Evolution-Informed Neural Networks for Microbiome Data Analysis

Michael Ito
Information and Computer Sciences
University of Hawai'i at Mānoa
Honolulu HI, USA
mi24@hawaii.edu

Yannik Glaser
Information and Computer Sciences
University of Hawai'i at Mānoa
Honolulu HI, USA
yglaser@hawaii.edu

Peter Sadowski
Information and Computer Sciences
University of Hawai'i at Mānoa
Honolulu HI, USA
peter.sadowski@hawaii.edu

***Abstract*—Advances in metagenomic sequencing have provided an unprecedented view of the microbial world, but untangling the web of microbe interdependencies and the complex relationship between microbiome and host is a major challenge in biology. New statistical methods are needed to analyze metagenomic data and infer these relationships. Focusing on amplicon sequencing data, we present methods for leveraging phylogenetic information in deep neural network models and for transfer learning from large data repositories. This approach is demonstrated in experiments using data from the Earth Microbiome Project (EMP) and a dataset of 1500 samples from Waimea Valley on the island of Oahu, Hawaii.**

***Index Terms*—deep learning, neural networks, phylogeny**

I. Introduction

Rapid advances in metagenomic sequencing technology have enabled scientists to quantify the composition of microbial communities from environmental samples. These methods enable us to quantify the relative abundance of hundreds of thousands of different bacteria and/or fungi variants. From these measurements, biologists attempt to infer microbe-microbe or microbe-host relationships. Understanding the dependencies within these microbe communities, their stability, and their interactions with their host are fundamental problems in biology.

There is significant interest in developing better machine learning methods for metagenomic analysis [1], [2]. However, non-linear machine learning methods quickly overfit due to the relatively small sample size and high dimensionality of the data: the well-known "large P, small N" problem. Thus, for most microbiome data analysis applications, deep learning is not a practical approach [3]. However, there are two sources of additional information that could change this situation: (1) phylogenetic information (which results in correlations between related microbes, similar to correlations among neighboring pixels in images); and (2) large repositories of metagenomic data, such as the EMP [4], [5] and the Human Microbiome Project (HMP) [6], [7], which could potentially be used for transfer learning if good pre-trained deep learning models were developed. In this work we propose a novel neural network architecture that enables us to leverage both sources of information.

Neural architecture design provides a source of inductive bias in deep learning by constraining the hypothesis space. Important examples include convolutional architectures for exploiting symmetries in sequences and images, graph architectures for capturing spatial relationships in small molecules [8], [9], and "physics-informed" neural networks for incorporating physics domain knowledge [10]–[12]. An analogous source of domain knowledge in microbiology is phylogeny: the evolutionary history of microbes. Traits have a tendency to be conserved among related microbes despite the prevalence of lateral gene transfer [13], so a phylogenetic tree provides a hierarchical clustering of correlated features for analysis — indeed, it is common to cluster related microbes into "operational taxonomic units"' (OTUs) to reduce the data dimensionality and simplify analysis.

We propose a new deep learning architecture for microbiome data analysis that incorporates this phylogenetic information — an *Evolution-Informed Neural Network* (EINN). Typically tree-shaped, the architecture has edges corresponding to phylogenetic relationships (Fig. 1). A neural network module at each node maps vector-valued representations of its children to a new vector-valued representation, summarizing the taxon of related microbes included in the subtree. This local connectivity reduces overfitting by constraining the function space, just as convolutional neural networks in computer vision learn local features that summarize patches of neighboring (and highly-correlated) pixels.

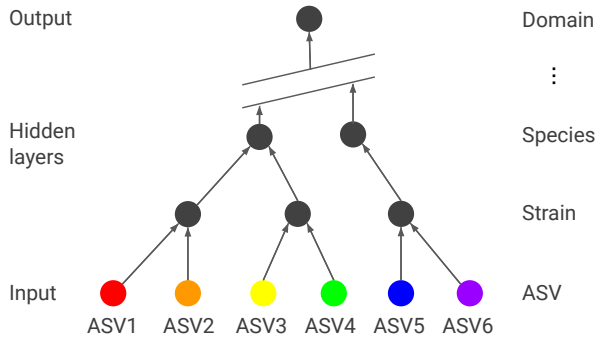

Fig. 1. Evolution-Informed Neural Networks use phylogenetic information such as taxonomy to define sparse connectivity. This reduces overfitting on high-dimensional metagenomic data.

