



Gut Microbiome

Probiotics Strike Back

Naif Ganadily, Martha Kaiser,
Tatum Dykstra, Asiful Arefeen

Gut Microbiome

- Diverse microorganisms found within the gastrointestinal tract
- Influence the digestion of nutrients and health of individuals
- Impacted by many factors including diet, exercise, sleep patterns, and medications¹⁻³

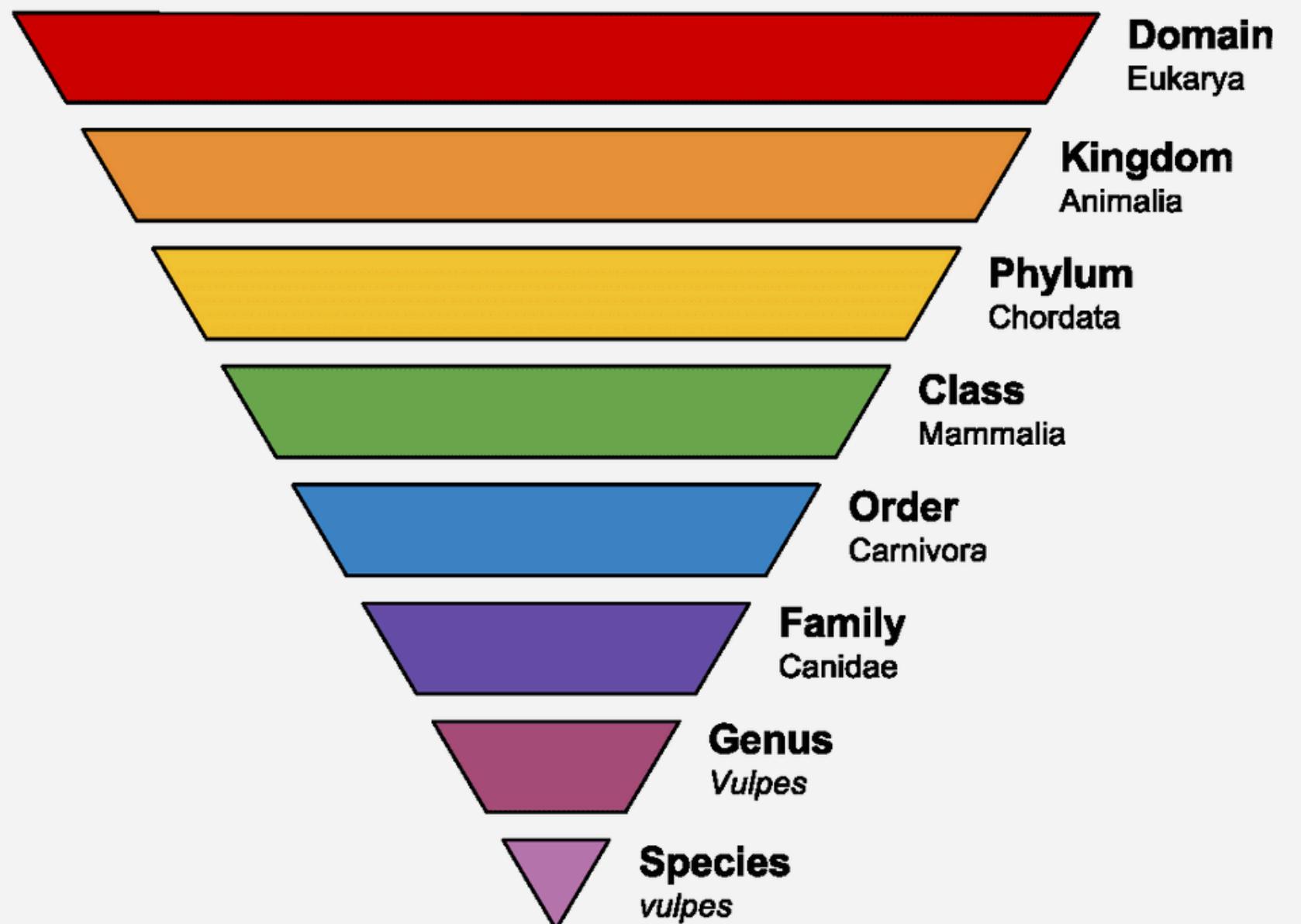


Statement of Work

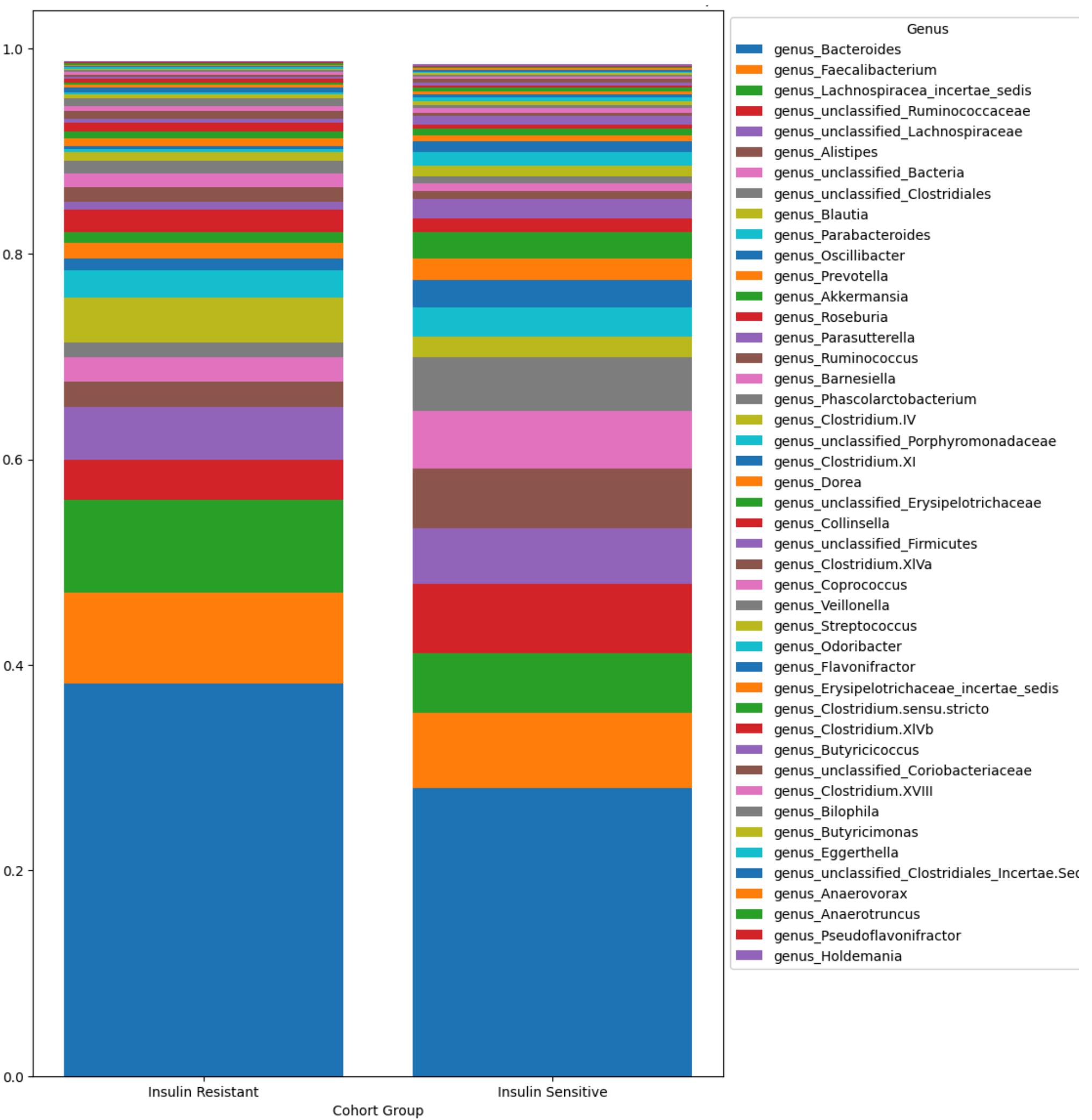
Prediabetic individuals may have insulin resistance before the development of type 2 diabetes. Considering this, we aim to understand the differences in the gut microbiome composition between those with insulin sensitivity and those with insulin resistance to inform potential nutritional applications. In our analysis, we built machine and deep learning models to model IR and IS based on gut microbiome composition.

```
phylum_Actinobacteria
└ class_Actinobacteria
  └ order_Coriobacterales
    └ family_Coriobacteriaceae
      └ genus_Collinsella
phylum_Bacteroidetes
└ class_Bacteroidia
  └ order_Bacteroidales
    └ family_Bacteroidaceae
      └ genus_Bacteroides
phylum_Firmicutes
└ class_Bacilli
  └ order_Lactobacillales
└ class_Clostridia
  └ order_Clostridiales
    └ family_Clostridiaceae
      └ genus_Clostridium
    └ family_Lachnospiraceae
      └ genus_Roseburia
phylum_Proteobacteria
└ class_Betaproteobacteria
  └ order_Burkholderiales
└ class_Gammaproteobacteria
  └ order_Enterobacteriales
    └ family_Enterobacteriaceae
      └ genus_Escherichia
phylum_Verrucomicrobia
└ class_Verrucomicrobiae
  └ order_Verrucomicrobiales
    └ family_Verrucomicrobiaceae
      └ genus_Akkermansia
phylum_unclassified_Bacteria
```

