A Review on the Use of Artificial Intelligence for Human Microbiota Analysis in Clinical Tasks

R. Janwadkar[1], L.T. Ramos[2,3], N. Payahuala-Díaz[4], F. Rivas-Echeverría[1,3], E. Diaz[1] and E. Casas[3]

1. Dept. of Comp. Sci. and Eng., The University of Texas at Arlington, Arlington, TX 76019, USA
2. Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, 08193, Spain
3. Kauel Inc., Menlo Park, Silicon Valley, CA 94025, USA
4. Servicio de Salud Osorno, Osorno, 825, Chile

rxj6469@mavs.uta.edu, ltramos@cvc.uab.cat, nidia.payahuala@redsalud.gob.cl, francklin.rivas@uta.edu, elizabeth.diaz@uta.edu, edmundo.casas@kauel.com

Abstract— **This review briefly investigates how artificial intelligence (AI) methods are being applied to microbiota-based clinical tasks. We analyse a curated set of peer-reviewed studies to characterize the models used, the nature of microbiota-derived inputs, and the clinical goals addressed. Our findings show a preference for classical machine learning approaches, especially random forests, due to their robustness and interpretability. Deep learning methods are less frequent and primarily employed in multimodal contexts. Most studies focus on disease prediction or classification, though some explore treatment response or drug-microbiota interactions. Gut-derived profiles dominate the input data, with limited exploration of other microbiota niches. Key challenges include the lack of external validation, inconsistent preprocessing practices, and limited use of explainability techniques. These observations point to the need for more standardized, transparent, and clinically grounded research to advance the integration of AI with microbiome science.**

Keywords— microbiota, microbiome, machine learning, deep learning, computational biology, diagnosis

I. Introduction

The human microbiota consists of trillions of microorganisms that inhabit different body sites, including the gut, skin, oral cavity, and urogenital tract [1, 2]. These microbial communities play essential roles in digestion, immune regulation, and protection against pathogens [3]. Disruptions to this microbial ecosystem, known as dysbiosis, have been associated with a wide range of medical conditions, including colorectal cancer, type 2 diabetes, and inflammatory bowel disease [4, 5]. Beyond disease association, information derived from the microbiota can also support research into treatment response, or patient stratification, broadening its potential use in medical tasks [6, 7].

Despite the increasing interest in leveraging microbiota information for clinical purposes, traditional analysis techniques still face important limitations [8]. Statistical and bioinformatic methods such as 16S rRNA gene sequencing and metagenomic profiling have enabled significant descriptive insights, but often fall short when applied to predictive or diagnostic tasks [9]. These approaches struggle with high-dimensional, sparse, and noisy data, as well as inter-individual variability [8]. Moreover, the compositional nature of microbial profiles and the lack of standardized reference frameworks complicate reproducibility and scalability, hindering clinical translation [10].

Artificial intelligence (AI), encompassing machine learning (ML) and deep learning (DL) techniques, offers a powerful framework to address many of the challenges associated with medical data analysis [11, 12]. By learning patterns from complex, high-dimensional datasets [13, 14], these methods can support tasks such as classification, prediction, segmentation, and feature selection [15, 16], potentially overcoming issues of variability, sparsity, and non-linearity. In clinical contexts, AI has the potential to assist in hypothesis generation, patient stratification [17], and the identification of microbiota-based signatures, paving the way for more personalized and data-driven medical interventions.

Given the emerging role of AI in medical research, this review explores its integration with microbiota analysis. While microbiome data holds increasing relevance for understanding health and disease, the use of AI to process, interpret, and extract meaningful insights from such data remains a growing area of interest. In this context, the present work offers a concise overview of how AI methods are being applied to microbiota-based studies, with attention to their clinical orientation, methodological approaches, and application domains. The objectives of this review are the following:

- Identify how AI is currently being used in conjunction with microbiota data.
- Characterize the clinical purposes it supports.
- Examine the types of data and models involved.
- Discuss the limitations and open challenges reported in the literature.
- Provide insights into the future direction of AI in microbiota research.

II. Methodology

This review follows a structured process inspired by the principles of the Preferred Reporting Items for

Systematic Reviews and Meta-Analyses (PRISMA) [18]. In line with these guidelines, the study was designed around explicit research questions, clear eligibility criteria, and a systematic strategy for identifying and organizing the literature.

First, we defined the following research questions to guide the review:

- What artificial intelligence techniques have been applied to human microbiota data for clinical purposes?
- How do traditional machine learning approaches compare to deep learning methods in this domain?
- What clinical tasks are addressed using microbiota-based AI?
- What methodological challenges and limitations are reported across studies?
- What open challenges and future directions emerge for AI-driven microbiota analysis in clinical research?

