AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

Revealing General Patterns of Microbiomes That Transcend Systems: Potential and Challenges of Deep Transfer Learning

Maude M. David,[a,b] Christine Tataru,[a] Quintin Pope,[c] Lydia J. Baker,[a] Mary K. English,[a] Hannah E. Epstein,[a] Austin Hammer,[a] Michael Kent,[a] Michael J. Sieler Jr,[a] Ryan S. Mueller,[a] Thomas J. Sharpton,[a,d] Fiona Tomas,[e] Rebecca Vega Thurber,[a] Xiaoli Z. Fern[c]

[a]Department of Microbiology, Oregon State University, Corvallis, Oregon, USA
[b]Department of Pharmaceutical Sciences, Oregon State University, Corvallis, Oregon, USA
[c]School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, Oregon, USA
[d]Department of Statistics, Oregon State University, Corvallis, Oregon, USA
[e]Instituto Mediterráneo de Estudios Avanzados, IMEDEA, Esporles, Balearic Islands, Spain

ABSTRACT A growing body of research has established that the microbiome can mediate the dynamics and functional capacities of diverse biological systems. Yet, we understand little about what governs the response of these microbial communities to host or environmental changes. Most efforts to model microbiomes focus on defining the relationships between the microbiome, host, and environmental features within a specified study system and therefore fail to capture those that may be evident across multiple systems. In parallel with these developments in microbiome research, computer scientists have developed a variety of machine learning tools that can identify subtle, but informative, patterns from complex data. Here, we recommend using deep transfer learning to resolve microbiome patterns that transcend study systems. By leveraging diverse public data sets in an unsupervised way, such models can learn contextual relationships between features and build on those patterns to perform subsequent tasks (e.g., classification) within specific biological contexts.

KEYWORDS deep learning, embeddings, machine learning, microbial ecology

It is now apparent that microbiomes frequently mediate the dynamics and functional capacities of environmental and biological systems and in so doing can impact system health and homeostasis (1). Diverse sources of biotic and abiotic variation can affect these microbial communities, which in turn can reciprocally impact the health and functioning of their host or habitat. In this era of rapid, human-induced environmental change, our ability to manage biological and environmental systems may very well depend upon our understanding of how microbes interact with one another as well as with their host or environment and how exogenous variation (i.e., variation external to the host or environment microbiome itself) impacts these interactions. To propel this understanding, microbiologists have developed technologies that catalog and quantify measurable features of the microbiota, their host, and the environment. From environmental DNA sequencing to metabolomics, microbial ecologists increasingly generate, analyze, and integrate massive multi-omic data sets to characterize communities, understand the mechanisms they use to interact and drive their environment, and determine how these community components respond to environmental changes (2). In doing so, microbial ecologists often focus solely on their specific data sets and do not leverage the vast repository of publicly available microbial data to extract complex but generalizable patterns in relation to specific scientific questions.

Researchers currently apply powerful statistical and machine learning methods to a specific microbiome data set to identify microbial features (i.e., the measurable

TABLE 1 Lexicon of key terms and their definitions

Term	Definition
Machine learning	Data analytics methods that use a variety of algorithms learning from available data to optimize the parameters of models, which are often predictive of classes if used in a supervised context.
Deep learning	A class of machine learning algorithms originally inspired by the brain. Deep models have multiple layers that sequentially process inputs, have high parameter counts, and use large amounts of data to train. In training, deep models learn to extract high-level features that are useful for both the trained model and potentially other models down the analytical pipeline.
Pretrained models	Models trained on a "pretraining" objective for which there is a very large amount of data. Such models can then be adapted to a "downstream" objective for which there are far fewer data and are often able to leverage information learned from the pretraining data to more easily perform the downstream objective.
Transformer models	A specific type of deep learning model originally proposed for language but later applied to many other domains, including vision, graphs, and sets.
Embeddings	Higher-level features derived from deep learning algorithms. They can efficiently represent statistical patterns in microbial communities, aiding in data processing for such models.
Transfer learning	The process of training a model on one task/data set and then fine-tuning it to perform a different task on a different data set is known as transfer learning.
Unlabeled data	Use of microbiome-related data (may be of different data types) without requirement for specific metadata associated with a given prediction task.
Language model	Models based on algorithms originally developed by the field of natural language processing, trained to model natural language and estimate the probability of a given word appearing in a certain context.
Task-specific model	Models developed to predict a specific set of features or outcomes and almost always trained on a prediction task with labeled data.

