Infant Gut Microbial Biomarkers With Future Neurodevelopmental disorders

Huaye Zhan [1]

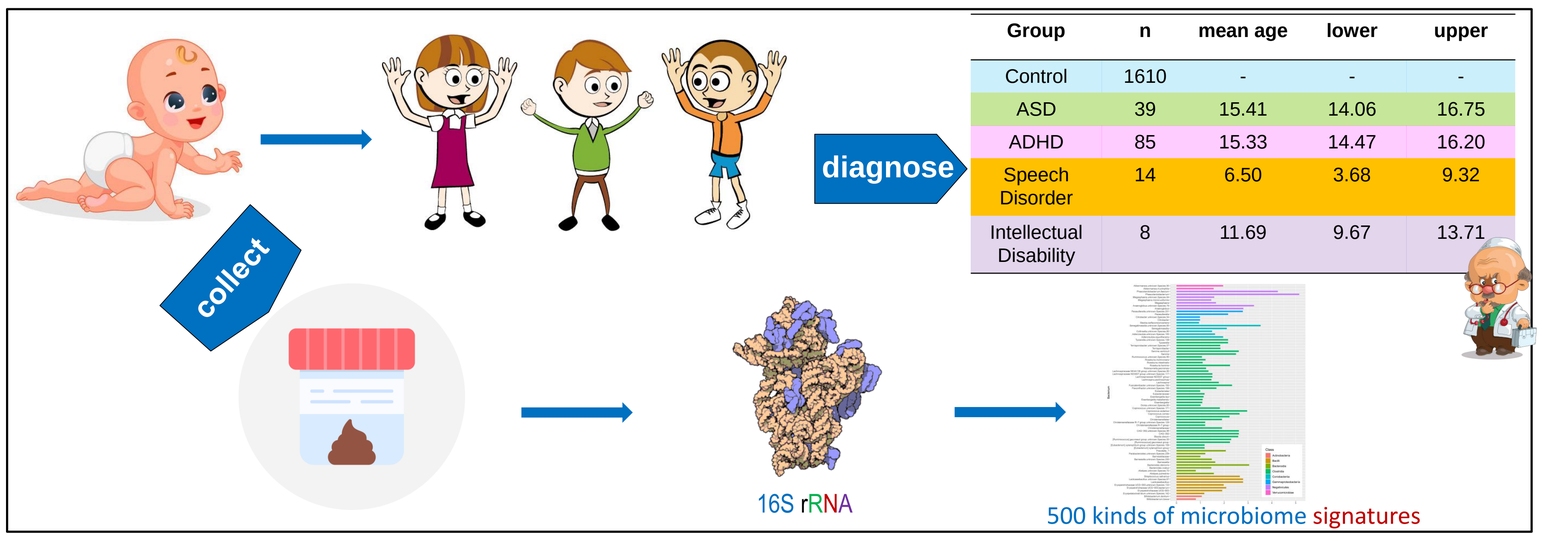
[1] Artificial Intelligence - Software engineer technology, Centennial College

Background

- Neurodevelopmental disorders (NDDs) affect brain and nervous system development.
- Alternations in infant gut microbiome biomarkers would be associated with their future NDDs.
- However, prospective early life data of infant gut microbial biomarkers are limited.
- Based on prospective data, I develop predictive models of infant future NDDs with present gut microbial biomarkers, which can foresee their future NDDs outcomes.

Method & Result

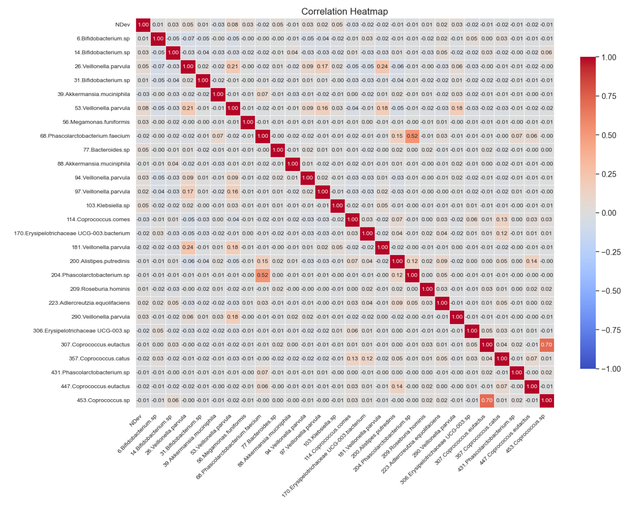
Dataset



Group	n	mean age	lower	upper
Control	1610	-	-	-
ASD	39	15.41	14.06	16.75
ADHD	85	15.33	14.47	16.20
Speech Disorder	14	6.50	3.68	9.32
Intellectual Disability	8	11.69	9.67	13.71

