Machine Learning provides preliminary snapshots of disease-specific variations in gut microbial aging-trajectories

by

Seerapthynath T R

Roll No. MT23236

Under the Supervision of

Dr. Tarini Shankar Ghosh

Assistant Professor,

Department of Computational Biology

Submitted

in partial fulfillment of the requirements for the degree of

Master of Technology

to

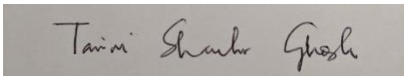
Indraprastha Institute of Information Technology Delhi

August, 2025

CERTIFICATE

This is to certify that the thesis titled "**Machine Learning provides preliminary snapshots of disease-specific variations in gut microbial aging-trajectories**" being submitted by **Seerapthynath T R** to the Indraprastha Institute of Information Technology Delhi, for the award of the Master of Technology, is an original research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.



August, 2025

Dr. Tarini Shankar Ghosh

Department of Computational Biology

Indraprastha Institute of Information Technology Delhi

New Delhi 110 020

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my advisor, Dr. Tarini Shankar Ghosh, for his invaluable guidance, mentorship, and support throughout this research endeavor. His expertise in microbiome science, insightful discussions, and constant encouragement have had a profound impact on shaping my work. I feel highly fortunate to have derived benefit from his scientific expertise, sharp thinking, and enthusiasm for research excellence.

I would like to extend my sincere thanks to my laboratory colleagues, Shivangi Verma and Sourav Goswami, for their unwavering support, constructive criticism, and collaborative efforts. Their involvement significantly enriched this work, whether in idea generation, technical challenges, sharing laboratory work, or moral support in times of adversity. Their combined effort has been an extremely crucial part of this work.

I wish to extend my heartfelt gratitude to all the faculty members, staff, and colleagues of the Department of Computational Biology at IIIT Delhi for providing such rich intellectual environments. Inputs, discussions, and motivation gained through lab sessions, departmental seminars, and informal talks contributed significantly to enriching the research. I further extend my gratitude to the technical support staff and computational resources at IIIT Delhi, as processing of large-scale data required for this research would not have been possible without them.

Personally, I would like to offer my warmest gratitude to my family. I am extending my sincerest gratitude to my parents, whose love without condition, unfailing faith, and countless sacrifices have been the foundation of my life; for that, I am forever thankful. Their values of determination and honesty have shaped my life at every turn. I am thankful to friends, too, who have provided emotional support, encouragement, and humor, guiding me through the ordeal that comes with my MTech journey.

Date: August 2025

Place: New Delhi

ABSTRACT

The human gut microbiome plays a critical role in metabolism, immunity, and disease resistance, with its composition shifting notably with age. These changes have positioned the microbiome as a potential “biological clock” capable of predicting host age. While numerous studies highlight strong associations between microbiome profiles and chronological age, most existing models are developed on healthy populations and rarely address how diseases alter microbial aging trajectories.

Here, we compiled 19,342 gut microbiome samples from 67 study cohorts in 31 nations on both 16S and shotgun metagenomic sequencing. With Random Forest regression models trained on control samples only, we made robust age-microbiome correlations and computed interpretable species-level SHAP values to define aging signals. Test on disease cohorts showed systematic departures from healthy aging patterns, where numerous diseases started accelerated, decelerated, or disrupted microbial aging patterns. In 16 diseases, classification analyses also showed that aging-related features continue to distinguish diseased patients from controls, albeit with both cross-cohort consistencies and disease-specific variations.

Our observations represent among the first large-scale demonstrations that microbiome-based indices of aging are replicable across populations but severely disrupted in disease. This research provides a foundation for the creation of disease-sensitive microbiome aging clocks and the discovery of microbial species responsible for healthy versus pathological aging trajectories with future biological interpretation and potential therapeutic implication.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

INTRODUCTION:

Human gut harbors a complex and dynamic community of microorganisms, often called the gut microbiome, that participate in numerous physiological functions, metabolic processes, and immune reactions. This microbial community is complex, and its structure and functional capability alter both during a person's lifespan from the dynamic variability characteristic of infancy to those alterations that occur during old age. Perhaps the most fascinating aspect of this interaction is the highly specific way in which the microbiome evolves over time, actually acting as a biological clock that informs a person's physiological status.

Recent studies have conclusively made the gut microbiome a key aspect of aging. Experiments reveal that the architecture of microbial communities has a direct effect on declining health and the onset of numerous age-related, non-communicable diseases (Ghosh et al., 2022). Notably, aging is not a straightforward, linear process but is a multifaceted phenomenon with periodic molecular shifts at several phases of life, each of which affects biological functions and makes one increasingly susceptible to disease (Shen et al., 2024).Although the adult microbiome is relatively constant compared to the dynamic conditions observed during early and late life stages (Odamaki et al., 2016), the decades between 40 and 80 are of special significance. It is during this time that the incidence of chronic disease increases drastically and emerging evidence increasingly suggests that the composition of the gut microbiome during these decades significantly affects long-term health consequence (Simbirtseva & O'Toole, 2025)

This has led to "microbiome aging clocks," computer algorithms that estimate a human's chronological age from the composition of their gut microbiota(Chen et al., 2021; Galkin et al., 2020; Wilmanski et al., 2022) .Promising, these models face a significant hurdle: they have been trained and validated largely on healthy cohorts. As a result, we have little information about the effect of disease on microbial aging pathways (Ratiner et al., 2022). To our knowledge, no large-scale, systematic study of microbial aging patterns across disease and across all global populations exists. These questions such as the extent of reproducibility of aging-related microbial signals and the specific mechanisms by which different states of disease perturb these signals remain largely unanswered.

This dissertation aims to fill this gap by investigating the interaction between microbial aging and disease at a scale that has never before been attempted. Ultimately, the goal is to develop a new framework of analysis explaining how disease disrupts the microbiome's natural process of aging. To this end, we constructed and normalized one of the largest global data sets in existence: more than 19,000 gut microbiome samples gathered from 31 countries, healthy patients and patients with a range of health conditions.

Using the dataset above, we applied a machine learning approach specifically, a Random Forest regression model not only to predict age, but to reveal a more interpretable and informative signature of microbial aging. By applying explainable AI techniques, i.e., SHAP (SHapley Additive exPlanations) values, we quantified the contribution of single microbial taxa towards aging predictions (Lundberg & Lee, 2017). The signatures obtained by SHAP provided a useful tool to describe how heterogeneous diseases such as metabolic and cardiovascular diseases, and cancer and infectious diseases are affecting the microbial clock. The report begins with an introduction to the process of developing the global repository and feature selection. This is succeeded by detailing the process of developing and validating the machine learning framework

to be used in aging predictions. The key findings are then summarized to show that disease states consistently deviate from normal microbial aging paths in disease-specific patterns. The report concludes by discussing the broader implications of the findings and outlining future directions for the development of disease-sensing biological clocks and developing possible therapeutic strategies.

CHAPTER 2

BACKGROUND AND OBJECTIVES:

2.1 The Gut Microbiome: A Dynamic Ecosystem Linked with Host Health

The human gastrointestinal tract harbors trillions of microorganisms, such as archaea, viruses, fungi, and bacteria, which together are known as the gut microbiome. This dynamic, intricate community is not just a passive player but an active participant in host physiology, contributing to important functions such as the degradative breakdown of complex carbohydrates, the synthesis of essential vitamins, the maturation and regulation of the immune system, and protection against pathogenic invaders (Valdes et al., 2018). The assembly of this microbial community begins at birth and undergoes explosive changes during infancy and early childhood before stabilizing towards a relatively stable, adult-like composition. But this established microbiome is not invariant; it is repeatedly altered by a host of factors such as diet, lifestyle, pharmacologic therapy, and, importantly, the aging process.

