Precision Microbiome Medicine and AI-Driven Infectious Disease Diagnostics

Editors

Ali Salah Jumaah
Department of Biology, College of Science, University of Basra, Iraq

Ali Al-Murtadha Ahmed Hassan
Department of Biology, College of Science, University of Basra, Iraq

Forqan Sameer Jasim
Department of Biology, College of Science, University of Basra, Iraq

Bright Sky Publications ™
New Delhi

Published By: *Bright Sky Publications*

Bright Sky Publication
Office No. 3, 1st Floor,
Pocket - H34, SEC-3,
Rohini, Delhi, 110085, India

Editors: *Ali Salah Jumaah Ali, Al-Murtadha Ahmed Hassan and Forqan Sameer Jasim*

The author/publisher has attempted to trace and acknowledge the materials reproduced in this publication and apologize if permission and acknowledgements to publish in this form have not been given. If any material has not been acknowledged please write and let us know so that we may rectify it.

© **Bright Sky Publications**

Edition: *1st*
Publication Year: *2025*
Pages: *101*
Paperback ISBN: *978-93-6233-738-2*
E-Book ISBN: *978-93-6233-069-7*
DOI: *https://doi.org/10.62906/bs.book.456*
Price: *₹654/-*

Abstract

Microbial infections continue to emerge as a leading cause of widespread morbidity and significant mortality on a global scale, impacting millions of lives each year. Timely and accurate detection of various pathogens is absolutely critical for ensuring effective treatment options and for curtailing the further spread of these infections within communities and populations. The human microbiome-the incredibly complex collective of genomes and diverse interactions of the microbial communities that exist on and within the human body-plays a pivotal role in shaping infectious disease dynamics and overall health outcomes. The specific composition of the microbiome significantly influences not only pathogen carriage rates but also the intricate transmission dynamics and overall expression of various diseases. However, the full predictive power of microbiome data has yet to be completely harnessed for practical diagnostic applications that could revolutionize patient care. This groundbreaking work proposes an innovative modular suite of artificial intelligence (AI) tools that are uniquely capable of seamlessly integrating existing microbiome-disease knowledge alongside a wide array of diverse multi-omic datasets. This integration substantially enhances the overall identification of viral, bacterial, and fungal pathogens, regardless of any previous sequencing or culture results that may have been obtained in routine laboratory practices. These advanced and sophisticated approaches work diligently to accelerate the robust development of crucial microbiome-informed diagnostic models. Such models will take into account specific pathogen trajectories and broaden the application of genomics-guided decision support systems within the expansive field of infectious disease management.

Ultimately, this will lead to significantly better patient outcomes, more effective public health strategies, and the potential to save countless lives by addressing the challenges posed by microbial infections more effectively [1, 2, 3].

Contents

Chapter - 1

Introduction to the Human Microbiome

The human microbiome refers to the collection of organisms that inhabit the human body, which includes bacteria, archaea, fungi, and viruses. These organisms colonize the skin, oral cavity, and gut after birth and play important roles in bodily functions, such as digestion and immune function. The microbiome contains approximately 100 times as many genes as the human genome; thus, it profoundly influences a range of biological processes. Imbalances to health-promoting microbiome functions are associated with numerous health issues, ranging from inflammatory bowel disease, diabetes, and antimicrobial-resistant infections. Systematic investigation of microbial communities using high-throughput sequencing techniques has yielded an extensive catalogue of species prevalent in diverse human populations and has revealed how microbial community composition changes from healthy states. The human microbiome is an extraordinarily promising target for early detection and therapeutic intervention.

The microbiome influences both the innate and adaptive immune systems, which helps dictate the success of vaccine and therapy responses. Specific immune-related signals generated by the microbiota regulate front-line innate immune defences, including the production of antimicrobial peptides and mucus, and shape the composition and function of distinct CD4+ T helper, regulatory T-cell, and T follicular helper populations. Many of the same microbiota-derived signals and populations remain important for managing the severity of subsequent viral, bacterial, and parasitic infections.

The literature describes observations linking shifts in microbiome community structure to numerous health and disease phenotypes. Healthy, disease-predictive, and infection-predictive microbiota have also been associated with infections by species such as Clostridium difficile, Pseudomonas aeruginosa, and pathogens capable of causing sepsis. Pathogen-host-microbiome triads have been defined to describe these interconnections [4, 5, 6].

Overview of the human microbiome (gut, skin, oral, etc.)

Different ecosystems colonize the human body such as the gut, skin, vagina, and mouth. The gut microbiome and its functioning have been studied extensively and provide insights into the ecological dynamics of the other ecosystems. The dominant phyla in the gut are Firmicutes and Bacteroidetes, and their relative abundance varies according to age, diet, and other factors. Microbiome shifts impact the host's health and are linked to chronic diseases. Knowledge of the microbiota extends to the skin and oral environments, and ongoing investigations continue to explore these ecosystems. Microbiome disturbance is expected to provide information about the health state of their corresponding sites and other anatomical areas [7, 8, 5].

Role in health and disease

Dysbiosis, an alteration in the composition of the microbiota, is increasingly associated with inflammatory and infectious diseases [5]. Rodent studies identified microbial community changes linked with conditions ranging from obesity to stress, while disease signatures have been connected with fibrosis and Crohn's disease in humans. Microbiota independence is defined as the absence of assembly and functional correlation between the gut microbiota and the disease. Immunological processes triggered by microbiota can affect pathogen clearance, and a multi-kingdom docking model connects dysbiotic communities to infection susceptibility. Several studies have generated association maps that link communities, edge counts, and genotype variations across major infectious diseases, highlighting the early microbiota establishment window [9].

Importance in immune system regulation

The human microbiota comprises a vast array of microorganisms, including bacteria, viruses, archaea, fungi, and microbial eukaryotes, and occupies all body surfaces. Despite their small proportions, these microorganisms significantly influence a wide range of physiological processes [9]. During birth, the microbiota starts to develop based on maternal factors, environmental exposure, and other factors. The configuration continues to change throughout life, driven by many factors, including environmental, diet, antibiotics, hygiene, contraceptives, health issues, climate, and more. There is a need to obtain high-quality and high-throughput metagenomics data to predict the health status and health conditions of humans.

There are four critical and significant elements that must be thoroughly considered when addressing any form of infection: the pathogen, the host, the immune system, and the microbiome. Of these, the dynamics of the microbiome surrounding infections contain vital and essential information about both the pathogen and the host. Gaining a deep understanding of these intricate interactions, particularly in the periods before treatment initiation and during the treatment process itself, can lead to innovative and effective solutions aimed at improving infection control and overall management strategies. The human microbiome is currently studied in its entirety through a comprehensive approach known as multi-omics, which includes metagenomics, metatranscriptomics, metabolomics, resistomics, and various other methods. This extensive research framework is employed to obtain as much detailed information as possible about a wide range of certain medical conditions and diseases, enhancing our understanding of how infections can be managed more successfully [10, 11, 12].

Chapter - 2

The Era of Precision Medicine

Definition and evolution of precision medicine

A new generation of personalized frameworks divides individuals into groups that share common features for stratification and clinical decision support. Personalization now extends beyond the genomics of single genes and diseases. A growing family of signals, encompassing virome, resistome, immunome, and metabolic features, together with artificial intelligence (AI), excels at rapidly categorizing individuals and calibrating treatment strategies for infectious pathogens. These extensions complement precision medicine approaches and surmount existing gaps in standard stratification systems, which do not yet accommodate the diversity of pathogen, transmission route, and clinical presentation arising from the same human source [4].

Precision medicine is fundamentally about ensuring that the correct treatment is delivered to the appropriate patient at the optimal moment. However, traditional medical frameworks frequently treat various pathogens, including fungal, viral, and bacterial, in a somewhat undifferentiated manner. This lack of specificity can lead to suboptimal treatment outcomes. In contrast, precision microbiome medicine introduces an essential new dimension by incorporating the microbiome into the discussion. This approach offers a complementary and clinically significant way to characterize the individual patient. It emphasizes the critical role that the microbiome plays in modulating the body's response to various pathogens, ultimately influencing treatment efficacy. By understanding these interactions, healthcare providers can better tailor interventions that take into account the unique microbiome composition of each patient, leading to improved therapeutic strategies and individualized patient care [13, 14, 15, 16].

Integration with genomics and personalized healthcare

The diagnostics of infectious diseases have significantly benefited from the integration of advanced genomic technologies and-more recently-from a more expansive definition of precision medicine that incorporates not only host and pathogen data but also comprehensive microbiome data. During an outbreak of an infectious disease, obtaining genomic sequence information from the pathogen often takes precedence over the patient’s own genomic data when it comes to determining the most appropriate treatment strategies that can be implemented. A thorough and detailed temporal along with spatial understanding of pathogen transmission dynamics is also absolutely critical for designing and implementing effective containment measures that can help curb the spread of the disease. By utilizing pathogen genotypic information for not just the diagnosis but also the treatment and prevention of infectious diseases, it becomes possible to adopt and implement personalized approaches that are strikingly similar to those utilized in genomically driven precision medicine. Technological advances in pathogen nucleic acid extraction techniques and high-throughput sequencing methods have enabled these innovative concepts to be seamlessly translated into routine clinical applications. Moreover, the incorporation of machine learning algorithms can further enhance the accuracy and efficacy of these diagnostic methods, leading to improvements in patient outcomes [17, 18, 19, 20].

Limitations of traditional one-size-fits-all models

Various definitions of precision medicine abound, yet they emphasize personalization: diagnoses, disease mechanisms, treatments, or intervention outcomes individualized to the patient. The precision medicine framework is increasingly embraced for infectious diseases, as artificial intelligence (AI) augments stratification and decision-support capacities [21]. Traditional one-size-fits-all metrics, however, continue to predominate in clinical microbiome diagnostics. Pathogen acquisition risk, microbial community structure, and potential therapy responses are a few characteristics presently modeled with generic, population-averaged parameters [22].

These metrics-along with microbiome-guided pre-, pro-, and post-antibiotic treatments-form the basis of many clinical applications. Precision microbiome medicine addresses systemic gaps in the individual-level diagnostic strategies available for complex infections and other microbiome-influenced diseases. For infectious diseases in particular, the objective is to supplement current one-size-fits-all approaches with highly controlled, individualized guidance that draws upon a richer combination of microbial signals and attributes and encompasses pathogen-screening activity [23, 24, 25].

Chapter - 3

Infectious Diseases in the Age of Microbiome Science

How microbiome imbalances contribute to infection risk

The human microbiome, the community of microbes inhabiting the human body, governs numerous health aspects. A balanced microbiome contributes to immunity, hosts pathogen competition, and protects against infectious diseases. Conversely, a dysbiotic microbiome disrupts immune homeostasis, harbours pathogen reserves, and impedes successful treatment. These imbalances partly explain the high infectious disease burden, wherein existing pathogens exploit medically privileged body parts to proliferate unchecked, highlighting the need for improved diagnostics to aid efficient decision-making. Targeting one of the foremost infection risk factors, the World Health Organization selected pathogens accountable for a significant global mortality share as the priority list for diagnostic aid [26]. Such diagnostics favour the identification of at-risk individuals over the pathogens directly responsible for transmission.

Contemporary models in the field of microbiology suggest an imbalance-competitor-restoration triad as a fundamental framework underpinning the risk associated with pathogen development. This triadic relationship illustrates how a dysbiotic microbiota can lead to the elimination of competently controlled bacteria, allowing for the unchecked proliferation of infection-prone pathogens. Upon recruitment, driven by specific behaviors or environmental factors, a verified pathogen can proliferate without effective natural competition. Several classical examples, including notorious pathogens such as Clostridium difficile, urinary tract infections, and sepsis, aptly illustrate the complexities and nuances of this triadic relationship. These are not merely isolated incidents but rather fundamental representations that underpin microbiome-aware predictive approaches in modern research. The explanatory frameworks derived from these examples inform and guide the choice of microbiome signals, which are closely associated with infection risk. Furthermore, they illuminate the various modalities that render the datasets amenable to machine learning techniques. This involves methodically identifying the types of inputs required, as well as the necessary properties that these inputs should possess to provide accurate predictions and enhance our understanding of pathogen dynamics in relation to the gut microbiome [5, 27].

Pathogen-host-microbiome interactions

Understanding pathogen-host-microbiome interactions is crucial for the design of precision microbiome medicine strategies. These interactions can be direct or indirect, are highly dependent on host and microbiota characteristics, and influence infection outcome. Understanding them helps to identify microbiome-based diagnostic targets and informs the design of microbiota-targeted infection prevention strategies. Two classical infection prevention examples are C. difficile-associated disease and uropathogenic E. coli (UPEC) urinary tract infections (UTIs) [28]. The microbiome protects against recurrent C. difficile infection, while the presence of specific commensal strains appears to protect against symptomatic UPEC colonization. A third example highlights the potential use of C. albicans to monitor the risk of sepsis due to dysfunction of key bacterial genera [29, 15, 13, 30]

Examples: C. difficile, UTIs, sepsis

C. difficile, UTIs, and sepsis have been extensively investigated for pathogen-host-microbiome interactions, making them ideal case studies to illustrate the contributions of microbiome data to pathogen identification.

C. difficile, a common cause of antibiotic-associated diarrhoea, has a complex relationship with the host and microbiome. Antibiotic treatments can cause microbiome dysbiosis, leading to C. difficile overgrowth and recurrent infection. In cases of nosocomial C. difficile infection, strains from environment-colonizing patients may establish secondary infections in susceptible individuals. These infection cycles highlight the utility of microbiome information in predicting possible C. difficile infections, underscore the opportunity for AI support, and demonstrate the complementarity of 16S and metagenomic data [9].

Urinary tract infections (UTIs) are among the most prevalent hospital-acquired infections. The urinary microbiome varies significantly among individuals, introducing the possibility of microbiome-guided UTI diagnostics. Considering causality, UTI-associated pathogens are likely to overpopulate the urinary microbiome just before symptom onset. These dynamics establish a clear target for predictive modelling of UTI risk and demonstrate compatibility with 16S and metagenomic data.

Sepsis, frequently triggered by bacterial infection of the lung, abdomen, urinary tract, or skin, is the most severe and lethal form of hospital-acquired infection. The interplay among invading pathogens, host immune response, and microbiome modulation has been well mapped through microbiome profiling over the disease trajectory.

Chapter - 4

Microbiome Profiling Technologies

16S rRNA sequencing

Microbiome diagnostic applications of 16S rRNA data entail amplifying and sequencing selected hypervariable regions of the 16S rRNA gene via polymerase chain reaction (PCR). 16S rRNA gene amplicons are relatively resilient to storage conditions, freezers, and thawing, allowing the use of samples archived from polymicrobial soaking, dilute plating, and other preservation methods [31]. Certain physiological processes (e.g., antimicrobial treatment) and lifestyle factors (e.g., diet, probiotics, and other dietary supplements) curtail time windows throughout which 16S rRNA data remain biologically relevant. A variety of database-independent and database-dependent bioinformatic pipelines are available to process and analyze raw 16S rRNA sequence data, each with distinct performance characteristics, system requirements, and available computational resources and software.

The primary limitations of pathogen detection from 16S rRNA data largely stem from the uncertain source attribution of the microbiota that is being sampled and the imperfect selection of the hypervariable region during the analysis process. Sequences that are originating from pathogen 16S rRNA genes occasionally fail to cluster effectively with adjacent terminal restriction fragments in database-independent analyses, which complicates the interpretation of the results. Additionally, these sequences cannot consistently serve as reliable filters in database-dependent scenarios, leading to further complications and uncertainty in the detection process. Furthermore, some pathogen ribosomal operons may be found to differentiate multiple copies of the 16S rRNA gene per genome, which means that the compatibility with broadly useful bioinformatic workflows continues to decline and poses challenges for researchers striving for accurate pathogen detection. This intricate situation highlights the need for improved methodologies in microbial analysis [32, 33, 34].

Studying microorganisms that inhabit natural environments, extreme conditions, foodstuff, and living organisms has fascinated scientific minds since Leeuwenhoek and Pasteur first observed them. The microbiome refers to the microorganisms in a given environment, the interactions between these microorganisms, and the environment itself. The corresponding, high-resolution characterization of all genomic material derived from a community of microorganisms is defined as metagenomics. Over the past two decades, advances in nucleic acid-based sequencing technologies have inspired the development of multiple microbiome-profiling methods; notable progress among these scientific efforts has driven a surge of microbiome research investigating host physiology, disease aetiology, biogeochemical cycling, and ecosystem functioning [1].

