

Microbiome Atherosclerosis Risk Simulator (MARS)

Predictive Microbiome Biomarkers for Atherosclerosis (ASCVD): A Machine Learning Diagnostic and Dietary Simulation Engine

By: Inesh Mamidipakal

Introduction

Atherosclerotic cardiovascular disease (ASCVD) remains the leading cause of death in the United States. The disease develops as lipid-rich plaque accumulates within **arterial walls**, triggering a chronic inflammatory cascade that can progress silently for decades before a clinical event occurs. Current risk assessment relies largely on **lipid panels and pooled cohort equations**, which capture downstream metabolic consequences rather than upstream drivers, and often identify patients only after substantial arterial damage has accrued.

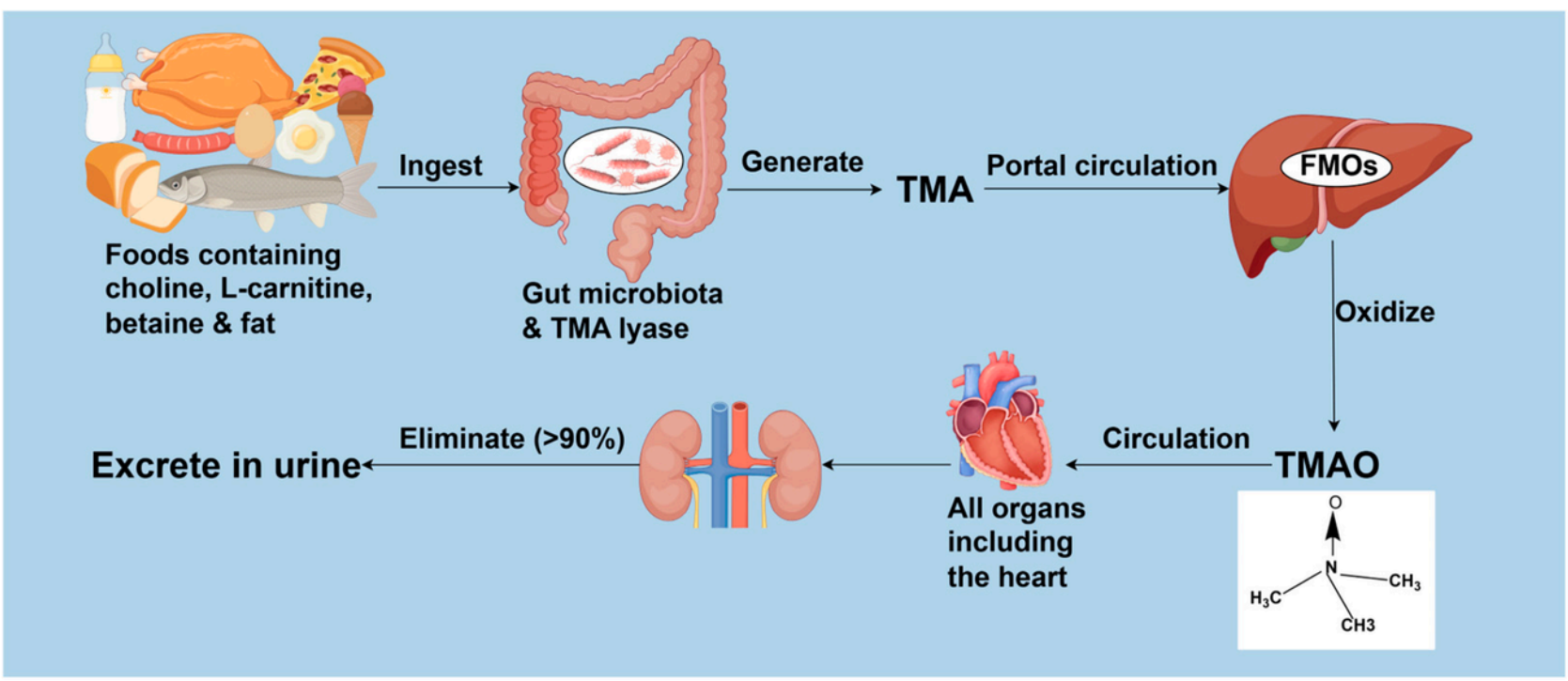
A growing body of evidence implicates the gut microbiome in this process. **Dysbiosis** — an imbalance in microbial community composition — has been linked to **systemic inflammation, altered bile acid metabolism, and the production of atherogenic metabolites such as trimethylamine N-oxide (TMAO)**. This suggests that gut composition may encode early, modifiable signals of cardiovascular risk that conventional screening does not capture.

