Machine Learning Groupings of Gut Microbes to Predict Dietary Weight Loss

Huw Evans
SCTP3669
The University of Sydney

June 30, 2018

1 Introduction

Differences in gut microbiomes between humans have been implicated in a variety of health problems, including obesity[1], irritable bowel syndrome, cardiovascular disease, inflammatory bowel disease[2] and colorectal cancer[3]. The causes and mechanisms of these issues, however, is less understood due to the large amounts of information available in complete genome sequences of microbiome samples. Such a problem is intractable to traditional methods of analysis, which can only handle small numbers of variables, and is better suited to computational, machine learned methods.

While Hjorth et al.[1] were able to distinguish weight losers on a 6-month diet based on broad categories of enteric microbes, the study does not explain its choice of categories, nor how future studies of different diets or other metabolic functions (such as glycaemic responses[4]) would be able to make similar distinctions. In particular, it may be the case that other dietary strategies can accommodate for the remainder of the participants, and that a future personalised health treatment would involve selecting the best of a number of diets for a patient. This study therefore aims to provide a more general method of categorising gut microbiomes in order to predict human weight loss on three different 3-month diets: Low glycaemic-index, high protein, or Mediterranean.

At least one stool sample was collected from each participant ($n = 90$, 30 per diet) before and after each of these diets, as well as their weight loss (as a percentage of their starting body weight) over this period. Each stool sample was represented by a subset of $13,107$ Operational Taxonomic Units (OTUs; a proxy for microbial species[5]), each expressed as a percentage of the total genetic material in the sample (so that all of the proportions in a sample summed to 1). Since the aim of this study is prediction, only the data from the samples collected before the dietary period were used. This dataset will be referred to as the 3diet data.

Previous attempts by Haunton[6] to predict weight loss from the samples did not find significant predictive power. The study also inappropriately applied Correlation-based Feature Selection (CFS) before cross-validation, which meant that the subset of features selected was created using data that was also used to test it. By doing so, CFS was able to select a subset of the OTUs which happened to correlate well with the result, decreasing the cross-validation error between the predicted and actual weight losses. When randomising the weight loss labels, the regressor was able to produce a similarly low error, implying that CFS selected a different subset of OTUs that happened to correlate well with the now-random result.

Gut microbiome data has three properties that reduce the effectiveness of machine learning: sparsity, noise, and small numbers of samples. Sparsity occurs when most of the values in a dataset are zero, implying that the meaningful or useful data is spread out among a large number of features. Many machine learning techniques assume that all features are equally important and struggle with data where the useful features are hard to determine. The problem is compounded by the large amounts of noise in most biological data, which randomly make features appear more or less useful than they are in the population distribution. In human data, it is also much harder to control their environments or whether they stick to their dietary plans. When machine learning models train on these noisy features, they will then often attempt to learn patterns that appear to emerge from the noise (known as overfitting). It is also difficult to obtain large numbers of samples due to the costs of working with humans and animals, which means that it will be more difficult to approximate the population distribution of data.

When dealing with gut microbiome data, these limitations can be mitigated by reducing the number of features. In doing so, the main options are to categorise by genetic similarity, or by indicators of functional similarity. For predicting bodily responses to dietary changes, it would be better to cluster together OTUs based on their function, especially since Read and Holmes[7] state that genetic similarity is unlikely to provide much information about the function of two different microbes.

In order to do so, OTUs were clustered on a second, larger dataset of 30 genetically identical mice. 3 mice in a cage were fed the same 12 diets in succession, spending 5 days on a diet followed by 1 day on standard chow. Faeces was collected and sequenced into $2,519$ OTUs on the 3rd and 5th days of each diet. This dataset will be referred to as the mouse data.

Since this data is more controlled with many more samples, it will reduce the amount and effects of noise. It will be able to capture a wider range of responses to different diets than 3diet, meaning that OTUs that respond similarly to dietary changes are more likely to belong to the same functional grouping. These functional groupings of OTUs could be transferred to the OTUs in 3diet to improve its predictive power.

Code for this study can be found at `https://github.sydney.edu.au/heva9329/sctp3669`, using Jupyter Notebooks. Packages used include `scipy`, `skbio`, `sklearn`, `tensorflow` and `numpy` (cited where relevant).

