

Key lifestyle factors are predictive of the low gut microbiota diversity status in two cohorts

Alexander Hasson¹, Zachariah Henseler², Abigail J Johnson^{2,3}, Tim Lamberton¹, Tonya Ward², Nashmil Emami⁴, Chieh Jason Chou⁴, Lutz Krause¹, Dan Knights^{2,3}, Shaillay Kumar Dogra⁴

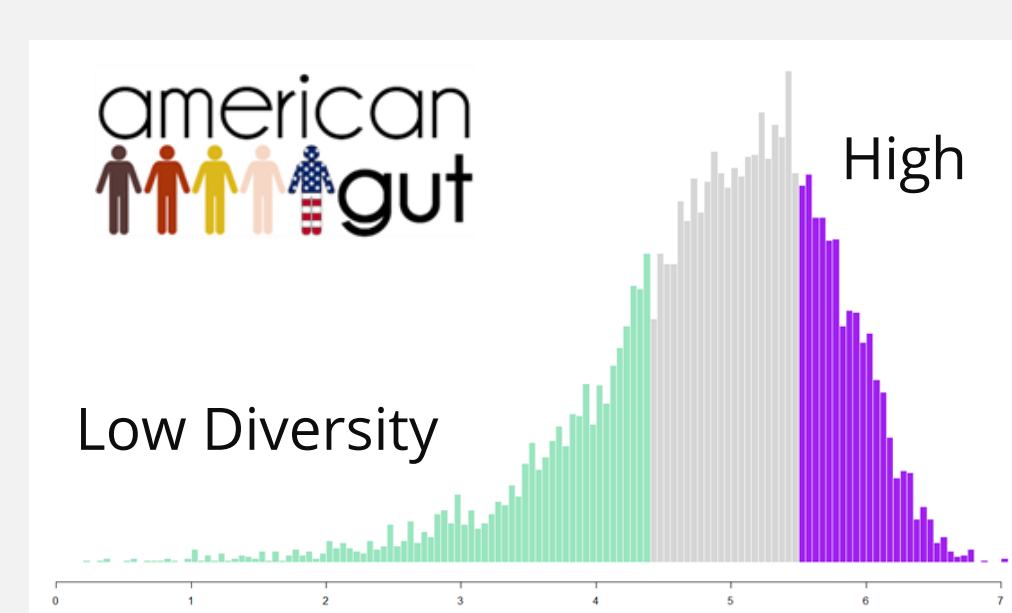
- Background:** Link between gut microbiota diversity and health has been reported, with lower diversity associated with risk of infections and dysbiosis, observed in diseases such as obesity, Type 2 diabetes, and Crohn's disease & higher diversity associated with diet, lifestyle and exercise.
- Methods:** Using Machine Learning based algorithms, we classified participants of American Gut Project (AGP, n=9951) and Microba Discovery Database (MDD, n=6000) into Low (first quartile) and notLow groups, on alpha-diversity indices.
- Results:** We obtained an 'Area Under the Curve' of the 'Receiver Operating Characteristic' (AUC-ROC) value of 0.70 (AGP) and 0.90 (MDD). Questions on age, BMI, and antibiotics usage were predictive in both cohorts (AGP, MDD), while specific measures from 'The Depression, Anxiety and Stress Scale - 21 Items' (DASS21) led to an 8% increase in classification accuracy for MDD.
- Impact:** Our work could inform general population on their gut microbiota diversity status through a questionnaire-based approach (no fecal sampling).