A key insight is that EINN architectures can be adapted

to a variety of applications. In addition to predicting host phenotypes and microbiome-level properties such as stability, EINNs can predict properties of individual microbes using a dynamic, recursive architecture that changes for each input example (Fig. 2), analogous to work in natural language processing where semantic parse trees can be used to construct a unique tree architecture for any given sentence [14], [15]. This has applications in cataloging microbial functional traits, which are important for understanding the *mechanisms* of microbe-microbe and microbe-host interactions [16].

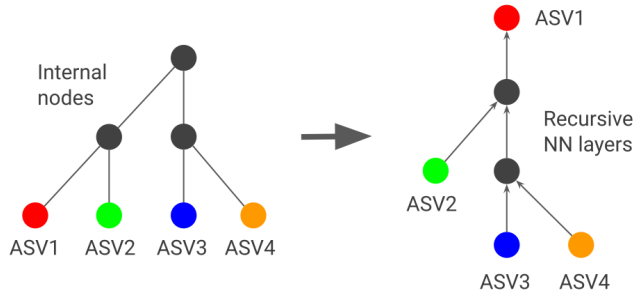

Fig. 2. A dynamic EINN can be used to predict properties of microbes from those of related microbes. The use of weight-sharing enables a unique architecture to be constructed for each example by transforming the phylogenetic tree (left) to a tree with the target microbe at the root (right)

Another important advantage of EINNs is that they enable the use of transfer learning to improve performance using external data [17]. EINNs can be pre-trained on large repositories such as the EMP (for environmental studies) or the HMP (for human microbiome studies), then fine-tuned to solve a related task on a smaller target dataset. EINNs provide a natural approach to transferring pre-trained models even when the target dataset contains a different set of amplicon sequence variants (ASVs), as the tree architecture can simply be extended to include any new ASVs, with similar ASVs automatically clustered within the tree.

The contribution of this work is to propose EINNs as a general framework for leveraging phylogenetic information in microbiome data analysis. Efficient software implementations that leverage sparse matrix multiplications are provided in Tensorflow [18] (for static architectures) and Pytorch [19] and Deep Graph Library [20] (for dynamic architectures) (to be released upon acceptance). Experiments demonstrate the advantages of EINNs over fully-connected neural networks and standard shallow machine learning approaches on microbiome classification tasks for two 16S rRNA sequencing datasets of different sizes: EMP (30,000 samples) and Waimea Valley (1,500 samples). The advantage of transfer learning is demonstrated by pre-training on the EMP dataset and fine-tuning on the Waimea Valley dataset. Finally, we demonstrate the use of dynamic EINNs to predict the microbe metabolism from features of related microbes.

II. Related Work

Previous work has demonstrated that phylogenetic information can be successfully incorporated into deep learning models. Reiman, *et al.* 2017 [21], Fioravanti, *et al.* 2018 [22], Wassan, *et al.* 2018 [23], and Manning, *et al.* 2018 [24], each incorporate phylogenetic information into deep neural networks by first embedding amplicon sequence profiles into images using various strategies before applying standard convolutional neural networks. In contrast, the most similar approach to ours is that of Khan and Kelly, 2020 [25] who use a phylogeny adjacency matrix to define a graph, then apply graph convolution neural networks [26]. They demonstrate the performance advantage of their deep learning architecture on a disease classification task with a dataset of 5,643 samples. EINNs are similar to this approach in that information is propagated along a computational graph determined by the phylogenetic tree, but differ in their use of weight-sharing (not required in EINNs) and computational efficiency. EINNs are significantly more efficient in terms of memory and computation by reducing the number of nodes in subsequent layers, as opposed to representing the entire graph at each layer. Furthermore, Khan and Kelly do not discuss microbe property prediction, even though their graph convolution network could be directly applied to the application. While we do not perform any experimental comparisons in this work, we expect the results to be similar, with any performance differences due to choices such as weight-sharing, which will be task-dependent.