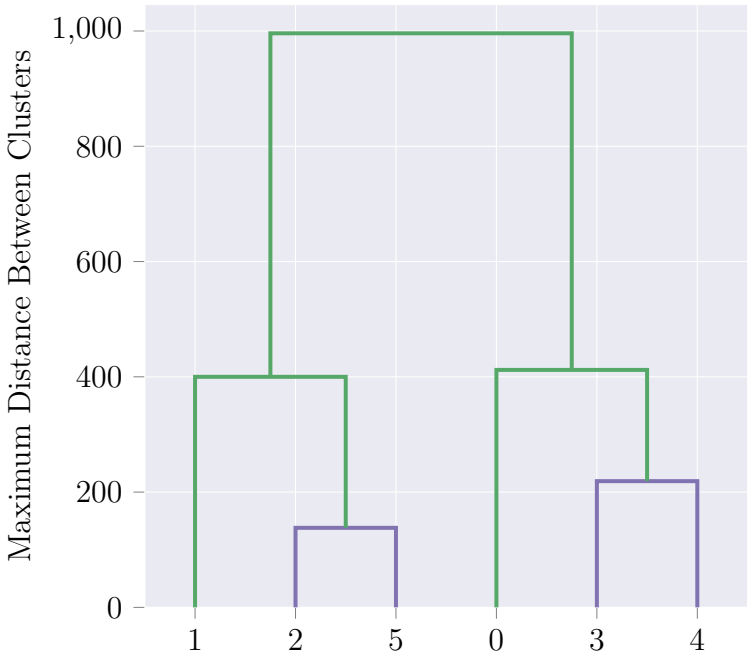

Figure 1: Example of a dendrogram using complete link clustering. Cutting off the dendrogram at $y = 300$ would produce the highlighted clusters $\{\{1\}, \{2, 5\}, \{0\}, \{3, 4\}\}$

2 Method

2.1 Forming Functional Clusters

There are a number of methods that can be used to cluster together OTUs. Many of them are unsuitable for microbiome data because they either require the number of clusters k to be provided *a priori* (which we had no reasonable estimate of), or they cluster over all of the supplied dimensions — where clustering is impossible due to many dimensions being zero or noisily distributed. Hierarchical clustering was chosen, which repeatedly merges clusters based on a supplied distance metric and does not have these issues. At each step, it compares each pair of elements between two clusters, and either merges them if the minimum (single link clustering), maximum (complete link), or mean (average link) distance is below the threshold for that step. Complete link clustering was chosen to ensure that different clusters were maximally separated.

This produces a dendrogram (Figure 1), which can be cut off on its y-axis to return the set of clusters that existed before that step. Since Euclidean distance suffered from the curse of dimensionality mentioned above, two different metrics were considered: Partial correlations, and a novel biologically-motivated δ-measure.

2.1.1 Partial Correlations Between OTUs

One way of comparing two OTUs is to determine to what extent the values of one OTU can be expressed as a linear function of another — i.e. how much they correlate, expressed as a correlation coefficient $-1 \leq \rho \leq 1$. However, this comes with a number of pitfalls that need to be mitigated. Existing approaches to OTU data are prone to statistical biases due to compositional data and deficiencies in correlation coefficients.

Compositional data (where the values in every sample sum to the same number) exhibit both a negative correlational bias and a different correlational structure to the population distributions[8, 9]. Gloor et al.[8] recommend that the data is first transformed from the unit simplex to the real space $\mathbb{R}^d$ by taking the centred log-ratio (clr) transform over the sample vector $\vec{x} = [x_1, x_2, \ldots, x_n]$:

$$\vec{x}_{\text{clr}} = [\log \frac{x_1}{G(\vec{x})}, \log \frac{x_2}{G(\vec{x})}, \ldots, \log \frac{x_n}{G(\vec{x})}] \tag{1}$$

$$\text{where } G(x) \text{ is the geometric mean } \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n}$$

Correlation coefficients are biased because they fail to take into account the effects of other variables on the pair being explicitly compared[8, 9]. If OTUs A and B correlate highly, but are actually conditionally independent when the effect of OTU C is subtracted, then the correlation matrix be positively biased toward them. In OTU data, where complex networks of cause and effect are prevalent, use of the correlation matrix will lead to incorrect results. Instead, Kurtz et al.[9] recommend using the inverse covariance matrix to compute partial correlations between pairs of OTUs, where a partial correlation first subtracts the effects of all other variables on each pair.

