Supplementary Material

Oshrit Shtossel[1], Omry Koren[2], and Yoram Louzoun[1,*]

[1]Department of Mathematics, Bar-Ilan University, Ramat Gan 52900, Israel
[2]The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel
[*]Corresponding author: Yoram Louzoun, louzouy@math.biu.ac.il

June 29, 2024

1 ML Nomenclature

In order to facilitate the understanding of the more ML-oriented terms in the text, we here provide a short description of the main ML terms used in the manuscript.

- **Model.** A *model* is the mathematical relation between any input (in our case microbiome ASVs) and the appropriate output (in our case the class of the sample/the phenotype). In ML, the model usually contains a set of parameters called weights, and the model is trained by finding the weights for which the model best matches the relation between the input and output in the *training set*.
- **Training set.** The *training set* is the part of the data used to train the model. The quality of the fit between the input and output data on the training set is not a good measure of the quality of the model, since it may lead to *overfitting*.
- **Overfitting.** *Overfitting* is a problem occurring when a model produces good results on data in the training set (usually due to too many parameters) but produces poor results on unseen data.
- **Validation Set.** The *validation set* is a separate set from the training set that is used to monitor the training process, but is not used for the actual training. This set can be used to optimize some parts of the learning process, including setting the *hyperparameters*.
- **Model Hyperparameters.** *Model hyperparameters* are adjustable values that are not considered part of the model itself in that they are not updated during training, but they still have an impact on the training of the model and its performance. To ensure that these are not fitted to maximize the test set performance, the hyperparameters are optimized using an internal validation set.
- **10-Fold Cross-Validation** *10-Fold Cross-Validation* (referred to as 10 CVs) is a resampling procedure used to evaluate machine learning models on a limited data sample. The data is first partitioned into 10 equally (or nearly equally) sized segments or folds. Subsequently, 10 iterations of training and validation are performed such that within each iteration a different fold of the data is held out for validation while the remaining nine folds are used for training.

- **Receiver Operating Characteristic Curve (ROC).** The *Receiver Operating Characteristic (ROC) Curve* is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate (TPR = the probability that an actual positive will test positive) and False Positive Rate (FPR = the probability that an actual negative will test positive).
- **Area Under the ROC Curve (AUC)** The *Area Under the ROC Curve (AUC)* is a single scalar value that measures the overall performance of a binary classifier. The AUC value is within the range [0.5–1.0], where the minimum value represents the performance of a random classifier and the maximum value corresponds to a perfect classifier (e.g., with a classification error rate equivalent to zero). It measures the area under the ROC curve defined above.
- **FFT (Fast Fourier Transform).** *FFT (Fast Fourier Transform)* is a mathematical algorithm that converts complex signals (like biological data) from their original form into a set of simpler signals. Thus making it easier to analyze patterns and trends in biological data.
- **Low Pass Filter.** A *low pass filter* is a filter that allows signals with a frequency lower than a certain cutoff frequency to pass through and attenuates frequencies higher than the cutoff frequency. It is commonly used to remove high-frequency noise from a signal.
- **SAM (Spectral Angle Mapper) metric.** *SAM* determines the spectral similarity between given reference spectra and the spectra found at each pixel (in our case each tree-based image pixel). The result of the comparison is reported as the angular difference (in radians) between the two spectra.
- **MSE (Mean Square Error) metric.** *MSE metric* measures the average squared difference between two vectors or images, in our case between the two tree-based images.

2 Supplementary Tables

Table 1: Acronym table

Acronym	**Meaning**
ML	Machine learning
DAA	Differential Abundance Analysis
SOTA	State-Of-The-Art
ASV	Amplicon Sequence Variants
CLR	Centered Log-Ratio
RCLR	Robust Centered Log Ratio
GIMIC	Smoothed Graph IMages of the MICrobiome
iMic	iMage Microbiome
MSE	Mean Squared Error
SAM	Spectral Angle Mapper
GGMP	Guangdong Gut Microbiome Project
FFT	Fast Fourier Transform
AUC	Area under the ROC Curve
SCC	Spearman Correlation Coefficient
WGS	Whole-genome sequencing
UMAP	Uniform Manifold Approximation and Projection for Dimension Reduction
KNN	K Nearest Neighbors
PCA	Principal Component Analysis
CDI	Clostridioides Difficile Infection
IBD	Inflammatory Bowel Disease
IBS	Inflammatory Bowel Syndrom
BMI	Body Mass Index
NN	Neural Network

Table 2: Different metrics solutions to the challenges of the microbiome analysis

	Problems solved by method.
Bray Curtis	Simplicity - it is the simplest dissimilarity metric based on the counts ratio only.
Jaccard	Simplicity- it is the simplest dissimilarity metric based on the the presence vs absence of the taxa only.
Unweighted Unifrac and weighted Unifrac	Inner relations - gives a different weight to different taxa considering the phylogenetic structure.
PhyloRPCA	1. Inner relations - gives a different weight to different taxa considering the phylogenetic structure. 2. Sparsity - utilizes an expectation-maximization algorithm that iteratively imputes missing values based on the available information and updates the loadings accordingly. This allows for robust analysis even when there are missing values in the data. 3. Scales - applies the RCLR normalization to the values.
Aitchison	1. Scales - applies the RCLR normalization to the values. 2. Compositionality -
2D-log(composition) and 2D-SAM	1. Inner relations - gives a different weight to different taxa considering the phylogenetic structure by the Cladogram of means. 2. Sparsity - computes the Cladogram of means puts a non zero value to taxon's anscestors, since one of the taxon's sisters is likely to be non-zero. 3. Scales - applies log normalization to the values (Supp. Mat. Fig. S17). 4. Smoothing - builds microbial images based on the Cladogram of means (referred as 'image smoothing').
GIMIC	1. Inner relations - gives a different weight to different taxa considering the phylogenetic structure by the Cladogram of means. 2. Sparsity - computes the Cladogram of means puts a non zero value to taxon's anscestors, since one of the taxon's sisters is likely to be non-zero 3. Scales - applies log normalization to the values (Supp. Mat. Fig. S17). 4. Smoothing - * builds microbial images based on the Cladogram of means (referred to as 'image smoothing'). applies FFT with cutoff on the processed images, such that the high signals above the cutoff are removed from the image, (referred as 'FFT smoothing').

Table 3: Datasets details used during the analyses

Dataset name	16S-WGS	Samples	Condition	Origin	Reference
Allergy	16S	274	Milk Nut Peanut allergy	Stool	[1]
Cirrhosis	16S	130	Cirrhosis	Stool	[2]
Cirrhosis (WGS)	WGS	232	Cirrhosis	Stool	[3]
GC (WGS) (ERAWIJANTARI)	WGS	96	Gastrectomy gastric cancer	Stool	[4]
IBD-4 (ERP020401)	16S	684	IBD	Stool	ERP020401 in the NCBI
IBD-3 (ERP021216)	16S	86	IBD	Stool	[5]
GDM	16S	1607	GDM	Stool	[6]
IBD (WGS) (FRANZOSA)	WGS	220	IBD	Stool	[7]
GGMP	16S	7008	Many conditions	Stool	[8]
Breast feeding	16S	277	Breastfeeding vs formula	Stool	[9]
IBD-2	16S	257	IBD	Stool	[10]
IBD-5 (JACOB)	16S	90	IBD	Stool	[11]
CRC (KIM)	16S	240	CRC	Stool	[12]
MF	16S	190	Male vs. female	Stool	[2]
IBS (WGS) (MARS)	WGS	444	IBS	Stool	[13]
Nugent	16S	342	Binary nugent score	Vagina	[14]
Obesity (WGS)	WGS	253	Obesity	Stool	[3]
IBD-1 (OK94)	16S	132	CD vs. UC	Stool	[10]
CDI-1 (PRJNA353587)	16S	83	CDI	Stool	PRJNA353587 in the NCBI
CDI-2 (PRJNA412501)	16S	71	CDI	Stool	PRJNA412501 in the NCBI
CDI-3 (PRJNA419097)	16S	104	CDI	Stool	PRJNA419097 in the NCBI
T2D (WGS)	WGS	342	Diabetes-2	Stool	[3]
ESRD (WANG)	WGS	287	End-stage renal disease (ESRD)	Stool	[15]
WB	16S	200	White vs. black	Vagina	[14]
CRC (WGS) (YACHIDA)	WGS	347	CRC	Stool	[16]
GC-1 (PRJNA428883)	16S	746	Gastrectomy gastric cancer	Stool	PRJNA428883 in the NCBI
GC-2 (PRJNA778008)	16S	451	Gastrectomy gastric cancer	Stool	PRJNA778008 in the NCBI
GC-3 (PRJNA310127)	16S	291	Gastrectomy gastric cancer	Stool	PRJNA310127 in the NCBI
GC-4 (PRJNA561290)	16S	380	Gastrectomy gastric cancer	Stool	PRJNA561290 in the NCBI

GC-5 (PRJNA481413)	16S	445	Gastrectomy gastric cancer	Stool	PRJNA481413 in the NCBI
GC-6 (PRJNA532731)	16S	123	Gastrectomy gastric cancer	Stool	PRJNA532731 in the NCBI
IBS-1 (PRJNA46339)	16S	354	Irritable bowel syndrome	Stool	PRJNA46339 in the NCBI
IBS-2 (PRJNA604466)	16S	52	Irritable bowel syndrome	Stool	PRJNA604466 in the NCBI
IBS-3 (PRJNA391149)	16S	205	Irritable bowel syndrome	Stool	PRJNA391149 in the NCBI
IBS-4 (PRJDB5442)	16S	39	Irritable bowel syndrome	Stool	PRJDB5442 in the NCBI
IBS-5	16S	42	Irritable bowel syndrome	Stool	[17]
IBS-6 (PRJNA268708)	16S	289	Irritable bowel syndrome	Stool	PRJNA268708 in the NCBI
IBS-7 (PRJNA373876)	16S	51	Irritable bowel syndrome	Stool	PRJNA373876 in the NCBI
Autism-1 (PRJNA578223)	16S	96	Autism	Stool	PRJNA578223 in the NCBI
Autism-2 (PRJEB42687)	16S	77	Autism	Stool	PRJEB42687 in the NCBI
Autism-3 (PRJNA589343)	16S	127	Autism	Stool	PRJNA589343 in the NCBI
Autism-4 (PRJNA529598)	16S	484	Autism	Stool	PRJNA529598 in the NCBI
Depression-1 (PRJNA778934)	16S	127	Depression	Stool	PRJNA778934 in the NCBI
Depression-2 (PRJNA776170)	16S	85	Depression	Stool	PRJNA776170 in the NCBI
Depression-3 (PRJNA591924)	16S	90	Depression	Stool	PRJNA591924 in the NCBI
Depression-4 (PRJNA724918)	16S	305	Depression	Stool	PRJNA724918 in the NCBI
Depression-5 (PRJNA685916)	16S	106	Depression	Stool	PRJNA685916 in the NCBI