Next, we established inclusion and exclusion criteria to ensure consistency in the selection of works. The inclusion criteria is as follows:

- Empirical studies presenting original experimental work.
- Articles published in peer-reviewed scientific journals or presented at scientific conferences.
- Articles explicitly applying AI methods (machine learning or deep learning) to human microbiota data with clinical relevance.
- Articles written in English.

Then, we established the following exclusion criteria to filter out works that did not meet the scope of the review:

- Reviews, editorials, meta-analyses, commentaries, or tutorials.
- Non-peer-reviewed sources (e.g., preprints, technical reports).
- Articles lacking sufficient methodological or experimental detail to allow meaningful interpretation.

The literature analyzed in this review was retrieved from reputable databases to guarantee the quality and reliability of the included works. Specifically, we consulted PubMed, IEEE Xplore, ScienceDirect, and SpringerLink. In addition, Google Scholar was employed as a complementary tool to locate further relevant studies. The search strategy relied on targeted keywords and Boolean combinations, including: 'human microbiota,' 'microbiome,' 'artificial intelligence,' 'machine learning,' 'deep learning,' and 'clinical applications.'

All retrieved articles were manually screened for relevance according to the inclusion and exclusion criteria. For each selected work, we extracted information regarding: (i) the type of AI method employed, (ii) the nature of the microbiota data analysed, (iii) the clinical objective (e.g., diagnosis, prognosis, treatment prediction), (iv) reported performance metrics, and (v) acknowledged limitations or challenges. This structured extraction provided the basis for the synthesis and comparative analysis presented in the following sections.

III. Surveyed Works by Learning Paradigm

This section provides an overview of each selected study, summarizing their main objectives, methodologies, and key findings. To facilitate clarity and comparison, the works are grouped according to the type of AI technique employed. Two primary categories were identified: studies based on traditional ML algorithms, and those leveraging DL methods.

III-A. Machine Learning

Most of the studies included in this review adopt ML as their primary strategy, reflecting a notable preference for classical and interpretable approaches over more complex DL architectures.

The application of microbiota-derived features for cancer-related prediction tasks gains notable traction within the reviewed literature. In [19], gut microbial strains are explored as pretreatment biomarkers to predict patient response to combination immune checkpoint blockade (CICB) therapies across multiple cancer types. Using strain-level metagenomic profiles and random forest (RF) classifiers, the study identifies a core set of 22 strains capable of achieving AUCs of 0.73 and 0.70 for response and 12-month progression-free survival, respectively.

Complementarily, [20] presents DeepMicroCancer, a multiclass classification framework trained on microbial abundance profiles from over 11,000 tumor tissue samples covering 21 cancer types. The method achieves an AUC of 0.95 and demonstrates the utility of transfer learning to improve predictions in blood-derived data, notably increasing AUC for lung adenocarcinoma from 0.80 to 0.89. Finally, [21] addresses the use of gut microbiota as a noninvasive biomarker for early-stage lung cancer, training a support vector machine (SVM) on OTUs selected through mRMR and achieving AUCs of 97.6% in the discovery cohort and 76.4% in validation. This work further highlights microbial shifts in key taxa and functional alterations in bile acid metabolism.

Other studies address liver-related diseases using microbiota data and ML approaches. In [22], the authors benchmark a range of ML algorithms for microbiome-based host trait prediction, including cases of liver cirrhosis. Using OTU tables from 16S rRNA sequencing across 17 datasets, the study compares RF, SVM, logistic regression, gradient boosting, and neural networks (feed-forward multilayer neural networks). RF consistently yields strong performance, while neural networks require substantial tuning.

Additionally, a Hierarchical Feature Engineering (HFE) strategy is proposed to reduce OTU dimensionality and improve accuracy on sparse datasets. In a more specific context, [23] analyses gut microbiota associations with liver transcriptomic profiles and clinical outcomes in HBV-related hepatocellular carcinoma (HCC). By correlating 310 microbial OTUs with over 5,000 liver-expressed genes, the study identifies genera like Bacteroides and Clostridium XIVa as linked to smaller tumors and favorable immune expression. These features are then used in RF and SVM models to predict prognosis, with AUCs of 0.81 for 5-year survival and 0.70 for 2-year disease-free survival.

Gastrointestinal disorders are another focus of microbiota-based ML. In [24], gut microbiota composition is used to predict neratinib-induced diarrhea in elderly patients with HER2-positive breast cancer. Using 16S rRNA data from 50 stool samples across 11 individuals, an XGBoost classifier with SHAP-based interpretability achieves AUROC = 0.88 and AUPRC = 0.95, identifying protective species like Ruminiclostridium 9 and Bacteroides sp. HPS0048. However, the study is limited by sample size and lack of external validation. A broader effort is found in [25], which applies ridge logistic regression to distinguish Crohn's disease (CD) from ulcerative colitis (UC) using whole-metagenome data from 482 individuals. Reported AUCs are 0.873 (training), 0.778 (test), and 0.633 (external). Distinct biomarkers include E. coli and Shigella dysenteriae in CD, and Prevotella spp. in UC, though challenges such as confounding, batch effects, and microbial heterogeneity persist.