Microbial diversity and functioning can be assessed through characterizing microbiome composition, metagenome content, and metatranscriptome quantity. Characterization encompasses ad hoc isolation plus cultivation or deploying cultivation-independent approaches. The former relies on inoculating specimens onto predetermined selective media and subsequent growth to one or more colonies by free propagation. This process broadly delivers the identity of predominant community members using classical microbiological techniques and preserved isolates for further screening. Assembly and reconstruction via currently available genome-resolved metagenomics techniques can yield genome sequences for suspected isolated microbiome members. Although these pioneering methods have greatly accelerated the understanding of microbial diversity and ultimately bacterial taxonomic identification, taxonomic classification remains behind biogeochemical, ecosystem-functional, and ecological-function characterization of the microbiome [2].

The term microbiome refers to all microbial genomes present in a given habitat [1]. A metagenome is the ensemble of all the microbial genomes in a given community. It can be many times larger than the metagenomes of individual taxa combined, due to the presence of dissimilar and incomplete genomes. Community profiling quantifies taxonomic or functional metagenomic composition to assess community change. Profiling remains popular because accessibility is high and complementary-current sequencing costs favour larger numbers of samples over greater depth, and light rain is straightforward to assess. Nevertheless, specific analysis of metagenomic, metatranscriptomic, or metaproteomic sequences gathers further insights of primary ecological and molecular interest.

Microbiome-sampling efforts date back to 1991, when microbiota-tailored nucleic acid extraction and amplification lowered contamination and amplification bias risk. The Scientific Foundation “Microbiology Grant No. 5905” propelled initial extension of cultivation-independent specimen characterisation beyond the 20-character ribosomal DNA target and machine-learning Rectors towards comprehensive empirical water characterisation through automated sequence-based extraction of total and functional water profiles. The 1990s saw taxonomic-objective 5S amplification-internal standards and custom automated extraction-transcription-by-analytical copying of prev-XXXX removal proceed, followed by blend-in replacement of macro-organisms by micro-organisms [2].

Microbial communities have colonized and influence practically every ecosystem on Earth, impacting environmental sciences, agriculture, and human health. The field of microbiome science or microbiomics focuses on microbial composition, diversity, and function as they interact with environmental features. Microbiomics has become a large interdisciplinary field incorporating microbiology, chemistry, metatranscriptomics, metaproteomics, metabolomics, cultivaromics, ecology, phylogenetics, systems biology, and much more. It is rapidly growing, leading to a rise in novel analytical approaches to parse, curate, and analyze multi-omic data. These tools and data-types make it increasingly difficult to interpret, compare, standardize, and benchmark the quality of data and methods in a consistent way. These issues confound the ability to translate multi-omics research into clinical applications. Precise computational techniques are needed to process, normalize, and analyze microbiomics datasets to support reproducible research into the role of microbiomes in health and the environment.

Measuring the composition of a microbiome is addressed by amplicon-based or marker-gene sequencing approaches to perform a microbial census. The functional potential of a microbial community can be inferred indirectly by marker-gene surveys, or through direct observation of functional genes and pathways by whole-metagenome sequencing surveys. Measurements of functional activity in a microbiome can be derived through metabolomics, proteomics, and transcriptomics [3].

Amplicon sequencing of variable-length, hyper-variable regions of the 16S rRNA gene or the internal transcribed spacer regions (ITS) of rRNA operons allows assessment of microbiome taxonomic composition in a single PCR amplification [4]. In particular, choice of 16S rRNA-targeted primers influences microbiome composition estimates, especially in prior samples that lacked enriching bias control. Both amplicon length and choice of ITS region shape community structure recovery. Use of variable-rich regions or super-primer strategies enhances bioinformatic usability but amplifies primer-specific recovery bias. Finally, incomplete 16S rRNA-containing genomes and tRNA-rich regions in some groups impede precise quantification of taxonomic units at certain levels.

Metagenomic shotgun sequencing of environment origin nucleic acids enables un-biased analysis of taxonomical and functional genomic constituents. For assemblages comprising only one or a few organisms, efficient full-genome assembly permits taxonomic identification and functional-potential characterization. From nucleic-acid extraction to sequencing in less than six hours and total recovery of high-throughput sequencing generated data and clustered communities, a simple-deduplication approach competes with premature and sample-bias-propagating 16S rRNA gene-targeted strategies. Since shotgun-microbiome construction nor recovery steps guarantee assembly-continuity, metagenomic community data serve as operational taxonomic unit presence-absence tables, like amplicon-16S profiling. Nevertheless, relative-abundance unaffected by initial-microbiota complexity, origin exclusion, or cell-growth altering extraction methods still observes reproduction of previously claimed temporal and site-variation observations at phylogenetic and metagenomic-eubacteria resolution.

Amplicon sequencing technology has played an essential role in profiling the composition of microbial communities and is thus a widely applied method for microbiome studies. The general principle of amplicon sequencing is to generate sequencing libraries through the amplification of specific regions of interest via the polymerase chain reaction (PCR). For microbiome analysis, the most commonly targeted regions are portions of the 16S ribosomal RNA (16S rRNA) gene for bacteria and archaea, the internal transcribed spacer (ITS) regions for fungi, and other specific loci for other microorganisms. Amplicon sequencing delivers a high level of accuracy for characterizing microbial community composition at the phylum and genus ranks. The chosen marker region clearly determines the maximum taxonomic resolution that can be achieved and different regions generate various coverage biases and taxonomic markers, which can lead to dissimilar community structures when identical samples are characterized. Extensive studies have validated amplification primers for different marker regions, improving taxonomic coverage and minimizing bias. The U.S. National Institute of Health has established a 16S rRNA gene sequencing protocol, with widely adopted primers to amplify a region of approximately 500 base pairs [5]. Specific protocols for eukaryotic ribosomal RNA and second-generation sequencing technologies have been developed, enabling the characterization of various microbiomes, including soil, rhizosphere, and plant samples [6]. A distinct aspect of amplicon sequencing lies in its filling-sequence capability, where a central region remains unblocked and elongation can proceed, allowing multiplexing with various barcodes and the generation of sequence data for the forward and reverse primers [7].

Profiling methods based on metagenomic shotgun sequencing collect nonspecific gene- or taxon-centric data without relying on amplification of target gene sequences. The functional composition of the healthy microbiome as well as shifts associated with various diseases, especially within the gut, are now approaching cross-sectional completeness. These data underscore the importance of sequencing depth, which should be calibrated according to desired resolution at the functional as well as taxonomic physiotype level. Collectively, however, major shortfalls still need to be addressed ahead of microbiome-guided diagnostics or therapeutics, including link-from-function to health and disease states, extensive accuracy validation of functional predictions, improved representation of species from underrepresented phyla in reference collections, and integration of diverse other -omic data layers.

Temporary sampling may also enable metagenomic shotgun sequencing of short, low-budget, or easily cultured investigations of portable devices such as eDNA cages or automated ultrasound systems. Nevertheless, outside these highly constrained conditions, the accumulation of public sequencing data currently far outweighs that of public raw short-read data, and the development of tailored portable devices with proven comparative classification performance has not yet kept pace. For amplicon sequence classification, accurate training and testing of machine-learning classifiers remains critically dependent on the incorporation of commercial Noise-Nextclade cloud support and detected bias or limitation zones.

Of diverse microbial phyla present in the gut, only a small fraction is usually active, though activity levels vary among individuals. Functional characterizations solely based on metagenomic data may therefore be misleading. Capturing gene expression profiles through metatranscriptomic sequencing determines whether particular species and functions are actively expressed. Information on transcript levels further enables inferences about the relative activity of organisms and pathways within the community [8]. Integration with metagenomic, metaproteomic, and metabolomic datasets allows reconstruction of the complete microbiome activity biocycle, wherein gene-centred analysis highlights transcripts governing degradation, turnover, or absorption of specific substrates; metabolomic data indicates substrate supply; and metaproteomic observations provide complementary, verification-oriented support for genome-centric interpretations [9].

Proteomic and metabolomic data augments gene-centric activity inferences by providing a pathway-level view of microbiome functioning. Metaproteomics resolves the highly multifactorial taxonomic diversity of microbiomes but, due to high costs and other technical constraints, remains relatively niche. Metaproteomic-based profiling attempts have been attempted in marine ecosystems and associated with disease recovery transitions in biological wastewater treatment plants utilizing a diet shift to gluten-free food and reducing transplant-associated diarrhea incidence. Such focused applications may become more common as methods are continuously refined and an increasing body of publicly available sequence databases of noncultured microorganisms accumulated. Integrative studies comparing existing metabolomic profiles with metaproteomic data may also help identify the specific bacterial activity in metabolic reprogramming.

Third-generation sequencing methods and nanopore sequencing devices offer the unique ability to sequence nucleic acids in real time and generate ultra-long reads of hundreds of kilobases or even megabases that can completely span entire plasmids. The ultra-high temporal resolution offered by portable sequencers enables point-of-care microbiome profiling for various clinical and non-clinical applications. Such hold time resolution capability has the potential to provide relative quantification of rapidly growing bacterial pathogens in screening samples for wound infections and urinary tract infections at point-of-care. The success and deployment of these technologies, however, remain constrained by sequencing error rates, particularly with nanopore sequencing, potential amplification bias during library preparation, and the relevance of sequencing readout for clinical decision making.

As a conceptual extension of metagenomic shotgun sequencing, Hi-C-an approach originally designed for intact eukaryotic genomes-enables genome-resolved metagenomics to aggregate fragmented prokaryotic assemblies into genome bins by inferring proximity from, for example, unique read pairs and mate pairs that originate from the same genomic locus [10]. Metagenomic assemblies, therefore, can be linked spontaneously into genome bins through Hi-C or cross-platform technology combinations (e.g., PacBio long-read assemblies linked with Hi-C). A linked assembly or binning scheme clarifies subsequent interpretation by delineating which loci coparticipate in the same biological cycle. Access to, or interpretation of, Hi-C data may benefit from referring to ‘Metagenomic shotgun sequencing’ and ‘Hi-C and genome-resolved metagenomics’ when strategies aiming for multiple information layers are considered [1]. Technologies for single-cell genomics can likewise pinpoint biological activity in specific community members, yet highly targeted amplification schemes usually yield limited genomic or transcriptomic input per event, imposing throughput/cost trade-offs on larger collections. Conveying single-cell approaches is enhanced by two-way cross-referencing with ‘Metagenomic shotgun sequencing’ and ‘Single-cell genomics’ to highlight similar target organisms or complementary interpretative architectures.

All Hi-C experiments are based on the creation of large chimeric DNA molecules. Individual genomes are cross-linked to preserve their spatial organization. Genomic DNA is then enzymatically fragmented and biotinylated. Biotinylated ends from different DNA fragments are subsequently joined together, forming a mixture of endogenous chromatin molecules. A sequencing library is prepared and sequenced, and the resulting chimeric reads provide information about relative proximity between DNA loci. When the process is performed on bulk microbial communities, the mixture consists predominantly of DNA from multiple organisms. The participating species can be identified, and the likelihood of interconnectedness between each set of two taxa can be quantified. This type of information is exploited in MetaHiC to cluster or bin contigs [11]. Clustering combines contigs into groups preferentially containing several pieces from the same organism, while binning assigns individual contigs to distinct taxa groups. The approach does not yield complete genomes but enables the accurate reconstruction of substantial portions of tens to hundreds of individual genomes [12]. Meta3C, a variant of Hi-C targeting only a few abundant species, allows complete genomes to be assembled from highly complex communities [13]. A major limitation of metagenome-analysis approaches characterized by unconnected contigs is that two separately binned pieces may originate from the same genome, leading to misinterpretation of evolutionary and ecological processes. Various concepts related to community assembly, multifactorial analysis of community distance metrics, evolution of mobile genetic elements, community-wide phylogenetic trees, and metabolic networks depend on the presence of contig connectivity. These Hi-C methods solve the long-fragment problem addressed by other approaches, which generally output shorter contigs at more moderate coverage.

Xenobiotic degradation, plant nourishment, and even pathogenesis are influenced by ecological interactions of anaerobic arginine-degrading bacteria in anaerobic wastewater treatment systems. Inactivated catabolism or microbe-microbe cross-feeding of aromatic amino acids impedes anammox. Environmental sciences and bioremediation strategies targeted at small deodorizers, paints, rubber, pharmaceuticals, personal-care products, or household products are consortia comprising hydrocarbon-degrading Bacillus, corynebacterium, cometabolic Rhodococcus and Gordonia, and aromatic-compound-attaining

Microbacterium. Microbiomes of synthetic-biodegradable polyesters characterized. Use of bioplastics and biodegradable plastics naturally derives from carbon-neutral alternative substrate or, more generally, nonpetroleum feedstock. The economical sustainability along with contamination of such a system where co-incorporated plastic-fixating bacteria [14] improves aerobic degradation of polylactic acid versus using a species amalgam without, can be a promising prospect.

Emerging platforms for profiling microbiomes and ecosystems embrace several concepts, including community composition and gene function, through many promising investigational approaches. Nano- and third-generation sequencing technologies offer compelling advantages for microbiome and ecosystem profiling, including much longer read lengths and real-time data acquisition. Current applications encompass whole-community profiling in water and soil samples, targeted analysis of clinically relevant pathogens in human and food samples, portable epidemiological analyses for airborne SARS-CoV-2 data in public transportation, and early-stage study of the oral microbiome. Numerous limitations remain, including limited depth and bioinformatic analysis capabilities, and concurrent validation with established platforms is essential for method and data interpretation.

At the cutting-edge of sample handling and analysis, portable or point-of-care profiling technologies have received rising interest. Recent applications employ a wide range of sequencing technologies, including second-generation nanopore sequencing combined with portable personal computers to explore oral microbiomes and microbial links between electric vehicles and underground electrochemical environments. While the market for pack-sized benchtop sequencers has grown rapidly in diverse laboratories, read length and sequencing throughput remain paramount. Most applications rely exclusively on long-read sequencing, on which extensive genomic and transcriptomic analyses have focused, and many established protocols remain fruitful solo endeavours [1].

Emerging nano- and third-generation sequencing platforms that offer long read lengths and continuous data acquisition open new opportunities for portable, point-of-care microbiome profiling. Third-generation sequencing platforms produce reads up to 20 kb long, facilitating assembly of repetitive regions and structural variation. These sequencers provide rapid, real-time data in a compact, easy-to-use format, critical for outbreak surveillance or ecological monitoring. Bioinformatics frameworks such as MinION2Metagenomics streamline analysis by translating raw signal data into fastq files for taxonomic and functional characterization [15]. Last-generation platforms integrate sequencing, amplification, and library preparation into a single disposable flow cell capable of on-device analysis and display].

Portable and point-of-care microbiome profiling. Current ultra-short and portable DNA sequencers present enticing opportunities for portable and point-of-care microbiome profiling. Potential case studies include human microbiome analysis and monitoring of environmental microbiomes in cases of predecessors to pandemics. The technique remains in its infancy and requires further validation before practical applications can be safely considered.

Emerging third-generation sequencing technologies, such as those from Oxford Nanopore and PacBio, offer read lengths up to 100 kbp and the distinct prospect of real-time access to data. Although nanopore sequencing is currently associated with lower throughput and higher error rates than Illumina sequencing, a considerable number of study designs are compatible with a trade-off between immediate data access and other metrics [1, 16].

Taxonomic classification refers to the computational assignment of observed sequence-based operational taxonomic units (OTUs), unsequenced reads, or metagenomic contigs (metagenome-resolved coverage) to biological entities at a specified taxonomic clade level [17]. Generally, classification aims to satisfy a specified relative accuracy criterion while optimizing for coverage and computational efficiency. Functional annotation denotes the assignment of observed sequence-based sequence features or metagenomic contigs to statistical models of putative biological functions, with a possible secondary goal of pathway inference. Comparative analysis refers to the derivation of statistically significant differences in taxonomic or functional profiles from two or more groups of community profiling samples, subsequently characterized by associated metadata [18]. Longitudinal analysis designates the extraction of statistically significant changes in taxonomic or functional profiles within the same community profiling sample over time. Data integration involves the combination of correlated measurements from distinct community profiling datasets acquired at the same time and environment or from different but temporally matched instruments. These considerations are closely related to the design of experiments and interpretation of results, which incorporate a wide variety of interdisciplinary factors. Despite the availability of data integration across omics layers, taxonomic classification remains the major enrichment and exploratory analysis complement for amplicon datasets, and cross-sectional metagenomics subsequently suggests gene-level functional insights about bacteria in subgingival plaque microbiomes without accessibility to direct metagenomic reads.

Taxonomic classification is the assignment of sequences to taxonomic groups according to a reference database of similar sequences. Clade resolution extends classification by assigning each group to a clade in a hierarchical phylogeny that may not correspond to the taxonomy of the reference database. PhyloRelief identifies taxa that distinguish different sets of metagenomic samples [19]. To overcome reliance on an external taxonomy, it applies phylogenetic relationships of the taxa contained in the reference database to group them into clades and ranks those clades according to their contribution to sample differentiation.