Assessment of Diversity w/o fecal sampling

Discovery and Validation cohorts

American Gut Project (AGP):

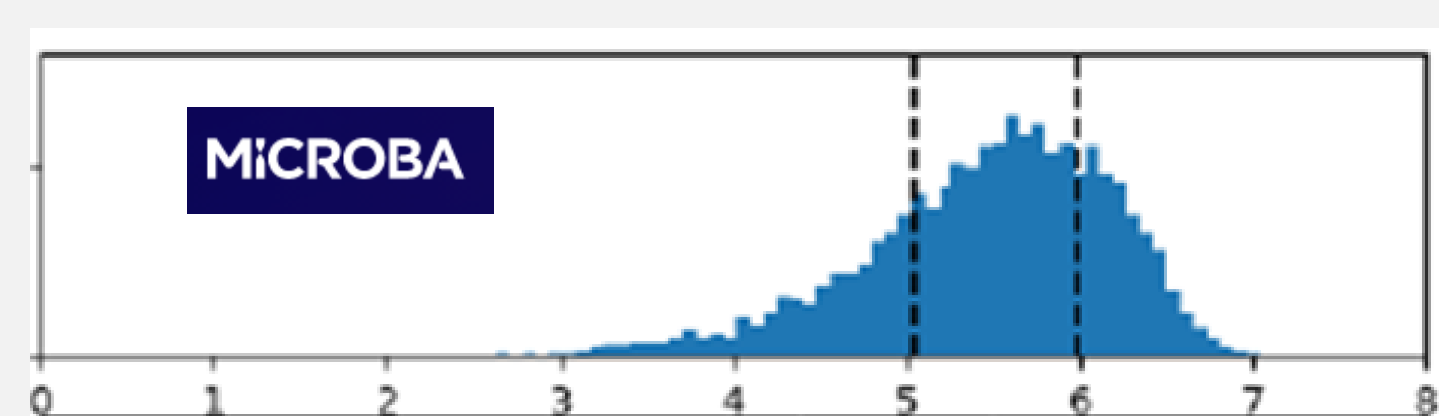
- n=9951, 16s sequencing
- ~200 general questions:
 - Individual traits & Health conditions
 - Lifestyle & High level food habits



Distribution of Shannon diversity (AGP) with lower and upper quartiles shown in color as green and purple, respectively.

Microba Discovery Database (MDD):

- n=6000, shotgun metagenomics
- General Questionnaire
- Medical History



Distribution of Shannon diversity (MDD) with lower and upper quartiles shown as vertical dashed lines