Other studies use oral microbiota as input for clinical prediction. In [26], oral microbiota and salivary metabolites are combined to diagnose oral lichen planus (OLP). Using 16S rRNA sequencing and untargeted metabolomics from 200 saliva samples, several ML models (LASSO, SVM-RFE, XGBoost, RF, ANN) are trained. A panel including Pseudomonas, Rhodococcus, and the metabolite (±)10-HDoHE yields an AUC of 0.890. OLP samples show reduced diversity and altered metabolite profiles, with microbe-metabolite associations pointing to inflammatory pathways.

In [27], oral microbiota is used to predict preterm birth (PTB) from 59 pregnant women. An RF classifier trained on selected taxa achieves 0.765 balanced accuracy and 87.5% accuracy on a twin pregnancy subset. PTB is linked to the depletion of protective taxa rather than pathogenic enrichment, indicating a preventive microbiota signature.

Beyond organ-specific disorders, several studies illustrate the versatility of microbiome-based ML in diverse health contexts. In [4], data from over 4,100 individuals is used to examine microbial circadian rhythms in type 2 diabetes (T2D). Loss of diurnal rhythmicity in specific taxa enables the construction of a logistic regression model (AUC = 0.79), improved to 0.87 with BMI and clinical features. External validation supports microbial rhythmicity loss as an early biomarker of metabolic dysfunction. For cardiovascular disease (CVD), [28] trains a RF model on fecal microbiota, achieving an AUC of 0.70 using the top 25 high-contribution OTUs. CVD samples show enrichment of Bacteroides and Veillonella, while Faecalibacterium and Alistipes dominate in controls, underscoring the value of OTU-level features.

Microbiome data have also been used to predict pharmacological effects. [29] develops a ML classifier to determine whether small-molecule drugs are susceptible to microbial depletion via gut microbiota, through either metabolism or bioaccumulation. The study compiles a dataset of 455 compounds labeled and encodes them using Morgan fingerprints and physicochemical properties derived from SMILES strings. Among the tested algorithms, the extra trees model performs best, reaching an AUROC of 75.1%, with 80.2% precision and 79.2% recall after hyperparameter tuning. This enables non-experimental, preclinical assessment of drug-microbiota interactions, supporting informed decisions in early-stage drug development.

Finally, [30] addresses multiclass disease classification using a ML framework trained on fecal metagenomic profiles to distinguish nine disease phenotypes within a single model. Among the tested algorithms, RF yields the best results, with AUROC scores between 0.90 and 0.99 on the internal test set. External validation across 12 datasets shows reasonable generalization, albeit with some performance decline. The model identifies over 360 microbial species linked to specific diseases, enabling detailed microbiota-phenotype mapping. Remaining challenges include confounding variables, limited phenotype diversity, and low biological interpretability.

III-B. Deep Learning

A small subset of the reviewed studies adopts DL as the primary modeling strategy. Notably, these works rely on convolutional neural networks (CNNs) as their core predictive approach, often adapted to exploit the structural and compositional characteristics of microbiome data.

The study in [31] presents MDeep, a phylogeny-aware CNN that integrates microbial taxonomic hierarchies into its architecture by aligning convolutional layers with taxonomic ranks. Evaluated on both simulated and real 16S rRNA datasets, MDeep outperforms baselines like RF and lasso in classification and regression tasks, including disease prediction (e.g., rheumatoid arthritis, kwashiorkor) and demographic traits (age, gender), achieving higher F1 and R^2 scores and showing robustness to preprocessing. However, its performance declines with sparse microbial signals, and interpretability is limited. In a different context, [32] develops a CNN-based model for staging alcohol-associated liver disease (ALD) and MASLD from gut microbiota profiles. Among several classifiers, CNN with PCA achieves the best results, with AUCs of 0.94-0.97 for ALD and 0.93 for MASLD. External validation confirms generalizability (AUC > 0.90 for ALD, 0.88 for MASLD), highlighting CNNs' effectiveness in modeling complex microbial patterns.

