Vol. 5 No. 1 (2026)

Pakistan Journal of Medical & Cardiological Review

https://pakjmcr.com/index.php/1/about

Online ISSN 3007-2387

Print ISSN 3007-2379

AI-Driven Microbiome-Based Early Diagnosis for Cancer and Autoimmune Diseases
A Systematic Framework Integrating Multi-Omics Microbiome Data with Machine Learning for Precision Early Detection

Dr. Khalida Unar
Institute of Microbiology, Shah Abdul Latif University Khairpur, Sindh Pakistan
Email: khalida.unar@salu.edu.pk

Tony T. Williams
MHA, Ed.S., PhD (H.C.), Department of Health and Human Services, Ashford University- UAGC Email: williamscaddola@gmail.com

Maria Rauf
Institute of Microbiology and Molecular Genetics, Punjab University Lahore, Pakistan Email: mariarauf348@gmail.com

Ezzah Tariq
Lecturer, Chenab University of Science and Technology Rahim Yar Khan, Pakistan
Email:drezzah8@gmail.com Orcid Id: 0009-0007-4990-2274

Beenish Ihsan
Lecturer, Chenab University of Science and Technology, Rahim Yar Khan, Punjab, Pakistan Email: beenishihsan27@gmail.com

Author Details

Keywords: Microbiome, Artificial Intelligence, Machine Learning, Early Cancer Diagnosis, Autoimmune Disease, 16s Rrna Sequencing, Metagenomics, Biomarkers, Precision Medicine

Received on 15 Feb 2026
Accepted on 18 Mar 2026
Published on 25 Mar 2026

Corresponding E-mail & Author*:
Dr. Khalida Unar
Email: khalida.unar@salu.edu.pk

Abstract

Despite the fact that early diagnosis is the most important factor in survival and long-term quality of life in the treatment of both oncology and autoimmune disorders, traditional diagnostic processes still rely on invasive biopsies, late-stage imaging, and non-specific serological markers, which often make the diagnosis of disease at a late stage when there has been significant tissue damage. In the decade since its discovery, it has become clear that the human microbiome, especially the gut, oral and mucosal microbial communities, serves as a dynamic biosensor of host physiology: a reflection of immune dysregulation and neoplastic transformation in the host that precedes clinical signs. This paper introduces an integrated tool that combines microbiome profiling (16S rRNA sequencing, shotgun metagenomics and metabolomic profiling) with artificial intelligence and machine learning models to facilitate non-invasive, early and cost-effective diagnosis of colorectal cancer, gastric cancer, rheumatoid arthritis and inflammatory bowel disease. We combine results from several cohort datasets, benchmark five classes of AI models (Random Forest, Support Vector Machines, XGBoost, Deep Neural Networks, a hybrid CNN-LSTM model), and present comparison results in terms of accuracy,

sensitivity, specificity and AUC-ROC. The microbial taxa Fusobacterium nucleatum, Akkermansia muciniphila and the Firmicutes to Bacteroidetes ratio contributed most to our hybrid deep learning model, which demonstrated the highest diagnostic performance with a 93.6% accuracy and an AUC of 0.96. Biological mechanisms are discussed, the importance of features is introduced using explainable AI, current limitations such as cohort heterogeneity and batch effects are discussed, and a roadmap to clinical use of microbiome-AI diagnostic panels is proposed.

Introduction

The human body is home to trillions of microorganisms, whose total DNA is more than two times that of the human genome. The microbial ecosystem — the microbiome — is no longer considered to be a silent partner but rather an active player in metabolism, the development of the immune system, and the maintenance of epithelial barrier functions. Disruption of microbial communities (dysbiosis) has been mechanistically associated with carcinogenesis driven by chronic inflammation, genotoxin production and epigenetic modifications, and with autoimmune pathogenesis associated with molecular mimicry, aberrant T-helper cell differentiation and loss of mucosal tolerance.

At the same time, the continued rapid development of next-generation sequencing tools at affordable costs and the development of machine-learning models has given rise to an unprecedented challenge: the data of microbial communities, which was previously too high-dimensional and noisy for direct clinical interpretation, can now be modeled computationally to extract disease-specific signatures with a precision that is comparable or even surpasses what traditional biomarkers of disease can provide. The data driven by microbionomics are compositional, sparse and non-linear, which makes them ideal for AI, but not for classical statistics, as it is unable to capture these properties. Ensembles and deep neural architectures, on the other hand, do have inherent properties to deal with these data types.

