ML4FG Project Midpoint Report: Using Transfer Learning to Enhance Microbiome Analysis

George Austin, gia2105

March 2021

1 Introduction

Analysis of human microbiomes have improved our understanding of many diseases, including Inflammatory Bowel Disease, Type 2 Diabetes, and colorectal cancer (1). However, there remains a challenge in applying deep neural networks to this kind of data, which are typically high dimensional with a small number of datapoints. This means that any network needs lots of parameters, but can rarely have enough data to extract signals without overfitting. While it has been suggested that as more datasets become publicly available, there will be an opportunity to alleviate this issue by applying transfer learning methods (2), there are very few examples to date.

In this work, we explore transfer learning techniques such as hard parameter sharing (3) and Model Agnostic Meta-Learning (4) on publicly available microbial datasets. This is done by extending the results of DeepMicro(5), a study applying various deep learning methods to 6 different collections of fecal gut samples, demontrating that signals can be improved by using autoencoders to reduce the datasets' dimension. We use their autoencoders, in addition to other models such as feedforward neural networks, random forests and SVMs, as baselines which we use to determine if transfer learning techniques can lead to measurable improvements in predictive performance.

2 Methods

2.1 Data Summary

In total, we are working with six datasets, spanning five different prediction tasks: inflammatory bowel disease (6), type 2 diabetes(7)(8), obesity(9), liver cirrhosis(10), and colorectal cancer(11). The datasets are from six distinct studies, all of which provide publicly available metagenomic data. All datasets were processed during the DeepMicro (5) study using MetaPhlAn2 (12) and MetAML (13) to create two representations:

1. The Markers data, which illustrates the strain-level markers that appeared in the samples. These datasets have $O(100,000)$ distinct strains per sample
2. The Abundance data, which shows the species-level abundance for each sample. These datasets tend to have around 500 taxa per sample.

All baseline analyses are performed separately on each dataset, while the transfer learning techniques use combinations of either marker or abundance data.

Figure 1: Dataset Summaries [1]

	Nationality	Year	Positive Samples	Negative Controls
IBD	European	2010	25	85
EW-T2D	European Women	2013	53	43
C-T2D	Chinese	2012	170	174
Obesity	Danish	2013	164	89
Cirrhosis	Chinese	2014	118	114
Colorectal	French	2014	74	47

[1] Our Colorectal positive/negative numbers are different from what DeepMicro uses. This is because we classify our adenomas as positive, as these detections might warrant a colonoscopy, even if they are not cancerous. We might switch this back to DeepMicro's definitions.

2.2 Baseline Models

For our baseline, we recreated the models, training, any hyperparameter tuning scheme explored by DeepMicro. The models used are short autoencoder (SAE), deep autoencoders (DAE), variational autoencoders (VAE), random forests (RF), and support vector machines (SVM). We also add a deep feedforward neural network (FFNN), as an alternative baseline model structure that can use our transfer learning methods. Every model's tuned hyperparameters were the same as those used by DeepMicro (except FFNN), and were analyzed in the same train/validation/test scheme used in their analysis. As an alternative to their gridsearch tuning scheme, we reduced the runtime by running Facebook's AX tuning (14), which optimizes the search across the hyperparameter space. For this baseline analysis, all datasets are trained separately, obtaining a different set of hyperparameters and performance metrics for each.

2.3 Transfer Learning

The first step to applying the transfer learning methods is to select the datasets which are likely to share important signals. Figure S1 shows the percent of elements that are overlapping across datasets, suggesting some pairs of datasets are better candidates for transfer learning than others. PCA and PCOA analyses were also explored as a way to visualize batch effects. Figure S2 clearly shows the Cirhhosis and C-T2D datasets forming their own cluster. This might stem from the fact that these datasets originate from China (Figure 1), which could contribute to differences in the underlying microbial samples, or differences in processing techniques.

Due to these batch effects, we will start the transfer learning approaches by only considering either the European datasets, or the Chinese datasets. While for our purposes, we don't need to perfectly eliminate the batch effects to get some use out of the transfer learning, we can assume that we won't see many improvements by incorporating data representations that are completely different. For the time being, it does not appear that we will need to do any data transformations to further reduce batch effects, since the PCA representations of just the European data don't produce distinct clusters. Furthermore, it would be better to avoid transformations, if we can help it, so we can isolate the impact of just the transfer learning methods in comparison to DeepMicro's published results.