Unfortunately, as Cao, Lin, and Li[10] remark, the sample covariance matrix is singular when the number of variables is larger than the sample size, and therefore not invertible. There are a range of estimation techniques available, but many assume the population covariance matrix to be sparse, which runs contrary to our assumption that OTUs form complex networks of interactions. The Maximum Likelihood Estimator (provided in Scikit-learn as `EmpiricalCovariance`) was chosen, as of the methods that remained it was sufficiently unbiased when the number of variables greatly exceeded the number of samples[11]. Figure 2 compares the final estimated partial correlation matrix with the sample uncorrected correlation matrix. Code to produce this figure can be found in the Jupyter Notebook at `notebook/Correlations vs Partial Correlations.ipynb`.

For hierarchical clustering, the metric $d = 1 - \rho_{AB}$ was used to compute the difference between OTUs A and B. If A and B were maximally similar ($\rho = 1$), then $d = 0$, and if they were maximally dissimilar ($\rho = -1$), then $d = 2$.

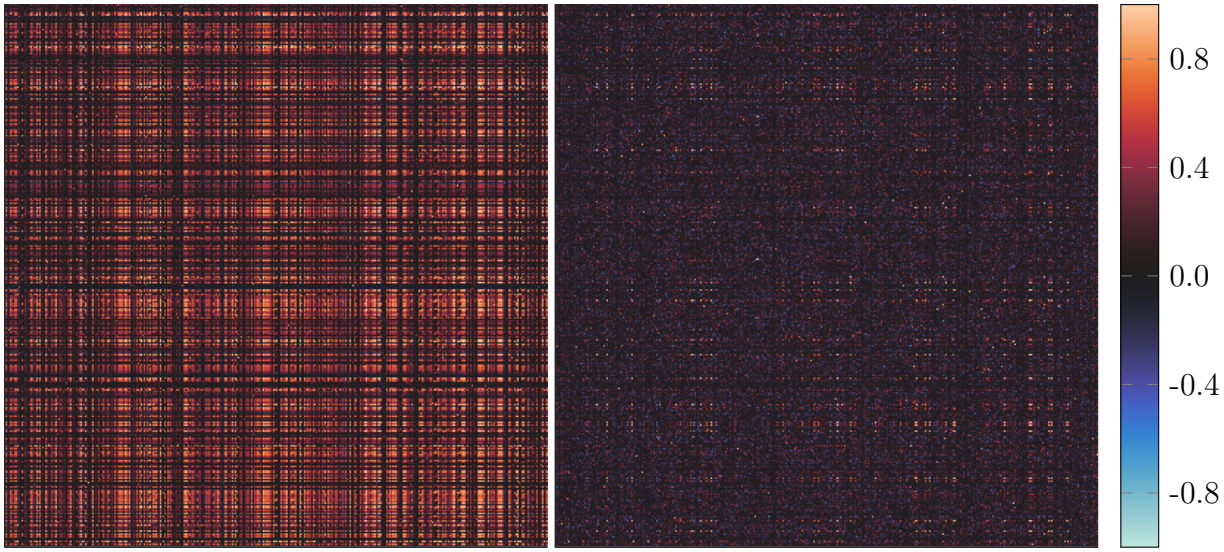

Figure 2: Pearson correlation matrix compared to partial correlation matrix. The Pearson matrix exhibits a strong positive bias. Some of its structure is still present in the partial correlation matrix.

2.1.2 δ-measure: Selecting Similarly Responsive OTUs

The δ-measure leverages the specifics of the mouse study to create a more biologically-motivated distance metric. If two OTUs A and B were to respond similarly to a change in diet (i.e. increase or decrease together), then it is likely that they perform the same function. This is because microbes in the gut process different components of different diets, and are able to populate more of the of the microbiome when fed those components. If A and B had similar responses to all of the dietary transitions, then it is very likely that they belong to the same functional grouping and should be clustered together.

Since the OTU data is compositional, care must be taken when interpreting a change in relative OTU abundance. If an OTU increases in relative abundance, that may be because it has increased in absolute abundance, or it may be because other OTUs have decreased in absolute abundance. However, if the relative abundances of two OTUs move in different directions we can be confident that they did not respond to the change in diet in the same way. For a pair of OTUs, the distance between them is the number of dissimilar responses in relative abundance (either an increase, decrease, or stagnation; stagnation only occurs when an OTU is below detection threshold for two diets). As each OTU is considered over 109 possible transitions $\times$ 3 mice per transition, noise in these responses should largely cancel out.

