2026, 66(10): 4840-4854
CSTR: 32112.14.j.AMS.20260244
DOI: 10.13343/j.cnki.wsxb.20260244

微生物学报
Acta Microbiologica Sinica
http://journals.im.ac.cn/actamicrocn

Review 综述

智能共生：基于人工智能的益生菌研究与微生态个性化营养

时凤翠[1*]，李泽阳[1]，景艺凡[1]，李咏诗[1]，马文建[1,2*]

1 齐鲁理工学院 生物与化学工程学院，山东 济南
2 天津科技大学 生物工程学院，天津

时凤翠, 李泽阳, 景艺凡, 李咏诗, 马文建. 智能共生：基于人工智能的益生菌研究与微生态个性化营养[J]. 微生物学报, 2026, 66(10): 4840-4854.
SHI Fengcui, LI Zeyang, JING Yifan, LI Yongshi, MA Wenjian. Intelligent symbiosis: artificial intelligence-enabled probiotic research and microecological personalized nutrition[J]. Acta Microbiologica Sinica, 2026, 66(10): 4840-4854.

摘　要：人体微生物组在维持健康方面发挥着至关重要的作用，其失调与多种疾病密切相关，包括肠道炎症、代谢紊乱和神经系统疾病。随着微生物组数据集规模和复杂性的不断增加，传统统计方法通常无法捕捉这些高维、稀疏且结构化数据中的复杂模式。本文系统性地探讨了人工智能(artificial intelligence, AI)，特别是机器学习(machine learning, ML)和深度学习(deep learning, DL)技术在益生菌研究及微生物生态个性化营养领域的整合应用；评估了 AI 如何通过元基因组数据加速新型益生菌候选物的发现，如何提高对益生菌功效及作用机制的预测能力，如何通过多组学整合解析复杂的宿主-微生物相互作用，以及如何实现个性化干预策略的设计。本文强调了微生物组健康管理领域正经历从经验性、一刀切方法向数据驱动、个性化策略的范式转变，为精准营养和疾病预防提供了新的机遇。

关键词：人工智能；益生菌；微生物组；个性化营养；机器学习；深度学习

资助项目：齐鲁理工学院科研计划(QIT24NN005)；山东省自然科学基金(ZR2022MC057)
This work was supported by the Research Program of Qilu Institute of Technology (QIT24NN005) and the Natural Science Foundation of Shandong Province (ZR2022MC057).
*Corresponding authors. E-mail: SHI Fengcui, 15066119131@163.com; MA Wenjian, ma_wj@tust.edu.cn
Received: 2026-03-25; Accepted: 2026-05-23; Published online: 2026-06-16

Intelligent symbiosis: artificial intelligence-enabled probiotic research and microecological personalized nutrition

SHI Fengcui[1*], LI Zeyang[1], JING Yifan[1], LI Yongshi[1], MA Wenjian[1,2*]

1 College of Biological and Chemical Engineering, Qilu Institute of Technology, Jinan, Shandong, China
2 College of Biotechnology, Tianjin University of Science and Technology, Tianjin, China

Abstract: The human microbiome plays a crucial role in maintaining health, and its dysregulation is strongly associated with various diseases, including intestinal inflammation, metabolic disorders, and neurological conditions. As microbiome datasets grow in size and complexity, conventional statistical methods often fail to capture the intricate patterns within these high-dimensional, sparse, and structured data. This review systematically examines the integration of artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, in probiotic research and microecological personalized nutrition. We evaluate how AI accelerates the discovery of novel probiotic candidates from metagenomic data, improves prediction of probiotic efficacy and mechanism of action, deciphers complex host-microbe interactions through multi-omics integration, and enables the design of personalized intervention strategies. This review highlights a paradigm shift from empirical, one-size-fits-all approaches toward data-driven, personalized strategies in microbiome-based health management, offering new opportunities for precision nutrition and disease prevention.

Keywords: artificial intelligence; probiotic; microbiome; personalized nutrition; machine learning; deep learning

The human gut microbiome, a complex ecosystem of trillions of microorganisms, plays a fundamental role in maintaining host health. It is intricately involved in critical physiological processes, including nutrient metabolism, immune system development, and protection against pathogens[1-2]. Consequently, disruptions to this delicate microbial balance lead to dysbiosis, a state characterized by alterations in microbial composition, reduced diversity, and disrupted metabolic functions[3-4]. Dysbiosis is strongly associated with a wide spectrum of diseases, ranging from gastrointestinal disorders like inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS) to metabolic conditions such as obesity and type 2 diabetes, and even neurological and cardiovascular pathologies[5-7].

In this context, probiotics—live microorganisms that confer a health benefit on the host when administered in adequate amounts—have emerged as a promising therapeutic strategy to restore eubiosis and ameliorate disease states[8]. Despite their potential, the clinical application of probiotics faces significant hurdles. A major challenge is the substantial inter-individual variability in treatment response, in which a specific probiotic strain or formulation may be highly effective in one person but completely ineffective or even detrimental in another[9]. This heterogeneity underscores the limitations of the traditional "one-size-fits-all" approach and highlights the urgent need for more precise and personalized intervention strategies.

Addressing this need requires navigating an unprecedented level of biological complexity.

Modern microbiome research generates vast and high-dimensional datasets encompassing metagenomics, metatranscriptomics, metabolomics, and host genomic and clinical metadata. The intricate, non-linear interactions within and between these multi-omics layers are beyond the capacity of conventional statistical methods to fully decipher[10]. This data deluge has created a critical bottleneck, necessitating the adoption of advanced computational frameworks capable of extracting meaningful patterns and generating testable hypotheses. Against this backdrop, the concept of "intelligent symbiosis" has emerged, referring to the deep integration of artificial intelligence (AI) algorithms with biological microecosystems. For example, interpretable machine learning (ML) models, by analyzing multi-omics data (such as metagenomic and metabolomic data), can reveal the dynamic mechanisms of host-microbe interactions, thereby driving a paradigm shift in this field from "empirical intervention" to "precision regulation"[11].

To this end, AI has become an indispensable tool. AI, broadly defined as the science of creating intelligent machines, offers powerful solutions for handling complex biological data. Within AI, ML enables algorithms to automatically learn from data and identify hidden patterns without being explicitly programmed for every rule. For instance, ML classifiers can predict a bacterial strain's immunomodulatory potential based on its genomic features[12]. An even more potent subset, deep learning (DL), utilizes artificial neural networks with multiple layers to model highly complex, non-linear relationships. DL models have been successfully applied to integrate diverse data types, such as metagenomic sequencing and host metabolomics, to uncover novel host-microbe interactions and to predict personalized responses to dietary or probiotic interventions[13].

Previous research by Shi et al.[14-15] has shown that specific bacterial strains hold significant potential for improving cognitive impairment. Based on the results of our microbiome data analysis project, traditional screening methods have reached a bottleneck, and there is an urgent need to incorporate artificial intelligence technologies to enable systematic and critical evaluation. Therefore, this review aims to comprehensively examine the current landscape of AI integration in probiotic science and personalized nutrition. We will first explore how AI is revolutionizing the analysis of the gut microbiome itself. We will then critically evaluate its applications in the discovery and functional prediction of next-generation probiotics. Finally, we will discuss the design and implementation of AI-driven personalized nutritional and probiotic intervention strategies.