Table 4: Running times of different metrics in seconds on

Metric	Bray-Curtis	Jaccard	Aitchison	Weighted UniFrac	1D log (composition)	2D log (composition)	2D SAM	GIMIC MSE	GIMIC SAM
Allergy	15.73	23.6	3.933	150.78	4.089	55.3	70.03	140.83	141.6
Cirrhosis	4.948	7.423	1.23	88.68	1.67	19.56	21.03	41.52	44.54
Cirrhosis-pop	4.847	7.27	1.454	67.089	2.03	8.45	15.37	41.44	43.63
ERAWIJANTARI	38.209	57.31	9.55	427.09	8.989	84.23	100.23	295.83	343.886
ERP020401	4.915	6.613	0.321	40	0.657	7.049	19.968	34.84	36.128
ERP021216	0.115	0.124	0.241	8	0.408	0.02	0.064	0.809	0.861
GDM	4.857	7.285	1.457	74.089	1.289	6.08	5.23	39.25	43.715
FRANZOSA	149	223.5	24.833	1698.08	27.098	169.34	200.467	1148.936	1341.307
HE	3.131	4.697	1.256	42.089	2.035	6.04	15.73	23.54	28.184
IBD	0.904	1.207	0.173	50	0.567	6.07	8.05	17.08	20.227
JACOB	0.2	0.236	0.09	15	0.05	0.263	0.324	2.12	2.41
KIM	2.57	3.857	0.964	26.089	0.76	5.987	6.78	18.569	23.145
MF	5.75	8.637	2.044	76.08	2.05	7.068	18.698	46.69	51.825
MARS	251.55	377.33	37.73	2400	38.069	127.98	137.89	1762.678	2264.002
Nugent	4.735	7.103	3.089	58.09	4.089	8.05	10.89	37.53	42.623
Obesity	5.628	8.44	2.111	68.09	2.0798	17.08	22.08	44.59	50.656
OK94	1.453	2.18	1.308	16.897	2.098	5.089	6.789	11.06	13.08
PRJNA353587	0.126	0.159	0.06	3.5	0.05	0.119	0.323	3.068	3.339
PRJNA412501	0.03	0.04	0.03	2	0.03	0.019	0.05	0.06	0.1
PRJNA419097	0.202	0.247	0.07	10	0.1	0.199	0.477	4.798	5.34
T2D	13.12	19.68	5.589	136.08	5.897	52.567	60.76	115.07	118.08
WANG	248.258	372.33	148.993	2513.07	151.98	197.789	234.8	1858.907	2234.329
WB	1.72	2.58	1.29	18.09	2.06	4.598	7.089	13.478	15.49
YACHIDA	742.77	1114.16	445.708	7089	443.07	389.06	400.89	6130.427	6685.628

3 Supplementary Figures

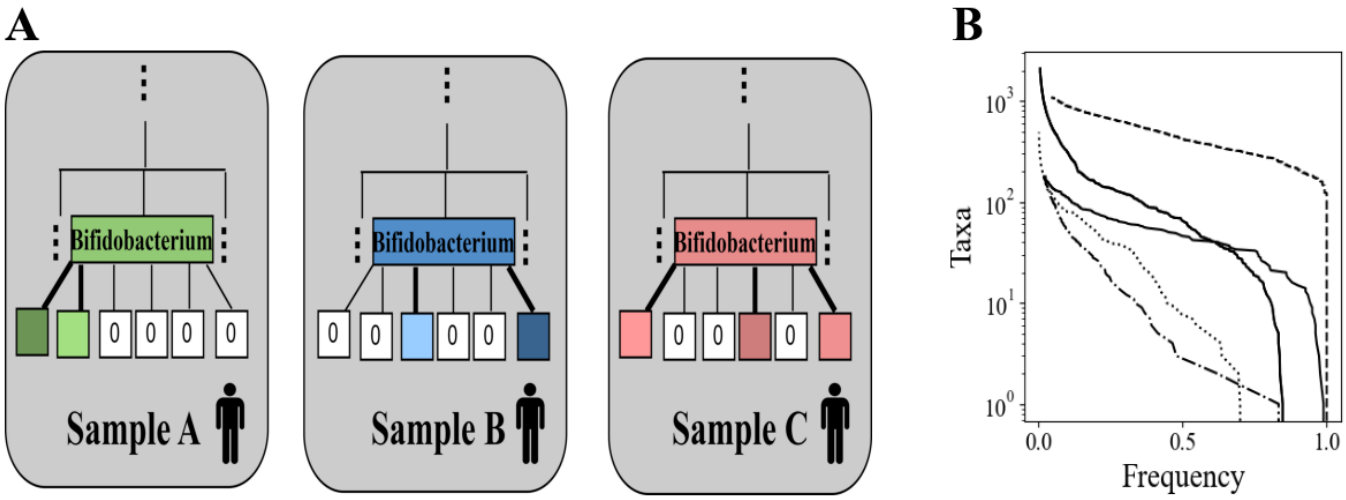


Figure 1: Microbiome challenges **A.** Shared sisters within a certain genus, while a specific species is unique. **B.** Different scales of microbiome taxa over different datasets. The x-axis represents the frequency and the y-axis represents the number of taxa with this frequency.

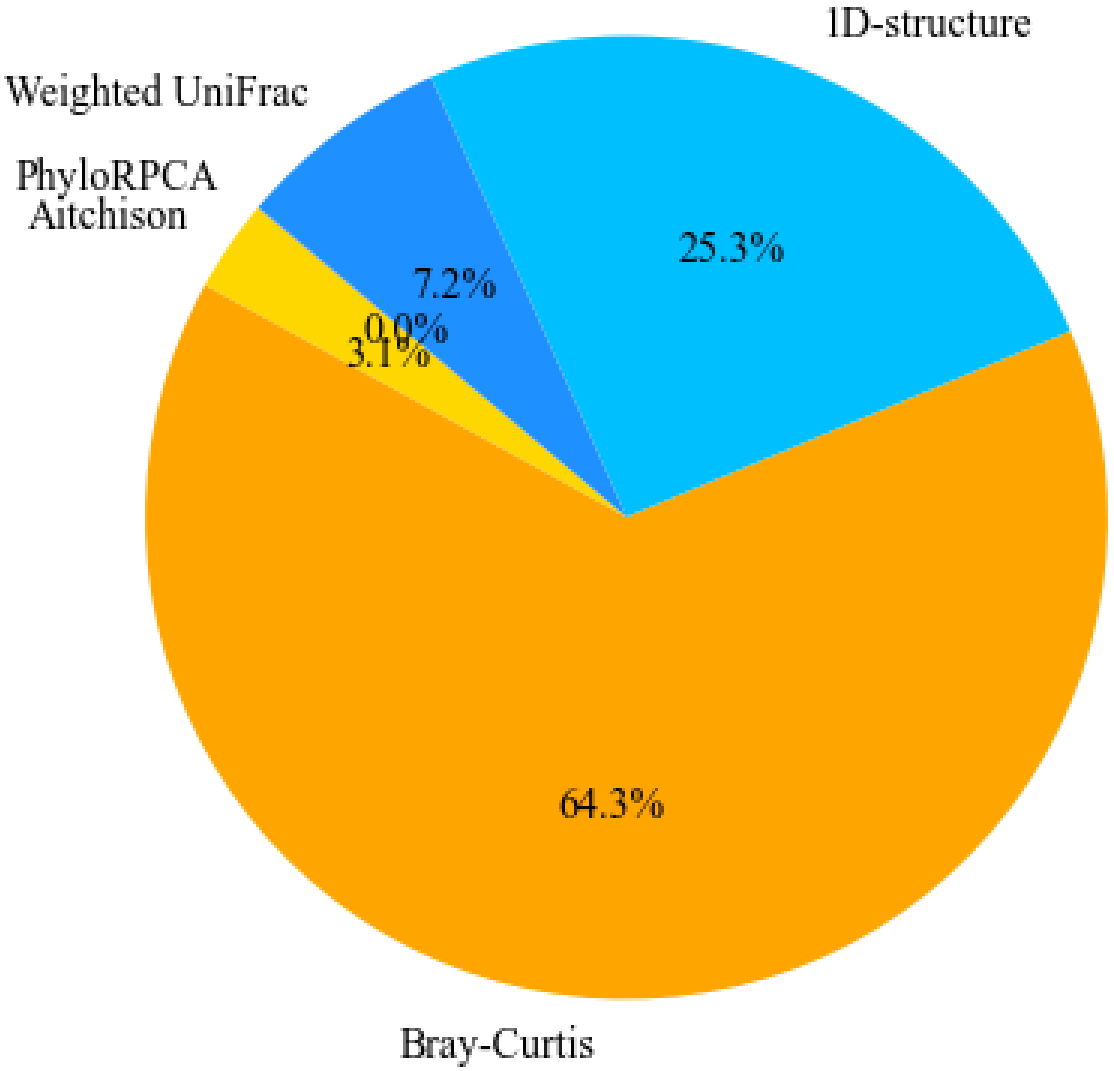


Figure 2: State-of-the-art common metrics citation distribution based on more than 20,000 citations from Google Scholar, such that the gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, royal blue represents the PhyloRPCA.

