

AI EMPOWERED HUMAN MICROBIOME

N. Leela Lakshmi Kumari ;B.L.D.V. Narsaveni ; Sk.Noorsaba ,

Dr. Y. Manjulatha (HOD) Dr. V. Bhaskar Raju(Principal)

Department of Microbiology

Sri Vasavi Institute Of Pharmaceutical Science Tadepalligudem , Andhra Pradesh ,
India

students of final year B pharmacy

Abstract :

The human microbiome is a complex ecosystem consisting of bacteria, fungi, viruses, archaea, and other microorganisms that play important roles in human health and disease. Rapid advances in high-throughput sequencing technologies, including 16S rRNA sequencing and shotgun metagenomics, have generated large and complex microbiome datasets. However, the high dimensionality, sparsity, and complexity of these datasets create significant challenges for conventional analytical approaches. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), provides powerful tools for analysing microbiome data and identifying meaningful biological patterns. This review highlights the applications of AI in human microbiome research, including microbiome data analysis, dimensionality reduction, microbial classification, biomarker identification, metagenomics, metatranscriptomics, metaproteomics, and metabolomics. Various AI algorithms, such as Decision Trees, Random Forest, convolutional neural networks, and other deep learning approaches, can be used to identify disease-associated microbial signatures and predict health outcomes. AI also has potential applications in microbiome-based diagnostics and therapeutics, including disease prediction, faecal microbiota transplantation, phage therapy, personalised nutrition, and probiotic development. Furthermore, integration of AI with multi-omics data may improve understanding of host-microbiome interactions and support precision medicine. Despite its potential, challenges such as data heterogeneity, model interpretability, standardisation, reproducibility, and clinical validation remain. Overall, AI represents a promising approach for advancing human microbiome research and may contribute to the development of personalised diagnostic and therapeutic strategies.

Introduction :

The microbiome refers to the collection of micro organisms that live in a specific environment such as human body , soil or oceans. As for the human health, the microbiome includes the wide variety of microbes that live inside and on our bodies, especially in areas like the gut, skin and mouth. The Artificial intelligence(AI) provides powerful tool for analysing these complex microbiome datasets, marking a promising direction for both basic research and practical applications in microbiome science. In this review, we use a structured approach where AI represents a broad category of computational methods designed to imitate human intelligence. Machine learning (ML) is a part of AI that focuses on learning from data, while deep learning is a type of ML that uses multiple layers of neural networks. We clearly define and consistently apply these terms. We describe a shift in microbiome research from traditional statistical and alignment- based methods to AI- driven approaches, and we summarise strategies throughout the analytical process, from data preparation and modelling to interpretation. We also examine AI applications at various biological levels, from microbial communities to molecular functions, and highlight recent progress, especially large language models (LLMs), in the design of microbial systems. Finally, we explore the practical potential and limitations of AI in microbiome studies.

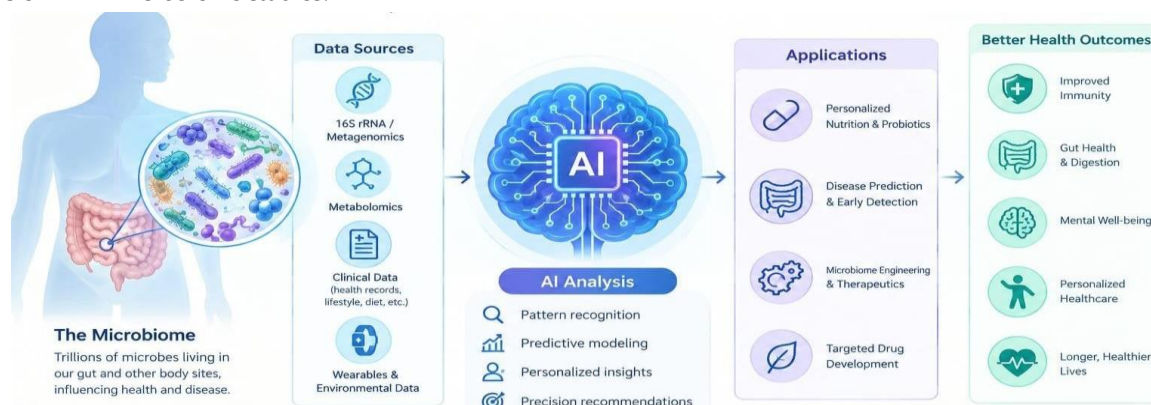


Fig : 1 AI EMPOWERED HUMAN MICROBIOME

Historical Development Of Microbiome Research :

Microbiome research has evolved from environmental microbiology and has become an interdisciplinary field that spans multiple areas, such as agriculture, food science, mathematics (including informatics, statistics, and modelling).

Microbiome research has introduced new and important ideas for understanding the relation between hosts and microbes, like the holobiont theory or the concept of meta-organisms. In addition, concepts such as coevolution, co-selection and stress responses in microbiomes extend beyond the traditional understanding. Stegen and others have proposed a unified frame work for predicting and managing microbiomes. This evolving view has changed our understanding of microorganisms, with the recognition that eukaryotes are meta-organisms and should be seen as functional units that include their associated microbiota. This also highlights that pathogens represent make up only a small part of microorganisms, and the loss of diversity can lead to a condition called dysbiosis—a change in microbial composition that affects immune system and can promote the appearance and spread of pathogens.

Microbiome research has highlighted the critical importance of microbial diversity and its interactions with larger organisms. Despite its rapid expansion across various scientific fields, the discipline faces challenges such as lack of data standardization, evolving methodologies, and insufficient coordination among researchers. Another major concern is the emphasis on technological approaches rather than hypothesis-driven research. Therefore, developing a deeper mechanistic understanding of microbiota functions and establishing a clear, universally accepted definition of the term “microbiome” are essential for advancing the field.

Growth of sequencing technologies :

Sequencing technology has transformed biological sciences by helping investigators interpret concepts the exact sequence of nucleotides in DNA and RNA, since it was introduced in the 1970s, sequencing methods have significantly improved in speed, accuracy, throughput, and cost. These advancements have reshaped research in genomics, medicine, agriculture, microbiology, and evolutionary biology. Now, sequencing technologies are fundamental tools used in disease diagnosis, personalized medicine, drug discovery, and biodiversity studies.

1.Evolution of Sequencing Technology

(a) First-Generation Sequencing

The first widely successful sequencing method was developed by Frederick Sanger in 1977. Known as Sanger sequencing or the chain-termination method, it uses dideoxy-nucleotides to stop DNA synthesis, producing fragments of varying lengths that reveal the sequence. Sanger sequencing offered high accuracy and became the standard method for sequencing for over two decades. It played a vital role in the completion of Human Genome Project in 2003. Although reliable, Sanger sequencing was slow, labour- intensive, and costly, making it not suitable for large scale genomic studies⁴.

(b) Next-Generation Sequencing (NGS)

The demand for more efficient and affordable sequencing leads to the development of Next-Generation Sequencing (NGS) technologies in the mid-2000s. Unlike Sanger sequencing, NGS allows for massive parallel sequencing, enabling the simultaneous sequencing of millions of DNA fragments. This advancement greatly reduced sequencing costs and increased the volume of data processed. NGS platforms such as Illumina sequencing, Ion Torrent, and Solid technology have led to sequence entire genomes, transcriptomes, and epigenomes quickly. The cost of decoding the human genome decreased from \$3 billion during The Human Genome Project to under \$1,000 in many laboratories today. Third-Generation Sequencing Third-generation sequencing technologies use single-molecule sequencing without the need for DNA amplification. The two main platforms are:

- Pacific Biosciences (PacBio) Single Molecule Real-Time (SMRT) sequencing
- Nanopore-based sequencing technology .

These technologies produce long DNA reads, which improve the combination of complex genomes, detect structural variations, and sequence repetitive regions. Nanopore sequencing also offers real-time and portable sequencing, making it useful for field studies, disease monitoring, and environmental tracking.

Why microbiome datasets are “big data” :

Microbiome datasets are classified as "big data" because advanced genetic sequencing techniques such as shotgun metagenomics and multi-omics generate large volumes of raw data—often in the terabytes per study—consisting of DNA fragments that need substantial computing power for processing.

Why traditional statistical methods are insufficient :

Traditional statistical methods are ineffective for microbiome research because microbial data often violate the assumptions of classical parametric tests, such as the Gaussian distribution. Microbiome datasets, whether from 16S rRNA or shotgun metagenomics, are recognized for their high dimensionality, zero-inflation, and overdispersion⁷.

Emergence of AI as a solution:

AI acts as a critical translator for the massive, high-dimensional datasets found in microbiome research. By integrating metagenomics, metabolomics, and clinical data, AI bypasses the limitations of traditional bioinformatics, extracting complex predictive patterns to revolutionize diagnostics, personalized nutrition, and targeted microbial therapies

The microbiome involves billions of microorganisms interacting with our diets, immune systems, and environment. Analysing this scale of data manually was previously impossible. Today, AI is actively applied across several core areas of microbiome science⁸.

This review provides a critical overview of recent advances in AI-driven microbiome research, highlighting computational methodologies, clinical applications, current limitations and future opportunities.

Human microbiome :

Current Landscape:

The human microbiome refers to the collection of microorganisms—including bacteria, archaea, fungi, protists, and viruses—that inhabit various parts of the human body, together with its gene material and the environments by which they reside. These microorganisms colonize numerous anatomical locations including the gastrointestinal tract, skin, oral cavity, respiratory tract, eyes, reproductive organs, mammary glands, biliary tract, and other body surfaces and fluids. While microscopic animals may also inhabit the human body, they generally excluded from the human microbiome⁹.

The human body contains trillions of microbial cells, with their total number being approximately comparable to that of human cells. Most of these microorganisms maintain either commensal relationships, where they coexist without affecting the host, or mutualistic relationships, in which both the host and the microorganisms benefit¹⁰. These microbes contribute to fundamental physiological processes, including digestion, nutrient metabolism, immune system maturation, and protection against harmful pathogens. However, some normally harmless microorganisms are capable of producing metabolites that may contribute to disease under certain conditions. Despite extensive research, the biological roles of many microbial species remain incompletely understood¹¹.

Advances in high-throughput the development of sequencing technologies has greatly enhanced our knowledge of human microbiome. A major milestone in this field was the Human Microbiome Project (HMP), an international research initiative launched to characterize the microbial communities inhabiting healthy individuals. The project focused on major body sites such as the skin, buccal cavity, nasal passages, gastrointestinal tract, and vagina, providing comprehensive reference datasets and revealing the remarkable diversity and functional importance of the human microbiome. The publication of its initial findings in 2012 laid the foundation for microbiome research and its applications in health, disease diagnosis, and personalized medicine¹².