1 Fundamentals of the application of artificial intelligence in microecological research

1.1 AI Methods for high-dimensional microbiome data

Microbiome data derived from technologies such as 16S rRNA gene sequencing, metagenomic shotgun sequencing, macrotranscriptomics, macroproteomics, and metabolomics are inherently highly complex[16-18]. AI/ML algorithms are particularly well-suited to handle these data features. They are capable of handling massive amounts of data, identifying hidden patterns, performing feature extraction/selection, reducing data dimensionality, and modeling complex nonlinear interactions that may be ignored by traditional methods[19]. AI also helps to integrate multi-omics data for a more comprehensive system-level understanding[20]. Traditional approaches often fail to predict medium- and long-term outcomes with sufficient accuracy, resulting in suboptimal patient management (Figure 1). In contrast, AI-driven personalized nutrition integrates multidimensional data—including metagenomic, metabolomic, phenotypic, lifestyle, and genetic information—to provide individuals with tailored health plans, enabling precise interventions such as improving cardiovascular function, enhancing

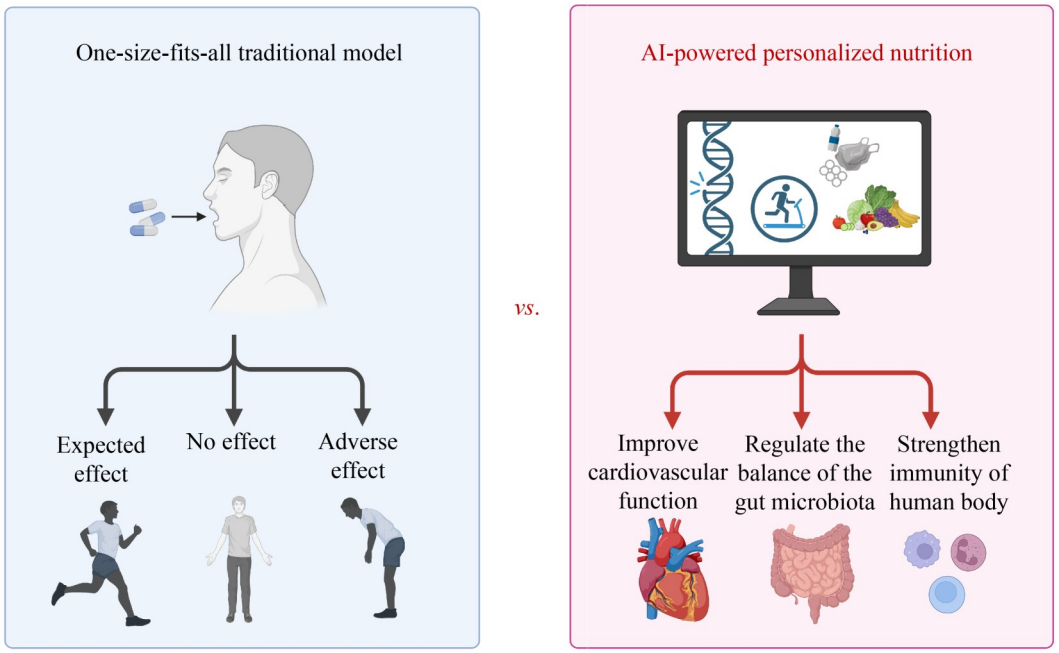


Figure 1 Comparison between traditional one-size-fits-all nutrition and AI-powered personalized nutrition. The left side illustrates how standard dietary interventions typically yield varying outcomes: some individuals have the expected effect, others show no effect, and some experience adverse effects due to inter-individual differences. The right side demonstrates how an AI-driven model integrates multi-omics and lifestyle data to generate personalized nutritional recommendations, thereby producing consistent positive health outcomes such as improved cardiovascular function, enhanced sensory perception, and strengthened immunity across diverse individuals. AI: Artificial intelligence; *vs.*: Versus.

sensory perception, and strengthening the immune system (Figure 1).

Current research widely employs various artificial intelligence architectures, with their selection highly dependent on data types and scientific problems (Table 1). In supervised learning tasks, classical machine learning models such as random forests (RF), support vector machines (SVM), and gradient boosting trees (GBDT) demonstrate robustness in small-to-medium sample scenarios and are frequently used for disease risk prediction, biomarker identification, or patient stratification[25-26]. However, when the sample size is significantly smaller than the feature dimension ($n<p$), such as in high-dimensional metagenomic data, these models are prone to overfitting[26-27]. Random forests may even fail due to feature redundancy. For sequence or image-type data, deep learning methods demonstrate superior expressive power: convolutional neural networks (CNNs) can automatically extract discriminative features from sequence patterns like k-mers[28]. Tools like DeepMicrobes significantly enhance metagenomic species classification accuracy, though their performance heavily relies on large-scale and high-quality annotated data[29]. Recurrent neural networks (RNNs) and their variant long short-term memory (LSTM) networks excel at modeling temporal dynamics in the microbiome, though they similarly require intensive longitudinal sampling or sufficiently high sampling rate[30].

Table 1 Classification, input data types, typical applications, and performance characteristics of artificial intelligence and machine learning methods in human gut microbiome research

AI/ML approach	Functions	Applications in probiotic research	Example tools	Advantages	Limitations	References
Classical machine learning	Learns patterns from data to classify samples or select important features	Disease risk prediction; identifying microbial biomarkers; patient stratification	RF, SVM, LASSO	It requires minimal training data and offers high interpretability, making it suitable for clinical data with limited samples	When dealing with extremely high-dimensional data ($n<p$), overfitting is common, and it is difficult to capture complex nonlinear relationships	[7,21]
Deep learning	Automatically extracts complex patterns from large datasets	Taxonomic profiling from sequences; predicting host-microbe interactions; functional annotation	CNN, LSTM, transformer	Automatic feature extraction; capable of handling large-scale data; suitable for images and time series	"Black-box" models lack interpretability; they require vast amounts of labeled data; and they consume significant computational resources	[22]
Generative models	Creates synthetic data to address data limitations	Augmenting small datasets; correcting batch effects; imputing missing values	VAE, GAN	Addressing the issue of sample imbalance allows for the generation of simulated data to test hypotheses	Verifying the authenticity of the generated data is challenging, and the training process is unstable	[23]
Graph neural networks	Models relationships and networks between entities	Mapping microbial interaction networks; integrating multi-omics data	GNN, GraphMB	It addresses the issue of sample imbalance and can generate simulated data to test hypotheses	It is difficult to verify the authenticity of the generated data, and the training process is unstable	[24]
Probiotic-specific tools	Purpose-built tools for probiotic discovery	Screening probiotic candidates from metagenomic data; predicting metabolite production	MetaProbiotics, MMETHANE, MelonnPan, MGM	MGM demonstrates exceptional accuracy and strong generalization capabilities in traceability tasks; MelonnPan fills gaps in metabolomics data; and metaProbiotics precisely identifies the functions of probiotics	They are typically designed for specific tasks and are less versatile than general-purpose algorithm frameworks; some tools (such as MGM) require significant computational resources or offer limited interpretability	[23-25]

Data in this table are presented descriptively; No statistical analyses were performed as the content summarizes methodological categories rather than empirical results. RF: Random forest; SVM: Support vector machine; LASSO: Least absolute shrinkage and selection operator; CNN: Convolutional neural network; LSTM: Long short-term memory recurrent neural network; VAE: Variational autoencoder; GAN: Generative adversarial network; GNN: Graph neural network; GraphMB: Graph-based microbiome biomarker detection; MetaProbiotics: Tool for probiotic candidate discovery from metagenomes; MMETHANE: Metagenomic method for efficient taxonomic and functional hierarchy annotation and exploration; MelonnPan: Metabolite prediction from microbiome profiles using neural networks; MGM: Microbial genome mining (a foundation model for microbiome analysis). This table provides a non-exhaustive overview; specific implementations, input data requirements (e.g., 16S rRNA gene sequencing, shotgun metagenomics, metabolomics), and model performance [e.g., accuracy, area under the curve (AUC)] vary substantially across studies depending on cohort size, data quality, and biological context.