[1] https://www.nature.com/articles/s41598-020-63159-5/figures/2

3 Baseline Results

Figure 2 shows that we have for the most part been able to reproduce results within the uncertainty ranges reported by DeepMicro [1]. We performed the train/valid/test scheme DeepMicro propose, although we have so far only run it for one test group per dataset, which could account for some of our discrepancies with their results. Our encoders also do slightly worse for the colorectal data, which could stem from our classification of adenomas as positive. While this does allow for more potential improvements via transfer learning, we are considering adjusting this to keep our results comparable to DeepMicro's. We also see SVM occasionally fail, which will warrant a few additional tests. One surprising detail from our baseline results is that we often see our deep FFNN model outperform DeepMicro's autoencoder structures. This is unexpected, since it wasn't as big of a focus in the original paper, although it could bode well for this project, since the FFNN structure can lend itself a little more naturally to the transfer learning techniques proposed. While we will still explore hard parameter sharing and MAML for all the neural networks, these baseline results suggest our final results might focus a little more on our FFNN results.
Our exploration of the similarities across datasets suggests that we can evaluate transfer learning techniques by using the European datasets and the Chinese datasets separately (although it will also be a worthwile test to see if using all 6 datasets together leads to any improvements). While there are some noticeable batch effects, we should still be able to fairly test our methods without the need for any further transformations.

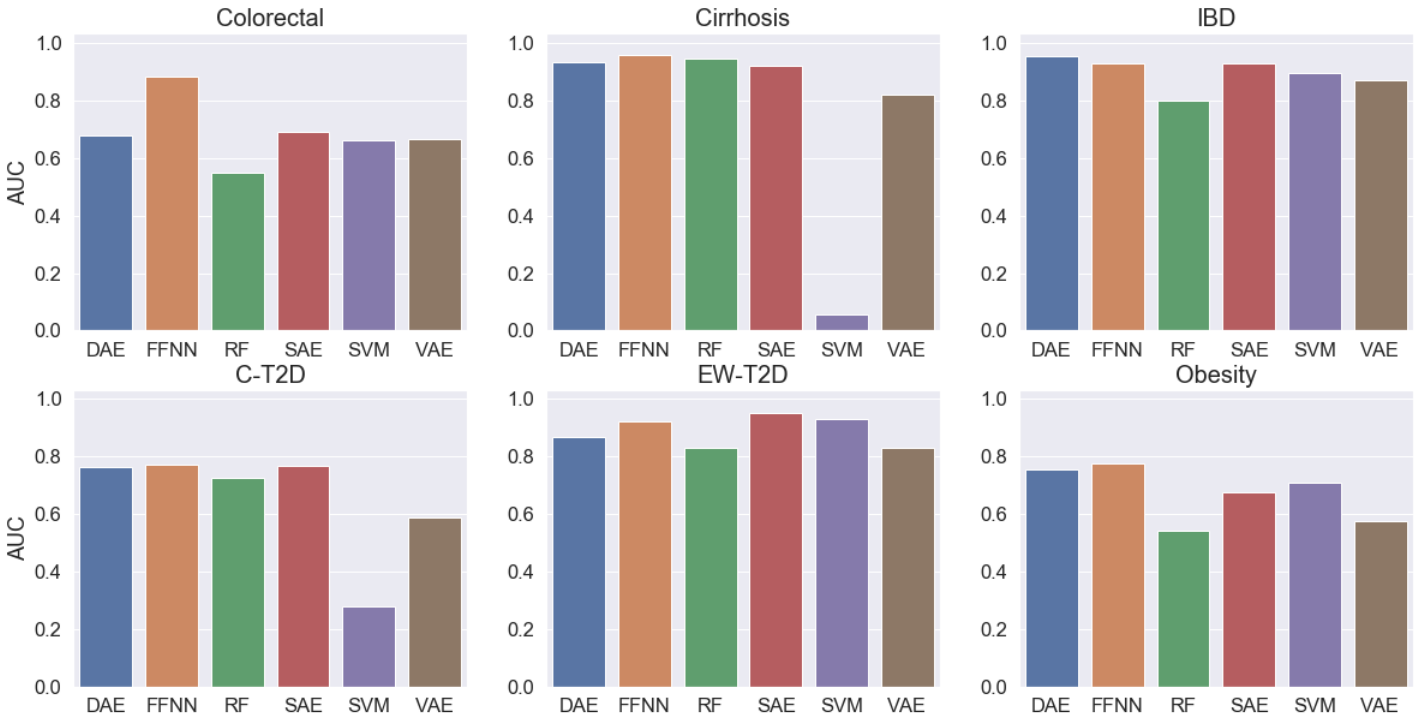
For all code, results and other figures obtained so far, see our github repository:
`https://github.com/gaustin15/Microb_Transer_Learning`

4 Next Steps

In general, the project is going well. All the goals for the midway point have been completed, and we have a plan in place to apply the proposed transfer learning techniques. Once those have been implemented, we will have a few options for how to proceed, including exploring transformations to reduce batch effects, and interpreting what signals are easier to detect using our methods, which would involve approaches like LIME (15) and Deeplift (16).