variables of a system: taxa, genes, metabolites) that stratify groups of samples or that explain the variation of a continuous covariate across samples (2–7). Traditional methods may also use summary metrics like beta-dispersion and alpha- and beta-diversity to characterize microbial community patterns (8). However, neither the single feature approach nor the community summary statistic approach incorporates patterns across studies or systems, nor do they take into account the interdependencies between features. Thus, they will fail to capture basic principles governing how microbial communities assemble, diversify, and respond to environmental variation.

Currently, meta-analyses are used to apply these traditional analysis approaches across studies. However, while meta-analyses of microbiome data sets generally support relationships between microbiome changes in community structure or function and disease observed in individual studies (9, 10), these interactions are often relatively weak and confounded by interstudy and individual microbiome variation (11–14). This may be due to a limitation in the traditional analysis methods, which treat each study independently and consequently may miss key insights into a system's more generalizable biological properties. To overcome these limitations, prior work has sought to robustly link specific microbiome features to outcomes and phenotypes across studies (11, 15). These studies go beyond traditional meta-analyses by identifying generalizable properties of microbial communities using network analysis or by accounting for noise across multiple data sets when developing predictive models. Such work shows that (i) there are generalizable patterns among diverse microbiomes that could be better exploited and (ii) many of these patterns or structures may be undetectable within individual studies due to the contextual nature of the patterns and limitation of the study sample size.

We believe deep transfer learning can expand on these approaches and is particularly well suited to generalize patterns of microbial ecology across studies and biological systems. We and others are leveraging deep learning methods (Table 1) both to

capture microbe-microbe and microbe-environment interactions across systems and to generate models and transformed features informed by aggregations of all available data. Deep learning approaches are specifically designed to analyze diverse, high-dimensional data in ways that can detect complex patterns of association between multiple features and covariates of interest. As a result, deep learning approaches can discover associations between a feature and a covariate that are contextually dependent upon other features in the system. Moreover, deep neural networks have proved capable of transfer learning, the process of learning generalizable patterns from diverse data and then fine-tuning that general model on study- or system-specific tasks (16–18). By developing deep learning approaches that resolve shared patterns across studies, researchers can draw novel insights about well-studied communities and aid discovery in understudied environments.

In particular, we advocate using deep learning models that have already proven adept at transfer learning in natural language processing (NLP) to better understand baseline interactions present in microbiome data (19–21). There exist easily drawn parallels between natural language data and microbiome data, namely, that documents are equivalent to biological samples, words to taxa, and topics to microbial neighborhoods (22–24). While other language-inspired algorithms such as topic modeling methods have been employed to identify latent variables in microbiomes, deep learning approaches offer the unique advantage that they scale with the amount of data available, allowing our understanding and our predictive models to scale alongside the genomic revolution (25).

One simple NLP algorithm is embedding, where a low-dimensional space is learned that preserves information about the cooccurrences between features. Embedding-based analysis of gut microbiomes has previously enabled a more accurate and generalizable differentiation between microbiomes associated with inflammatory bowel disease and healthy gut microbiomes than does analysis based on individual taxon counts (26). It was also observed that a word embedding algorithm applied to 16S k-mers resulted in meaningful numeric feature representations that both bolstered downstream classification performance and offered insight into the correspondence of microbial taxa to particular body sites (27).

