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

Rohith Janwadkar · Leo Thomas Ramos · Nidia Payahuala-Díaz

Elizabeth Diaz · Francklin Rivas-Echeverría · Edmundo Casas

Department of Computer Science and Engineering, The University of Texas at Arlington,Arlington, TX 76019, USA

Computer Vision Center, Universitat Autònoma De Barcelona, Barcelona, 08193, Spain

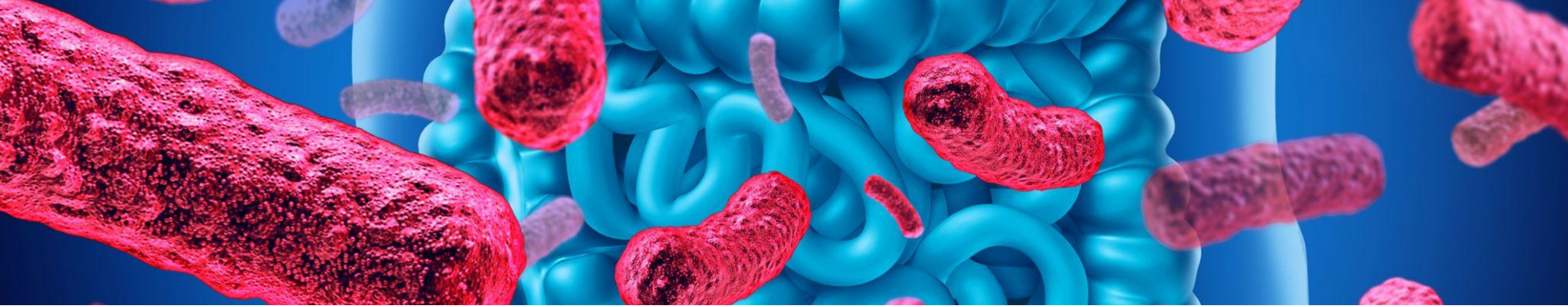
Kauel Inc., Menlo Park, Silicon Valley, CA 94025, USA

Servicio de Salud Osorno, Osorno, 825, Chile

IEEE SPMB | 2025

Agenda

- Background
- Why AI for Microbiota
- Methodology
- Overview of Selected Studies
- Machine Learning Approaches
- Deep Learning Approaches
- Clinical Applications
- Limitations
- Future Directions
- Conclusion

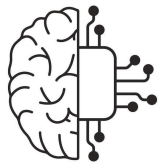


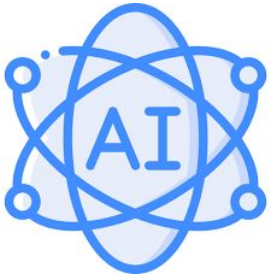
Background : Microbiota

1. Trillions of microorganisms across gut, oral, skin, urogenital sites
2. Key roles: Immunity, Metabolism, Pathogen resistance
3. Dysbiosis linked to major diseases:
   a. Colorectal cancer
   b. T2 Diabetes
   c. IBD
   d. Liver disease
4. High-dimensional, sparse, noisy data → analytical challenges

Why AI For Microbiota?