2.2 Microbial Markers of Aging: The Gastrointestinal Tract as a Biological Clock

One of the most intriguing discoveries in the field of microbiome studies is the predictable and regular manner in which the gut microbiota changes over a person's lifetime. This discovery has led to the idea of the microbiome as a "biological clock," an objective measure of a person's physiological age, possibly more revealing than their chronological age. Basic science has shown that the microbiota of elderly people has a completely different constitution than that of young people, typically evident by a loss of significant beneficial bacteria (e.g., Bifidobacterium), a decrease in total diversity, and an increase in opportunistic pathogens or "pro-inflammatory" species (Odamaki et al., 2016).

The aging shift is not a correlation. There is considerable evidence that has strongly established that the gut microbiome is a powerful modulator and a causative agent in the decline in health and the onset of many of the non-communicable diseases of aging (Ghosh, Shanahan, & O'Toole, 2022). This process is not a straight, linear decline. Recent studies by using multi-omics profiling have shown that human aging is a nonlinear, complex process with enormous waves of molecular changes at the fourth and sixth decades of life, which redefine biological functions and predispose to age-related diseases like cardiovascular disease and diabetes (Shen et al., 2024). The course of one's health during the latter years is increasingly known to be tightly connected with the composition and function of their gut microbiome at these critical decades (Simbirtseva & O'Toole, 2024).

2.3 Machine Learning and the Rise of Microbiome Aging Clocks:

The finding of consistent, age-related microbial patterns has enabled the development of "microbiome aging clocks." These are mathematical models, traditionally developed using machine learning methods, that can predict the chronological age of an individual from the gut microbiome composition. The first models were very accurate, with some deep learning models predicting age to within less than six years on average (Galkin et al., 2020). The models have since been improved by follow-up studies with multi-view learning approaches that combine taxonomic and functional profiles to improve predictability and to discover potential microbial biomarkers for aging (Chen et al., 2021).

These clocks are more than a subject of scientific curiosity; they represent a useful tool for the quantification of biological aging. The age difference between microbiome-predicted age and chronological age, also known as "gut-microbiome age" or "microbial age discrepancy," has been proposed as a novel biomarker for health evaluation. A "younger" microbial age, for instance, is associated with better aging trajectories, and an "older" microbial age with increased frailty and increased risk of death (Wilmanski, Gibbons, & Price, 2022).

2.4 The Critical Gap: The Effect of Disease on Microbial Senescence

While promising, existing microbiome aging clocks are afflicted by a fundamental limitation that is the underlying motivation of this thesis: they have been primarily designed and validated based on samples from healthy populations. This "healthy bias" means we have models that can describe the normal aging process, but there is a large knowledge gap about how the microbiome aging trajectories are disrupted upon the onset of disease (Ratiner et al., 2022). Existing research does not provide a comprehensive, systematic analysis of microbial patterns of aging across many diseases and within a variety of worldwide populations. It is not clear whether specific diseases trigger a common expression of "disturbed aging" or whether each disease causes a specific, condition-specified divergence from the normal aging trajectory. The consistency of aging-related microbial markers and the exact manner in which they are altered by various pathological conditions have yet to be comprehensively quantified. This dissertation bridges this vast knowledge gap by developing an innovative analytical approach designed to produce the first preliminary findings on such disease-specific divergences of gut microbial aging within a worldwide population.

2.5 Objective

1. Collating large datasets from different locations around the world.
2. To develop microbiome-based models for age prediction and to characterize how aging trajectories are altered in disease conditions.
3. To use SHAP values to track how microbial species alter aging trajectories in disease samples
4. Provide a foundation for understanding host-microbiome interactions in the context of age-associated trajectories in associated with different diseases

CHAPTER 3

MATERIAL AND METHODOLOGY:

3.1. Creating a Data Repository:

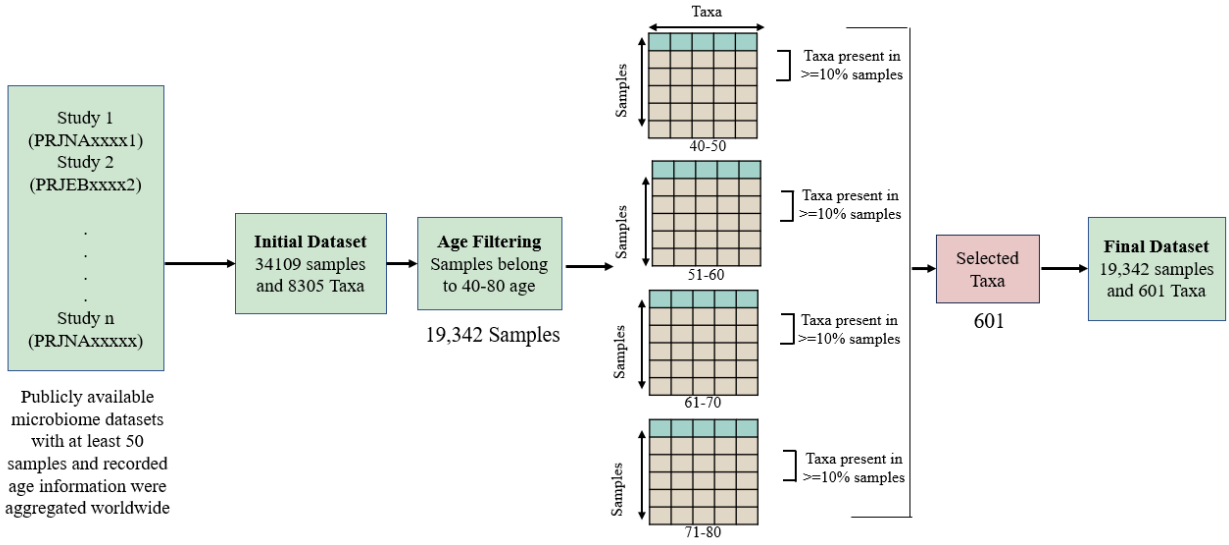


Fig 3.1.1 Creation of Data Repository

To address the limitations of previous studies, the foundational step of this research was the creation of a large-scale, diverse, and globally representative data repository. This was a monumental undertaking that involved the aggregation, curation, and harmonization of publicly available gut microbiome datasets from around the world.

The process began with a systematic search of public repositories to identify studies containing gut microbiome data (both 16S rRNA gene amplicon and shotgun metagenomic sequencing) with associated host age information and a minimum of 50 samples. This initial aggregation yielded a massive dataset of 34,109 samples and 8,305 distinct microbial taxa.

Recognizing that the most significant interplay between microbial aging and chronic disease onset occurs during mid-to-late adulthood, we focused our analysis on a specific and critical age window. All samples were filtered to include only individuals between the ages of 40 and 80. This age-filtering step resulted in a refined dataset of 19,342 samples. This dataset, to our knowledge, represents one of the largest of its kind, spanning 67 distinct study cohorts from 31 different countries. It includes 9,555 shotgun metagenomic samples from 42 studies and 9,787 16S amplicon samples from 26 studies, providing a robust and heterogeneous collection for analysis. The final repository contains data from 10,872 healthy individuals and 8,470 individuals with a wide range of diseases, making it uniquely suited to investigate the impact of disease on microbial aging trajectories.