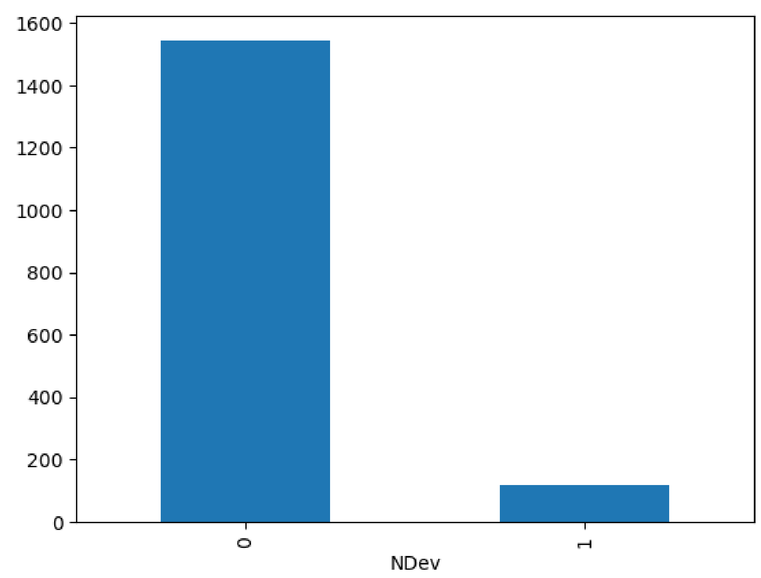
- Stool was collected from **1748** infants at **11.9 ± 2.9** months.
- The result of diagnose happened **a few years later**, when infants grew up.
- Our **longitudinal prospective** studies on **early-life microbiomes** and **future ND diagnoses**.

EDA: Correlation Matrix

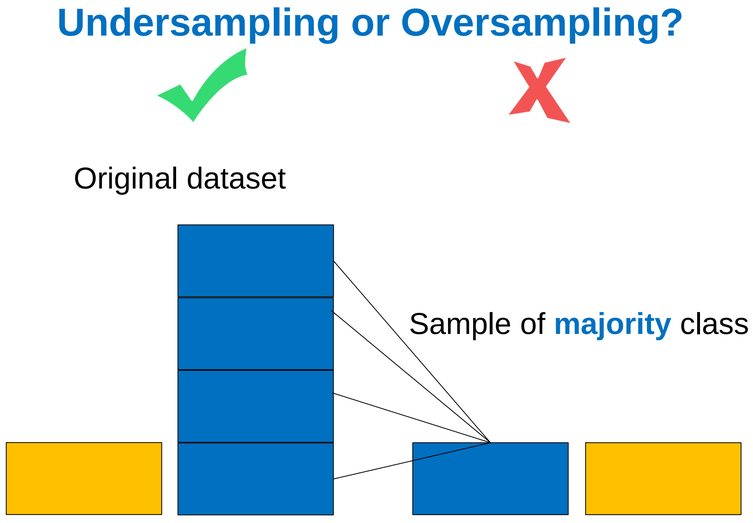


Source from dataset: Bacteria enriched in controls included ***Akkermansia muciniphila***, ***Roseburia hominis***, ***Erysipelotrichiaceae UCG-003*** **spp.**, ***Adlercreutzia equolifaciens***, ***Alistipes putredinis***, ***Phascolarctobacterium***, ***Coprococcus***, **and** ***Bifidobacterium*** **sp.** ***Veillonella parvula***, ***Megamonas funiformis***, **ASV-77** ***Bacteroides*** **sp., and ASV-103** ***Klebsiella*** **sp.** were higher in future cases.

Imbalanced Target Variable

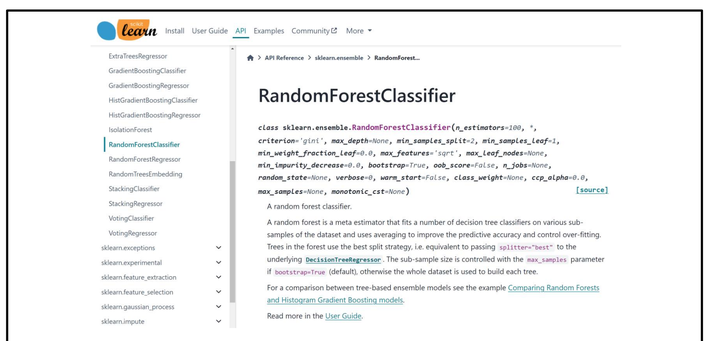


Deal with imbalanced data method: Undersampling or Oversampling?