Classification Based Predictive Modeling

Definition of Low and notLow gut microbiota diversity status

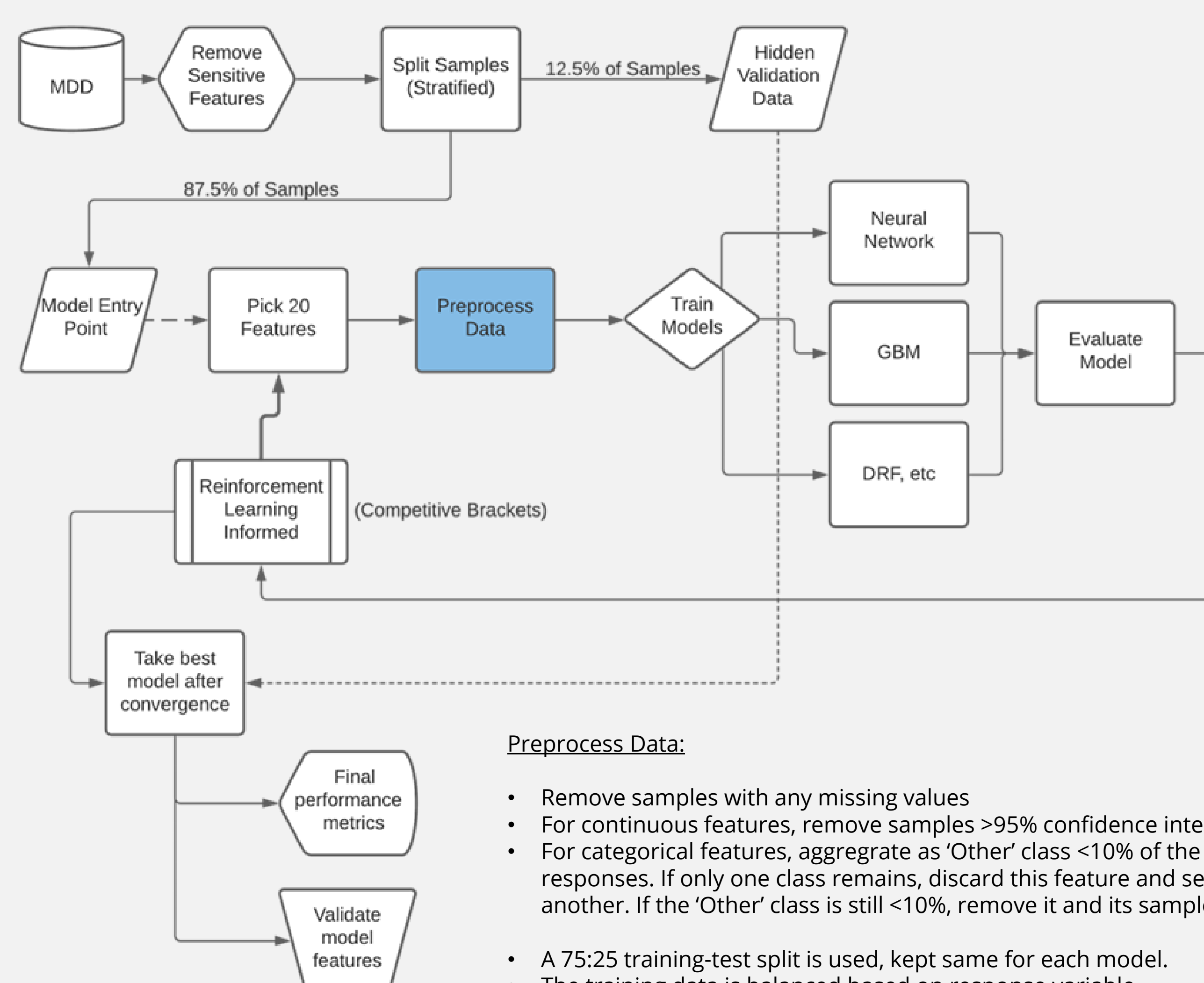
For American Gut Project (AGP):

- Low were subjects in the first quartile of all three measures of alpha-diversity (Shannon Index, Richness, Faith's Phylogenetic Diversity)
- notLow were subjects above first quartile of all these three alpha-diversity measures

For Microba Discovery Database (MDD):