Taxonomic classification of microbiome community profiles is frequently combined with functional annotation of the corresponding metagenomic samples. In this case, activity profiling focuses on metagenomic features, whereas pathway inference seeks to establish links between functional modules. Comprehensive comparative analyses-including longitudinal studies-are essential to reveal community dynamics associated with a given condition. These inferences can be enriched through integration of cross-omics data, such as multi-kingdom metagenomes and metatranscriptomes, metagenomes and metaproteomes, or metagenomes and metabolomes. The resulting views of community activity are complemented by complementary microbiome studies, including systematics, amplicon library profiling, and shotgun sequencing of purified genomes from single cells or contiguity-based linking of Hi-C clones from cultivated resources [1].

Functional gene annotation and metabolic pathway inference employ various bioinformatics approaches, each with unique strengths and limitations. Existing metabolic databases, which assign functional annotations to genes, provide a simple but imperfect means of inferring pathway-level functional potential across a community. More advanced methods, such as Kaiju and DIAMOND-based approaches, enable clade-specific functional annotation, while functional network construction scaffolds subsequent activity-driven analyses. Gene set enrichment analysis and pathway activity score calculation provide inferential shortcuts based on known biological activity across the metabolites represented within a profile.

In addition to kingdom-level detection of eukaryotic sequences, clade-specific annotation approaches add ecological value, especially where cryptic Eukaryota are present. Despite its biological complexity, the functional properties of the eukaryotic community remain less actively interrogated than their prokaryotic counterparts. Metabolite profiles have been applied to such communities, but derived pathways often represent basic building blocks of life. For investigations where the metabolic profiles of Fungi, Animalia, Archaea, or other kingdoms are key, a clade-resolved activity-framework combining multiple -omic layers would provide superior interpretive resolution. Attributes essential for comparative or longitudinal analysis of metagenomes and their functional signatures across environmental or health-related samples are also vital for evaluating temporal shifts during targeted time-series studies. The inferred function of integrated-deep sequencing-based profiles can be strengthened through deliberate connection with periodic -omic reports.

Compiling microbiome data into regulatory frameworks that enable comparative and longitudinal analyses further enhances their interpretive value [12]. Taxonomic classification-the assignment of each sequence to a lineage based on similarity to reference databases-provides the foundation for such analyses. Phylogenomics links sequence information to a structured reference tree. Microbiome profiling generally introduces sufficient diversity that metagenomic assemblies reliably recover core isotopic units from an emerging set of Universal Common Ancestors [1]. Sequence similarity within these January 18, 2021 83 Assemblers 78 frameworks helps establish ignorance of sequencing condition and preserves core insights even when taxonomies switch or evolve. Several systems now enable core facility access to phylogenomics, providing a viable technical option for microbiome datasets.

Another widely used facility targets functional characterization, assigning sequences to gene families based on their sequence and motif similarity. This approach, however, can obscure biological specificity because gene families are generally broader than the equivalent lineage measures already deployed. Instead, statistics on the component building blocks of assembly and their shared compositional patterns within individual contigs or fragments-abundance stratified by building-block oligo-length or other widely varied building-block plug-in token frequencies-better capture the sameness of isolative specification for subsequent Rousseau or other ulterior exponential growth modelling. Supporting data on biomass per strand conversion keeps metagenomic calculation simplicity while directly linking back to building block oligo stepping lengths and their size-discrepant active-rich group generators, ultimately probing a remaining proxy of equivalent isolation beyond banded whole-probe embedding for cross-core inter-system insights.

Simultaneous profiling of multiple biological layers presents opportunities to decipher complex community interactions, yet straightforward methods for integrating info across diverse -omics types and longitudinal datasets remain scarce. Popular approaches incorporate metagenomic and metatranscriptomic data. Transcript/gene ratios indicate transcriptional activation or repression of particular genes; a metagenomic database assists metaproteomic protein identification; and gene abundance/proteomic activity connections emerge when comparing shotgun metagenomics with metaproteomics. Foxtail millet rhizospheric soil and microbial communities tagged with metagenomics plus 16S rRNA amplicon sequencing enable co-variation exploration of microbial functions and 93 environmental factors [20].

Integrating multi-omics data across different biological layers helps characterize a microbiome's collective function and metabolic activity. Such integration is crucial for diseases-e.g., Type 1 diabetes, diabetes, or inflammatory bowel disease-that exhibit microbiome alterations unlikely linked to single bacterial species or individual molecules. Because 16S rRNA/ITS amplicon sequencing and shotgun metagenomics continue driving microbiome analyses, assemblers and pipelines tailored to long reads, multi-omics compatibility, and specific -omics combinations are in demand [21].

Human health is tightly integrated with microbiome composition and status. The gut microbiome of individuals suffering from inflammatory bowel disease, irritable bowel syndrome, or obesity exhibits notable deviations from healthy counterparts. The oral microbiome serves as a reliable indicator for periodontal disease and certain systemic diseases. On the skin, elevated allergen-specific immunoglobulin E and disease-indicative inflammatory cytokines correlate with eczema-associated communities [1]. Specimens of the gut, oral cavity, and skin deliver sufficient microbial DNA for various profiling approaches.

Profiling methods have unique strengths and weaknesses for health-related diagnostics and therapeutics. Amplicon sequencing, a cost-effective option with limited taxonomic resolution, proves suitable for the gut microbiome; oral and skin-site profiling require shotgun metagenomic sequencing or higher-precision approaches. Shotgun and full-length 16S rRNA sequencing are deployable across all body sites; the former is favored when functional insights are beneficial or for systems exhibiting substantial strain or genus diversity. Combinations of amplicon and shotgun sequencing leverage the former's low cost with the latter's functional characterization and strain-resolution potential. To develop microbiome-guided diagnostics and therapeutic microbial applications, accelerated R&D cycles demand integrated amplicon sequencing and shotgun metagenomics complemented by cultivability and resistance profiling.

The gut, however, is the best-studied human microbiome, changing over periods of days, childhood, and aging and in response to diet or illness. It is modified in unknown ways by travel, chemical additives, and antibiotics [22]. Although no two people share identical microbiota and microbiomes vary between host species, studies of wild-type vertebrates such as mice and zebrafish have identified similarities in microbiotic function and host interactions. Germ-free organisms have elucidated the microbiota's roles in maturation of the immune system and, surprisingly, anatomical development of the intestine. Gnotobiotic systems in which an organism's natural microbiota are replaced with the human analog are also being used to study the phenotypic consequences of controlled perturbations. High-profile demonstrations of this technique include an ongoing investigation of the microbiome in obesity.

Growing interest in the microbiome and its association with human health and disease has resulted in characterization of the gut, oral, and skin microbiomes across numerous diseases, enabling microbiome-guided therapeutics and diagnostics. Example disease states include irritable bowel syndrome or Crohn's disease, where fecal microbiome transplant using metabolites as functional proxies shows therapeutic effects; periodontitis, which can be diagnosed based on the oral microbiome; and skin diseases like atopic dermatitis and psoriasis, which display dysbiosis patterns. Guided by these associations, amplicon sequencing mainly targeting 16S rRNA and general function primers has been the dominant method due to its cost-efficiency and applicability in population-scale studies. However, detection of these associations requires high sequencing depth to discover rare taxa, and the limited resolution of the 16S rRNA marker leads to false-positive signals. As a result, long-read metagenomic shotgun sequencing is being applied to disease cohorts and proposed as the standard for microbiome-characteristic studies.

Recent advances in culturomics, machine-learning-based classifiers, and standardization approaches are further improving the resolution and reproducibility of microbiome diagnostics. With these developments, the gut, oral, and skin microbiomes are showing connections to medicine, forensic science, nutrition, and personalized medicine, and the resulting products are starting to reach the consumer market.

Agricultural microbiomes include those associated with soil, plants, and animals. Seed and root-associated microbiomes are of special interest because they affect plant health and influence the microbiome assembly of tissues further downstream [17]. Microbiome monitoring of wastewater treatment processes and ecosystem surveillances can inform decision-makers and guide public policies on water safety, pollution, and ecosystem health [23]. Soil contamination and the loss of soil biodiversity through erosion compromise nutrient cycling and crop yields. Monitoring microbial community diversity helps identify best practices for soil maintenance.

Plants form associations with diverse assemblages of microorganisms that affect growth, nutrient acquisition, and resistance to biotic and abiotic stress. A large proportion of rhizosphere-associated bacteria neatly establish microscale biofilms on root surfaces, which contain active colonies. In roots, endophytic communities thrive in tissues such as epidermis and hairs, while other locations remain inaccessible due to highly impermeable barriers. Disentangling individual metabolic contributions in plant-associated microbiome species is difficult but crucial due to often subtle phenotype effects and multi-species dependencies. As single-cell transcriptomics enables large-scale genome-wide exploration of individual microbial functions in plants, the bottleneck remains studying complex whole-community transcriptomes that obscure species-level functions [24].

Biota regulate essential ecosystem services in soil. Changes in these processes affect terrestrial C stores and aggravate climate change. Nutrient availability and plant growth are crucial for sustainable agriculture worldwide. Biogeochemical cycles of C, N, P, K, S, and Fe dominate nutrient saturation in soils. Soil microbiomes participating in these cycles span diverse genera and species that remain poorly understood across multiple systems. Tracking phylogeny, composition, and structure in soil microbiomes is vital for managing ecosystem health and agricultural productivity. Temporal variations in soil microbiomes remain understudied, impeding understanding of their public health impact and isotope technology for ecosystem-wide nutrient cycles at species resolution. After experimental separation of microbial fractions and nano-enrichment of genomes, high-throughput sequencing characterizes plant-associated organisms and explores direct roles in mineral-weathering and gas-fluxes across soil microbiomes [25].

Historically, microbial parameters have been used to monitor environmental pollution since the 1980s, focusing initially on drinking water quality. These early frameworks were rooted in the fact that aquatic ecosystems of concern were generally subject to unprocessed faecal discharges or uncontrolled effluents from large wastewater treatment plants (WWTPs). In the intervening decades, specific bioassays based on the physicochemical changes caused by one or multiple toxic chemicals, along with whole-cell biosensor devices employing living organisms for biological anal - yte detection, have become available. Bacteriophages and viruses have also been moved into the spotlight as highly appealing proxies for wastewater tracking and control. Detection methodologies for enteric pathogens have evolved to increasingly sensitive nucleic-acid-based methods, primarily quantitative PCR [26].

Nonetheless, these methodologies suffer from limitations that hinder the identification of sources and the assessment of pollution state. False positives may arise from the persistence of non-viable cells or naked DNA in water samples. Viable but non-culturable states may not be accounted for. Meanwhile, constant exposure to a major stressor can drive the loss of essential pathogenic traits, and toxic organic pollutants have been shown to accelerate plasmid loss. One of the main challenges is therefore the development and implementation of sensitive in-field monitoring tools for pollutant impact and sublethal effects. Within eco-toxicology, much of the methodological innovation has focused on advanced molecular frameworks-particularly those that fill the gaps left by classical microbial techniques-designed to clarify the principal drivers and pressures acting on aquatic ecosystems [27].

Technical considerations in microbiome profiling center on three factors: sequencing depth, sequencing cost, and desired resolution. The deeper the sampling, the more comprehensive the coverage of both abundant and rare taxa [1]. On the practical side, strategies exist that make shallow profiling viable when the budget does not allow extensive community surveys. This is especially true in environments in which only a fraction of the present taxa is expected to be in the samples. In such cases, the cost of sequencing can be kept low while still providing content-rich, informative community coverage. Sequences illuminating the relevant community structure can return larger insights than an ill-planned or poorly designed high-coverage survey.

Microbiome profiling is characterised by practical trade-offs between sequencing depth, overage, cost and information content. Microbiome studies typically contain many samples, each of which can be characterised by low-cost, low-depth shotgun sequencing at ~ 101-105 reads per sample. Lower sequencing depth does not necessarily entail lower information content; at full width, the information content of shallow shotgun metagenomics studied at ~ 103 reads per sample ends to be greater than that of deep amplicon sequencing. Cost-effective amplification-free and primer-free approaches to mapping the microbiome are expected to become increasingly feasible to complement amplification-based strategies.

Sample numbers and sequencing depth strongly determine cost per sample. For microbially dominated samples, deep single genome and multi-genome shotgun sequencing of ~ 105-107 3 -HTs or < 104 3 -HTs are typically applied, respectively, to obtain many large, isolated genomes and maximise information [28]. 107 whole-genome shotgun reads provide about the same information as < 106 metagenomics 2 -HTs. Similar considerations also apply to metatranscriptomics and when data on proteomics or metabolites can directly address microbiome questions.

The microbiome is the collective assemblage of microorganisms in a defined environment, particularly the human body [29]. Characterisation of an individual's microbiome raises privacy concerns because such information has the potential to reveal sensitive data about the individual. Microbiome information may serve to further elucidate an individual's susceptibility and predisposition to disease, linking such information with factors like obesity that are already well-studied in the field of human genomics. As many diseases have a clear microbiome signature, like inflammatory bowel disease, microbiome measurements could also inform on a patient’s future health trajectory. The concern is amplified by the fact that, alone, information concerning an individual’s genetic makeup does not allow precise and comprehensive estimation of a person’s life expectancy or likely future health conditions. Microbiome-based information, on the other hand, especially when merged with genomic data, could lead to even more informed and possibly harmful insights about factors such as socioeconomic status and exposure to certain geographical areas and countries. Such data could lead to a more granular and even discriminatory understanding of a person’s microbiome; in some societies, predictions about one’s caste or class could be made based on the information potentially contained in microbiome analysis. A further level of complexity accompanies the domain of participatory research and citizen-science projects. The public may be willing to participate in such frameworks, willing to contribute to art, science, or other realms of knowledge. But there are still challenges: individuals may cede control over their data but remain highly concerned about how it is to be used subsequently [30]. Finally, there are still pressing issues regarding equity. Microbiome research and understanding has advanced rapidly in developed parts of the world without engaging directly with work conducted elsewhere. Microbes shape human health and human health shapes microbial evolution. Consequently, research conducted in a global context cannot be concerned solely with one region. To aid in sustaining the necessary equitable flow of funding between regions, projects need not be exclusively urban-centric; smaller communities can also form rich veins of scientific study and discovery, permitting a “many-to-many” approach to research strategies. Highly localised efforts with several active research connections can therefore facilitate patterns of outreach and collaboration across increasingly broad networks. To engage on a wide scale with entirely different ecosystems, data-sharing protocols remain essential. Maintaining respectful values around the flow of information - whether scientific, political, social, or cultural - can nurture and safeguard the multifaceted process of building competency and understanding.

Despite widespread interest in microbiome research, reproducibility remains a critical issue in sequencing studies. Disparities in the taxonomic or functional profiles of nominally identical samples can emerge from even minor variations in preparation or sequencing protocols [1]. To address these challenges, it is essential to systematically document input material properties, wet-bench steps, sequencing methods, and bioinformatics analyses. Control reagents, such as mock communities, enable labs to characterize protocol performance and monitor consistency throughout the workflow [31]. Community profiling thus reflects shared technical and ethical considerations in biotechnology, where data privacy and equitable access must accompany scientific advancement.

Microbiome profiling continues to evolve at a rapid pace, and several core themes emerge from the progress to date. First, the spectrum of available technologies and methods has matured greatly, addressing the desiderata identified in recent surveys [1] and meeting specific needs in response to emerging biological questions. Second, the integration of longitudinal and cross-omics datasets is gaining traction and increasingly informing microbiome studies, while ongoing efforts to establish clearer standards and protocols will continue to enhance comparability and reproducibility [17]. Third, tomorrow’s microbiome innovation will follow the trajectory of today’s research, concentrating on more elaborate applications in environmental monitoring, wastewater surveillance, and agriculture, targeting both fundamental understanding and preventive or therapeutic options.

The transition from discovery-driven to implementation-oriented microbiome studies entails a fresh examination of technical and ethical considerations. As the range and sophistication of profiling technologies broaden, careful attention to privacy, consent, and equitable access will help build and sustain public trust, while standardization of procedures and documentation will facilitate responsible stewardship and dissemination of information. Throughout this international effort, the microcosm of the microbiome remains a constant source of wonder, reshaping planets, transforming ecosystems, and influencing processes and evolution on scales far beyond what any individual species can accomplish.

Metagenomics and metatranscriptomics

Metagenomics enables a comprehensive analysis of all microbial DNA present within a given sample, providing a broad and detailed view of the various constituents, communities, and functional potential of a complex microbiome. Similar insights can be gleaned from metatranscriptomics, which specifically focuses on the diverse RNA repertoire. Both of these cutting-edge techniques represent immensely powerful tools for understanding the intricate influence of the microbiome on overall health and for feeding advanced predictive models aimed at diagnosing or anticipating infections effectively. Combining the insights gained from these methodologies could lead to groundbreaking advancements in microbiome research and its applications in healthcare [35].