In unsupervised or semi-supervised scenarios, autoencoders (AE) and their probabilistic extension—variational autoencoders (VAE), are employed for data denoising, dimensionality reduction, and clustering[22]. For modeling complex systems, graph neural networks (GNNs) effectively characterize microbial ecological interactions or integrate multi-omics heterogeneous data, demonstrating unique advantages in deciphering microbial network structures and host associations. However, current AI models face a core bottleneck: the lack of rigorous external validation[31]. Numerous studies evaluate performance solely within a single cohort, casting doubt on the models' generalization capabilities. For instance, machine learning models trained on dermatological datasets primarily composed of white patients typically achieve only 5%−10% representativeness for black patients[32]. To address this issue, the field urgently needs to establish standardized benchmark frameworks and explore systematic solutions such as federated learning. The former can systematically identify and correct model bias by establishing dynamic evaluation benchmarks that incorporate diverse racial and geographic characteristics; the latter can integrate heterogeneous data to train more robust and equitable models while protecting the privacy of data from multiple centers, thereby truly enhancing the clinical translation potential of artificial intelligence.

1.2 Computational platforms and software for AI-driven microbiome analysis

Current microbiome analysis tools span the entire workflow from sequence annotation [e.g., Kraken, quantitative insights into microbial ecology (QIIME 2)] to machine learning modeling [e.g., statistical inference of associations between microbial communities and host phenotypes (SIAMCAT), mikropml][33-34]. Emerging foundational models—such as microbial genome mining (MGM)[24] and BigRNA[35] based on large-scale pre-training—further drive paradigm shifts by enabling transfer learning for diverse downstream tasks (Table 1). However, the highly fragmented tool ecosystem results in poor comparability between methods, inconsistent evaluation metrics, and a lack of unified best practice guidelines. To enhance reproducibility and practicality, there is an urgent need to develop integrated, user-friendly analysis platforms and establish community-driven standardized benchmarking systems. These systems should enable systematic comparison of algorithm performance and guide method selection.

Microbiome data presents three core mathematical challenges: high dimensionality (far more features than samples), sparsity (a large number of zero values), and combinatoriality (complex interactions among microbial communities). General-purpose models (such as standard RF or SVM) often misinterpret "zero" as "no signal", overlooking the distinction between "detection limits" and "biological absence", and struggle to capture strong and nonlinear combinatorial effects[36]. In contrast, specialized tools are designed with greater specificity. For example, PhyloMix incorporates phylogenetic tree information as prior knowledge, using evolutionary relationships to fill gaps in sparse data and compress high-dimensional data into a low-dimensional manifold[37-38]; models such as MMETHANE employ specialized zero-value handling mechanisms (e.g., hurdle models or zero-inflation models) to distinguish technical noise from true biological zeros, thereby more accurately resolving causal relationships between community structure and function[36]. This design philosophy of "domain-knowledge-embedded algorithms" is the key reason why specialized tools outperform general-purpose models.

2 Applications of AI in probiotic discovery, microbiome analysis, and personalized nutrition

2.1 Accelerating discovery: using AI to identify novel probiotic strains and their functions

Traditional probiotic discovery methods

usually rely on isolating strains from known beneficial genera (e.g., *Lactobacillus* spp., *Bifidobacterium* spp.) and testing them *in vitro* or *in vivo*, a process that is time-consuming and laborious, and may miss many promising novel candidate strains[39]. Artificial intelligence/machine learning offers the possibility of systematically mining large-scale metagenomic and genomic data to enable the identification of potentially novel probiotic strains or microorganisms with probiotic functions without relying on traditional culture methods[25]. For example, the metaProbiotics tool utilizes natural language processing models and random forest algorithms to enable rapid detection of potential probiotics directly from macroeconomic bins (metagenomic bins)[25]. A study achieved a prediction accuracy of 96.88% by identifying *Lactobacillus* using both unsupervised and supervised artificial neural networks[40]. However, this model is based solely on *in vitro* isolates and has not been validated *in vivo*, raising concerns about its physiological relevance. These approaches overcome the limitations of traditional taxonomy and cultivation, providing a viable pathway for developing next-generation precision live biotherapeutic products (LBPs)[41] targeting specific host functional defects such as compromised intestinal barriers and immune dysregulation.

AI can also be used to analyze massive amounts of data to link microbial characteristics (such as taxonomic units, genes, and metabolites) with host phenotypes (such as health status, disease risk, and response to interventions) to pinpoint microbes associated with positive health outcomes and use them as potential probiotic candidates[19]. AI tools such as metaProbiotics[25] and methods described in National Institutes of Health (NIH) funded programs[13,36] aim to mine precisely this metagenomic data for new strains associated with probiotic functions. Additionally, researchers integrated gut microbiome characteristics from obese individuals to construct a predictive model for screening potential anti-obesity strains. They validated its ability to reduce fat accumulation in mouse models, though its efficacy in humans remains unknown[42]. This approach exemplifies the current typical technical pathway for microbiome AI—including RF for feature selection, GNNs for modeling gene adjacency relationships, and CNNs for processing sequence k-mers—which has been widely applied in classification, functional prediction, and dynamic modeling[43].

2.2 Predictive function: using AI models to predict probiotic efficacy and mechanism of action

AI models can establish quantitative mappings from strain genomes or phenotypic traits to their *in vivo* functional outputs, significantly reducing reliance on costly, low-throughput animal experiments. Existing research demonstrates that machine learning models can effectively predict the immune-enhancing effects of *Lactobacillus* strains in snail infection models based on easily measurable *in vitro* parameters such as surface hydrophobicity and self-aggregation capacity[44]. Another study trained ML models using RNA-seq dynamic data to successfully identify key cultivation conditions that activate *Lactobacillus reuteri* to produce beneficial metabolites like histamine and lactic acid[45]. In a study investigating probiotic mixture interventions, multiple machine learning-based classification models demonstrated 100% accuracy in predicting inflammatory biomarkers. Their results showed high consistency with microbial biomarkers identified by linear discriminant analysis effect size (LEfSe) and with *Bifidobacterium* species within the probiotic formulations[46]. Collectively, these cases demonstrate AI's core value: translating measurable *in vitro* characteristics into reliable predictions of complex *in vivo* efficacy, thereby enabling early-stage prioritization of high-potential candidate strains and optimizing research and development pipelines.