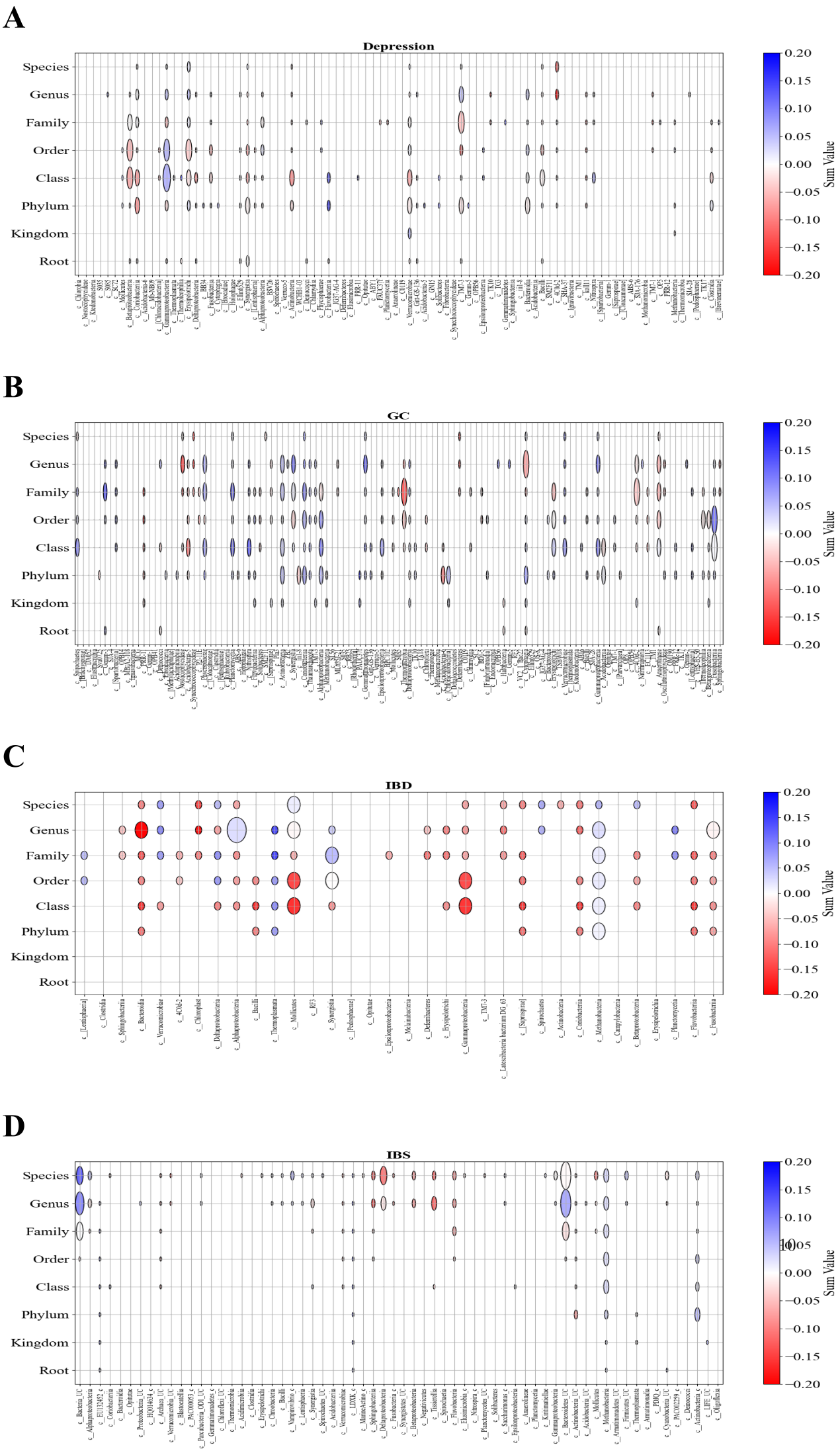

A
Depression
Species
Genus
Family
Order
Class
Phylum
Kingdom
Root
Sum Value
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
B
GC
C
IBD
D
IBS

Figure 3: GIMIC meta-analysis over various cohorts from different phenotypes. This schematic illustrates the GIMIC meta-analysis conducted over four distinct groups of phenotypes Depression **(A)**, GC **(B)**, IBD **(C)** and IBS **(D)**. Each circle in the figure represents a significant integrated cluster at various taxonomic levels. The color of the circles denotes the sum of differences across the different cohorts, with red indicating negative differences and blue indicating positive differences. The size of each circle corresponds to the number of cohorts in which the difference is consistently significant. There are phenotypes that are less related to the microbiome, like depression, while others are more related, like IBD and GC.

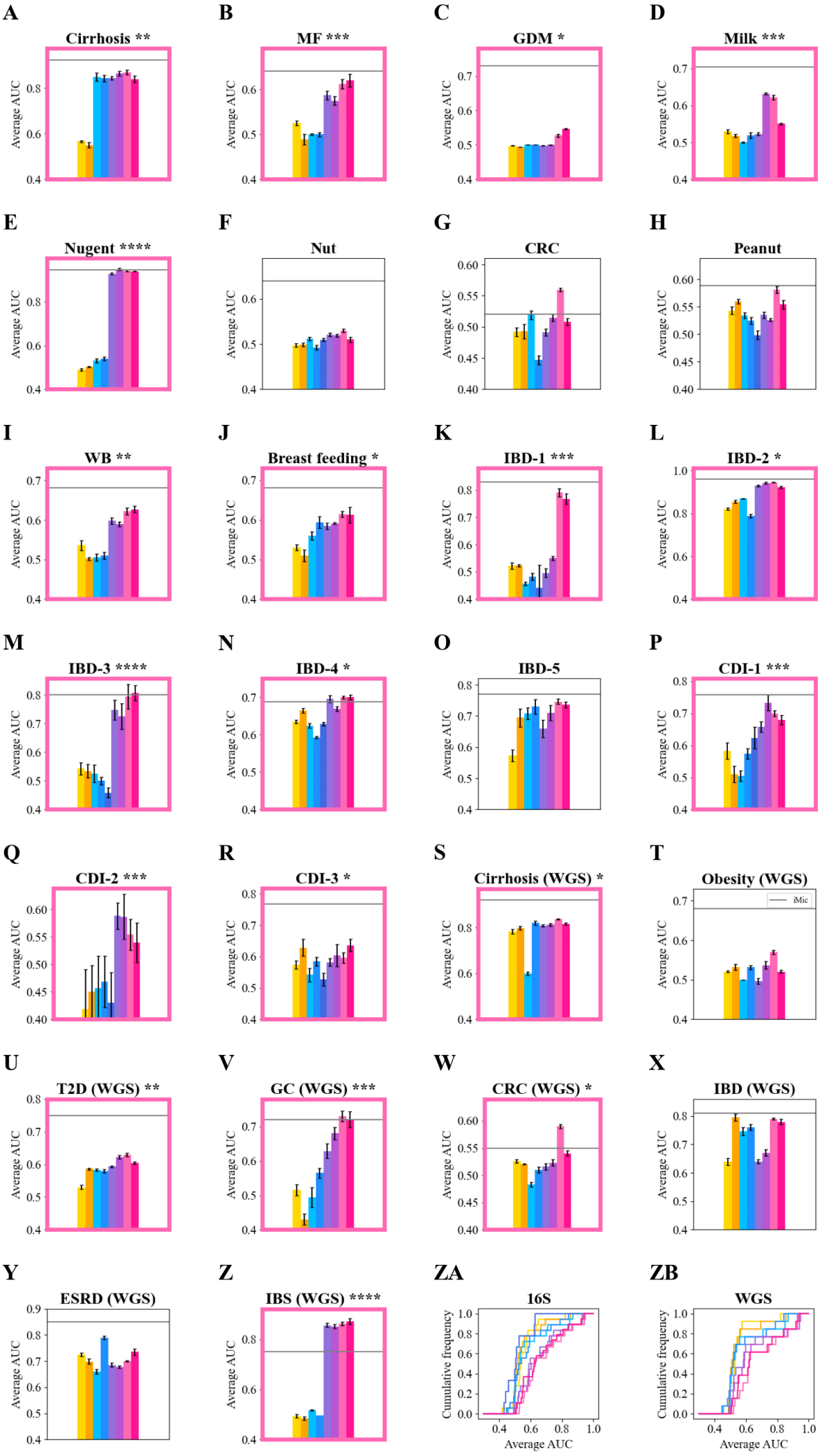

A
Cirrhosis **
B
MF ***
C
GDM *
D
Milk ***
E
Nugent ****
F
Nut
G
CRC
H
Peanut
I
WB **
J
Breast feeding *
K
IBD-1 ***
L
IBD-2 *
M
IBD-3 ****
N
IBD-4 *
O
IBD-5
P
CDI-1 ***
Q
CDI-2 ***
R
CDI-3 *
S
Cirrhosis (WGS) *
T
Obesity (WGS)
iMic
U
T2D (WGS) **
V
GC (WGS) ***
W
CRC (WGS) *
X
IBD (WGS)
Y
ESRD (WGS)
Z
IBS (WGS) ****
ZA
16S
ZB
WGS
Average AUC
Cumulative frequency

Figure 4: GIMIC creates the best KNN classification measured by the average AUC in 16S cohorts as well as WGS cohorts. **A - R.** All average AUCs over 10 runs of different metrics of the 16S datasets. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, royal blue represents the PhyloRPCA, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance. For similar results on more cohorts see Supp. Mat. Fig. S5. **S- Z.** All average AUCs over 10 runs of different metrics of the WGS datasets. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, royal blue represents the PhyloRPCA, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance. The black lines represent the std errors among the 10 test runs. The grey line represents the average AUC of iMic which is the latest state-of-the-art in the field of microbiome predictions. There is a pink frame where GIMIC is better than all the state-of-the-art methods, and the stars represent the $p - value$ of GIMIC vs. all the other methods, such that *-p<0.05, **-p<0.01, ***-p<0.001. **ZA - ZB.** Cumulative histogram of average AUC performance over all the 16S cohorts **(ZA)** and all the WGS cohorts **(ZB)**.