Multi-omics integration

Integrating data from multiple omics technologies, including genomics, transcriptomics, proteomics, and metabolomics, offers a much more comprehensive and nuanced understanding of overall biological states. This integration is crucial as it enhances the interpretation of host-microbiome interactions, leading to insights that are not possible when examining any single layer of biological information alone. Precision microbiome approaches, therefore, necessitate a sophisticated data fusion process that harmonizes diverse molecular profiles from these omics layers, creating a more complete picture of biological systems. For instance, a study conducted by Heintz-Buschart *et al.* explored patients suffering from type 1 diabetes utilizing a combination of metagenomics, metatranscriptomics, and metaproteomics techniques. Their findings from metagenomic sequencing, interestingly, showed that there were no significant shifts found in the microbial composition following the onset of the disease. However, the transcriptomic analyses conducted as part of this comprehensive study revealed noteworthy insights, specifically an enrichment of differentially abundant transcripts during this condition. These findings underscore the importance of integrating multiple omics technologies to capture the complexities of biological responses accurately. Similar guiding principles have shaped additional research efforts that integrate recoverable gut microbiome functions with various other modalities, aiming to model intricate multi-omics association patterns effectively. More broadly, machine-learning methods have been increasingly applied to both evaluate different multi-omics data integration techniques and develop cohesive strategies that explicitly incorporate biological knowledge, thus facilitating richer analyses of complex biological systems [35, 36, 37].

Chapter - 5

AI in Infectious Disease Diagnostics

Machine learning vs deep learning approaches

Machine Learning and deep learning are prevalent methodologies in data analytics and artificial intelligence research. While the distinction between these two approaches is often blurred, understanding their differences is critical for selecting the appropriate modeling technique. In the fields of pathogen identification, machine Learning and deep learning methods exhibit distinct differences regarding the quantity and nature of training data required, complexities involved in model interpretation, and the variety of microbiome data signals leveraged. Accordingly, the most suitable approach to employ depends on the specifics of the acquisition process and the characteristics of the microbiome features involved [38].

In the domain of microbiome-based diagnostic systems, pathogen identification often focuses on the species or strain of the infectious agent present in a biological specimen; the likelihood of specific resistance genes; or both. These outputs can be predicted from a range of dataset types, including sequence-based, Biosensor, and gene-resistome observations [36]. The data collected through these methods can support either machine Learning or deep learning frameworks, depending on the intended feature set and temporal constraints. For example, biosensor datasets generally yield signals that are more amenable to machine Learning methods, while the higher dimensionality of extracted features from genomes and resistomes linked to metagenomics or gene-resistome datasets is more suited to deep learning approaches. Moreover, algorithmic intricacies associated with employing deep learning necessitate larger sample sizes for effective training, rendering machine learning the favored option when the data acquisition timeline is condensed [39, 40, 41, 42].

Predictive modeling for pathogen identification

The human microbiome serves as an unexplored reservoir of information to predict microbiome-dependent pathogen infections and guide antibiotic selection. Recent work has demonstrated that infectious diseases alter the composition of microbiome communities. As such, microbiome sequencing data has the potential to inform pathogen detection, knowledge that can be efficiently incorporated into predictive models. The targeted output of these models includes species, strain, and antibiotic-resistance predictions, with analyzable features derived from metagenomic DNA sequencing and other unintrusive biosensors capable of quantifying metabolites, volatile compounds, and other by-products associated with human microbiome activities [39] By accelerating the identification of infectious pathogens, such models have the potential to inform and personalize treatment and improve patient outcomes.

Specifically, real-time monitoring of microbiome compositions through various sensors can facilitate the timely detection of community shifts brought on by pathogen-host interactions. Target species and specific interaction pairs have been linked to certain conditions, ranging from Clostridium difficile transmission and urinary tract infections (UTIs) to sepsis [43] Classifying these interactions allows models to tackle pathogen identification at the patient level and to evaluate infection risk on the basis of microbiome data along with spatiotemporal movement patterns. By combining these diverse predictive signals with data streams that cover pathogen spread through microbial, waterborne, and fomitic routes, a comprehensive system for infection monitoring emerges.

AI in outbreak detection and surveillance

Machine learning applications for infectious disease testing showcase healthcare disruptions and include outbreak detection and proactive surveillance. Healthcare records, demographic information, environmental data on humidity, temperature, precipitation, wind speed, and population density fuel forecasting models [9].

AI-based early warning systems are continuously scrutinizing vast amounts of data sourced from various online health portals, advanced environmental sensors, and sophisticated flow cytometers to identify potentially alarming microbial infection patterns that could pose serious risks. A stand-alone item is expertly connected to local and broader infectious disease reporting databases that effectively follow targeted pathogens of interest that merit close attention. Microbiological count databases meticulously track species associations, relevant modification timelines, and broader health screening links that are instrumental in assessing public health. Such advanced systems facilitate prompt and thorough public health investigations, which trigger essential biotechnological diagnostics that prioritize forthcoming species of concern that could impact community health [44, 45, 46, 47].

Chapter - 6

Integrating AI with Microbiome Data

Data preprocessing and normalization

Scaling and transformation serve as alternatives to rarefying [48]. Scaling adjusts the data relative to a reference sample, using approaches such as total sum, trimmed mean, geometric mean, upper quartile, or data-driven threshold. Scaling deals with sparsity and is compatible with most statistical methods, yet it does not fully resolve compositional data challenges. Transformation replaces the original counts with normalized ones, and various methods, including additive, centered, and isometric log-ratio transformations, exist for microbiome data. Transformation does not reduce sample size but can complicate the estimation of diversity, feature selection, or correlation between samples when precision remains constrained. Again, zero counts pose challenges during transformation, and employing pseudo-counts or imputation has advantages and limitations. The addition of a pseudo-count can inflate low-abundance features, while imputation can induce spurious correlations among zeros [38].

Feature selection and dimensionality reduction

Constitute fundamental preparatory steps in Microbiome-AI fusion, applicable to various microbiome data types. While large datasets can provide a plethora of features, many of these are irrelevant or excessively noisy for supervised-machine-learning models such as those aimed at predicting infections. Reducing the feature space to focus on the most relevant signals is thus essential prior to the automatic identification of informative patterns. Moreover, maintaining biological interpretability during feature-reduction procedures is crucial to ensure that the selected features not only aid prediction but also support the diagnosis of infection routes and inform subsequent clinical decisions.

Work on multidimensional feature reduction specifically tailored to microbiome data is limited. Studies employing standard dimensionality-reduction techniques, such as PCA and t-SNE, often disregard the unique biological characteristics of the datasets and may yield results that remain difficult to interpret from a microbiological perspective (ref: 977bbbce-7aec-4a4b-8390-6eb346ea2b06). Adopting well-targeted strategies for feature selection and dimensionality reduction, thus combining signal concentration with biological relevance, enhances the utility of Microbiome-AI applications. The need to tackle both aspects simultaneously arises from the distinct procedures involved in the two activities and the site-specific nature of the signals available for infection prediction (ref: 40da36a5-7323-4478-b408-b99d424554c9) [49, 50, 51].

Challenges in microbiome-AI fusion

Artificial intelligence (AI) holds promise for enhancing medical diagnostics-a particular focus in digital healthcare-and improving national disease outbreak surveillance. In parallel, precision medicine is evolving from the broader establishment of clinical pathways, protocols, and guidelines into smaller units of therapy from populations into varied and individualized treatment plans. Within this context, precision microbiome medicine merges microbiome data into diagnostics and treatment options; a particularly tractable challenge within the broader precision medicine landscape. Holistic models of infectious disease have primarily characterized the interactions of pathogen, host, and environment. However, efforts to build models of the clinical microbiome and its relationship to infectious disease are now evolving [9].

Chapter - 7

Real-Time Pathogen Detection and Monitoring

AI-based biosensors

A broad range of sensing modalities supports biosensor-based diagnostics, facilitating point-of-care (POC) implementation [9]. AI-especially machine learning-fundamentally reshapes PIC epidemiology, equipping experts with new capabilities for health monitoring, pathogen detection, and infection prediction [52]. When combined with conventional resources and numerical data, biosensor signals enhance detection and prevention of infection risks. The demand for biosensors rises with global efforts in epidemic monitoring and COVID-19-related precautions. Alongside conventional monitoring and detection technologies, biosensors serve to effectively validate both local and global health transmissions.

Further PI-related sensors lend vital support to multidisciplinary approaches that encompass a wide array of disciplines, including microbial, metabolic, and lifestyle monitoring. These advanced technologies enable an unprecedented level of precision in data gathering. The accessibility of such accurate, longitudinal, and lifestyle-related data collection holds immense potential to significantly enhance the existing pandemic model by capturing critical insights that were previously underestimated or overlooked in the context of PI transmissions. Various aspects of monitoring signals arise from conventional detection methods, early identification of outbreaks, comprehensive infection risk assessments, evaluations of nosocomial infections, occasional infection sites, and precise forecasts regarding the progression of infections over time. This comprehensive data collection approach not only aids in understanding current infection trends but also lays the groundwork for proactive responses in the domain of public health management and epidemiology. [44, 53, 54]

Wearable diagnostics

Wearable devices facilitate longitudinal and nonintrusive monitoring of various biophysical markers, promoting adherence and integration into clinical workflows [55]. They are particularly well suited to monitoring at-risk populations, such as patients on antibiotic treatment or those exposed to infectious outbreaks, when early detection of secondary nosocomial infections or pathogen carriage is crucial [56]. Microbiome signals that indicate increased risk of infection can therefore be coupled with smart biosensors to implement a comprehensive infectious disease management workflow.

Point-of-care tools using microbiome signals

Appropriately, microbiome signals are particularly compelling for point-of-care (POC) diagnostics, where simple binary decisions based on low-complexity input types must be made rapidly, sometimes even without the assistance of a laboratory specialist [55]. When used in conjunction with externally complemented sensing technologies, these signals can inform immediate decisions about antibiotic use, microbiome-modulating therapies, and therapeutic resistance-choices that otherwise require time-consuming laboratory intervention. The target specifications for such POC diagnostics in infectious disease agreements with established studies on urgency and decision complexity [57]. POC diagnostics deal with clear pathogen identification and resistance forecasting from intact biological samples to components such as circulating DNA or

RNA shed by active pathogens remain viable options. By utilizing AI-based models, the corresponding data support a higher-frequency early-warning system that might prevent the onset of infection or act preemptively following environmental exposure.

Chapter - 8

Personalized Antimicrobial Therapies

Microbiome-informed antibiotic selection

Precision microbiome medicine offers a paradigm shift from conventional one-size-fits-all approaches by tailoring diagnostics and therapeutics to an individual’s unique microbial communities and host characteristics. Consequently, microbiome-informed models for pathogen and resistance prediction hold considerable promise for enhancing infectious disease diagnostics. Specifically, advanced analytics applied to multi-omics and biosensor data can pinpoint pathogens, forecast resistances, and identify target pathogens and mechanism-based therapies. However, regions such as low- and middle-income countries (LMICs), where such intricate data may not be readily available, risk being sidelined from the benefits of microbiome intelligence. Accordingly, models that extract microbiome signatures from routine culture data, thus enabling large-scale monitoring of microbial resistance, could enhance diagnostic equality across diverse geographies and enable precision guidance toward effective treatments.

AI plays a crucial and pivotal role in significantly enriching and enhancing antibiotic selection processes. Advanced models that are meticulously trained on datasets encompassing metagenomics, resistome, and comprehensive microbiota data have the capability to accurately predict the minimum inhibitory concentration (MIC) ranges. These models can also identify the likely resistance determinants for various key infection pathogens. Moreover, when it comes to the analysis of microbiota alone, such assessments can effectively steer proposals for targeted therapy and inform the selections of fecal microbiota transplant (FMT) donors. This is achieved by identifying patients who exhibit significant microbial deviations prior to their transplants, leading to better therapeutic outcomes and personalized treatment strategies [9, 58, 59, 60].

AI for predicting resistance profiles

The widespread utilization of antibiotics has been plagued by the emergence of resistance. The accurate distinction between susceptible and resistant microorganisms is paramount for effective antibiotic selection and the reduction of resistance propagation. The increasing sophistication of next-generation sequencing technologies (NGS) has enabled the determination of the genomic features of pathogens, paving the way for the selection of antibiotics on a genomic basis. Furthermore, machine-learning models can predict the likelihood of antimicrobial resistance directly from the metagenomic sequencing of clinical samples. These network-based models have made it possible to generate a comprehensive overview of the resistome, thereby offering decision-support capabilities for the sustainable application of antibiotics.

Microbial metabolites play a pivotal role in the modulation of antibiotic activity and the prediction of pharmacological effects. Machine-learning algorithms can forecast the metabolization of drugs through the gut microbiota from their structural characteristics. The generated knowledge on pharmacokinetic profiles holds the potential for preclinical screening of drug candidates, ultimately minimizing the risk of late-stage failures associated with in vivo studies. Similarly, ongoing research endeavors aim to elucidate the association of broad-spectrum antibiotics with the microbiome to guide clinically pertinent therapeutic decisions [9].

Fecal microbiota transplantation (FMT) and probiotics

Microbiota-based interventions seek to restore healthy microbiota composition and function. The World Health Organization has listed fecal microbiota transplantation (FMT) as an essential therapy, and a practical diagnostic framework exists for identifying candidates. Models trained with 16S amplicon sequencing data indicate a preventive effect against Clostridium difficile. Experimental and clinical studies demonstrate that modeling the microbiota community enables AI-assisted donor-recipient matching; patterns of predicted donor impact correlate with clinical success.

Probiotics, prebiotics, and synbiotics constitute additional therapeutic avenues. When the microbiota is a target and specific candidates are under consideration, a framework for scrutinizing their suitability according to patient context has been articulated. AI-enhanced matching for pivotal microbiota-based treatments is an emerging field where diagnostic input can refine strategy development and increase the probability of favorable outcomes.

Chapter - 9

AI-Enhanced Drug Discovery Targeting the Microbiome

Screening of microbiota-based therapeutics

Microbiota-based therapeutics have emerged as a promising strategy to combat infections and restore microbiome balance. The performance of microbiota-based therapeutics, such as fecal microbiota transplantation (FMT) or dietary supplements, can vary considerably among individuals. This variability often hinders the identification of optimal therapeutic options or candidate donors. A systematic assessment of microbial and host markers could facilitate the selection of microbiota-modulating strategies in precision medicine frameworks.

Microbiome parameters influencing the efficacy of microbiota-targeting treatments can be predicted using artificial intelligence (AI) methods. To address the clinical need for personalized microbiota modulation recommendations, machine learning (ML) tools for predicting treatment outcome were developed. These models leverage diverse pre- and post-treatment datasets and capture the effects of dietary changes, probiotics, and prebiotics on the microbiome [61].

Predicting microbial drug metabolism

The human gut microbiome harbors a vast array of metabolic enzymes capable of modulating human physiology and the metabolism of drug molecules. Microbiome-mediated modification of drug metabolism can substantially alter pharmacokinetics and pharmacodynamics, leading to interindividual variations in therapeutic outcomes, adverse drug reactions, and even drug efficacy. It has been demonstrated experimentally that the biotransformation of hundreds of drugs could be catalyzed by the human gut microbiome. Nevertheless, a rapid in silico approach for accurately predicting the microbiome-enzyme-drug relationships before in vitro studies remains important, especially for drug development pipelines that require fast action [62]. The microbiome-ecosystem status prior to exposure to a drug can also be predicted, as the drug itself may alter the metabolism of other molecules present in the ecosystem [63]. A tool called DrugBug-2 has been developed, which employs a machine learning framework trained on a large experimental dataset of 475 drugs that have been biotransformed by the gut microbiota and their associated microbiome-encoded metabolic enzymes.

AI-guided microbiome modulation strategies

Design interventions-dietary changes, prebiotics, or probiotics-with simultaneous tracking of pathogen presence or related signals, to enhance effectiveness and suggest adjustments as conditions evolve [9]. Optimization employs formalisms such as response surface methods, combining preliminary experiments with feedback-control algorithms. Systems biology models extract community-scaled interactions, enabling trials with multiple microbes, to gauge interventions targeting frequently overrepresented taxa, or multi-species replacements that quell competition or foster beneficent partners. AI interfacing describes structure-activity relationships across food compounds, prebiotics, and probiotics, aiding therapeutic selections for specific objectives.

Chapter - 10

Host-Microbiome-Pathogen Triad Modeling

Systems biology approaches

Microbiomes contribute to human health and well-being and are involved in diseases. Disruption of microbiota increases risks of opportunistic infections and can alter pathogen life cycles, yet a comprehensive understanding of the links between microbiota and infections is still lacking. Systems biology enables the integration of biological data from multiple sources into mathematical models that describe systems-level activities governed by the interacting components. Building systems biology models of pathogen-host-microbiome interactions can provide an integrative viewpoint on infection risk mechanisms, guide experimental studies refining diagnostic targeting and treatment selection, and enable the prediction of the effects of selected interventions.