Taxonomic Classification



Mean Relative Abundance of Genera in Insulin and Insulin Sensitive Groups



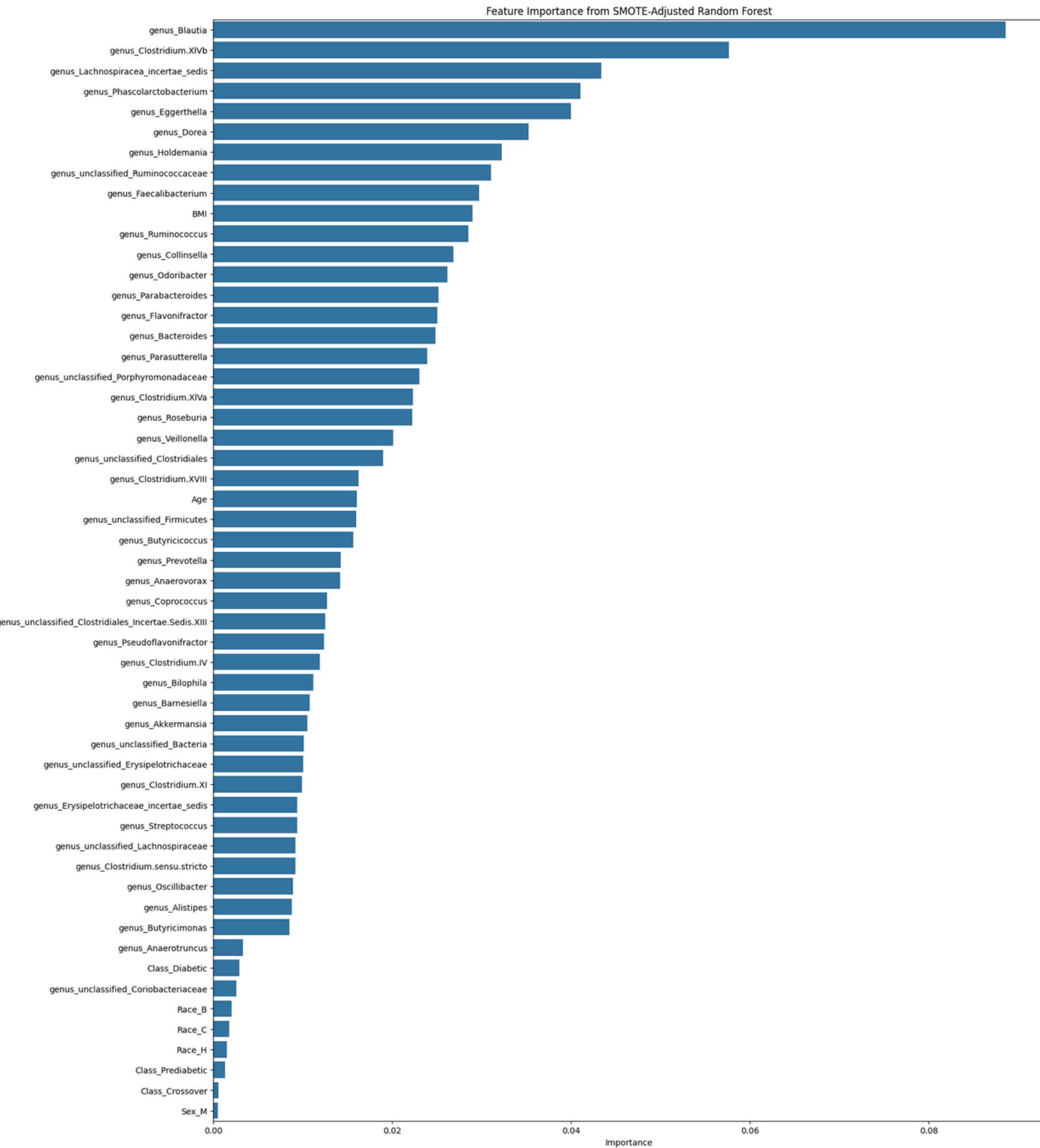
Preprocessing

- Merged data
- Removed subjects without data for IR and IS
- One-hot encoding of the categorical variables
- CLR
 - Commonly used in microbiome research to account for compositionality of data
- Sparse PCA
- Standard scaler
- Dropped SSPG
- Limited sample to genus taxonomic classification

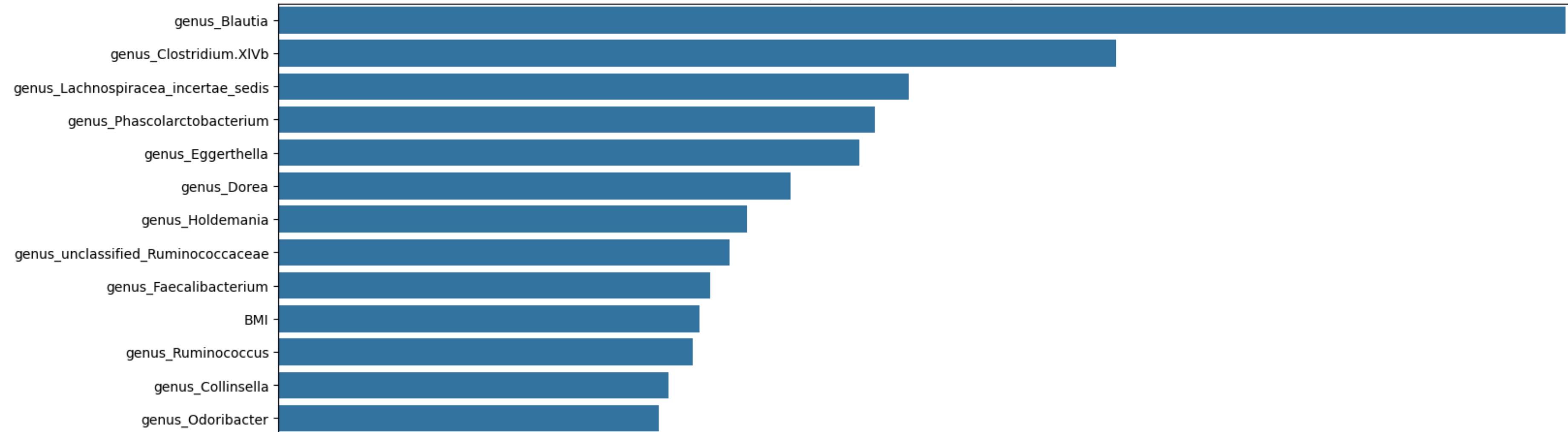
Random Forest Model

Classification Report with SMOTE:

	precision	recall	f1-score	support
0.0	0.71	0.83	0.77	6
1.0	0.80	0.67	0.73	6
accuracy			0.75	12
macro avg	0.76	0.75	0.75	12
weighted avg	0.76	0.75	0.75	12



Feature Importance from SMOTE-Adjusted Random Forest



SVM Model

```
param_grid = {
    'C': [0.01, 0.1, 1, 1.5, 10, 100],
    'gamma': [1, 0.1, 0.01, 0.001],
    'kernel': ['rbf', 'poly', 'sigmoid', 'linear'],
    'class_weight': ['balanced', None]
}

[CV] END C=100, class_weight=None, gamma=0.001, kernel='linear'; total
{'C': 0.1, 'class_weight': None, 'gamma': 0.1, 'kernel': 'sigmoid'}
```