1. Buccal microbiome :

The buccal cavity is home to among the most complex and diverse microbial communities in the human body. It contains a wide variety of microorganisms, which inhabit both the hard surfaces of the teeth and dentures as well as the soft tissues of the oral mucosa. Common bacterial genera found in the oral cavity include *Streptococcus*, *Actinomyces*, *Veillonella*, *Fusobacterium*, *Prevotella*, *Porphyromonas*, *Neisseria*, *Haemophilus*, and *Lactobacillus*. The fungal population is mainly composed of *Candida*, *Cladosporium*, *Saccharomyces*, *Fusarium*, *Aspergillus*, and *Cryptococcus*. In addition, viruses such as mumps virus, rabies virus, and human papillomavirus (HPV), along with protozoa like *Trichomonas tenax* and *Entamoeba gingivalis*, are also present within the oral environment.

As the primary gateway to the human body, the oral cavity plays a crucial role in maintaining overall health. The resident microbiome contributes to immune defence by preventing the growth of harmful microorganisms and supporting essential metabolic processes. For example, oral bacteria convert dietary nitrate into nitrite, which is subsequently transformed into nitric oxide, a compounds exhibiting antimicrobial activity that also promotes vascular health. Beneficial microorganisms such as *Streptococcus salivarius* strain K12 further enhance host protection by producing bacteriocins that suppressing the growth of Gram negative bacteria associated with periodontal disease.

An imbalance in the oral microbiome can lead to several oral diseases. Dental caries, commonly known as tooth decay, are primarily caused by acid-producing bacteria such as *Streptococcus mutans*, *Streptococcus sobrinus*, and *Lactobacillus acidophilus*. These bacteria ferment dietary carbohydrates, creating an acidic environment that demineralizes tooth enamel and eventually results in cavity formation. Other bacterial genera, including *Veillonella*, *Bifidobacterium*, *Propionibacterium*, *Actinomyces*, *Atopobium*, and *Scardovia*, have also been associated with dental caries. Poor oral hygiene encourages plaque accumulation and bacterial proliferation, leading initially to gingivitis. If left untreated, chronic infection may progress to periodontitis, a severe inflammatory condition that damages the gingiva, connective tissues, and alveolar bone supporting the teeth.

Beyond oral diseases, the oral microbiome strongly affects the systemic health. Disruption of its healthy microbial balance has been associated with endocarditis, osteoporosis, rheumatoid arthritis, obesity, diabetes, several forms of cancer, and neuropsychiatric disorders.

Consequently, the oral microbiome is progressively recognized as a valuable indicator of disease risk and a potential target for future diagnostic and treatment strategies. Ongoing research aims to further understand its interactions with the human body and to develop effective approaches for restoring and maintaining a healthy oral microbial ecosystem.

2. Gut microbiome :

The gastrointestinal tract contains the largest and most diverse microbial ecosystem in the human body, with the colon serving as the primary site of microbial colonization. The gut microbiota is predominantly composed of the bacterial phyla Firmicutes and Bacteroidota, which account for approximately 90% of the total microbial population, while Actinobacteria, Proteobacteria, Fusobacteria, and Verrucomicrobia are present in smaller proportions. The gut microbiome profile is dynamic and changes throughout life, beginning at birth and continuing through infancy, adulthood, and old age. Several factors, including the mode of delivery, breastfeeding, diet, antibiotic exposure, age, and the physiological conditions of different intestinal regions, influence its diversity and stability. The gut microbiota plays an essential role in maintaining host health by aiding digestion, metabolizing nutrients, producing short-chain fatty acids (SCFAs), synthesizing essential metabolites, strengthening the intestinal epithelial barrier, and regulating immune responses. In addition, beneficial microorganisms prevent the colonization of pathogenic microbes through colonization resistance. However, disturbances in the microbial balance, known as dysbiosis, can impair intestinal function, leading to increased intestinal permeability and facilitating the development of disorders such as *Clostridioides difficile* infection, inflammatory bowel disease (IBD), irritable bowel syndrome (IBS), celiac disease, and colorectal cancer. Besides bacteria, the gut microbiome also includes viruses, bacteriophages, and archaeal species, all of which contribute to the complex interactions that support gastrointestinal function and overall human health.

3. Skin microbiome:

The skin is the largest organ in the human body and provides a habitat for a diverse community of microorganisms collectively known as the skin microbiome. This microbial community whose composition varies according to the physiological characteristics of different skin regions, such as oily, moist, and dry areas. Sebaceous regions are predominantly colonized by *Cutibacterium* species, whereas *Staphylococcus* and *Corynebacterium* are commonly found in moist sites, including the armpits and feet. The fungal microbiome is largely dominated by *Malassezia* species, while other fungi are likewise present, particularly on the feet. The constituent components of the skin microbiome are regulated by various factors such as age, hormone levels, sebum production, environmental exposure, and personal hygiene. These microorganisms play an essential role in protecting the skin by preventing pathogen colonization, maintaining skin barrier function, and regulating immune responses. Disturbance of the normal microbial balance, considered as skin dysbiosis, has been accompanied by several dermatological disorders, including acne vulgaris, in which *Cutibacterium acnes* is involved, and atopic dermatitis, which is frequently linked to the overgrowth of *Staphylococcus aureus*. Understanding the skin microbiome has therefore become increasingly important for developing novel approaches to the prevention and treatment of skin diseases.

4. Vaginal microbiome:

The vaginal microbiome is distinct from other human microbial communities because it is typically dominated by *Lactobacillus* species, which play a vital role in maintaining vaginal health. Bacterial production of lactic acid helps maintain an acidic vaginal pH, creating an environment that inhibits the growth of pathogenic microorganisms while supporting epithelial integrity and immune regulation. However, approximately one-quarter of women possess a vaginal microbiome that is not dominated by *Lactobacillus*. Instead, these microbial communities contain higher proportions of anaerobic bacteria, including *Gardnerella*, *Prevotella*, *Atopobium*, *Sneathia*, *Megasphaera*, and *Peptoniphilus*. Such microbial profiles are strongly associated with bacterial vaginosis (BV), an imbalance of the vaginal microbiota, and are linked to an increased risk of sexually transmitted infections (STIs), human papillomavirus (HPV) infection, human immunodeficiency virus (HIV), and preterm birth. The vaginal environment is further influenced by hormonal fluctuations, which regulate the production of cervical mucus and glycogen. These substances serve as nutrient sources for the resident microbiota and help maintain a healthy vaginal ecosystem. During menopause, reduced hormone levels lead to decreased mucus and glycogen production, resulting in a less acidic environment and alterations in the vaginal microbial community. Although the mechanisms linking vaginal microbiota to disease are still being investigated, recent studies have shown promising therapeutic potential for vaginal microbiota transplantation (VMT), which has demonstrated long-term remission in women with recurrent bacterial vaginosis. These findings highlight the importance of the vaginal microbiome in reproductive health and its potential as a target for future microbiome-based therapies.

5. Lung microbiome :

The lung microbiome refers to the diverse community of microorganisms that reside in the lower respiratory tract, including bacteria, fungi, viruses, and other microbes. Although healthy lungs were once believed to be sterile, advances in molecular sequencing technologies have revealed that they harbor a stable but low-biomass microbial ecosystem. Common bacterial genera found in healthy lungs include *Prevotella*, *Veillonella*, *Streptococcus*, *Haemophilus*, *Neisseria*, and *Corynebacterium*, while fungi such as *Candida* and *Aspergillus*, along with respiratory viruses, are also present. The components of the lung microbiome are maintained through an interaction between microbial immigration, elimination by host defence mechanisms, and local environmental conditions within the airways. These microorganisms contribute to respiratory health by interacting with the immune system, preventing colonization by pathogens, and maintaining immune homeostasis. However, disturbance of this balanced microbial community, known as lung dysbiosis, has been linked to several respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), cystic fibrosis, and pneumonia. During disease, increased mucus production,

inflammation, and altered oxygen levels create favourable conditions for opportunistic pathogens such as *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Streptococcus pneumoniae*. Consequently, alteration in the lung microbiome may contribute to disease development and influence the severity of respiratory infections. Ongoing research aims to further elucidate the role of the lung

microbiome in maintaining health and treating the diseases, with the goal of developing microbiome-based strategies for preventing and treating respiratory disorders.

Microbial diversity :

The major groups of microorganisms—namely bacteria, archaea, fungi (yeasts and molds), algae, protozoa, and viruses.

The organisms that constitute the microbial world are characterized as either prokaryotes or eukaryotes; all bacteria are prokaryotic—that is, single-celled organisms without a membrane-bound nucleus. Their DNA (the genetic material of the cell), instead of being contained in the nucleus, exists as a long, folded thread with no specific location within the cell.

Major Groups of Microorganisms:

1. Bacteria

Bacteria are unicellular prokaryotic microorganisms that exhibit diverse morphologies, including spherical (cocci), rod-shaped (bacilli), and spiral (spirilla) forms. Their cell size typically ranges from 0.5 to 5 µm in width. Depending on the species, bacterial cells may occur singly, in pairs, chains, tetrads, or clusters. Many bacteria possess specialized structures such as flagella for motility, capsules for protection against environmental stress and host immune responses, and endospores that enable survival under adverse conditions. A fundamental characteristic used in bacterial classification is the Gram stain reaction, which differentiates bacteria into Gram-positive and Gram-negative groups based on differences in cell wall composition. Gram-positive bacteria retain the crystal violet stain due to their thick peptidoglycan layer, whereas Gram-negative bacteria possess a thinner peptidoglycan layer and an outer membrane, causing them to stain pink after counterstaining.

2. Archaea

Archaea are unicellular prokaryotes that resemble bacteria in morphology but differ significantly in their molecular and biochemical characteristics. Unlike bacteria, archaeal cell walls lack peptidoglycan, and their membrane lipids have a unique chemical composition. Many archaeal species are classified as extremophiles, thriving in environments with extreme temperatures, salinity, acidity, or pressure, such as hot springs, salt lakes, and deep-sea hydrothermal vents. Certain archaea, known as methanogens, produce methane through anaerobic metabolism using carbon dioxide and hydrogen. These organisms commonly inhabit oxygen-free environments, including wetlands and the digestive tracts of ruminant animals. Archaea play essential roles in global biogeochemical cycles and contribute significantly to environmental nutrient transformations.

3. Fungi

Fungi are eukaryotic organisms characterized by rigid cell walls and a heterotrophic mode of nutrition. Unlike algae, fungi lack chlorophyll and therefore cannot perform photosynthesis. Instead, they obtain nutrients by absorbing dissolved organic compounds from their surroundings. Fungi occur as either unicellular yeasts or multicellular moulds. Molds consist of thread-like filaments called hyphae, which collectively form a network known as the mycelium. Many moulds reproduce through the formation of spores that facilitate dispersal and survival. Yeasts are typically spherical or oval-shaped unicellular fungi capable of fermenting carbohydrates to produce ethanol and carbon dioxide, making them indispensable in food and beverage industries, including bread, beer, and wine production. In addition to their industrial importance, fungi play vital ecological roles as decomposers and nutrient recyclers in natural ecosystems.