Another powerful approach is to pretrain a model using "self-supervised" learning strategies on a large, diverse corpus (21, 28). In natural language, this can be achieved by (i) blocking out words from the input text and asking the model to recreate the hidden portion or (ii) asking the model to generate the next word in the sentence. This forces the model to identify properties in the input data that are useful for inferring the missing information and can lead the model to learn linguistic and semantic patterns that govern the composition of natural text beyond simple word cooccurrences (e.g., grammatical rules, semantic relationships, and other statistical patterns that aid in understanding text). Analogously, self-supervised learning can be applied to learn a "language" model of the microbiome, which captures the interactions among different microbial species or metabolic processes and can be used to understand shared composition rules of microbial communities. The pretrained language models can encode each word, or measured bacterial feature, into an embedding, which can subsequently be used for supervised prediction tasks (to differentiate two systems, or the same perturbation across two systems, for example) (Fig. 1) with greatly reduced dimension, hence reducing the risk of overfitting. These models offer distinct advantages over conventional methods because they use large data sets of bacterial features from diverse species and environments. Such models may be used to generalize across microbial systems to reduce input dimensionality and find useful latent properties in microbial communities.

There are important distinctions between natural language and microbiome data. Language has a natural sequential structure that is not present in most omics-based microbiome data. Comparatively, microbiome data also has rich biological information about taxa and functions that helps generate ecological and evolution inferences into
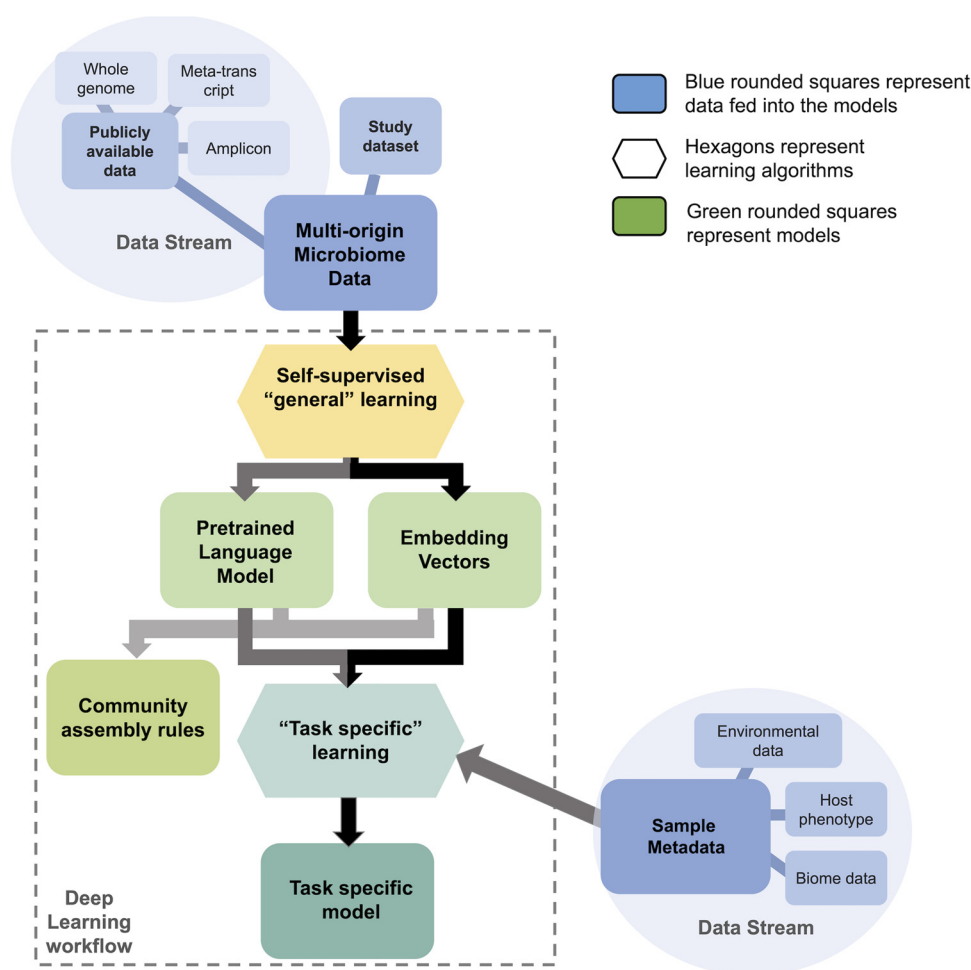
FIG 1 Workflow of microbial data transformation using deep learning approaches used to capture cross-system properties and apply to a specific task. Hexagons represent algorithms, fed data (in blue) and producing models (squares). Measured microbial features from publicly available data sets are input into the deep learning algorithms (in yellow) to produce pretrained models constituting new features (in green). During pretraining, a model is trained to reconstruct original input from distorted or partial input, using unlabeled data. The process outputs a trained model and/or a set of feature embeddings (which may be interpreted to represent community assembly rules and microbial ecology principles). The general model may then be fine-tuned on metadata (study data set in blue) to answer a biome- or system-specific question of interest. The process of training a model on one task/data set and then fine-tuning it on another task/data set is known as transfer learning.