Other work has explored the use of deep learning for microbiome analysis, but without incorporating phylogenetic information. Ditzler, *et al.* 2015 [27] propose the use of static, recursive tree architectures to predict microbiome-level properties. Oh and Zhang 2020 [28] use fully-connected autoencoders to perform dimensionality reduction, an approach that could be used for transfer learning. However, without leveraging phylogenetic information as a source of inductive bias, pre-trained models that take ASV-level inputs are less-likely to transfer well to datasets with different ASVs.

III. Methods

A. Static EINNs

A frequent application of microbiome data analysis is to predict some property associated with the microbiome as a whole, such as sample provenance, host phenotype, or the stability of the composition. In this scenario an EINN model will be a feed-forward neural network organized as a tree, with each leaf node input associated with a particular microbe (ASV, or alternatively OTU, strain, etc.), and a prediction of the target variable is read-out at the root node. Intermediate nodes in the tree can be single neurons, but more typically would be vector-valued functions represented by one or more fully-connected layers. In this work each intermediate node is represented by a single fully-connected neural network layer. Here we assume a tree architecture, but this is not required in general, and non-tree architectures could be used to account for phylogenetic uncertainty or lateral gene transfer.

The particular choice of phylogenetic tree will impact both inductive bias and computational efficiency. In practice, shallower neural networks are faster to train and easier to optimize, and in general we expect trees that are roughly

balanced in terms of height, degree, and microbes to have better inductive bias. In our experiments with static EINNs, we start from a simple taxonomic tree in which most of the differentiation takes place at the lower levels (ASV, strain, and species), then reduce the tree height by compressing the higher levels (domain, kingdom, etc.) in order to achieve a more balanced tree, treating the compressed tree height as a hyperparameter to be optimized on a validation set. We leave further investigation of the effects of tree balancing for future work.

B. Dynamic EINNs

Dynamic EINNs are proposed as a method for predicting properties of microbes from features of related microbes. Here the tree architecture changes with each input, which can be accomplished using weight sharing (across every edge in the same layer) or recursive architectures (across every edge in the entire tree). We demonstrate the latter, using a single neural network layer repeated for each node in the dynamic tree architecture. The architecture when predicting a trait for a target microbe will be the same phylogenetic tree graph rearranged so that the target microbe is the root, with all other microbes at the leaves. Fig. 2 shows this for a simple tree, with the root of the phylogenetic tree removed in the dynamic EINN for compactness.

Such an architecture requires a node to represent a function of a variably-sized *set* of inputs. This can be accomplished by mean-pooling over the vector-valued representations of the set elements [9], [29]. In the model used for experiments, each intermediate node of the dynamic EINN takes a set of input representations $\mathbf{x}_k$ and outputs a hidden representation $\mathbf{z}$, with

$$\bar{\mathbf{x}} = \frac{1}{|K|} \sum_{k=1..K} \mathbf{x}_k \tag{1}$$

$$\mathbf{z} = \tanh(W \cdot \bar{\mathbf{x}} + b) \tag{2}$$

where K is the number of inputs, b is a bias term, and the hyperbolic tangent non-linearity is applied element-wise to the vector-value input. At the final layer (the root node), a special readout layer is used to predict the target variable.

We found this architecture to be quite slow when using the full tree, as it is harder to exploit parallelism in dynamic graph neural networks [30]. In experiments we greatly simplified the model by restricting the input to only the k nearest microbes, where k was treated as a hyperparameter and optimized over the range [1,50]. Thus a microbe trait is predicted by first finding the k nearest microbe neighbors according to the graph distance in the phylogenetic tree, constructing the dynamic EINN for this subgraph, then forward-propagating the features of the k microbes through this EINN to the root (output) node.

