ABSTRACT

The gut microbiota, the large community of bacteria and other microorganisms living in our digestive system, has been recently considered a great potential source of understanding factors affecting human health, and thus novel therapeutic approaches. A healthy gut microbiota helps our body in crucial physiological functions, but an imbalanced microbiota has been linked to many diseases. It is strongly subject to many factors and thus varies greatly across individual. We thus set out to explore how well gut microbiome compositional data can predict an individual's demographic characteristics using various multivariate analysis and machine learning techniques. We also discuss and test methods of dealing with the unique statistical and analytical challenges that arise from working with such a large and complex data set.