The final consideration was whether to group the mice in a cage or consider them individually. Since the mice in each cage were genetically identical and experienced exactly the same environment, taking a majority vote for the difference in an OTU's change would cancel a lot of noise (as it was assumed that any differences left over would very likely be noise). Hence this approach was selected.

2.2 Evaluation of Clusters on the 3diet Dataset

The method outlined above will produce a set of clusters, where each OTU in the mouse data belongs to exactly one group. In order to evaluate sets of clusters, they were transferred to the 3diet dataset and used as features to predict weight loss. For a set of clusters to be considered valuable, it has to make better predictions than a baseline where all of the OTUs were considered individually.

2.2.1 Baseline Evaluation

As each participant may have had more than one sample taken, all 3diet results used leave-one-group-out cross-validation. 90 models were successively trained, where the k^{th} model used all of the data except participant k's. Each model would then predict the weight loss for each of the samples it did not train on. Three techniques were used in baseline evaluation:

- Random Forests (varying the number of decision trees generated and the number of variables randomly selected to train each tree on)
- Neural Networks (varying the learning rate α, the regularisation term β, and the the number of hidden nodes h)
- Linear Regressors (varying whether a regularisation term β was applied, and the value of that term)

Selecting only the $x\%$ most abundant variables and leave-k-groups-out cross-validation were attempted, but did not produce meaningfully different results so were not considered in the final baselines.

During evaluation, one model was trained on each of the three component diets and had its predictions combined, while another was trained on the entire dataset. Each model's predictions were compared to the true weight loss by taking both the mean-squared error and a correlation coefficient. Code to produce the baseline evaluations can be found in `notebook/3diet Baseline.ipynb`.

2.2.2 Transferring Clusters to 3diet

To combine OTUs in a cluster into a single metric, they should be summed together to capture the intuition that OTUs are proportional. As they are from different species, the 3diet and mouse datasets only share 127 OTUs. One way of transferring clusters to 3diet was to only use these OTUs, combining them whenever two or more appeared in a cluster from the mouse data together.

The alternative solution was to assign each OTU in the 3diet data to a cluster based on the OTU's known metabolic capacities from the KEGG database[12]. Every OTU in both the mouse and 3diet studies has a set of known structures in its genome, some of which it uses to perform metabolic functions. We can create a 'signature' for each cluster by finding the proportion of OTUs in the cluster that have each structure, and assign new OTUs to each cluster by finding the closest match.

To find the score for whether an OTU belongs to a cluster, take the sum of the probabilities that each metabolic structure in the OTU belongs to the cluster, and divide by the base rate of that structure's prevalence. If half of the OTUs in a cluster had a certain structure but only one-tenth of the OTUs overall had it, then that structure would receive a score of $\frac{1}{2} \div \frac{1}{10} = 5$ for the cluster. The following situations are ranked in order of their score:

1. An OTU shares a lot of structures with a cluster, and those structures are not common in the population
2. An OTU shares a lot of structures with a cluster, but they're all common
3. An OTU shares few structures, but they're uncommon (and therefore distinguishing)
4. An OTU shares few structures, and they're common

To evaluate the transferred clusters, the cutoff for clustering on the dendrogram (t; the maximum difference between two clusters on the dendrogram) was varied between 0 and the maximum value that appeared on the dendrogram. At $t = 0$, each OTU is in its own cluster, and at $t = t_{\max}$, all of the OTUs are in one cluster.

The combined values were trained on a random forest regressor with 30 trees and the full number of features (1 per cluster) each time. Once again, the data was alternated between being split by diet and not. Code to produce these evaluations can be found in `notebook/3diet with Clusters.ipynb`.

3 Results

3.1 Baseline

Table 1 summarises the best baseline results for each method and hyperparameter tuned. The best possible mean-squared error achieved was 8.26, using random forests with # trees = 30 and # features = 8278 and splitting the dataset into 3. The predicted weight loss is compared with the actual weight loss in Figure 3. The full baseline results are included in the Appendix in Tables 3, 4 & 5.

Method	Parameter	Together	Split
RF	# features	8.30	8.47
	# trees	8.35	8.26
NN	α	19.63	18.57
	β	15.26	18.54
	h	8.71	8.48
Linear	β	8.42	8.29

Table 1: Sample baseline results for 3diet. For each method and hyperparameter, the best mean-squared error achieved during tuning is included.