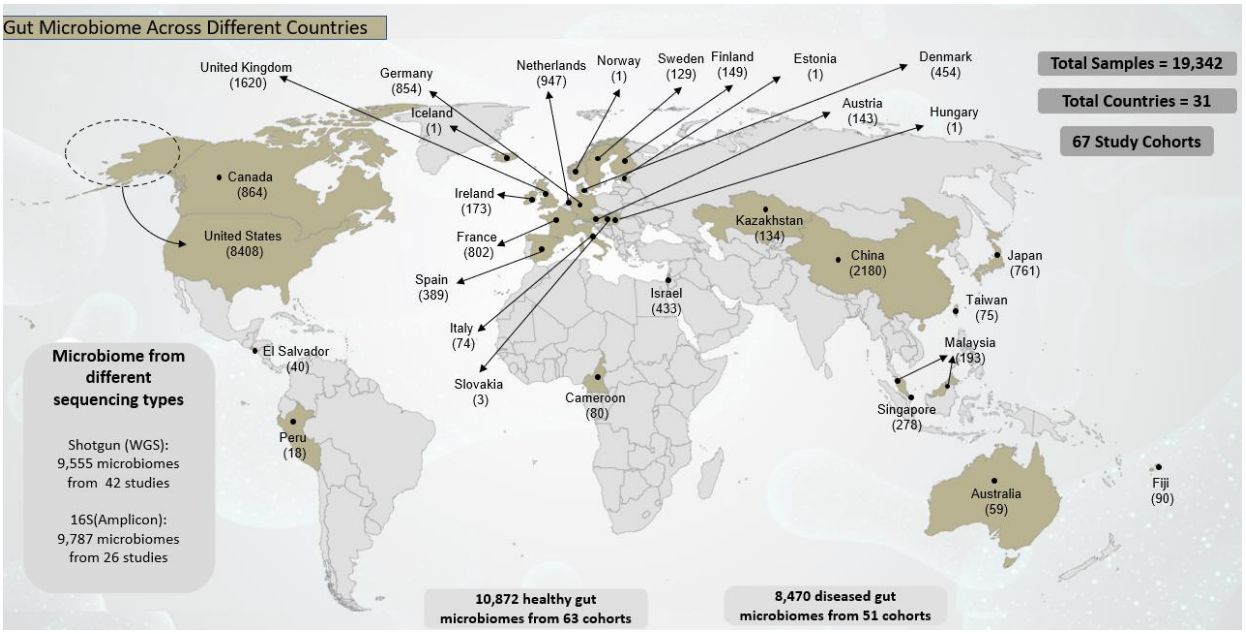


Fig 3.1.2 Gut Microbiome across different countries

3.2. Downloading and Pre-processing the Data:

Only sequence data from the filtered literature was downloaded. Further, based on the sequence strategy used (Amplicon or Shotgun), these sequence files were taken for phylogenetic classification. For amplicon-based sequences, the SPINGO tool (Allard et al., 2015) was used, while for Shotgun sequenced data, the MetaPhlAn version 3 tool was used(*Integrating Taxonomic, Functional, and Strain-Level Profiling of Diverse Microbial Communities with bioBakery 3 | eLife*, n.d.).

The output of these tools was the species names with their abundance for each sample. The species abundance for each sample was stored in the form of a table where columns define the species names and rows define sample names.

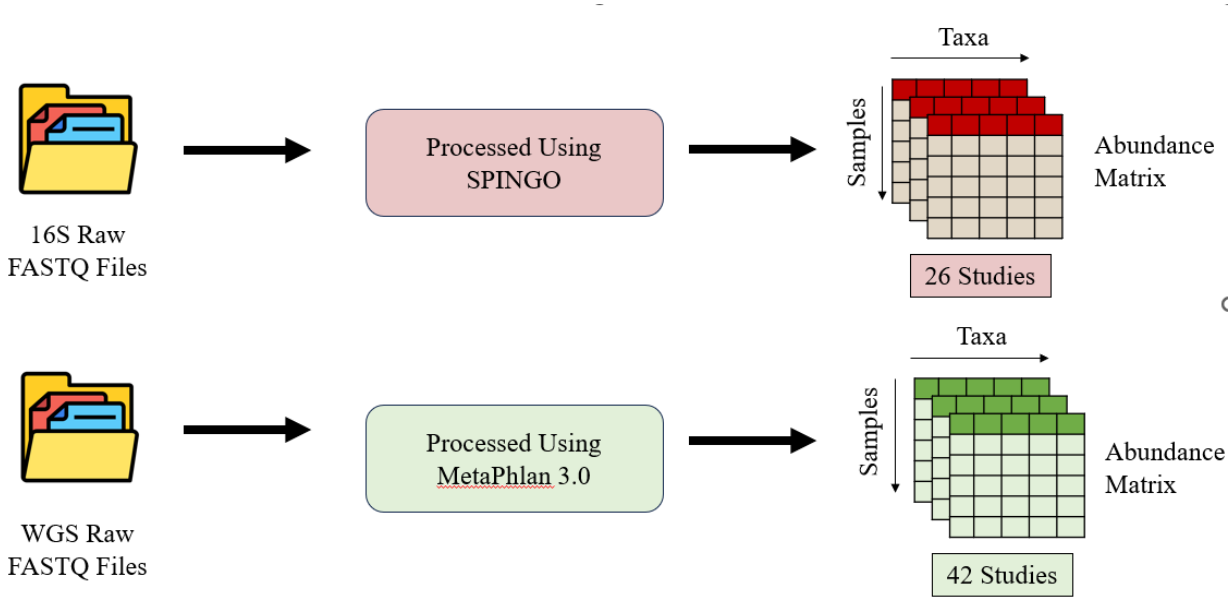


Fig 3.2.1 Collection and preprocessing of Data

3.3 Machine Learning Model of Age Regression

The machine learning regression model was our main analytical tool for this thesis. We trained the model to predict chronological age based on the relative abundances of the 601 chosen microbial taxa. We chose the Random Forest algorithm for this purpose because it has high performance, resistance to overfitting, and intrinsic ability to deal with complicated, non-linear interactions between features, which is typical for microbial ecosystems. The model was run with out-of-bag (OOB) error estimation turned on to allow an unbiased performance measure during training.