Results

LR Mean train score: 1.0
LR Mean test score: 0.7136363636363636

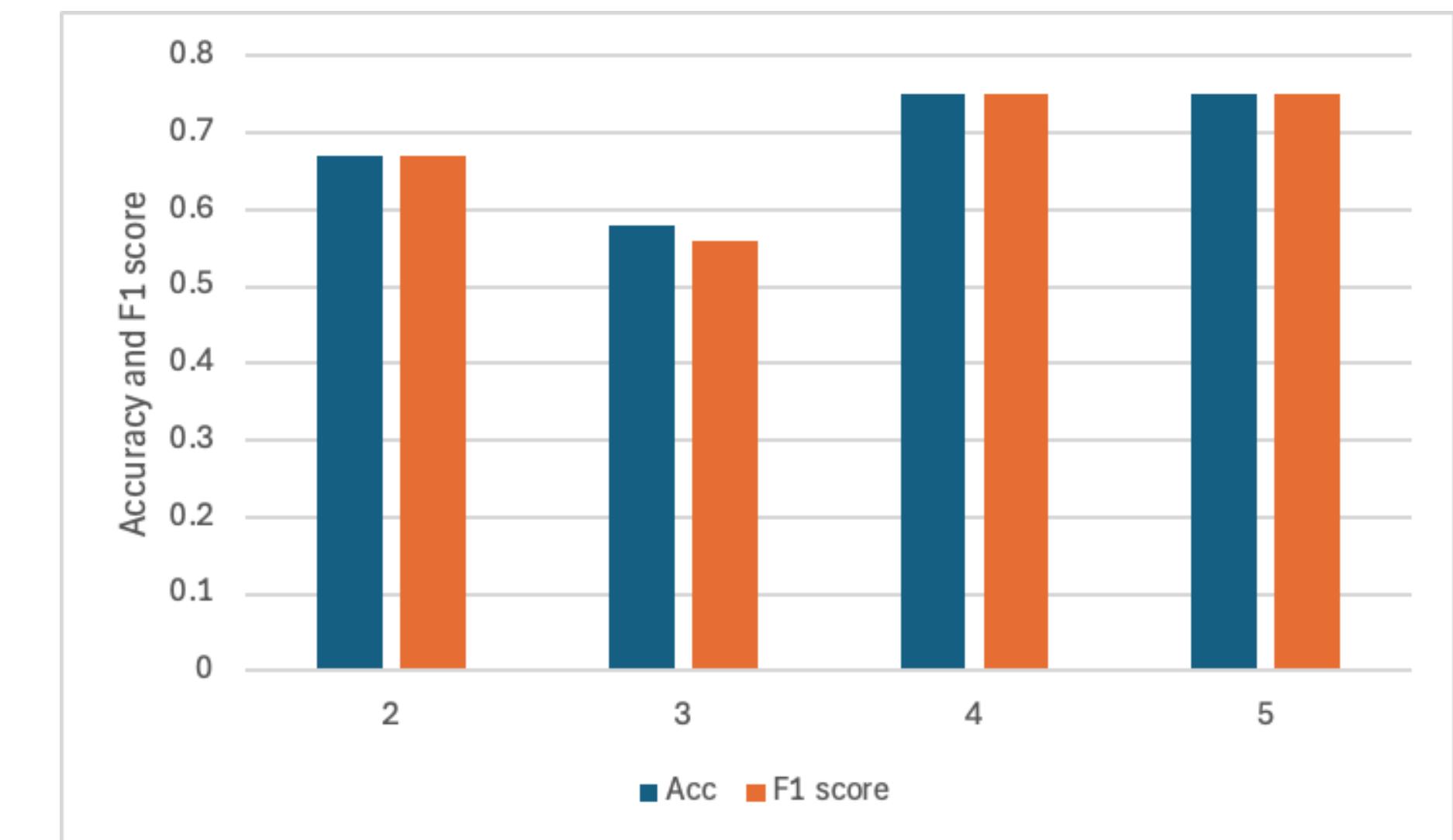