	Intersection		Structures	
Diets	Partial	δ	Partial	δ
Together	8.97	10.81	8.44	8.66
Split	9.46	18.57	8.80	8.70

Table 2: Sample baseline results for the transferred clusters. For each variation of the method, the best mean-squared error is recorded.

Splitting the data into three diets produced significantly lower mean-squared errors on the baseline ($p = 0.018$; Paired t-test where H_0: $\text{MSE}_{split} < \text{MSE}_{together}$).

3.2 Transferring Clusters

Table 2 summarises the results when transferring OTUs from the mouse study into 3diet. The best possible mean-squared error achieved was 8.44, by transferring OTUs using their metabolic structures, clustering using partial correlations, and not splitting the data into diets. This result used a clustering cutoff of $t = 1.4$ (70% of the maximum distance on the dendrogram). The predicted weight loss for this configuration is compared with the actual weight loss in Figure 3. The full results for transferred clusters are included in the Appendix in Table 6.

Using this method, there was no significant difference in mean-squared error when splitting the data or using it all together ($p = 0.445$; Paired t-test where H_0: $\text{MSE}_{split} \neq \text{MSE}_{together}$). However, using partial correlations as a metric was significantly more performant than the δ-measure ($p < 10^{-5}$; Paired t-test where H_0: $\text{MSE}_{partial} < \text{MSE}_{\delta}$), and transferring clusters was significantly more performant than using the intersection ($p < 10^{-7}$; Paired t-test where H_0: $\text{MSE}_{structures} < \text{MSE}_{intersection}$).

In whole, transferring clusters performed significantly worse than the baseline results for random forests ($p < 10^{-22}$; Paired t-test where H_0: $\text{MSE}_{baseline} < \text{MSE}_{transfer}$, assuming unequal variances). The best method for transferring clusters (metabolic structures, partial correlations, not splitting) also performed significantly worse than the baseline results ($p = 0.022$; Paired t-test where H_0: $\text{MSE}_{baseline} < \text{MSE}_{transfer}$, assuming unequal

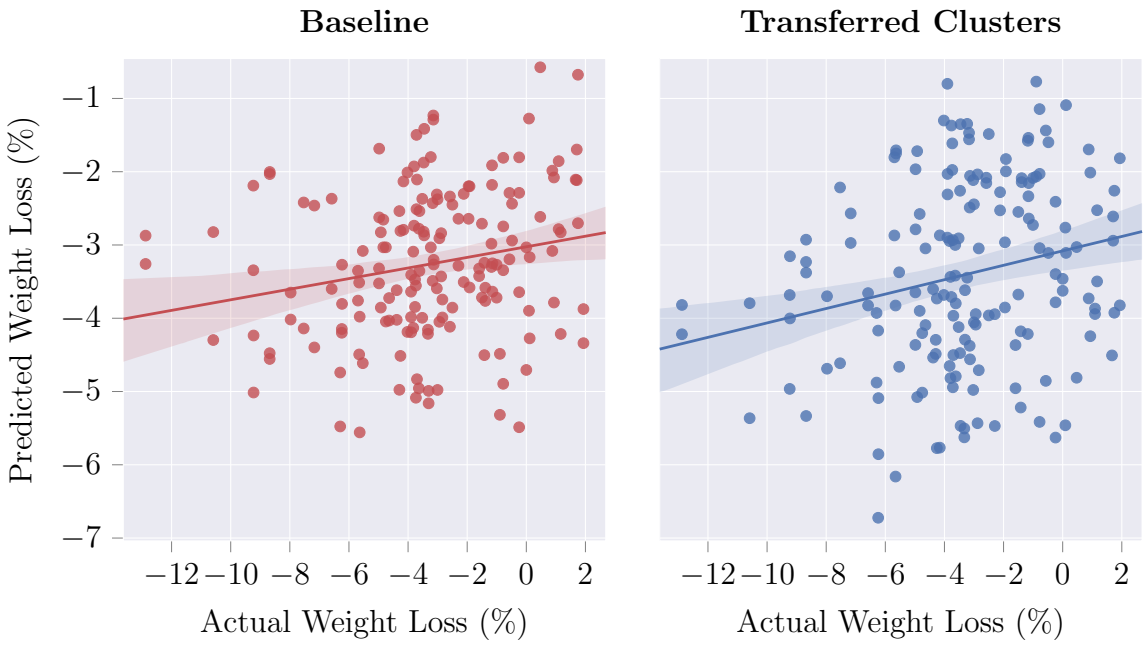

Figure 3: Predicted and actual weight loss on the best-performing models.

variances).