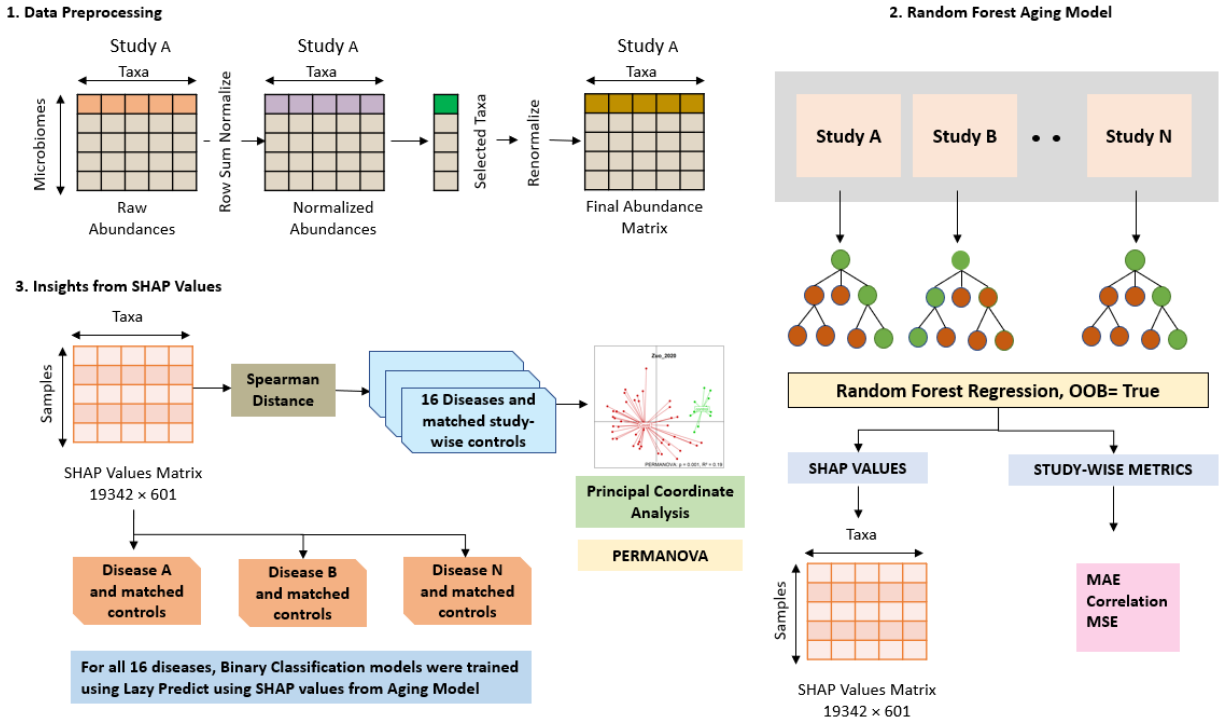


Fig 3.3.1 Complete Workflow

3.4 Model Validation Strategy: Unmixing the Disease Signal

The underlying hypothesis of this work is that disease disrupts the healthy microbial aging course. To validate this, a specific and intentional validation plan was developed. Rather than training on the full dataset, we trained only on a subset of healthy people. In particular, two-thirds (2/3) of the healthy control samples were taken to train the Random Forest model to learn to identify the microbial patterns that reflect healthy aging.

The model was then tested on the remaining one-third (1/3) of healthy controls (to establish a baseline performance) and, crucially, on all disease samples. This approach ensures that the model has no prior knowledge of disease-associated microbial profiles, allowing us to quantify how diseased microbiomes deviate from the healthy aging patterns the model has learned. The primary metric for this deviation was the "reverse residual," calculated as (Predicted Age - Actual Age).

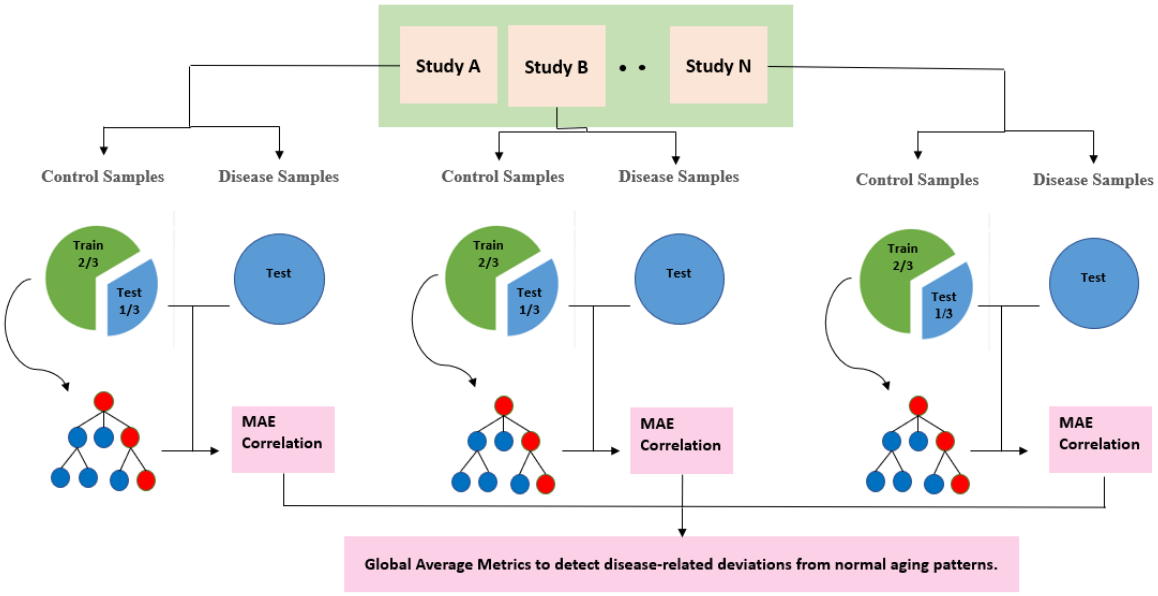


Fig 3.4.1 Model Validation: Unmixing the Disease Signal

3.5 Explainable AI for Signature Extraction: SHAP Values

To go beyond prediction and understand the microbial drivers behind the aging signature, we used an explainable AI (XAI) method known as SHAP (SHapley Additive exPlanations). For each sample, SHAP values were obtained for each of the 601 microbial taxa. A SHAP value represents the contribution of every feature (taxon) towards the model's final age prediction for the given sample.

This step converted our initial abundance matrix into a SHAP value matrix of equal size (19,342 samples × 601 taxa). This matrix is a more understandable "aging signature," in which the value represents the importance of a microbe to the aging process, not just its abundance. This SHAP matrix was used as the main input for all downstream analyses used to compare the aging signatures between healthy and diseased subjects.

CHAPTER 4

RESULTS:

4.1 A Universal Microbial Aging Signal is Evident Across Global Cohorts Before assessing the impact of disease, it was essential to first validate that a consistent, age-related microbial signal existed within our large and heterogeneous dataset. A study-wise correlation analysis was performed between the microbial compositions (based on the 601 selected taxa) and host chronological age for each of the 67 cohorts. The results confirmed a strong and reproducible aging signal. Nearly 60% of the cohorts exhibited a statistically significant positive correlation between the gut microbiome and host age. This foundational finding demonstrates that despite significant geographical, technical, and lifestyle-related variations across the studies, a universal microbial aging pattern is clearly discernible, providing a robust baseline for investigating disease-related perturbations.