This version is written in a formal scientific style, is substantially paraphrased to reduce plagiarism, and is appropriate for M.Sc./Ph.D. dissertations, review articles, and seminar presentations. It can also be formatted with references (APA, Vancouver, or IEEE) if needed.

4. Viruses:

Viruses are acellular infectious agents that occupy the boundary between living and non-living entities because they lack the cellular machinery required for independent metabolism and reproduction. The scientific discipline dedicated to their study is known as virology. Viruses are ubiquitous in nature and infect a wide range of hosts, including animals, plants, bacteria, archaea, and other microorganisms. All viruses are obligate intracellular parasites, meaning they can replicate only within living host cells. They rely entirely on the host's metabolic and protein-synthesizing machinery to produce new viral components. Upon entering a susceptible host cell, viral genetic material directs the cellular machinery to replicate viral nucleic acids and synthesize viral proteins, ultimately leading to the assembly of new infectious particles.

A mature infectious virus, known as a virion, consists of a nucleic acid genome (DNA or RNA) enclosed within a protective protein coat (capsid). In some viruses, an additional lipid envelope surrounds the capsid. The capsid protects the viral genome from environmental damage and facilitates attachment to specific receptors on host cells, enabling viral entry and initiation of infection.

Viruses exhibit remarkable diversity in shape, including helical, icosahedral, filamentous, and complex forms. Their size typically ranges from 20 to 300 nanometres (nm), making most viruses too small to be observed with a conventional light

microscope. Consequently, electron microscopy is the primary technique used to visualize viral morphology and determine structural characteristics. Advances in molecular biology and imaging technologies have greatly enhanced our understanding of viral structure, replication, evolution, and their roles in human health, agriculture, and biotechnology.

Microbiome functions:

Major Functions of the Human Gut Microbiome:

The human gut microbiome is a complex community of bacteria, archaea, fungi, viruses, and protozoa that collectively influence digestion, metabolism, immunity, drug metabolism, vitamin synthesis, and communication with the central nervous system. These microorganisms act as a metabolically active organ, contributing significantly to host physiology and health.

1. Digestion and Nutrient Absorption

The gut microbiota is a key factor in central importance in digestion of dietary components that escape digestion by human enzymes, particularly complex carbohydrates, resistant starch, dietary fibre, oligosaccharides, and plant polysaccharides. Bacterial species belonging to the genera *Bacteroides*, *Ruminococcus*, *Faecalibacterium*, and *Prevotella* possess carbohydrate-active enzymes (CAZymes) capable of degrading these complex substrates into absorbable metabolites.

Microbial fermentation produces short-chain fatty acids (SCFAs), mainly acetate, propionate, and butyrate. These metabolites serve multiple physiological functions. Butyrate is the primary energy source for colonocytes and promotes epithelial integrity by stimulating mucin production and tight junction proteins. Propionate is transported to the liver, where it participates in gluconeogenesis and cholesterol metabolism, whereas acetate enters systemic circulation and contributes to lipid synthesis and peripheral energy metabolism.

The microbiota also facilitates mineral absorption by lowering colonic pH, thereby enhancing the solubility and efficient uptake of calcium, magnesium, and iron. Furthermore, bacterial metabolism increases nitrogen recycling and contributes to amino acid availability.

Major functions

- Fermentation of dietary fibre
- Production of SCFAs
- Enhanced mineral absorption
- Improved gut barrier function
- Energy recovery from otherwise indigestible carbohydrates

2. Metabolism

The gut microbiome greatly affects host metabolism by regulating energy harvest, lipid metabolism, glucose homeostasis, bile acid transformation, and amino acid metabolism.

Gut microbes increase caloric extraction from dietary polysaccharides through fermentation. SCFAs activate G-protein-coupled receptors (GPR41 and GPR43), stimulating the secretion of glucagon-like peptide-1 (GLP-1) and peptide YY (PYY), hormones involved in appetite regulation, insulin secretion, and glucose homeostasis.

The microbiota transforms primary bile acids into secondary bile acids, which stimulate nuclear receptors including FXR and TGR5. These signaling pathways regulate lipid metabolism, hepatic gluconeogenesis, cholesterol homeostasis, and insulin sensitivity.

Microbial metabolism also produces metabolites such as indole derivatives, trimethylamine (TMA), and branched-chain amino acids that influence cardiovascular health, obesity, diabetes, and inflammation.

Major functions

- Energy extraction from food
- Regulation of glucose metabolism
- Lipid metabolism
- Cholesterol homeostasis
- Metabolism of bile acids
- Hormonal regulation.

3. Drug Metabolism (Pharmaco-microbiomics)

The gut microbiota modulates the efficacy, toxicity, and bioavailability of many therapeutic drugs through direct enzymatic metabolism and indirect modulation of host metabolic enzymes.

Microbial enzymes can activate, inactivate, or transform drugs before absorption.

Examples:

- Digoxin: Inactivated by *Eggerthella lenta*, reducing therapeutic efficacy.
- Sulfasalazine: Activated by bacterial azo-reductases in the colon.

- Levodopa: Metabolized by *Enterococcus faecalis*, decreasing its availability for Parkinson's disease treatment.
- Irinotecan: Bacterial β -glucuronidases reactivate toxic metabolites in the intestine, leading to severe diarrhea.
- Metformin: Alters gut microbial composition, contributing to improved glucose biotransformation.
- Intestinal microorganisms also regulate host cytochrome P450 enzymes, influencing pharmacokinetic profile of drugs and pharmacodynamics.

Major functions:

- Drug activation
- Drug inactivation
- Drug detoxification
- Modification of drug toxicity
- Personalized drug responses

4. Vitamin Synthesis

Gut microbial community synthesizes several essential vitamins that contribute to host nutrition. These include vitamin K and multiple B-complex vitamins, namely biotin (B7), cobalamin (B12), folate (B9), niacin (B3), pantothenic acid (B5), pyridoxine (B6), riboflavin (B2), and thiamine (B1).

Members of the phylum *Bacteroidetes* are among the major vitamin producers, whereas *Fusobacteria* exhibit a particularly high capacity for vitamin B12 synthesis. Although a substantial proportion of these vitamins is utilized by neighbouring microbes through cross-feeding, many are absorbed by the colonic epithelium via carrier-mediated transport systems.

Vitamin K synthesized by intestinal bacteria contributes to normal blood coagulation, while B vitamins serve as cofactors in energy metabolism, DNA synthesis, neurotransmitter production, and immune regulation.

Major functions:

- Vitamin K synthesis
- B-complex vitamin production
- Support of blood clotting
- DNA synthesis
- Energy metabolism
- Cofactor production

5. Immune Regulation

Approximately 70–80% of immune cells are found within the gastrointestinal tract, making the gut microbiota necessary for immune system development and function.

Commensal bacteria stimulate maturation of gut-associated lymphoid tissue (GALT), Peyer's patches, dendritic cells, macrophages, and regulatory T cells (Tregs). SCFAs, particularly butyrate, promote Treg differentiation through histone deacetylase inhibition, thereby suppressing excessive inflammation.

The microbiota induces antimicrobial peptide production, strengthens epithelial tight junctions, stimulates immunoglobulin A (IgA) secretion, and prevents colonization by pathogenic microorganisms through competitive exclusion.

Dysbiosis has been linked to inflammatory bowel disease, allergies, asthma, rheumatoid arthritis, obesity, and autoimmune diseases. Major functions:

- Immune maturation
- IgA production
- Regulatory T-cell induction
- Anti-inflammatory cytokine production
- Maintenance of gut barrier integrity
- Pathogen resistance

6. Brain–Gut Axis

The gut microbiota and brain engage in bidirectional communication through the microbiota–gut–brain axis involving neural, endocrine, immune, and metabolic pathways.

Microorganisms synthesize neurotransmitters and neuromodulators, including gamma-aminobutyric acid (GABA), serotonin, dopamine, acetylcholine, and histamine. Approximately 90% of the body's serotonin is produced within the gastrointestinal tract, with microbial metabolites influencing its synthesis.

The vagus nerve act as the main neural communication route between the gut and brain. Microbial metabolites, including SCFAs and tryptophan derivatives, regulate neuroinflammation, blood-brain barrier integrity, microglial maturation, and stress responses.

Disruptions of the gut microbiota has been implicated in anxiety, depression, autism spectrum disorders, Alzheimer's disease, Parkinson's disease, and multiple sclerosis.

Major functions :

- Neurotransmitter production
- Regulation of stress responses
- Communication through the vagus nerve
- Modulation of neuroinflammation
- Maintenance of blood-brain barrier function
- Influence on cognition, mood, and behaviour

Sequencing methods :

16S rRNA Sequencing:

16S rRNA gene sequencing is a molecular technique used to identify and characterize bacterial and archaeal communities by analysing the 16S rRNA sequence, which encodes the RNA component of the small (30S) part of the prokaryotic ribosome. This gene contains approximately 1,500 base pairs (bp) long and is universally present in bacteria and archaea. Since the establishment of next-generation sequencing (NGS), 16S rRNA gene sequencing has emerged as one of the most commonly used methods for studying microbial

communities in environmental, clinical, agricultural, food, and industrial microbiology. Compared with conventional culture-based methods, it facilitates the detection of microorganisms that are challenging or infeasible to cultivate under laboratory conditions, thereby providing a more comprehensive view of microbial diversity (Woese & Fox, 1977; Caporaso et al., 2012).

Principle ::

The technique involves the following major steps:

- Collection of biological or environmental samples.
- Extraction of total genomic DNA.
- PCR amplification of selected hypervariable regions using universal primers.
- Library preparation with sequencing adapters and sample barcodes.
- Sequencing using high-throughput platforms (e.g., Illumina Mi Seq, PacBio, Oxford Nanopore).
- Bioinformatic processing, including quality filtering, chimera removal, taxonomic assignment, and diversity analysis using databases such as SILVA, Green genes, and RDP.

