

Infant Gut Microbial Biomarkers With Future Neurodevelopmental disorders

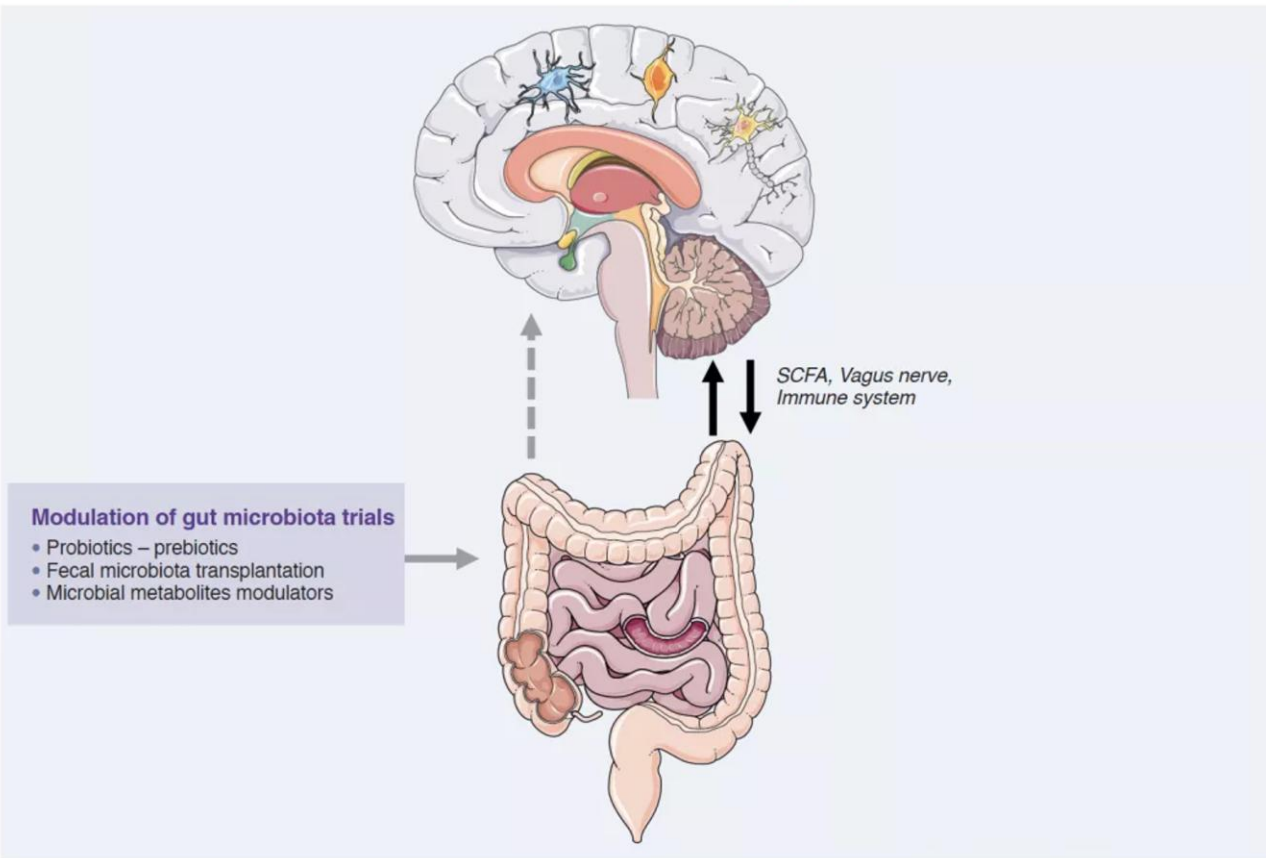
Huaye Zhan ¹

¹ Artificial Intelligence - Software engineer technology, Centennial College

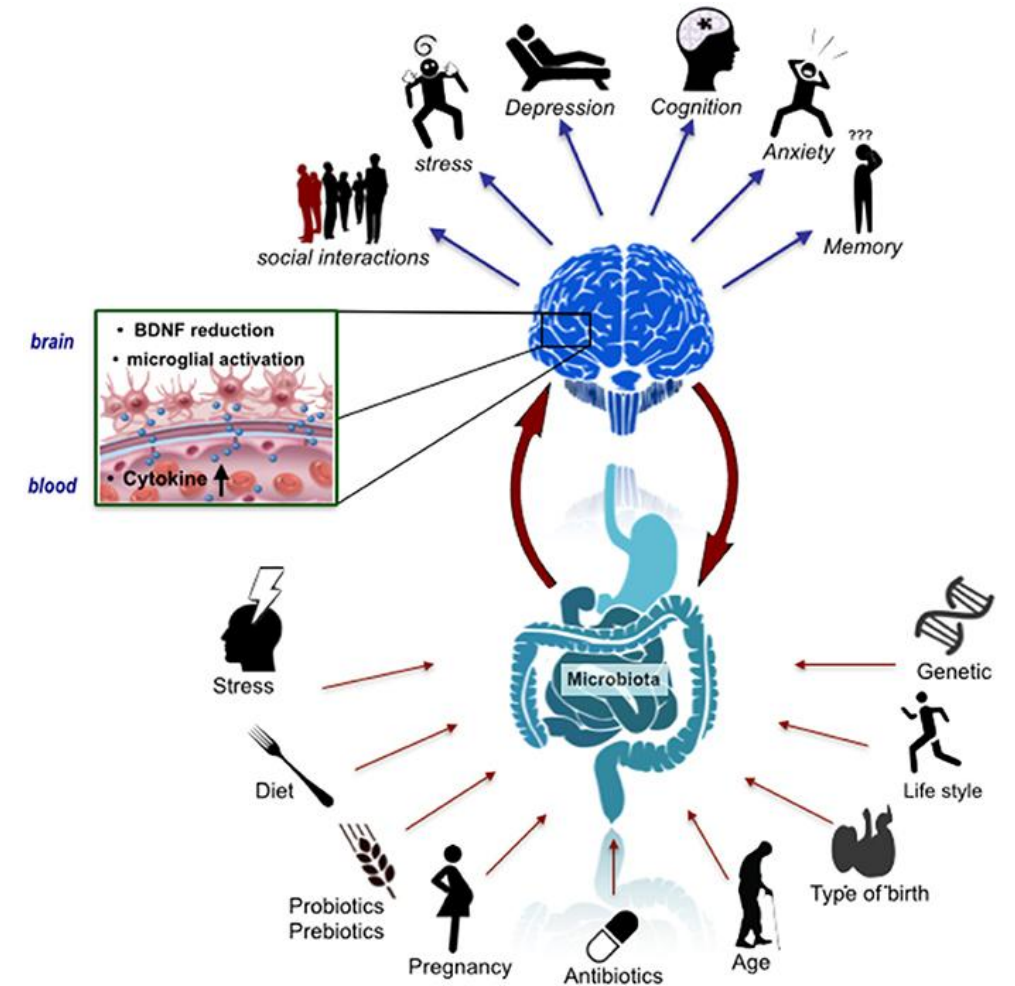
content

1. Introduction
2. Dataset
3. Method and Model
4. Result and discussion
5. Conclusion

1 Introduction



- Gut microbial and brain affection.
- Image from: <https://www.biocodexmicrobiotainstitute.com/en/pro/new-perspectives-autism-role-microbiota-social-communication>



- The Microbiota and Neurodevelopmental Disorders.
- Image from: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2017.00490/full>

1 Introduction

Based on prospective data, can we develop predictive models of infant future NDDs with present gut microbial biomarkers, which can foresee their future NDDs outcomes?

2 dataset

[Explore data](#)[Who we are](#)[What we do](#)[Join us](#)[Help](#)[Login](#)

Data from: Infant microbes and metabolites point to childhood neurodevelopmental disorders

Ahrens, Angelica ¹ ; Hyötyläinen, Tuulia ² ; Petrone, Joseph ¹ ; Igelström, Kajsa ³ ; George, Christian ¹ ; Garrett, Timothy ¹ ; Orešič, Matej ² ; Triplett, Eric ¹ ; Ludvigsson, Johnny ³

[Author affiliations](#)

Published Apr 09, 2024 on Dryad. <https://doi.org/10.5061/dryad.ghx3ffbwj>

Cite this dataset

Ahrens, Angelica et al. (2024). Data from: Infant microbes and metabolites point to childhood neurodevelopmental disorders [Dataset]. Dryad. <https://doi.org/10.5061/dryad.ghx3ffbwj>