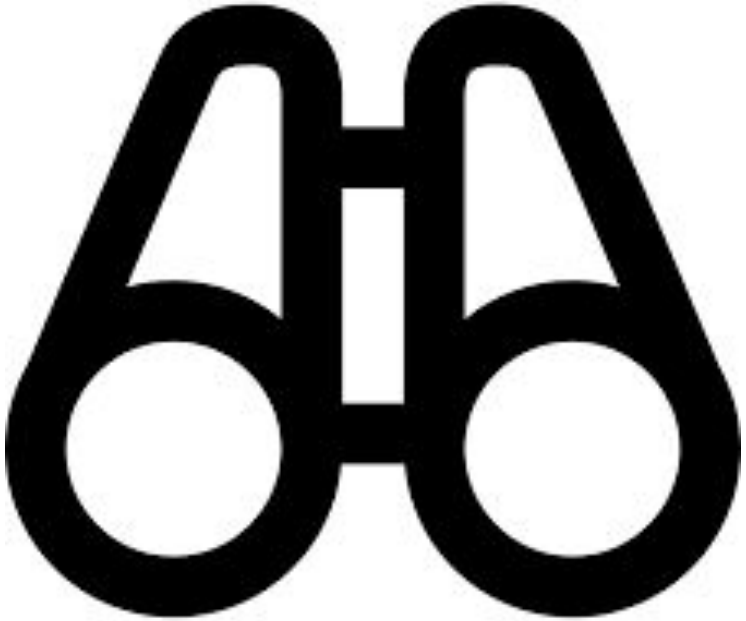
- Learns patterns in complex microbial ecosystems
- Handles high dimensionality and non-linearity
- Supports:
  - Disease prediction
  - Prognosis
  - Patient stratification
  - Biomarker discovery
- Potential pathway to precision medicine





Methodology(PRISMA INSPIRED)

- Databases: PubMed, IEEE Xplore, ScienceDirect, SpringerLink
- Keywords: microbiota · microbiome · ML · DL · clinical
- Inclusion:
- • Human data
- • Peer-reviewed
- • Empirical AI models
- Extracted: model type, data source, clinical goal, performance, limitations



Overview Of Selected Studies

- Majority use classical ML
- Smaller but growing set uses DL, especially CNNs
- Data types:
  - 16S rRNA
  - Whole-metagenome
  - Metabolomics
  - Imaging + microbiota
- Clinical focus: cancer, liver disease, IBD, oral disease, dermatology, drug interactions

Machine Learning Approaches

- Dominant: Random Forest (robust, interpretable)
- Others: SVM · Logistic Regression · XGBoost · Extra Trees
- Applications:
  - Cancer subtype prediction
  - Liver disease prognosis
  - CD vs UC classification
  - Preterm birth prediction
  - CVD screening
  - Drug–microbiota interaction modeling

Deep Learning Approaches

- Mostly CNN-based architectures
- Phylogeny-aware CNNs using taxonomic structure
- Used for:
  - Liver disease staging
  - Dermatology + microbiota multimodal tasks
  - High-resolution ASV identification
- Strength: feature representation
- Weakness: interpretability, data requirements

Clinical Applications of AI

- Diagnosis / Classification
- Colorectal cancer · Liver disease · IBD · Oral lichen planus · Dermatology
- Prognosis / Risk Prediction
- Survival · CVD · T2D (microbial rhythmicity)
- Therapeutic Support
- Immunotherapy response · Drug metabolism/bioaccumulation
- Multi-disease frameworks emerging