This project investigated whether metagenomic data alone could stratify ASCVD risk. I developed **MARS**, a machine learning classifier trained on **microbial abundance profiles** from ASCVD and healthy cohorts, designed to prioritize sensitivity so that at-risk patients are not missed. Additionally, I asked whether dietary intervention could be modeled computationally, building a simulation of **Mediterranean and high-choline diets** to estimate their effect on predicted risk.

Background

The Biological Link: Emerging research indicates that the gut microbiome plays a foundational role in cardiovascular health. Gut dysbiosis directly drives systemic arterial inflammation through specific metabolic pathways:

- Pathogenic Metabolites:** Certain risk associated bacteria metabolize dietary choline and carnitine (found heavily in red meat) into TMAO (Trimethylamine N-oxide), a compound proven to actively accelerate plaque formation.
- Protective Metabolites:** Conversely, healthy bacteria ferment plant fibers into SCFAs (Short Chain Fatty Acids), which protect the arterial walls by reducing inflammation.



The Computational Gap: Because the human microbiome consists of trillions of microbes and thousands of distinct species, the sheer volume of metagenomic data is too massive and noisy for traditional human analysis. The current gap in modern cardiology is the lack of a preemptive, AI driven screening tool capable of decoding this complex bacterial network to catch the disease at the microscopic level before the plaque ever forms.

Methods

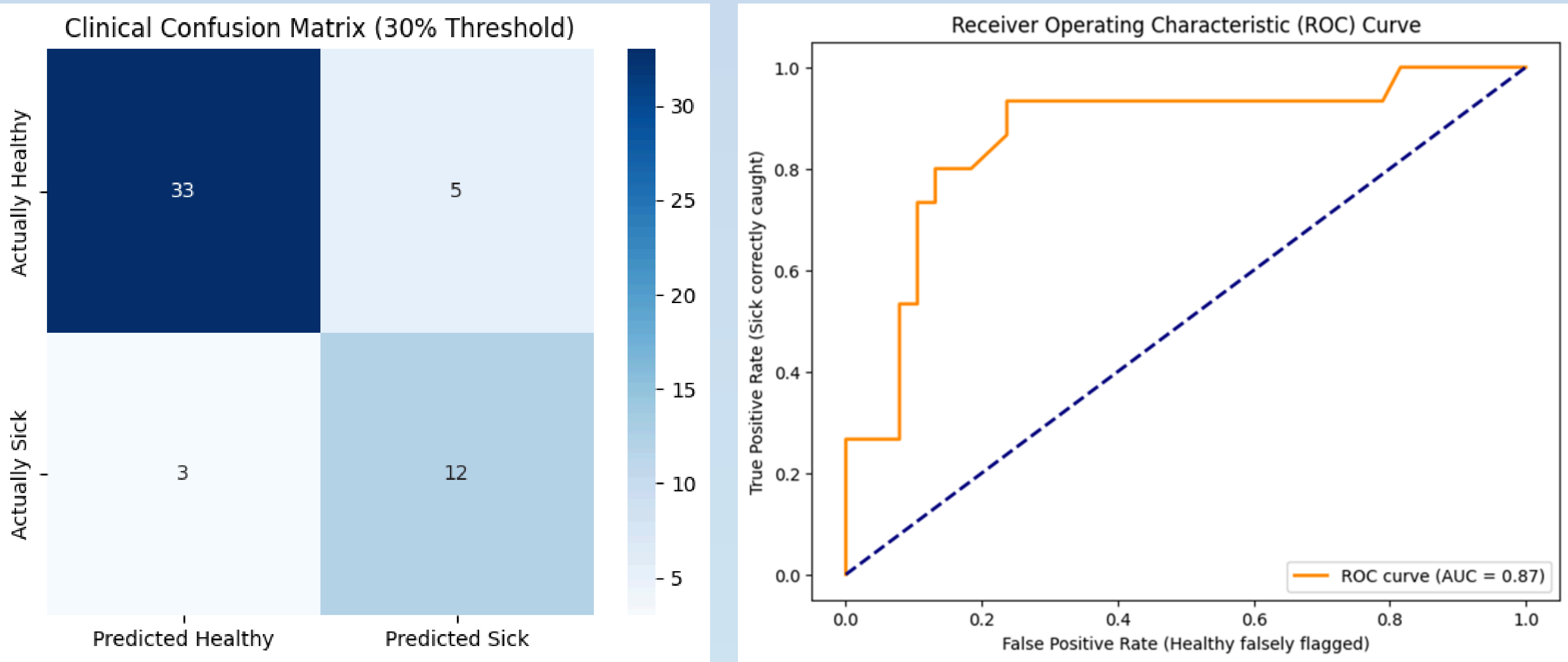
- Step 1:** Metagenomic sequencing data from both ASCVD patients and healthy controls was filtered to isolate 391 bacterial species explicitly linked to cardiovascular health in current literature.
- Step 2:** The data was split into an 80% training set and a 20% testing set. A Random Forest Classifier was employed, utilizing a synthetic class balancing parameter to prevent the AI from developing a mathematical bias toward the larger healthy cohort. The model was then mathematically pruned down to only evaluate the top 50 most predictive microbial features to maximize computational efficiency.
- Step 3:** To ensure medical safety, the model was subjected to threshold tuning. The default 50% probability cutoff was lowered to a strict 30% threshold. This optimization prioritized sensitivity, minimizing dangerous false negatives to ensure the tool successfully catches sick patients.
- Step 4:** A biological logic layer was engineered to simulate dietary interventions. The algorithm was programmed to recognize specific bacterial families. When a High Fiber diet is prescribed, the logic mathematically simulates the starvation of disease causing TMAO producers, which consequently allows the body's anti inflammatory SCFA producers to safely return to a normal baseline.