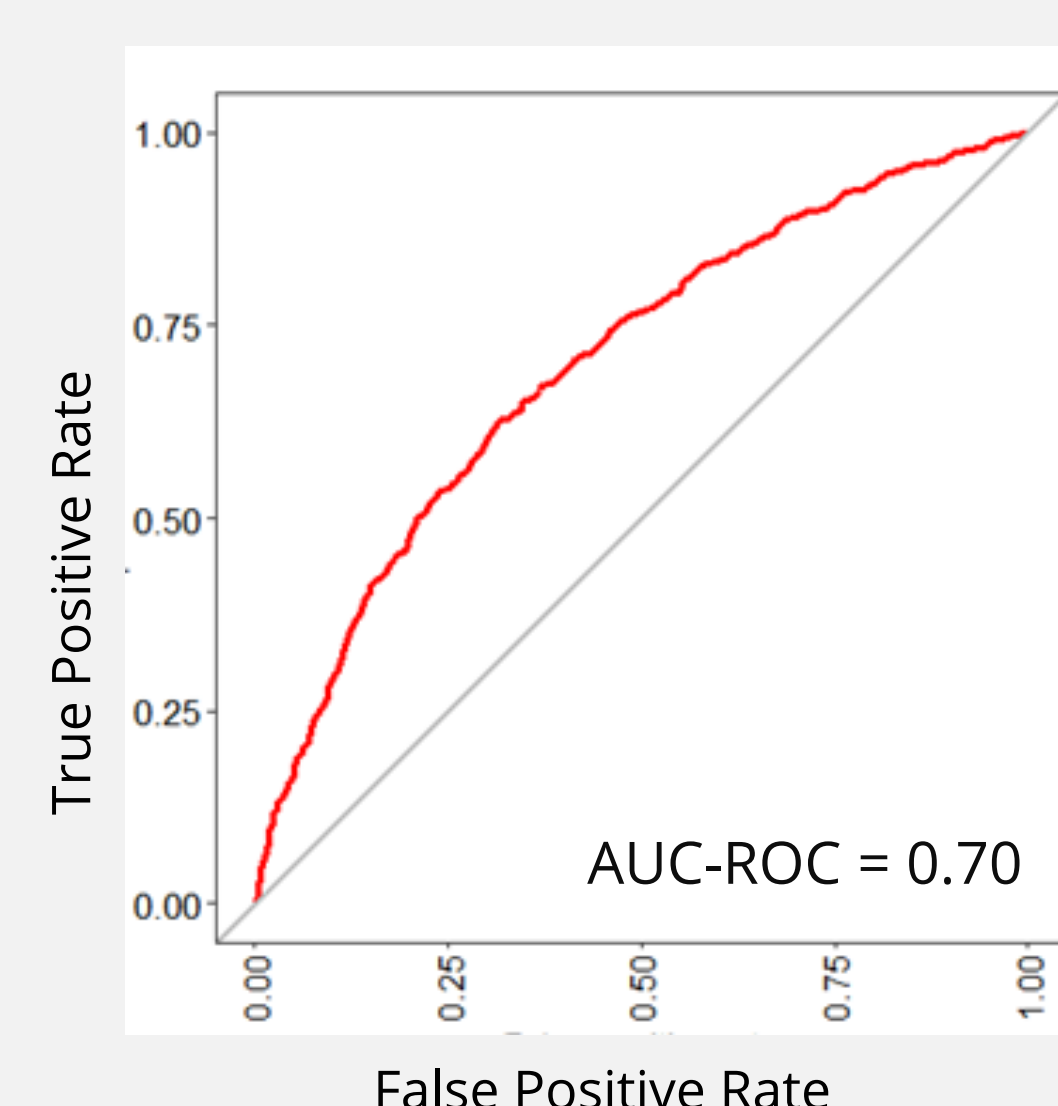
- Low were subjects in first quartile of both measures of alpha-diversity (Shannon Index, Richness)
- notLow were all the remaining subjects

Exemplar workflow to determine best model (in MDD)