AI models formally represent knowledge, make predictions, and provide insights into pathogen-host-microbiome interactions-formulating complex biological problems in quantitative terms and capturing diverse operating principles. By permitting in silico experiments, they facilitate the generation of and testing hypotheses, thus accelerating cautionary evaluation of clinically and biologically relevant scenarios-hypothetical perturbations of the microbiome, microbiome shifts induced by vaccines or therapies, and the predicted outcomes of such events [5, 21]. Cross-reference the title ‘AI models simulating interactions’ when planning ‘Implications for immune therapies.’

AI models simulating interactions

Machine learning models provide unique tools for scientific exploration of pathogen-host-microbiome interactions after infection. In silico experiments can test hypotheses about transmission pathways and quantify their likely rapidity. Data-driven, dynamical models of microbiome evolution inform the potential for microbiota-targeted therapies. One example of these models is “Relationships between commensal bacteria and co-infecting pathogens determine the severity of intestinal infection,” which investigates how key pathogens subvert host defenses and condition microbiota structure to facilitate transmission of superinfections and vice versa, influencing treatment and infection risk. The system is formulated as a compartmental model where populations are subject to mass-action transmission, horizontal gene transfer, and class-based competition. Individual microorganisms govern their optimal acquisition depending on microbiota structure, readouts are time series of abundances, and the corresponding procedures enable discerning the modelling and mechanistic processes with maximum predictive, likelihood-based accuracy [64].

Implications for immune therapies

Insights into the triadic pathogen-host-microbiome interactions outlined above may unlock novel immunomodulatory strategies for infections and sepsis [9]. Interactions with the microbiome shape safety and efficacy outcomes of vaccines and therapies based on these immune mechanisms. Reasoning around these influences could therefore inform precision diagnostics targeting the most susceptible pathways. Pathogen-host-microbiome modelling of C. difficile, uropathogenic E. coli, and the fungal triad of Candida albicans, Candida glabrata, and Saccharomyces cerevisiae illustrates the varied and significant impact of microbiota on immune system regulation. Within this framework one can further probe the relevance of pathway-specific immune signatures for different infections.

Chapter - 11

Ethical, Legal, and Social Implications (ESI)

Data privacy in microbiome and AI applications

Drawing relevant lessons from the human microbiome, disease sensitivity and immune regulation emerge as two fundamental aspects of infectious diseases. Multiple mechanisms illustrate how microbiome perturbations contribute to heightened infection susceptibility. Conversely, the interplay between pathogens, microbiome and host-specific immune regulation presents a basis for predictive diagnostics regarding infection outcomes [65]. AI adoption offers promising avenues for diagnostic tool development, accelerating detection of dysbiosis and biotic pathways linked to intervention strategies. The use of artificial intelligence (AI) techniques is further distinguished by its capacity to handle big data sets, elevating their potential as tools for guiding microbiome models and delivery systems.

Bias in AI models

These frameworks encounter population bias when the characteristics of the training data differ significantly from those of the target population. For instance, a model trained mainly on White Americans may yield higher error rates for other demographics. Within cardiac imaging and digital histology, variations in equipment and protocols among institutions introduce site-specific signatures in the data that can also lead to population bias when models are deployed elsewhere [66].

Informed consent for microbiome-based diagnostics

The integration of next-generation sequencing to profile the microbiome holds great promise for improved diagnostics across a variety of infectious diseases, particularly bacterial infections. However, microbiota profiling is still an emergent field and microbiome-based diagnostics have yet to be adopted for clinical use. Rigorous evaluation of the accuracy of these diagnostics, as well as their integration in clinical practice, is still ongoing. Understanding the evolution, distribution, and growth of co-infecting pathogens, notably during co-infection with common pathogens, is critical for guiding treatment strategies. Additional modelling efforts aimed at understanding pathogen spread, end-points, and optimally-targeted treatment strategies also offer potential. Cross-jurisdictional spread surveillance is also critical, and AI models have the potential to play a pivotal role in such investigations.

The optimisation of epidemic- and intervention-targeting agent deployment across tightly co-dependent networks represents an additional modelling challenge with far-reaching implications for tackling future pandemics. The emergence of bulk omics technologies presents another important opportunity; the integration of host/viral and virome data on the same sample will enable the study of the interplay between human host response and virus (including bacteria) interactions in infectious diseases. Microbiota studies have so far exclusively focused on a single compartment (e.g., blood or gut); tracking pathogens, microbiota and viral pathogens across multiple samples and compartments is crucial in unravelling the community responses to pathogen challenges and virulence factors such as antimicrobial resistance. Companion datasets, including virome, metabolome, and metabolome-soil datasets, may aid in further elucidating microbial network-level understanding of the impact of soil on the microbiome and the role of viral co-infection on pathogenicity and spread. Across Pandemic 1, Pandemic 2, and Pandemic 3 studies, the impact of individual pathogens and their respective interactions from circadian, annual, and pandemic cycles can be investigated [67].

Chapter - 12

Case Studies in Precision Microbiome Medicine

COVID-19 and microbiome interactions

COVID-19 emerged in December 2019 as a serious pandemic; its causative agent, SARS-CoV-2, is transmitted mainly via respiratory droplets. The microbiome plays key roles in the onset, development, and recovery from COVID-19. Perturbations of the gut microbiome have been associated with an increased risk of acute respiratory distress syndrome and COVID-19 severity. Concomitant infection with pathogens such as influenza a virus has been correlated with more severe disease. The accumulation of harmful pathogens can, on the other hand, promote gut dysbiosis, compromising the barrier against SARS-CoV-2 and favouring replication. Continuous monitoring of the microbiome may provide complementary insights into COVID-19 progression [68].

Recent advances in rapid and accurate detection of a pathogen's presence or absence have paved the way for an independent pathway towards reliable diagnosis and ongoing monitoring of COVID-19. The pathogen-host-microbiome triadic interactions associated with COVID-19 allow for serious yet timely surveillance, as existing co-infections consume portions of valuable information used in previous models. Early screening has emerged as a priority target; the combination of key features from low-cost instruments with multi-omics approaches empowers strategic decision-making around not only COVID-19 but also other respiratory infectious diseases.

AI diagnostics in tuberculosis and HIV

Applications of Artificial Intelligence in Microbial Diagnosis The diagnosis is an important factor in healthcare care, and it is essential to identify microorganisms that cause infections and diseases. The application of artificial intelligence (AI) systems can improve disease management, drug development, antibiotic resistance prediction, and epidemiological monitoring in the field of microbial diagnosis. AI systems can quickly and accurately detect infections, including new and drug-resistant strains, and enable early detection of antibiotic resistance and improved diagnostic techniques. The application of AI in bacterial diagnosis focuses on the speed, precision, and identification of pathogens and the ability to predict antibiotic resistance. Microbial diagnosis entails the identification of microorganisms through techniques such as culture, molecular analysis, and imaging, which constitute a pivotal domain within the realm of healthcare. It starts with appropriate sample collection and runs into several problems with conventional procedures, including sample handling, difficulty in culture, incorrect identification, and antimicrobial susceptibility testing difficulties. These traditional methods require manpower, and treatment is often delayed. Artificial intelligence (AI) has revolutionized the field of microbial diagnostics by providing more precise and current findings. AI analyzes data, pattern recognition, and diagnostic processes faster. It is essential for early identification of the disease, advancement of treatment, custom treatment, and epidemic monitoring. Advanced data sets are analyzed by AI-driven algorithms to detect infections rapidly, anticipate disease outbreaks, and improve treatment approaches and outcomes. The use of AI in microbial diagnosis raises concerns about ethics, including protecting patient privacy, addressing algorithmic biases, maintaining data security, promoting transparency, and ensuring equal treatment.

Machine learning of the Whole Genome Sequence of Mycobacterium Tuberculosis: A Scoping PRISMA-based Review This review analyzes AI models for predicting DR-TB profiles, including AI techniques, datasets, and performance metrics, to understand strengths and limitations. It discusses experimental setups, data sources, and evaluation methodologies to gain insights into advancements and challenges and suggests future research directions. Proposed avenues include integrating multiple AI techniques with other modalities, developing robust models for diverse, noisy data, predicting resistance to additional drugs, combining AI with molecular diagnostics or biosensors for rapid detection, and evaluating cost-effectiveness and ethics in low- and middle-income countries. The review also emphasizes expanding diverse training datasets across geographic regions and genetic variants of M. tuberculosis to improve generalizability, and developing user-friendly bioinformatics tools to interpret whole-genome sequencing data. Limitations include diversity of ML approaches and inconsistent model naming, which can hinder article discovery; to mitigate this, a manual search of references and tools like Research Rabbit were used, though omissions may still occur.

Nosocomial infections and microbiome disruption

Healthcare environments are associated with distinct microbial communities that differ significantly from those found in the community [5]. The hospital environment comprises complex ecosystems composed of microbes of human, animal, and environmental origins, including those that are pathogenic to humans [9]. Although many hospital pathogens are already well-established in the community at the time of admission, disruption of the microbiome due to factors such as surgery, chemotherapy, antibiotic treatment, and prosthetic implants enhances the risk of infection. Such disruption enables pathogens to exploit vacant niches, spread through food and water, or, in the case of opportunistic pathogens such as C. difficile, take advantage of decreased microbial competition. Early disruption therefore represents a critical period that must be monitored and curbed to decrease infection risk. The hospital environment offers both opportunities for mitigation, through the implementation of effective and rapid diagnostics, and challenges, such as enabling low-cost standalone tools that are simple to administer and incorporate into busy clinical workflows.

Chapter - 13

Regulatory Frameworks and Clinical Implementation

FDA and EMA guidelines on AI/microbiome tools

Regulatory agencies have issued important statements on the use of AI tools in healthcare. The FDA's discussion paper on "Artificial Intelligence-Based Software as a Medical Device" provides guidance for software that proposes patient diagnoses, the selection of therapeutic options, or any other clinical decision that determines patient management. Similar considerations may apply to diagnostics involving severe disease phenotypes or life-threatening conditions, which require high reliability and accuracy, especially when the result implies a change in the prescribed treatment or care.

Alongside the necessity of conducting clinical trials, the continuous and ongoing evaluation of AI systems throughout their use is crucially important to ensure patient safety effectively. This aspect is especially relevant in the context of the European Union, where the Medical Device Regulation explicitly states that post-market surveillance is an essential component for gathering real-world evidence regarding the performance and associated risks of medical devices. Furthermore, the stringent requirement for the traceability of all AI decisions necessitates clear and comprehensive reporting as well as justifications of the various data-dependent factors that drive the decision-making process. The European Medicines Agency's draft discussion on "Guidelines on the Use of Artificial

Intelligence in Medicinal Product Development" provides detailed delineation regarding the use of AI for medicines that are reliant upon artificial intelligence at any stage in the comprehensive process: including design, development, manufacturing, and quality control. AI applications must adhere meticulously to a comprehensive set of principles of trustworthiness; namely, that the system operates effectively within clearly defined and established boundaries and is demonstrably accurate, reliable, valid, safe, secure, and explainable. In addition, it must be able to be audited thoroughly, avoid or minimize unintended biases effectively, and remain under human control throughout its operation [69, 70, 71, 72].

The advent of machine learning (ML) in microbiome science paves the way for clinical AI-enabled microbiome tools with the potential to improve patient health outcomes [1]. These tools can support the prescription of microbiome-targeted therapies by predicting the effect of specific microbiome modulations on host traits. Regulatory guidelines anticipate widespread adoption of ML enzyme prediction tools in the bio-manufacturing sector, suggesting similar uptake in clinical microbiome applications.

Microbiome tools use data such as 16S rRNA gene sequencing, shotgun metagenomics, and metabolomics [2]. The explanatory notes from the FDA state that microbiome data constitutes an additional diverse input and output layer that is not the de-facto narrative of the patient (e.g., diagnosis) or the prescription (e.g., drug selection). Microbiome analysis serves as an exploratory scientific investigation of the relationship between host and microbiota. Four dimensions of microbiome data and their heterogeneous nature necessitate elucidation of analytical validity, clinical validity, and clinical utility. Detailed evidence standards for microbiome-targeted AI systems will be addressed later.

Multiple literature sources report secure handling of electronic health record (EHR) data during AI system development, and other domains demonstrate similar security assurances with genomic data. Thus, these insights are applicable when accompanying microbiome data is integrated into a broader multi-omics clinical testing framework. As a consequence, concern over the provision of explanatory information and stimulus for data-provenance strategies will be highlighted in subsequent sections.

Artificial Intelligence (AI) is progressively being integrated into the research and application of microbiome science. The microbiome is increasingly recognized as a key player in human health. A wealth of microbiome data encompassing gut, oral, vaginal, skin, and biogeographical distinctions is available. Appropriate anonymized and aggregated microbiome samples and associated metadata can be shared with transparent protocols for controlled research. Until now, the standard for microbiome data sharing has been driven by scientific collaboration. Commercial organizations are increasingly entering the microbiome data domain, highlighting a need to align commercial interests with freedom of scientific research and open access sharing to support these interests [3, 1].

Microbiome data and analytics underpin the practice of machine-learning-enhanced microbiota analysis to assess and provide clinically actionable insights into human health. Data types such as taxonomic and functional abundances captured through 16S ribosomal RNA (rRNA) gene amplicon sequencing, shotgun metagenomic sequencing, metatranscriptomic sequencing, or metabolomic profiling represent the core of this domain. Additional data that provide valuable contextual information for microbiome-targeted analysis include clinical metadata, dietary intake records and commonly co-analyzed dietary profiles, multi-omics measurements from host biosamples, gut transit time and body mass index derived from subject questionnaires or wearable devices, and other microbiome-independent time series observations and widely available longitudinal datasets [1].

Artificial intelligence (AI)-the field of computer science encompassing machine-learning (ML) and deep-learning (DL) methods-has rapidly evolved to become a pivotal element of microbiome science across diverse biomedical domains. AI-driven solutions capable of predicting microbiome composition, microbial functionality, and host-microbiome interactions empower practitioners to efficiently derive clinically actionable insights from complex microbiome datasets. These AI-enabled microbiome tools address the well-documented need for advanced analysis methods that facilitate broader adoption of microbiome science in clinical practice. Designed to augment, not replace, professional judgement, they assist healthcare practitioners in gaining a deeper understanding of the relation between human health and the microbiome.

The U.S. FDA recognizes microbiome-focused artificial intelligence (AI) tools as devices or software as a medical device (SaMD) for microbial characterization that derive clinical insights without further data processing. Estimating the relevance of raw microbiome information retains analytical validity. Similarly, AI-enabled interpretation of microbiomes generally remains classified as SaMD rather than an in vitro diagnostic device [4]. Microbiome-targeted AI analysis occupies a unique niche within restricted-device pathways by verifying analytical validity through content-provider assurances on data generation and instrument operability. Clinical-validation strategies for microbiome-centric AI vary widely; uncontrolled studies are commonplace, and a definitive gold standard remains elusive [5].

The FDA endorses continuous validation of artificial intelligence/machine learning (AI/ML) techniques and utilizes real-world evidence to inform maintenance, alteration, and constrained expansion of indications, with expectations for ongoing safety and performance adherence. Regulatory discussions surrounding microbiome-centric tools emphasize performance and safety alongside real-world evidence and continuous-validation provisions found throughout the FDA's AI/ML framework. The FDA prioritizes risk management and clinical intervention over conventional clinical-utility metrics, although observational studies capturing clinical impacts remain acceptable. Therefore, microbiome modeling following AI/ML principles will likely align with FDA regulatory intent.

The US Food and Drug Administration (FDA) consolidates medical devices, such as microbiome-system tools, vaccination coverage, or weight loss applications. Regulatory attention for software-based assistance products depends on the specific claims being made. These items can still be regulated via the software-as-a-medical-device (SaMD) pathway.

The FDA lists four guiding principles for SaMD classification:

1) The device's intended use.
2) Data inputs and the outputs produced by the software
3) The effect of those outputs on clinical decision-making.
4) The workflow associated with implementing outputs or recommendations.

Artificial intelligence/machine learning (AI/ML) platforms that adapt and improve over time are already subject to regulation on account of predictive-publication use cases. The agency has further stipulated that regulatory frameworks for AI/ML tools must account for concurrent collection of real-world data, provided that such evidence continues to assure safety and efficacy.