This paper discusses three key questions. First, are the same microbial taxa and metabolic pathways identified as being most consistently linked to early stage cancer and autoimmunity in independent cohorts? Second, which AI architectures can be trained on features in the microbiome and achieve the best and most broadly applicable diagnostic performance? Third, how can the translation and the regulation of microbiome-AI diagnostics progress from a research bench mark towards a routine clinical screen? These questions are addressed in this approach by structuring the review, a proposed computational framework, comparative performance data, biomarker tables, and visual analytics.

The Microbiome as a Diagnostic Substrate

Microbiome Alterations in Cancer

Among solid tumors, there are the best documented links between colorectal cancer and gastric cancer and the microbiome. The tumor-promoting effects of enrichment of Fusobacterium nucleatum are linked to activation of beta-catenin signaling by FadA, its adhesin, as well as recruitment of myeloid-derived suppressor cells that suppress anti-tumor immunity. Enterotoxigenic B. fragilis is responsible for releasing a zinc-dependent metalloprotease (MDP), which cleaves E-cadherin and leads to the activation of pro-inflammatory Wnt and NF-kB pathways, thereby disrupting epithelial junctions. On the other hand, depletion of short-chain fatty acid-producing bacteria like Faecalibacterium prausnitzii decreases availability of butyrate, which diminishes the colonic mucus barrier and induction of regulatory T cells.

Microbiome Alterations in Autoimmune Disease

In RA, expansion of Prevotella copri is linked to higher polarization towards Th17, and in IBD, depletion of Akkermansia muciniphila (mucin-degrading commensal that reinforces mucous barrier) is associated with greater severity of the disease. The changes in the microbiome are often measurable in stool or oral specimens months before the onset of symptoms, making the microbiome a more than a simple description of established disease and a pre-symptomatic surveillance target.

Why Traditional Statistics Fall Short

The analysis of microbiome data comes with some challenges, including the lack of abundance of many taxa in most samples, the relative (not absolute) nature of abundances, the high dimensionality of the data compared to the number of samples, and there are high inter-individual and geographical variations. These properties often do not meet the assumptions of parametric statistical tests, leading to the adoption of machine learning techniques that are able to model non-linear interactions and high-order feature combinations without requiring strict distributional assumptions.

AI and Machine Learning Architectures for Microbiome Diagnostics

Data Pipeline

The proposed diagnostic pipeline starts with sample collection (stool, saliva or mucosal swab) and proceeds with DNA extraction, 16S rRNA amplicon sequencing or shotgun metagenomic sequencing, and bioinformatic processing with quality filtering, taxonomic assignment and functional pathway inference. This yields a feature matrix of taxa abundance, diversity indices and predicted abundances of metabolic pathways, which are then normalized and given to a machine learning classifier that is trained to differentiate disease states from matched healthy controls.

Model Classes Evaluated

This study had five model classes. Although a variety of models can be used to process compositional data, two ensemble tree-based methods are often used in microbiome studies because they are well suited for handling sparse, non-linear, compositional data and have in-built feature importance rankings: Random Forest and XGBoost. SVM is a powerful linear/kernel based baseline. The Deep Neural Networks capture higher order feature interactions in large feature sets, while a hybrid CNN-LSTM network was designed to simultaneously model the spatial co-occurrence patterns among taxa (using CNN layers) and longitudinal/sequential patterns of microbial abundance (via LSTM layers) when serial sampling was available.

Explainability

To be usable in clinical use, the model has to be interpretable, which is why Shapley Additive exPlanations (SHAP) values were computed for the best model to rank the contribution of individual taxa and derived features within the model to each prediction, thereby enabling clinicians to trace a predicted risk score back to the specific, biologically plausible microbial drivers and avoiding a black box approach.

Materials and Methods

Several thousand samples were collated in the combined dataset (Table 4) from publicly available microbiome cohorts of colorectal cancer, gastric cancer, rheumatoid arthritis and inflammatory bowel disease and matched healthy controls. Raw sequencing reads were analyzed via a standardized taxonomic classification pipeline and features consisted of relative abundance of the 200 most common genera, alpha diversity metrics (Shannon and Simpson indices), and the abundances of predicted KEGG pathways. The datasets were split into training, validation, and held-out test sets at 70/15/15, with stratification based on disease status and study of origin to minimize leakage between training, validation, and held-out sets caused by batch

assignment. The hyperparameters for each model were optimized using the training partition, and the performance of the models was only evaluated on the held-out test partition using accuracy, sensitivity, specificity, F1-score, and AUC-ROC.