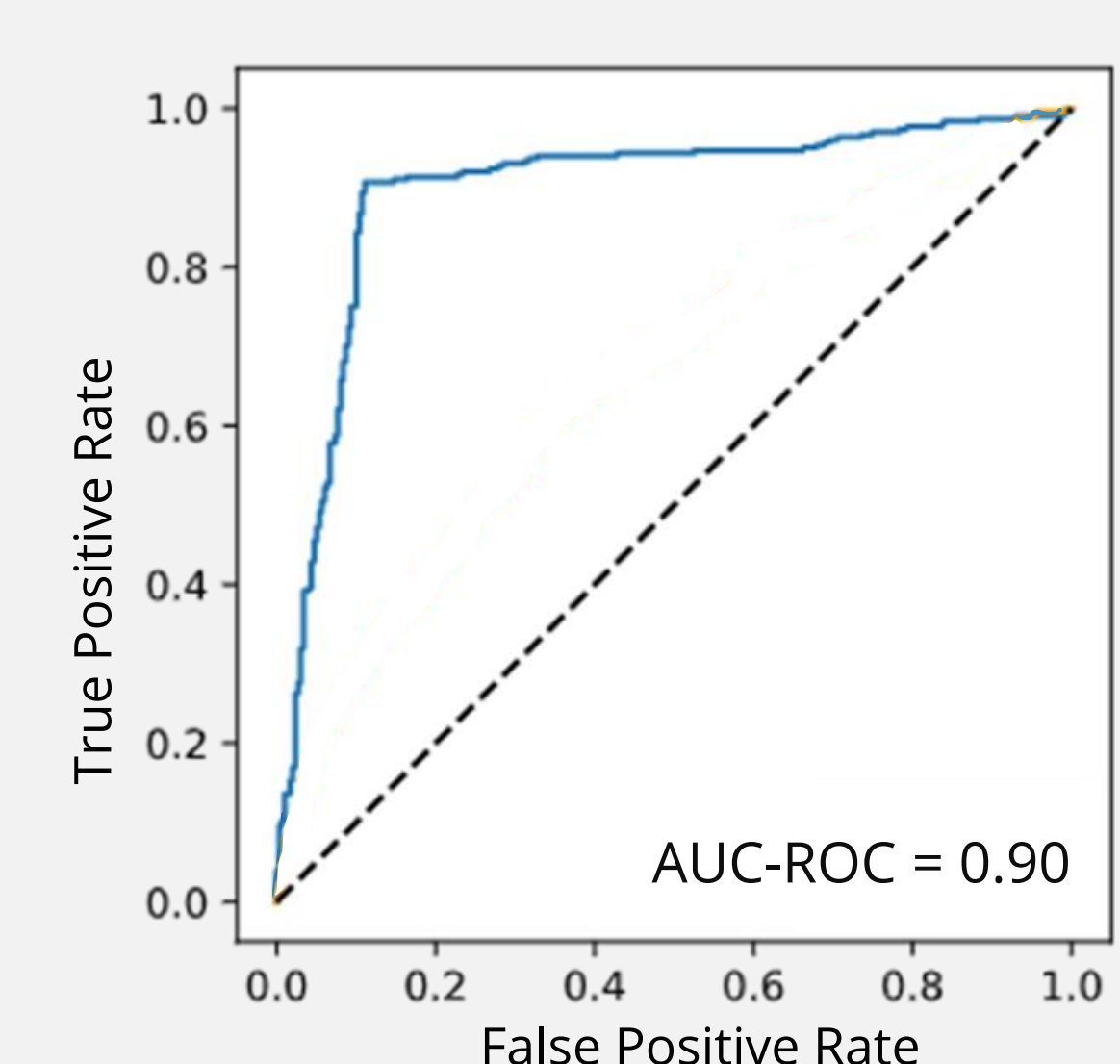
Limited Question Set Predicts Diversity Status