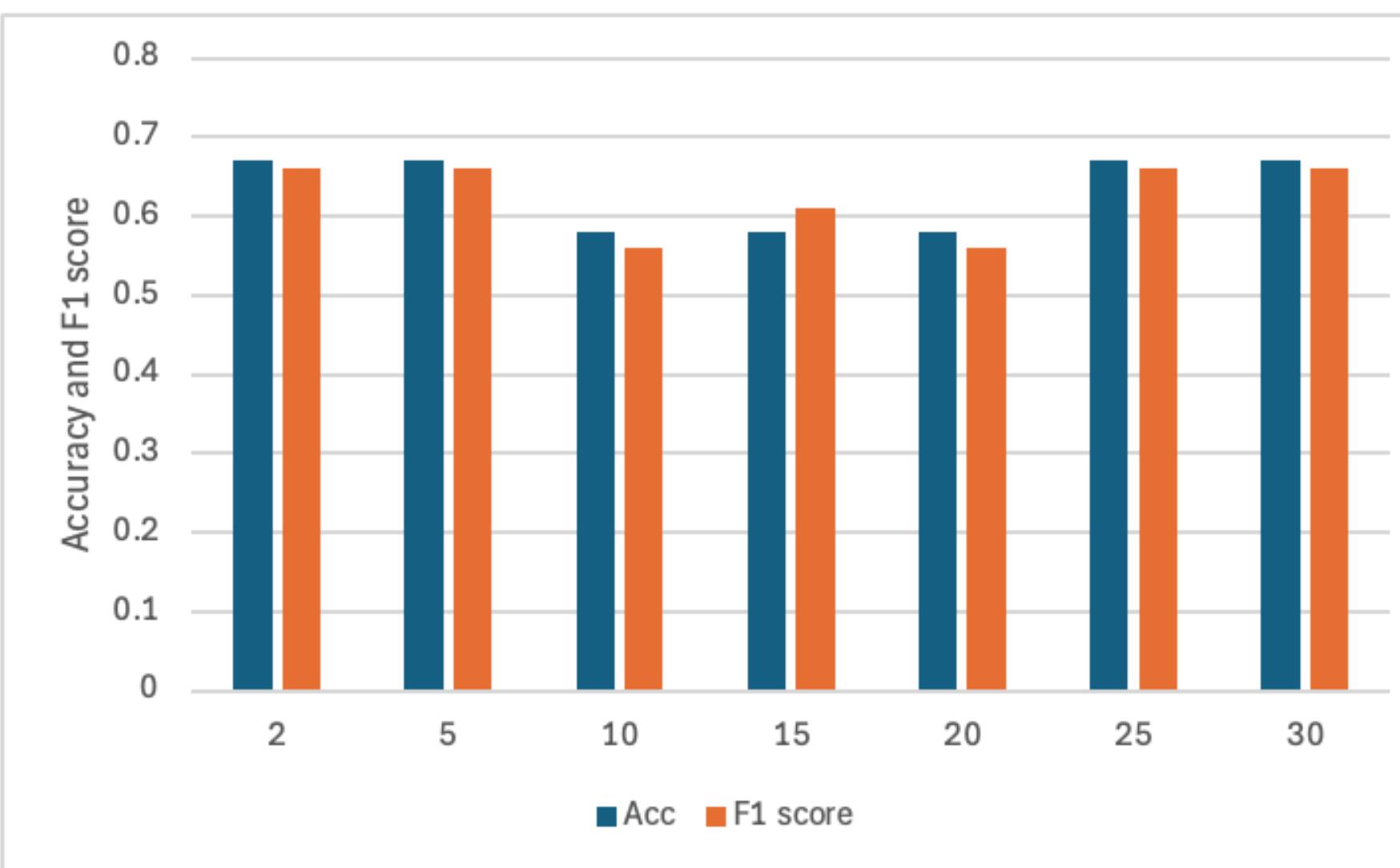
KNN Mean train score: 0.8304078014184396
KNN Mean test score: 0.7454545454545454