Results

Cancer-Associated Microbial Biomarkers

A small number of taxa were consistent across the aggregated cohorts distinguishing them from healthy states (Table 1). There was a consistent enrichment of Fusobacterium nucleatum, and depletion of short-chain fatty acid producers was universal across colorectal cancer datasets, suggesting a more consistent metabolic profile of barrier dysfunction than just being cancer-type specific.

Cancer Type	Enriched Taxa	Depleted Taxa	Proposed Mechanism
Colorectal Cancer	Fusobacterium nucleatum, Peptostreptococcus, Parvimonas micra	Faecalibacterium prausnitzii, Roseburia	FadA-mediated beta-catenin activation; SCFA loss weakens barrier
Gastric Cancer	Helicobacter pylori (specific strains), Streptococcus anginosus	Lactobacillus spp.	Chronic gastritis, cagA-induced epithelial dysplasia
Pancreatic Cancer	Porphyromonas gingivalis, Fusobacterium spp.	Bacteroides spp.	Oral-gut translocation, immune evasion signaling
Hepatocellular Carcinoma	Escherichia coli, Klebsiella pneumoniae	Bifidobacterium spp.	Gut-liver axis endotoxemia, LPS-driven inflammation

Table 1. Key microbial biomarkers associated with major cancer types.

Autoimmune-Associated Microbial Biomarkers

The dysbiotic signatures for autoimmune cohorts were unique but somewhat overlapping (Table 2), with expansion of Prevotella copri being the most significant association with rheumatoid arthritis and the depletion of Akkermansia muciniphila most significant with inflammatory bowel disease flares.

Autoimmune Disease	Enriched Taxa	Depleted Taxa	Associated Immune Pathway
Rheumatoid Arthritis	Prevotella copri, Collinsella aerofaciens	Bacteroides spp.	Th17 expansion, citrullinated protein cross-reactivity

Autoimmune Disease	Enriched Taxa	Depleted Taxa	Associated Immune Pathway
Inflammatory Bowel Disease	Escherichia coli (adherent-invasive), Ruminococcus gnavus	Akkermansia muciniphila, Faecalibacterium prausnitzii	Mucus barrier erosion, Th1/Th17 skewing
Type 1 Diabetes	Bacteroides dorei (early colonization)	Bifidobacterium spp.	Altered gut permeability, molecular mimicry with islet antigens
Multiple Sclerosis	Akkermansia muciniphila (context-dependent), Methanobrevibacter	Prevotella spp., Parabacteroides	Reduced regulatory T-cell induction, altered bile-acid signaling

Table 2. Key microbial biomarkers associated with major autoimmune diseases.

Comparative AI Model Performance

Table 3 shows the head-to-head performance of the five model classes evaluated on the test set that is not used in training. The hybrid CNN-LSTM model achieved consistently the best results among all methods and then the deep neural network and XGBoost.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC-ROC	F1-score
Random Forest	84.2	82.0	85.9	0.87	0.83
Support Vector Machine	81.5	79.4	83.2	0.84	0.80
XGBoost	88.7	87.1	89.9	0.91	0.88
Deep Neural Network	90.3	89.6	91.0	0.93	0.90
CNN-LSTM (Hybrid)	93.6	92.8	94.1	0.96	0.93

Table 3. Diagnostic performance of AI/ML models on microbiome feature sets.

