Applying Feed-Forward Neural Networks to Gut Microbial Abundance Data for Disease Classification

Max Z. Grice, BS, April 21st, 2024
McWilliams School of Biomedical Informatics, Houston, Texas

Abstract

As the leading cause of death and disability in the U.S, the continued rise of chronic conditions like cancer and irritable bowel syndrome (IBD) continues to prompt preventative efforts like early diagnosis and/or identification of prominent risk factors involved in disease progression. Since many of these diseases are partially characterized by modifications in composition and function of gut microbial species, understanding the link between our microbiome and disease onset may be helpful in informing early diagnosis and targeted treatment approaches. Due to the high dimensionality and complexity of data from the human gut microbiome, this investigation sought to determine the feasibility of deep learning approaches to predict disease states given microbial abundance data. Specifically, a Feedforward Neural Network (FNN) was trained on 762 samples from patients that were classified as either Healthy, with Cancer, or with IBD. When compared to non-deep learning approaches like pixel similarity and linear modeling, the FNN approach was found to outperform all others with an accuracy of ~72% for cancer detection and an accuracy of ~85% for IBD detection. While these disease detection models require future improvements, the relatively high performance of the FNN compared to other approaches suggest that deep learning models should be studied more and are valuable in the context of microbial data and its role in disease progression.

Introduction

The human gut microbiota is highly diverse and contains trillions of microorganisms at various points along the digestive system, the shaping of which depends on various genetic, nutritional, and environmental factors (Gomaa, 2020). In recent years, research has found that dysbiosis, defined by modifications in the composition and function of gut microbiota, can be linked to diseases like cancer, cardiovascular disease, and IBD (Madhogaria, 2022). With 51.8% of U.S. adults currently diagnosed with a chronic condition, understanding specifically how the microbiota changes in disease states could help to uncover novel pathways involved in disease progression and motivate new therapeutic approaches or early diagnosis (Boersma, 2020). While linear modeling of high dimensional genomic data has traditionally been used to investigate mechanisms of disease, a growing number of studies are beginning to experiment with deep learning models for microbial data (Reiman, 2021; Nguyen, 2017). Using the wide array of microbial genus abundance levels to investigate and predict disease states with deep learning, specific microbes implicated in disease states could be uncovered and inspire new treatment approaches. Because microbial communities are directly influenced by environment and other host-associated factors, they provide an important link in understanding how the external world can influence our internal health. Gaining insight into how the human microbiome interacts with various chronic diseases and their progression may have clinical utility through disease prediction models which may in turn result in better deployment of preventative therapies, and better patient outcomes.

With this background in mind, the goal of this project was to conduct a preliminary investigation into whether deep learning could be used to classify patients by a given chronic health condition given gut microbial abundance data. Based on these microbe abundances, patient samples could be classified as one of three possible categories which included Healthy patients, patients with Cancer, and patients with IBD. In comparing three different classification approaches, including prediction based on pixel similarity, a simple linear prediction model, and a feedforward neural network, one could determine the extent to which deep learning was helpful in modeling microbial data. In doing so, this investigation can assist in paving the way for microbial data models to have clinical utility in disease prediction as well as in discovering novel mechanisms involved in disease progression.

Methods

Data Description

In total, there were 1,124 samples from 6 different studies (and 1,124 distinct patients), each of which collected fecal samples which were analyzed with amplicon or metagenomic sequencing to quantify gut microbial abundances. As

shown in table 1 below, the number of microbes quantified ranged from 89 to nearly 12,000, yielding a vast range of species and genera. While two of the studies focused specifically on characterizing the microbiota of patients with IBD (Chron's Disease and Ulcerative Colitis), the remaining four studies focused on patients with different types and stages of cancer. Control samples were present in all 6 studies and made up 482 of the 1,124 samples. Due to limitations in the number of available samples, I chose to classify samples into one of three possible conditions with included healthy patients, patients with some kind of gastrointestinal cancer, and patients with IBS. These classification categories are denoted by the coloring in Table 1, where yellow highlights denote all healthy sample groupings, orange coloring denotes all cancer sample groupings, and the purple coloring denotes all IBD sample groupings. The conditions that are not highlighted, specifically HS (history of surgery for cancer) as well as adenomas, were filtered from the dataset because they were more ambiguous and did not fit well into one specific disease condition. In total, this left 952 samples, 482 of which were healthy controls, 270 of which were patients with some type of cancer, and 200 of which were patients with some type of IBD. Due to differences in the methodology employed by these studies to quantify gut microbes, only microbes that were common between all 6 studies were used to predict disease classification. After filtering microbes that were not common between studies, this left a total of 45 microbial genera which were later used as features in the FNN and other modeling approaches.