2.3 Decoding interactions: using AI to understand host-microbe and inter-microbe dynamics

The individual variability in probiotic efficacy stems from the highly dynamic and nonlinear interactive network among probiotics, indigenous microbiota, and the host. By integrating multi-omics data, AI is progressively revealing the underlying mechanisms of this "tripartite dialogue". Further integration of individual dietary exposure data (such as using machine learning to automatically quantify processing levels from food labels or ingredients[47]) can more accurately simulate the gut microenvironment, thereby enhancing the precision of translating *in vitro* characteristics into *in vivo* efficacy. Explainable models like MMETHANE[36] further explicitly identified specific bacteria-metabolite pairs (e.g., *Faecalibacterium*-butyrate) significantly correlated with host inflammatory states. Additional research combined GNN with temporal models (LSTM) to simulate the dynamic restructuring of microbiota-host interaction networks under antibiotic or dietary perturbations[22]. However, such advanced models often rely on invasive samples like intestinal mucosal biopsies or tissue transcriptomes, severely limiting their application in routine clinical settings. The future requires balancing performance and feasibility while relying on robust computational infrastructure to support the entire chain from multi-source data collection to personalized intervention generation—this is key to achieving the closed-loop transition "from data to nutrition"[48].

2.4 Personalized diagnosis: disease prediction and stratification using microbiome characteristics

The human microbiome exhibits significant inter-individual heterogeneity, making disease biomarkers based on population-average effects difficult to generalize. AI/ML technologies excel at identifying individual-specific microbial characteristics (such as specific microbial communities, functional genes, or metabolite combinations) from high-dimensional metagenomic and metabolomic data for disease risk prediction, early diagnosis, or patient stratification (Table 2 and Figure 2). For instance, a study combining unsupervised clustering with machine learning classified short-chain fatty acid (SCFA)-associated microbial subpopulations, revealing their composition and metabolic potential. It confirmed that the random forest model demonstrated optimal performance in binary classification (91.05% for agglomerative receiver operating characteristic (ROC), 87.74% for k-means), highlighting its robust predictive capability within SCFA-related clusters[59]. Another study proposed MicroAI biome—a machine learning-based AI workflow designed to overcome challenges such as high dimensionality, structural complexity, and poor model interpretability in microbiome data, enabling classification across five cancer types; Among the algorithms tested, XGBoost demonstrated optimal performance (78.23% accuracy), and Shapley additive Explanations (SHAP) analysis revealed category-specific microbial biomarkers[60]. Additionally, a study integrating five 16S rRNA gene datasets from North America and East Asia identified 109 discriminative microbial taxa and constructed a random forest model that effectively distinguished healthy, adenomatous, and cyclic redundancy check (CRC) conditions (internal AUC=0.90, external AUC=0.82). This demonstrates the potential of combining the gut microbiome with machine learning for early detection across populations, though clinical validation remains necessary[61]. Overall, while current AI-driven microbiome diagnostic tools offer a non-invasive window into host physiological states, their clinical translation remains constrained by insufficient generalization, high sample invasiveness, and a lack of dynamic data, rendering them unsuitable for large-scale deployment.

Table 2 AI/ML applications in human microbiome studies with health-related outcomes

Population/ sample size	AI/ML method	Intervention/target outcome	Key findings	Performance metrics	References
254 children (111 ASD, 143 TD)	XGBoost, RF, DT+SHAP	Identification of ASD-associated microbial biomarkers	Personalized microbial signatures distinguish ASD from typically developing controls	AU-ROC= 0.965±0.005; AU-PRC=0.967±0.008	[49]
78 adults	Microbiome-guided formulation	Consti-Biome (constipation) or Sensi-Biome (diarrhea) for GI symptom relief	Enrichment of *Lachnospiraceae* and *Lactobacillaceae*; modulation of pain/inflammation-related taxa	Clinically meaningful symptom improvement+ microbiota shifts	[50]
442 fecal samples	XGBoost, RF, SVM+SHAP	CRC biomarker discovery	Discriminative microbial taxa identified for CRC detection	Best model (RF): accuracy=0.729±0.038; AU-PRC=0.668±0.016	[51]
79 individuals (49 glioma, 30 controls)	Gradient boosting machine (GBM)	Predict glioma status from gut microbiome	Gut microbial profile significantly associated with brain tumor presence	AUC-ROC=0.79	[52]
1 429 subjects (729 IBD, 700 controls)	RF, DT, Elastic net, SVM, Neural networks	Classify IBD subtypes from 16S rRNA data	High-variance OTUs enable robust classification	AUC>0.90 on held-out test set	[53]
130 melanoma patients	DeepGeni (deep generalized interpretable autoencoder)	Predict response to immune checkpoint inhibitors	Identified microbial features predictive of treatment response; potential targets for FMT/probiotics	Improved interpretability *via* latent encoding	[54]
147 CRC patients	RF, MLP	Predict lymph node metastasis (LNM) from microbiome	*Bacteroides plebeius* linked to LNM *via* immune/metabolic pathways	Good predictive performance (metric not quantified)	[55]
2 698 samples across 13 public datasets	MicroHDF (microbiome-based hierarchical deep forest)	Host phenotype prediction from metagenomics	Effectively handles class imbalance and high dimensionality	AU-ROC= 0.918 2±0.009 8; AU-PRC= 0.946 9±0.007 6	[56]
1 043 patients with 10 autoimmune diseases (AIDs)	RF, SVM, KNN, MLP, XGBoost	Discriminate autoimmune disease types based on microbiome	Disease-specific microbial signatures reflect underlying immune dysregulation	XGBoost AU-ROC: 0.75−0.99; sensitivity: 0.66−1.00; specificity: 0.70−0.96	[57]
137 patients	XGBoost, RF+SHAP	Identify fibrosis-associated microbial signatures	Gut taxa linked to liver fibrosis progression	SHAP value=0.128 for top predictive feature	[58]

Data are presented as reported in the original studies; no pooled statistical analysis was performed in this summary table. Continuous variables (e.g., age) are shown as ranges; categorical variables (e.g., sex, diagnosis) are reported as counts. AI: Artificial intelligence; ML: Machine learning; ASD: Autism spectrum disorder; TD: Typically developing; XGBoost: Extreme gradient boosting; RF: Random forest; DT: Decision tree; SHAP: Shapley additive explanations; GI: Gastrointestinal; CRC: Colorectal cancer; SVM: Support vector machine; GBM: Gradient boosting machine; IBD: Inflammatory bowel disease; OTU: Operational taxonomic unit; DeepGeni: Deep generalized interpretable autoencoder; FMT: Fecal microbiota transplantation; MLP: Multilayer perceptron; LNM: Lymph node metastasis; MicroHDF: Microbiome-based hierarchical deep forest; KNN: K-nearest neighbors; HBV: Hepatitis B virus; AIDs: Autoimmune diseases AU-ROC: Area under the receiver operating characteristic curve; AU-PRC: Area under the precision-recall curve; AUC: Area under the curve.