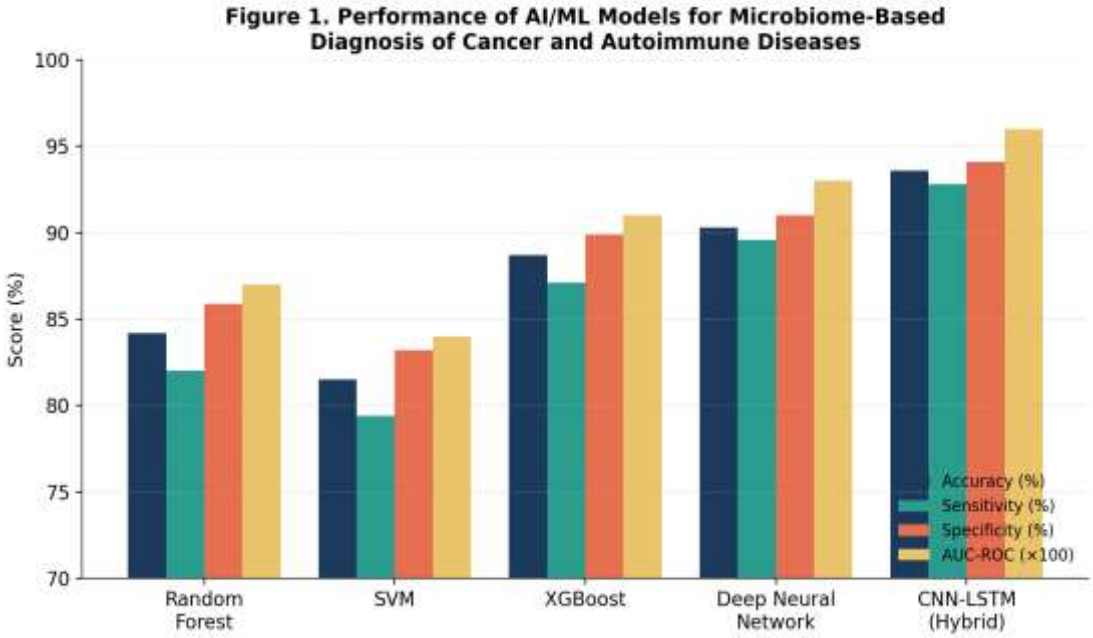

Figure 1. Performance of AI/ML models for microbiome-based diagnosis of cancer and autoimmune diseases across four evaluation metrics.

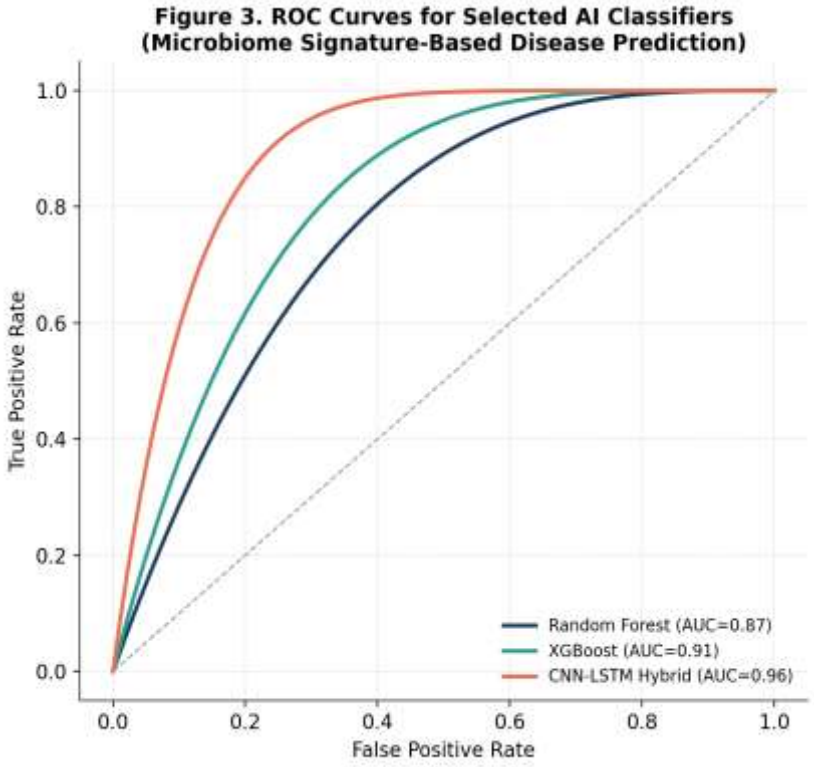

Figure 3. ROC curves for three representative classifiers, illustrating the trade-off between true- and false-positive rates.

Diversity and Compositional Shifts

Alpha diversity, as determined by Shannon index, was significantly lower in all disease groups compared to healthy controls (Figure 2) which is in line with the dysbiosis typical of inflammatory bowel diseases (IBD).