Figure 2: Baseline Marker Results from Tuned Models

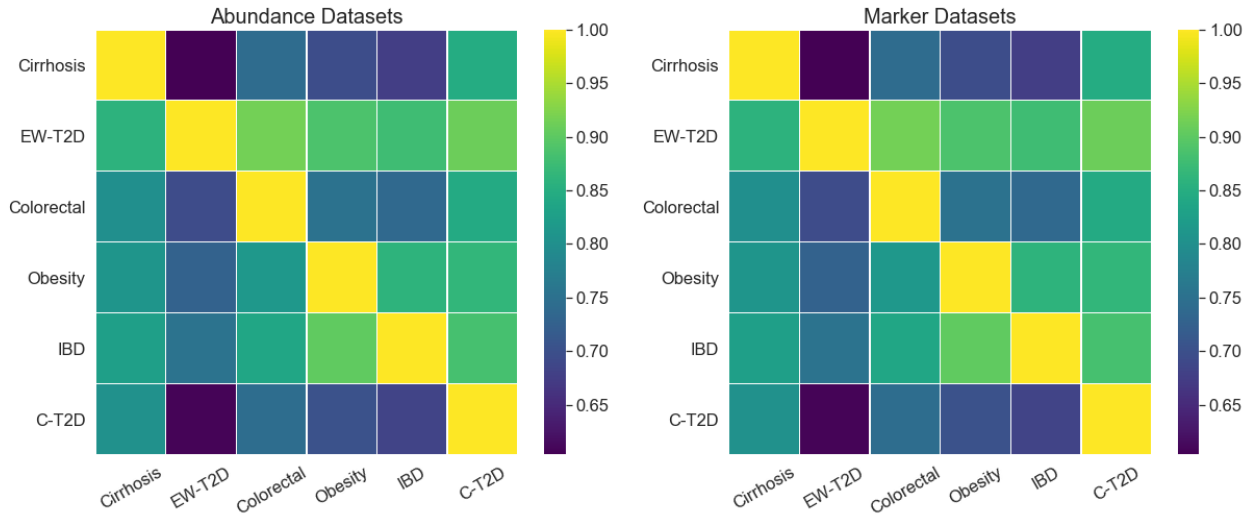
Showing resulting AUCs on the test set for deep autoencoder (DAE), Feedforward neural network (FFNN), random forest (RF), short autoencoder (SAE), support vector machines (SVM), and variational autoencoder (VAE).



5 Appendix

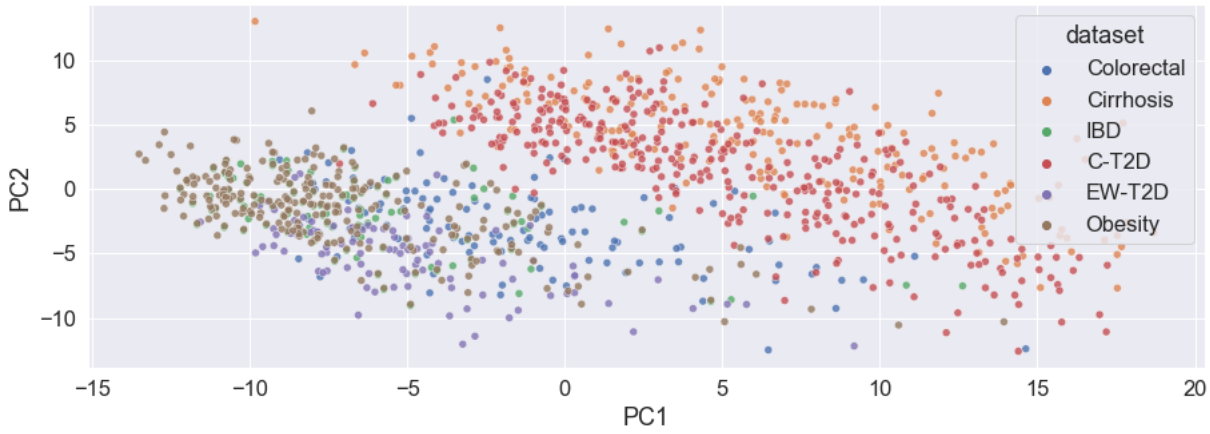
5.1 Supplementary Figures

Figure S1: Percent of Elements Overlapping Datasets



Element (i,j) of each heat map illustrates what percentage of elements from dataset i appear in dataset j. In general, there is a high rate of overlap across groups, with some exceptions. We also see very similar patterns across the marker and abundance datasets.