Abstract

This study has followed a birth cohort for over twenty years to find factors associated with neurodevelopmental disorder (ND) diagnosis. Detailed, early-life longitudinal questionnaires captured infection and antibiotic events, stress, prenatal factors, family history, and more. Biomarkers including cord serum metabolome and lipidome, HLA genotype, infant microbiota, and stool metabolome were assessed. Among the 16,440 Swedish children followed across time, 1,197 developed an ND. Significant associations emerged for future ND diagnosis in general and for specific ND subtypes, spanning intellectual disability,

Data files

> Apr 09, 2024 version files

19.10 MB



Download full dataset

Related works

Primary article

<https://doi.org/10.10...6/j.cell.2024.02.035>

Share



Desktop >

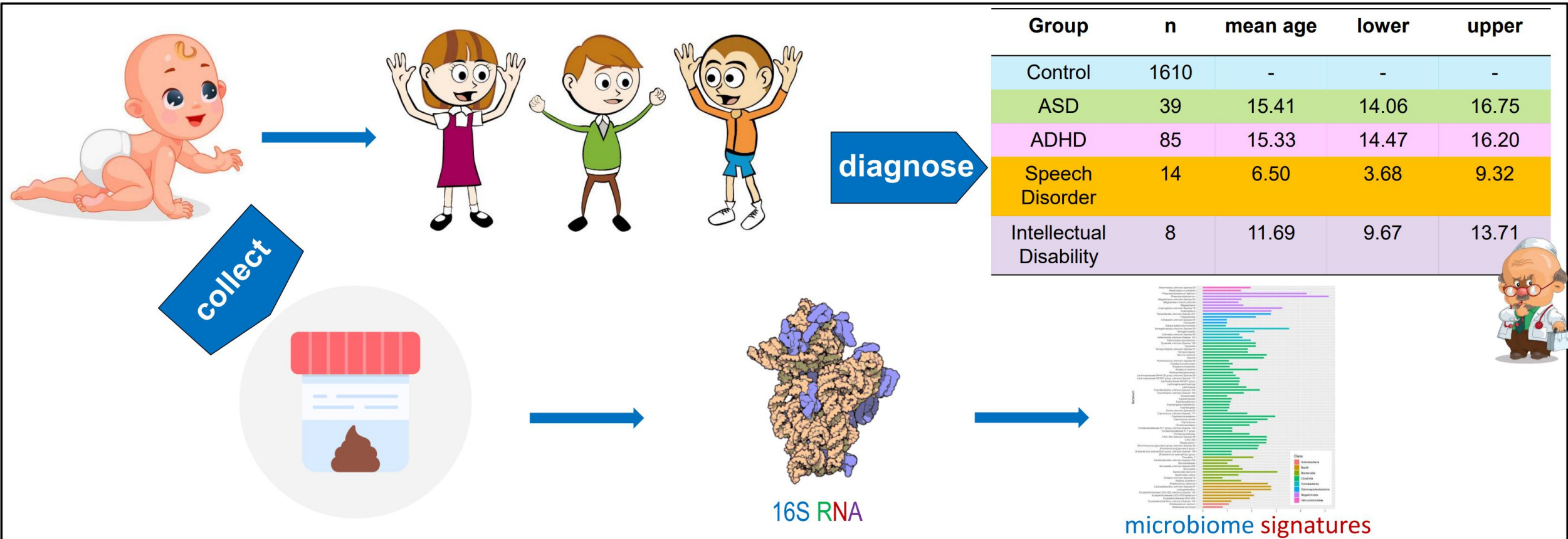
doi_10_5061_dryad_ghx3ffbwj_v20240409

Search doi_10_5061_dryad_ghx3ffbwj_...