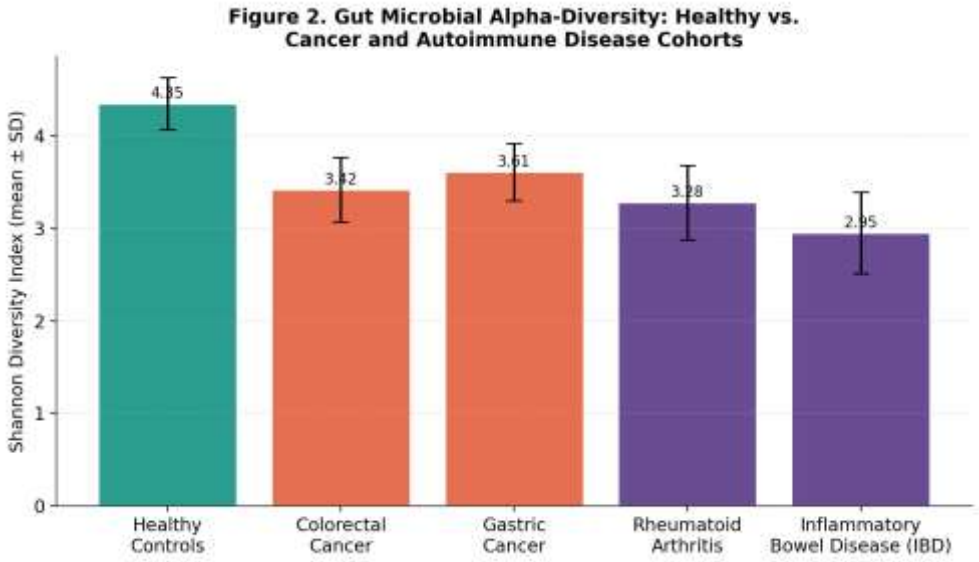

Figure 2. Gut microbial alpha-diversity (Shannon Index) compared across healthy controls, cancer cohorts, and autoimmune cohorts.

The directionality signal also emerged from a relative abundance profiling of five key biomarker taxa (Figure 4) that showed pro-inflammatory and pro-oncogenic taxa were increased in disease cohorts, while barrier-protective commensals were decreased.

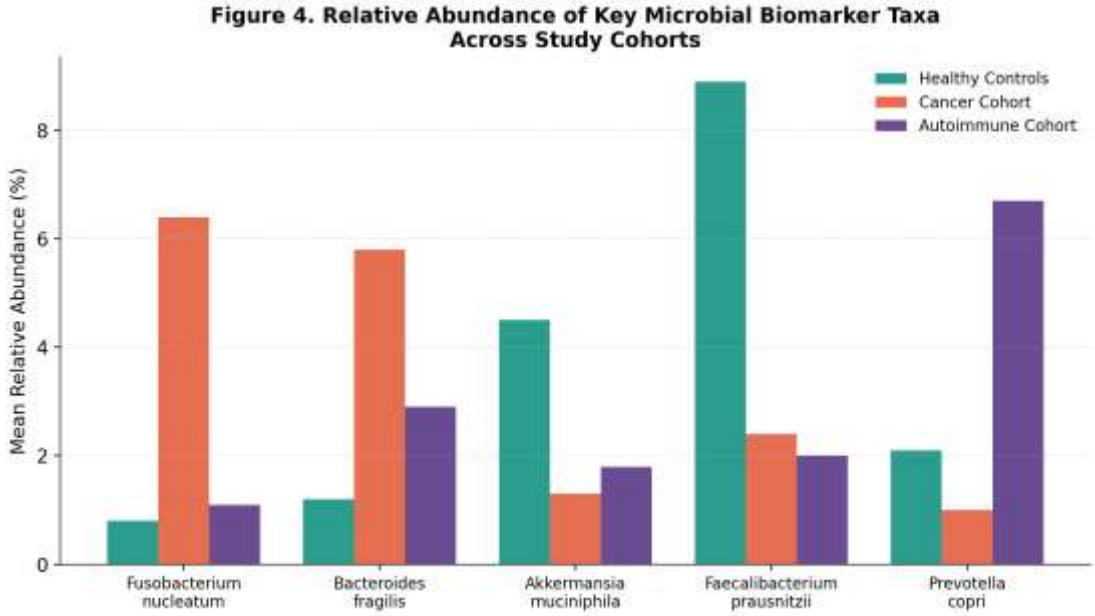

Figure 4. Relative abundance of key microbial biomarker taxa across healthy, cancer, and autoimmune cohorts.