4 Discussion

The best performance on the 3diet dataset was unable to predict weight loss with a reasonable degree of accuracy. This is clear from the performance measures (mean-squared errors and correlation coefficients), and the sampled scatter plots, which also show highly divergent values for pairs of samples from the same patient. Since these results were evident in all tests, it is likely that the 3diet dataset is intractable with current methods. The reasons for this (including sparsity, noise, and a lack of samples) are discussed in the introduction, and widely apply to all microbiome datasets.

On the baseline results, using different methods did not hinder results after tuning. In all cases, tuned models achieved mean-squared errors in the range of 8.29 to 8.48. It was therefore acceptable to only use a single method (random forests) when transferring clusters. However, a model created to always predict the mean weight loss of its training data obtained mean-squared errors of 8.47 and 8.29 for the diets together and split, respectively. This implies that none of the models performed better than the most naïve guess.

Although splitting the dataset into the 3 diets achieved significantly smaller mean-squared errors, this is could be due to the difference in the two methods' naïve performance. No conclusions can be drawn about whether splitting the dataset into 3 was valuable.

Transferring OTUs from the mouse data to 3diet was unable to achieve significantly better results, even when comparing the best-performing transferred model to all of the baseline results (including untuned models). This cannot be explained by a difference in naïve guesses, so could be due to deficiencies in the transferring method. One side effect of combining features could be that with less features, there is less information for the model to learn from, so it produces worse predictions. A methodological reason could be that the transferred models relied on the tuned parameters from the baseline model, and that these parameters may no longer be the best possible for the new models.

When conducting the transfer, assigning each OTU in 3diet to a cluster based on its metabolic structure was better than using the 127 intersecting OTUs. This was expected, as using the full set of OTUs would mean that there was more information to learn from. This doesn't capture the full picture, since once transferred each method produced the same number of clusters, so there were no more variables to learn from. Instead, the likely reason for this performance increase is that the extra OTUs provided more fidelity to the existing clusters, many of which had no members from the intersection.

The most promising result is that using partial correlations performed significantly better than the δ-measure. As it was more biologically motivated, it was surprising that the δ-measure performed worse. This could be due to the fact that it binarises the change in OTU abundance between diets, and therefore does not provide any information about the magnitude of that change (unlike partial correlations, which are able to capture that). It could also be because the δ-measure does not seek to subtract the effect of all other OTUs.

5 Conclusion

Since the 3diet dataset was relatively intractable to machine learning methods, it was a poor candidate for evaluating this methodology. However, it was still able to produce some directions for future microbiome data analysis.

Firstly, microbiome studies are likely to benefit from the considerations outlined regarding partial correlations. Reasons were given in the introduction about why it is important to treat compositional data differently and take partial correlations over Pearson correlations, as well as specific methods for overcoming issues such as the singularity of the sample covariance matrix.

Secondly, new methods such as these should be evaluated on larger, more biologically similar datasets. The best way to counteract the issues outlined with microbiome data is to take more samples. As this can be expensive in humans, it would have been more useful for this study to evaluate the clustering method on another dataset from mice — not only because more samples would be able to be obtained, but because the clustering method was being performed on a dataset from mice. Using the same species for both datasets would have ensured that similar OTUs were present in both (instead of the much smaller intersection in this study), and that the underlying biological processes were more

similar. Further, using mice in both datasets would have ensured more control over the experiment, which would have reduced noise.

The methodology of transferring functional OTU groupings was inconclusive, because although it produced worse results than a naïve guess, the dataset being used to evaluate it was not tractable. If a future study were to repeat this method, it should cluster the OTUs using partial correlations, and transfer them using the metabolic structures. It should also ensure that its models are fully tuned before making any comparisons.

References

[1] M F Hjorth et al. "Pre-treatment microbial Prevotella-to-Bacteroides ratio, determines body fat loss success during a 6-month randomized controlled diet intervention". In: *Nature Publishing Group* 42.3 (Oct. 2017), pp. 580–583. DOI: 10.1038/ijo.2017.220. URL: http://dx.doi.org/10.1038/ijo.2017.220.