whole-system dynamics. Such distinctions offer exciting opportunities for new innovations to adapt NLP models to microbiome data, similar to what has been done in computer vision (29) and learning with graphs (30) and sets (31) (Table 2).

While the application of deep learning to study microbiomes holds great promise, there exist significant challenges. Deep learning requires large amounts of data. For example, BERT, a commonly used pretrained language model for natural text, was trained on 3.3 billion words (23). While the scientific community has been generating microbiome data exponentially, these data are often not publicly available or remain unprocessed or uncurated. For example, while NCBI counts around 690 million sequence records (as of August 2021), the vast majority of records do not contain the metadata information sufficient to include in training data sets. Conversely, well-curated databases often lack sufficient depth of records. For example, databases like Qiita, which aim to enable cross-system meta-analyses of omics data, contain on the order of tens of millions of individual sequence records (e.g., amplicon sequence variants [ASVs] from 16S amplicon studies) from a few hundred thousand (160,000 to 200,000)

TABLE 2 Nonexhaustive list of relevant examples of deep learning algorithms and their ecological representations, when applicable[a]

Deep learning algorithm	Description	Application to microbial ecology	Papers
Autoencoder	Condense a long vector of input features into a dense mathematical space and then validate by regenerating the input from the condensed space.	16S amplicon, gene, metabolite, or protein counts can be condensed, thereby defining latent variables that drive observed feature counts. Ecological representation: latent variables may represent environmental factors like nutrient availability or pH that dictate observed microbial features.	"Using autoencoders for predicting latent microbiome community shifts responding to dietary changes" (37); "DeepGeni: deep generalized interpretable autoencoder elucidates gut microbiota for better cancer immunotherapy" (38); "Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via Long Short-Term Memory networks" (39); "Predicting microbiomes through a deep latent space" (40); "DeepMicro: deep representation learning for disease prediction based on microbiome data" (41)
Convolutional neural networks	A class of models that use convolution kernels or filters that slide along input features to learn features between input segments with a given spatial relationship	Integrate spatial information about relative and global microbial locations within a system. Interpret nucleotide text. Ecological representation: metabolic relationships dependent on location within a system. Motif function dependent on location relative to other motifs.	"TaxoNN: ensemble of neural networks on stratified microbiome data for disease prediction" (42); "Learning, visualizing and exploring 16S rRNA structure using an attention-based deep neural network" (43)
Long short-term memory (LSTM)	A class of deep neural networks that process data sequentially. They have a memory that updates every time the network processes a sequence entry.	Learn patterns from microbial genetic sequence (e.g., shotgun metagenomics or transcriptomics) which hold information about gene function. Represent longitudinal count data and changing microbiome dynamics along a nutrient or time gradient. Ecological representation: gene clusters of interreliant genes. Changes in specific taxon abundances over time are indicative of system state.	**"Transfer learning improves antibiotic resistance class prediction"** (33); "Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via Long Short-Term Memory networks" (39); **"Shedding light on microbial dark matter with a universal language of life"** (34)
Embedding	Learn word vectors so the inner product of vectors i and j reflects the cooccurrence probability for words i and j. Specific example: GloVe.	Consider taxon or any microbial feature cooccurrences. Ecological representation: metabolic corporations or partnership.	"GloVe: global vectors for word representation" (44); **"Decoding the language of microbiomes using word-embedding techniques, and applications in inflammatory bowel disease"** (26)
Continuous bag of words	Generate one static embedding per word. Training algorithm captures semantic and lexical information of each word based on that word's context—i.e., its neighboring set of words. Specific example: Word2vec.	k-mer representations of sequences could be embedded in such a way that their context is preserved. Ecological representation: conserved motifs (k-mers) represent evolutionary and possibly functional similarities across sequences.	Word2vec (45); "Learning representations of microbe-metabolite interactions" (46)
Transformer	Guess the next word or part of a sentence based on all the words present in a sentence. Mask input text tokens, then train the model to predict original tokens from unmasked context. Distort text by replacing input text tokens with fake but plausible substitutions. Then train model to identify the fake tokens based on the surrounding context. Specific examples: BERT, ELECTRA	Provide a view of a sample by considering all neighbors. Identify key microbial features association within samples. Ecological representation: community assembly and/or metabolic cooperation at a higher degree.	"BERT: pretraining of deep bidirectional transformers for language understanding" (21); "ELECTRA: pretraining text encoders as discriminators rather than generators" (47)