C. Datasets

Two datasets of environmental samples were used to evaluate the performance of static EINNs for sample-level prediction tasks. The EMP dataset contains 27,751 microbiome samples from 97 independent studies capturing diverse environmental types, geographies, and chemistries [4], [5]. The Waimea dataset [31] contains 1,513 samples from a study conducted at the Waimea Valley on the island of Oahu, Hawaii, with samples taken at regular intervals from the mountainous ridge at the top to the coral reef in the bay below. Amplicon sequencing of the 16S rRNA gene provide sequences for each sample, resulting in a combined total of 307,572 unique 90-basepair sequences from the two datasets. Samples in both the EMP and Waimea datasets are then represented as a 307,572 dimensional vector of ASV counts. All samples are rarefied to 5,000 reads by bootstrap sampling with replacement across the ASV vector. Each sample in both datasets is annotated with metadata regarding the source of the sample, and five of these properties were selected as targets for classification: habitat, environmental biome, environmental feature, environmental material, and sample type. Descriptions of these variables are given in Table I.

TABLE I
CLASSIFICATION TASKS

Task	Description	N Classes (EMP/Waimea)
Habitat	Global ecological context of a sample (eg. Marine, Terrestrial)	13 / 4
Biome	Broad ecological context of a sample (eg. freshwater, desert, woodland)	43 / 3
Feature	Local ecological context of a sample (eg. harbor, sandy beach, cliff)	97 / 11
Material	General material displaced by the sample (eg. air, water, soil)	45 / 8
Sample Type	Specific material displaced by the sample (gill tissue, bird egg shell)	120 / 54

A separate evaluation of dynamic EINNs was conducted using a microbe trait dataset. This dataset contains 14,887 microbes and (incomplete) labels for 330 microbial traits including metabolism type, genome size, and cell shape. We focus on the metabolism type feature, for which the dataset contains labels (one of six classes) for 9,869 microbes. Experiments with dynamic EINNs are conducted on this subset and classification task.

IV. Experimental Results

A. EINNs for Microbiome Classification

The five classification tasks were used to compare EINNs against four other machine learning classifiers: LASSO classification (LC), Ridge Classification (RC), Random Forests (RFs), and fully-connected neural networks (FCNNs). For each experiment the data was split 60/20/20% into train/validation/test sets, and the hyperparameters in Table II were optimized using the validation set before final evaluation on the test set. Hyperparameters for LC, RC, and RF models were obtained by using grid search, while hyperparameters for deep learning models were obtained using a random search of 50 trials using SHERPA on a GPU cluster [32].

The deep learning models were optimized using Adam [33]. Bootstrap sampling of the ASV counts was used as a form

of data augmentation during training — rather than rarefying the relative abundance profiles as a pre-processing step, we sampled (with replacement) 5,000 ASVs from the empirical distribution of each sample at each training iteration. At test time for each model, the data was rarefied five times in the same way in order to expand the size of the test set and obtain a better measure of performance. Training was stopped when no improvement on the validation set was observed over 10 epochs (passes through the training set). In all tasks, EINNs were either the first or second best performing method (Table III).

TABLE II
HYPERPARAMETER SEARCH SPACE

Model	Parameter	Min	Max
LC	L1 factor	10^{-5}	1
RC	L2 factor	10^{-5}	1
RF	Max depth	5	50
FCNN/EINN	Learning rate	10^{-5}	10^{-2}
	Batch size	16	128
	Dropout rate	0	0.5
	N neurons	256	2048
	N layers	1	6