In [33], CNNs are used to classify dermatological conditions from microbiome and image data. Models like VGG16 and InceptionV3 trained on nail images reach AUCs up to 0.960 for fungal infection detection, while gut microbiota profiles achieve 0.929 accuracy for vitiligo diagnosis. These results showcase CNNs' adaptability across modalities, though performance declines with low-resolution inputs and underrepresented groups, highlighting the need for broader datasets. Similarly, [34] introduces MetaDR, a DL framework that integrates taxonomic structure with microbial abundance vectors using EPCNNs and a weighted RF for interpretability. Applied to liver cirrhosis, colorectal cancer, and T2D prediction, it achieves AUCs of 0.9535 and 0.9063 for the first two, outperforming models like DeepMicro. Despite identifying key microbial biomarkers, the lack of external validation and model complexity raise overfitting and generalization concerns.

Beyond phenotype classification, DL has also been applied to high-resolution analysis of microbiota composition. [35] presents DERSI, a DL approach for high-resolution identification of amplicon sequence variants (ASVs) from 16S rRNA data. A CNN projects input sequences into a 10D latent space trained on 117,000 reference V4 sequences, followed by denoising and abundance estimation. DERSI outperforms VSEARCH, DADA2, and UNOISE3 in precision and recall on mock datasets (e.g., 99/95 on Mock-16), achieves Bhattacharyya coefficients up to 99.78 for abundance estimates, and enables phenotype separation via PCA. It also reduces processing time on large-scale datasets.

IV. Clinical Purpose of AI Integration

A review of the selected studies reveals a diverse set of clinical objectives underlying the integration of AI with microbiota-related data. While disease detection remains a common focus, it is far from the only application explored. Broadly, the studies reflect a spectrum of diagnostic, prognostic, and therapeutic support goals, highlighting the expanding role of microbiota in multiple stages of clinical decision-making.

Most notably, several works center on disease diagnosis or classification, targeting conditions such as colorectal cancer, liver disease, inflammatory bowel disease, oral lichen planus, and even dermatological conditions [21, 25-27, 32-34] In these cases, microbiota data is typically used as a non-invasive biomarker source to distinguish between healthy and pathological states. However, other studies adopt more prognostic or predictive objectives, such as therapy response, patient survival, or cardiovascular risk prediction [4, 19, 23, 28].

A particularly relevant trend is the emergence of multi-disease frameworks, where models attempt to handle multiple diagnostic targets simultaneously, as seen in studies on cancer subtype classification or general disease phenotype identification [20, 30]. It is also important to note that not all models aim to detect pathology. Some are designed to distinguish between disease stages, evaluate liver enzyme elevation, or assess risk of adverse events such as drug-induced diarrhea or preterm birth [24, 27, 31]. These objectives suggest that AI is being used not only for binary classification but also to refine diagnostic granularity or guide preventive interventions in complex clinical scenarios.

In addition, the predictive modeling of drug-microbiota interactions represents a distinct use case, emphasizing the relevance of AI in pharmacological contexts and early-stage therapeutic screening [29]. Overall, the reviewed literature shows that AI applications in microbiota studies are clinically varied and evolving, with a clear trend toward supporting personalized and predictive medicine across diagnostic, prognostic, and therapeutic dimensions.

V. Analysis of Models and Approaches

The reviewed studies span both classical ML and DL paradigms, with a clear predominance of the former. RF appears as the most frequently employed model across studies, reflecting its flexibility and robustness to noise in compositional and sparse microbiome datasets [19, 20, 22, 23, 27, 28, 30]. Its widespread use may also relate to its low tuning requirements and compatibility with typical preprocessing pipelines such as OTU filtering and log-ratio transformations. Moreover, RF models often allow for the extraction of interpretable feature importance scores, making them attractive in biomarker discovery contexts.

Other classical models such as SVM, logistic regression, and gradient boosting are also commonly used, albeit with more modest adoption. SVMs tend to appear in studies with smaller or more curated feature sets [21, 26], where margin-based separation is advantageous. Logistic regression is used mainly in baseline comparisons or when simplicity and interpretability are prioritized [4, 25]. Ensemble techniques like XGBoost and Extra Trees have also been adopted, particularly when paired with SHAP or other model-agnostic interpretability tools [24, 29]. Nevertheless, while ML offers strong baseline performance and interpretability, its reliance on paradigms such as decision trees may hinder scalability when dealing with very large feature sets.

In contrast, DL models are less common and typically reserved for tasks requiring high representational capacity or when multimodal data is involved. Most of these studies rely on CNNs adapted to microbiome structures or combined with dimensionality reduction techniques like PCA [31-34]. Notably, several works embed phylogenetic information directly into the network architecture, enabling structured feature extraction from metagenomic profiles [31, 34]. Despite achieving strong performance in specific cases, deep models remain limited by interpretability and risk of overfitting, especially in settings with limited data or class imbalance.