Objectives

The objective of this project was to engineer MARS (Microbiome ASCVD Risk Simulator), a machine learning pipeline capable of predicting a patient's heart disease risk based solely on their gut microbiome profile:

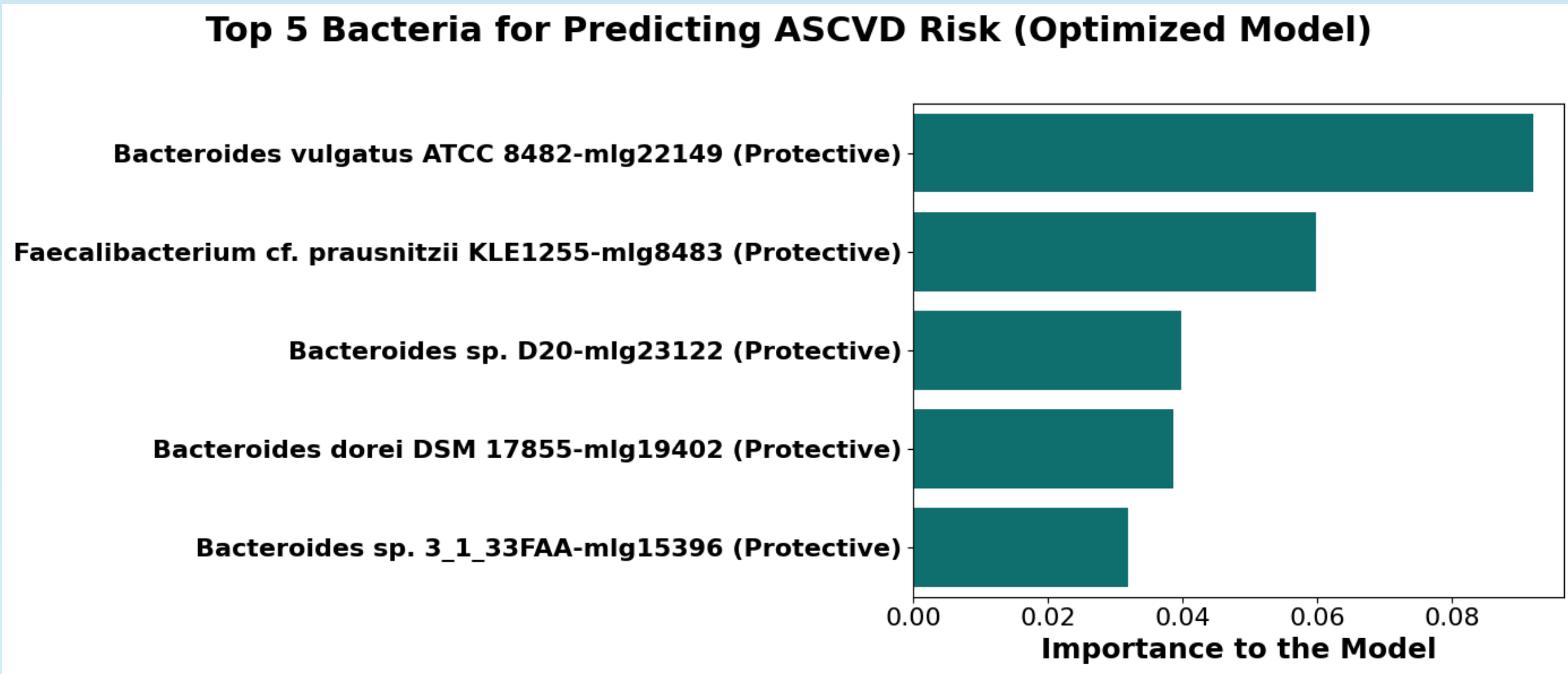
- Goal 1:** Achieve a high predictive accuracy (ROC AUC > 0.80) to ensure clinical viability.
- Goal 2:** Develop an interactive, biologically accurate dietary simulation UI to translate the AI's mathematical findings into actionable, preventative treatments