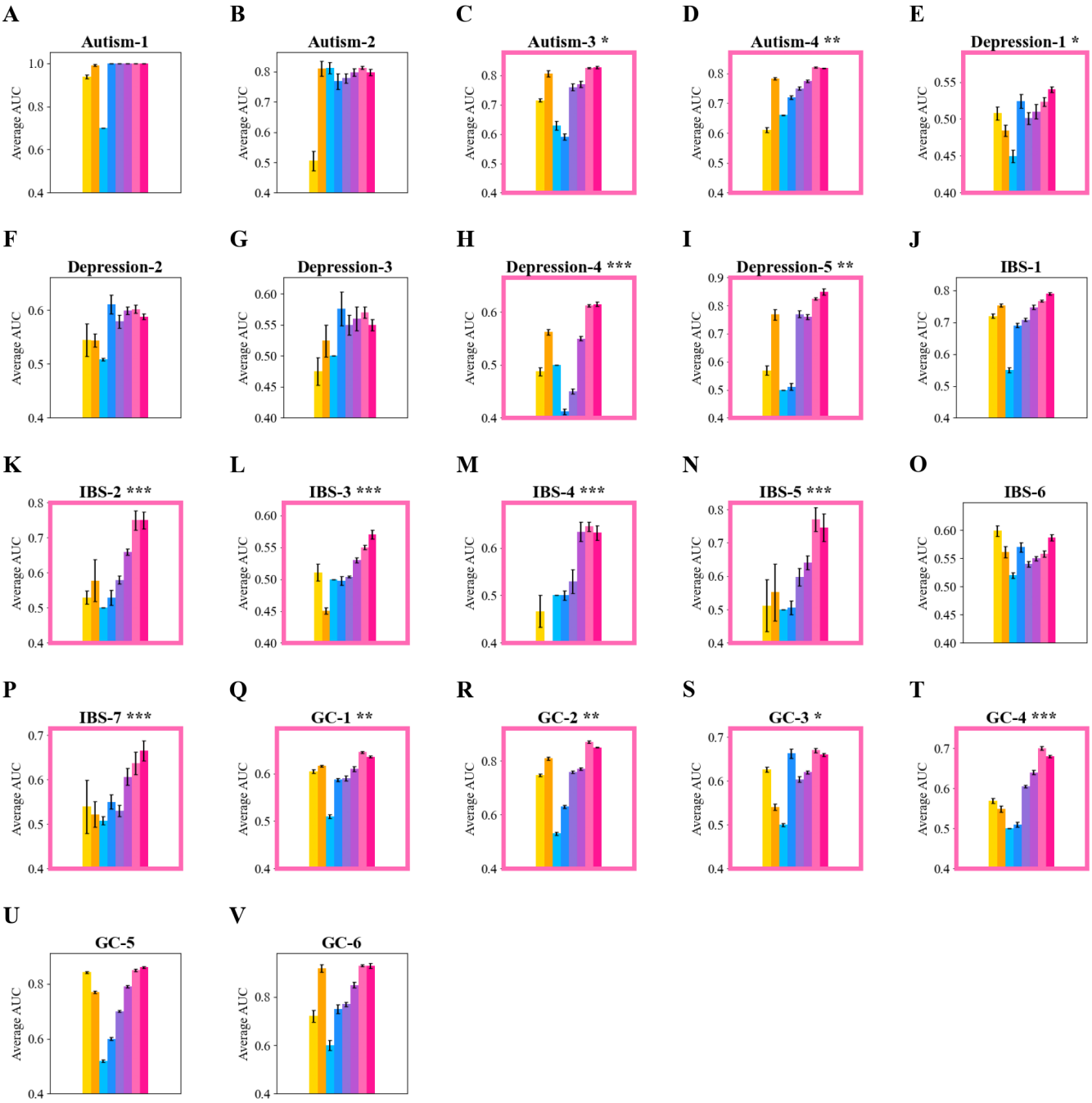


Figure 5: GIMIC creates the best KNN classification measured by the average AUC in 16S cohorts. **A - V.** All average AUCs over 10 runs of different metrics of the 16S datasets. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, royal blue represents the PhyloRPCA, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance. There is a pink frame where GIMIC is better than all the state-of-the-art methods, and the stars represent the $p - value$ of GIMIC vs. all the other methods, such that *-p<0.05, **-p<0.01, ***-p<0.001.


A
Cirrhosis
B
MF
C
CRC
D
Milk
E
Nugent
F
Nut
G
Peanut
H
GDM
I
WB
J
Breast feeding
K
IBD-1
L
IBD-2
M
IBD-3
N
IBD-4
O
IBD-5
P
CDI-1
Q
CDI-2
R
CDI-3
S
Autism-1
T
Autism-2
U
Autism-3
V
Autism-4
W
Depression-1
X
Depression-2
Y
Depression-3
Z
Depression-4
ZA
Depression-5
ZB
IBS-1
ZC
IBS-2
ZD
IBS-3
ZE
IBS-4
ZF
IBS-5
ZG
IBS-6
ZH
IBS-7
ZI
GC-1
ZJ
GC-2
ZK
GC-3
ZL
GC-4
ZM
GC-5
ZM
GC-6
GIMIC MSE
GIMIC SAM
2D-SAM
2D-log(composition)
PhyloRPCA
Weighted UniFrac
1D-structure
Bray-Curtis
Aitchison
Out/in distance

Figure 6: All average out/in distances over 10 runs of different metrics of the 16S datasets. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance.

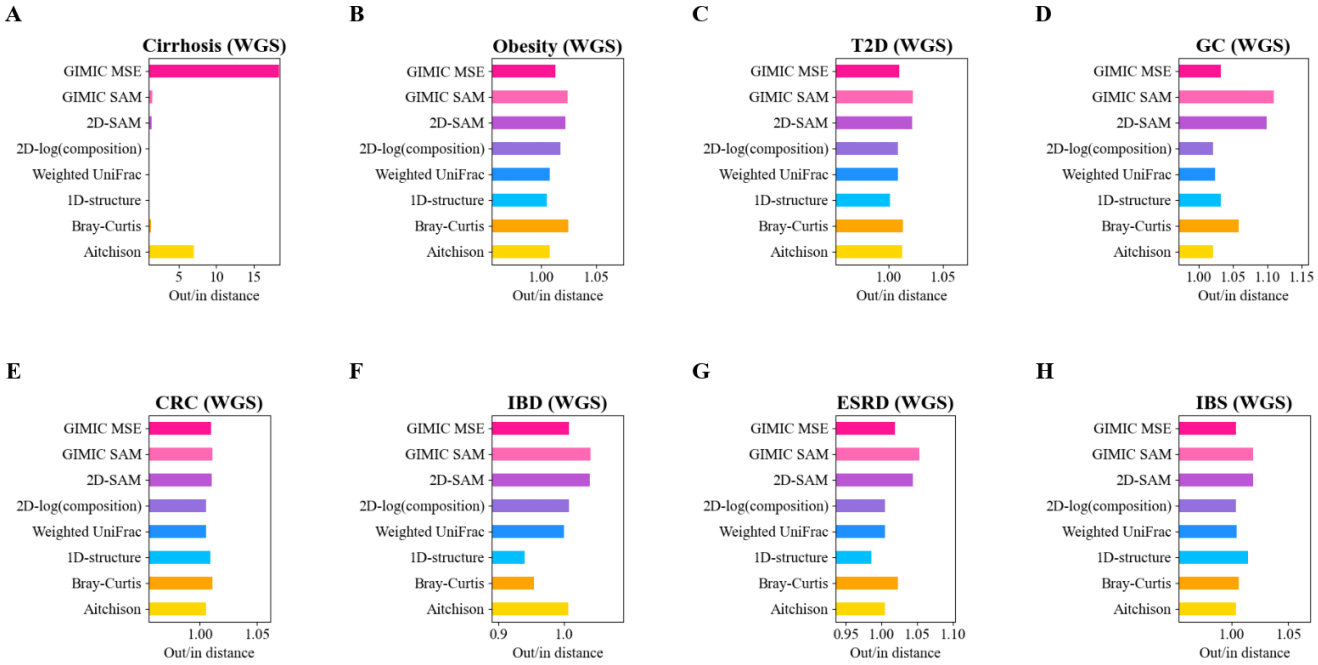


Figure 7: All average out/in distances over 10 runs of different metrics of the WGS datasets. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance.


A
Cirrhosis
B
MF
C
CRC
D
Milk
E
Nugent
F
Nut
G
Peanut
H
GDM
I
WB
J
Breast feeding
K
IBD-1
L
IBD-2
M
IBD-3
N
IBD-4
O
IBD-5
P
CDI-1
Q
CDI-2
R
CDI-3
S
Autism-1
T
Autism-2
U
Autism-3
V
Autism-4
W
Depression-1
X
Depression-2
Y
Depression-3
Z
Depression-4
ZA
Depression-5
ZB
IBS-1
ZC
IBS-2
ZD
IBS-3
ZE
IBS-4
ZF
IBS-5
ZG
IBS-6
ZH
IBS-7
ZI
GC-1
ZJ
GC-2
ZK
GC-3
ZL
GC-4
ZM
GC-5
ZM
GC-6
GIMIC MSE
GIMIC SAM
2D-SAM
2D-log(composition)
PhyloRPCA
Weighted UniFrac
1D-structure
Bray-Curtis
Aitchison
Average silhouette score

Figure 8: All average Silhouette scores over 10 runs. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance. The black lines represent the std errors among the 10 test runs.

