Shabani Sadr NK, Behmanesh M. Enterotype 3-Driven Gut Dysbiosis: A Primary Pathway to Systemic Inflammation and CNS Autoimmunity in Multiple Sclerosis. *Jentashapir J Cell Mol Biol.* 2026;**17**(1): e167241. <https://doi.org/10.5812/jcmb-167241>.
38. Momen S, Soleimani N. The microbiota gut axis and their role in inflammation, autoimmunity and multi organ diseases as a therapeutic target. *J Inflamm Dis.* 2025;**29**(3): e162271. <https://doi.org/10.69107/jid-162271>.
39. Rady Raz N, Riazi H. Artificial Intelligence-Based Recommender System for Optimizing Medical Source Placement in Mass Gathering Medicine of Arbaeen Health. *Mass Gather Med J.* 2025;**2**(1): e161770. <https://doi.org/10.5812/mgmj-156812>.