With the centre dedicated to digital health opening in 2020, FDA resources aimed at fostering responsible innovation have been released as part of the Digital Health Innovation Action Plan. The Digital Health Innovation Action Plan includes the Software Precertification Pilot Program and an AI/ML-driven Software as a Medical Device action plan outlining the development of a possible future regulatory model for oversight of software-based medical devices and their real-world performance. Both initiatives are accompanied by collaborative efforts with patient groups, health care providers, academics, and industry, aiming to determine thresholds and performance evaluations for metrics most critical to real-world operation of AI/ML-enabled devices. By virtue of their capacity to learn from real-world data and enhance performance on an ongoing basis, AI/ML systems stand to benefit markedly from data collected after deployment [6].

Marketing authorization pathways for devices, software, and AI/ML tools at the US FDA are accessible via the Industry Guidance and Digital Health Center of Excellence portals. Medical device regulations cover grounds for classification as a device and labeling requirements that apply whether intended for direct use by healthcare professionals or patients [5]. Marketing submission options comprise 510(k) and De Novo routes predicated on substantial equivalence and exemption from premarket review, respectively, as well as Pre-Market Approval which addresses a broader scope of safety and effectiveness criteria.

Software that incorporates AI/ML or following human training is classified as Software as a Medical Device (SaMD) if it meets Medical Device Definition for Equipment and Software. The Digital Health Center of Excellence articulates a conservative position on SaMD regulation based on a consensus that the premarket review framework has not kept pace with innovation. The Center defines AI/ML SaMD and stipulates a threshold that unregulated SaMDs may not alter clinical information to qualify for exemption.

The EU approach to medical-device regulation, particularly in the post-marketing phase, is best understood in the context of the category of products subject to the Medical Device Regulation (MDR). In the absence of better-defined legal categories within the MDR, AI-driven microbiome products will typically qualify as software classified as a Class II-device or a Class II-a device addressed by the General Safety and Performance Requirements outlined in Annex I of the MDR. The main requirements apply not only during the pre-market phase but also in the post-market phase, where continuous assessments of performance and safety are required. The main difference of the requirements concerning the safety and performance of AI-driven devices in the pre-market and post-marketing phases is that the latter must take into account data from actual clinical use [3].

The majority of AI medical software operating and being introduced in the European Union falls under the classification of class IIa or IIb. Both class IIa and IIb devices require an external audit and certification process to gain market access [7, 4]. The regulations governing artificial intelligence also stipulate that simplified and enhanced post-market surveillance should be an integral part of the conformity assessment process [3].

The European regulations use the CE mark to signify conformity with health, safety, and environmental standards. The Medical Device Regulation (EU) 2017/745 (MDR) classifies devices into four classes: I, IIa, IIb, and III, with higher classes indicating higher risk. Class I devices require self-certification, while most AI medical software falls under class IIa or IIb, which require external audits and certification for introduction [7].

Regulatory Frameworks and Clinical Implementation of AI-Driven Microbiome Tools outline the regulatory background of Artificial Intelligence-enabled microbiome analysis tools, which supplement conventional laboratory analyses of gut microbes with insight into health impact based on artificial intelligence (AI). This technology can enhance clinician decision-making after microbiome examination; help to select diet and supplements favourably influencing microbiome and health; and support the analysis of microbial evolution in response to diet, supplements, environmental exposure, and other factors. Practical AI-based digital approaches address medical needs: microbiome-aware selection of diets and supplements; pies, plots, and histograms illustrating expected health impact on a scale of severity and days after intervention; evolution charts; and microbiome health impact reporting. Internal and external oversight ensures acceptable levels of clinical relevance, health impact, and safety.

The study designs for AI that targets microbiome data typically encompass clinical and preclinical investigations. Clinical investigations can vary in complexity, and a well-defined baseline study design is crucial. A robust study should be designed to evaluate the AI software's analytical and clinical validity, as well as the clinical utility of its output. Analytical validity examines whether the AI accurately predicts features from microbiome datasets and verifies that the ground truth is well-defined. Importantly, in a microbiome-focused AI implementation, the predicted features ought to involve test-results-related microbiome signatures prior to evaluating clinical validity [8]. Clinical validity investigates how well the AI outcome correlates with clinical parameters of interest, reflected through clinical situations that permit sound evaluation. Clinical utility offers assurance that the AI output enhances clinical decision-making, potentially through a controlled intervention study that demonstrates AI-backed decision-making leads to superior or more consistent outcomes compared to prevailing practices.

The studies outlined primarily focus on microbiome-oriented AI directed at clinical validation. Addressing the AI tool's suitability while ensuring privacy and security requires a multi-faceted design. Beyond individual consent, de-identification of personal data, governance of data handling, and clear description of data provenance constitute essential factors. The use of self-reported questionnaires specifically addressing these dimensions, along with data lineage consideration, is paramount in these designs, which allows restrictive application of proprietary AI tools without compromising copyright protection. Ethical considerations and data governance, alongside privacy and security, emerge as prominent issues faced during the utilization of microbiome-related data.

Microbiome-targeted artificial intelligence can enhance analytical, clinical, and translational validity, and few data types warrant such investments. Essential analytical capabilities include variety (spanning clinical, metagenomic, and phenotypic data) and correlation (the degree to which model features relate to existing knowledge). Vigilance is especially important regarding clinical validity, as models may raise unanticipated concerns beyond those planned for testing.

Study designs that secure analytical validity yet prematurely constrain clinical validity encourage exploration of microbiome data spaces that expand on existing knowledge ([9]). Examples can include investigating target-disease spaces lacking direct clinical associations or leveraging rich phenotype data.

Analytical validity, clinical validity, and clinical utility. The successful deployment of microbiome analytics in clinical settings is predicated upon well-defined standards for analytical validity, clinical validity, and clinical utility. Such standards are further refined when regulatory reviews are applicable. Microbiome-targeted datasets afford the practical option of establishing analytical validity by extensive data provenance and system-agnostic data lineage. Accordingly, analytic tools can be deployed without prior clinical validation. With respect to clinical tools, analytical validity refers to tool application in accordance with described case studies and formal usage instructions, rather than comparative performance, whereas modelling is a prerequisite for clinical validity in non-regulated scenarios. Principles of clinical utility in microbiome analytics remain under exploration. Design frameworks mapping directly from clinical questions and case study characteristics support audit trail and provenance capture during real-world operation, thus ensuring alignment with system-independent validation principles.

Workflow integration presents additional control over data transmission, provenance specification, and consent tracking, all of which are critical to regulatory compliance before system implementation, as detailed elsewhere [1].

Microbiome-focused data, characterised by high dimensionality and compositional parameters, impose further constraints on validation strategies. Data-focussed frameworks align seamlessly with established decision-making and clinical control concepts, facilitating rigorous validation exposition.

The ethical implications of microbiome-targeted AI have earned increasing interest along with the rapid evolution of regulation and clinical applications. The need for ethics guidance in the AI domain is widely recognized across disciplines [3]. Ethical considerations encompass a diverse array of topics, including: inserting the regulatory focus on the content that follows. professional conduct in algorithm development; respect for the rights and welfare of data subjects; establishment of appropriate institutional governance; and equitable access to tools that can potentially generate health impact.

Frameworks for ethics in AI-based systems are being instituted to facilitate description and interpretation of ethics challenges. The ASIMOV project, undertaken under the auspices of a European University Association initiative involving the EU Commission on AI, provides a structured framework for recording ethical issues encountered during the development of university-based AI systems, whether in the education or research contexts. A key element of the project concerns the capacity to delineate systems under development that operate with an AI component de facto or de jure, thus amplifying system risk. The biomedical context of microbiome-targeted AI, coupled with regulatory scrutiny, composition of the overall development team, and patient-benefitting mission associated with the projects, all diminish the ethical burden of microbiome-focused algorithmic initiatives. The simplicity of the regulatory landscape further lightens the ethical workload. Given these strong incentive properties associated with microbiome-focused AI development, obscuring both ethical and regulatory challenges, precise articulation of existing concerns is both essential and beneficial.

De-identification and consent play pivotal roles in securing microbiome data from the introduction of AI-enabled algorithms. De-identification refers to the removal of direct identifiers, such as names and Social Security numbers, that could enable the identification of study participants [10]. Re-identification risk increases when individuals possess sensitive information about themselves, and expansive datasets raise the likelihood of uncovering identifiable attributes. Data governance encompasses the authority and control over data assets and establishes data- and process-specific policies, procedures, and standards for the data lifecycle [11]. Governance prioritises across-the-lifecycle management of personal data to mitigate the risks of misuse and unintended consequences. Consent signifies the agreement by an individual to the processing of their data under the specified terms. Consent enables organisations to engage in legally permissible data processing activities while preserving transparency, autonomy, and agency for the individual.

De-identification, data governance, and consent are prerequisite design considerations for governance of individuals' microbiome data and implementation of AI-driven microbiome applications. De-identification renders datasets non-attributable to individuals while preserving analytical value for modelling microbe-host interactions. Data governance supports de-identification by establishing responsibilities for compliance with regulations and organizational guidelines, formulating appropriate policies, and monitoring adherence to them, as well as delineating data access rights and uses. These issues are closely intertwined with data provenance, analytical population definitions, and participant consent, described in further detail in [12].

Governance of microbiome data-which may derive from health records, treatments, public health data, self-reports, or direct measurements-is compounded by process intricacy and diversity of potential hosts. The transparency required for enabling the governance framework often limits the data-access restrictions needed to fulfil an identified use. For many microbiome applications, therefore, temporally delimited access to the data required for identification of the analytical population and the population itself may suffice in addition to the de-identification necessary for the modelling stage. Governance thus extends to specification of the microbiome-access policy matrix and documentation of the data lineage throughout the workflow [11]. A coarse, time-correlated delineation of the models' expected domain aids in determining which records, treatments, or measurements constitute the initial data set when the governing policy permits access to the underlying data.

Audit trails and cybersecurity measures are essential for protecting the safety and integrity of clinical algorithms and establishing clinician and patient trust [13]. Patient safety may be compromised when an algorithm operates on unexpected or unintentionally tampered data or is substituted with a different or defective algorithm. Furthermore, the algorithm deployment process with an Electronic Health Record (EHR)-prescribing user interface adds further avenues for unintended tampering. Industry best practices should therefore uphold a minimum audit log, tracking the algorithm, the dataset, and the deployment environment. The ability to prevent tampering with the algorithm, either intent-driven or by another source, in accordance with industry standards reinforces the EHR-prescribing framework [14]. Consequently, maintaining a secure link within the EHR infrastructure that prevents unconsented alterations before deployment and trace-based communication once in execution strengthens algorithm trustworthiness.

Clinical microbiome science has seen rapid growth of AI-enabled tools spanning microbiome sequencing, a priori research, and cohort selection [15]. The potential to address high dimensionality, sparsity, and complex interactions in microbiome data highlights the importance of study design, regulatory engagement, legal requirements, proprietary data use, and product lifecycle considerations. Workflow integration and human-in-the-loop design optimise demand, address clinician concerns on algorithm limitations, and facilitate abiding by regulatory standards for continuous monitoring, algorithm updates, and audit trails. Tools can enhance Electronic Health Record (EHR) data collection, reduce data handling risks, and ensure patient consent compliance.

Microbiome tools achieve clinical integration through two avenues: connection with electronic health records (EHR) and attachment to laboratory workflows. EHR integration sits at the forefront of clinical decision-support effort and is often the most desirable mechanism for any AI-based tool [3]. Such design aligns data intake with the clinical environment and allows the sourced information to remain in-place, thus strengthening data governance. These properties also lend themselves well to human-in-the-loop modalities, where clinician and administrator interaction constitutes a core source of knowledge [16]. Clean input channels further assist algorithm monitoring [17]. The lab-centric counterpart, while often better suited to convenience and fidelity of microbiome transmission, demonstrates greater risk of disrupted data lineages and algorithm degradation through cycle iterations.

Linking microbiome resources directly to a laboratory information management system is the initial step to laboratory integration. Laboratory solutions emerge from defined missions: clarification of the use case, establishment of input heralds, and consideration of post-analysis output, which can take the form of stand-alone reports or direct feed-through into a patient- or sample-bound log. Beyond report mechanisms, successful microbiome-centric integrations secure multiple pathways into entry-level options. A solution with limited lab compatibility can therefore function as a contingency while reviews of linking possibilities unfold.

Microbiome-targeted AI tools are typically introduced into clinical workflows via software interfaces that integrate into electronic health record and laboratory information management systems. These clinician-facing tools are designed to enhance microbiome analysis, aid clinical decision-making, and reduce the cognitive burden on busy practitioners. Human-in-the-loop design principles enable iterative co-development and formative evaluation through user studies, which support practice-based decision-making and reduce the risk of developing mismatched systems [18]. Aligning the intended use, user needs, and input data with every revision promotes relevance, while keeping clinicians informed about the state of AI-driven analyses fosters transparency and trust. Integration of microbiome-targeted AI tools with existing electronic health record (EHR) systems and laboratory workflows supports streamlined data exchange and secure transfer of sensitive information while preserving clinician control over the analytical process. For microbiome-targeted AI tools, preliminary training and monitoring can occur outside clinical settings prior to integration, and detailed change logs facilitate regulatory compliance by documenting version updates and adjustments to datasets and methodologies. Such integration not only aligns with human-centered design and regulatory requirements but also addresses specific practitioner workflows, allowing tools to better reflect expert expectations and deliver clinically relevant analyses [19]. Human-in-the-loop design and clinician-facing interfaces represent foundational elements of the development process, ensuring that AI systems support rather than supplant practitioner expertise and that regulatory integration occurs early in the tool’s lifecycle.

An implementation strategy that aligns with the clinical AI lifecycle promotes the rapid integration of new microbiome-focused tools. Data streams from the extensive electronic health record and laboratory information management systems residing in healthcare institutions can complement microbiome-focused models, as can curated databases of public microbiome studies. These data sources include microbiome samples, clinical outcomes, and details about patient diagnoses, laboratory tests, medications, and treatments-and they capture the progression of the healthcare system toward the specified patient outcome. Regular creation of these data rounds and their scientific accessibility present a significant research avenue to facilitate data-hungry microbiome AI. A human-in-the-loop design enables submission of pertinent information for a tool that uses new data types and alterations of existing data types while assisting physicians in comprehending the tool's functioning and the motivation behind inputs provided [20].

Post-market performance of AI-enabled tools targeting the human microbiome remains of legitimate regulatory interest because data-generating technology must change to remain relevant, data-generation methods evolve, and the nature of algorithm-based predictions naturally changes when refitted. Adjustments are anticipated, and these changes hold implications for safety-performance linked to an individual tool or model could alter-although such model changes disrupt overall functionality only in rare instances that also require fresh regulatory review. Hence, in well-designed implementations, versioning of algorithms differs from a product-like scenario in which the final configuration is expected to persist.

Secure systems can maintain an audit trail of configuration settings, enabling patients and providers to verify that an instrument operates as expected and linking constitution of the AI-driven tool specifically to post-market risk assessment as established earlier [21]. Afterward, routine performance evolution has implications from a reporting perspective. In the case of conventional software, updates deemed not to compromise safety or effectiveness need not be communicated to the regulatory authority but generally must continue to address validated risks. External audit ensures that all substantial changes are precisely recorded [6] and Version Control marks alterations separately; only repetitions of core configurations must then undergo fresh scrutiny.

Post-market performance monitoring focuses on translating microbiome research into clinical practice and validating tools in real-world settings [22]. For AI-driven microbiome tools, translating research to practice typically entails integrating clinical data with microbiome data. Translating the complex relationship between microbiome composition and phenotypes directly into an AI algorithm can be challenging, particularly for data-prioritized methods such as deep learning. To preserve experimental design and leverage assisting data, an alternative is to utilize lightweight predictive models to estimate microbiome composition and apply a second model to relate composition to phenotypic response [9]. Because the first model has no access to experimental or clinical features, variant autotrophic models or biogeographical models that do not use prior knowledge of the full metabolic graph are preferred, although reinforcing from biochemical data is also possible.

Risk controls around decision-making are therefore essential, and a robust approach ensures they remain effective throughout the algorithm lifecycle. Continuous monitoring of statistical risks, performance metrics, and data sources can indicate when real-world drift or new information may affect safety and efficacy [3].

Algorithm lifecycles, including planned updates and ongoing monitoring, warrant regulatory consideration because they influence evidence generation. All algorithm modifications should follow a pre-defined process that assesses and articulates their impact on safety, performance, and applicability to existing use cases. Documenting the scope of changes and the standard applied for assessment, along with all associated information, supports transparency during review and audit.

AI introduces unique complexities in assessing and mitigating safety, clinical, and operational risks that must be methodically addressed for effective microbiome applications. The European Commission outlines an AI-specific risk framework: “minimal” (e.g., crowdsourcing applications), “limited” (e.g., risk evaluation of honorary co-authorship in scientific papers), “high” (e.g., systems that assist with substantial judicial decisions), and “unacceptable” (e.g., social scoring systems). Microbiome-targeted AI could be positioned as either high risk influenced by clinical context (e.g., disparate healthcare professional inability to provide timely treatment to patients with acute kidney injury) or as limited risk (e.g., food-medication interactions) [3].