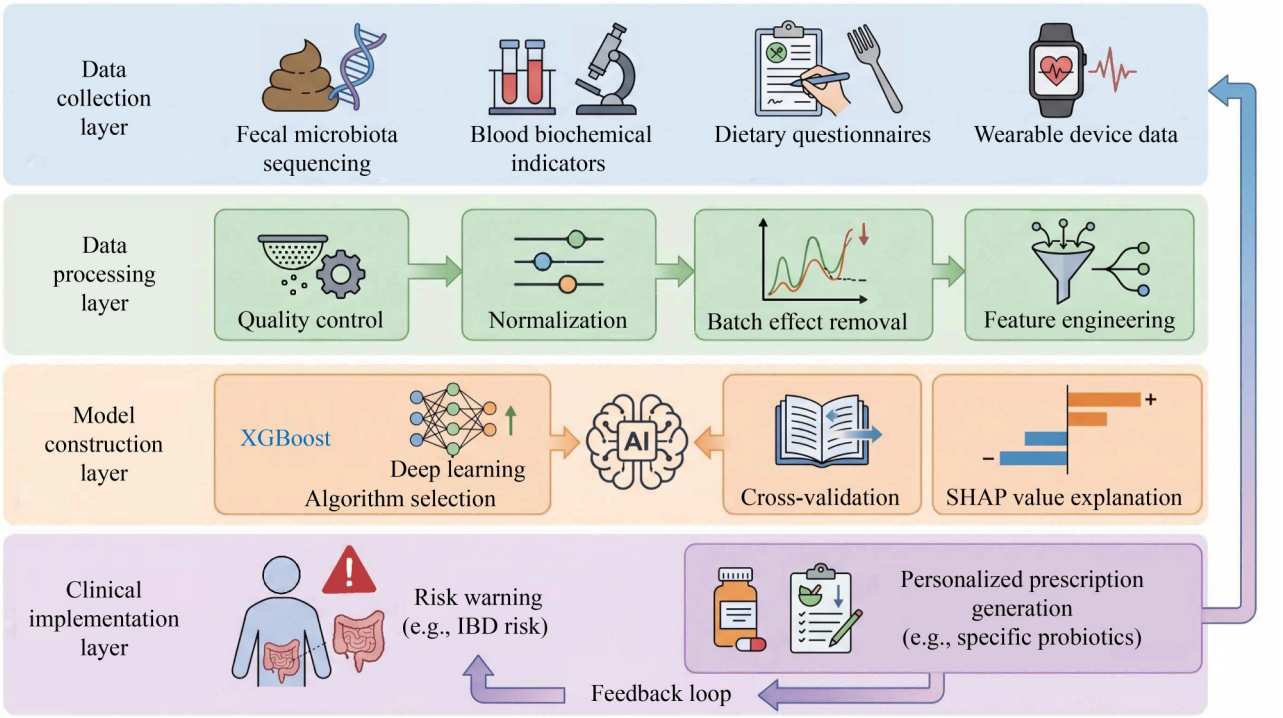


Figure 2 Schematic overview of the AI-driven personalized nutrition intervention framework. The framework operates through a four-layer architecture: The data collection layer integrates multi-source inputs, including fecal microbiota sequencing, blood biochemical indicators, dietary questionnaires, and wearable device data. The data processing layer performs essential preprocessing steps such as quality control, normalization, batch effect removal, and feature engineering. The model construction layer employs machine learning algorithms (e.g., XGBoost, Deep learning) with cross-validation and SHAP value explanation to build predictive models. Finally, the clinical implementation layer translates model outputs into actionable risk warnings (e.g., IBD risk) and personalized prescriptions (e.g., specific probiotics), establishing a continuous feedback loop. AI: Artificial intelligence; XGBoost: Extreme gradient boosting; SHAP: Shapley additive explanations; IBD: inflammatory bowel disease.

2.5 Customized interventions: using AI to design personalized probiotics and dietary strategies

One-size-fits-all probiotic or dietary interventions often yield inconsistent results due to individual microbiota variations, driving research toward AI-enabled precision matching (Figure 1). The core approach involves dynamically coupling interventions (specific probiotic strains, dietary components) with an individual's microbiota, diet, genetics, and medication history to realize the concept of precision probiotics (PPC) (Figure 3). AI serves as an "intelligent matchmaker" here: by learning intervention-response patterns from large-scale multimodal data, it predicts individual responses to specific strains or dietary protocols. For instance, a study based on data from 12 068 participants found that a high gut microbiome dietary index was associated with a 25% reduced risk of developing advanced cardiovascular-renal metabolic syndrome (each 1-point increase in risk reduced by 7%). This association was mediated by inflammatory biomarkers (contributing 5.59%−42.88%). XGBoost combined with SHAP further identified coffee,

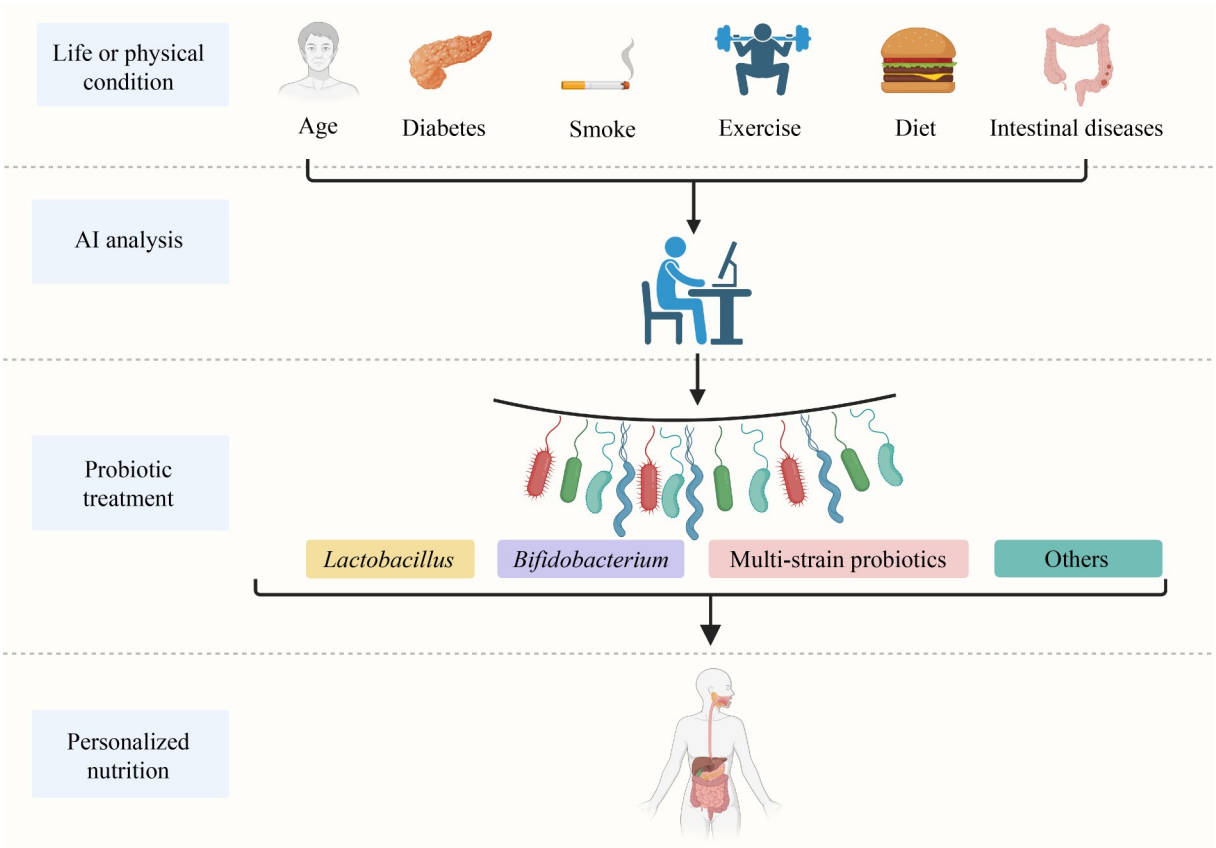


Figure 3 AI identifies host context for probiotic personalized nutrition by analyzing individual data or predicting responses. Based on individual factors such as age, diabetes, smoking, exercise, diet, and intestinal diseases, probiotic treatment plans (such as *Lactobacillus*, *Bifidobacterium*, multi-strain probiotics, etc.) are formulated through AI analysis to achieve precise personalized nutrition intervention and improve intestinal health.