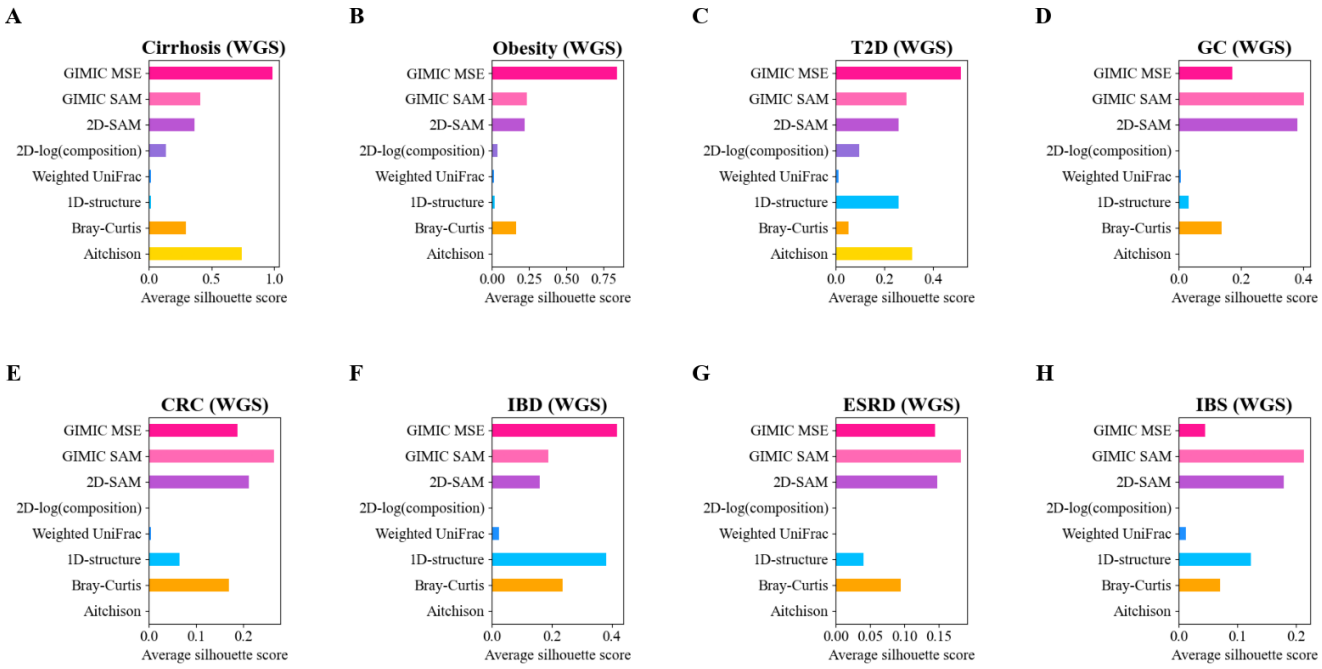


Figure 9: All average Silhouette scores over 10 runs of different metrics of the WGS datasets. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance. The black lines represent the std errors among the 10 test runs.

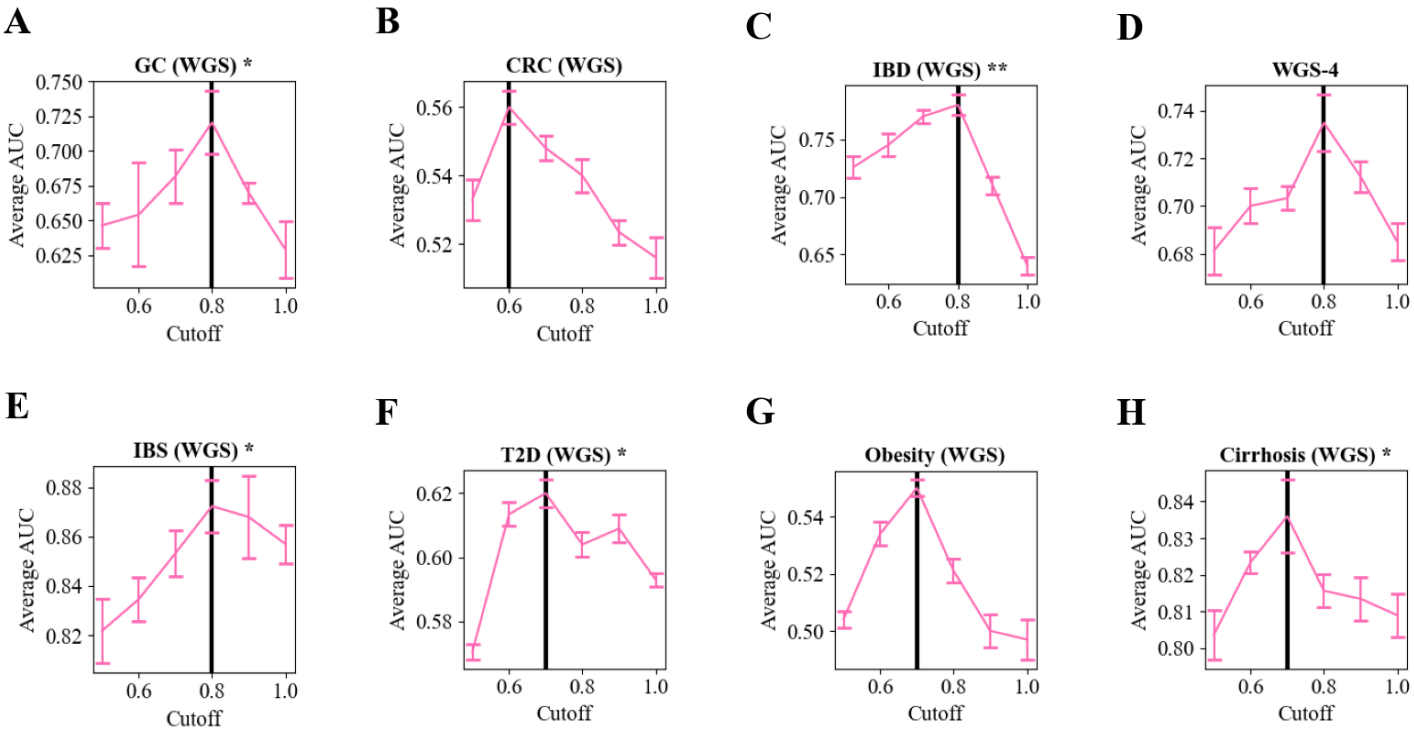


Figure 10: Different cutoffs affect the KNN performance over different tasks in the WGS cohorts using the MSE metric on the FFT transformed images. The x-axis represents the different cutoffs and the y-axis represents the average AUC over 10 runs on the test sets. The black line represents the maximum cutoff for a certain task.

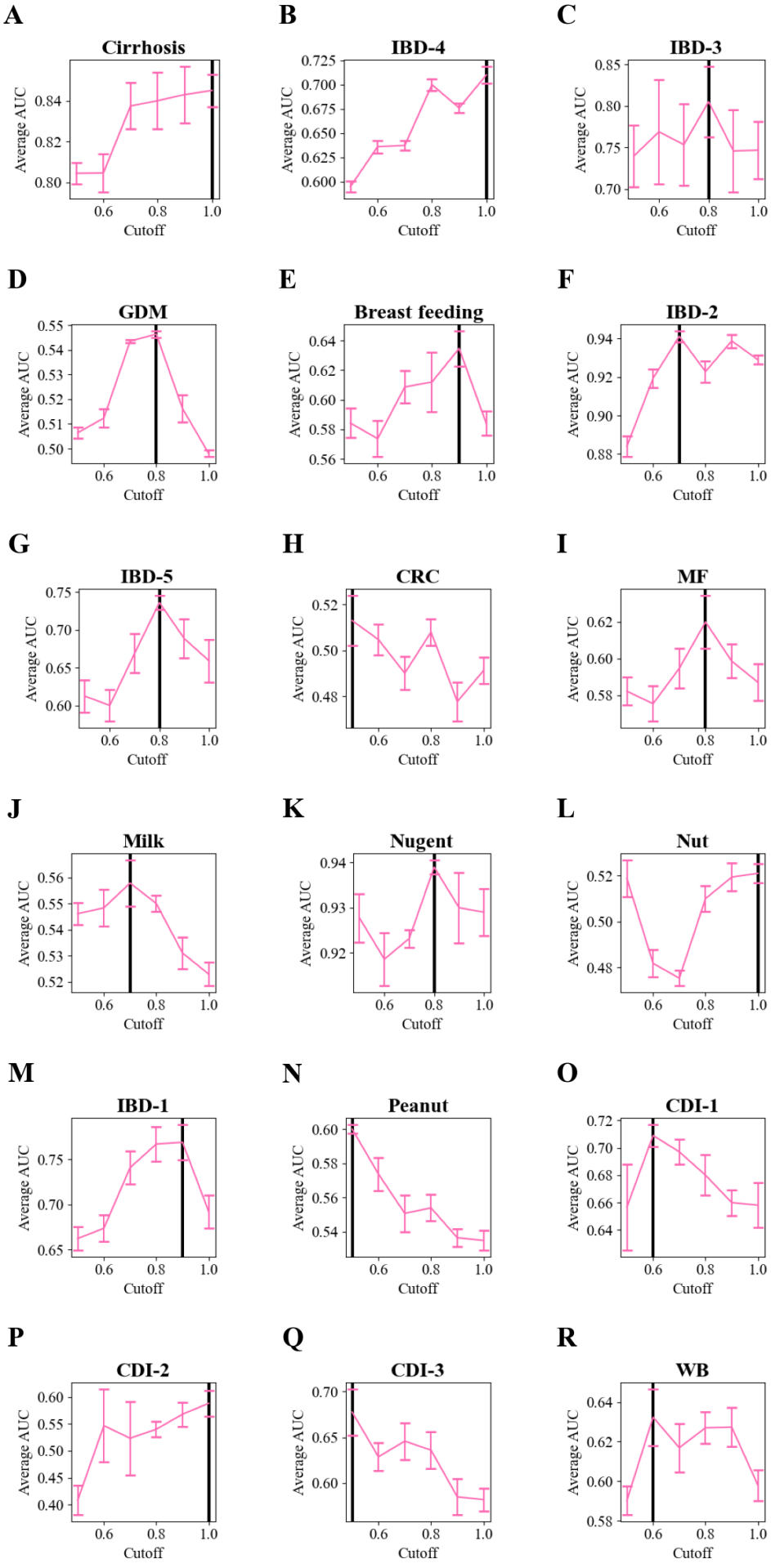


Figure 11: Different cutoffs affect the KNN performance over different tasks in the 16S cohorts using the MSE metric on the FFT transformed images. The x-axis represents the different cutoffs and the y-axis represents the average AUC over 10 runs on the test sets. The black line represents the maximum cutoff for a certain task.