Table 1. Number of total microbes, total samples, control samples, and other conditions for each study

Study	Microbes	Control Samples	Other Conditions	Total Samples
(Erawijantari, 2020)	10, 528	54	**Gastrectomy for GC (42)**	96
(Yachida, 2019)	11,943	127	**Colorectal Cancer (180)** - HS (3) - **Stage 0 (27)** - **Stage I-II (69)** - **Stage III-IV (54)** Adenoma (40)	347
(Kim, 2020)	500	102	**Colorectal Cancer (36)** Adenoma (102)	240
(Franzosa, 2019)	11,721	56	**Crohn's Disease (88)** **Ulcerative Colitis (76)**	220
(Jacobs, 2016)	509	54	**Crohn's Disease (26)** **Ulcerative Colitis (10)**	90
(Sinha, 2016)	87	89	**Colorectal Cancer (42)**	131

After filtering microbes and ambiguous conditions from the original data, the remaining samples were shuffled and then split into two groups for training and testing as shown in Figure 1 below.

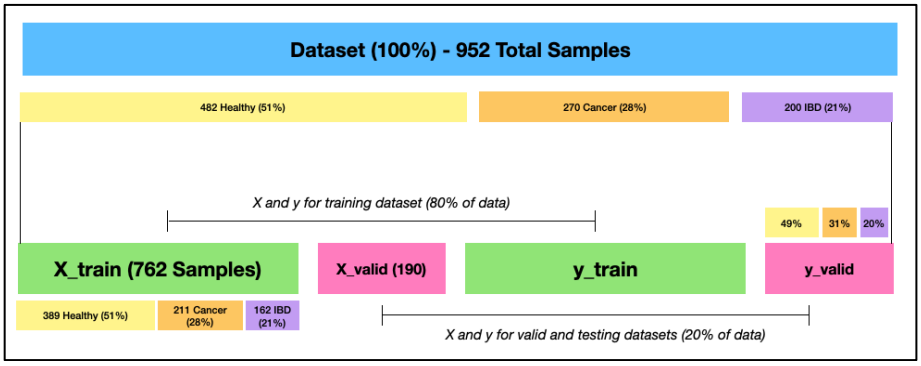

Figure 1. Total number of samples and class composition in original, training, and valid sets.

After shuffling the data to make sure that the training and test sets contained relatively equal portions of data from all six studies, 80% of the shuffled data (762 samples) were then used as training data while the remaining 20% (190 samples) were used as test data to compare the performance of modeling approaches. To further validate this split, I

made sure that the proportion of classes in the training and test sets (49-51% - 28-31% - 20-21%) were roughly equivalent to the proportion of classes in the original dataset (51% - 28% - 21%).

Data Modeling

To classify samples by condition (Healthy, Cancer, IBD), I tested three different approaches, including a pixel similarity prediction model, a linear model, and a feed forward neural network. To visualize the training data, I converted the format of each sample from a row in a pandas data frame to a list of 9×5 tensors with each value in the tensor representing the abundance (a number between 0 and 1) for one of the 45 possible microbes. With this format, it was more intuitive to use pixel similarity as an approach for categorizing images (i.e. healthy, cancer, or IBD). To measure pixel similarity, an "average image" for each disease category was needed so that a given "unknown" image could be compared to all three average images and the one it was most "similar" to (using absolute mean error) would be its predicted classification. To create these average images, tensors in the training set were separated by disease classification and the average tensor in each of the three groups of stacked tensors could then be calculated. After computing these averages, all images in the validation set were compared to the three averages to predict their disease classification based on the average image they were most similar to. The overall accuracy metrics for each class type based on these class predictions was then calculated and recorded.