[a]Papers highlighted in bold engage in transfer learning where an independent pretraining task and data set are used to build a model, which is then fine-tuned on a different prediction task and data set.

samples of various omics types (32). Because of this, transfer learning in the microbial space thus far has largely focused on using pretraining tasks that do not require curation, such as predicting the next segment of genetic sequence or the next amino acid of a protein sequence, from minimally processed reads (33, 34). Examples of transfer learning using predefined microbial features (e.g., ASVs, genes, or metabolites) are rare. However, our first attempt determined that approximately equal numbers of features (ASVs) and samples were sufficient to capture meaningful information in embedding vectors while drastically reducing the dimensionality from ~23,000 ASVs to 100 dimensions (26).

Given the complexity and noisiness of biological systems, integration of different omics data representing different processes will be key in the development of more generalizable models that transcend individual systems. Accordingly, understanding the biological relationships between these data types and what limitations and insights they provide into biological processes is a large and important challenge. The scientific community needs to (i) ensure metadata is available alongside omics; (ii) adopt a common ontology for metadata type and sequencing data type, such as those being developed by the National Microbiome Data Collaboration (35); and (iii) allow universal and easy access to those data. There is an urgent need to organize data catalogs, or at least to categorize any new data generated during published research. These efforts should follow principles such as the FAIR data standards (36) of Findability, Accessibility, Interoperability, and Reuse of digital assets. In addition to requiring large amounts of data, deep models are often difficult for humans to understand. Although the learned patterns of neural network models can detect useful and general statistical patterns in microbial data, interpreting those models remains challenging. How can we extract information from properties or dimensions that humans have not conceptually defined yet? Work on interpreting neural network models is well under way in the field of computer science, and cross-disciplinary partnerships that leverage that work are becoming increasingly valuable, especially as the amount of microbiome omics data grows year over year.

To conclude, we call upon the scientific community to collaboratively accelerate these endeavors: the algorithms developed in various branches of computer science, including natural language processing and explainable artificial intelligence (AI), offer us unprecedented opportunities to learn from complex and massive data sets and gain transformative insight into the rules that govern how microbial communities assemble, diversify, and respond to environmental variation. This effort is particularly relevant in an age where our ability to synthesize data on model systems outstrips our ability to sample across every microbial system on Earth.

REFERENCES

1. Ruff WE, Greiling TM, Kriegel MA. 2020. Host-microbiota interactions in immune-mediated diseases. Nat Rev Microbiol 18:521–538. https://doi.org/10.1038/s41579-020-0367-2.

2. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall L-I, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG, Dorrestein PC. 2018. Best practices for analysing microbiomes. Nat Rev Microbiol 16:410–422. https://doi.org/10.1038/s41579-018-0029-9.

3. Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, Berland M, Gruca A, Hasic J, Hron K, Klammsteiner T, Kolev M, Lahti L, Lopes MB, Moreno V, Naskinova I, Org E, Paciência I, Papoutsoglou G, Shigdel R, Stres B, Vilne B, Yousef M, Zdravevski E, Tsamardinos I, Carrillo de Santa Pau E, Claesson MJ, Moreno-Indias I, Truu J.

2021. Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. Front Microbiol 12:634511. https://doi.org/10.3389/fmicb.2021.634511.

4. Zhou Y-H, Gallins P. 2019. A review and tutorial of machine learning methods for microbiome host trait prediction. Front Genet 10:579. https://doi.org/10.3389/fgene.2019.00579.

5. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, Tickle TL, Weingart G, Ren B, Schwager EH, Chatterjee S, Thompson KN, Wilkinson JE, Subramanian A, Lu Y, Waldron L, Paulson JN, Franzosa EA, Bravo HC, Huttenhower C. 2021. Multivariable association discovery in population-scale meta-omics studies. bioRxiv https://www.biorxiv.org/content/10.1101/2021.01.20.427420v1.

6. Flannery JE, Stagaman K, Burns AR, Hickey RJ, Roos LE, Giuliano RJ, Fisher PA, Sharpton TJ. 2020. Gut feelings begin in childhood: the gut

metagenome correlates with early environment, caregiving, and behavior. mBio 11:e02780-19. https://doi.org/10.1128/mBio.02780-19.

7. Vega Thurber R, Mydlarz LD, Brandt M, Harvell D, Weil E, Raymundo L, Willis BL, Langevin S, Tracy AM, Littman R, Kemp KM, Dawkins P, Prager KC, Garren M, Lamb J. 2020. Deciphering coral disease dynamics: integrating host, microbiome, and the changing environment. Front Ecol Evol 8:402. https://doi.org/10.3389/fevo.2020.575927.

8. Zaneveld JR, McMinds R, Vega Thurber R. 2017. Stress and stability: applying the Anna Karenina principle to animal microbiomes. Nat Microbiol 2:17121. https://doi.org/10.1038/nmicrobiol.2017.121.

9. Sharpton TJ, Stagaman K, Sieler MJ, Jr, Arnold HK, Davis EW, II. 2021. Phylogenetic integration reveals the zebrafish core microbiome and its sensitivity to environmental exposures. Toxics 9:10. https://doi.org/10.3390/toxics9010010.

10. Armour CR, Nayfach S, Pollard KS, Sharpton TJ. 2019. A metagenomic meta-analysis reveals functional signatures of health and disease in the human gut microbiome. mSystems 4:e00332-18. https://doi.org/10.1128/mSystems.00332-18.

11. Sze MA, Schloss PD. 2016. Looking for a signal in the noise: revisiting obesity and the microbiome. mBio 7:e01018-16. https://doi.org/10.1128/mBio.01018-16.

12. Holman DB, Gzyl KE. 2019. A meta-analysis of the bovine gastrointestinal tract microbiota. FEMS Microbiol Ecol 95:fiz072. https://doi.org/10.1093/femsec/fiz072.

13. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. 2017. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun 8:1784. https://doi.org/10.1038/s41467-017-01973-8.

14. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, Bork P, Sunagawa S, Zeller G. 2021. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol 22:93. https://doi.org/10.1186/s13059-021-02306-1.