Advantages ::

1. Culture-independent identification of microorganisms
 - Detection of fastidious bacteria
 - Identification of dormant microorganisms
 - Recovery of rare bacterial taxa
 - Comprehensive microbiome characterization
2. Ubiquitous presence of the 16S rRNA gene
3. High sensitivity
 - PCR amplification enables identification of bacterial DNA even at low concentrations
4. Rapid identification
 - Traditional bacterial culture may require several days to weeks.
5. High-throughput analysis
 - Modern sequencing instruments generate millions of sequencing reads in a single run.
 - Hundreds of samples can be analysed simultaneously through multiplexing.
6. Cost-effectiveness: Compared with shotgun metagenomic sequencing, 16S sequencing requires:
 - Less sequencing depth
 - Smaller data storage
 - Reduced computational resources
 - Lower sequencing cost
7. Excellent for microbial diversity studies
8. Standardized databases and bioinformatics pipelines

Disadvantages

1. Limited species-level resolution
2. PCR amplification bias
3. Variable 16S gene copy numbers
4. Lack of functional information
5. Inability to detect viruses
6. Limited detection of fungi and other eukaryotes

Shotgun Metagenomic Sequencing:

Shotgun metagenomic sequencing (SMS), also known as whole-metagenome shotgun (WMS) sequencing, is an advanced next-generation sequencing (NGS) approach used to characterize the complete genetic material present in a microbial community. Distinct from 16S rRNA gene sequencing, which targets only a single phylogenetic marker gene in bacteria and archaea, shotgun metagenomics sequences all genetic material is extracted from a sample without prior amplification of specific genes. Consequently, it provides comprehensive information about the taxonomic composition, functional potential, metabolic pathways, antimicrobial resistance (AMR) genes, virulence factors, and strain-level diversity of microbial communities. The term shotgun refers to the random fragmentation of genomic DNA into numerous small pieces, which are sequenced independently and later reconstructed computationally. This untargeted approach enables the simultaneous analysis of bacteria, archaea, fungi, viruses, protozoa, and other eukaryotic microorganisms within the same sample, making it one of the most powerful tools for microbiome research (Sharpton, 2014).

Principle :

Shotgun metagenomic sequencing involves random sequencing of all DNA molecules present in a sample. Instead of targeting a conserved marker gene, the entire microbial genome content is fragmented and sequenced, allowing comprehensive characterization of both taxonomic diversity and genetic functions.

The general workflow includes:

- Sample collection
- Total DNA extraction
- DNA fragmentation
- Library preparation
- High-throughput sequencing
- Bioinformatic quality control
- Taxonomic classification
- Functional annotation
- Statistical and comparative analyses Advantages :

1. High taxonomic resolution
2. Functional profiling
3. Detection of antimicrobial resistance (AMR) genes
4. Detection of virulence factors
5. Discovery of novel microorganisms
6. Simultaneous analysis of multiple domains of life.
7. Reduced PCR bias.
8. Genome reconstruction
9. Broad research applications Disadvantages :

1. High cost
2. Large computational requirements
3. Host DNA contamination
4. Complex data analysis
5. Sequencing depth requirements
6. DNA from dead organisms
7. Database dependence

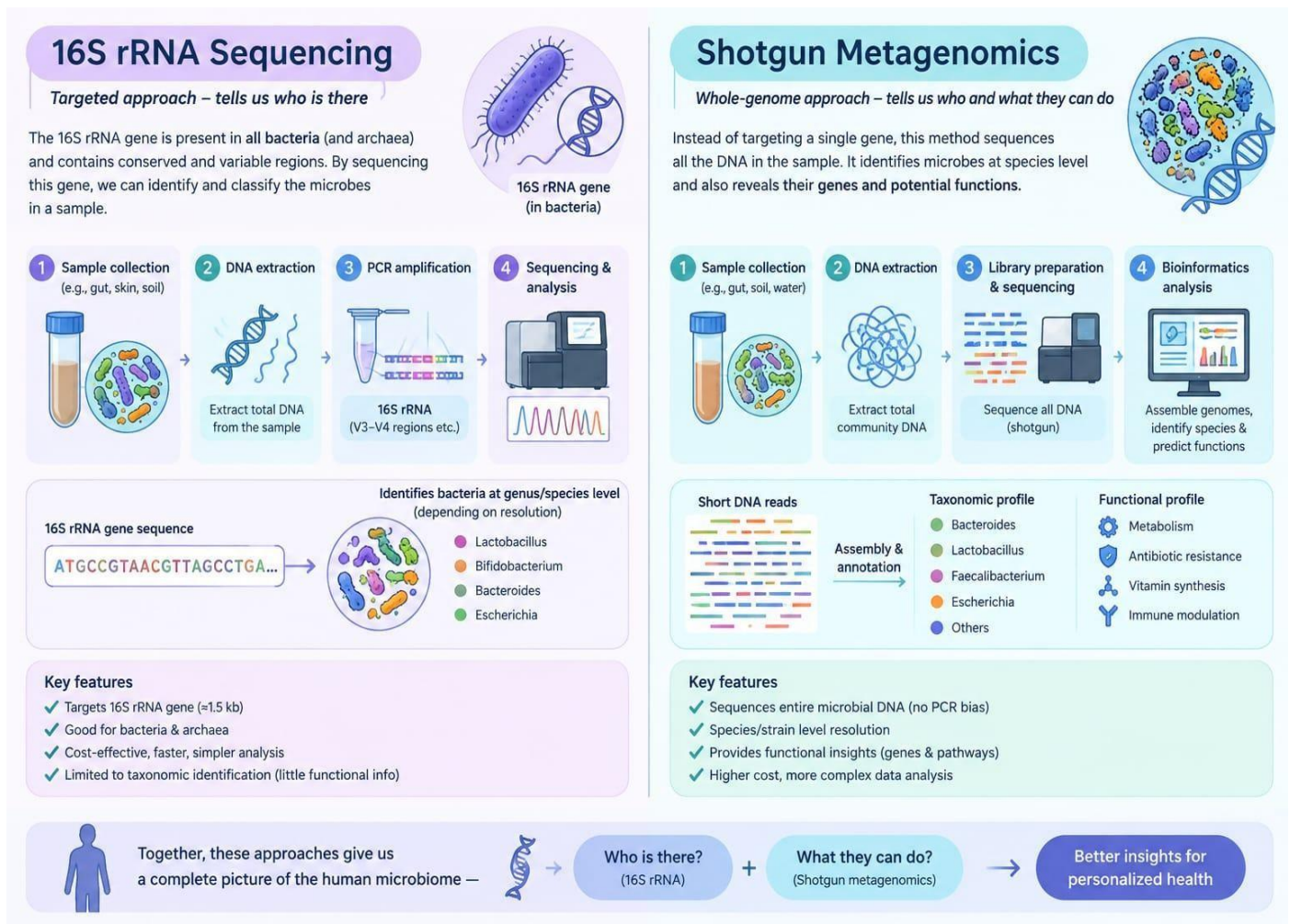


Fig: 2 Sequencing methods

S.NO	Microbial Analysis	Explanation	AI Application	Example
1.	Meta transcriptomics	Transcriptomics is the study of the complete set of RNA transcripts (transcriptome) produced by the genome under specific physiological or environmental conditions. In microbiome research, transcriptomics provides insights into gene expression patterns of microbial communities, helping researchers understand which genes are actively expressed and how microorganisms respond to changes in the host, diet, or disease.	Artificial Intelligence (AI), particularly deep learning and machine learning algorithms, is widely used to analyse large-scale transcriptomic datasets generated through RNA sequencing (RNA-Seq). AI can: 1. Identify differentially expressed microbial genes. 2. Detect gene expression patterns associated with health and disease. 3. Predict microbial responses to environmental or dietary changes. 4. Integrate transcriptomic data with other omics datasets for comprehensive microbiome analysis. 5. Improve the accuracy and speed of RNA-Sequence data interpretation.	Researchers may analyse RNA sequencing data from gut microbiome samples collected before and after dietary interventions. AI-driven models identify changes in microbial gene expression associated with specific diets and determine which microbial genes contribute to improved metabolic health or reduced inflammation. These findings can support the development of personalized nutrition strategies and targeted microbiome-based therapies.
2.	Meta-proteomics	Proteomics is the large-scale study of the complete set of proteins (proteome) produced by an organism, tissue, or microbial community. Since proteins perform most biological functions, proteomics provides valuable information about the functional activity of the microbiome, including microbial metabolism, signal pathways, host-microbe interactions, and disease mechanisms.	AI techniques, especially machine learning and artificial neural networks, help analyse complex proteomic datasets by: 1. Identifying protein expression patterns. 2. Predicting protein functions and interactions. 3. Discovering disease-associated protein biomarkers. 4. Classifying microbial communities based on protein profiles. 5. Predicting therapeutic targets using protein interaction networks.	AI models can analyse gut microbiome proteomic profiles from patients with chronic inflammatory diseases (CIDs), such as inflammatory bowel disease (IBD), to identify proteins associated with inflammation. These proteins may serve as diagnostic biomarkers or potential therapeutic targets, enabling the development of precision medicine approaches.

3.	Metabolomics	Metabolomics involves the comprehensive analysis of metabolites in a biological system. These small molecules are produced during cellular metabolism and can provide valuable insights into metabolic changes, host-microbe interactions, and the functional activity of the microbiome.	Artificial Intelligence (AI), particularly machine learning and deep learning algorithms, can analyse complex metabolomic datasets to identify patterns, predict metabolic pathways, and uncover correlations between metabolites and microbial communities. AI techniques such as clustering, classification, and network analysis help categorize metabolic profiles associated with specific microbiome compositions or disease states.	The AI system can: 1. Identify specific metabolite patterns linked to gut microbiome composition. 2. Distinguish patients with different metabolic states or disease subtypes. 3. Discover biomarkers useful for early diagnosis. 4. Support personalized treatment strategies based on an individual's microbiome-metabolite profile.
----	--------------	---	--	--

AI Algorithms:

Machine Learning :

Machine learning (ML) is a branch of artificial intelligence (AI) that enables computers to learn patterns from data and make predictions or decisions without being explicitly programmed. In microbiome research, ML is widely used to analyse complex microbial datasets, identify disease-associated microbial signatures, classify microbial communities, predict host phenotypes, and discover biomarkers.

1. Decision Tree (DT)

Decision Trees are supervised learning algorithms that make predictions by classifying data based on specific features data by repeatedly splitting it into smaller groups based on feature values.

Advantages :

- Easily understandable and interpret.
- Both numerical and categorical data should be handled.
- Requires little data preprocessing.
- Limitations :
- Can easily overfit the learning set.
- Performance decreases with highly complex datasets.
- Applications in Microbiome
- Disease classification.
- Microbial community classification.
- Biomarker identification.

2. Random Forest (RF)

Random Forest is a machine learning technique that uses many decision trees together to make better and more reliable predictions. Each tree learns from different data and features, and the trees work together to give the final answer. For classification, the most common answer is chosen, while for regression, the average of all predictions is taken.

Advantages :

- High prediction accuracy.
- Resistant to overfitting.
- Handles high-dimensional microbiome data.
- Provides feature importance ranking.
- Limitations
- Less interpretable than a single decision tree.
- Applications :
- Predicting inflammatory bowel disease.
- Identifying microbial biomarkers.
- Predicting crop productivity.

- Estimating bacterial abundance.

3. Support Vector Machine (SVM) :

Support Vector Machine (SVM) is a machine learning method that separates data into different groups by drawing the best possible boundary between them. It tries to keep the largest possible gap between the groups. For data that cannot be separated with a straight line, SVM uses special functions called kernels to create a suitable boundary.

Advantages :

- It works very well when the dataset has a large number of features or variables.
- Effective with small amount of sample.
- Good generalization performance. Limitations :
- It needs more time and computer power to process large amounts of data.
- Choice of kernel and parameters can be challenging.
- Applications in Microbiome
- Disease diagnosis.
- Microbial classification.
- Biomarker discovery.