RF train score: 1.0
RF Mean test score: 0.74696969696969697

SVM Mean train score: 0.7458333333333333
SVM Mean test score: 0.7303030303030303

```
#SVM
#-----
svm = SVC(probability=True, random_state=42, C=0.1, gamma=0.1, kernel='sigmoid', class_weight=None).fit(X_train, y_train)
score_tr_SVM = svm.score(X_train, y_train)
score_te_SVM = svm.score(X_test, y_test)
scores_train_SVM.append(score_tr_SVM)
scores_test_SVM.append(score_te_SVM)
y_pred_SVM = svm.predict(X_test)
print("SVM Predictions:", y_pred_SVM)
```

PCA vs Sparse PCA



Model used: Logistic regression

Deep Learning Model

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout
from tensorflow.keras.regularizers import l2

model = Sequential()
model.add(Dense(8, activation='relu', input_shape=(X_train_sparse_pca.shape[1],)))
model.add(Dropout(0.4))
model.add(Dense(4, activation='relu'))
model.add(Dropout(0.4))
model.add(Dense(1, activation='sigmoid')) # Output layer for binary classification

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

model.fit(X_train_sparse_pca, y_train, epochs=200, batch_size=8, validation_data=(X_test_sparse_pca, y_test))

_, accuracy = model.evaluate(X_test_sparse_pca, y_test)
print('Accuracy: %.2f' % (accuracy*100))
```

Deep Learning Model

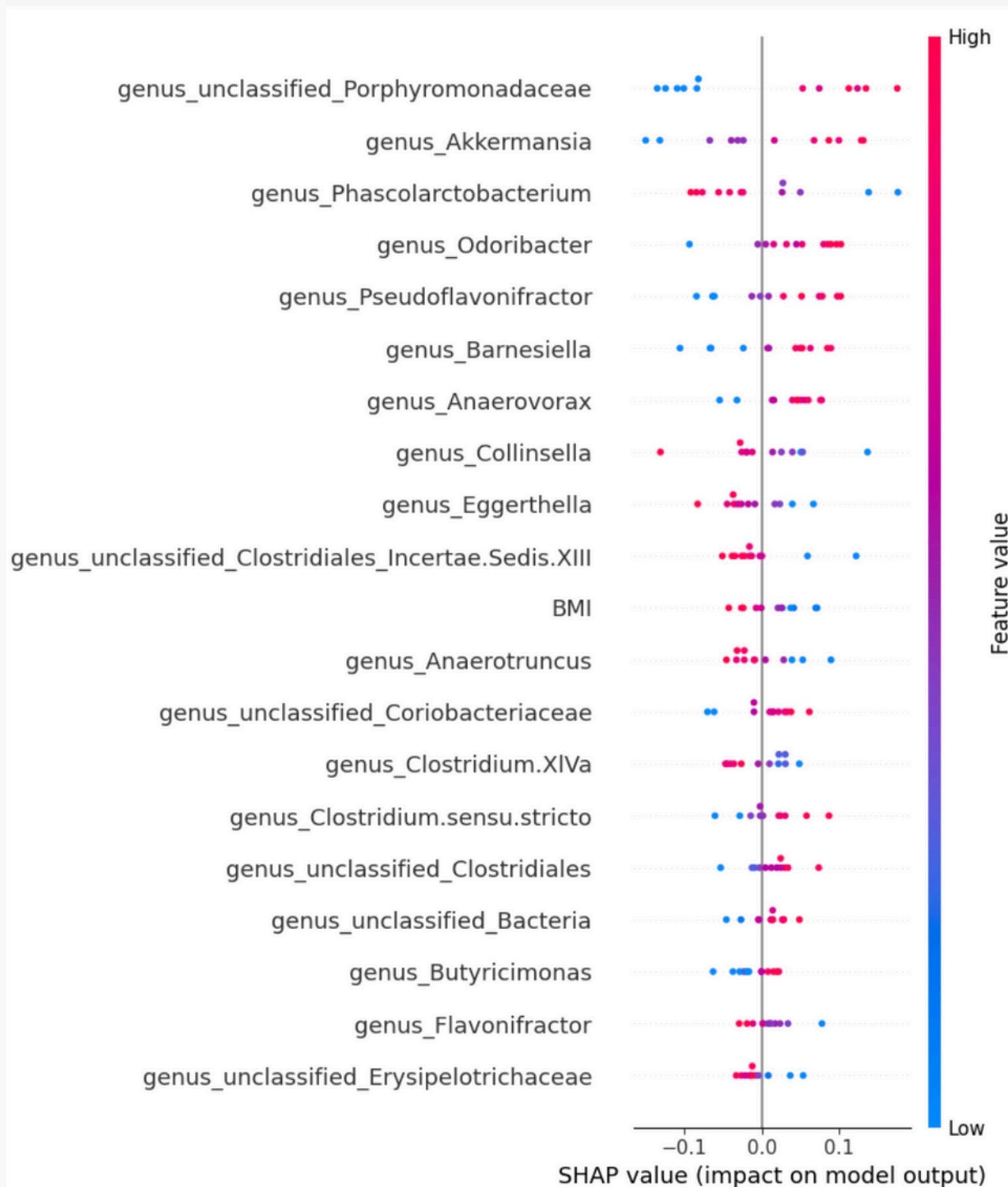
	train_loss	test_loss	train_acc	test_acc
DL Perfomance	0.52	0.49	0.84	0.83

Impute the NaN values

SubjectID	Sex	Age	Class	BMI	Race	IR_IS	SSPG	phylum_Actinobacteria	phylum_Bacteroidetes	phylum_Firmicutes	...	genus
ZJTKAE3	F	58.65	Control	31.24	C	IR	162.00	0.013645	0.582611	0.298357	...	
ZJXC41N	F	49.69	Prediabetic	28.24	B	IS	75.00	0.002822	0.260515	0.663827	...	
ZK112BX	F	57.72	Crossover	28.89	A	IR	234.00	0.047996	0.134357	0.419235	...	
ZK4CK8Y	F	50.02	Prediabetic	26.94	C	IS	40.00	0.005159	0.333832	0.531276	...	
ZKFV71L	F	66.08	Prediabetic	32.54	B	IS	65.00	0.003147	0.353603	0.516195	...	
ZKVR426	F	61.95	Prediabetic	30.59	C	IR	174.00	0.000936	0.444029	0.493188	...	
ZL63I8R	M	62.88	Prediabetic	32.94	C	IR	220.00	0.004373	0.695161	0.241594	...	

	train_acc	test_acc
After imputing	0.74	0.72

SHAP Model Interpretability



- SHAP (SHapley Additive exPlanations) Values
- Ran on Deep learning model
- Based on game theory and provides a way to fairly distribute the "gain" (or impact) of each feature across different possible feature combinations.