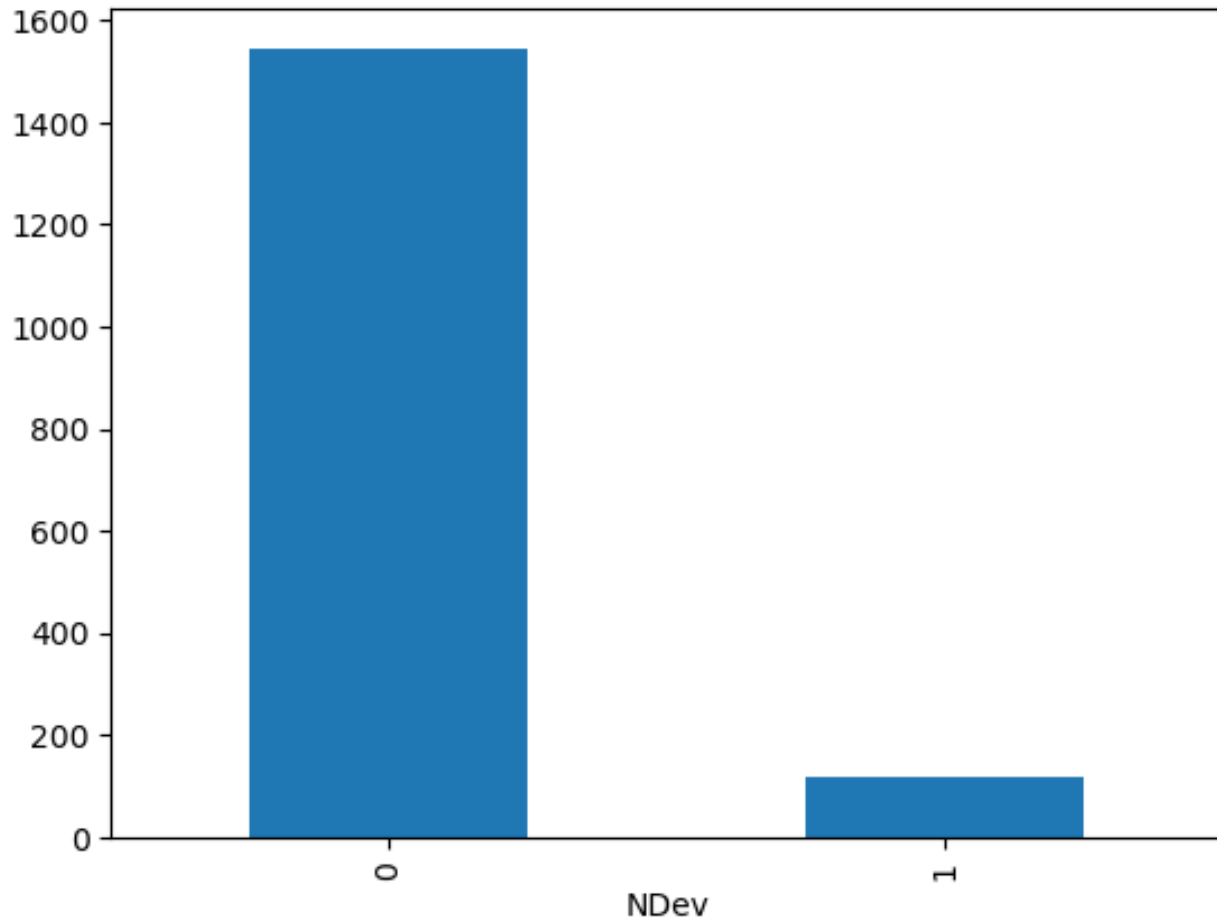
Metrics

Untargeted_Positive_Dryad.csv Microsoft Excel Comma Separate... 145 KB	Untargeted_Negative_Dryad.csv Microsoft Excel Comma Separate... 162 KB	TopGen_StoolNegativeMode_Dryad... Microsoft Excel Comma Separate... 1.97 MB	Sym_Data_Dryad.csv Microsoft Excel Comma Separate... 13.8 KB
StoolMetabolites_SCFAs_Dryad.csv Microsoft Excel Comma Separate... 2.27 KB	StoolMetabolites_Fattyacids_Dryad... Microsoft Excel Comma Separate... 487 KB	SpeechDis_DRYAD_v3.csv Microsoft Excel Comma Separate... 487 KB	PoslonStool_CodeBook_Dryad.csv Microsoft Excel Comma Separate... 13.8 KB
NeglonStool_CodeBook_Dryad.csv Microsoft Excel Comma Separate... 15.1 KB	ND_DRYAD_v2.csv Microsoft Excel Comma Separate... 2.14 MB	Microbiome_NDOutcomes_Dryad_s... Microsoft Excel Comma Separate... 1.24 KB	Microbiome_ASD_earlylatediagnosi... Microsoft Excel Comma Separate... 1.75 MB
IntDis_DRYAD_v3.csv Microsoft Excel Comma Separate... 1.43 MB	Identification_CordBloodMetabolite... Microsoft Excel Comma Separate... 42.3 KB	Factors_Data_Dryad.csv Microsoft Excel Comma Separate... 2.06 MB	equol_Dryad.csv Microsoft Excel Comma Separate... 1.24 KB
CordBlood_ProMatch_Metadata_Dr... Microsoft Excel Comma Separate... 48.9 KB	CordBlood_Microbiome_DRYAD.csv Microsoft Excel Comma Separate... 42.3 KB	ASD_DRYAD_v2.csv Microsoft Excel Comma Separate... 1.90 MB	ADHD_DRYAD_v2.csv Microsoft Excel Comma Separate... 1.75 MB
README.md MD File 48.9 KB			