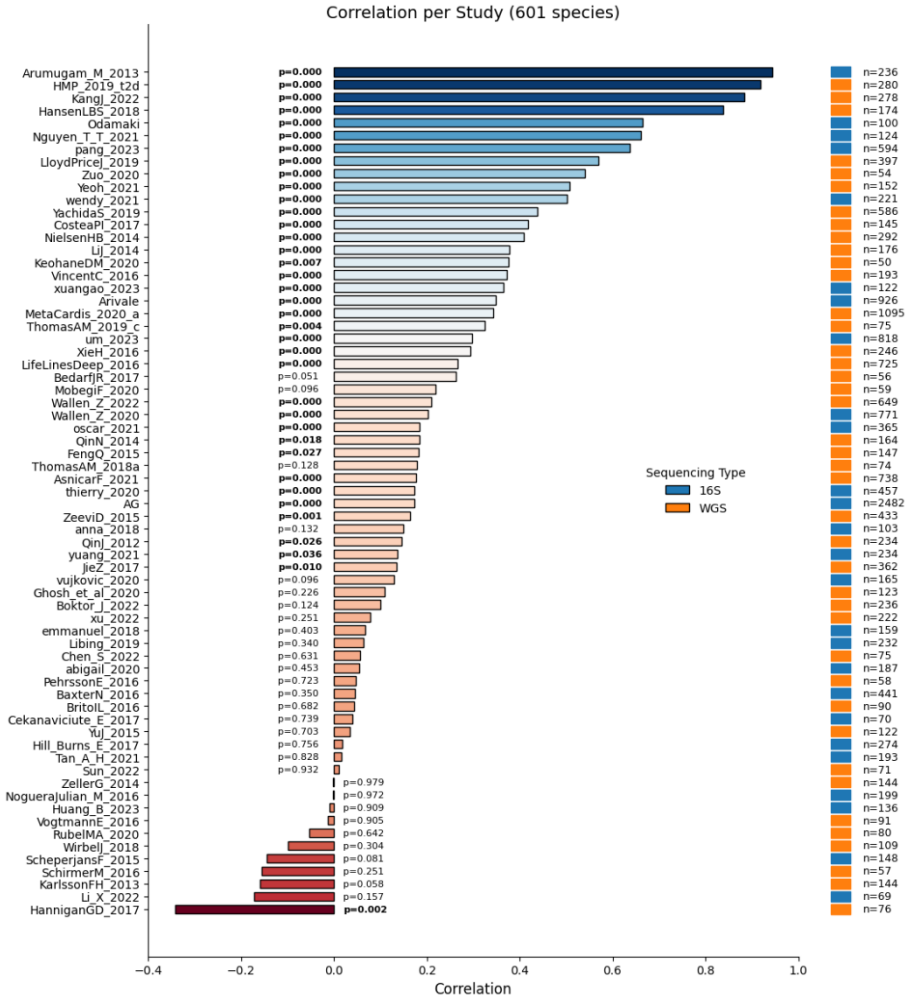


Fig 4.1.1 Strong Age-Microbiome Associations across cohorts

4.2 Disease Fundamentally Disrupts the Healthy Microbial Aging Trajectory

The core experiment of this thesis examined how a model trained only on healthy aging patterns would view the microbiomes of sick people. The outcome showed a dramatic and uniform pattern of disruption. When tested on the held-out healthy control samples, the Random Forest model behaved as it should, with the reverse residuals (Predicted Age - Actual Age) symmetric around a median of zero.

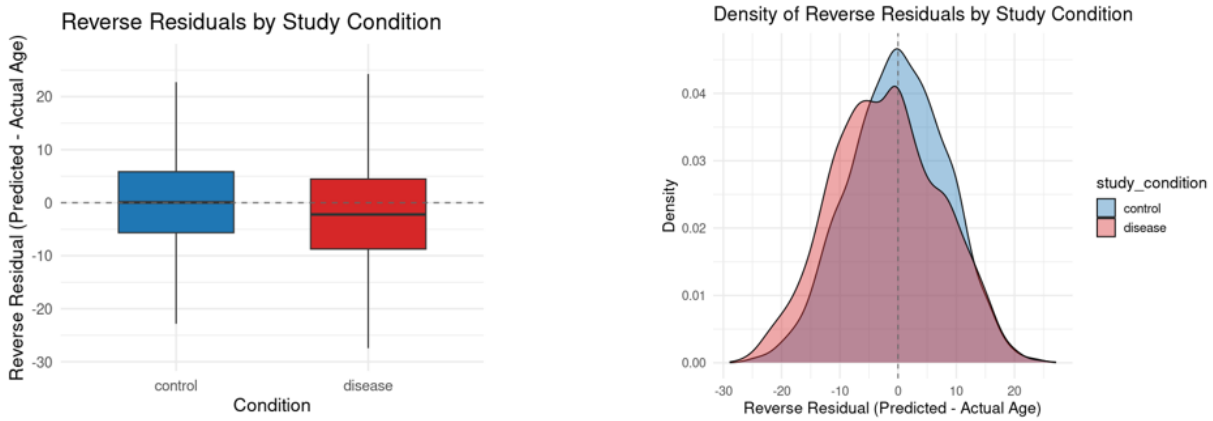


Fig 4.2.1 Figure showing disruptions in disease

In a stark departure, when the model was challenged with the 8,470 disease samples samples it had not been exposed to during training a coherent and consistent deviation was found. The reverse residual distribution of the disease cohort had a strong shift, with a statistically significant negative median, indicating that in every case, the model had systematically underestimated the age of the

	Controls	Diseased
MAE	6.7	7.7
Correlation	0.59	0.38

	Controls	Diseased
MAE	7	8.1
Correlation	0.53	0.34

	Controls	Diseased
MAE	6.9	8.1
Correlation	0.55	0.36

Fig 4.2.2 Comparison of different models for disease disruptions

disease-affected individuals, reading their microbiomes as being biologically "younger" than their chronological age. This is not an indication of rejuvenation but a quantitative measure of dysbiosis. The microbial communities in disease-affected individuals have strayed from the established, healthy aging pathway, leading to compositions that are no longer consistent with the patterns the model was designed to identify as representative of healthy aging.

To confirm the robustness of this underlying finding, the complete experimental process was replicated using two additional and sophisticated machine learning models: XGBoost and LightGBM. The results showed a remarkable consistency across the three models. All three models showed a substantially negative median residual for the disease group, along with a greater Mean Absolute Error (MAE) and a lower correlation between predicted and true value age when compared with the control group. This cross-validation across multiple models confirms that the observed disruption is a true biological signal and not an artifact of a specific algorithm.

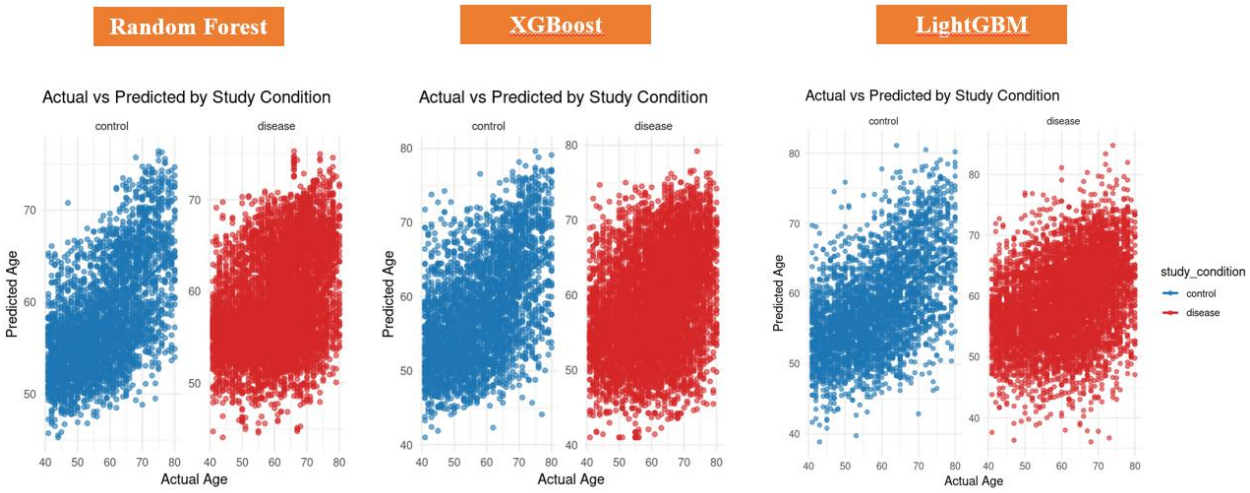


Fig 4.2.3 Scatter plots for actual and predicted ages comparing two study conditions

4.3 Disease-Specific Signatures and the Level of Disturbance

The disruption of the aging trajectory was a near-universal feature of disease; the magnitude of this impact, however, was astonishingly heterogeneous between diseases. The strong correlation between predicted and actual ages observed in healthy controls (Pearson's r = 0.59) decreased gradually with the impact of various diseases. The Prediabetes and Respiratory Infection, for instance, had a moderate correlation (r = 0.44 and r = 0.34, respectively) reflecting a blunting of the aging process. Severe systemic illnesses, on the other hand, like cardiovascular disease (CVD) and COVID-19, obliterated the aging signal completely, with correlations close to zero (r = -0.01 and r = -0.04, respectively).

The impact of this disease on the specific effects was similarly observable in the reverse residuals. Although the majority of disease groups exhibited a negative median residual, the degree of underestimation associated with age differed significantly. Disorders like Inflammatory Bowel Disease (IBD) and HIV demonstrated a marked negative shift, in contrast to conditions such as Type 2 Diabetes (T2D) and Prediabetes, which revealed a comparatively mild divergence from the healthy baseline. This indicates that various pathologies may influence unique quantitative characteristics on the microbial aging clock.