refined grains, and fat as key risk predictors[62]. Based on 8 677 metagenomic samples from 65 studies, Yang et al. developed the nonlinear gut microbiome health predictor (NGMHP). By employing kernel principal component analysis (KPCA) to capture complex relationships between the microbiome and health, the model achieved an accuracy of 0.791 4 and an AUC of 0.860 8 in health/disease classification, demonstrating potential for health monitoring[63].

However, these static models face two limitations: first, their reliance on static data limits their ability to reflect gut plasticity under dynamic perturbations such as diet or medication[63]; second, the lack of a procedural model that directly links "interventions" with "features" makes it difficult to achieve true dynamic regulation. To bridge this gap, future work must combine continuous monitoring data (e.g., from wearable devices and repeated sampling) with advanced AI paradigms. Specifically, reinforcement learning must be introduced to optimize dynamic intervention strategies—treating probiotic interventions as a sequential decision-making process that dynamically fine-tunes intervention protocols based on real-time host feedback (such as continuous metabolomic or microbiome profiles). Concurrently, recommendation algorithms should be integrated to generate actionable personalized plans, while causal inference is utilized to identify

effective targets. This "AI nutrition matchmaker" which integrates time-series models (such as LSTMs) with reinforcement learning, will evolve intervention strategies from "static prescriptions" into adaptive navigation systems capable of adapting to fluctuations in host physiology. This will not only enhance the timeliness and robustness of intervention recommendations but also truly establish a closed-loop "monitoring-intervention" system, driving the clinical application of AI-driven, personalized nutrition.

Although AI-driven personalized probiotics (such as PPC) demonstrate significant clinical prospects in translating personalized nutrition into routine medical practice, it is essential to establish a prudent ethical and regulatory framework. First, personalized nutrition relies on highly sensitive personal data, such as whole-genome and metabolomic data; therefore, mechanisms for data de-identification and encryption that comply with relevant standards must be established to effectively safeguard data privacy and security. Second, for vulnerable populations such as the elderly and immunocompromised individuals, probiotic formulations recommended by artificial intelligence may pose potential unknown risks; regulatory authorities should establish a dedicated approval pathway for live biotherapeutic products (LBPs) and rigorously evaluate the safety and colonization capacity of bacterial strains. Therefore, to ensure the universality and fairness of precision nutrition services, future research should focus on incorporating data from multicenter, diverse populations to continuously optimize the applicability and equity of the models.

3 Conclusion

The past decade has witnessed a deep convergence of microbiome science and artificial intelligence. Research has clarified the fundamental role of the microbiome in sustaining human health, and the proliferation of high-throughput sequencing technologies has led to an explosion of data. AI/ML technologies have become a central driver in analyzing this complex data, with significant progress in biomarker discovery, disease prediction, exploration of functional mechanisms, and initial attempts at personalized interventions. The field as a whole is progressively moving from characterizing correlations to exploring causality and understanding molecular mechanisms. It is expected that larger and higher-quality datasets, combined with more advanced AI models (e.g., base models, multimodal AI) and increasing computational power, will drive greater breakthroughs in research. The integration of the microbiome with AI is driving personalized nutrition from a therapeutic approach toward proactive prevention. However, achieving reliable clinical applications requires overcoming bottlenecks related to the standardization of multi-omics data, model interpretability, and generalizability, as well as establishing a robust ethical and regulatory framework to ensure that the technology is safe, equitable, and effective. The focus of research will be more inclined toward mechanism modeling, causal inference, and the development of truly personalized, dynamically adapted intervention strategies.

Credit authorship contribution statement

SHI Fengcui: Conceptualization, writing—original draft, writing—review and editing, supervision, project administration, funding acquisition; LI Zeyang: Software, formal analysis; JING Yifan: Software, investigation; LI Yongshi: Validation, visualization; MA Wenjian: Resources, writing—review and editing, project administration, funding acquisition.

Declaration of Competing Interest

The authors declare no competing financial interest.

References

[1] Ahn J, Hayes RB. Environmental influences on the human microbiome and implications for noncommunicable disease[J]. Annual Review of Public Health, 2021, 42: 277-292.

[2] Khalil M, Di Ciaula A, Mahdi L, Jaber N, Di Palo DM, Graziani A, Baffy G, Portincasa P. Unraveling the role of the human gut microbiome in health and diseases[J].