Interestingly, hybrid approaches that combine DL with simpler interpretable models (e.g., using CNNs for representation followed by RF for classification) begin to emerge as a way to balance expressiveness and transparency [34].

VI. Data types and Processing

Across the reviewed studies, the most prevalent source of microbiota data is the gut, obtained predominantly via stool samples. This is expected given its rich microbial diversity and established relevance in systemic health. For instance, in cancer and liver-related tasks [19, 23], fecal samples were used to build predictive models for treatment response and patient prognosis. A smaller group of studies used oral [26, 27] or skin-derived [33] microbiota, showing that alternative microbial niches are gaining relevance in disease detection but remain less explored.

In terms of profiling methods, 16S rRNA sequencing was the most common approach [21, 24, 25], largely due to its lower cost and availability. However, its taxonomic resolution typically stops at the genus level, which introduces limitations in clinical tasks that demand strain-level specificity. This is evident in [21], where the model's performance dropped substantially in external validation, partly due to the inability to resolve microbial features at finer taxonomic levels. In contrast, studies using whole-metagenome sequencing [19, 25] achieved better generalization and deeper insights, albeit with increased cost and complexity.

A few works integrated microbiome data with complementary modalities. [26] combined microbial abundance with metabolomics to diagnose oral conditions, while [33] incorporated nail images for skin-related classification tasks. Additionally, some studies enriched microbial profiles with clinical metadata [4, 25], improving interpretability and robustness. These combinations illustrate a trend toward multi-source integration to compensate for the limitations of microbiome data alone.

Finally, preprocessing practices varied widely across studies. Some used total-sum scaling or centered log-ratio transformations [21, 25], but details were often poorly reported, affecting reproducibility. Feature selection was inconsistently applied, ranging from simple abundance thresholds to complex statistical filtering [30]. This lack of standardization complicates cross-study comparisons and model generalization. Greater transparency in data handling remains a critical area for improvement in microbiota-based ML.

VII. Limitations and Open Challenges

VII-A. Generalization and external validation

A key limitation across the reviewed studies is the scarce use of external validation datasets. While most models report high internal performance (often AUCs above 0.85), only a minority evaluate their generalizability on independent cohorts. For instance, the multiclass model in [30] showed performance drops when tested on 12 external datasets, despite strong internal metrics. Similarly, [25] observed a decline from 0.873 AUC on training to 0.633 on an external cohort. These findings suggest overfitting to cohort-specific patterns, limiting translational utility. Geographic, demographic, and technical diversity in validation sets remains largely unaddressed.

VII-B. Data size and heterogeneity

Many studies are constrained by limited sample sizes and heterogeneous data sources. Works like [24] and [27] rely on fewer than 60 subjects, restricting statistical power and robustness. Even in larger datasets, variations in sequencing platforms, preprocessing methods, and taxonomic resolutions introduce noise that complicates model training and reproducibility. Studies such as [25] and [30] acknowledge batch effects and inter-study differences as sources of performance degradation, which hinder benchmarking and the development of unified pipelines.

VII-C. Interpretability and transparency

Although some studies incorporate explainability tools (e.g., SHAP in [24]), most lack systematic strategies for interpreting predictions or validating selected features. DL approaches, in particular, often prioritize performance over interpretability. For instance, CNN-based models like those in [32] and [34] achieve high AUCs but provide little insight into which microbial features drive predictions. Moreover, few studies release their code, feature selection procedures, or trained models, limiting transparency and reproducibility.

VII-D. Biological noise and batch effects

Biological variability, sequencing artifacts, and the absence of standardized procedures contribute to inconsistent findings. In multiclass classifiers such as [30], authors explicitly mention confounding effects and restricted phenotype diversity as challenges. Likewise, [25] notes heterogeneity in microbial communities and the influence of clinical covariates. Without harmonized preprocessing and annotation standards, microbiota-based AI models remain vulnerable to batch effects that compromise robustness.

VIII. Future Prospects

The challenges outlined above naturally point to the next steps required to strengthen the field. Several directions emerge as priorities for advancing the integration of AI in microbiota-based clinical research.

First, the standardization of data processing workflows is critical. Current pipelines for microbiota preprocessing, feature selection, and taxonomic resolution remain inconsistent, hampering reproducibility and comparability across studies. Establishing community-endorsed protocols would help ensure methodological consistency and facilitate the aggregation of results across cohorts and institutions. Specific practices such as adopting log-ratio transformations (e.g., CLR, ALR) to handle compositionality, and transitioning from OTU-based to ASV-based pipelines to improve taxonomic resolution and reduce ambiguity, could serve as concrete steps toward such standardization.