[2] Andrew B Shreiner, John Y Kao, and Vincent B Young. "The gut microbiome in health and in disease". English. In: *Current Opinion in Gastroenterology* 31.1 (Jan. 2015), pp. 69–75. DOI: 10.1097/MOG.0000000000000139. URL: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001574-201501000-00012.

[3] Nielson T Baxter et al. "Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions". English. In: *Genome Medicine* 8.1 (Apr. 2016), p. 104. DOI: 10.1186/s13073-016-0290-3. URL: http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0290-3.

[4] David Zeevi et al. "Personalized Nutrition by Prediction of Glycemic Responses". English. In: *Cell* 163.5 (Nov. 2015), pp. 1079–1094. DOI: 10.1016/j.cell.2015.11.001. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867415014816.

[5] Sarah P Preheim et al. "Distribution-Based Clustering: Using Ecology To Refine the Operational Taxonomic Unit". English. In: *Applied and Environmental Microbiology* 79.21 (Oct. 2013), pp. 6593–6603. DOI: 10.1128/AEM.00342-13. URL: http://aem.asm.org/lookup/doi/10.1128/AEM.00342-13.

[6] Charlotte Haunton. *Microbes, maladies and machine learning : an overview of the emerging field of microbiome-based machine learning*. Tech. rep. The University of Sydney, Oct. 2017.

[7] Mark N Read and Andrew J Holmes. "Towards an Integrative Understanding of Diet-Host-Gut Microbiome Interactions". In: *Frontiers in Immunology* 8 (May 2017), p. 361. DOI: 10.3389/fimmu.2017.00538. URL: http://journal.frontiersin.org/article/10.3389/fimmu.2017.00538/full.

[8] Gregory B Gloor et al. "Microbiome Datasets Are Compositional: And This Is Not Optional". In: *Frontiers in Microbiology* 8.November 2017 (Nov. 2017), pp. 1–6. DOI: 10.3389/fmicb.2017.02224.

[9] Zachary D Kurtz et al. “Sparse and Compositionally Robust Inference of Microbial Ecological Networks”. English. In: *PLoS computational biology* 11.5 (Apr. 2015), pp. 1–25. DOI: 10.1371/journal.pcbi.1004226. URL: http://dx.doi.org/10.1371/journal.pcbi.1004226.

[10] Yuanpei Cao, Wei Lin, and Hongzhe Li. “Large Covariance Estimation for Compositional Data via Composition-Adjusted Thresholding”. In: *arXiv.org* (Jan. 2016), pp. 1–36. URL: https://arxiv.org/pdf/1601.04397.pdf.

[11] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *arXiv.org* (Jan. 2012), arXiv:1201.0490. arXiv: 1201.0490 [cs.LG]. URL: http://arxiv.org/abs/1201.0490v4.

[12] Minoru Kanehisa and Susumu Goto. “KEGG: Kyoto Encyclopedia of Genes and Genomes”. In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 27–30. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10592173&retmode=ref&cmd=prlinks.

Appendix

Diets		Together		Split	
Parameter	Value	MSE	ρ	MSE	ρ
# features	1	8.46	0.094	8.49	0.120
	1380	8.60	0.066	8.64	0.101
	2760	8.56	0.083	8.51	0.145
	4139	8.57	0.083	8.47	0.157
	5519	8.56	0.087	8.53	0.143
	6898	8.76	0.046	8.85	0.094
	8278	8.30	0.158	8.65	0.135
	9658	8.62	0.085	8.82	0.107
	11037	8.74	0.055	8.71	0.127
	12417	8.72	0.073	8.92	0.099
# trees	10	8.98	0.083	9.32	0.042
	30	8.48	0.126	8.26	0.187
	50	8.41	0.135	8.34	0.167
	70	8.41	0.127	8.43	0.146
	90	8.35	0.138	8.36	0.154
	110	8.48	0.106	8.39	0.146
	130	8.46	0.110	8.38	0.148
	150	8.44	0.113	8.36	0.149
	170	8.44	0.115	8.37	0.146
	190	8.48	0.102	8.45	0.130

Table 3: Baseline Random Forest results for 3diet. Records mean-squared error (MSE) and correlation coefficient ρ between predicted and actual weight loss. The # features model was trained with # trees = 90, and the # trees model was trained with the best performing # features.

