[BIOINFO 2025]

Machine Learning and Graph-based Approaches for Microbiome Biomarker Discovery in Inflammatory Bowel Disease

Jaeyoung Jang[1], Seungpyo Hong[1,2*]

[1] *Department of Bioactive Material Sciences, Jeonbuk National University, Republic of Korea*
[2] *Department of Molecular Biology, Jeonbuk National University, Republic of Korea.*

**Corresponding author: seungpyo.hong@jbnu.ac.kr*

Inflammatory bowel disease (IBD), comprising ulcerative colitis (UC) and Crohn's disease (CD), is a chronic immune-mediated disorder arising from complex interactions among host genetics, environmental factors, immune responses, and gut microbiome. The microbiome plays a critical role in IBD pathogenesis; however, clinically reliable and non-invasive biomarkers have yet to be established. Therefore, developing microbiome-based biomarkers to predict disease risk, progression, postoperative recurrence, and therapeutic response is an unmet need. We applied traditional machine learning methods (logistic regression, random forest, XGBoost) and deep neural network (DNN) models to predict IBD status and identify associated microbial signatures. Machine learning models outperformed a graph-based deep learning approach (graph convolutional network, GCN). Feature importance analyses revealed model-specific differences but consistently highlighted taxa previously linked to IBD. Shapley analysis demonstrated model- and class-specific differences in feature contribution, and further identified multiple IBD-associated taxa that were consistently detected across models and classes. The relatively lower performance of the DNN model may reflect class imbalance and the sparsity of microbiome data, which constrain the ability of models with large parameter sets to generalize effectively. Despite these challenges, the overlap of key taxa identified across models is noteworthy. Our findings suggest that integrating predictive modeling with feature evaluation across multiple algorithms can help define consensus microbial features. These features hold potential for development into cost-effective, clinically applicable diagnostic biomarkers for IBD.