Second, external validation must become routine. Many models remain confined to single-center datasets with limited demographic or geographic diversity, which reduces generalizability. Incorporating multicenter cohorts with well-characterized metadata would allow robust cross-population evaluations and help identify sources of variability. Beyond improving performance assessment, such validation is essential to ensure that models trained on microbiota data can be reliably translated into diverse clinical settings.

Third, interpretability should be more systematically integrated. Although some works employ post hoc explanation methods, the field still lacks unified approaches to make predictions transparent and clinically actionable. Embedding interpretability frameworks into modeling pipelines is especially important in high-stakes applications such as early cancer detection, cardiovascular risk assessment, or psychiatric screening. Methods such as LIME, SHAP, or attention-based mechanisms could provide complementary insights and represent promising directions for improving trust and clinical adoption.

In addition, predictive modeling of therapeutic responses remains underexplored. Only a small number of studies address treatment stratification using microbiota data, and most are limited to small cohorts or binary outcomes. Expanding this line of work to include longitudinal designs, multiclass endpoints, and causal inference frameworks could provide actionable insights for personalized treatment planning. Incorporating time-resolved microbiota dynamics and integrating treatment metadata would further strengthen this line of research, enabling more precise identification of responders and non-responders in real-world clinical contexts.

Beyond classification, future research should also address causality. Most current models remain focused on correlational tasks such as disease classification or prognosis, but causal machine learning and counterfactual analysis could provide deeper insights into the mechanisms linking microbiota to clinical outcomes. These approaches would help move from association to causation, enabling the identification of drivers rather than markers, which is essential for actionable interventions.

In addition, research should expand beyond the gut to systematically include oral, skin, vaginal, and respiratory microbiota, which remain underrepresented despite their clinical relevance. Methodologically, new directions such as federated learning could facilitate cross-cohort studies without compromising data privacy, while graph-based models offer powerful ways to capture microbial interaction networks. Together with causal inference, these approaches represent promising avenues for broadening both the scope of microbiota-based AI and its methodological toolkit.

Finally, multimodal learning represents a promising but fragmented frontier. While some studies have combined microbiota data with clinical variables, metabolomics, or imaging, data fusion methods are still largely ad hoc and lack standardized evaluation frameworks. Future research should prioritize the development of unified architectures capable of learning coherent representations across modalities and systematically benchmarking their performance. Such approaches could better capture the interplay between microbiota and host factors, ultimately supporting more comprehensive models of disease risk, progression, and treatment response.

IX. CONCLUSIONS

This review examines how artificial intelligence techniques are being applied to microbiota data for clinical purposes. To this end, we conducted a structured analysis of selected peer-reviewed studies, focusing on their clinical objectives, types of microbiota-derived inputs, modeling strategies, and validation procedures. Works were categorized to identify methodological patterns and assess the scope of current approaches in this emerging intersection between microbiome research and computational methods.

Findings reveal that most studies focus on disease classification or prognosis, often favoring interpretable models such as decision trees and random forests. Deep learning remains less common but is employed in multimodal settings. Despite promising performance, external validation is infrequent, and dataset diversity is limited, reducing generalizability. Many studies lack standardized pipelines and report minimal reproducibility. Interpretability tools are rarely implemented, limiting clinical trust.

Likewise, microbiota data are mainly derived from gut samples, though oral and skin microbiomes are gaining attention. Integration with metabolomics or imaging remains experimental, and predictive modeling of therapeutic response is still incipient. These trends highlight the need for more rigorous validation, broader data inclusion, and increased focus on explainability to support clinical translation.

REFERENCES

[1] A. A. Singh and M. Okpeku, “Emerging methods of human microbiome analysis and its forensic applications: Review,” *Forensic Science International: Reports*, vol. 9, p. 100355, 2024.

[2] X. Peng, L. Cheng, Y. You, C. Tang, B. Ren, Y. Li, X. Xu, and X. Zhou, “Oral microbiota in human systematic diseases,” *International Journal of Oral Science*, vol. 14, no. 1, p. 1–11, Mar 2022.

[3] G. Caballero-Flores, J. M. Pickard, and G. Núñez, “Microbiota-mediated colonization resistance: mechanisms and regulation,” *Nature Reviews Microbiology*, vol. 21, no. 6, p. 1–14, Dec 2022.

[4] S. Reitmeier, S. Kiessling, T. Clavel, M. List, E. L. Almeida, T. S. Ghosh, K. Neuhaus, H. Grallert, J. Linseisen, T. Skurk, B. Brandl, T. A. Breuninger, M. Troll, W. Rathmann, B. Linkohr, H. Hauner, M. Laudes, A. Franke, C. I. Le Roy, J. T. Bell, T. Spector, J. Baumbach, P. W. O’Toole, A. Peters, and D. Haller, “Arrhythmic gut microbiome signatures predict risk of type 2 diabetes,” *Cell Host & Microbe*, vol. 28, no. 2, pp. 258–272.e6, 2020.