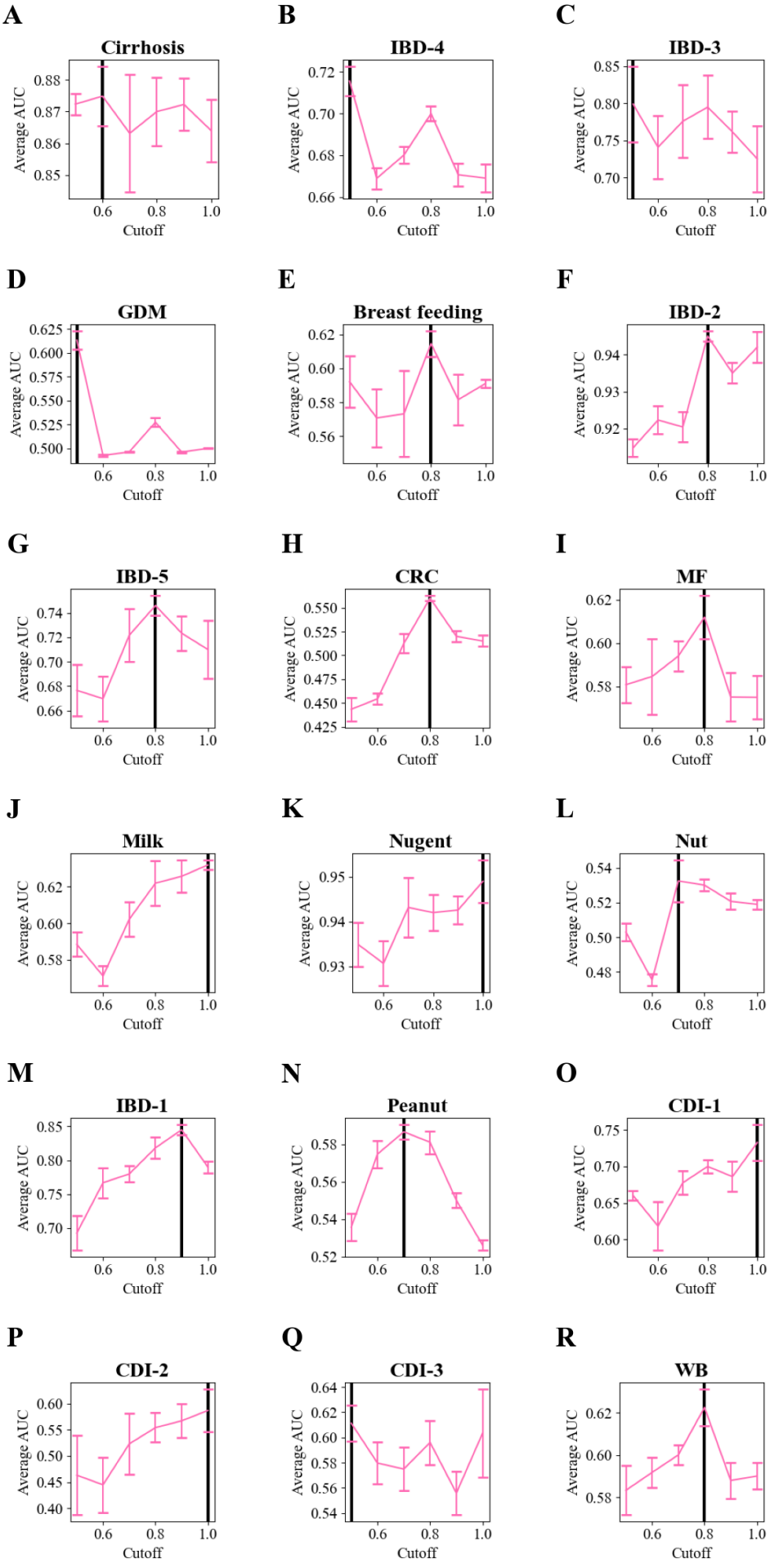


Figure 12: Different cutoffs affect the KNN performance over different tasks in the 16S cohorts using the SAM metric on the FFT transformed images. The x-axis represents the different cutoffs and the y-axis represents the average AUC over 10 runs on the test sets. The black line represents the maximum cutoff for a certain task.

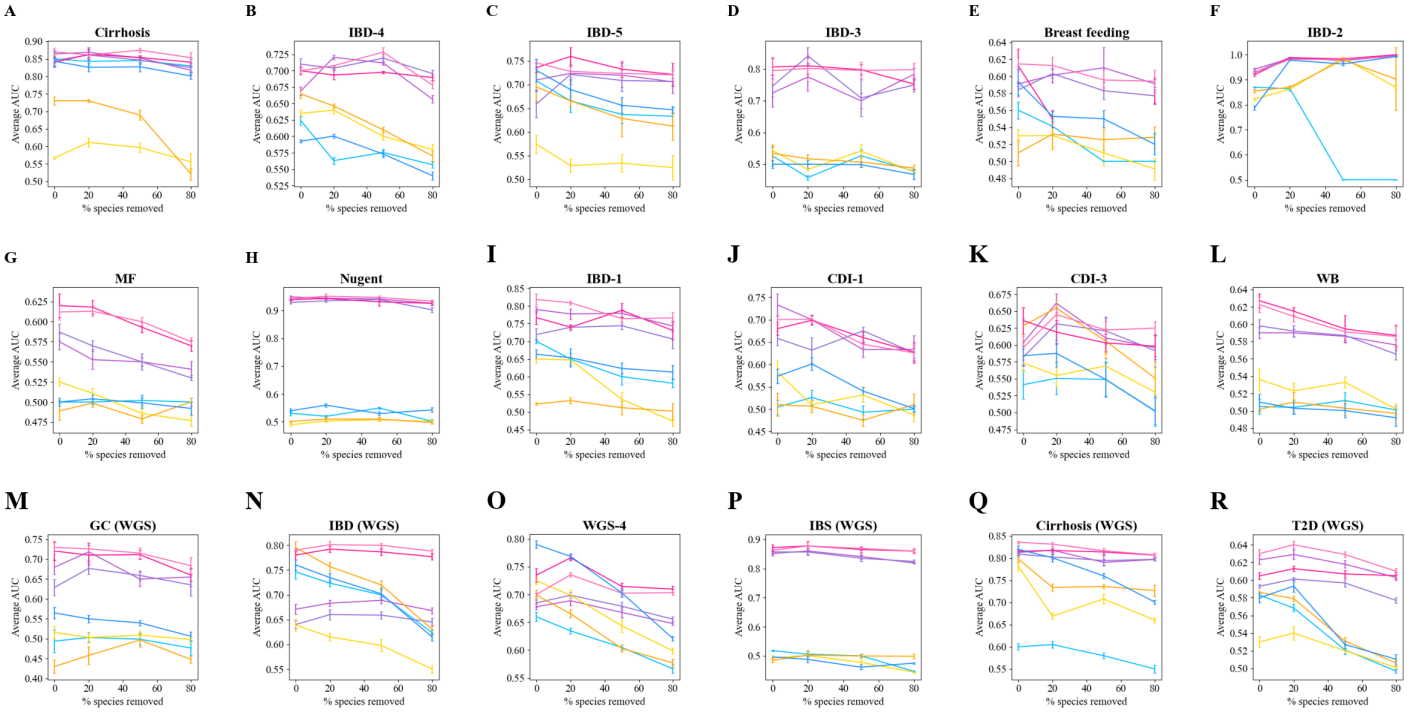
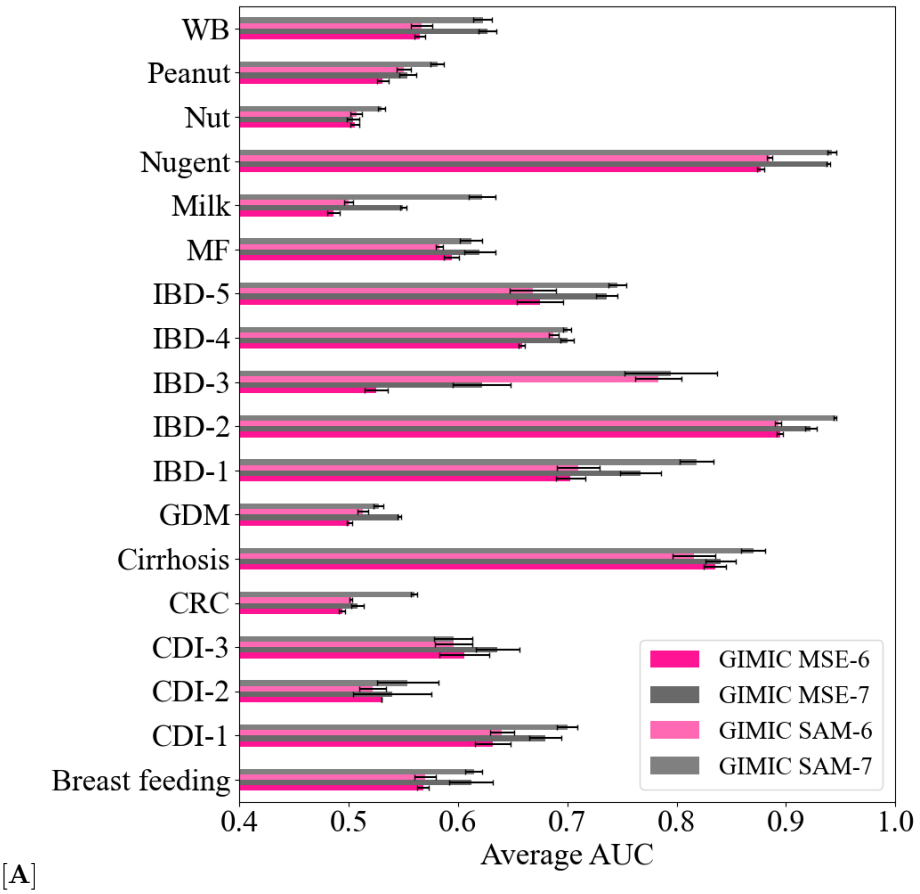
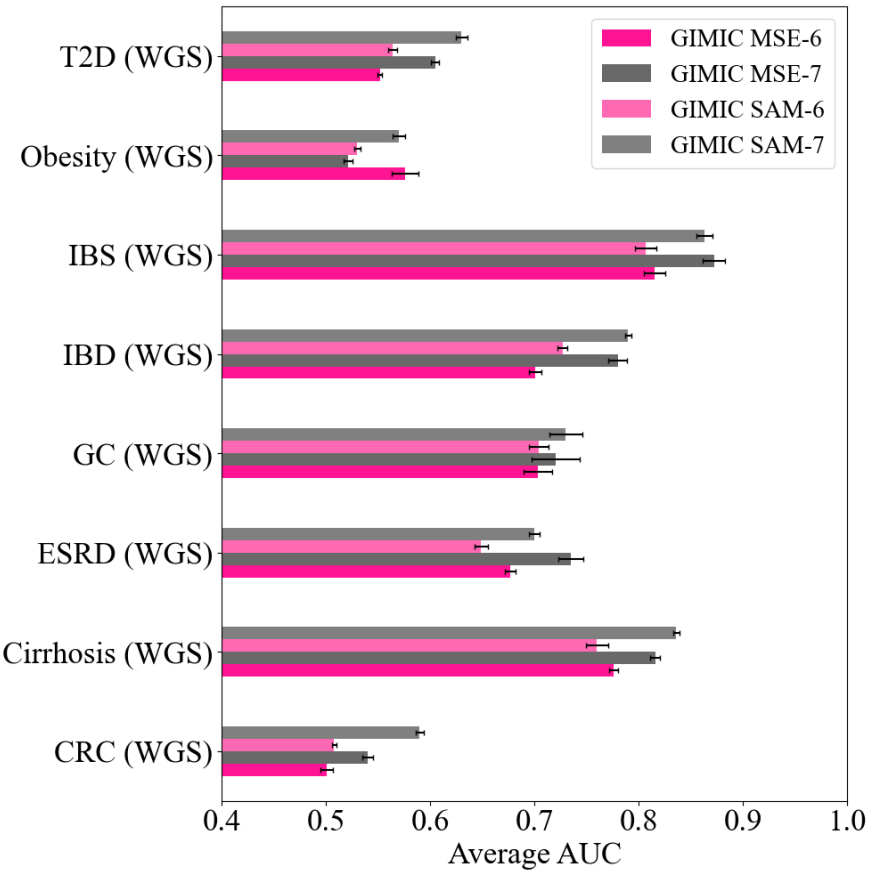