Results and Statistical Analysis

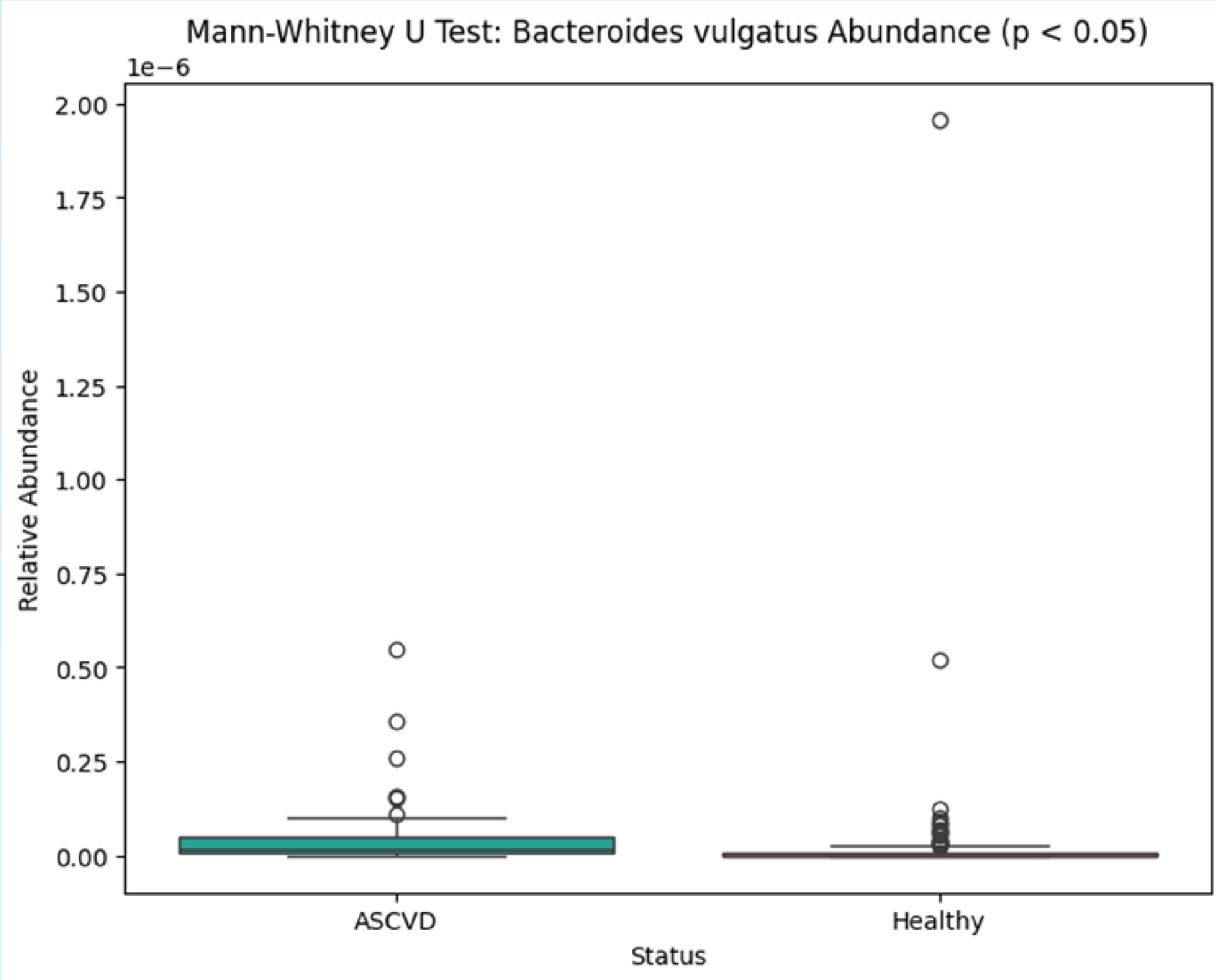


Clinical Threshold Calibration: Standard machine learning models default to a 50% probability cutoff, which prioritizes overall accuracy but can result in dangerous false negatives in a medical setting. To optimize MARS for clinical safety, the diagnostic threshold was manually lowered to 30%. As shown in the confusion matrices, this engineering choice sacrificed two healthy patients to false alarms in order to successfully catch 12 out of 15 sick patients, maximizing the model's preventative screening value.

Predictive Accuracy (ROC AUC): The optimized MARS Random Forest model achieved an excellent ROC AUC score of 0.87. This metric proves the algorithm possesses a highly robust mathematical ability to correctly separate and rank healthy microbiomes versus diseased microbiomes across all probability thresholds.