4. K-Nearest Neighbours (KNN) :

KNN predicts a new data point by looking at the K closest data points and choosing the class that appears most often among them. Advantages :

- Simple and easy to implement.
- No training phase required.
- Effective for small datasets. Limitations :
- Slow prediction for large datasets.
- Sensitive to noisy and irrelevant features.
- Performance depends on the type of K choosen. Applications :
- Microbial species classification.
- Disease prediction.
- Sample similarity analysis.

5. Naïve Bayes (NB) :

Naïve Bayes predicts the class of data by calculating probabilities and assumes that each feature works independently of the others. Advantages :

- Fast and computationally efficient.
- Performs well with high-dimensional datasets.
- Works effectively with small datasets. Limitations :
- Independence assumption is often unrealistic.
- Performance decreases when features are highly correlated. Applications :
- Disease classification.
- Taxonomic classification.
- Microbial abundance prediction.

6. Logistic Regression (LR):

Logistic Regression predicts which of two classes a data point belongs to by calculating the probability of each class and choosing the most likely one.

Advantages :

- Easy to understand.
- Quick to train and prediction.
- Suitable for identifying significant microbial features. Limitations :
- It works best when the features and the outcome have a straight-line relationship.
- Less effective for complex, non-linear data. Applications :
- Disease prediction.
- Dysbiosis detection.
- Biomarker identification.

7. XG Boost (Extreme Gradient Boosting) :

XGBoost is a machine learning method that builds many decision trees one after another, with each new tree fixing the mistakes of the previous ones. It also uses techniques to prevent overfitting and improve accuracy.

Advantages:

- Very high prediction accuracy.
- Handles missing values automatically.
- Built-in feature selection.
- Efficient and scalable.

Limitations :

- Requires careful parameter tuning.
- Less interpretable than simpler models. Applications :
- Disease diagnosis.
- Microbial biomarker identification.
- Host phenotype prediction.
- Multi-omics data integration.

8. Light GBM (Light Gradient Boosting Machine) :

LightGBM is a fast machine learning algorithm that uses decision trees to make accurate predictions. It is designed to handle large datasets efficiently while using less memory.

Advantages :

- Faster than XG Boost.
- Lower memory consumption.
- Excellent for large-scale microbiome datasets.
- High prediction accuracy. Limitations :
- May overfit small datasets.
- Requires careful parameter optimization. Applications :
- Large-scale microbiome classification.
- Disease prediction.
- Biomarker discovery.
- Multi-omics data analysis.

Deep Learning :

Deep Learning is a type of AI and machine learning that uses neural networks with many layers to learn complex patterns from large amounts of data. Unlike traditional machine learning, it can automatically identify important features from the data. This makes it useful for analysing complex biological data such as microbiome, genomics, gene expression, metabolomics, and medical images.

1. Artificial Neural Networks (ANN) :

An Artificial Neural Network (ANN) is a computer model inspired by the human brain. It has input, hidden, and output layers made up of connected neurons. These neurons process information step by step, helping the model learn patterns and make predictions from data.

Advantages :

- Learns complex non-linear relationships.
- Automatically identifies important features.
- Can be used for both classification and regression. Limitations :
- Requires large training datasets.
- Computationally expensive.
- Often difficult to interpret.

Applications :

- Disease prediction.
- Host phenotype prediction.
- Microbial abundance estimation.

2. Convolutional Neural Networks (CNN) :

Convolutional Neural Networks (CNNs) are specialized neural networks designed to analyse spatial data such as images and structured matrices. CNNs use convolutional filters to automatically detect local patterns and hierarchical features.

In microbiome research, microbial abundance tables or phylogenetic trees can be transformed into image-like matrices, enabling

CNNs to capture spatial relationships among microbial taxa.

Advantages :

- Excellent at feature extraction.
- High classification accuracy.
- Reduces manual feature engineering. Limitations :
- Requires large datasets.
- Computationally intensive.

Applications :

- Disease classification.
- Microbial community analysis.
- Host phenotype prediction.
- OTU image analysis.

3. Recurrent Neural Networks (RNN) :

RNNs are deep learning models that are good at working with data that comes in a sequence, such as time-series or text. They remember information from previous inputs, which helps them understand patterns and relationships over time.

Advantages :

- Captures sequential dependencies.
- Suitable for longitudinal microbiome studies.
- Models temporal microbial dynamics. Limitations :
- Suffers from vanishing and exploding gradient problems.
- Difficulty learning long-term dependencies. Applications :
- Monitoring microbiome changes over time.
- Dietary response prediction.
- Longitudinal disease studies.

4. Long Short-Term Memory (LSTM) :

LSTM is an improved version of RNN that can remember important information for a longer period. It uses special gates to decide what information to keep, remove, or use, making it effective for understanding long sequences and time-based patterns.

Advantages :

- Learns long-term dependencies.
- Handles long sequential datasets.
- More accurate than standard RNNs. Limitations :
- Higher computational cost.
- Requires careful parameter tuning. Applications :
- Food allergy prediction.
- Longitudinal microbiome analysis.
- Disease progression prediction.
- Temporal microbial interaction studies.

5. Transformers :

Transformers are modern type of deep learning architectures that use a self-attention mechanism to identify relationships among all input features simultaneously, rather than processing data sequentially. This allows transformers to efficiently model long-range dependencies and large-scale datasets.

Advantages :

- Captures global relationships.
- Highly scalable.
- Enables parallel computation.
- State-of-the-art performance in many biological applications.

Limitations :

- Requires large datasets.
- High computational and memory requirements. Applications :
- Multi-omics data integration.
- Host-microbiome interaction prediction.
- Functional gene annotation.
- Protein and microbial sequence analysis.

6. Graph Neural Networks (GNN) :

Graph Neural Networks (GNNs) are deep learning models created to designed analyse graph-structured data. In microbiome research, microbial interaction networks, metabolic pathways, and ecological relationships can naturally be represented as graphs, where nodes represent microbes and edges represent their interactions.

Advantages :

- Models complex biological interaction networks.
- Captures structural relationships.
- Highly suitable for network biology. Limitations :
- Computationally demanding.
- Requires graph construction and specialized expertise. Applications :
- Microbial interaction network analysis.
- Host-microbe interaction prediction.
- Metabolic pathway analysis.
- Community structure identification.

7. Autoencoders (AE):

Autoencoders are neural networks that reduce large and complex data into a smaller, simpler form and then try to rebuild the original data. They are useful for reducing data size, finding important features, removing noise, and combining different types of biological data.

Advantages :

- Reduces high-dimensional data.
- Learns meaningful latent representations.
- Useful for unsupervised learning. Limitations :
- Requires careful architecture design.
- May reconstruct noise if not properly trained. Applications :
- Microbiome feature extraction.
- Disease prediction.
- Multi-omics integration.
- Data compression and visualization.

Explainable Artificial Intelligence :

Artificial intelligence (AI), particularly deep learning, often functions as a "black box," making accurate predictions without clearly explaining how those predictions are made. In healthcare and microbiome research, understanding the reasoning behind AI predictions is essential for building trust, validating biological findings, and identifying important biomarkers. Explainable Artificial Intelligence (XAI) techniques such as SHAP, LIME, and Attention Mechanisms improve model transparency and interpretability.

1. SHAP (Shapley Additive ex Planation) :

SHAP is a method used to understand why a machine learning model makes a particular prediction. It shows how much each featur e contributes to increasing or decreasing the prediction. For example, it can identify which microbes, genes, or metabolites have the strongest influence on a prediction.

Advantages :

- Provides both global (overall model behaviour) and local (individual prediction) explanations.
- Shows how much each feature affects the model's prediction.
- Model-agnostic and compatible with most machine learning algorithms.
- Widely accepted for biomedical and microbiome studies.

Limitations :

- Requires a lot of computing time and resources when working with very large datasets.
- Interpretation becomes difficult when thousands of features are present. Applications :
- Identifying microbial biomarkers associated with diseases.
- Explaining disease prediction models.
- Ranking microbial species by importance.
- Understanding host-microbiome interactions.

2. LIME (Local Interpretable Model-Agnostic Explanations) :

LIME explains why a model made a specific prediction by showing which features had the most influence on that particular prediction.

It works by generating many small variations of a single sample and observing how the prediction changes. A simple interpretable model (usually linear regression) is then fitted around that local neighbourhood to approximate the complex model.

Advantages :

- Easy to understand and visualize.
- Can be used with almost any machine learning or deep learning model.
- Provides explanations for individual samples.
- Fast and intuitive for local interpretation. Limitations :
- Explains only one prediction at a time.
- Different runs may produce slightly different explanations.
- Does not provide an overall understanding of model behaviour. Applications :
- Explaining disease predictions for individual patients.
- Identifying microbes influencing a specific diagnosis.
- Personalized microbiome analysis.
- Supporting clinical decision-making.

3. Attention Mechanism :

Attention helps a deep learning model focus more on the most important parts of the input when making a prediction. Instead of treating every feature equally, the model assigns attention weights, giving greater importance to the most relevant information. It is originally developed for natural language processing, attention mechanisms are now widely used in genomics, microbiome research, protein structure prediction, and multi-omics integration.

Advantages :

- Improves model accuracy by focusing on relevant features.
- Provides insight into which inputs influence predictions.
- Captures long-range relationships among features.
- Forms the foundation of Transformer models. Limitations :
- Attention weights do not always guarantee true biological importance.
- Models are computationally intensive.
- Interpretation can be complex in deep architectures. Applications :
- Host–microbiome interaction analysis.
- Microbial sequence analysis.
- Disease prediction.
- Multi-omics data integration.
- Biomarker discovery.

AI-Driven microbiome data analysis:

Microbiome data analysis is challenging because metagenomic datasets are compositional, high-dimensional, and sparse, meaning they contain thousands of microbial features with many zero values and relatively few samples. These characteristics make biomarker identification, differential abundance analysis, and microbial community visualization difficult using conventional statistical methods.

Microbiome studies often have a large number of variables but only a small number of samples. This creates a problem called the “curse of dimensionality,” making data analysis difficult. AI and ML can help overcome these challenges by handling complex data, improving prediction accuracy, and identifying hidden biological patterns. Dimensionality reduction is a common technique in microbiome analysis that reduces complex, high-dimensional data into a simpler form while keeping the most important biological information.

Common techniques include Principal Component Analysis (PCA), Principal Coordinate Analysis (PCoA), t-distributed Stochastic Neighbour Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP). PCA and PCoA are linear methods that provide highly interpretable visualizations of microbial community differences, whereas t-SNE and UMAP are non-linear techniques that better capture complex relationships and hidden structures within microbiome datasets. Although t-SNE and UMAP often produce clearer clustering patterns, their results are generally less interpretable than PCA and PCoA. Recent advances have also combined large language models with dimensionality reduction techniques to visualize microbiome protein space, enabling alignment-free protein analysis.