15. Rodrigues RR, Gurung M, Li Z, García-Jaramillo M, Greer R, Gaulke C, Bauchinger F, You H, Pederson JW, Vasquez-Perez S, White KD, Frink B, Philmus B, Jump DB, Trinchieri G, Berry D, Sharpton TJ, Dzutsev A, Morgun A, Shulzhenko N. 2021. Transkingdom interactions between lactobacilli and hepatic mitochondria attenuate western diet-induced diabetes. Nat Commun 12:101. https://doi.org/10.1038/s41467-020-20313-x.

16. Hernandez D, Kaplan J, Henighan T, McCandlish S. 2021. Scaling laws for transfer. arXiv 2102.01293 [cs.LG]. https://arxiv.org/abs/2102.01293.

17. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. 2019. A comprehensive survey on transfer learning. arXiv 1911.02685 [cs.LG]. https://arxiv.org/abs/1911.02685.

18. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv 1910.10683 [cs.LG]. https://arxiv.org/abs/1910.10683.

19. Ruder S, Peters ME, Swayamdipta S, Wolf T. 2019. Transfer learning in natural language processing, p 15–18. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: tutorials. Association for Computational Linguistics, Minneapolis, MN. https://doi.org/10.18653/v1/N19-5004.

20. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. 2019. RoBERTa: a robustly optimized BERT pretraining approach. arXiv 1907.11692 [cs.CL]. https://arxiv.org/abs/1907.11692.

21. Devlin J, Chang M-W, Lee K, Toutanova K. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv 1810.04805 [cs.CL]. https://arxiv.org/abs/1810.04805.

22. Hosoda S, Nishijima S, Fukunaga T, Hattori M, Hamada M. 2020. Revealing the microbial assemblage structure in the human gut microbiome using latent Dirichlet allocation. Microbiome 8:95. https://doi.org/10.1186/s40168-020-00864-3.

23. Shenhav L, Furman O, Briscoe L, Thompson M, Silverman JD, Mizrahi I, Halperin E. 2019. Modeling the temporal dynamics of the gut microbial community in adults and infants. PLoS Comput Biol 15:e1006960. https://doi.org/10.1371/journal.pcbi.1006960.

24. Sankaran K, Holmes SP. 2019. Latent variable modeling for the microbiome. Biostatistics 20:599–614. https://doi.org/10.1093/biostatistics/kxy018.

25. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Hasan M, Van Essen BC, Awwal AAS, Asari VK. 2019. A state-of-the-art survey on deep learning theory and architectures. Electronics 8:292. https://doi.org/10.3390/electronics8030292.

26. Tataru CA, David MM. 2020. Decoding the language of microbiomes using word-embedding techniques, and applications in inflammatory bowel disease. PLoS Comput Biol 16:e1007859. https://doi.org/10.1371/journal.pcbi.1007859.

27. Woloszynek S, Zhao Z, Chen J, Rosen GL. 2019. 16S rRNA sequence embeddings: meaningful numeric feature representations of nucleotide sequences that are convenient for downstream analyses. PLoS Comput Biol 15:e1006721. https://doi.org/10.1371/journal.pcbi.1006721.

28. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. 2017. Attention is all you need. arXiv 1706.03762 [cs.CL]. https://arxiv.org/abs/1706.03762.

29. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. 2021. Transformers in vision: a survey. arXiv 2101.01169 [cs.CV]. https://arxiv.org/abs/2101.01169.

30. Yun S, Jeong M, Kim R, Kang J, Kim HJ. 2019. Graph transformer networks. Adv Neural Inf Proc Syst 32:11983–11993.

31. Lee J, Lee Y, Kim J, Kosiorek A, Choi S, Teh YW. 2019. Set transformer: a framework for attention-based permutation-invariant neural networks. Proc Mach Learn Res 97:3744–3753.

32. Gonzalez A, Navas-Molina JA, Kosciolek T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, Sanders JG, Shorenstein J, Holste H, Petrus S, Robbins-Pianka A, Brislawn CJ, Wang M, Rideout JR, Bolyen E, Dillon M, Caporaso JG, Dorrestein PC, Knight R. 2018. Qiita: rapid, web-enabled microbiome meta-analysis. Nat Methods 15:796–798. https://doi.org/10.1038/s41592-018-0141-9.