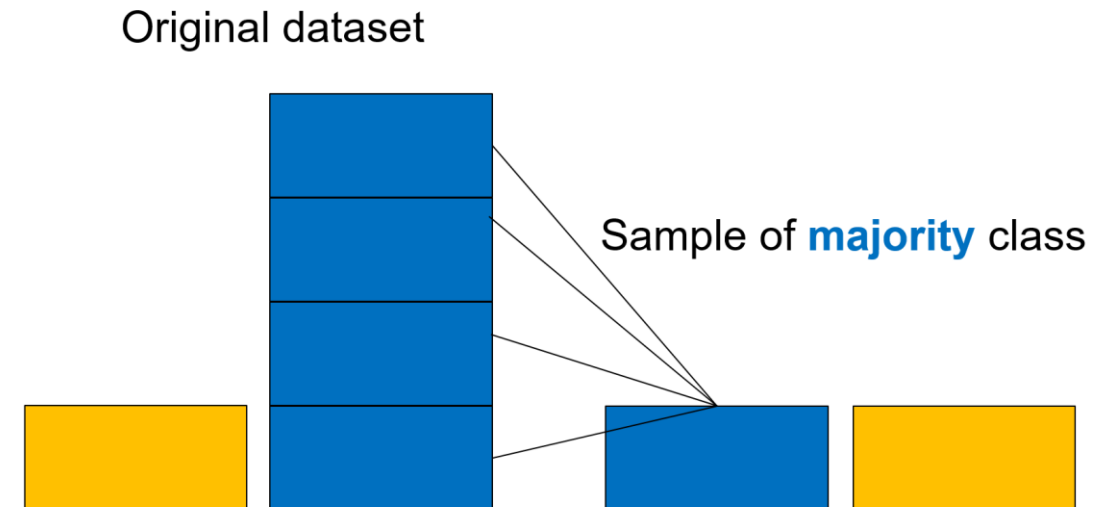
2 dataset



3 Method and model

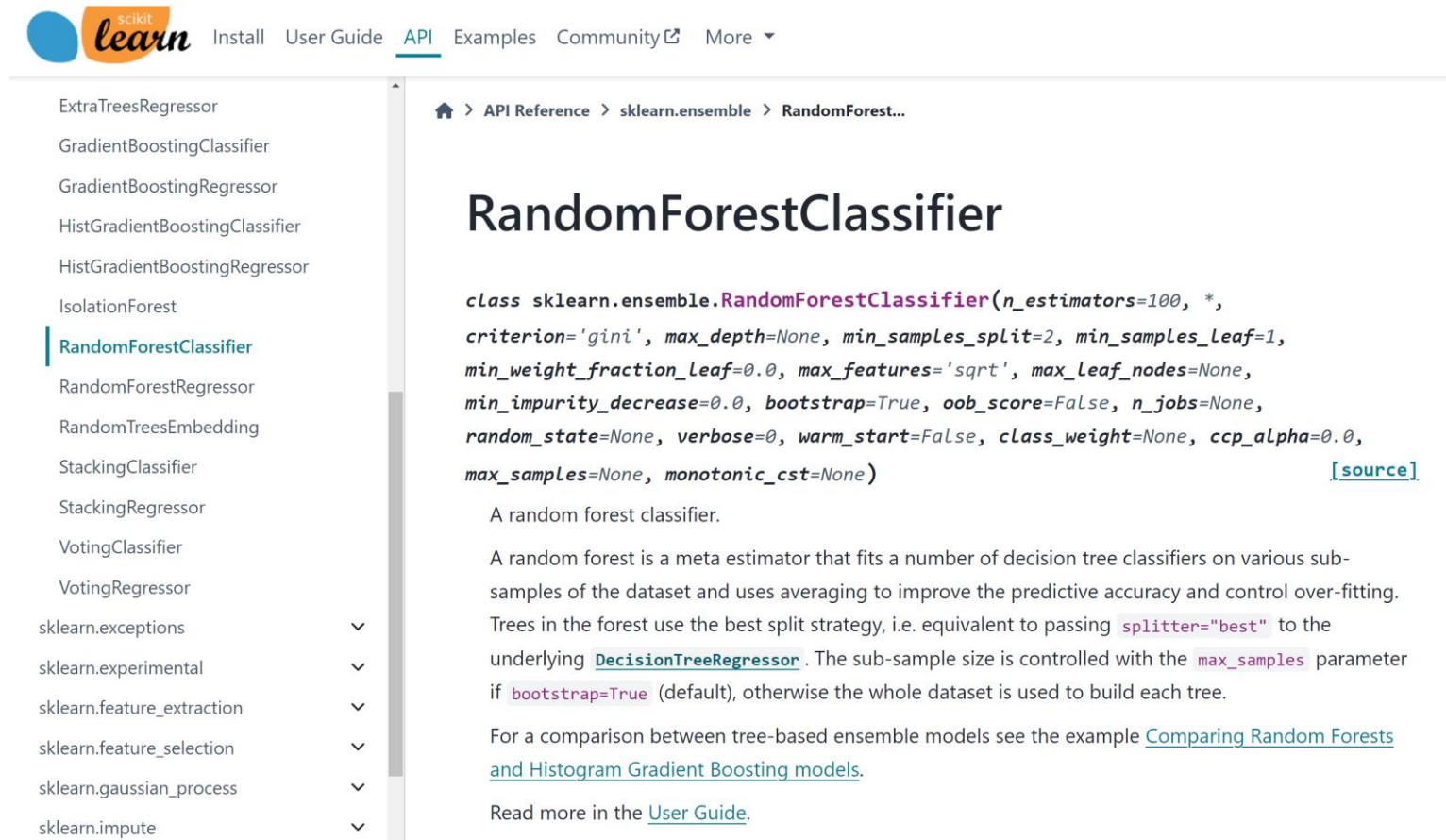


Imbalanced Target Variable



Undersampling

3 Method and model



The screenshot shows the scikit-learn website's API reference for the `RandomForestClassifier` class. The left sidebar lists various models, with `RandomForestClassifier` highlighted. The main content area shows the class name, its full signature with default parameters, a brief description, and a detailed explanation of how it works as a meta-estimator.

scikit-learn [Install](#) [User Guide](#) [API](#) [Examples](#) [Community](#) [More](#)

[API Reference](#) > [sklearn.ensemble](#) > [RandomForest...](#)

RandomForestClassifier

```
class sklearn.ensemble.RandomForestClassifier(n_estimators=100, *,
      criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1,
      min_weight_fraction_leaf=0.0, max_features='sqrt', max_leaf_nodes=None,
      min_impurity_decrease=0.0, bootstrap=True, oob_score=False, n_jobs=None,
      random_state=None, verbose=0, warm_start=False, class_weight=None, ccp_alpha=0.0,
      max_samples=None, monotonic_cst=None)
```

[\[source\]](#)

A random forest classifier.

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Trees in the forest use the best split strategy, i.e. equivalent to passing `splitter="best"` to the underlying [DecisionTreeRegressor](#). The sub-sample size is controlled with the `max_samples` parameter if `bootstrap=True` (default), otherwise the whole dataset is used to build each tree.