[5] M. Qin, Z. Huang, Y. Huang, X. Huang, C. Chen, Y. Wu, Z. Wang, F. He, B. Tang, C. Long, X. Mo, J. Liu, and W. Tang, “Association analysis of gut microbiota with ldl-c metabolism and microbial pathogenicity in colorectal cancer patients,” *Lipids in Health and Disease*, vol. 23, no. 1, Nov 2024.

[6] S. J. Blake, Y. Wolf, B. Boursi, and D. J. Lynn, “Role of the microbiota in response to and recovery from cancer therapy,” *Nature Reviews Immunology*, vol. 24, Nov 2023.

[7] C. C. Wong and J. Yu, “Gut microbiota in colorectal cancer development and therapy,” *Nature Reviews Clinical Oncology*, vol. 20, no. 7, p. 429–452, Jul 2023.

[8] S. Kodikara, S. Ellul, and K.-A. Lê Cao, “Statistical challenges in longitudinal microbiome data analysis,” *Briefings in Bioinformatics*, vol. 23, no. 4, p. bbac273, 07 2022.

[9] R. Bharti and D. G. Grimm, “Current challenges and best- practice protocols for microbiome analysis,” *Briefings in Bioinformatics*, vol. 22, no. 1, pp. 178–193, 12 2019.

[10] D. Wickramaratne, C. Wijesinghe, and A. Weerasinghe, “A deep learning approach to predict health status using microbiome profiling,” *22nd International Conference on Advances in ICT for Emerging Regions (ICTer)*, pp. 075–079, 2022.

[11] M. G. Ragab, S. J. Abdulkadir, A. Muneer, A. Alqushaibi, E. H. Sumiea, R. Qureshi, S. M. Al-Selwi, and H. Alhussian, “A comprehensive systematic review of yolo for medical object detection (2018 to 2023),” *IEEE Access*, vol. 12, pp. 57 815– 57 836, 2024.

[12] C. J. Haug and J. M. Drazen, “Artificial intelligence and machine learning in clinical medicine, 2023,” *New England Journal of Medicine*, vol. 388, no. 13, pp. 1201–1208, 2023.

[13] L. T. Ramos and A. D. Sappa, “Multispectral semantic segmentation for land cover classification: An overview,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 14 295–14 336, 2024.

[14] S. K. Baduge, S. Thilakarathna, J. S. Perera, M. Arashpour, P. Sharafi, B. Teodosio, A. Shringi, and P. Mendis, “Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications,” *Automation in Construction*, vol. 141, p. 104440, 2022.

[15] W. Quanyang, H. Yao, W. Sicong, Q. Linlin, Z. Zewei, H. Donghui, L. Hongjia, and Z. Shijun, “Artificial intelligence in lung cancer screening: Detection, classification, prediction, and prognosis,” *Cancer medicine*, vol. 13, no. 7, Apr 2024.

[16] L. T. Ramos and A. D. Sappa, “Leveraging u-net and selective feature extraction for land cover classification using remote sensing imagery,” *Scientific Reports*, vol. 15, no. 1, Jan 2025.

[17] Z. Al-Taie, D. Liu, J. B. Mitchem, C. Papageorgiou, J. T. Kaifi, W. C. Warren, and C.-R. Shyu, “Explainable artificial intelligence in high-throughput drug repositioning for subgroup stratifications with interventionable potential,” *Journal of Biomedical Informatics*, vol. 118, p. 103792, 2021.

[18] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo- Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, “The prisma 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, 3 2021.

[19] A. Gunjur, Y. Shao, T. Rozday, O. Klein, A. Mu, B. W. Haak, B. Markman, D. Kee, M. S. Carlino, C. Underhill, S. Frentzas, M. Michael, B. Gao, J. Palmer, J. Cebon, A. Behren, D. J. Adams, and T. D. Lawley, “A gut microbial signature for combination immune checkpoint blockade across cancer types,” *Nature Medicine*, vol. 30, p. 1–13, Mar 2024.

[20] W. Xu, T. Wang, N. Wang, H. Zhang, Y. Zha, L. Ji, Y. Chu, and K. Ning, “Artificial intelligence-enabled microbiome-based diagnosis models for a broad spectrum of cancer types,” *Briefings in Bioinformatics*, vol. 24, no. 3, p. bbad178, 05 2023.

[21] Y. Zheng, Z. Fang, Y. Xue, J. Zhang, J. Zhu, R. Gao, S. Yao, Y. Ye, S. Wang, C. Lin, S. Chen, H. Huang, L. Hu, G.-N. Jiang, H. Qin, P. Zhang, J. Chen, and H. Ji, “Specific gut microbiome signature predicts the early-stage lung cancer,” *Gut microbes*, vol. 11, no. 4, p. 1030–1042, Apr 2020.