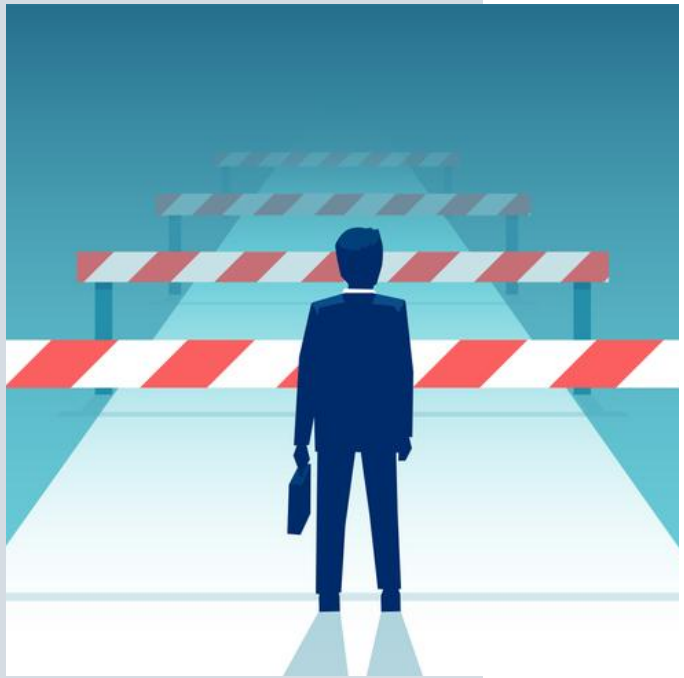
Data Types & Processing

- Mostly gut microbiota (stool samples)
- Growing use of oral & skin microbiota
- Profiling methods:
  - 16S rRNA (common, low resolution)
  - WGS (higher resolution, costly)
- Challenges:
  - Heterogeneous preprocessing
  - OTUs vs ASVs
  - Compositionality (CLR, ALR, TSS)

Key Limitations

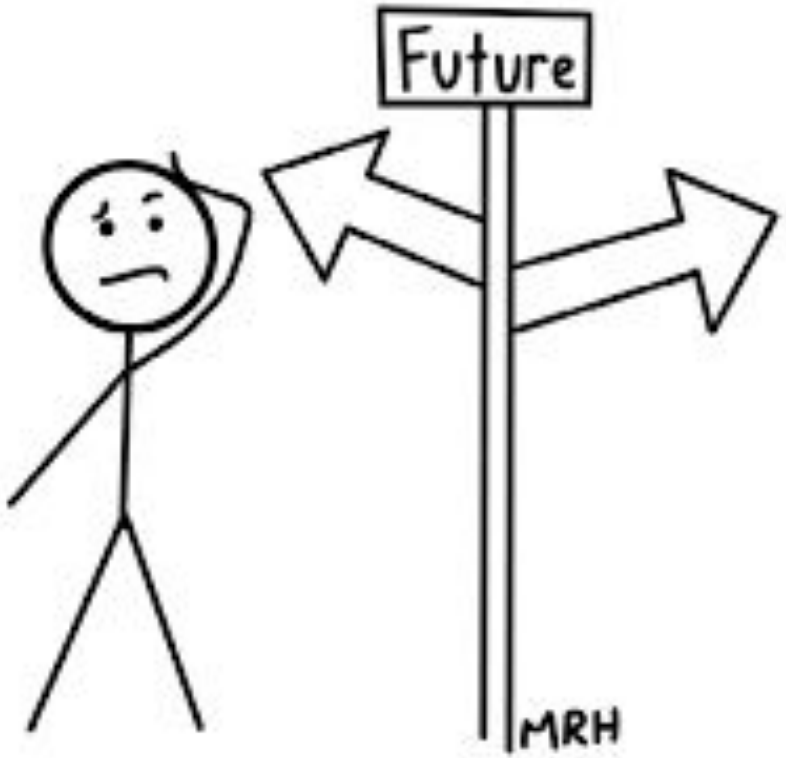
- Small, single-center datasets
- Weak external validation
- Batch effects & inter-study variability
- Inconsistent preprocessing pipelines
- Limited interpretability
- Minimal code / pipeline transparency

Challenges: What They Mean



- High AUC ≠ reliable model
- Models may learn site-specific artifacts instead of biology
- Poor generalization across regions, devices, populations
- Limited trust due to lack of explainability
- Biological noise + sequencing differences distort patterns

Future Directions



- Standardize microbiome preprocessing workflows
- Multi-center, diverse datasets for validation
- Integrate explainability (SHAP, LIME, attention)
- Expand to oral, skin, vaginal niches
- Multimodal learning (microbiota + metabolomics + imaging + metadata)
- Causal ML & temporal modeling
- Federated learning for privacy-preserving cross-site collaboration

Conclusions

- AI shows strong promise for microbiota-based clinical tasks
- ML dominates; DL excels in multimodal contexts
- Major gaps: generalization, standardization, interpretability
- Clinical translation requires rigorous validation and transparency
- Microbiota + AI → emerging pillar for personalized medicine
  -

Speaker: Rohith Janwadkar

IEEE SPMB | 2025