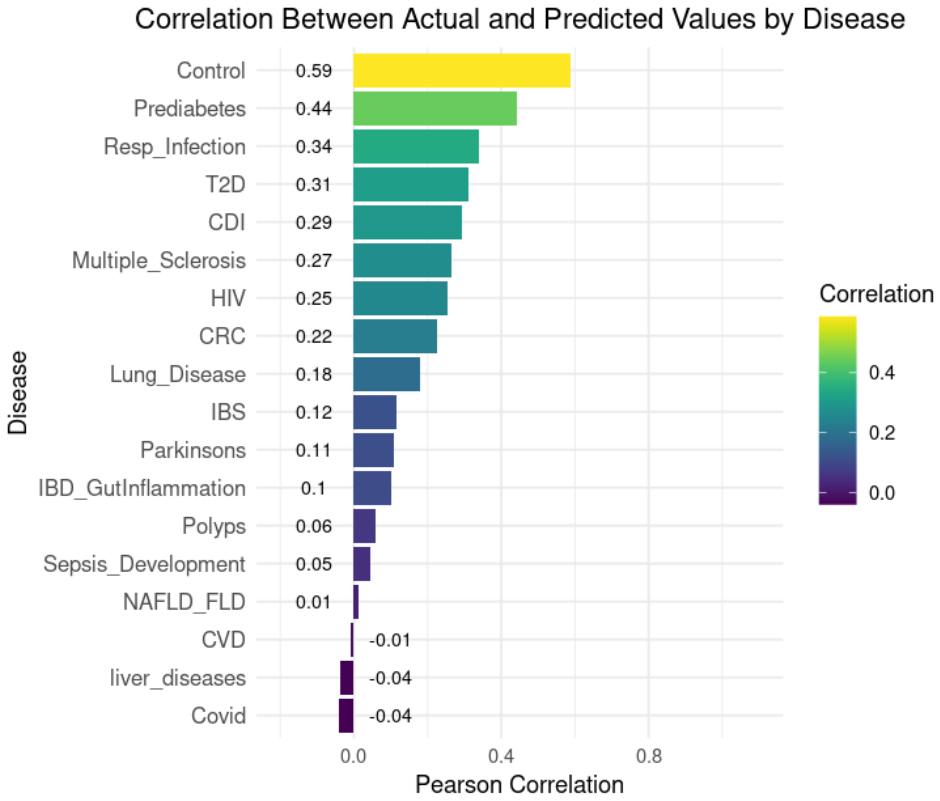


Fig 4.3.1 Correlation of actual and predicted ages of different diseases

4.4 Visualization of Disease-Specific Aging Signatures with Explainable AI

In our attempt to transcend simple quantitative differences and show the difference in the underlying patterns of microbes, we compared the aging signatures in terms of SHAP values. We used Principal Coordinate Analysis (PCoA) of Spearman distances between these SHAP signatures to visualize the samples in a lower-dimensional space, thereby enabling visual assessment of discrimination between diseased subjects and their respective healthy controls per study. For statistical assessment of whether these groups differed significantly, we used PERMANOVA. We found different patterns of discrimination that varied with the type of disease.

4.4.1 Strong and Consistent Aging Signature Disruptive Diseases

A number of diseases resulted in a significant and highly reproducible change of the microbial aging signature in many independent cohorts.

Metabolic Disorders (Prediabetes and Type 2 Diabetes): Prediabetes and Type 2 Diabetes both indicated a strongly consistent and statistically significant deviation from healthy controls across all cohorts examined. Such a strong signal suggests that metabolic dysregulation is inextricably linked with a recognizable and reproducible disturbance of the microbial aging process.

COVID-19 and Respiratory Infections: The acute infectious conditions had a significant impact. Among the two cohorts that were COVID-19 tested, the patient cohort formed a definite cluster distinct from the control group (PERMANOVA p=0.001), illustrating a significant and instant change in the aging profile. Likewise, a stable and strong divergence was observed for other respiratory infections.

Cardiovascular Disease (CVD) and Inflammatory Bowel Disease (IBD) both had strong and unambiguous divergence from control groups within their respective groups, indicating that chronic inflammatory diseases, whether systemic or local in the gastrointestinal tract, leave a strong and noticeable signature of disrupted microbial aging.

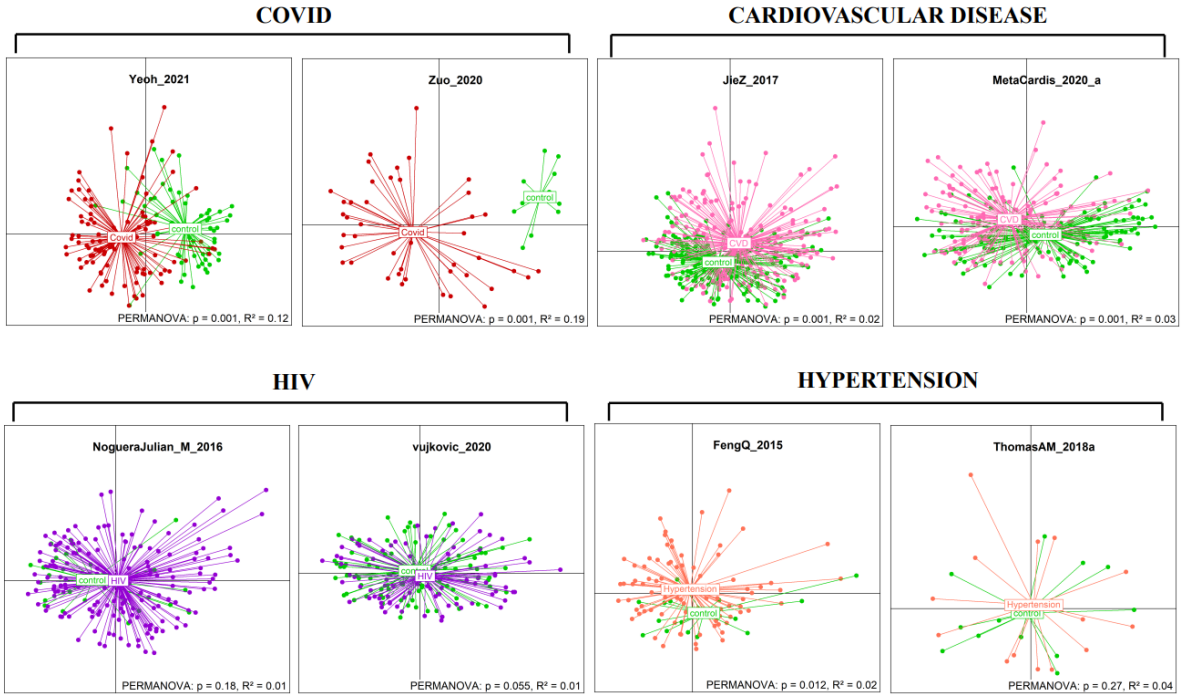


Fig 4.4.1.1 Plot showing significant differences in different diseases

4.4.2 Heterogeneous and Cohort-Dependent Signature Disorder

In contrast to the homogeneous patterns outlined above, the majority of complex diseases had a more heterogeneous pattern of disruption that varied between the study populations.

Colorectal Cancer (CRC): The effect of CRC on the aging signature was highly variable. In some cohorts, there was an apparent and statistically significant distinction between the CRC patients and controls. In others, however, the two groups were very overlapping with no distinction. This suggests that perturbation of microbial aging in CRC could be influenced by a variety of variables, such as the cancer stage, tumor location, specific driver mutations, or population diet and lifestyle interactions.

Parkinson's Disease (PD): There was also the same variability pattern for Parkinson's Disease. Although there was an obvious difference between PD patients and controls in some groups, this was not so in others. This may be due to the multifactorial etiology of PD, the effect of medication (e.g., L-DOPA), or heterogeneity in disease duration among participants.