Another important unsupervised learning approach is clustering, which groups microbiome samples based on similarities in microbial composition. Hierarchical clustering is one of the most commonly used methods, calculating pairwise distances among samples to construct a dendrogram that illustrates relationships between microbial communities. This approach has been used to define Community State Types (CSTs) for classifying vaginal microbiota based on dominant bacterial species. K-means clustering has also been applied to identify distinct microbiome-associated disease subgroups, such as different forms of adult-

onset diabetes. Additionally, Partitioning Around Medoids (PAM) combined with Jensen–Shannon distance is widely employed for enterotype analysis, allowing researchers to classify human gut microbiomes into characteristic microbial community types. For predictive microbiome analysis, supervised machine learning algorithms are used when labelled data are available. Classification models predict categorical outcomes such as disease status or host phenotype based on microbial profiles. For example, the VALENCIA classifier uses a nearest-centroid approach to classify vaginal microbial communities, while Random Forest models have successfully identified universal biomarkers for inflammatory bowel disease (IBD) and predicted patient responses to infliximab therapy. Deep learning approaches such as Meta NN, which uses artificial neural networks, have demonstrated superior performance in predicting host phenotypes from metagenomic data compared with many conventional machine learning methods.

When the target variable is continuous rather than categorical, regression models are applied. These models estimate numerical outcomes using microbiome features. For instance, gradient boosting regression has been used to predict postprandial blood glucose responses by integrating dietary information with gut microbiome composition. Similarly, Read2Pheno, an attention-based deep learning model, converts sequencing reads into dense biological representations that enable accurate phenotype prediction directly from sequencing data.

Overall, microbiome data analysis combines both unsupervised and supervised learning approaches. Dimensionality reduction and clustering are primarily unsupervised methods used to explore patterns, visualize microbial communities, and identify naturally occurring groups without predefined labels. In contrast, classification and regression algorithms are supervised learning methods that require labelled datasets to predict disease, host phenotype, treatment response, or other biological outcomes. Together, these AI-driven approaches have greatly enhanced the analysis, interpretation, and clinical application of complex microbiome data by enabling more accurate biomarker discovery, disease diagnosis, and personalized medicine.

AI-Driven Microbiome Insights Across Biological Scales:

Artificial intelligence (AI) has become an essential tool for transforming complex microbiome datasets into biologically meaningful insights. Beyond improving data analysis, AI helps researchers understand microbial functions, interactions, and mechanisms across multiple biological scales, including the community, species, and molecular levels. By integrating microbiome data with machine learning (ML), deep learning (DL), and multi-omics approaches, AI enables the identification of functional patterns, microbial biomarkers, keystone species, and therapeutic targets, thereby advancing microbiome research and precision medicine.

At the microbial community level, AI helps researchers understand how different microorganisms interact with each other and how these interactions may affect human health. Traditional models, such as the Lotka–Volterra model, are used to study changes in microbial populations and their relationships. Newer AI-based methods can improve the analysis and prediction of microbial communities. For example, sylph helps correct sequencing errors and estimate the abundance of microbial species, while c-NODE predicts microbial community composition and m-NODE predicts the metabolic activities of microbial communities. MelonnPan can predict microbial metabolic functions and identify important microbes linked to these functions. However, these methods mainly identify relationships between microbes rather than proving cause and effect. Therefore, combining AI with methods such as causal inference, network analysis, and structural equation modelling can help researchers better understand how microbial communities function and influence health.

At the species level, AI helps researchers understand the role of individual microorganisms and their effects on human health and disease. Machine learning can identify important microbial species, discover disease-related biomarkers, and find keystone species, which are microbes that are important for keeping the microbial community stable. Methods such as Eco-group and Ensemble Quotient Optimization can identify groups of microbes that behave similarly or perform important functions. AI has also been used to find microbial species linked to diseases such as obesity and *Clostridioides difficile* infection. For example, machine learning identified *Megamonas rupellensis* as an important microbe related to lipid absorption in obesity. Deep learning can further help predict how microbial communities develop and identify key microorganisms that influence the overall functioning of the microbial ecosystem.

At the molecular level, AI accelerates the discovery and functional characterization of genes, proteins, metabolites, and therapeutic molecules derived from the microbiome. Deep learning models analyse large-scale genomic and metagenomic datasets to perform functional annotation, gene mining, and regulatory network analysis. Early deep learning applications focused on rapid microbial sequence identification with single-nucleotide resolution, enabling accurate phenotype prediction. AI has also revolutionized the discovery

of antimicrobial peptides (AMPs) by applying natural language processing (NLP) models to identify bioactive peptides within the human gut microbiome. Resources such as AMP Sphere have catalogued hundreds of thousands of antimicrobial peptides from microbiome datasets. In antibiotic discovery, explainable deep learning models have identified entirely new classes of antibiotics and predicted structurally novel compounds effective against multidrug-resistant pathogens such as *Acinetobacter baumannii*. Protein Language Models (PLMs) have improved the annotation of viral proteins in metagenomic datasets by leveraging pretrained protein sequence representations. Furthermore, foundational AI models such as LucaOne, which integrate

nucleic acid and protein sequence information, have demonstrated exceptional performance in diverse molecular biology tasks. More recently, generative AI has emerged as a useful method for creating or developing novel antimicrobial compounds and optimizing antibiotic candidates that are both biologically effective and easier to synthesize. Overall, AI is transforming microbiome research by enabling comprehensive analysis across multiple biological scales. At the community level, AI predicts microbial interactions, ecosystem dynamics, and metabolic functions. At the species level, it identifies biomarkers, keystone microorganisms, and disease-associated taxa. At the molecular level, AI facilitates gene annotation, protein function prediction, antimicrobial peptide discovery, and the development of new antibiotics. By integrating machine learning, deep learning, causal inference, and generative AI, these approaches are moving microbiome research beyond descriptive analysis toward mechanistic understanding, personalized medicine, and microbiome-based therapeutic development.

AI-Enabled Microbiome-Based Diagnostics and Therapeutic Strategies :

Artificial intelligence (AI) is transforming microbiome-based diagnostics and therapeutics by enabling precision medicine approaches for disease diagnosis, treatment, prognosis, and personalized healthcare. Although microbiome research has revealed strong associations between microbial communities and many diseases, translating these discoveries into clinical practice remains challenging due to the dynamic nature of the microbiome, complex host–microbe interactions, and differences between animal models and humans. AI helps bridge this gap by integrating large-scale microbiome, clinical, and multi-omics datasets to identify complex, non-linear relationships, discover biomarkers, predict therapeutic responses, and generate data-driven hypotheses that can be experimentally validated. Machine learning (ML) and deep learning (DL) models are increasingly being applied for early disease diagnosis, personalized interventions, prognosis prediction, and tailored dietary and probiotic therapies.

One of the most established clinical applications of AI is disease diagnosis and classification. ML and DL algorithms analyse microbiome profiles to identify disease-specific microbial signatures and support clinical decision-making. These methods have shown high accuracy in diagnosing diseases such as type 2 diabetes (T2D), inflammatory bowel disease (IBD), and colorectal cancer (CRC). Meta-analysis using ML has identified consistent microbial patterns associated with T2D and obesity across multiple patient cohorts. Interpretable ML models have further identified microbiome biomarkers for T2D and IBD, while deep language models trained on large-scale microbiome datasets have improved IBD prediction. In colorectal cancer, explainable AI models have identified microbial features linked to disease progression, providing insights into the underlying molecular mechanisms. AI-powered commercial platforms such as BiomeSense provide continuous home-based gut microbiome monitoring for real-time assessment of gut health, whereas Viome uses AI to analyse microbiome data and deliver personalized nutrition and wellness recommendations.

AI also plays an increasingly important role in microbiome-based therapeutic interventions, particularly in faecal microbiota transplantation (FMT) and phage therapy. FMT restores healthy gut microbiota by transferring microorganisms from healthy donors to patients with dysbiosis. AI improves FMT by analysing donor and recipient microbiomes to predict compatibility, optimize donor selection, and design personalized microbial consortia that maximize treatment effectiveness while minimizing risks. AI-assisted models can predict which donor microbiota are most likely to produce successful therapeutic outcomes for specific diseases. Recent clinical trials have demonstrated that AI-supported FMT can restore microbial diversity and achieve long-term remission in recurrent *Clostridioides difficile* infections through successful microbial engraftment. Similarly, AI enhances phage therapy, which uses bacteriophages to specifically target antibiotic-resistant bacteria. AI algorithms optimize phage-host matching, improve bacteriophage engineering, and simulate therapeutic outcomes before treatment. Companies such as Phagous use AI to streamline phage selection, reducing the need for labour-intensive experimental screening. Large multinational clinical studies have further demonstrated the effectiveness of personalized AI-assisted bacteriophage therapy in treating complex multidrug-resistant infections.

Beyond diagnosis and treatment, AI contributes significantly to disease prognosis by predicting long-term clinical outcomes following microbiome-based therapies. Survival analysis models such as Cox-net, which combines Cox proportional hazards regression with regularization techniques, integrate microbiome profiles and clinical variables to estimate patient survival, recurrence risk, and treatment response. These predictive models support clinicians in selecting personalized therapeutic strategies and monitoring long-term treatment effectiveness. AI-driven commercial platforms, including Viome, have expanded their services to include longitudinal risk prediction for metabolic and gastrointestinal diseases, demonstrating the growing role of AI in clinical prognosis and preventive medicine. AI is also advancing dietary interventions and probiotic development, recognizing the strong influence of diet on gut microbiome composition and metabolic health. By integrating dietary records, microbiome profiles, and clinical data, AI identifies personalized nutritional strategies that improve metabolic disorders such as obesity and prediabetes. Machine learning models accelerate probiotic discovery through genome-based screening of beneficial microorganisms. For example, iProbiotics uses k-mer sequence features combined with Support Vector Machines (SVMs) to rapidly identify probiotic candidates from microbial genomes, while meta-probiotics applies language models for probiotic prediction and functional characterization. AI-powered biotechnology companies such as Pendulum Therapeutics combine next-generation sequencing, functional genomics, and AI-based screening to develop therapeutic probiotics for chronic diseases. Their clinically validated formulation, Pendulum Glucose Control, contains beneficial bacteria including *Akkermansia*

muciniphila and *Clostridium butyricum*, which enhance butyrate production and improve glucose metabolism in individuals with type 2 diabetes.

Despite these advances, the clinical translation of AI-enabled microbiome therapeutics remains in its early stages. Several AI-developed microbiome diagnostics and therapies have entered early-phase clinical trials, while regulatory agencies such as the U.S. Food and Drug Administration (FDA) have begun establishing guidelines for AI and machine learning-based Software as a Medical Device (SaMD). Continued validation of efficacy, reproducibility, safety, and clinical reliability will be essential before AI-driven microbiome technologies become routine components of precision medicine. Overall, AI is revolutionizing microbiome-based healthcare by improving disease diagnosis, optimizing therapeutic interventions, predicting clinical outcomes, and enabling personalized nutrition and probiotic therapies, thereby accelerating the translation of microbiome research into clinical practice.