TABLE III
EMP CLASSIFICATION TEST PERFORMANCE

Task	LC	RC	RF	FCNN	EINN
Habitat					
NLL	0.023	0.023	0.042	**0.010**	0.014
ACC	0.996	0.996	0.98.7	**0.997**	**0.997**
Biome					
NLL	0.243	0.248	0.498	0.237	**0.207**
ACC	0.935	0.935	0.888	0.928	**0.937**
Feature					
NLL	0.267	0.251	0.524	0.225	**0.195**
ACC	0.941	0.939	0.904	**0.950**	0.941
Material					
NLL	0.265	0.251	0.436	0.291	**0.208**
ACC	0.934	0.934	0.904	0.921	**0.937**
Sample					
NLL	0.340	0.321	0.568	0.309	**0.277**
ACC	0.920	**0.921**	0.882	0.918	0.920

B. Transfer Learning with EINNs

The deep representations learned by neural networks such as EINNs can be transferred between datasets, enabling us to leverage large data sets for improving predictions on smaller data sets. Here, we demonstrate this ability by pre-training EINNs on the EMP dataset then fine-tuning the model for the same classification task on the much smaller Waimea dataset. The transfer is performed by replacing the final softmax layer of the EINN and fine-tuning on the target dataset. The transferred EINN (T-EINN) is compared to the same set of models from Table III, using the same hyperparameter optimization procedure, except using 5-fold nested cross-validation instead of a single 60%/20%/20% split. Table IV shows the mean NLL and ACC over all cross validation splits. As expected, the linear models are more competitive on this smaller dataset. The transfer learning improves performance for two out of five tasks, and is only slightly worse than the best method in the other tasks.

TABLE IV
WAIMEA CLASSIFICATION TEST PERFORMANCE

Task	LC	RC	RF	FCNN	EINN	T-EINN
Habitat						
NLL	**0.142**	0.149	0.278	0.237	0.162	0.210
ACC	0.963	**0.968**	0.943	0.943	0.966	0.960
Biome						
NLL	0.178	0.185	0.259	0.162	**0.119**	0.147
ACC	0.959	0.957	0.931	0.968	**0.970**	0.967
Feature						
NLL	0.429	0.493	0.644	0.853	0.406	**0.395**
ACC	**0.897**	0.875	0.781	0.863	0.877	0.877
Material						
NLL	0.265	0.404	0.410	0.336	0.230	**0.203**
ACC	**0.951**	0.915	0.891	0.938	0.927	0.930
Sample						
NLL	1.120	1.189	1.694	1.330	**1.085**	1.136
ACC	**0.718**	0.693	0.626	0.678	0.708	0.698

C. Dynamic EINNs for Microbe Trait Imputation

Functional traits of microbes tend to be phylogenetically conserved, particularly at the shallowest branches of the tree [34], [35]. Thus, accounting for phylogenetic information is likely to help improve the performance of statistical models for inferring traits. We evaluate Dynamic EINNs and NN benchmarks on imputing microbial metabolism, a six-class classification task, given 324 microbial properties as input features. Table V shows the test set performance for this task in terms of negative log-likelihood (NLL) and accuracy (ACC).

To further investigate the benefit of the EINN architecture, we added explicit taxonomic information to each input node as a vector of engineered features. This was a one-hot vector for each taxonomic level excluding the species level, describing the location of the microbe with the tree. When including these engineered features, the total number of features per microbe was 4,040. Interestingly, the best performance was obtained by a Dynamic EINN trained with these features, suggesting that even though the dynamic EINN had phylogenetic information contained in the architecture, it still benefited from having the taxonomic features explicitly encoded.

TABLE V
MICROBE TRAIT CLASSIFICATION TEST PERFORMANCE

	Metabolism	
Model	NLL	ACC
FCNN	0.992	0.655
FCNN + Taxonomy Encoding	0.737	0.747
EINN	0.643	0.840
EINN + Taxonomy Encoding	0.469	0.932

V. CONCLUSION

The proposed Evolution-Informed Neural Network framework provides a general strategy for incorporating phylogenetic information into deep neural networks. We demonstrate

two different applications of EINNs: predicting properties of microbiomes and predicting microbe traits, the latter using a dynamic neural architecture. For both types of EINNs, we provide efficient software implementations that use sparse matrix multiplications. Furthermore, we demonstrate the use of transfer learning from the large EMP dataset to a smaller dataset.