Another method considers the foreseeable impact of AI-generated advice on essential decisions related to patients and organisations. An “ActU” score aggregates urgency and uncertainty into a single measure (A = urgency, C = consequences, T = beneficence, U = uncertainty). Real-world illustrations enhance perception of associated contingencies revealing the intricacy of effective operation. Displaying limitations for patient understanding is mandated and must accommodate the inherent uncertainty existing in all predictive applications to comply with existing legal frameworks [23].

Establishing fundamental principles for risk assessment of AI-driven tools in microbiome science helps ensure responsible and compliant use. Although in vitro assays and omics technologies underpin most microbiome characterizations, tools are increasingly incorporating AI that produces conclusions based on prior knowledge or human interpretation, alongside live microbiota interventions [24]. Regulatory frameworks for medical devices and software as a medical device must therefore extend to software that provides microbiome-related conclusions, justifying the need to characterize various aspects and limitations of AI-driven tools in the context of perceived risks to patients.

A comprehensive risk framework informs manufacturers and users of AI-powered instruments targeting the microbiome about salient risks and acceptable mitigations. Within this ecosystem, users-including healthcare providers, researchers and various technical experts-make risk-aware decisions on employing the contents of an AI-enabled software as part of a clinical workflow. The high stakes associated with microbiome analysis heighten interest in applying an assessment model to guide responsible deployment of AI-driven microbiome tools and frame a discourse on the fundamentally probabilistic nature of AI models and the uncertainties of digital data transmission. The following risk assessment approach aligns with FDA's draft guidance on artificial intelligence-based clinical decision support software regarding considerations for clear user communication of the inherent limitations and uncertainties of such models.

Artificial Intelligence (AI) performs a crucial role in the analysis of microbiome data, but the intricacy of microbiome science frequently leads to uncertainty regarding the predictions generated by AI-driven tools; hence, the limitations and uncertainty of such devices must be communicated to the patient. Risk-analyses are indispensable for engineers and regulators to manage AI-generated uncertainty, and these uncertainties can be qualitative, quantitative, or model-related [25]. Qualitative uncertainty encompasses the appropriate application of a model, for example, whether an inflammatory bowel disease (IBD) risk-prediction model is suitable for a patient that is not at risk of IBD. Quantitative uncertainty encapsulates population-based analyses that are location-specific. The concerning tendency to underestimate the popularity of disease is highlighted here contextually. Model uncertainty pertains to the additional estimations arising from, for instance, the adoption of a multi-layered auto-encoder that performs well across multiple datasets but still leads to an incomplete understanding of the observed data.

Uncertainty encompasses the limitations of the model-data fit, yet its quantification depends heavily on the modelling stage. This challenge is particularly acute when employing deep-learning models, hence Oncocount and Pathway models - which are based on generalised linear models (GLMs) - could be preferable from an uncertainty-communicating perspective. Classification uncertainty stems from the challenge of determining the correct label in supervised deep-learning functionality or evaluating the absence of documented detection. Nevertheless, even when stating that the prediction carries uncertainty, AI devices might still inadvertently convey reliability, perhaps because of the AI-driven, workshop-oriented, or entirely unsupervised manner of operation. Such error patterns illustrate the paramount importance of clearly conveying limitations and uncertainty to patients.

An examination of clinical scenarios and completed regulatory submissions for AI-powered microbiome tools reinforces the proposed delivery and validation frameworks, offering concrete illustrations to stimulate further development.

Regulatory pathways for Clinical Decision Support (CDS) tools using evidence from microbiome analysis are briefly outlined, followed by extraction of lessons learned from completed FDA regulatory submissions for an AI-enabled Microbiome-Metabolome Risk Assessment Capstone as a Clinical Decision Support tool. These considerations provide practical context for the regulatory-specific discussions. CDS or other AI tools that do not have a direct microbiome target application can still provide useful precedent examples, enabling machine learning models with interpretable outputs to characterise the human microbiome in an innovative manner. Regulatory submissions to document the inherent safety of decision support tools are indispensable. The importance of disclosing tool limitations and prediction uncertainty to patients cannot be overstated; AI-generated information can be misinterpreted and inadvertently harm patients if AI model uncertainty is misjudged by the end-user or clinician performing the interpretation

The Food and Drug Administration (FDA) defines a Medical Device as instruments, apparatuses, and in vitro reagents intended for use for the diagnosis, cure, mitigation, treatment, or prevention of disease. The FDA defines software as a medical device (SaMD) broadly as any software that is intended for one or more medical purposes without being considered a part of a hardware medical device, thus extending the definition of Medical Devices to encompass SaMD, which can range from highly complex to relatively simple [3]. AI/ML algorithms implemented in a microbiome-based model available as SaaMD can therefore be considered a Medical Device [26]. Serious biases can lead to over-reliance on models, especially if bias derives from misunderstood functions of microbiomes or related signalling. The psychiatric or psychological consequences-sustained anxiety worrying about physiological distress or chronic depression-of an inaccurate SaMD generated by a microbiome-based model benchmarked by observational data can be severe, justifying consideration of potential Harm as well as intended Benefits when prioritizing a particular de-bugging task.

In a review of regulatory submission trajectories at the FDA for medicinal products, [27] identified lessons that can inform the development of AI tools for microbiome analysis. Submission failures are often linked to the absence of a clearly articulated scientific rationale. Gaps in analytical or clinical validation are also major causes of deficiency in the submission package. The use of a digital or model-based framework to describe supplemental components of microbiome modelling can enhance understandability and straightforwardness of application. Having an AI-enabled microbiome modelling program and the corresponding regulatory pathway clearly defined before initial development can significantly economize resource use and accelerate access to the market.

Clinical microbiome tools enhancing patient care through insights on the gut ecosystem may be implemented safely and effectively within healthcare today. [3] Regulation-navigating validation studies matching clinical microbiome applications with appropriate FDA or EMA pathways demonstrate concurrent alignment of analytical, clinical, operational, privacy, security, and ethical responsibilities. Sufficiently understanding microbiome data diversity permits identification of suitable evidence standards-summary-based investigations support robust microbiome AI evidence, satisfying amenability to alternative regulatory regimes.

AI-driven systems can improve the quality of care and assist compliance with safety guidelines and legal obligations. Maintaining workflow integration, governance, and risk management ensures systematic alignment with agency standards and patient welfare. Continuous monitoring safeguards patient welfare amid algorithm updates. Patient data privacy expands access to microbiome insights in diverse settings and permits system-level workflow automation.

Clinical trial designs for microbiome-based diagnostics

AI-powered diagnostics of human health and disease rely on omics science to capture molecular fingerprints of biological systems. Such information facilitates precise and individualized medical decisions that positively affect patient health throughout the life cycle. Data-driven models for inferring microbial co-occurrence patterns from amplicon and metagenomic surveys allow the reconstruction of both taxonomy and function across eukaryotes, bacteria, and viruses, enabling an understanding of healthy and diseased states spanning microbiomes, food webs, and the flourishing ecosystem of Earth. Because their spatiotemporal features are navigable from diverse environmental niches to the human biogeome, biosensors capable of providing the status of microbial multitudes simultaneously support early and actionable detection of disease causation-empowering microbiome diagnostics to discover health threats and infection risk, guide therapeutic selection, and personalize interventions. At the interface of these health imperatives lies a convergence of precision microbiome medicine and microbial diagnostics. Specifically, listed below are clinical trial designs appropriate for testing various ecological or identified signals of pathogens or health states together with AI-powered algorithms for predicting the presence or physiological state of targeted microorganisms in microbiome samples or biosensors [73]. Microbiome diagnostics hold the potential to greatly benefit human health.

Standardization challenges

In order to compare and integrate findings on microbiome dynamics across studies utilizing different sequencing methods, techniques, analysis tools, and reporting formats, the effort to standardize microbiome reporting has gained wide interest in the research community. Since microbiome diagnostics and clinical applications are projected to increase, current methods and data formats must meet minimum levels of standardization [74].

Chapter - 14

Future Frontiers: Gut-Brain Axis, Virome, and Beyond

Emerging microbiome domains (fungome, virome, mycobiome)

Explorations of the human microbiota typically focus on archetypal members of the bacteria and archaea domains of life, especially their evolutionary relationships, community composition, and functional capacities. However, other microbiome components and higher-level microbial taxonomic units are increasingly recognized. In 2017, the term “fungome” appeared in a broad call for integrative resolution of simultaneous fungal surveys spanning culture assays and next-generation sequencing. The “virome” describes viruses and their hosts; as biogeochemical roles of noncellular, filterable life forms gained recognition, viromic portraits were crafted for the 2011 human virome project. Similarly, “mycobiome” denotes fungal functional or phylogenetic service avatars: such harvesting enables association of observed anomalies or fabulations (species presence/absence, abundance shifts) to precise community controllers, guiding articulation of cause/effect systems governing complex microbial systems [35]. A substantial catalog of structured training sets of rich, well-documented longitudinal multimixture biocensus time-series, curing forbidding generation-scattering in epidemiological-speciation inference-forecasting between booming generation-scattering types, has thus begun to accrue, nourished by a general emergence of actively curated sample bank services; representative sets are cited.

Microbiome and neuroinfectious diseases

Pathogen propagation can facilitate the crossing of multiple bacteria through the blood-brain barrier [75]. Infection risk can increase due to the dysregulation of the oral microbiome. A proposed shift in the composition and activity of the gut-brain microbial communities can influence physiology, thereby promoting changes beyond the gut. Subject-microbiome relationships, physiological response, and infection predisposition can be assessed through microbiome characterizations.

AI's role in decoding complex microbial ecosystems

Artificial intelligence (AI) is crucial for deciphering intricate microbial ecosystems due to the complexity and dynamism of microbial interactivity, along with the diverse formats of AI-enabled data acquisition. By analyzing microbial datasets gathered at various locations on an individual, AI elucidates the interplay of community structure, function, and health condition and suggests opportunistic pathogens when infections are underway. Even after an infection has been pinpointed, AI can project population dynamics; model the competition, cooperation, and network interplay of community members; and estimate the cohabitation time of predominant species. Furthermore, AI helps recognize the temporal evolution of ecosystems, picks pertinent characteristics from high-dimensional databases, and formulates hypotheses on ecological principles governing the emergence of specific interdependencies [9]. At a strategic level, a human microbiome-based focus driving rapid tooling development enables an examination and operationalization of microbiome inputs for diverse other fields including residues, surfaces, and extinction;

deciphering the influence of human behavior on plant growth through soil microbiomes; and characterizing single-cell microbial dynamism to optimize reactor efficiency.

Chapter - 15

Conclusion and Vision for the Future

Synergy between AI and microbiome science

Artificial Intelligence (AI) and Machine Learning (ML) approaches provide tremendous and invaluable insights into the gut microbiome as related to human health and the complexities surrounding it. Precision medicine, a tailored approach that customizes medical prevention and treatments to the unique characteristics of each individual patient, leads to personalized insights derived from the specific composition of the gut microbiome and its interactions with overall health. Automated Machine Learning (AutoML) models are exceptionally well suited for analyzing microbiome data due to its inherent size, sparsity, and complexity, as well as the commonly shared standardized data formats that facilitate the seamless merging of diverse datasets collected from robots and sensors with detailed genome sequencing data. The ability to provision these extensive datasets leads to the development of unique emergent models that are specific to each individual and can persist for the entire duration of that model's applicability. The realms of AI and the microbiome have emerged as crucially important areas of study, both independently and synergistically, raising significant issues related to privacy, security, and equity in the handling of data. AI and ML applications are pervasive across numerous domains, leading to exciting new opportunities for modeling human health by unlocking previously inaccessible microbiome data as well as auxiliary genomic sequence data. The provision of these datasets remains critically important to enable further experimentation using AI and ML methodologies, and datasets can be efficiently merged to enhance analysis. A central goal in optimizing microbiome health is to prevent the resuscitation and proliferation of harmful pathogens, particularly in sensitive environments. The deteriorating lung-care environment faced by patients on ventilators, when managed by healthy caregivers, significantly elevates the long-standing challenges associated with hospital-acquired pneumonia. The passive acquisition of pathogens from mortified skin surfaces, which reside on various human surfaces through bacterial, fungal, and spore attributes, necessitates the development of automatic alerts to maintain the safety and wellness of patients and caregivers, extending to the global community. Recent academic literature has been utilized to thoroughly assess the key drivers of AI and microbiome interactions across various disciplines. Ensuring regulatory compliance plays a crucial role in establishing retrospective national cyber assurance frameworks and guarantees high reliability for real-time alerts, thereby allowing for the retraction of non-compliant articles when necessary. Practical details of the implementation remain firmly on track with the overarching aim of facilitating the development of a working End-to-End live system that operates at human skin dimensions, ultimately benefiting all stakeholders involved [76, 77, 78].

Toward fully personalized infectious disease care

Infectious diseases pose a significant threat to global health, and antimicrobial resistance is threatening to render our most effective antibiotics useless. Many commercially available pathogen-detection methods still fail to identify the infecting organism more than 50% of the time, and diagnosis remains under-investigated compared to other medical specialties. As a result, critically ill patients are often started on empiric (guess) therapy that may be inappropriate. Infectious disease accounts for more than 30% of patients who develop acute-on-chronic kidney disease due to nephrotoxic antibiotics. These, coupled with significant logistical or analytical hurdles, remain deeply understudied [79].

Precision medicine strives to identify the right treatment for the right patient at the right time. Competent models that address the time dimension are thus critically needed [57]. By combining traditional micro-biology with the latest advances in AI and biosensors, a fully personalized approach to infectious disease is slowly emerging. However, two major limitations hinder their clinical implementation: the absence of fully structured training datasets to build and validate reliable predictive models, and the need for extensive data pre-processing and a defined machine-learning strategy at the implementation stage.

Challenges and opportunities ahead

Practical, ethical, and scientific challenges must be addressed to realise the full potential of microbiome-informed diagnostics. Advancing the microbiome field requires new experimental techniques, data types, and modelling paradigms. At the experimental level, more easily interpretable biosensors could reveal pathogen-specific signatures that other modalities fail to detect. Strain-resolved metagenomic data would clarify the precise roles of individual microbial members. Detailed host immune response readouts would directly link microbiome functions to infection and inform decoupled immune-targeted diagnostics. Multi-organ, interkingdom, or viral approaches could explore other parts of the body or different infectious agents to test for emergent phenomena through microbiome influences. Novel assay formats that non-invasively extract signals from unanticipated sources present interesting avenues for exploration.

Regulatory demands for algorithmic transparency are rising significantly in tandem with the accelerated adoption of artificial intelligence across various sectors and industries. One important avenue of this increased scrutiny focuses on the critical need for a broader and more comprehensive representation within training cohorts. This broader representation is essential to mitigate the risk of faction-differentiated rollout that may arise from biases in training data. To effectively address this challenge, proactive assessment methods should be employed to systematically gauge the generality of the underlying modelling assumptions, thereby identifying any remaining side effects that could potentially threaten equitable access to AI systems. Furthermore, interpretability techniques that trace model decisions back to specific input features are instrumental in allowing for thorough vulnerability checks and ensuring accountability. This process not only fosters greater trust in the classifiers but also unlocks crucial opportunities to delve deeper into understanding their operational principles. By committing to such precautionary readiness, stakeholders can ensure adequate preparedness for resource-limited scenarios. This includes exploring opportunities to streamline combinations of multisource inputs, enhancing overall system efficiency and outcomes. Additionally, it opens the door for potential redeployment of these algorithms for other microbiome-related applications, broadening the impact of AI in healthcare and related fields [80, 81, 82, 83].

Chapter - 16

Conclusion

A pandemic is a starting point for action but also an alarm bell for systemic transitions. Although it is impossible to predict whether such a drastic, yet necessary, transformation will occur, the experience of the COVID-19 pandemic has fundamentally altered perspectives on the impact of infectious diseases on society and the economy - on the micro-scale as well as the macro-scale. In the context of the human microbiome, pathogenic bacteria need to be tackled swiftly. The AI-based precision microbiome medicine and the AI-driven infectious disease diagnostic systems proposed in this research can serve as indispensable monitors, deterrent devices, rapid responders and precise protectors. They are of unprecedented importance in situations of growing antibiotic resistance, re-emerging infectious diseases and viruses incorporating bacterial systems. A third wave of dangers is propagating such infectious threats, and at the same time, rapidly communicating them through instant technologies intensifies their dissemination enormously. The combined authority of biotechnology and AI profoundly transforms international society.