In the second approach, I trained a simple linear model to separately model each classification. For example, to model cancer detection, I re-labeled all cancer images as 1 and all images without cancer as 0, creating a binary classification problem. Next, I created a matrix with 45 weights (one for each microbe feature), and then trained the model using stochastic gradient descent (SGD) as my optimizer and mean absolute error as a loss function. After training the model for 100 epochs, I recorded the accuracy of the model and repeated this process to model and detect images with IBD, and then to model and detect detect healthy images.

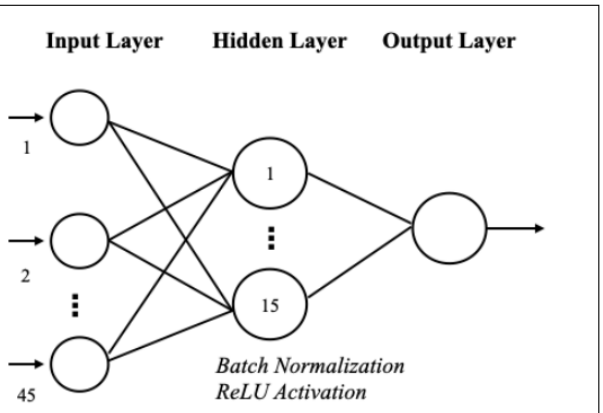

Figure 2: FNN Model Architecture

In the third and final approach, I used a feed forward neural network for the binary classification of each disease category (healthy, cancer, and IBD), resulting in three feed forward neural network models and their associated accuracies which could then be compared to the previous approaches outlined above. The architecture of the model was relatively simple with 3 layers, the first of which was the input layer which took in the 45 microbial features as a flattened vector in batches of 32. Next, there was a single hidden layer with 15 neurons, after which batch normalization was applied to normalize the activations of the previous layer which helped to reduce overfitting and accelerate the training process. A rectified linear unit (ReLU) activation function is then applied to introduce nonlinearity into the network, allowing it to learn the complex patterns that are characteristic of deep learning. Lastly, another linear transformation was applied to map the activations from the hidden layer to the output layer with one neuron used to produce the final classification. For more detail, a complete diagram of this FNN model architecture can be visualized above in Figure 2.

Results

Visualizations from the pixel similarity approach along with the accuracy metrics from each modeling approach (pixel similarity, linear modeling, and FNN) are given in the figures and tables below.

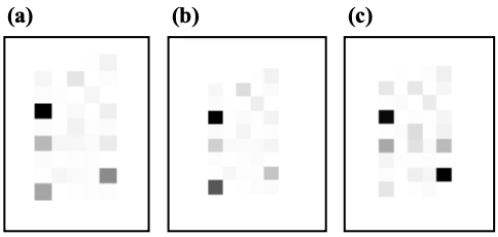

Figure 3: All images are 9 by 5 tensors with each cell representing the abundance level of one of the 45 different microbial genuses, with (a) being the average tensor image for Healthy samples, (b) being the average tensor image for samples with Cancer, and (c) being the average tensor image for samples with IBD.

Figure 3 shows the results from the pixel similarity approach described in the methods section. For this approach, each sample was converted to a 9×5 tensor with each cell representing one of the 45 microbes and its color indicating the level of abundance for that specific microbe (with darker cells denoting a higher abundance and lighter cells indicating lower abundances or no presence). The average image for each class could then be compared to all other images using the mean absolute error as a similarity metric and the average image that was most like a given image would be that images predicted class. In observing these images, the Healthy and Cancer average images look fairly similar to one another with only slight differences, while the IBD average image is much more distinctive in terms of microbial abundance patterns.

The accuracy for each approach (including pixel similarity, linear modeling, and FNN) are given in the table below and divided by class (with binary classification models created for all three classes in each approach).

Table 2. Accuracy Results for Each Modeling Approach and Class Type (Healthy, Cancer, & IBD)

APPROACH	ACCURACY
Healthy Class Detection	
Pixel Similarity	34.41
Linear Model	51.58
Feed Forward NN	60.5
Cancer Class Detection	
Pixel Similarity	47.46
Linear Model	68.95
Feed Forward NN	73.05
IBD Class Detection	
Pixel Similarity	81.58
Linear Model	80.00
Feed Forward NN	85.26

