

Gut Microbiome ML: Multi-Omics Integration for Dysbiosis Classification

Your Name¹ and Co-Author²

¹ Your Institution your.email@example.com

² Second Institution

Abstract. This article presents a machine learning approach to gut microbiome analysis, integrating metagenomics and fluxomics to identify dysbiosis patterns.

Keywords: Microbiome, Machine Learning, Multi-Omics, Dysbiosis

1 Introduction

2 Introduction

The human gut microbiome plays a critical role in health and disease, influencing metabolic processes, immune function, and even neurological pathways. Advances in multi-omics analysis, particularly metagenomics and fluxomics, allow for a deeper understanding of microbiome composition and functionality.

Recent studies have demonstrated that machine learning (ML) can be leveraged to analyze complex microbiome datasets, providing predictive insights into dysbiosis and associated health conditions. By integrating multiple layers of biological data, ML approaches can uncover hidden patterns and key biomarkers relevant to gut health.

In this study, we propose an ML-based framework that combines metagenomic and fluxomic data to classify gut microbiome states. Our approach aims to improve the identification of dysbiosis patterns, enhancing disease prediction and potential therapeutic interventions.

3 Methods

4 Methods

To classify gut microbiome health states, we employ a machine learning pipeline that integrates metagenomic and fluxomic data. Our methodology consists of the following steps:

4.1 Data Acquisition

Metagenomic sequencing data is collected from publicly available microbiome datasets, while fluxomic simulations are generated using genome-scale metabolic models (GEMs). The datasets are preprocessed to remove contaminants and ensure quality control.

4.2 Preprocessing

Metagenomic reads undergo taxonomic and functional profiling using tools like Kraken2 and HUMAnN. Fluxomic data is generated using constraint-based modeling approaches, such as flux balance analysis (FBA), to predict metabolic flux distributions.

4.3 Feature Engineering and Selection

Key microbial and metabolic features are extracted using dimensionality reduction techniques like principal component analysis (PCA) and feature selection algorithms such as recursive feature elimination (RFE). These features serve as inputs for the ML models.

4.4 Model Training and Evaluation

Supervised machine learning models, including Random Forest, XGBoost, and neural networks, are trained to classify healthy and dysbiotic microbiome states. Model performance is assessed using cross-validation and evaluation metrics such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC).

5 Results

6 Results

6.1 Data Distribution and Characteristics

6.2 Model Performance

6.3 Feature Importance

6.4 Comparison with Baseline Approaches

7 Discussion

8 Discussion

8.1 Interpretation of Findings

8.2 Biological and Clinical Implications

8.3 Limitations

8.4 Future Directions

9 Conclusion

This study demonstrates the potential of machine learning in microbiome-based disease prediction. Future work will explore additional omics layers.