References

1. H. Fongang, A. T. Mbaveng, and V. Kuete, "Global burden of bacterial infections and drug resistance," Advances in Botanical Research, 2023. [HTML]
2. C. J. L. Murray, K. S. Ikuta, F. Sharara, L. Swetschinski, *et al.*, "Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis," *The Lancet*, 2022. thelancet.com
3. D. W. Denning, "Global incidence and mortality of severe fungal disease," The Lancet Infectious Diseases, 2024. thelancet.com
4. T. M. Kuntz and J. A. Gilbert, "Introducing the microbiome into Precision Medicine," 2016. [PDF]
5. B. Peñalver Bernabé, L. Cralle, and J. A. Gilbert, "Systems Biology of the human microbiome," 2017. [PDF]
6. X. C. Morgan and C. Huttenhower, "Chapter 12: Human Microbiome Analysis," 2012. ncbi.nlm.nih.gov
7. T. Sun, X. Niu, Q. He, F. Chen *et al.*, "Artificial Intelligence in microbiomes analysis: A review of applications in dermatology," 2023. ncbi.nlm.nih.gov
8. W. K. Mousa, F. Chehadeh, and S. Husband, "Recent Advances in Understanding the Structure and Function of the Human Microbiome," 2022. ncbi.nlm.nih.gov
9. Y. P Shelke, A. K Badge, and N. J Bankar, "Applications of Artificial Intelligence in Microbial Diagnosis," 2023. ncbi.nlm.nih.gov
10. I. Ferrocino, K. Rantsiou, R. McClure, *et al.*, "The need for an integrated multi-OMICs approach in microbiome science

in the food system," *Comprehensive Reviews in Food Science and Food Safety*, vol. 2023, Wiley Online Library. wiley.com

11. S. Y. Yang, S. M. Han, J. Y. Lee, K. S. Kim, "Advancing Gut Microbiome Research: The Shift from Metagenomics to Multi-Omics and Future Perspectives," *Journal of Microbiology and...*, 2025. nih.gov
12. A. Mathuria, N. Ali, I. Mani, and V. Singh, "Overview on multi-omics research in microbiome analysis," in *Analysis of the Human Microbiome*, 2024, Springer. [HTML]
13. M. Kamel, S. Aleya, M. Alsubih, and L. Aleya, "Microbiome dynamics: a paradigm shift in combatting infectious diseases," Journal of personalized medicine, 2024. mdpi.com
14. W. K. Mousa and A. Al Ali, "The gut microbiome advances precision medicine and diagnostics for inflammatory bowel diseases," International journal of molecular sciences, 2024. mdpi.com
15. V. Shukla, S. Singh, S. Verma, S. Verma, A. A. Rizvi, "Targeting the microbiome to improve human health with the approach of personalized medicine: Latest aspects and current updates," Clinical Nutrition, vol. 2024, Elsevier. [HTML]
16. M. O. Yaqub, A. Jain, C. E. Joseph, and L. K. Edison, "Microbiome-driven therapeutics: from gut health to precision medicine," Gastrointestinal Disorders, 2025. mdpi.com
17. J. T. Ladner, N. D. Grubaugh, O. G. Pybus, and K. G. Andersen, "Precision epidemiology for infectious disease control," 2019. ncbi.nlm.nih.gov
18. Q. Liu, X. Jin, J. Cheng, H. Zhou, "Advances in the application of molecular diagnostic techniques for the

detection of infectious disease pathogens," *Molecular Medicine*, 2023. spandidos-publications.com

19. M. Altindiş and E. P. Kahraman Kilbaş, "Managing viral emerging infectious diseases via current and future molecular diagnostics," Diagnostics, 2023. mdpi.com
20. E. E. Hilt and P. Ferrieri, "Next generation and other sequencing technologies in diagnostic microbiology and infectious diseases," Genes, 2022. mdpi.com
21. J. Kumar Narayana, M. Mac Aogáin, W. Wen Bin Goh, K. Xia *et al.*, "Mathematical-based microbiome analytics for clinical translation," 2021. ncbi.nlm.nih.gov
22. S. M Gibbons, T. Gurry, J. W Lampe, A. Chakrabarti *et al.*, "Perspective: Leveraging the Gut Microbiota to Predict Personalized Responses to Dietary, Prebiotic, and Probiotic Interventions," 2022. ncbi.nlm.nih.gov
23. D. Traversi, A. Pulliero, A. Izzotti, E. Franchitti, "Precision medicine and public health: new challenges for effective and sustainable health," in *Personalized Medicine*, 2021. mdpi.com
24. M. C. Roberts, K. E. Holt, G. Del Fiol, A. A. Baccarelli, *et al.*, "Precision public health in the era of genomics and big data," *Nature Medicine*, 2024. nih.gov
25. A. Gerussi, M. Scaravaglio, L. Cristoferi, D. Verda, "Artificial intelligence for precision medicine in autoimmune liver disease," Frontiers in ..., 2022. frontiersin.org
26. C. K. Nkera-Gutabara, R. Kerr, J. Scholefield, S. Hazelhurst *et al.*, "Microbiomics: The Next Pillar of Precision Medicine and Its Role in African Healthcare," 2022. ncbi.nlm.nih.gov
27. X. Yu, X. Chen, and Z. Wang, "Characterizing the Personalized Microbiota Dynamics for Disease

Classification by Individual-Specific Edge-Network Analysis," 2019. [PDF]

28. P. Sudhakar, K. Machiels, B. Verstockt, T. Korcsmaros *et al.*, "Computational Biology and Machine Learning Approaches to Understand Mechanistic Microbiome-Host Interactions," 2021. ncbi.nlm.nih.gov
29. M. Ahmad Malla, A. Dubey, A. Kumar, S. Yadav *et al.*, "Exploring the Human Microbiome: The Potential Future Role of Next-Generation Sequencing in Disease Diagnosis and Treatment," 2019. [PDF]
30. K. Ratiner, D. Ciocan, S. K. Abdeen, and E. Elinav, "Utilization of the microbiome in personalized medicine," *Nature Reviews*, 2024. [HTML]
31. G. S. Watts, K. Youens-Clark, M. J. Slepian, D. M. Wolk *et al.*, "16S rRNA gene sequencing on a benchtop sequencer: accuracy for identification of clinically important bacteria," 2017. ncbi.nlm.nih.gov
32. M. D. Weinroth, A. D. Belk, C. Dean, N. Noyes, *et al.*, "Considerations and best practices in animal science 16S ribosomal RNA gene sequencing microbiome studies," *Journal of Animal Science*, vol. 2022. oup.com
33. A. Hiergeist, J. Ruelle, S. Emler, and A. Gessner, "Reliability of species detection in 16S microbiome analysis: Comparison of five widely used pipelines and recommendations for a more standardized approach," PLoS One, 2023. plos.org
34. D. Grinevich, L. Harden, S. Thakur, and B. Callahan, "Serovar-level identification of bacterial foodborne pathogens from full-length 16S rRNA gene sequencing," Msystems, 2024. asm.org
35. X. Zhang, L. Li, J. Butcher, A. Stintzi *et al.*, "Advancing functional and translational microbiome research using meta-omics approaches," 2019. ncbi.nlm.nih.gov

36. P. Li, H. Luo, B. Ji, and J. Nielsen, "Machine learning for data integration in human gut microbiome," 2022. ncbi.nlm.nih.gov
37. C. Fiocchi, "Omics and Multi-Omics in IBD: No Integration, No Breakthroughs," 2023. ncbi.nlm.nih.gov
38. M. Unal, E. Bostanci, C. Ozkul, K. Acici *et al.*, "Crohn’s Disease Prediction Using Sequence Based Machine Learning Analysis of Human Microbiome," 2023. ncbi.nlm.nih.gov
39. S. Jae Lee and M. Rho, "Multimodal deep learning applied to classify healthy and disease states of human microbiome," 2022. ncbi.nlm.nih.gov
40. V. Vashisht, A. Vashisht, A. K. Mondal, J. Farmaha, "Genomics for emerging pathogen identification and monitoring: Prospects and obstacles," 2023. mdpi.com
41. R. S. Jayawardena, K. D. Hyde, A. R. G. de Farias, *et al.*, "What is a species in fungal plant pathogens?," Fungal Diversity, vol. 2021, Springer. academia.edu
42. N. Li, Q. Cai, Q. Miao, Z. Song *et al.*, "High-throughput metagenomics for identification of pathogens in the clinical settings," Small methods, 2021. nih.gov
43. S. Acharjee Dip, U. Acharjee Shuvo, T. Chau, H. Song *et al.*, "PathoLM: Identifying pathogenicity from the DNA sequence through the Genome Foundation Model," 2024. [PDF]
44. A. B. Isiaka, V. N. Anakwenze, C. R. Ilodinso, "Harnessing artificial intelligence for early detection and management of infectious disease outbreaks," Journal of Innovative, 2024. ssrn.com
45. I. Villanueva-Miranda, G. Xiao, and Y. Xie, "Artificial Intelligence in Early Warning Systems for Infectious Disease

Surveillance: A Systematic Review," Frontiers in Public Health, 2025. frontiersin.org

46. C. El Morr, D. Ozdemir, Y. Asdaah, "AI-based epidemic and pandemic early warning systems: A systematic scoping review," Health Informatics, 2024. sagepub.com
47. H. Ali, "AI for pandemic preparedness and infectious disease surveillance: predicting outbreaks, modeling transmission, and optimizing public health interventions," Int J Res Publ Rev, 2024. researchgate.net
48. G. Papoutsoglou, S. Tarazona, M. B. Lopes, T. Klammsteiner *et al.*, "Machine learning approaches in microbiome research: challenges and best practices," 2023. ncbi.nlm.nih.gov
49. G. Armstrong, G. Rahman, C. Martino, "Applications and comparison of dimensionality reduction methods for microbiome data," *Frontiers in …*, 2022. frontiersin.org
50. L. J. Marcos-Zambrano *et al.*, "Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment," *Frontiers in Microbiology*, vol. 12, 2021. frontiersin.org
51. M. Teixeira, F. Silva, R. M. Ferreira, T. Pereira, "A review of machine learning methods for cancer characterization from microbiome data," NPJ Precision Medicine, 2024. nature.com
52. M. A. Mujawar, H. Gohel, S. K. Bhardwaj, S. Srinivasan *et al.*, "Nano-enabled biosensing systems for intelligent healthcare: towards COVID-19 management," 2020. ncbi.nlm.nih.gov
53. FC Udegbe, EI Nwankwo, GT Igwama, "Real-time data integration in diagnostic devices for predictive modeling of

infectious disease outbreaks," Computer Science & IT, 2023. researchgate.net

54. A. O. Olatunji, J. A. Olaboye, C. C. Maha, "Revolutionizing infectious disease management in low-resource settings: The impact of rapid diagnostic technologies and portable devices," Journal of Applied ..., vol. 2024. researchgate.net
55. M. P. McRae, K. S. Rajsri, T. M. Alcorn, and J. T. McDevitt, "Smart Diagnostics: Combining Artificial Intelligence and In Vitro Diagnostics," 2022. ncbi.nlm.nih.gov
56. N. Brasier, M. Osthoff, F. De Ieso, and J. Eckstein, "Next-Generation Digital Biomarkers for Tuberculosis and Antibiotic Stewardship: Perspective on Novel Molecular Digital Biomarkers in Sweat, Saliva, and Exhaled Breath," 2021. ncbi.nlm.nih.gov
57. L. Bissonnette and M. G. Bergeron, "Infectious Disease Management through Point-of-Care Personalized Medicine Molecular Diagnostic Technologies," 2012. ncbi.nlm.nih.gov
58. M. C. R. Melo and J. R. M. A. Maasch, "Accelerating antibiotic discovery through artificial intelligence," *Communications*, 2021. nature.com
59. AM Mohammed, JK Oleiwi, AF Osman, "Enhancing antimicrobial resistance strategies: Leveraging artificial intelligence for improved outcomes," South African Journal, 2025. journals.co.za
60. F. Pennisi, A. Pinto, G. E. Ricciardi, C. Signorelli, "The role of artificial intelligence and machine learning models in antimicrobial stewardship in public health: a narrative review," Antibiotics, 2025. mdpi.com
61. S. A. Boers, S. D. Hiltemann, A. P. Stubbs, R. Jansen *et al.*, "Development and evaluation of a culture-free microbiota profiling platform (MYcrobiota) for clinical diagnostics," 2018. ncbi.nlm.nih.gov

62. L. E. McCoubrey, S. Thomaidou, M. Elbadawi, S. Gaisford *et al.*, "Machine Learning Predicts Drug Metabolism and Bioaccumulation by Intestinal Microbiota," 2021. ncbi.nlm.nih.gov
63. L. E. McCoubrey, M. Elbadawi, M. Orlu, S. Gaisford *et al.*, "Machine Learning Uncovers Adverse Drug Effects on Intestinal Bacteria," 2021. ncbi.nlm.nih.gov
64. A. Z. Al Meslamani, I. Sobrino, and J. de la Fuente, "Machine learning in infectious diseases: potential applications and limitations," 2024. ncbi.nlm.nih.gov
65. B. van den Bogert, J. Boekhorst, W. Pirovano, and A. May, "On the Role of Bioinformatics and Data Science in Industrial Microbiome Applications," 2019. ncbi.nlm.nih.gov
66. Y. Yang, M. Lin, H. Zhao, Y. Peng *et al.*, "A survey of recent methods for addressing AI fairness and bias in biomedicine," 2024. ncbi.nlm.nih.gov
67. H. Cheng, F. Strouts, T. E Sweeney, T. Briese *et al.*, "Integration of Next-Generation Sequencing, Viral Sequencing, and Host-Response Profiling for the Diagnosis of Acute Infections," 2017. ncbi.nlm.nih.gov
68. M. Han, Y. Zha, H. Chong, C. Zhong *et al.*, "Utilizing microbiome approaches to assist source tracking, treatment and prevention of COVID-19: Review and assessment," 2020. ncbi.nlm.nih.gov
69. T. Khinvasara, N. Tzenios, and A. Shanker, "Post-market surveillance of medical devices using AI," in *… and Alternative Medical …*, 2024. researchgate.net
70. R. Cioeta, A. Cossu, E. Giovagnoni, M. Rigoni, "A new platform for post-marketing surveillance and real-world evidence data collection for substance-based medical

devices," Frontiers in Drug Safety, vol. 2022. frontiersin.org

71. T. Judd, "Post-market Surveillance (PMS) of Medical Devices: From a Clinical Engineering Perspective," in *… of Medical Devices: For Regulatory Purposes*, 2023, Springer. [HTML]
72. S. Manetti, E. Guidotti, F. Vola, and M. Vainieri, "A systematic literature review of real-world evidence (RWE) on post-market assessment of medical devices," Health Economics, Policy, vol. 2025, Cambridge University Press. cambridge.org
73. N. K. Dudek, M. Chakhvadze, S. Kobakhidze, O. Kantidze *et al.*, "Supervised machine learning for microbiomics: bridging the gap between current and best practices," 2024. [PDF]
74. D. D'Elia, J. Truu, L. Lahti, M. Berland *et al.*, "Advancing microbiome research with machine learning: key findings from the ML4Microbiome COST action," 2023. ncbi.nlm.nih.gov
75. M. Obrenovich, H. Jaworski, T. Tadimalla, A. Mistry *et al.*, "The Role of the Microbiota-Gut-Brain Axis and Antibiotics in ALS and Neurodegenerative Diseases," 2020. ncbi.nlm.nih.gov
76. A. Patil, N. Singh, M. Patwekar, F. Patwekar, and A. Patil, "AI-driven insights into the microbiota: Figuring out the mysterious world of the gut," Intelligent Systems, 2025. sciencedirect.com
77. T. C. Dakal, C. Xu, and A. Kumar, "Advanced computational tools, artificial intelligence and machine-learning approaches in gut microbiota and biomarker identification," Frontiers in Medical Technology, 2025. frontiersin.org
78. S. Malakar, P. Sutaoney, H. Madhyastha, "Understanding gut microbiome-based machine learning platforms: A review on

therapeutic approaches using deep learning," *Chemical Biology &...*, 2024. [HTML]

79. A. Savoldi, N. T. Mutters, and E. Tacconelli, "Personalized infection prevention and control: a concept whose time has arrived," 2023. ncbi.nlm.nih.gov
80. M. Busuioc, "AI algorithmic oversight: new frontiers in regulation," Handbook of regulatory authorities, 2022. elgaronline.com
81. G. Chaudhary, "Unveiling the black box: Bringing algorithmic transparency to AI," Masaryk University Journal of Law and Technology, 2024. muni.cz
82. B. C. Cheong, "Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making," Frontiers in Human Dynamics, 2024. frontiersin.org
83. H. B. Harvey and V. Gowda, "Regulatory issues and challenges to artificial intelligence adoption," Radiologic Clinics, 2021. [HTML]