February 18, 2020

**DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY**

1520A MEDICAL SCIENCE RESEARCH BUILDING I

ANN ARBOR, MICHIGAN 48109-0620

PHONE: (734) 763-3531

FAX: (734) 764-3562



Dear Editor,

My colleagues and I are happy to resubmit our manuscript, **A framework for effective application of machine learning to microbiome-based classification problems**, for your consideration. This is submitted as a Research Article. Our intention with this manuscript is to communicate to the research community how to apply machine learning methods reproducibly and reliably to microbiome data.

As you are aware, machine learning is a new and emerging area within microbiology and the broader society. With the adoption of any new techniques it is common to see practices that overestimate the value of the technique. Our manuscript describes a number of common pitfalls and solutions. Throughout the manuscript we use a clinically important dataset of 16S rRNA gene sequence collections obtained from 490 stool samples where the donors either had normal colons or screen relevant neoplasias. To facilitate our analysis, we created an open-source and reproducible machine learning pipeline to predict the presence of neoplasias using the relative abundance of bacterial populations. We showcase how to this pipeline can be used for this type of classification problem using seven different machine learning algorithms.

We presented these results at the 2019 Microbe and received considerable interest from researchers across the microbial ecology community. The manuscript was previously submitted to bioRxiv as a preprint.

Sincerely,



Patrick D. Schloss, PhD

Professor