Figure 13: Sister smoothing stability of different metrics. The x-axis represents the percent of species names that were randomly removed and the y-axis represents the average AUC over 10 runs on the test set of a KNN classifier with the different metrics.The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance. The horizontal lines represent the std errors among the 10 test runs.



[**A**]



[**B**]

Figure 14: GIMIC based on a full microbiome vector (including species) is significantly better than GIMIC based on information till the genus level. The x-axis represents the average test AUC over 10 runs of a KNN classifier based on the GIMIC metric. The pink bars represent the GIMIC based on the whole information, while the grey bars represent GIMIC based on information till the genus level. The black lines represent the std errors.

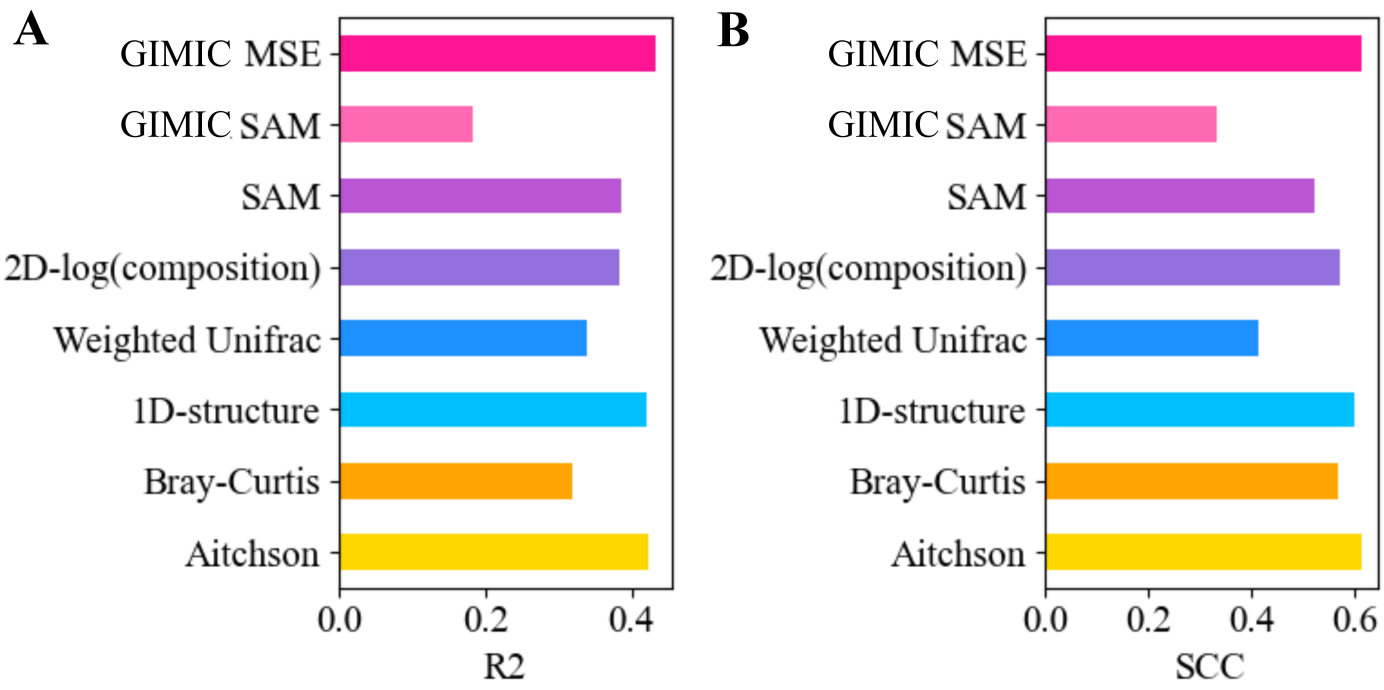


Figure 15: XGBOOST manages to predict the different metrics according to the differences in biological factors, such as BMI, metabolites, and illness as was measured by R2 score **(A)** and SCC **(B)**. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance.