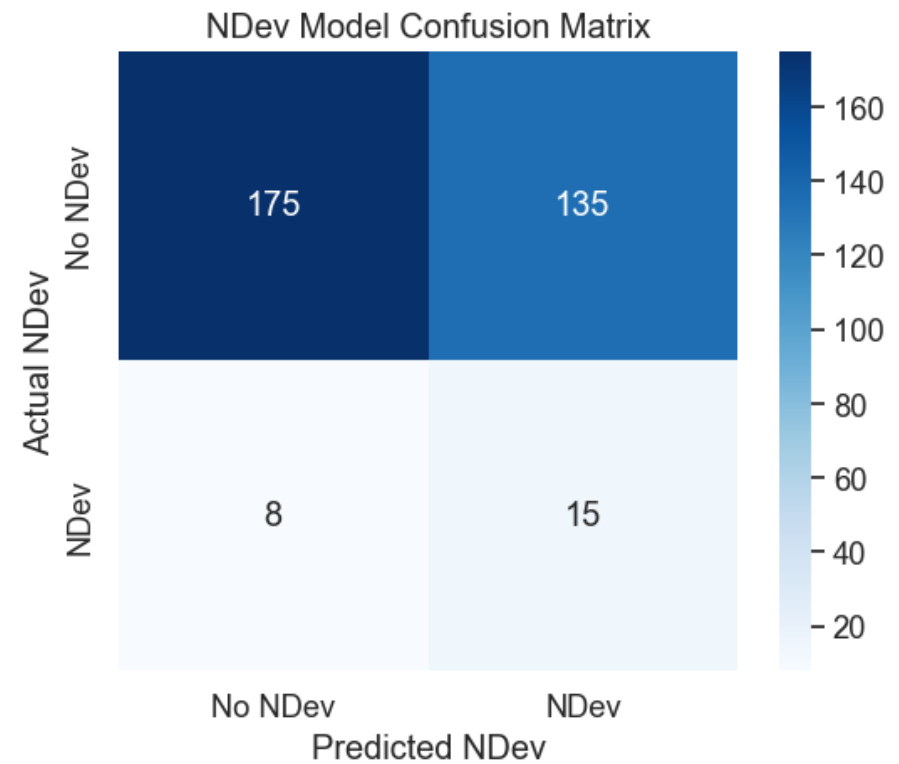
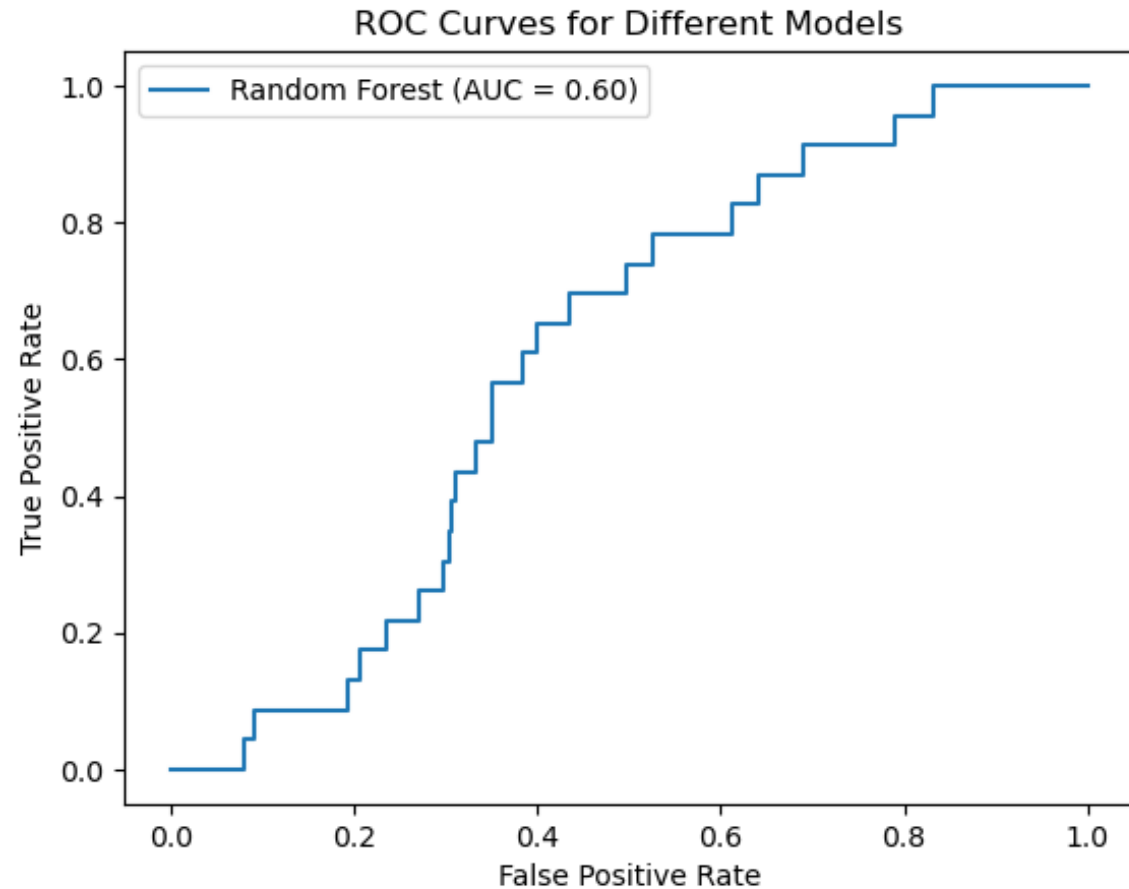
For a comparison between tree-based ensemble models see the example [Comparing Random Forests and Histogram Gradient Boosting models](#).