Figure S2: PCA of Markers Datasets



Similarly to Figure S1, we see that the Cirrhosis and the C-T2D datasets are separate from the other four groups. PCOA with Bray-Curtis similarity and abundances datasets give similar results.

References

[1] Marcos-Zambrano, Laura Judith, Kanita Karaduzovic-Hadziabdic, Tatjana Loncar Turukalo, Piotr Przymus, Vladimir Trajkovik, Oliver Aasmets, Magali Berland, et al. 2021. "Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment." Frontiers in Microbiology 12 (February). https://doi.org/10.3389/fmicb.2021.634511.

[2] Metwally, Ahmed A., Philip S. Yu, Derek Reiman, Yang Dai, Patricia W. Finn, and David L. Perkins. 2019. "Utilizing Longitudinal Microbiome Taxonomic Profiles to Predict Food Allergy via Long Short-Term Memory Networks." Edited by Yana Bromberg. PLOS Computational Biology 15 (2): e1006693. https://doi.org/10.1371/journal.pcbi.1006693.

[3] Sebastian Ruder. An Overview of Multi-Task Learning in Deep Neural Networks, 2017; arXiv:1706.05098.

[4] Chelsea Finn, Pieter Abbeel and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, 2017; arXiv:1703.03400.

[5] Oh, Min, and Liqing Zhang. 2020. "DeepMicro: Deep Representation Learning for Disease Prediction Based on Microbiome Data." Scientific Reports 10 (1). https://doi.org/10.1038/s41598-020-63159-5.

[6] Qin, Junjie, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, et al. 2010. "A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing." Nature 464 (7285): 59–65. https://doi.org/10.1038/nature08821.

[7] Karlsson, Fredrik H., Valentina Tremaroli, Intawat Nookaew, Göran Bergström, Carl Johan Behre, Björn Fagerberg, Jens Nielsen, and Fredrik Bäckhed. 2013. "Gut Metagenome in European Women with Normal, Impaired and Diabetic Glucose Control." Nature 498 (7452): 99–103. https://doi.org/10.1038/nature12198.

[8] Qin, Junjie, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, et al. 2012. "A Metagenome-Wide Association Study of Gut Microbiota in Type 2 Diabetes." Nature 490 (7418): 55–60. https://doi.org/10.1038/nature11450.

[9] Le Chatelier, Emmanuelle, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, et al. 2013. "Richness of Human Gut Microbiome Correlates with Metabolic Markers." Nature 500 (7464): 541–46. https://doi.org/10.1038/nature12506.

[10] Qin, Nan, Fengling Yang, Ang Li, Edi Prifti, Yanfei Chen, Li Shao, Jing Guo, et al. 2014. "Alterations of the Human Gut Microbiome in Liver Cirrhosis." Nature 513 (7516): 59–64. https://doi.org/10.1038/nature13568.

[11] Zeller, Georg, Julien Tap, Anita Y Voigt, Shinichi Sunagawa, Jens Roat Kultima, Paul I Costea, Aurélien Amiot, et al. 2014. "Potential of Fecal Microbiota for Early-stage Detection of Colorectal Cancer." Molecular Systems Biology 10 (11): 766. https://doi.org/10.15252/msb.20145645.

[12] Truong, Duy Tin, Eric A Franzosa, Timothy L Tickle, Matthias Scholz, George Weingart, Edoardo Pasolli, Adrian Tett, Curtis Huttenhower, and Nicola Segata. 2015. "MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling." Nature Methods 12 (10): 902–3. https://doi.org/10.1038/nmeth.3589.

[13] Pasolli, Edoardo, Duy Tin Truong, Faizan Malik, Levi Waldron, and Nicola Segata. 2016. "Machine Learning Meta-Analysis of Large Metagenomic Datasets: Tools and Biological Insights." Edited by Jonathan A. Eisen. PLOS Computational Biology 12 (7): e1004977. https://doi.org/10.1371/journal.pcbi.1004977.

[14] Eytan Bakshy, Lili Dworkin, Brian Karrer et at, AE: A domain-agnostic platform for adaptive experimentation; https://konstantinkashin.com/assets/papers/AE-NeurIPS2018.pdf

[15] Marco Tulio Ribeiro, Sameer Singh and Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier, 2016; arXiv:1602.04938.

[16] Avanti Shrikumar, Peyton Greenside and Anshul Kundaje. Learning Important Features Through Propagating Activation Differences, 2017, PMLR 70:3145-3153, 2017; arXiv:1704.02685.