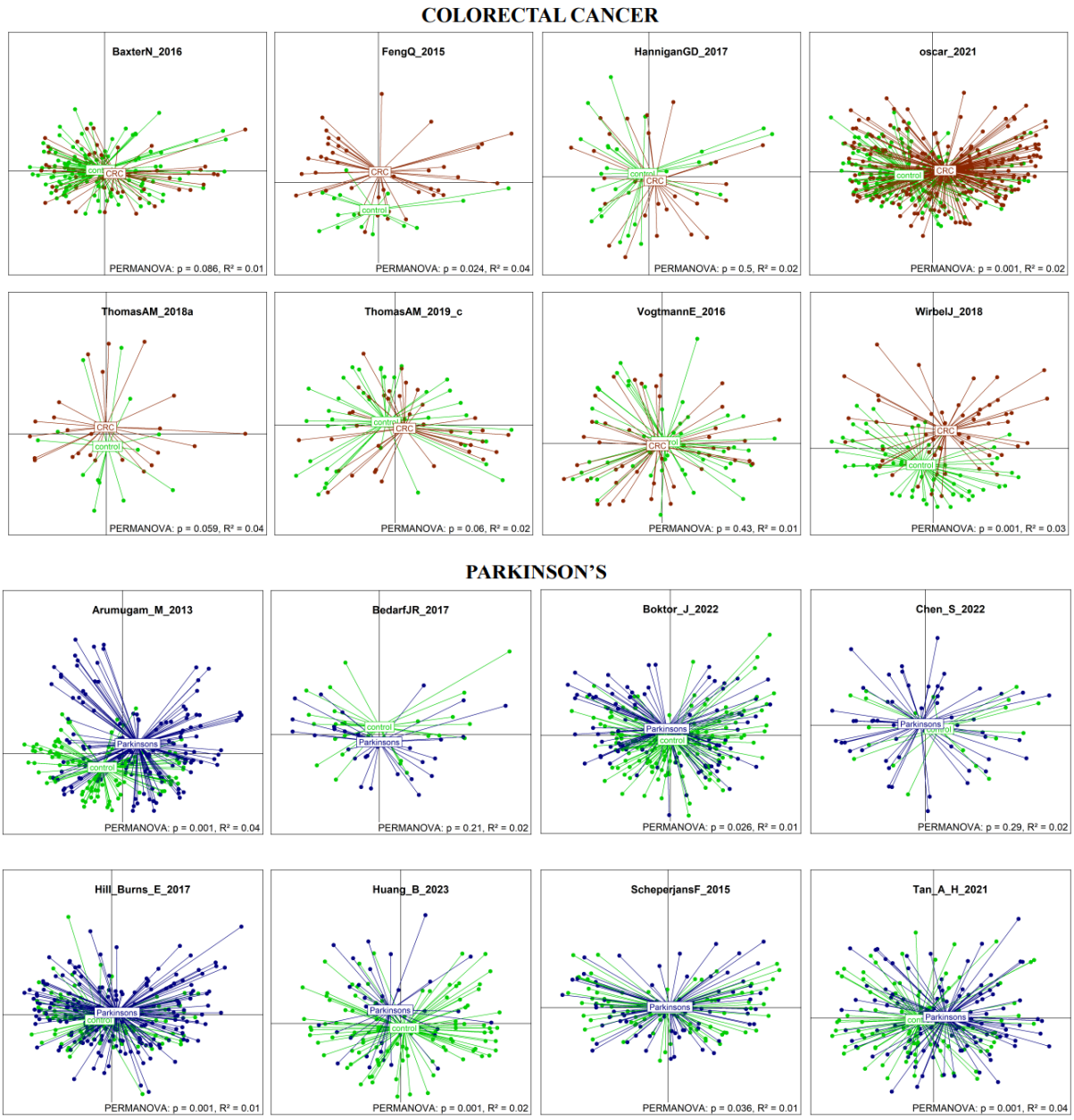


Fig 4.4.2.1 PCoA plots for CRC and Parkinson's

4.4.3 Disorders with Lower or Overlapping Signatures

Finally, some conditions had a reduced or non-existent difference in their aging markers relative to control groups.

HIV: In the cohorts examined, HIV-positive individuals did not exhibit a clear, consistent discrimination from healthy controls. The clusters overlapped to a great extent, and PERMANOVA tests were not always significant. This could imply that with highly potent antiretroviral therapy, the microbial aging signature is not as severely compromised as with other chronic diseases, although further studies are required to validate this. Hypertension: The hypertension indication was not consistent either, as there was disparity in one cohort, but overlap

in the other, indicating a variable or weaker effect on the microbial aging clock. 4.5 Aging-Related Characteristics are Strong Classifiers of Disease To finally validate the biological relevance of these aging signatures, we formulated the performance of SHAP-based features from our age-regression model to address a classification problem: diseased vs. healthy controls. Using the LazyPredict library to compare a range of classification algorithms, we found that the aging-specific SHAP features worked exceptionally well. For numerous diseases, such as Respiratory Infection, IBD, T2D, and COVID-19, the models had excellent classification performance, with ROC AUC values above 0.85 and, in some cases, above 0.90. This discovery strongly demonstrates that the microbial features most important for characterizing healthy aging are also most severely disrupted by disease, and thus represent ideal biomarkers for discrimination between states of health and disease.

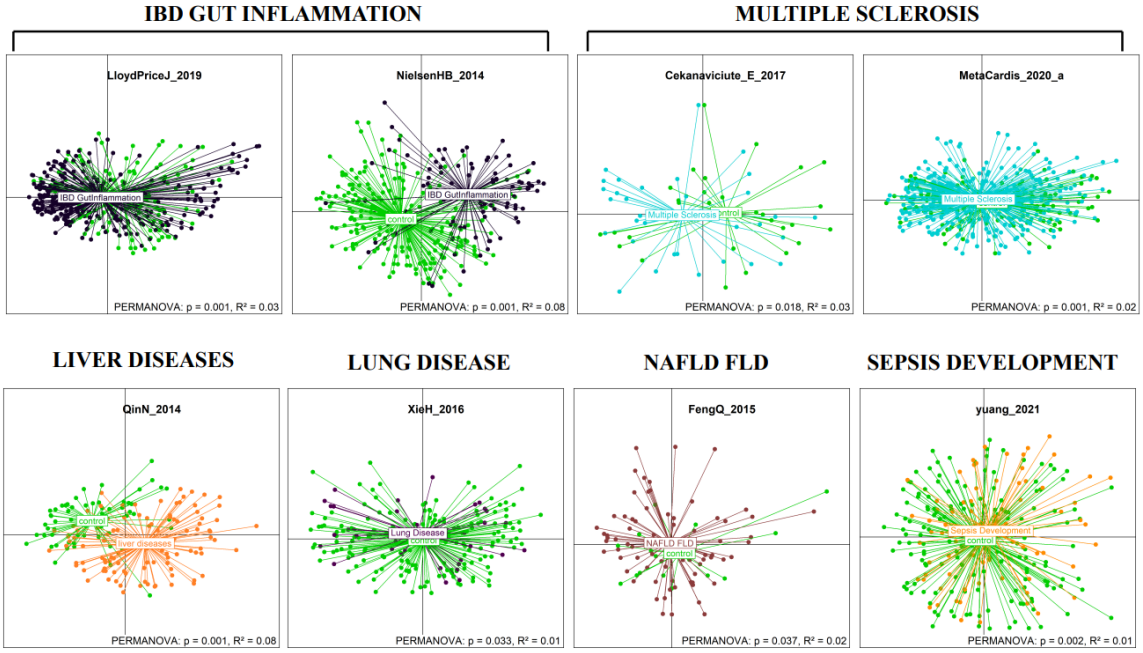
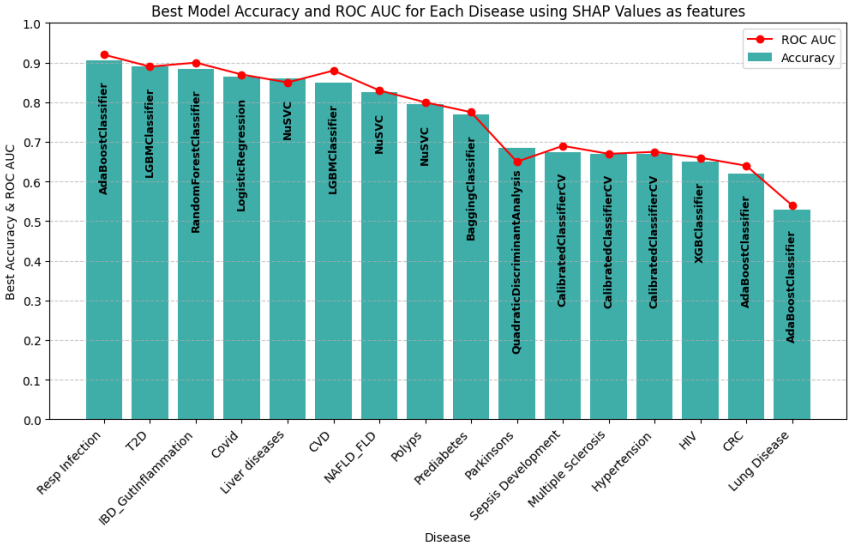