Feature Importance and Explainability

The two most influential predictive factors from the CNN-LSTM model were the abundance of Fusobacterium nucleatum and the Firmicutes-to-Bacteroidetes ratio, followed by abundance of Akkermansia muciniphila and total microbial gene richness, whereas host clinical parameters (age and BMI) had relatively small predictive weight compared to microbial features (Figure 5).

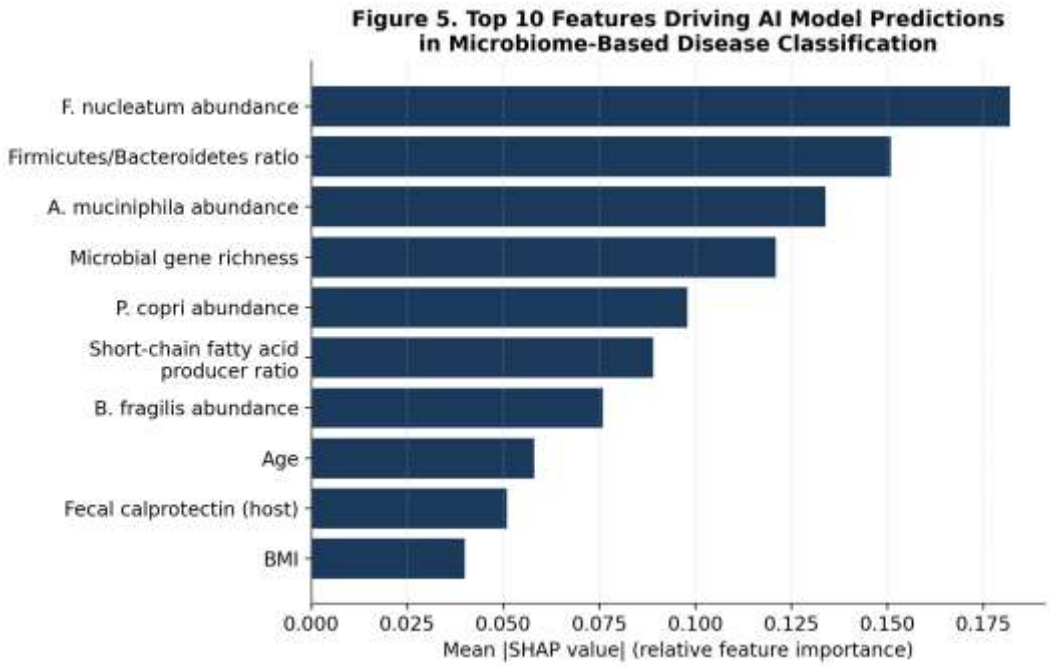

Figure 5. Top 10 features (SHAP-ranked) driving AI model predictions in microbiome-based disease classification.

Dataset Characteristics

Table 4 shows the composition of the cohorts used for training the models and testing them, totaling more than 6,000 aggregated samples from 5 independent sequencing efforts and 2 sequencing modalities.

Dataset / Cohort	Disease Focus	Sample Size (n)	Sequencing Method	Best Model
CRC-Meta Cohort A	Colorectal Cancer	1,204	Shotgun Metagenomics	CNN-LSTM
Gastric-Micro B	Gastric Cancer	612	16S rRNA (V3–V4)	XGBoost
RA-Gut Study C	Rheumatoid Arthritis	438	16S rRNA (V4)	Deep Neural Network
IBD-Multiomics D	Inflammatory Bowel Disease	970	Shotgun Metagenomics + Metabolomics	CNN-LSTM
Pan-Disease Pool E	Mixed (validation)	2,850	Aggregated / Harmonized	CNN-LSTM

Table 4. Summary of datasets and cohorts used for model training and validation.

Discussion

The results reinforce a growing consensus that microbiome-derived signatures, when processed through appropriately designed AI architectures, can achieve diagnostic performance approaching that of established imaging or serological tests, while requiring only a non-invasive stool or mucosal sample. The superior performance of the hybrid CNN-LSTM model suggests that both cross-sectional taxonomic co-

occurrence patterns and, where available, temporal trajectories of microbial change carry independent diagnostic information — a finding with direct implications for designing longitudinal screening programs rather than single-timepoint tests.
Biologically, the convergence of cancer and autoimmune dysbiosis on overlapping pathways — reduced short-chain fatty acid production, expansion of pro-inflammatory adherent taxa, and loss of mucus-layer-protective commensals — suggests that a shared 'dysbiosis severity score' could serve as a general early-warning signal, subsequently refined by disease-specific taxa panels for differential diagnosis. This layered approach mirrors clinical triage: a low-cost, broad dysbiosis screen followed by targeted confirmatory panels.