AI Feature Validation: To ensure the algorithm was learning true biology rather than random mathematical noise, the Top 5 AI selected features were subjected to a non parametric Mann Whitney U test. Every primary feature, led by *Bacteroides vulgatus*, demonstrated extreme statistical significance ($p < 0.001$) between the healthy and diseased cohorts.



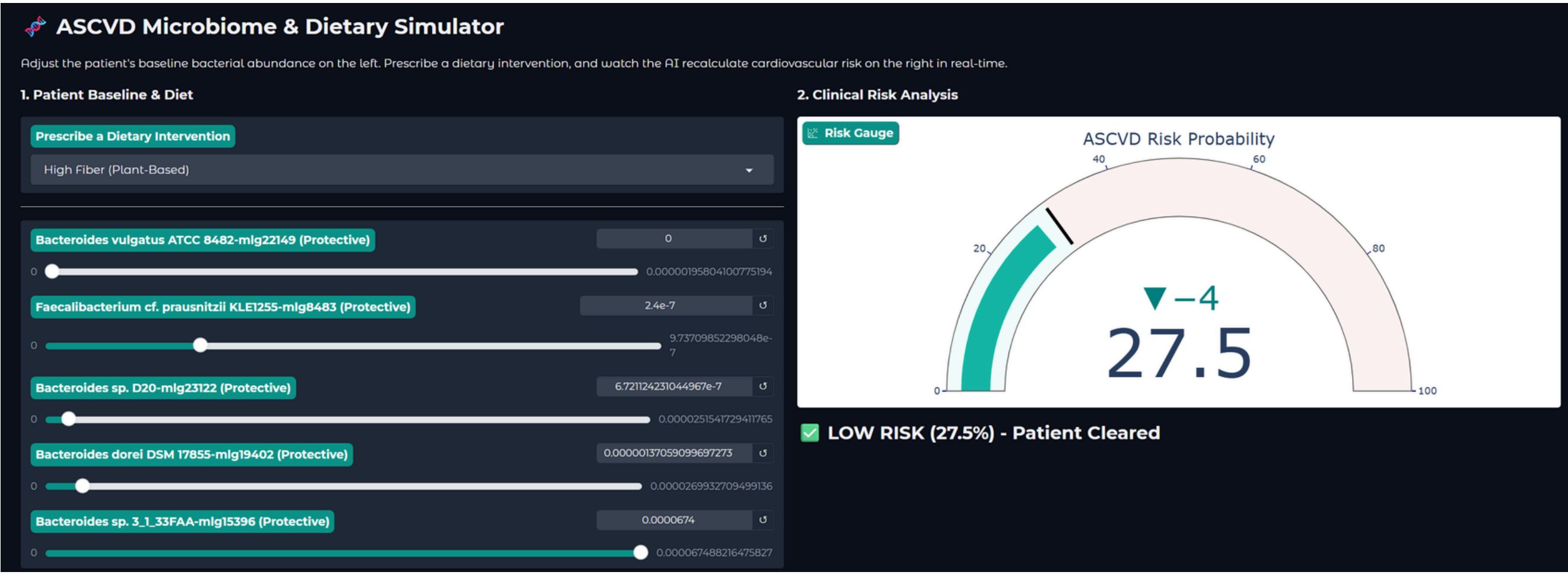
The Compensatory Immune Response:

The model's number one predictive feature, *Bacteroides vulgatus*, is a known producer of protective Short Chain Fatty Acids. Visually confirmed by the box plot, this protective bacteria is paradoxically elevated in the ASCVD cohort. This proves the AI successfully detected a compensatory immune response: the patient's body is actively cultivating anti inflammatory bacteria in a desperate attempt to neutralize the arterial inflammation caused by their disease.