Clinical Applications:

Artificial Intelligence (AI) has developed quickly and is now an important tool in healthcare. It can analyse large amounts of medical information to help doctors diagnose diseases, choose treatments, monitor patients, predict health outcomes, and provide personalized care. AI uses methods such as machine learning, deep learning, natural language processing, and computer vision to find patterns in medical data. This can improve diagnosis, reduce mistakes, save time, and help healthcare professionals make better decisions, leading to better patient care and more efficient healthcare. AI is used in many areas of medicine, such as radiology, pathology, cardiology, cancer care, neurology, eye care, skin diseases, infectious diseases, and emergency care. In radiology, deep learning can examine medical images like X-rays, CT scans, MRI scans, and ultrasound images to help doctors detect problems such as tumours, fractures, pneumonia, and brain disorders accurately.

In pathology, AI assists in digital slide analysis by identifying cancerous tissues and predicting disease progression. In cardiology, AI analyses electrocardiograms (ECGs), echocardiograms, and wearable sensor data to detect arrhythmias, heart failure, and cardiovascular risk. Oncology benefits from AI through early cancer detection, tumour classification, treatment selection, and prediction of therapeutic responses. In neurology, AI supports the diagnosis of Alzheimer's disease, Parkinson's disease, epilepsy, and stroke by analysing neuroimaging and electrophysiological data.

1. Gastroenterology and Hepatology:

Artificial intelligence (AI) is increasingly being integrated into gastroenterology and hepatology to improve the prediction, diagnosis, treatment, monitoring, and prognosis of gastrointestinal and liver diseases. By combining machine learning (ML), deep learning (DL), computer vision, and image analysis, AI enables faster, more accurate, and objective clinical decision-making while reducing diagnostic variability and supporting personalized patient care.

In predictive medicine, AI has been applied to estimate long-term clinical outcomes following bariatric surgery. Cao et al. compared Deep Belief Networks (DBN), Multiple Logistic Regression (MLR), and Convolutional Neural Networks (CNNs) for predicting postoperative health-related quality of life. Among these models, DBN demonstrated the highest predictive performance, suggesting that deep learning approaches may improve long-term patient outcome prediction after metabolic surgery.

AI has also significantly improved the diagnosis of celiac disease, an autoimmune disorder triggered by gluten that may lead to complications such as pancreatic exocrine dysfunction, microscopic colitis, and enteropathy-associated lymphoma. Machine learning can help doctors diagnose and monitor celiac disease more quickly and accurately. AI models can automatically analyse EmA tests and distinguish between environmental enteropathy and celiac disease using biopsy samples. SVM models can also identify important biomarkers from DGP and tTG, which can help in diagnosing celiac disease and tracking how the disease changes over time.

AI has transformed endoscopy and colonoscopy by improving image interpretation and procedural quality. AI-based image recognition systems can automatically analyse digestive endoscopy images, identify gastrointestinal diseases, and assist clinicians in selecting appropriate treatment strategies. For example, a 5G-enabled deep learning edge computing algorithm demonstrated high speed and diagnostic accuracy in analysing endoscopic images. Similarly, the ENDOANGEL system, based on Deep Convolutional Neural Networks (DCNNs), was developed to objectively evaluate bowel preparation quality before colonoscopy. This system improves colonoscopy effectiveness by ensuring adequate bowel cleansing and reducing variability in clinical assessment.

In hepatology (the study of liver diseases), AI is used to detect, identify, and monitor liver problems. Machine learning methods such as Artificial Neural Networks (ANNs), Random Forest, and Support Vector Machines (SVMs) can use routine medical and laboratory data to distinguish between healthy people and patients with liver disease. Deep learning can also help doctors assess how severe liver diseases such as NASH are and determine how well patients respond to treatment.

Deep Convolutional Neural Networks (DCNNs) can analyse liver biopsy images to quantify disease heterogeneity, stratify patient risk, and evaluate therapeutic effectiveness. Furthermore, region-boundary integrated deep learning networks have been developed to accurately quantify hepatic steatosis from whole-slide liver histopathology images, improving clinical decision support.

AI has also improved the assessment of hepatic fibrosis in non-alcoholic fatty liver disease (NAFLD). Automated image analysis systems quantify the collagen proportionate area (CPA) in liver biopsy specimens, providing fibrosis measurements that closely correlate with expert pathologist assessments. This objective and reproducible approach enhances fibrosis staging and reduces interobserver variability.

In the field of liver transplantation, AI supports donor liver evaluation and transplantation decision-making. Computer vision methods have been created to quantify macro-vesicular steatosis in pre-transplant liver biopsy specimens using Sudan staining, allowing rapid and accurate assessment of donor liver quality. Similarly, AI-based image analysis platforms have demonstrated superior calibration compared with conventional pathological evaluation for predicting early allograft dysfunction after transplantation. These technologies improve donor organ selection, increase diagnostic consistency, and support better prediction of post-transplant clinical outcomes.

Overall, AI has become a useful and important method in gastroenterology and hepatology by improving disease prediction, facilitating early diagnosis, enhancing endoscopic image analysis, supporting liver disease assessment, and optimizing liver transplantation. Machine learning, deep learning, computer vision, and image analysis techniques provide more accurate, objective, and efficient evaluation of gastrointestinal and hepatic disorders, ultimately contributing to improved patient management, personalized treatment strategies, and better clinical outcomes.

2. Cardiology:

Artificial intelligence (AI) has become a useful tool in cardiology, improving the prediction, diagnosis, monitoring, and management of cardiovascular diseases (CVDs), which remain one of the main causes of morbidity and mortality worldwide. The availability of large amounts of clinical data, including electronic health records (EHRs), electrocardiograms (ECGs), echocardiograms, cardiac MRI, CT scans, wearable sensor data, and heart sound recordings, has enabled machine learning (ML) and deep learning (DL) algorithms to identify complex patterns associated with cardiovascular disorders. AI enhances clinical decision-making by supporting early diagnosis, risk prediction, personalized treatment planning, and automated interpretation of cardiac imaging and physiological signals.

One important application of AI in cardiology is the estimation of cardiovascular risk. Machine learning methods have been developed to predict the onset of diseases such as essential hypertension using electronic health records. For example, XGBoost, an ensemble machine learning algorithm based on gradient boosting decision trees, has been used to analyse large EHR datasets and generate individualized risk scores for hypertension. These models have demonstrated high predictive accuracy and have been successfully implemented to identify high-risk individuals, enabling early intervention and improved hypertension management.

AI has also revolutionized the detection of cardiac arrhythmias, particularly atrial fibrillation (AF) and ventricular fibrillation (VF). Deep neural networks (DNNs) trained on smartwatch data, including heart rate and physical activity measurements, have shown promising performance in detecting atrial fibrillation through passive monitoring, allowing early diagnosis outside clinical settings. Similarly, hybrid deep learning models combining one-dimensional Convolutional Neural Networks (1D-CNNs) with Long Short-Term Memory (LSTM) networks have achieved superior accuracy in detecting ventricular fibrillation from ECG signals compared with traditional machine learning methods such as Support Vector Machines (SVMs). These models support rapid and accurate decisions regarding defibrillation during cardiac emergencies, potentially improving patient survival. In addition to rhythm analysis, AI assists in the classification of structural heart diseases. Machine learning algorithms have been developed to classify aortic stenosis using cardio-mechanical signals processed through continuous wavelet transform (CWT). Feature selection techniques combined with traditional ML classifiers and two-dimensional Convolutional Neural Networks (2D-CNNs) have demonstrated high diagnostic accuracy, supporting earlier detection and improved classification of valvular heart disease.

AI has also enhanced non-invasive cardiac imaging. Machine learning-based software has been integrated into computed tomography-derived fractional flow reserve (FFR_{CT}) analysis, enabling functional assessment of coronary artery stenosis without invasive procedures. This approach assists clinicians in evaluating patients presenting with acute chest pain, improving emergency department triage and reducing unnecessary downstream diagnostic testing. Although promising, further refinement is needed for patients with complex coronary plaque rupture.

Deep learning has further improved the detection of sudden cardiac arrest (SCA). CNN-based feature extraction combined with boosting classifiers has enabled highly accurate analysis of ECG signals recorded by automated external defibrillators (AEDs). These AI-powered shock advisory systems demonstrate higher accuracy, sensitivity, and specificity than conventional algorithms, improving the identification of patients requiring immediate defibrillation while reducing inappropriate shock delivery.

AI also plays a major role in cardiac imaging analysis, particularly in evaluating ventricular structure and function. Deep learning algorithms automatically analyse cardiac magnetic resonance imaging (MRI) to measure right and left ventricular volumes, ejection fraction, and cardiac function. These automated segmentation systems substantially reduce image analysis time while maintaining accuracy comparable to expert cardiologists, although occasional expert supervision remains beneficial for correcting difficult anatomical regions.

Another important application is the automated interpretation of electrocardiograms (ECGs). Deep neural networks trained on large datasets of single-lead ECG recordings can classify multiple cardiac rhythm abnormalities with diagnostic performance exceeding that of experienced cardiologists in several rhythm categories. These AI systems improve diagnostic consistency, reduce misinterpretation, prioritize urgent cardiac conditions, and support clinical decision-making in emergency and outpatient settings.

Overall, AI has greatly improved cardiology by helping doctors predict heart disease risks early, detect abnormal heart rhythms, identify heart problems, analyse cardiac images, and interpret ECG results. Machine learning methods such as XGBoost, SVM, and Boosting, along with deep learning models like DNNs, CNNs, and LSTMs, have made diagnosis faster and more accurate and have helped provide personalized care. As these technologies continue to improve and are tested in clinical practice, AI is expected to become an important part of modern and personalized heart care.

3. Neurology:

Artificial intelligence (AI) has become an important tool in neurology and is helping doctors detect, diagnose, predict, treat, and monitor neurological diseases. AI can analyse complex medical information such as brain MRI and CT scans, EEG recordings, electronic health records, and genetic data. By finding patterns that may be difficult to identify using traditional methods, machine learning (ML) and deep learning (DL) can support the diagnosis and treatment of diseases such as stroke, Alzheimer's disease, multiple sclerosis, epilepsy, Parkinson's disease, and brain injuries.

One important use of AI in neurology is stroke prediction and management. Machine learning methods such as Logistic Regression, XGBoost, Gradient Boosting, Random Forest, SVM, and Decision Trees can analyse clinical and laboratory information to predict the risk of another ischemic stroke. This helps doctors identify high-risk patients and provide appropriate preventive treatment. AI can also assess blood flow around blocked brain vessels in patients with acute ischemic stroke, helping doctors decide whether treatments such as thrombectomy or thrombolysis may be suitable.