Original Positive Class : 7%
Undersampling Positive Class : 49.2%

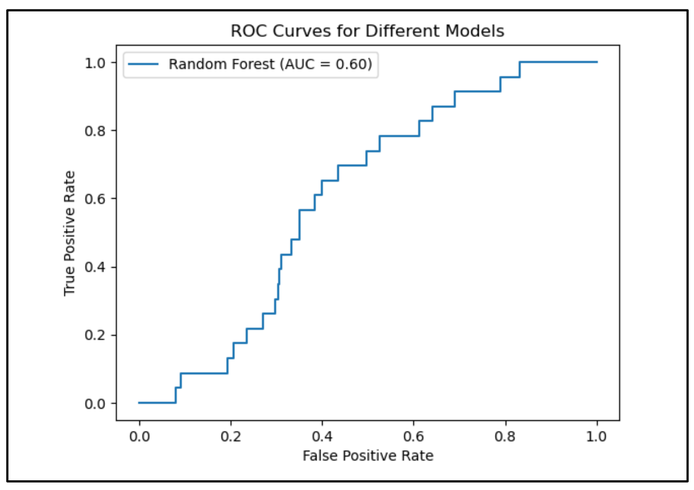
Model: Random Forest



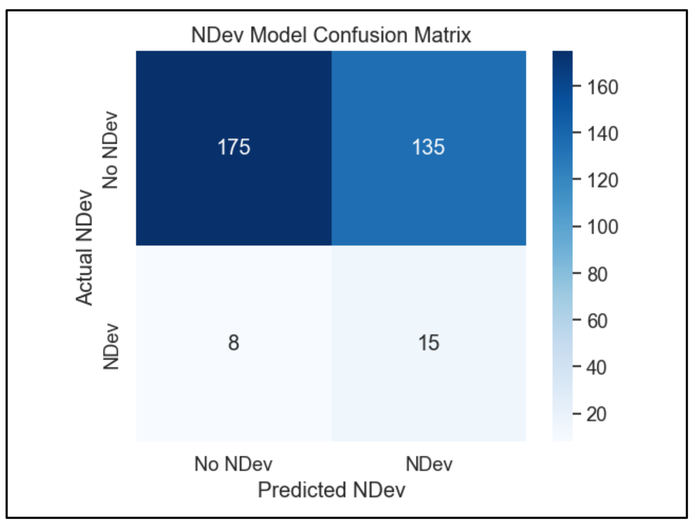
RandomizedSearchCV(RandomForestClassifier(), params)

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

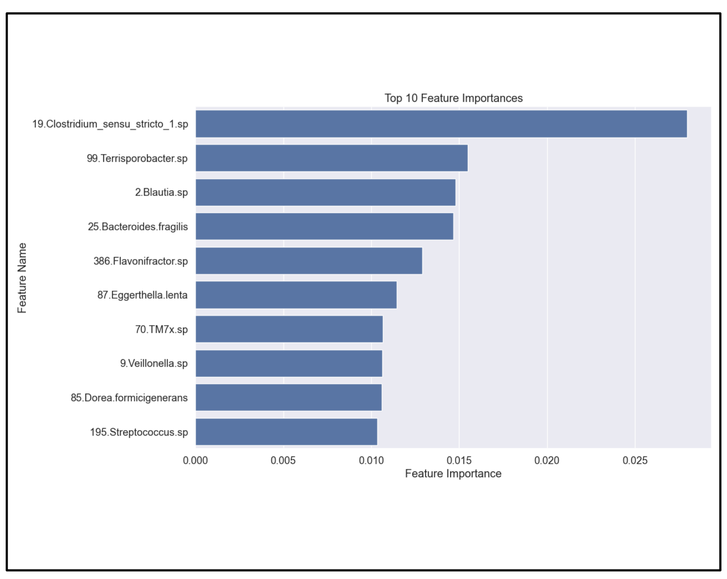
ROC AUC Curve



Confusion Matrix



Feature Importance



Conclusion

- Limitations of this study include few samples of future NDDs.
- Further studies focus on improving the model prediction and predicting specific future NDDs subtypes(ASD, Intellectual Disability, Speech Disorder, or ADHD) in the larger cases datasets.