[22] Y.-H. Zhou and P. Gallins, "A review and tutorial of machine learning methods for microbiome host trait prediction," *Frontiers in Genetics*, vol. 10, Jun 2019.

[23] H. Huang, Z. Ren, X. Gao, X. Hu, Y. Zhou, J. Jiang, H. Lu, S. Yin, J. Ji, L. Zhou, and S. Zheng, "Integrated analysis of microbiome and host transcriptome reveals correlations between gut microbiota and clinical outcomes in hbv-related hepatocellular carcinoma," *Genome Medicine*, vol. 12, no. 1, Nov 2020.

[24] C. W. Wong, S. E. Yost, J. S. Lee, J. D. Gillece, M. Folkerts, L. Reining, S. K. Highlander, Z. Eftekhari, J. Mortimer, and Y. Yuan, "Analysis of gut microbiome using explainable ma- chine learning predicts risk of diarrhea associated with tyrosine kinase inhibitor neratinib: A pilot study," *Frontiers in Oncology*, vol. 11, Mar 2021.

[25] D. Kang, J. Park, M. Yeo, S. B. Kang, J. M. Kim, J. S. Kim, and S. Kim, "Diagnosis of crohn's disease and ulcerative colitis using the microbiome," *BMC Microbiology*, vol. 23, no. 1, Nov 2023.

[26] L. Yan, J. Xu, F. Lou, Y. Dong, S. Lv, N. Kang, Z. Luo, Y. Liu, J. Pu, X. Zhong, P. Ji, P. Xie, and X. Jin, "Alterations of oral microbiome and metabolic signatures and their interaction in oral lichen planus," *Journal of Oral Microbiology*, vol. 16, no. 1, Oct 2024.

[27] Y. M. Hong, J. Lee, D. H. Cho, J. H. Jeon, J. Kang, M.-G. Kim, S. Lee, and J. K. Kim, "Predicting preterm birth using machine learning techniques in oral microbiome," *Scientific Reports*, vol. 13, no. 1, Nov 2023.

[28] S. Aryal, A. Alimadadi, I. Manandhar, B. Joe, and X. Cheng, "Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease," Hypertension, vol. 76, no. 5, Sep 2020.

[29] L. E. McCoubrey, S. Thomaidou, M. Elbadawi, S. Gais- ford, M. Orlu, and A. W. Basit, "Machine learning predicts drug metabolism and bioaccumulation by intestinal microbiota," *Pharmaceutics*, vol. 13, no. 12, p. 2001, Nov 2021.

[30] Q. Su, Q. Liu, R. I. Lau, J. Zhang, Z. Xu, Y. K. Yeoh, T. W. H. Leung, W. Tang, L. Zhang, J. Q. Y. Liang, Y. K. Yau, J. Zheng, C. Liu, M. Zhang, C. P. Cheung, J. Y. L. Ching, H. M. Tun, J. Yu, F. K. L. Chan, and S. C. Ng, "Faecal microbiome-based machine learning for multi-class disease diagnosis," *Nature Communications*, vol. 13, no. 1, p. 6818, Nov 2022.

[31] Y. Wang, T. Bhattacharya, Y. Jiang, X. Qin, Y. Wang, Y. Liu, A. J. Saykin, and L. Chen, "A novel deep learning method for predictive modeling of microbiome data," *Briefings in Bioinformatics*, vol. 22, no. 3, p. bbaa073, 05 2020.

[32] I.-g. Park, S. J. Yoon, S.-m. Won, K.-K. Oh, J. Y. Hyun, K. T. Suk, and U. Lee, "Gut microbiota-based machine-learning signature for the diagnosis of alcohol-associated and metabolic dysfunction-associated steatotic liver disease," *Scientific Reports*, vol. 14, no. 1, Jul 2024.

[33] T. Sun, X. Niu, Q. He, F. Chen, and R.-Q. Qi, "Artificial intelligence in microbiomes analysis: A review of applications in dermatology," *Frontiers in Microbiology*, vol. 14, Feb 2023.

[34] X. Chen, Z. Zhu, W. Zhang, Y. Wang, F. Wang, J. Yang, and K.-C. Wong, "Human disease prediction from microbiome data by multiple feature fusion and deep learning," *iScience*, vol. 25, no. 4, p. 104081, 2022.

[35] J. Borgman, K. Stark, J. Carson, and L. Hauser, "Deep learning encoding for rapid sequence identification on microbiome data," *Frontiers in Bioinformatics*, vol. 2, Jun 2022.