Read more in the [User Guide](#).

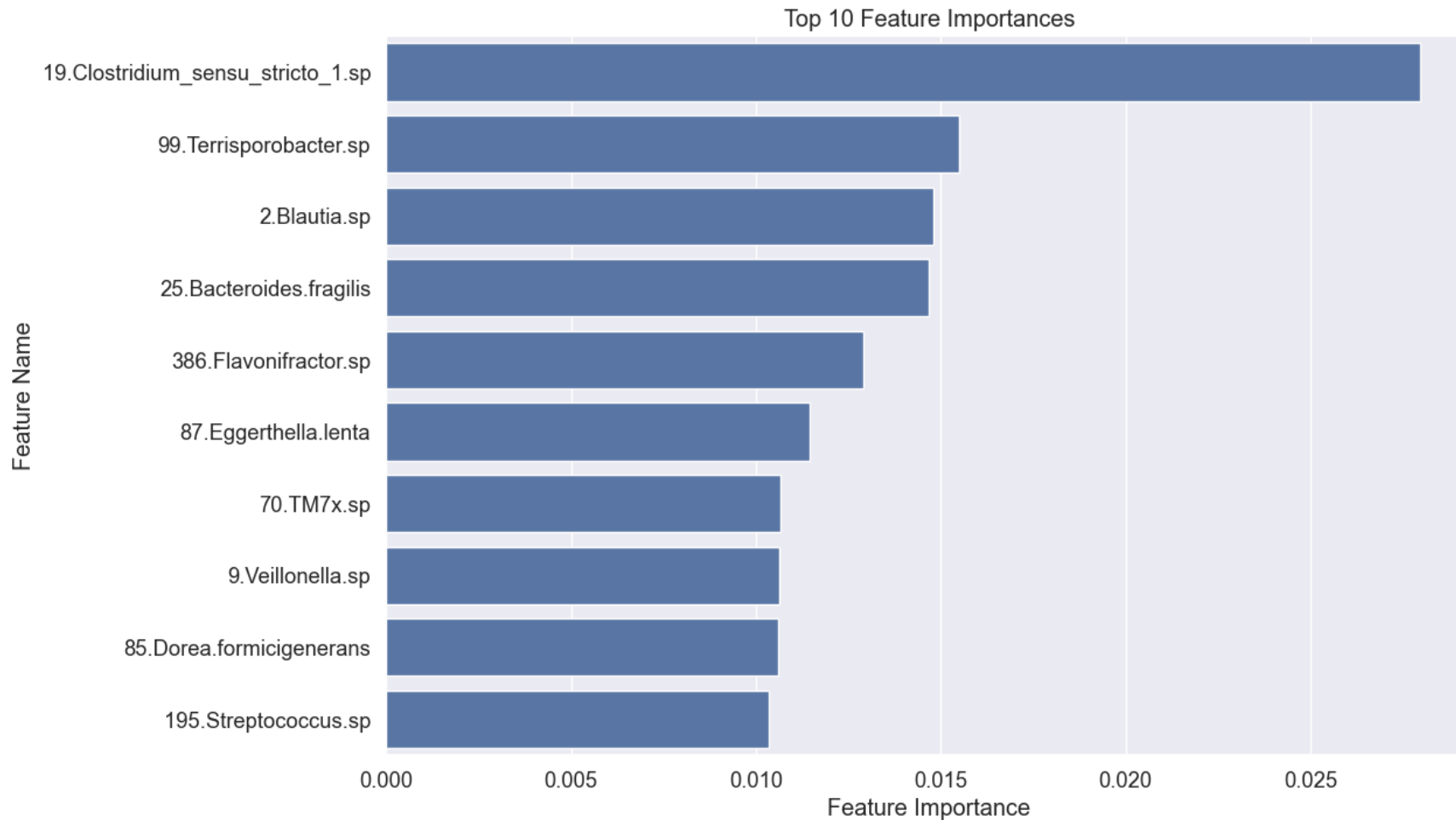
RandomizedSearchCV(RandomForestClassifier(), params)

RandomForestClassifier(n_estimators=100, max_depth=8, random_state=2024)

4 Result and discussion



4 Result and discussion



5 Conclusion: limitation

- Limitations of this study include few samples of future NDDs
- 60% accuracy, 65% recall
- Many other factors in the prospective data

Thank you!