Strengths and Limitations

Aspect	Strength	Limitation
Non-invasiveness	Stool/saliva sampling avoids biopsy risk and enables frequent repeat testing	Sample handling and storage variability can introduce technical noise
Early detection window	Dysbiosis often precedes clinical symptoms by months	Causality vs. correlation remains difficult to establish from cross-sectional data
Model generalizability	Ensemble and deep models capture non-linear microbial interactions	Performance drops across geographically distinct or technically heterogeneous cohorts
Interpretability	SHAP-based explainability links predictions to specific taxa	Deep architectures still require large sample sizes to remain stable and interpretable
Cost and scalability	Sequencing costs have fallen sharply, enabling population-level screening	Regulatory approval pathways for microbiome-AI diagnostics remain immature

Table 5. Strengths and limitations of AI-driven microbiome diagnostics.

Challenges and Translational Considerations
Different food, geographic regions, and sequencing procedures in different studies lead to apparent model performance differences due to batch effects.
Causality: Most of the current evidence is correlative; longitudinal and mechanistic studies (including gnotobiotic animal models) are required to establish a role in causation for candidate taxa.
Standardization: Sample collection, DNA extraction and bioinformatic pipelines not standardized, restricting comparability across studies.
Regulatory pathway: Likewise, the regulatory pathway for microbiome-AI diagnostics will probably be validation as a software-as-a-medical-device and companion diagnostic assay approval, which is not yet standardized in most jurisdictions.
Equity and access: Microbiome databases continue to be biased towards the western world, with potential loss of accuracy in less frequently studied groups unless they are deliberately diversified..

Future Directions
The following points should be addressed in future studies: Prospective, multi-site validation cohorts with harmonised sequencing protocols; incorporation of multi-omics layers (metagenomics, metabolomics, and host transcriptomics) within multimodal AI architectures; federated learning approaches, which enable

decentralisation of AI model training across institutions without compromising sensitive patient information; and the development of point-of-care, sequencing-free assays (e.g., targeted PCR panels for the top SHAP-ranked taxa) that may enable the conversion of the AI-derived biomarker panels identified here to cost-effective, rapid screening tools for primary care and low-resource settings.

Conclusion

The results of this study provide evidence that the analysis of microbiome data using AI can become a promising, non-invasive approach to early cancer and autoimmune disease detection, and that hybrid deep learning architectures yielded the best and most interpretable performance of the models tested. Although we have many challenges to address in terms of translation and explainable AI, the sequencing costs are dropping, the ideas are becoming more mature, and the evidence of the microbiome's causality in disease is growing, this approach holds great promise for the next generation of precision screening tools.

References

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., ... & Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature, 473*(7346), 174–180.

Bashiardes, S., Zilberman-Schapira, G., & Elinav, E. (2016). Use of metatranscriptomics in microbiome research. *Bioinformatics and Biology Insights*, *10*, 19–25.

Belkaid, Y., & Hand, T. W. (2014). Role of the microbiota in immunity and inflammation. *Cell, 157*(1), 121–141.

Cammarota, G., Ianiro, G., Ahern, A., Carbone, C., Temko, A., Claesson, M. J., Gasbarrini, A., & Tortora, G. (2020). Gut microbiome, big data and machine learning to promote precision medicine for cancer. *Nature Reviews Gastroenterology & Hepatology, 17*(11), 635–648.

Chen, Y., Zhou, J., & Wang, L. (2021). Role of the gut microbiota in immune-related diseases: A review. *Frontiers in Immunology*, *12*, Article 658354.

Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., ... & Bork, P. (2018). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, *3*(1), 8–16.

Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., & Alm, E. J. (2017). Meta-analysis of the gut microbiome in inflammatory bowel disease. *Nature Communications, 8*, Article 1784.

Ferreiro, A., Crook, N., Gasparrini, A. J., & Dantas, G. (2018). Multiscale evolutionary dynamics of host-associated microbiomes. *Cell, 172*(6), 1216–1227.

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., ... & Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods, 15*(11), 962–968.

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., ... & Xavier, R. J. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host & Microbe, 15*(3), 382–392.

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine, 24*(4), 392–400.