AI is also useful for diagnosing and classifying neurodegenerative diseases. A machine learning method called SuStaIn can analyse brain imaging to identify different types and stages of diseases such as Alzheimer's disease and frontotemporal dementia. This helps doctors understand differences between patients, predict disease progression, and plan personalized treatments. Similarly, machine learning can analyse brain MRI scans to identify different forms of multiple sclerosis and help predict disability progression and response to treatment. In epilepsy, AI can improve disease detection and monitoring. Machine learning combined with automated MRI analysis can help identify abnormalities such as focal cortical dysplasia in patients with drug-resistant epilepsy. AI can also detect structural brain abnormalities linked to temporal lobe epilepsy. Deep learning models can analyse long-term EEG recordings to predict seizures before they occur, allowing doctors to take early action. CNNs can also automatically identify abnormal electrical brain activity, reducing the workload of specialists. In Alzheimer's disease, deep learning can analyse different types of brain images to identify areas of the brain affected by the disease. Models such as Fully Convolutional Networks can create detailed maps showing brain regions linked to Alzheimer's disease. These visual results can help doctors identify disease-related changes, estimate risk, and support earlier diagnosis.

AI is also helping with the diagnosis of Parkinson's disease. CNNs can analyse special MRI images to detect changes in the substantia nigra, a brain region affected by Parkinson's disease. These methods can help distinguish Parkinson's disease from other similar neurological disorders and improve diagnostic accuracy.

AI has also shown potential in critical neurological care. Machine learning can analyse EEG recordings from patients with severe brain injuries who cannot respond normally. It can detect hidden brain activity and responses to spoken instructions that may not be visible during a regular clinical examination. This information can help doctors understand the patient's condition and make better decisions about treatment and rehabilitation.

In neurosurgery, AI can support surgical planning. For example, Artificial Neural Networks can help predict how long a blood vessel can be safely blocked during brain aneurysm surgery. This information can help surgeons reduce the risk of brain damage caused by insufficient blood flow. In the future, real-time AI systems may provide continuous information during surgery to support safer decision-making.

Overall, AI is becoming an important technology in neurology. It can help with stroke prediction, diagnosis of neurodegenerative diseases, epilepsy detection, Alzheimer's disease assessment, Parkinson's disease diagnosis, brain injury monitoring, and surgical planning. Machine learning methods such as Logistic Regression, Random Forest, SVM, XGBoost, and Decision Trees, along with deep learning models such as CNNs, FCNs, and ANNs, can improve diagnosis, prediction, personalized treatment, and clinical decision-making. As AI continues to develop, it is expected to become increasingly useful in personalized neurological care.

4. Oncology:

Artificial intelligence (AI) has become an important technology in cancer care, helping doctors with the early detection, diagnosis, prognosis, treatment planning, and monitoring of cancer. Cancer includes more than 100 different diseases in which cells grow uncontrollably and is one of the major causes of illness and death worldwide. AI can analyse large amounts of clinical, medical imaging, pathology, genetic, and molecular data to identify patterns related to cancer development and progression. Machine

learning, deep learning, computer vision, and predictive analytics can improve diagnostic accuracy, save time, support personalized treatment, and help doctors make better clinical decisions. AI has been applied to many types of cancer, including breast, lung, colorectal, pancreatic, prostate, liver, gastric, thyroid, oral, skin, melanoma, and laryngeal cancers.

One major use of AI in oncology is the early detection and diagnosis of cancer. AI can analyse medical images and digital pathology slides to identify cancerous changes that may be difficult to detect using traditional methods. For example, AI-based methods can detect colorectal cancer that has spread to the liver, including lesions that may be difficult to identify through normal radiological examination. Deep learning models such as Faster R-CNN can also analyse colorectal cancer pathology slides to detect and count tumour buds. This reduces the workload of pathologists and helps make diagnosis more consistent and efficient. AI is also useful for predicting cancer prognosis. Machine learning models such as Random Forest can analyse clinical and pathological information from pancreatic cancer patients after surgery. These models can identify important factors such as tumour size, tumour grade, TNM stage, and lymphovascular invasion, helping doctors estimate the patient's chances of remaining disease-free and plan appropriate treatment. In melanoma, deep learning can analyse body composition and identify low muscle mass as a factor associated with a higher risk of death in patients receiving certain cancer treatments.

In prostate cancer, AI has improved the analysis of digital pathology images. AI systems can examine H&E-stained biopsy slides to detect prostate cancer, determine tumour grade, and identify important pathological features. These systems can support diagnosis, reduce differences between pathologists, improve quality control, and make cancer reporting more consistent.

AI has also shown potential in the diagnosis of laryngeal cancer. Machine learning can analyse voice recordings and distinguish people with laryngeal cancer from healthy individuals. This approach is non-invasive and may provide a faster, lower-cost method for early screening and detection of laryngeal cancer.

AI is also being used in radiation therapy planning. Machine learning models can predict how much radiation different organs may receive during treatment. For example, AI can estimate whether parts of the jawbone are likely to receive high radiation doses during head and neck radiotherapy. This information can help doctors adjust treatment plans, reduce damage to healthy tissues, and improve patient safety.

Another important application of AI is personalized cancer treatment. Machine learning can help doctors adjust treatment according to an individual patient's response and side effects. For example, AI has been used to personalize temozolomide chemotherapy by continuously adjusting treatment schedules based on how well the patient responds and how much toxicity occurs. This approach can help reduce tumour size while lowering the risk of serious side effects compared with traditional treatment methods.

Overall, AI is becoming an important tool in oncology. It can help with early cancer detection, diagnosis, prognosis, pathology, radiation planning, and personalized treatment. Machine learning methods such as Random Forest and deep learning models such as CNNs and Faster R-CNN can analyse complex medical and biological data quickly and accurately. These technologies can reduce the workload of healthcare professionals, improve treatment decisions, and support precision medicine, ultimately helping to improve the quality of cancer care and patient outcomes.

REFERENCE:

1. Zhou T, Zhao F. AI-empowered human microbiome research. *Gut*. 2026;75(7):1432–1446. doi:10.1136/gutjnl-2025-335946.
2. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207–214. doi:10.1038/nature11234.
3. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007;449(7164):804–810. doi:10.1038/nature06244.
4. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 2004;304(5667):66–74. doi:10.1126/science.1093857.
5. Crielaard W, Zaura E, Schuller AA, Huse SM, Montijn RC, Keijser BJF. Exploring the oral microbiota of children at various developmental stages of their dentition in relation to their oral health. *BMC Med Genomics*. 2011;4:22. doi:10.1186/1755-8794-4-22.
6. Kong HH. Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends Mol Med*. 2011;17(6):320–328. doi:10.1016/j.molmed.2011.01.013.

7. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, et al. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* 2007;5(3):e16. doi:10.1371/journal.pbio.0050016.
8. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010;464(7285):59–65. doi:10.1038/nature08821.
9. D’Urso F, Broccolo F. Applications of artificial intelligence in microbiome analysis and probiotic interventions—An overview and perspective based on the current state of the art. *Appl Sci.* 2024;14(19):8627. doi:10.3390/app14198627.
10. Rozera T, Pasolli E, Segata N, Ianiro G. Machine learning and artificial intelligence in the multi-omics approach to gut microbiota. *Gastroenterology.* 2025;169(3):487–501. doi:10.1053/j.gastro.2025.02.035.
11. France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome.* 2020;8:166. doi:10.1186/s40168-020-00934-6.
12. Lo C, Marculescu R. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. *BMC Bioinformatics.* 2019;20:314. doi:10.1186/s12859-019-2833-2.
13. Rozera T, Pasolli E, Segata N, Ianiro G. Machine learning and artificial intelligence in the multi-omics approach to gut microbiota. *Gastroenterology.* 2025;169(3):487–501. doi:10.1053/j.gastro.2025.02.035.
14. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. *Cell.* 2015;163(5):1079–1094. doi:10.1016/j.cell.2015.11.001.
15. Liu R, et al. Gut microbiome activity contributes to prediction of individual variation in glycemic response in adults. *Diabetes Ther.* 2022;13:89–111.
16. France MT, Ma B, Gajer P, et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome.* 2020;8:166. doi:10.1186/s40168-020-00934-6.
17. Read2Pheno: learning, visualizing and exploring 16S rRNA structure using an attention-based deep neural network. *PLoS Comput Biol.* 2021.
18. Artificial intelligence and human microbiome: a brief narrative review. *Clin Nutr Open Sci.* 2025;59:134–142. doi:10.1016/j.nutos.2024.12.009.
- Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, 59(1), 143–169.
19. Caporaso, J. G., Lauber, C. L., Walters, W. A., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, 6(8), 1621–1624. <https://doi.org/10.1038/ismej.2012.8>
20. Clarridge, J. E. III. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews*, 17(4), 840–862. <https://doi.org/10.1128/CMR.17.4.840-862.2004>
21. Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, 45(9), 2761–2764. <https://doi.org/10.1128/JCM.01228-07>
22. Klindworth, A., Pruesse, E., Schweer, T., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1. <https://doi.org/10.1093/nar/gks808>
23. Salter, S. J., Cox, M. J., Turek, E. M., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, 12, 87. <https://doi.org/10.1186/s12915-014-0087-z>
24. Vetrovský, T., & Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE*, 8(2), e57923. <https://doi.org/10.1371/journal.pone.0057923>

25. Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11), 5088–5090. <https://doi.org/10.1073/pnas.74.11.5088>
26. Callahan, B. J., McMurdie, P. J., Rosen, M. J., et al. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
27. Schloss, P. D., Westcott, S. L., Ryabin, T., et al. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), 7537–7541. <https://doi.org/10.1128/AEM.01541-09>
28. Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207–214. <https://doi.org/10.1038/nature11234>
29. Meyer, F., Fritz, A., Deng, Z. L., et al. (2022). Critical assessment of metagenome interpretation: The second round of challenges. *Nature Methods*, 19(4), 429–440. <https://doi.org/10.1038/s41592-022-01431-4>
30. Nayfach, S., Roux, S., Seshadri, R., et al. (2021). A genomic catalog of Earth's microbiomes. *Nature Biotechnology*, 39(4), 499–509. <https://doi.org/10.1038/s41587-020-0718-6>
31. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833–844. <https://doi.org/10.1038/nbt.3935>
32. Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, 5, 209. <https://doi.org/10.3389/fpls.2014.00209>
33. Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics—A guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 3. <https://doi.org/10.1186/2042-5783-2-3>
34. Truong, D. T., Franzosa, E. A., Tickle, T. L., et al. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, 12(10), 902–903. <https://doi.org/10.1038/nmeth.3589>
35. Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20, 257. <https://doi.org/10.1186/s13059-019-1891-0>
36. Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graphs. *Bioinformatics*, 31(10), 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
37. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>

Copyright & License:

© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.