Our experimental results show that this deep learning approach works best on large datasets with tens of thousands of samples, much larger than most microbiome studies. However, the results show a clear improvement over fully-connected architectures, suggesting that the phylogenetic information provides useful inductive bias, in agreement with previous studies. Compared with previous work, EINNs provide a more flexible and computationally efficient approach to incorporating phylogenetic information.

There are many details of EINN design that are left for future work. One important strategy to be explored is weight-sharing, where the nodes in a single layer could be constrained to share weights (or partially share weights) with every other node in the same layer, a technique that has provided extremely useful inductive biases in computer vision [36]. Previous work with graph convolutional networks [25] is a special case of EINNs where strict weight-sharing is enforced. Other details to be explored are the choice of phylogenetic tree, the importance of tree balance (in terms of height, degree, and/or samples), and how these choices affect transfer learning.

ACKNOWLEDGMENT

This work was supported by NSF 2124922. The authors thank Mahdi Belcaid for many helpful discussions about this work. The technical support and advanced computing resources from the University of Hawaii Information Technology Services – Cyberinfrastructure are gratefully acknowledged.

REFERENCES

[1] H. Li, "Microbiome, metagenomics, and high-dimensional compositional data analysis," *Annual Review of Statistics and Its Application*, vol. 2, pp. 73–94, 2015.

[2] Y.-H. Zhou and P. Gallins, "A review and tutorial of machine learning methods for microbiome host trait prediction," *Frontiers in Genetics*, vol. 10, p. 579, 2019.

[3] L. J. Marcos-Zambrano, K. Karaduzovic-Hadziabdic, T. Loncar Turukalo, P. Przymus, V. Trajkovik, O. Aasmets, M. Berland, A. Gruca, J. Hasic, K. Hron *et al.*, "Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment," *Frontiers in microbiology*, vol. 12, p. 313, 2021.

[4] J. A. Gilbert, J. K. Jansson, and R. Knight, "The earth microbiome project: successes and aspirations," *BMC biology*, vol. 12, no. 1, pp. 1–4, 2014.

[5] L. R. Thompson, J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann *et al.*, "A communal catalogue reveals earth's multiscale microbial diversity," *Nature*, vol. 551, no. 7681, pp. 457–463, 2017.

[6] B. A. Methé, K. E. Nelson, M. Pop, H. H. Creasy, M. G. Giglio, C. Huttenhower, D. Gevers, J. F. Petrosino, S. Abubucker, J. H. Badger *et al.*, "A framework for human microbiome research," *nature*, vol. 486, no. 7402, p. 215, 2012.

[7] C. Huttenhower, D. Gevers, R. Knight, S. Abubucker, J. H. Badger, A. T. Chinwalla, H. H. Creasy, A. M. Earl, M. G. FitzGerald, R. S. Fulton *et al.*, "Structure, function and diversity of the healthy human microbiome," *nature*, vol. 486, no. 7402, p. 207, 2012.

[8] A. Lusci, G. Pollastri, and P. Baldi, "Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules," *Journal of chemical information and modeling*, vol. 53, no. 7, pp. 1563–1575, 2013.

[9] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper/2015/file/f9be311e65d81a9ad8150a60844bb94c-Paper.pdf

[10] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[11] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[12] T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine, "Enforcing analytic constraints in neural networks emulating physical systems," *Physical Review Letters*, vol. 126, no. 9, p. 098302, 2021.

[13] A. C. Martiny, K. Treseder, and G. Pusch, "Phylogenetic conservatism of functional traits in microorganisms," *The ISME journal*, vol. 7, no. 4, pp. 830–838, 2013.

[14] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.

[15] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.

[16] P. Trivedi, J. E. Leach, S. G. Tringe, T. Sa, and B. K. Singh, "Plant–microbiome interactions: From community assembly to plant health," *Nature Reviews Microbiology*, vol. 18, no. 11, pp. 607–621, 2020.

