

Assessing the Predictive Power of the Oral Microbiome for Cardiometabolic Risk Using Machine Learning Approaches



SUMAYYAH BORDERS^{1,2,3}, Yan Wang⁴

¹BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA; ²Department of Biological Sciences, University of Pittsburgh; ³Department of Computer Science, University of Pittsburgh; ⁴Public and Population Health, Division of Oral and Systemic Health, School of Dentistry, UCLA

Background

- **CVD burden:** In 2023, over 900,000 Americans died from cardiovascular disease (CVD), representing 1 in 3 deaths.
- **Gut microbiome link:** Gut microbiota composition is known to differ significantly in individuals with and without CVD.
- **Emerging oral microbiome evidence:** Recent studies suggest that oral dysbiosis may contribute to systemic inflammation, endothelial dysfunction, and atherosclerotic plaque formation.
- **Data gap:** While associations between the gut microbiome and CVD are well studied, the oral microbiome's predictive power for CVD risk is less understood.
- **NHANES:** The National Health and Nutrition Examination Survey (NHANES) 2009-2012 oral microbiome project provides an opportunity to explore microbial associations with CVD.

Conclusions & Future Directions

- XGBoost (eXtreme Gradient Boosting) achieved the highest predictive performance, indicating strong discriminative ability.
- Random Forest revealed interpretable microbial features, including several phyla among its top predictors.
- Discrepancies in feature importance between models suggest different pathways to prediction.

Future Directions

- Train on half of the NHANES dataset to reduce overfitting
- Higher-resolution microbiome profiling at the genus/species level for greater biological insight
- Integrate systemic biomarker data to identify biologically grounded associations with CVD.

Acknowledgements

Special thanks to Dr. Yan Wang and the Bruins-in-Genomics (BIG) Summer Program for their guidance in this research. This project was funded by the BIG for Dental, Oral, and Craniofacial research NIH grant.

References

Centers for Disease Control and Prevention. (2025, August 7). *National Health and Nutrition Examination Survey*. National Health and Nutrition Examination Survey. <https://www.cdc.gov/nchs/nhanes/index.html>

Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Medical Informatics and Decision Making*, 19(1). <https://doi.org/10.1186/s12911-019-0918-5>

Li, Y., Zhu, M., Liu, Y., Luo, B., Cui, J., Huang, L., Chen, K., & Liu, Y. (2022). The oral microbiota and cardiometabolic health: A comprehensive review and emerging insights. *Frontiers in Immunology*, 13. <https://doi.org/10.3389/fimmu.2022.1010368>

Methods & Results