AGP

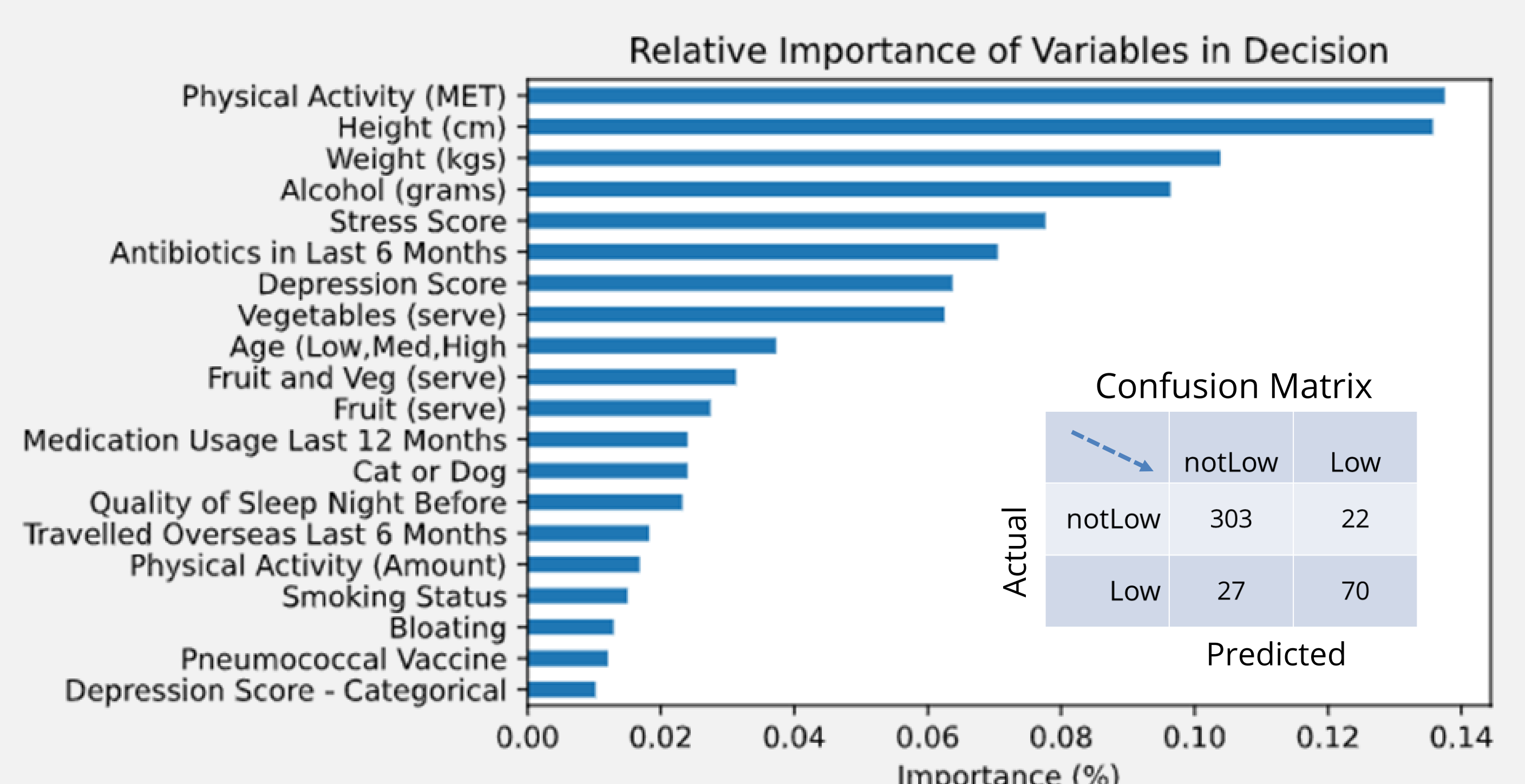


'Receiver Operating Characteristic' (ROC) on American Gut Project (AGP) Train data

MDD



'Receiver Operating Characteristic' (ROC) on Microba Discovery Database (MDD) Train data



Relative importance of questionnaire-based features used by the Low-notLow classification model. Inset – model's confusion matrix (MDD Validation set, n=422, accuracy: Low - 76.09%, notLow - 91.82%)

Key Outcomes

- Gut microbiota diversity status could be assessed using inputs from a general questionnaire via a classification problem to predict low or notLow diversity
- Model features such as age, BMI, physical activity, sleep, pets, vegetable and fruit consumption, alcohol intake, recent antibiotics usage, recent travel abroad were similar in both American Gut Project and Microba Discovery Database
- Some unique measures from 'The Depression, Anxiety and Stress Scale - 21 Items' (DASS21) led to an increased prediction accuracy in Microba Discovery Database

Conclusion

We demonstrated a questionnaire-based approach to assess Low gut microbiota diversity status in two cohorts using Machine Learning based classification models.

References

- AGP- McDonald D, et al. mSystems 2018
- High diversity: Losasso C, et al. Front Microbiol. 2018; Vangay P, et al. Cell. 2018; Clarke SF, et al. Gut. 2014; Estaki M, et al. Microbiome. 2016
- Low diversity and health: Turnbaugh PJ, et al. Nature. 2009; Menni C, et al. Int J Obes (Lond). 2017; Li Q, et al. Sci Rep. 2020, Manichanh C, et al. Gut. 2006; Clooney AG, et al. Gut. 2021; Kampmann C, et al. Clin Microbiol Infect. 2016
- Patents filed: EU - EP20209382A (2020-11-24) & USA - US202163228668P (2021-08-03)