

Fuzzy Forest for Microbiome-Driven Diagnosis of Cardiovascular Disease

Eliana Ibrahimi^{1*}, Annamaria Porreca^{2,3}, Alise Ponsero⁴

¹Department of Biology, University of Tirana, Albania

²Department of Human Sciences and Promotion of the Quality of Life, San Raffaele, Rome, Italy

³Unit of Clinical and Molecular Epidemiology IRCCS San Raffaele Roma, Rome, Italy

⁴Quadram Institute of Biosciences, UK

*Correspondence: eliana.ibrahimi@fshn.edu.al

Abstract

The gut microbiome is increasingly recognized as a key modulator in the early pathogenesis of cardiovascular disease such as ischemic heart disease (IHD), offering both diagnostic potential and mechanistic insights before clinical symptom onset. Translating this promise into clinical tools requires advanced computational methods capable of extracting meaningful patterns from complex and noisy microbiome data. In this study, we compare three supervised machine learning algorithms, Lasso-regularized generalized linear models (GLM), Random Forests (RF), and Fuzzy Forests (FF), to classify IHD from healthy controls, using gut microbiome profiles obtained from the Metacardis project (metacardis.net). The dataset included 375 patients with IHD and 275 healthy controls. Preprocessing included normalization, filtering of low-abundance taxa, and stratified train and test splitting. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), with FF outperforming RF and Lasso-GLM in models classifying IHD from healthy subjects. The results from the FF algorithm are consistent with previously reported studies, which emphasize significant alterations in the abundance of various microbial taxa in IHD patients. Taxa such as *Prevotella*, *Bacteroides*, and *Ruminococcus* were among those exhibiting marked differences in abundance between IHD patients and healthy controls, suggesting their potential role in the disease's etiology as previously reported. The ability of FF to identify key microbial taxa contributing to this separation provides valuable insights into potential biomarkers for early diagnosis and therapeutic targets. The superior performance of FF highlights its robustness in handling uncertainty and high dimensionality inherent in microbiome data, making it a promising tool for early IHD prediction and microbiome-related biomarker discovery.

Keywords: gut microbiome, ischaemic heart disease, fuzzy random forest, machine learning, microbial biomarkers, dysbiosis.