[17] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *arXiv preprint arXiv:1411.1792*, 2014.

[18] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[20] M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai, T. Xiao, T. He, G. Karypis, J. Li, and Z. Zhang, "Deep graph library: A graph-centric, highly-performant package for graph neural networks," *arXiv preprint arXiv:1909.01315*, 2019.

[21] D. Reiman, A. Metwally, and Y. Dai, "Using convolutional neural networks to explore the microbiome," in *2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2017, pp. 4269–4272.

[22] D. Fioravanti, Y. Giarratano, V. Maggio, C. Agostinelli, M. Chierici, G. Jurman, and C. Furlanello, "Phylogenetic convolutional neural networks in metagenomics," *BMC bioinformatics*, vol. 19, no. 2, pp. 1–13, 2018. [Online]. Available: https://bmcbioinformatics.biomedcentral.com/track/pdf/10.1186/s12859-018-2033-5.pdf

[23] J. T. Wassan, H. Wang, F. Browne, and H. Zheng, "Paam-ml: a novel phylogeny and abundance aware machine learning modelling approach

for microbiome classification," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 44–49.

[24] T. Manning, J. T. Wassan, C. Palu, H. Wang, F. Browne, H. Zheng, B. Kelly, and P. Walsh, "Phylogeny-aware deep 1-dimensional convolutional neural network for the classification of metagenomes," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 1826–1831.

[25] S. Khan and L. Kelly, "Multiclass disease classification from microbial whole-community metagenomes," in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2020*. World Scientific, 2019, pp. 55–66. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7120658/pdf/nihms-1061135.pdf

[26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations (ICLR)*, 2017.

[27] G. Ditzler, R. Polikar, and G. Rosen, "Multi-layer and recursive neural networks for metagenomic classification," *IEEE transactions on nanobioscience*, vol. 14, no. 6, pp. 608–616, 2015.

[28] M. Oh and L. Zhang, "Deepmicro: deep representation learning for disease prediction based on microbiome data," *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020. [Online]. Available: https://www.nature.com/articles/s41598-020-63159-5.pdf

[29] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, "Deep sets," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf

[30] M. Looks, M. Herreshoff, D. Hutchins, and P. Norvig, "Deep learning with dynamic computation graphs," in *International Conference on Learning Representations (ICLR)*, 2017.

[31] S. Swift, M. Belcaid, N. Centraro, J. Darcy, K. Frank, K. Fraiola, T. McDermott, M. McFall-Ngai, C. Mora, M. Medeiros, K. Nakayama, C. Nelson, N. Nguyen, R. Rollins, P. Sadowski, W. Sparagon, M. Tefit, J. Yew, N. Hynson, J. Buchanan, D. Yogi, and K. Kajihara, "A ridge-to-reef ecosystem microbial census characterizes environmental reservoirs for animal and plant microbiomes," *Science (submitted)*, 2021.

[32] L. Hertel, J. Collado, P. Sadowski, J. Ott, and P. Baldi, "Sherpa: Robust hyperparameter optimization for machine learning," *SoftwareX*, vol. 12, p. 100591, 2020, arXiv:2005.04048 [cs.LG]. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352711020303046

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[34] M. Saifuddin, J. M. Bhatnagar, D. Segrè, and A. C. Finzi, "Microbial carbon use efficiency predicted from genome-scale metabolic models," *Nature communications*, vol. 10, no. 1, pp. 1–10, 2019.

[35] J. B. Martiny, S. E. Jones, J. T. Lennon, and A. C. Martiny, "Microbiomes in light of traits: a phylogenetic perspective," *Science*, vol. 350, no. 6261, 2015.

[36] G. Elsayed, P. Ramachandran, J. Shlens, and S. Kornblith, "Revisiting spatial invariance with low-rank local connectivity," in *International Conference on Machine Learning*. PMLR, 2020, pp. 2868–2879.