Discussion

The original purpose of this investigation was to investigate whether deep learning models could be effectively utilized to understand and possibly predict how the microbial environment of a given sample may impact a disease state like cancer or irritable bowel syndrome. Using microbial abundance data from 6 different studies and 952 samples, non-deep learning approaches used to classify samples as either Healthy, Cancer, or IBD were compared to the results of a feed forward neural network. In observing the results of Table 2, it is clear that the FNN performs significantly better than all other non-deep learning approaches for all three binary classification models, with FNN IBD class detection showing the greatest level of accuracy (~85%) compared to all other models. This makes sense when looking at the average pixel image for IBD samples in Figure 3 since the image is much more distinctive and shows more activity than the average images for samples classified as either Healthy or Cancer. The IBD samples also came from only 2 different datasets (as opposed from Healthy samples coming from all 6 studies and cancer samples coming from 4 studies) which may explain why it was easier for the model to correctly classify them as IBD.

The binary classification model for cancer detection was also relatively accurate (~73%) especially when compared to the results from pixel similarity prediction for cancer (~47%). The large difference between these results support the claim that deep learning can be effectively used to predict how a given microbial environment may impact diseases like Cancer. Unlike the FNN detection models for Cancer and IBD, the accuracy using FNN to classify Healthy samples was relatively low (~60%). This may be due to the fact that Healthy samples are often relatively

heterogeneous since the research has been unable to find any ideal microbiota composition (instead, healthy microbiomes are characterized by diversity). As a result, it would have been more difficult to detect Healthy samples as opposed to Cancer and IBD samples which may have had more of a distinct microbial signature. While all models require future improvement in terms of performance metrics, these preliminary results suggest that distinct microbial signatures of Cancer and IBD do exist and should be explored further with the help of deep learning models.

Limitations

There are two general limitations to this investigation, the first of which includes problems arising of the merging and integration of data from different sources. One specific problem that arises from this limitation is that different studies often use different methods to analyze the microbiome, the most prominent of which include amplicon and metagenomic sequencing. While both approaches yield microbial abundance data, differences between targeted and non-specific sequencing can change the type and number of microbes that are quantified. While I tried to compensate for this limitation by only selecting microbiota that are common between datasets, there is still some inherent uncertainty that comes with comparing data with different quantification methodologies. Selecting only common microbiota may have also limited this investigation further by decreasing the number of available features that could be utilized by the deep learning model to classify samples. In future work, it would be helpful to compare different filtering approaches that attempt to compensate for this limitation.

The second general limitation lies in the assumption that data categorizations and structure behave in simple and predicable ways. While this assumption was helpful for the purpose of modeling, in reality the three categories of Healthy, Cancer, and IBD can be broken down into many more subcategories, including the stage or type of cancer along with specific subtypes of IBD. The true complexity of these categories may have in turn impacted model performance and may explain some of the incorrectly categorized samples. While I assumed that the number of features was small (45) compared to the number of samples (952), there is also the reality that most of the datasets have a large number of features (over 1000) compared to a small number of samples (100 or less) which also contributes some uncertainty to the modeling results and should be accounted for in the future.

Conclusions

In total, 952 samples from 6 different studies were extracted and pre-processed, with each sample classified as being either Healthy, with Cancer, or with IBD. With the goal of determining whether deep learning models could be effectively utilized to predict how microbial environments contribute to disease states, the results of non-deep learning approaches such as pixel similarity and linear models were compared to a feed forward neural network for each of the three classes. In all three classes, the FNN for binary classification significantly outperformed pixel similarity and linear modeling approaches in terms of detection accuracy. Using the FNN to detect samples with IBD (accuracy=~85%) and to detect samples with Cancer (accuracy=~72%) performed relatively well while using the FNN to detect healthy samples (accuracy=~60%) had a much lower accuracy which could be attributed to the heterogeneity of Healthy samples.

While there are some limitations that should be addressed in future work, including accounting for problems in merging data from different sources, small sample sizes, and complexities in disease classification, the success of this simple deep learning model shows promise and should be expanded upon to include more complex models such as Convolutional Neural Networks (CNNs) and Tabular Learners. Using these models to classify samples and predict disease states may provide the tools necessary to uncover specific microbial features relevant to disease progression which could be targeted in therapy to promote better health outcomes for these conditions.