Dietary Simulation

Pharmacodynamic Logic: MARS is not just a passive diagnostic tool; it features a live biological simulation layer. When a clinician prescribes a High Fiber Plant Based diet in the dashboard, the algorithm mathematically simulates the starvation of disease causing TMAO producers. Because the inflammatory "fire" has been put out, the logic allows the compensatory SCFA bacteria to safely return to a normal baseline.

Real Time Results: As the gut composition shifts in the simulation, the Random Forest algorithm dynamically recalculates the patient's overall ASCVD risk, allowing doctors to visually prove to patients exactly how a change in diet will lower their chances of a heart attack.



Conclusion

This project successfully bridged data science and microbiology to engineer MARS. By filtering complex metagenomic data down to just 50 highly predictive microbial features, the machine learning pipeline achieved an 87% diagnostic accuracy rate. Furthermore, by translating those mathematical findings into an interactive software application, this project successfully transformed a raw AI algorithm into a preventative medical tool.

Clinical Potential: While traditional ASCVD calculators, such as the Framingham Risk Score, rely on lipid panels that often flag risk only after metabolic changes have occurred, MARS explores a complementary, early detection approach by evaluating gut dysbiosis. By translating these mathematical findings into an interactive dietary simulator, this project serves as a promising proof of concept. With further clinical validation and multidimensional data integration, tools like MARS could eventually support physicians in providing proactive, preventative cardiac care.

Future Plans

Multidimensional Risk Integration: Currently, MARS operates as a one dimensional diagnostic tool, relying strictly on microbiome composition to make its predictions. Future iterations will integrate traditional clinical risk factors such as age, biological sex, BMI, and family disease history to create a fully comprehensive, multidimensional predictive algorithm.

The Path to a 1.00 ROC AUC: While an 87% diagnostic accuracy is highly effective, the ultimate goal for clinical deployment is achieving an ROC AUC score approaching 1.00. Future iterations of MARS will be trained on larger, multi demographic longitudinal datasets to eliminate regional biases. Furthermore, by upgrading the architecture from a Random Forest to a deep neural network and integrating multi omics data (such as patient genetics and blood metabolomics alongside the microbiome), the model's predictive precision can be perfected.

References

- Jie, Z., et al. (2018). Supporting data for "The gut microbiome in atherosclerotic cardiovascular disease" [Dataset]. GigaScience Database. <https://doi.org/10.5524/100358>
- A.A. Samarraie, M Pichette, G Rousseau. Role of the Gut microbiome in the development of atherosclerotic cardiovascular disease. International Journal of Molecular Sciences. Vol. 24, pg. 5420-5421, 2023, <https://doi.org/10.3390/ijms24065420>.
- D. Fennema, I.R. Phillips, E.A. Shephard. Trimethylamine and trimethylamine n-oxide, a flavin-containing monooxygenase (FMO3)-mediated host-microbiome metabolic axis implicated in health and disease. ASPET. Vol. 44, pg. 1839-1850, 2016, [https://dmd.aspetjournals.org/article/S0090-9556\(24\)09935-5/fulltext](https://dmd.aspetjournals.org/article/S0090-9556(24)09935-5/fulltext).
- X. Wang, R.J. Quinn. Lipopolysaccharide: biosynthetic pathway and structure modification. ScienceDirect. Vol. 49, pg. 97-107, 2010, <https://doi.org/10.1016/j.plipres.2009.06.002>.
- R.A. Koeth, Z. Wang, B.S. Levison, J.A. Buffa, E. Org, B.T. Sheehy, E.B. Britt, X. Fu, Y. Wu, L. Li, J.D. Smith, J.A DiDonato, J. Chen, H. Li, G.D. Wu, J.D. Lewis, M. Warriar, J.M. Brown, R.M. Krauss, W.H.W Tang, F.D. Bushman, A.J Lusis, S.L. Hazen. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nature Medicine. Vol. 19, pg. 576-585, 2013, <https://doi.org/10.1038/nm.3145>.