33. Hamid M-N, Friedberg I. 2020. Transfer learning improves antibiotic resistance class prediction. bioRxiv https://www.biorxiv.org/content/biorxiv/early/2020/04/18/2020.04.17.047316.full.pdf.

34. Hoarfrost A, Aptekmann A, Farfañuk G, Bromberg Y. 2020. Shedding light on microbial dark matter with a universal language of life. bioRxiv https://www.biorxiv.org/content/10.1101/2020.12.23.424215v2.full.

35. Wood-Charlson EM, Anubhav, Auberry D, Blanco H, Borkum MI, Corilo YE, Davenport KW, Deshpande S, Devarakonda R, Drake M, Duncan WD, Flynn MC, Hays D, Hu B, Huntemann M, Li P-E, Lipton M, Lo C-C, Millard D, Miller K, Piehowski PD, Purvine S, Reddy TBK, Shakya M, Sundaramurthi JC, Vangay P, Wei Y, Wilson BE, Canon S, Chain PSG, Fagnan K, Martin S, McCue LA, Mungall CJ, Mouncey NJ, Maxon ME, Eloe-Fadrosh EA. 2020. The National Microbiome Data Collaborative: enabling microbiome science. Nat Rev Microbiol 18:313–314. https://doi.org/10.1038/s41579-020-0377-0.

36. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18.

37. Reiman D, Dai Y. 2019. Using autoencoders for predicting latent microbiome community shifts responding to dietary changes, p 1884–1891. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, New York, NY.

38. Oh M, Zhang L. 2021. DeepGeni: deep generalized interpretable autoencoder elucidates gut microbiota for better cancer immunotherapy. bioRxiv https://www.biorxiv.org/content/10.1101/2021.05.06.443032v1.

39. Metwally AA, Yu PS, Reiman D, Dai Y, Finn PW, Perkins DL. 2019. Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via Long Short-Term Memory networks. PLoS Comput Biol 15:e1006693. https://doi.org/10.1371/journal.pcbi.1006693.

40. García-Jiménez B, Muñoz J, Cabello S, Medina J, Wilkinson MD. 2021. Predicting microbiomes through a deep latent space. Bioinformatics 37:1444–1451. https://doi.org/10.1093/bioinformatics/btaa971.

41. Oh M, Zhang L. 2020. DeepMicro: deep representation learning for disease prediction based on microbiome data. Sci Rep 10:6026. https://doi.org/10.1038/s41598-020-63159-5.

42. Sharma D, Paterson AD, Xu W. 2020. TaxoNN: ensemble of neural networks on stratified microbiome data for disease prediction. Bioinformatics 36:4544–4550. https://doi.org/10.1093/bioinformatics/btaa542.

43. Zhao Z, Woloszynek S, Agbavor F, Mell JC, Sokhansanj BA, Rosen GL. 2021. Learning, visualizing and exploring 16S rRNA structure using an

mSystems®

attention-based deep neural network. PLoS Comput Biol 17:e1009345. https://doi.org/10.1371/journal.pcbi.1009345.

44. Pennington J, Socher R, Manning C. 2014. GloVe: global vectors for word representation, p 1532–1543. *In* Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Minneapolis, MN.

45. Mikolov T, Chen K, Corrado G, Dean J. 2013. Efficient estimation of word representations in vector space. arXiv https://arxiv.org/abs/1301.3781.

46. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, Swenson TL, Van Goethem MW, Northen TR, Vazquez-Baeza Y, Wang M, Bokulich NA, Watters A, Song SJ, Bonneau R, Dorrestein PC, Knight R. 2019. Learning representations of microbe-metabolite interactions. Nat Methods 16:1306–1314. https://doi.org/10.1038/s41592-019-0616-3.

47. Clark K, Luong M-T, Le QV, Manning CD. 2020. ELECTRA: pre-training text encoders as discriminators rather than generators. arXiv https://arxiv.org/abs/2003.10555.