

Machine Learning Results

performance measured by the area under the receiver-operator characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC).

Mean AUROC 0.95 (s.d. 0.029)

Mean AUPRC 0.85 (s.d. 0.039)

TODO feature importance

Figure 5

TODO caption

Machine Learning Methods

TODO describe pipeline (1)

mikropml version 1.2.1 (2)]

References

1. **Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD.** 2020. A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. *mBio* **11**. doi:10.1128/mBio.00434-20.
2. **Topçuoğlu BD, Lapp Z, Sovacool KL, Snitkin E, Wiens J, Schloss PD.** 2021. Mikropml: User-Friendly R Package for Supervised Machine Learning Pipelines. *JOSS* **6**:3073. doi:10.21105/joss.03073.