Fig 4.4.3.1 Significant separation is evident across gut, neurological, metabolic, and systemic diseases, from IBD to Sepsis



CHAPTER 5

CONCLUSION AND LIMITATIONS:

5.1 Discussion and Conclusions

This thesis seeks to bridge a vital knowledge gap in our understanding of the interplay between the gut microbiome, aging, and disease states. By building an unprecedentedly wide and dense global dataset, and applying a new explanatory analytical framework based on explainable artificial intelligence, this study reports the first systematic insights into how various diseases shape the microbial aging trajectory.

The most significant result of this research is that illness profoundly modifies the predetermined course of microbial aging. Our working definition of "healthy aging," when applied to microbiomes of sick patients, consistently underestimated their chronological ages. This frequent observation of a "broken clock" is a significant finding, suggesting that many pathological states drive the gut microbiome into a dysbiotic state in which its development does not follow the rules of healthy aging. This provides a new understanding of disease: not simply as an isolated state, but as a process that disrupts the normal, age-related course of our microbial partners.

In addition, this study illustrates that the nature of this perturbation is not random but is disease-specific and patterned. We found conditions such as T2D and COVID-19 that confer a strong and reproducible signature of perturbed aging in numerous populations. In contrast, we determined that multifactorial diseases such as CRC and Parkinson's Disease confer a more heterogeneous signature, a critical observation that underscores the necessity of more nuanced, more stratified analyses in future studies.

Lastly, the successful application of aging-related traits to disease state categorization provides strong evidence in support of the theory. It affirms that the most significant microbial taxa engaged in aging are also most engaged in maintaining healthy states. This provides a novel framework in terms of aging that is able to allow comparisons of the microbial underpinnings of different apparently unrelated diseases and identify shared key species for further investigation.

5.2 Limitations of the Study

While this study offers much new information, it is worth mentioning its limitations, which simultaneously offer possible directions for further research.

Cross-Sectional Design: The current analysis applies the cross-sectional meta-analysis technique, combining information from different individuals at a single time point. Although this design is very good at uncovering strong associations, it cannot discriminate on causal relationships or track the emerging pattern of an individual's microbial aging trajectory over time. Longitudinal studies tracking individuals through disease inception and evolution are thus required to confirm these findings.

Uncontrolled Confounding Variables: Being a meta-analysis of publicly available data, it was not feasible to control for all the confounding variables. Diet, use of medications (e.g., metformin, antibiotics, L-DOPA), disease severity, and certain subtypes of diseases were not consistently available across all 67 cohorts. These variables most definitely affect the microbiome and may at least be partly responsible for some of the heterogeneity seen, especially in diseases such as CRC and Parkinson's.

Technical Heterogeneity: Despite the application of standardized bioinformatic pipelines, the database has samples that have been derived from various DNA extraction methods and sequencing technologies, including 16S rRNA and shotgun metagenomics. Although the large number of samples reduces some types of variability, there can still be inherent technical biases.

Taxonomic Resolution: The study was conducted predominantly at the species level. This does not include variation at the strain level, which could be of functional importance. Different strains of the same species can have very different metabolic processes and impact upon the host, an extent of specificity not included in this general study. Overall, this thesis sets a firm and thorough foundation for the new field of disease-conscious microbial aging. The "preliminary snapshots" worked out in this thesis provide a new paradigm and an array of hypotheses for further, more focused research into the mechanisms by which disease derails our microbial clock and the possible avenues to its eventual restoration.

CHAPTER 6

REFERENCES:

1. Allard, G., Ryan, F. J., Jeffery, I. B., & Claesson, M. J. (2015). SPINGO: A rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics*, *16*(1), 324. https://doi.org/10.1186/s12859-015-0747-1
2. Chen, Y., Zhou, J., & Wang, L. (2021). Role and Mechanism of Gut Microbiota in Human Disease. *Frontiers in Cellular and Infection Microbiology*, *11*, 625913. https://doi.org/10.3389/fcimb.2021.625913
3. Galkin, F., Mamoshina, P., Aliper, A., Putin, E., Moskalev, V., Gladyshev, V. N., & Zhavoronkov, A. (2020). Human Gut Microbiome Aging Clock Based on Taxonomic Profiling and Deep Learning. *iScience*, *23*(6), 101199. https://doi.org/10.1016/j.isci.2020.101199
4. Ghosh, T. S., Shanahan, F., & O'Toole, P. W. (2022). The gut microbiome as a modulator of healthy ageing. *Nature Reviews Gastroenterology & Hepatology*, *19*(9), 565–584. https://doi.org/10.1038/s41575-022-00605-x
5. *Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3 | eLife*. (n.d.). Retrieved August 21, 2025, from https://elifesciences.org/articles/65088
6. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.
7. Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J., Abe, F., & Osawa, R. (2016). Age-related changes in gut microbiota composition from newborn to centenarian: A cross-sectional study. *BMC Microbiology*, *16*(1), 90. https://doi.org/10.1186/s12866-016-0708-5
8. Ratiner, K., Abdeen, S. K., Goldenberg, K., & Elinav, E. (2022). Utilization of Host and Microbiome Features in Determination of Biological Aging. *Microorganisms*, *10*(3), 668. https://doi.org/10.3390/microorganisms10030668
9. Shen, X., Wang, C., Zhou, X., Zhou, W., Hornburg, D., Wu, S., & Snyder, M. P. (2024). Nonlinear dynamics of multi-omics profiles during human aging. *Nature Aging*, *4*(11), 1619–1634. https://doi.org/10.1038/s43587-024-00692-2
10. Simbirtseva, K. Y., & O'Toole, P. W. (2025). Healthy and Unhealthy Aging and the Human Microbiome. *Annual Review of Medicine*, *76*(Volume 76, 2025), 115–127. https://doi.org/10.1146/annurev-med-042423-042542
11. Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ (Clinical Research Ed.)*, *361*, k2179. https://doi.org/10.1136/bmj.k2179
12. Wilmanski, T., Gibbons, S. M., & Price, N. D. (2022). Healthy aging and the human gut microbiome: Why we cannot just turn back the clock. *Nature Aging*, *2*(10), 869–871. https://doi.org/10.1038/s43587-022-00294-w