- **Positive Values:** Positive SHAP values indicate that the feature contributes to increasing the predicted value (i.e., pushing the prediction towards a higher outcome).
- **Negative Values:** Negative SHAP values mean the feature contributes to decreasing the predicted value (i.e., pushing the prediction towards a lower outcome).
- Higher probability of the positive class (IS) classification, higher score in regression

Nutritional Applications

- Act as markers of insulin resistance for early detection¹
- Improve therapeutic responses by understanding interactions between the gut microbiome and medications
- Inform potential dietary recommendations
 - Increase fiber intake^{4,5}
 - Probiotic and Prebiotic Supplementation⁶

Limitations

- Small sample size
- Complexity and symbiotic relationships between different microorganisms making up the gut microbiome

Future Works

- Use all data across the different visits
- Look at different statistical methods for microbiome

References

1. Crudele L, Gadaleta RM, Cariello M, Moschetta A. Gut microbiota in the pathogenesis and therapeutic approaches of diabetes. *eBioMedicine*. 2023;97. doi:10.1016/j.ebiom.2023.104821
2. Gurung M, Zhipeng L, You H, et al. Role of gut microbiota in type 2 diabetes pathophysiology. *eBioMedicine*. 2020;51. doi:doi.org/10.1016/j.ebiom.2019.11.051 2352-3964
3. Hills RD, Pontefract BA, Mishcon HR, Black CA, Sutton SC, Theberge CR. Gut microbiome: Profound implications for diet and disease. *Nutrients*. 2019;11(7):1613. doi:10.3390/nu11071613
4. Ojo O, Ojo OO, Zand N, Wang X. The effect of dietary fibre on gut microbiota, lipid profile, and inflammatory markers in patients with type 2 diabetes: A systematic review and meta-analysis of randomised controlled trials. *Nutrients*. 2021;13(6):1805. doi:10.3390/nu13061805
5. Paul P, Kaul R, Harfouche M, et al. The effect of microbiome-modulating probiotics, prebiotics and synbiotics on glucose homeostasis in type 2 diabetes: A systematic review, meta-analysis, and meta-regression of clinical trials. *Pharmacological Research*. 2022;185:106520. doi:10.1016/j.phrs.2022.106520
6. Zhao L, Zhang F, Ding X, et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science*. 2018;359(6380):1151-1156. doi:10.1126/science.aa05774