References

1. Boersma, P., Black, L. I., & Ward, B. W. (2020). Prevalence of Multiple Chronic Conditions Among US Adults, 2018. *Centers for Disease Prevention and Control*, *17*(17). https://doi.org/10.5888/pcd17.200130
2. Chen, X., Zhu, Z., Zhang, W., Wang, Y., Wang, F., Yang, J., & Wong, K. C. (2022). Human disease prediction from microbiome data by multiple feature fusion and deep learning. iScience, 25(4), 104081. https://doi.org/10.1016/j.isci.2022.104081
3. Erawijantari, P. P., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Saito, Y., Fukuda, S., Yachida, S., & Yamada, T. (2020). Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. *Gut*, *69*(8), 1404–1415. https://doi.org/10.1136/gutjnl-2019-319188
4. Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., Vatanen, T., Hall, A. B., Mallick, H., McIver, L. J., Sauk, J. S., Wilson, R. G., Stevens, B. W., Scott, J. M., Pierce, K., Deik, A. A., Bullock, K., Imhann, F., Porter, J. A., Zhernakova, A., … Xavier, R. J. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature microbiology*, *4*(2), 293–305. https://doi.org/10.1038/s41564-018-0306-4
5. Gomaa E. Z. (2020). Human gut microbiota/microbiome in health and diseases: a review. *Antonie van Leeuwenhoek*, *113*(12), 2019–2040. https://doi.org/10.1007/s10482-020-01474-7
6. Jacobs, J. P., Goudarzi, M., Singh, N., Tong, M., McHardy, I. H., Ruegger, P., Asadourian, M., Moon, B. H., Ayson, A., Borneman, J., McGovern, D. P., Fornace, A. J., Jr, Braun, J., & Dubinsky, M. (2016). A Disease-Associated Microbial and Metabolomics State in Relatives of Pediatric Inflammatory Bowel Disease Patients. *Cellular and molecular gastroenterology and hepatology*, *2*(6), 750–766. https://doi.org/10.1016/j.jcmgh.2016.06.004
7. Kim, M., Vogtmann, E., Ahlquist, D. A., Devens, M. E., Kisiel, J. B., Taylor, W. R., White, B. A., Hale, V. L., Sung, J., Chia, N., Sinha, R., & Chen, J. (2020). Fecal Metabolomic Signatures in Colorectal Adenoma Patients Are Associated with Gut Microbiota and Early Events of Colorectal Cancer Pathogenesis. *mBio*, *11*(1), e03186-19.
8. Liu, Y., Fachrul, M., Inouye, M., & Méric, G. (2024). Harnessing human microbiomes for disease prediction. Trends in microbiology, S0966-842X(23)00339-6. Advance online publication. https://doi.org/10.1016/j.tim.2023.12.004
9. Madhogaria, B., Bhowmik, P., & Kundu, A. (2022). Correlation between human gut microbiome and diseases. *Infectious medicine*, *1*(3), 180–191. https://doi.org/10.1016/j.imj.2022.08.004
10. Muller, E., Algavi, Y.M. & Borenstein, E. The gut microbiome-metabolome dataset collection: a curated resource for integrative meta-analysis. *npj Biofilms Microbiomes* **8**, 79 (2022). https://doi.org/10.1038/s41522-022-00345-5
11. Nguyen, T. H., Chevaleyre, Y., Prifti, E., Sokolovska, N., & Zucker, J.-D. (2017, December 1). *Deep Learning for Metagenomic Data: using 2D Embeddings and Convolutional Neural Networks*. ArXiv.org. https://doi.org/10.48550/arXiv.1712.00244
12. Reiman, D., Layden, B. T., & Dai, Y. (2021). MiMeNet: Exploring microbiome-metabolome relationships using neural networks. *PLoS computational biology*, *17*(5), e1009021. https://doi.org/10.1371/journal.pcbi.1009021
13. Sinha, R., Ahn, J., Sampson, J. N., Shi, J., Yu, G., Xiong, X., Hayes, R. B., & Goedert, J. J. (2016). Fecal Microbiota, Fecal Metabolome, and Colorectal Cancer Interrelations. *PloS one*, *11*(3), e0152126. https://doi.org/10.1371/journal.pone.0152126
14. Yachida, S., Mizutani, S., Shiroma, H. *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* **25**, 968–976 (2019). https://doi.org/10.1038/s41591-019-0458-7