Microorganisms, 2024, 12(11): 2333.
[3] He RS, Qi PQ, Shu LZ, Ding YD, Zeng P, Wen GS, Xiong Y, Deng H. Dysbiosis and extraintestinal cancers[J]. Journal of Experimental & Clinical Cancer Research, 2025, 44: 44.
[4] Zhang YJ, Qing JB, Saed YA, Li YF. Gut microbiota implication in diabetic kidney disease: mechanisms and novel therapeutic strategies[J]. Renal Failure, 2025, 47: 2517402.
[5] Chandrasekaran P, Weiskirchen S, Weiskirchen R. Effects of probiotics on gut microbiota: an overview[J]. International Journal of Molecular Sciences, 2024, 25(11): 6022.
[6] Mazziotta C, Tognon M, Martini F, Torreggiani E, Rotondo JC. Probiotics mechanism of action on immune cells and beneficial effects on human health[J]. Cells, 2023, 12(1): 184.
[7] Aryal S, Alimadadi A, Manandhar I, Joe B, Cheng X. Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease[J]. Hypertension, 2020, 76(5): 1555-1562.
[8] Hsieh MH, Versalovic J. The human microbiome and probiotics: implications for pediatrics[J]. Current Problems in Pediatric and Adolescent Health Care, 2008, 38(10): 309-327.
[9] Yadav M, Shukla P. Recent systems biology approaches for probiotics use in health aspects: a review[J]. 3 Biotech, 2019, 9(12): 448.
[10] Eetemadi A, Rai N, Pereira BMP, Kim M, Schmitz H, Tagkopoulos I. The computational diet: a review of computational methods across diet, microbiome, and health[J]. Frontiers in Microbiology, 2020, 11: 393.
[11] Laccourreye P, Bielza C, Larrañaga P. Explainable machine learning for longitudinal multi-omic microbiome[J]. Mathematics, 2022, 10(12): 1994.
[12] Charizani E, Dushku E, Kyritsi M, Metallinou ET, Karathodorou A, Amanetidou E, Kokkaleniou MM, Passalis N, Tefas A, Staikou A, Yiangou M. Predicting the immunomodulatory activity of probiotic lactic acid bacteria using supervised machine learning in a *Cornu aspersum* snail model[J]. Fish & Shellfish Immunology, 2024, 152: 109788.
[13] Wang SH, Song SN, Gao JX, Wu WM, Fu Y, Yuan T, Zhao WG. Dynamic prediction of postprandial glycemic response and personalized dietary interventions based on machine learning[J]. The Journal of Nutrition, 2025, 155(12): 4193-4208.
[14] Shi FC, Peng JW, Li HJ, Liu DH, Han L, Wang Y, Liu QL, Liu Q. Probiotics as a targeted intervention in anti-ageing: a review[J]. Biomarkers, 2024, 29(8): 577-585.
[15] Zhou F, Zhang QR, Zheng XH, Shi FC, Ma K, Ji F, Meng N, Li RD, Lv JW, Li QY. Antiaging effects of human fecal transplants with different combinations of *Bifidobacterium bifidum* LTBB21J1 and *Lactobacillus casei* LTL1361 in D-galactose-induced mice[J]. Journal of Agricultural and Food Chemistry, 2024, 72(17): 9818-9827.
[16] Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis[J]. Briefings in Bioinformatics, 2021, 22(1): 178-193.
[17] Deek RA, Ma SY, Lewis J, Li HZ. Statistical and computational methods for integrating microbiome, host genomics, and metabolomics data[J]. eLife, 2024, 13: e88956.
[18] Galloway-Peña J, Hanson B. Tools for analysis of the microbiome[J]. Digestive Diseases and Sciences, 2020, 65(3): 674-685.
[19] Majumder S, Sen K, Karanjai R. Artificial intelligence-based target for personalized interventions of atherosclerosis from gut microbiota signature[J]. SynBio, 2025, 3(1): 2.
[20] Abdelwahab SI, Taha MME, Ali Jerah A, Farasani A, Abdullah SM, Aljahdali IA, Ibrahim R, Oraibi O, Oraibi B, Ahmad Alfaifi H, Alzahrani AH, Babiker YOH. Artificial intelligence and microbiome research: evolution of hotspots, research trends, and thematic-based narrative review[J]. Cellular and Molecular Biology, 2024, 70(10): 182-192.
[21] Xiao LW, Zhao FQ. Exploring the frontier of microbiome biomarker discovery with artificial intelligence[J]. National Science Review, 2024, 11(11): nwae325.
[22] Przymus P, Rykaczewski K, Martín-Segura A, Truu J, Carrillo De Santa Pau E, Kolev M, Naskinova I, Gruca A, Sampri A, Frohme M, Nechyporenko A. Deep learning in microbiome analysis: a comprehensive review of neural network models[J]. Frontiers in Microbiology, 2025, 15: 1516667.
[23] Wu SF, Feng T, Tang WJ, Qi CC, Gao J, He XL, Wang JX, Zhou HW, Fang ZC. metaProbiotics: a tool for mining probiotic from metagenomic binning data based on a language model[J]. Briefings in Bioinformatics, 2024, 25(2): bbae085.
[24] Zhang HH, Zhang YL, Kang ZX, Song LL, Yang RH, Ning K. MGM as a large-scale pretrained foundation model for microbiome analyses in diverse contexts[EB/OL]. bioRxiv (2025-01-01). https://doi.org/10.1101/2024.12.30.630825.
[25] Sun Q, Ma MQ, Du X, Wan X, Cao CC, Liang XB. Explainable machine learning models for early prediction of acute kidney injury after cardiac surgery[J]. International Urology and Nephrology, 2026, 58(4): 1499-1509.
[26] Ahmad Ansari G, Shafi S, Ansari MD, Shadab A. Advanced supervised machine learning methods for precise diabetes mellitus prediction using feature selection[J]. Frontiers in Medicine, 2025, 12: 1620268.
[27] Babai AMA, Ehinola OA, Gebbayin OIMFA, Ibrahim MAE. Clastic facies classification using machine learning-based algorithms: a case study from Rawat Basin, Sudan[J]. Energy Geoscience, 2025, 6(1): 100353.
[28] Yu YW. On minimizers and convolutional filters: theoretical connections and applications to genome analysis[J]. Journal of Computational Biology, 2024, 31(5): 381-395.
[29] Wichmann A, Buschong E, Müller A, Jünger D, Hildebrandt A, Hankeln T, Schmidt B. MetaTransformer: deep metagenomic sequencing read classification using