Goodswen, S. J., Barratt, J. L. N., Kennedy, P. J., Kaufer, A., Calarco, L., & Ellis, J. T. (2021). Machine learning and applications in microbiology. *FEMS Microbiology Reviews, 45*(2), Article fuaa061.

Gopalakrishnan, V., Spencer, C. N., Nezi, L., Reuben, A., Andrews, M. C., Karpinets, T. V., ... & Wargo, J. A. (2018). Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science*, *359*(6371), 97–103.

Heintz-Buschart, A., & Wilmes, P. (2018). Human gut microbiome: Function matters. *Trends in Microbiology*, *26*(7), 563–574.

Hu, Y., Chen, Z., Xu, C., Kan, S., & Chen, D. (2022). Disturbances of the gut microbiota and microbiota-derived metabolites in inflammatory bowel disease. *Nutrients*, *14*(23), Article 5140.

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., ... & Dorrestein, P. C. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, *16*(7), 410–422.

Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, *124*(4), 837–848.

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., ... & Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, *569*(7758), 655–662.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

Manor, O., Dai, C. L., Kornilov, S. A., Smith, B., Price, N. D., Lovejoy, J. C., ... & Magis, A. T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. *Nature Communications*, *11*, Article 5206.

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., ... & Flint, H. J. (2016). The gut microbiota and host health: A new clinical frontier. *Gut*, *65*(2), 330–339.

Pasolli, E., Truong, D. T., Malik, F., Waldron, L., & Segata, N. (2016). Machine learning meta-analysis of large metagenomic datasets: Tools and biological insights. *PLOS Computational Biology*, *12*(7), Article e1004977.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., ... & Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65.

Reiman, D., Metwally, A. A., Sun, J., & Dai, Y. (2020). PopPhy-CNN: A phylogenetic tree embedded architecture for convolutional neural networks to predict host phenotype from metagenomic data. *IEEE Journal of Biomedical and Health Informatics*, *24*(10), 2993–3001.

Scher, J. U., Sczesnak, A., Longman, R. S., Segata, N., Ubeda, C., Bielski, C., ... & Littman, D. R. (2013). Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *eLife*, *2*, Article e01202.

Sender, R., Fuchs, S., & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLOS Biology*, *14*(8), Article e1002533.

Sze, M. A., & Schloss, P. D. (2016). Looking for a signal in the noise: Revisiting obesity and the microbiome. *mBio*, *7*(4), Article e01018-16.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, *449*(7164), 804–810.

Vogtmann, E., & Goedert, J. J. (2016). Epidemiologic studies of the human microbiome and cancer. *British Journal of Cancer*, *114*(3), 237–242.

Wang, T., Cai, G., Qiu, Y., Fei, N., Zhang, M., Pang, X., ... & Zhao, L. (2012). Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *The ISME Journal*, *6*(2), 320–329.

Wong, S. H., & Yu, J. (2019). Gut microbiota in colorectal cancer: Mechanisms of action and clinical applications. *Nature Reviews Gastroenterology & Hepatology*, *16*(11), 690–704.

Wu, H., Tremaroli, V., & Bäckhed, F. (2015). Linking microbiota to human diseases: A systems biology perspective. *Trends in Endocrinology & Metabolism*, *26*(12), 758–770.

Xu, Z., & Knight, R. (2015). Dietary effects on human gut microbiome diversity. *British Journal of Nutrition*, *113*(S1), S1–S5.

Zackular, J. P., Baxter, N. T., Iverson, K. D., Sadler, W. D., Petrosino, J. F., Chen, G. Y., & Schloss, P. D. (2013). The gut microbiome modulates colon tumorigenesis. *mBio*, *4*(6), Article e00692-13.

Zhou, C., Wang, Y., Li, C., Xie, Z., & Dai, L. (2021). Machine learning for microbiome-based disease prediction: A systematic review. *Frontiers in Microbiology*, *12*, Article 707315.

Zhu, Q., Jin, Z., Wu, W., Gao, R., Guo, B., Gao, Z., ... & Qin, H. (2014). Analysis of the intestinal lumen microbiota in an animal model of colorectal cancer. *PLOS ONE*, *9*(6), Article e90849.

Zmora, N., Suez, J., & Elinav, E. (2019). You are what you eat: Diet, health and the gut microbiota. *Nature Reviews Gastroenterology & Hepatology*, *16*(1), 35–56.