Diets		Together		Split	
Parameter	Value	MSE	ρ	MSE	ρ
α	10^{-5}	27.19	-0.106	19.73	0.065
	10^{-4}	26.03	-0.041	19.65	0.071
	10^{-3}	23.35	-0.014	18.57	0.131
	0.01	27.01	-0.010	19.61	0.109
	0.1	19.63	0.049	27.11	-0.141
β	10^{-8}	62.04	-0.080	24.49	-0.113
	10^{-7}	62.06	-0.080	21.45	0.004
	10^{-6}	62.07	-0.081	26.86	-0.137
	10^{-5}	62.01	-0.081	23.35	-0.111
	10^{-4}	62.05	-0.072	24.14	0.024
	10^{-3}	61.94	-0.072	24.63	-0.075
	0.01	60.07	-0.121	25.98	-0.096
	0.1	54.34	-0.214	22.41	0.059
	1.0	62.21	-0.192	21.53	0.093
	10	151.12	0.215	18.54	0.195
	100	373.29	-0.208	26.77	-0.021
	10^3	23.73	-0.215	18.99	0.135
	10^4	15.26	-0.075	21.84	-0.041
	10^5	16.44	-0.112	22.96	-0.023
	10^6	16.48	-0.114	21.14	0.020
	10^7	16.52	-0.115	19.63	0.110
	10^8	16.51	-0.115	27.07	-0.097
h	4	8.71	0.043	8.48	0.113
	8	9.28	-0.076	9.25	0.007
	16	9.38	0.010	9.18	0.074
	32	12.30	-0.235	9.33	0.144
	64	11.87	0.044	13.29	-0.047
	128	17.18	-0.003	18.12	-0.102
	256	23.77	-0.021	26.15	-0.086

Table 4: Baseline Neural Network results for 3diet. Records mean-squared error (MSE) and correlation coefficient ρ between predicted and actual weight loss. The α model was trained with $\beta = 10^3, h = 256$, successive models were trained with the tuned parameters from the previous ones.

Diets		Together		Split	
	Value	MSE	ρ	MSE	ρ
β	10^{-8}	8.75	0.029	8.42	0.125
	10^{-7}	8.75	0.029	8.42	0.125
	10^{-6}	8.75	0.029	8.42	0.125
	10^{-5}	8.75	0.029	8.42	0.125
	10^{-4}	8.75	0.029	8.42	0.125
	10^{-3}	8.75	0.029	8.42	0.125
	0.01	8.75	0.029	8.42	0.125
	0.1	8.75	0.029	8.42	0.125
	0 (No term)	8.75	0.029	8.42	0.125
	1.0	8.75	0.029	8.42	0.125
	10	8.75	0.029	8.42	0.125
	100	8.74	0.030	8.42	0.125
	10^3	8.68	0.035	8.39	0.130
	10^4	8.46	0.063	8.31	0.138
	10^5	8.42	0.070	8.29	0.138
	10^6	8.46	-0.062	8.29	0.137
	10^7	8.47	-0.098	8.29	0.136
	10^8	8.47	-0.101	8.29	0.136

Table 5: Baseline Linear Regression results for 3diet. Records mean-squared error (MSE) and correlation coefficient ρ between predicted and actual weight loss.

Transfer	Intersection				Structures			
Metric	Partial		δ		Partial		δ	
Diets	Together	Split	Together	Split	Together	Split	Together	Split
$t = 0\%$	9.96	10.10	11.00	10.48	9.01	8.81	8.96	8.94
10%	10.07	9.95	11.66	10.45	8.84	8.85	8.92	8.72
20%	10.14	10.10	11.99	11.70	8.59	8.80	8.72	8.70
30%	10.22	10.04	11.56	11.75	8.78	8.92	8.66	8.86
40%	10.09	10.17	11.56	12.14	8.81	9.20	9.01	9.84
50%	10.19	9.90	11.55	12.28	9.03	9.69	9.42	9.59
60%	10.25	9.76	11.51	12.28	8.82	9.83	9.75	10.17
70%	10.54	10.75	10.81	12.14	8.44	11.30	9.67	9.79
80%	9.83	9.46	11.58	11.97	10.24	11.08	10.65	9.75
90%	8.97	10.46	12.74	11.59	10.39	10.59	12.78	12.18
100%	12.74	11.59	12.74	11.59	12.77	12.18	12.78	12.18

Table 6: Results from transferring clusters. t is given as a percentage of the largest value on the dendrogram, such that $t = 100\%$ combines all of the OTUs into one cluster. Each score is a mean-squared error.