self-attention models[J]. NAR Genomics and Bioinformatics, 2023, 5(3): lqad082.
[30] Qamar WUR, Lee MH, Abibullaev B. Deep learning in intracranial EEG for seizure detection: advances, challenges, and clinical applications[J]. Frontiers in Neuroscience, 2025, 19: 1677898.
[31] Kong FH, Su MK, Guo JF, Wu HY, Chen HB, Qiu J, Huang JC. Implementation of convolutional neural networks for microbial colony recognition[J]. Microbiology Spectrum, 2025, 13(9): e02885-24.
[32] Iacucci M, Maeda Y, Ghosh S. A baby step or a real giant stride: histomic enabled by artificial intelligence to predict treatment response in pediatric patients with ulcerative colitis[J]. Gastroenterology, 2024, 166(5): 730-732.
[33] Marcos-Zambrano LJ, López-Molina VM, Bakir-Gungor B, Frohme M, Karaduzovic-Hadziabdic K, Klammsteiner T, Ibrahimi E, Lahti L, Loncar-Turukalo T, Dhamo X, Simeon A, Nechyporenko A, Pio G, Przymus P, Sampri A, Trajkovik V, Lacruz-Pleguezuelos B, Aasmets O, Araujo R, Anagnostopoulos I, et al. A toolbox of machine learning software to support microbiome analysis[J]. Frontiers in Microbiology, 2023, 14: 1250806.
[34] Miossec MJ, Valenzuela SL, Pérez-Losada M, Johnson WE, Crandall KA, Castro-Nallar E. Evaluation of computational methods for human microbiome analysis using simulated data[J]. PeerJ, 2020, 8: e9688.
[35] Leung MKK, Xiong HY, Lee LJ, Frey BJ. Deep learning of the tissue-regulated splicing code[J]. Bioinformatics, 2014, 30(12): i121-i129.
[36] Dawkins JJ, Gerber GK. MMETHANE: interpretable AI for predicting host status from microbial composition and metabolomics data[J]. Microbiome, 2026, 14: 21.
[37] Piraino P, Ricciardi A, Salzano G, Zotta T, Parente E. Use of unsupervised and supervised artificial neural networks for the identification of lactic acid bacteria on the basis of SDS-PAGE patterns of whole cell proteins[J]. Journal of Microbiological Methods, 2006, 66(2): 336-346.
[38] Maftei NM, Raileanu CR, Balta AA, Ambrose L, Boev M, Marin DB, Lisa EL. The potential impact of probiotics on human health: an update on their health-promoting properties[J]. Microorganisms, 2024, 12(2): 234.
[39] Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, Berland M, Gruca A, Hasic J, Hron K, Klammsteiner T, Kolev M, Lahti L, Lopes MB, Moreno V, Naskinova I, Org E, Paciência I, Papoutsoglou G, Shigdel R, et al. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment[J]. Frontiers in Microbiology, 2021, 12: 634511.
[40] Jiang YF, Liao DS, Zhu QY, Lu YY. PhyloMix: enhancing microbiome-trait association prediction through phylogeny-mixing augmentation[EB/OL]. BioRxiv (2024-12-15). https://doi.org/10.1101/2024.08.26.609661.
[41] McCoubrey LE, Elbadawi M, Orlu M, Gaisford S, Ashford FL, Basit AW. Harnessing machine learning for development of microbiome therapeutics[J]. Gut Microbes, 2021, 13: 1872323.
[42] Sulaimany S, Farahmandi K, Mafakheri A. Computational prediction of new therapeutic effects of probiotics[J]. Scientific Reports, 2024, 14: 11932.
[43] Hernández Medina R, Kutuzova S, Nielsen KN, Johansen J, Hansen LH, Nielsen M, Rasmussen S. Machine learning and deep learning applications in microbiome research[J]. ISME Communications, 2022, 2: 98.
[44] Zhai YT, Kim M, Fan PX, Rajeev S, Kim SA, Driver JD, Galvão KN, Boucher C, Jeong KC. Machine learning-enhanced assessment of potential probiotics from healthy calves for the treatment of neonatal calf diarrhea[J]. Frontiers in Microbiology, 2024, 15: 1507537.
[45] Wu H, Li Y, Jiang YX, Li XR, Wang SL, Zhao CL, Yang XM, Chang BC, Yang JH, Qiao JJ. Machine learning prediction of obesity-associated gut microbiota: identifying *Bifidobacterium pseudocatenulatum* as a potential therapeutic target[J]. Frontiers in Microbiology, 2025, 15: 1488656.
[46] Laterza L, Putignani L, Settanni CR, Petito V, Varca S, De Maio F, Macari G, Guarrasi V, Gremese E, Tolusso B, Wlderk G, Pirro MA, Fanali C, Scaldaferri F, Turchini L, Amatucci V, Sanguinetti M, Gasbarrini A. Ecology and machine learning-based classification models of gut microbiota and inflammatory markers may evaluate the effects of probiotic supplementation in patients recently recovered from COVID-19[J]. International Journal of Molecular Sciences, 2023, 24(7): 6623.
[47] Menichetti G, Ravandi B, Mozaffarian D, Barabási AL. Machine learning prediction of the degree of food processing[J]. Nature Communications, 2023, 14: 2312.
[48] Di Bitonto P, Magarelli M, Novielli P, Romano D, Diacono D, de Trizio L, Mariano A, Zoani C, Ferrero R, Manzin A, De Angelis M, Bellotti R, Tangaro S. From data to nutrition: the impact of computing infrastructure and artificial intelligence[J]. Exploration of Foods and Foodomics, 2024: 810-829.
[49] Novielli P, Romano D, Magarelli M, Diacono D, Monaco A, Amoroso N, Vacca M, De Angelis M, Bellotti R, Tangaro S. Personalized identification of autism-related bacteria in the gut microbiome using explainable artificial intelligence[J]. iScience, 2024, 27(9): 110709.
[50] Min U, Jin YJ, Jang YJ, Lim J, Kim BY. Personalized probiotic strategy considering bowel habits: impacts on gut microbiota composition and alleviation of gastrointestinal symptoms *via* Consti-Biome and Sensi-Biome[J]. Frontiers in Nutrition, 2024, 11: 1302093.
[51] Novielli P, Romano D, Magarelli M, Di Bitonto P, Diacono D, Chiatante A, Lopalco G, Sabella D, Venerito V, Filannino P, Bellotti R, De Angelis M, Iannone F, Tangaro S. Explainable artificial intelligence for microbiome data analysis in colorectal cancer biomarker identification[J]. Frontiers in Microbiology, 2024, 15: 1348974.
[52] Li Z, Zhao K, Liu HY, Liu JL, Chen X, Hu WT, Wen E, Zhang K, Chen L. Construction of predictive machine learning model of glioma-associated gut microbiota[J]. Brain and Behavior, 2025, 15(9): e70843.

[53] Manandhar I, Alimadadi A, Aryal S, Munroe PB, Joe B, Cheng X. Gut microbiome-based supervised machine learning for clinical diagnosis of inflammatory bowel diseases[J]. American Journal of Physiology-Gastrointestinal and Liver Physiology, 2021, 320(3): G328-G337.
[54] Oh M, Zhang LQ. DeepGeni: deep generalized interpretable autoencoder elucidates gut microbiota for better cancer immunotherapy[J]. Scientific Reports, 2023, 13: 4599.
[55] Wu YZ, Deng CG, Huang ZG, Huang YQ, Chen CB, Qin MJ, Wang Z, He FH, Liu SH, Zhong RM, Liu J, Long CY, Liu JG, Tang WZ, Huang XL. Identification and predictive machine learning models construction of gut microbiota associated with lymph node metastasis in colorectal cancer[J]. mSystems, 2025, 10(8): e00339-25.
[56] Shi K, Liu QH, Ji QR, He QS, Zhao XM. MicroHDF: predicting host phenotypes with metagenomic data using a deep forest-based framework[J]. Briefings in Bioinformatics, 2024, 25(6): bbae530.
[57] An TF, Zhang SY, Li JJ, Wang H, Chen L, Shi YR, Wang JY, Han SR, Wang RX, Wang LY, Huan ZJ, Yang RQ, Hao DS, Liu YF, Liu XH, Yuan C. Gut microbiota analysis reveals microbial signature for multi-autoimmune diseases based on machine learning model[J]. Frontiers in Microbiology, 2025, 16: 1660775.
[58] Zhu Y, Geng SY, Chen Y, Ru QJ, Zheng Y, Jiang N, Zhu FY, Zhang YS. Machine learning algorithms reveal gut microbiota signatures associated with chronic hepatitis B-related hepatic fibrosis[J]. World Journal of Gastroenterology, 2025, 31(16): 105985.
[59] Kasani PH, Yun CH, Cho KH, Jeong SJ. Neonatal gut microbiota stratification and identification of SCFA-associated microbial subgroups using unsupervised clustering and machine learning classification[J]. Frontiers in Microbiology, 2025, 16: 1668451.
[60] Rahman MM, Shokouhmand S, Rahman S, Tamzi NN, Bhatt S, Faezipour M. MicroAIbiome: decoding cancer types from microbial profiles using explainable machine learning[J]. Microorganisms, 2025, 13(9): 2210.
[61] Tsai YJ, Lyu WN, Liao NS, Chen PC, Tsai MH, Chuang EY. Gut microbiome-based machine learning model for early colorectal cancer and adenoma screening[J]. Gut Pathogens, 2025, 17: 80.
[62] Yan ZQ, Xu YF, Du XF. Gut microbiota-modulating diets attenuate cardiovascular-kidney-metabolic syndrome *via* inflammatory mediation: insights from NHANES and machine learning model[J]. Journal of Food Science, 2025, 90(10): e70634.
[63] Yang ZX, Gao XK, Wu YX, Xie HQ, Wang HC, Lu WW, Tian FW, Zhao JX, Zhang H, Chen W, Zhu JL. Nonlinear machine learning-based gut microbiota modelling for health prediction and dietary recommendation[J]. Food Bioscience, 2025, 74: 107962.