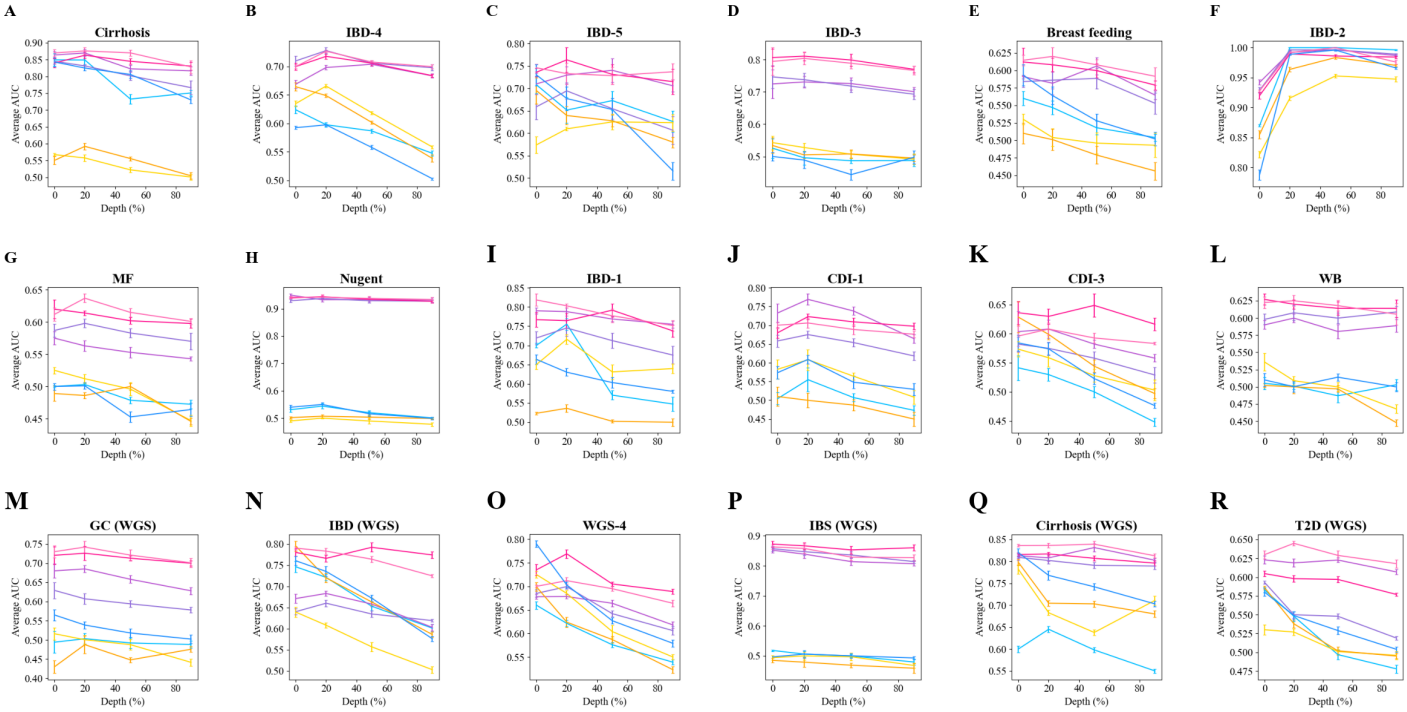
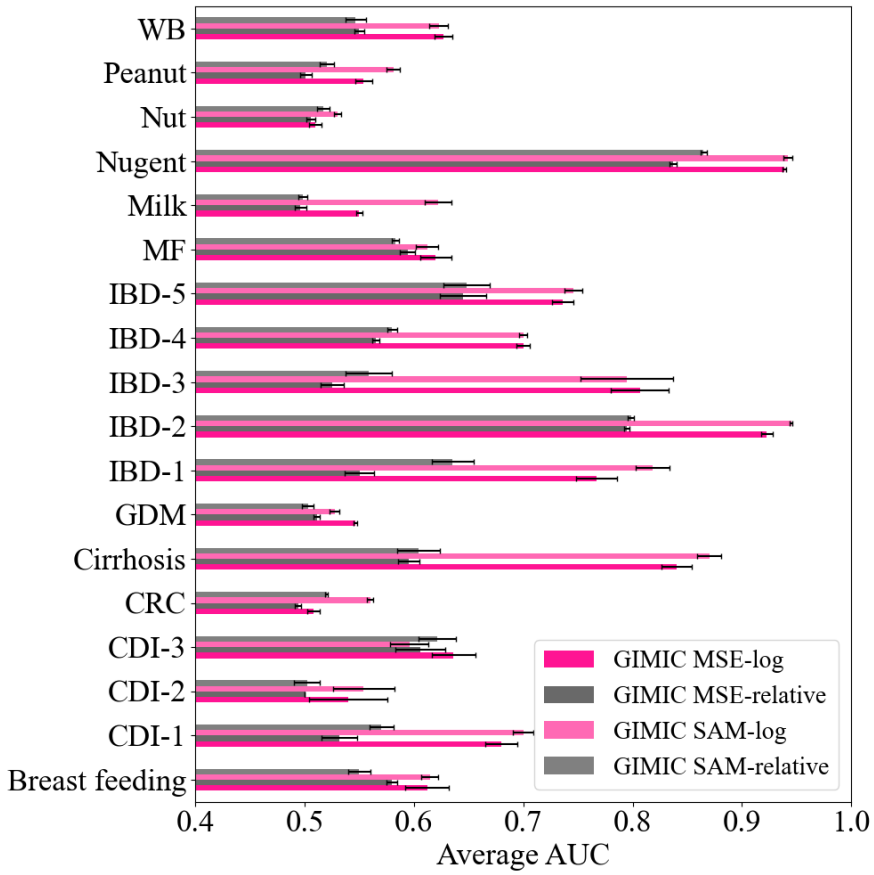
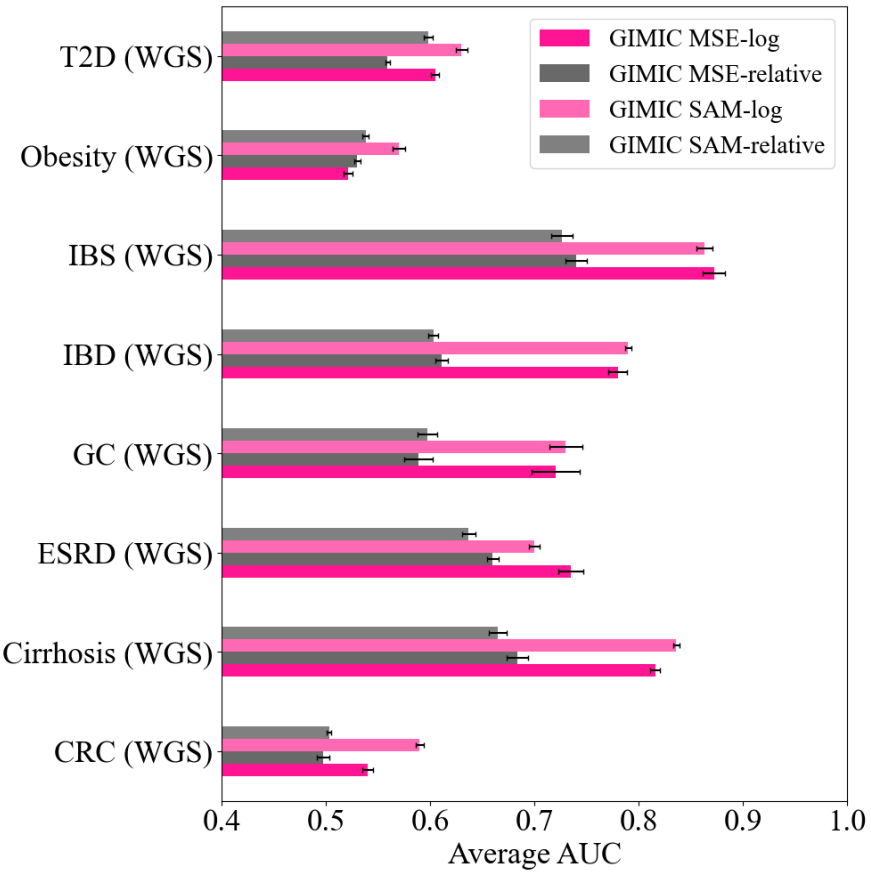


Figure 16: Depth sequencing effect on different metrics performance. The x-axis represents the percent of reads that were randomly removed and the y-axis represents the average AUC over 10 runs on the test set of a KNN classifier with the different metrics. The gold color represents the Aitchison distance, orange represents the Bray-Curtis distance, light blue represents the 1D-structure distance, blue represents the weighted Unifrac, light purple represents the 2D-log(composition) distance, purple represents the 2D-SAM distance, pink represents the GIMIC SAM distance and dark pink represents the GIMIC MSE distance. The horizontal lines represent the std errors among the 10 test runs.



[**A**]



[**B**]

Figure 17: GIMIC with log normalization is significantly better than GIMIC with relative normalization. The x-axis represents the average test AUC over 10 runs of a KNN classifier based on the GIMIC metric. The pink bars represent the GIMIC with the log normalization and the grey bars represent the GIMIC with the relative normalization. The black lines represent the std errors.

References

[1] Michael R Goldberg, Hadar Mor, Dafna Magid Neriya, Faiga Magzal, Efrat Muller, Michael Y Appel, Liat Nachshon, Elhanan Borenstein, Snait Tamir, Yoram Louzoun, et al. Microbial signature in ige-mediated food allergies. *Genome Medicine*, 12:1–18, 2020.

[2] Human Microbiome Project Consortium et al. Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207, 2012.

[3] Derek Reiman, Ahmed A Metwally, Jun Sun, and Yang Dai. Popphy-cnn: a phylogenetic tree embedded architecture for convolutional neural networks to predict host phenotype from metagenomic data. *IEEE Journal of Biomedical and Health Informatics*, 24(10):2993–3001, 2020.

[4] Pande Putu Erawijantari, Sayaka Mizutani, Hirotsugu Shiroma, Satoshi Shiba, Takeshi Nakajima, Taku Sakamoto, Yutaka Saito, Shinji Fukuda, Shinichi Yachida, and Takuji Yamada. Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. *Gut*, 69(8):1404–1415, 2020.

[5] Sahil Khanna, Yoshiki Vazquez-Baeza, Antonio González, Sophie Weiss, Bradley Schmidt, David A Muñiz-Pedrogo, John F Rainey, Patricia Kammer, Heidi Nelson, Michael Sadowsky, et al. Changes in microbial ecology after fecal microbiota transplantation for recurrent c. difficile infection affected by underlying inflammatory bowel disease. *Microbiome*, 5(1):1–8, 2017.

[6] Yishay Pinto, Sigal Frishman, Sondra Turjeman, Adi Eshel, Meital Nuriel-Ohayon, Oshrit Shtossel, Oren Ziv, William Walters, Julie Parsonnet, Catherine Ley, et al. Gestational diabetes is driven by microbiota-induced inflammation months before diagnosis. *Gut*, 2023.

[7] Eric A Franzosa, Alexandra Sirota-Madi, Julian Avila-Pacheco, Nadine Fornelos, Henry J Haiser, Stefan Reinker, Tommi Vatanen, A Brantley Hall, Himel Mallick, Lauren J McIver, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology*, 4(2):293–305, 2019.

[8] Yan He, Wei Wu, Hui-Min Zheng, Pan Li, Daniel McDonald, Hua-Fang Sheng, Mu-Xuan Chen, Zi-Hui Chen, Gui-Yuan Ji, Zhong-Dai-Xi Zheng, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nature Medicine*, 24(10):1532–1535, 2018.

[9] Xuan He, Mariana Parenti, Tove Grip, Bo Lönnerdal, Niklas Timby, Magnus Domellöf, Olle Hernell, and Carolyn M Slupsky. Fecal microbiome and metabolome of infants fed bovine mfgm supplemented formula or standard formula with breast-fed infants as reference: a randomized controlled trial. *Scientific Reports*, 9(1):11589, 2019.

[10] Janine van der Giessen, Dana Binyamin, Anna Belogolovski, Sigal Frishman, Kinneret Tenenbaum-Gavish, Eran Hadar, Yoram Louzoun, Maikel Petrus Peppelenbosch, Christien Janneke van der Woude, Omry Koren, et al. Modulation of cytokine patterns and microbiome during pregnancy in ibd. *Gut*, 69(3):473–486, 2020.

[11] Jonathan P Jacobs, Maryam Goudarzi, Namita Singh, Maomeng Tong, Ian H McHardy, Paul Ruegger, Miro Asadourian, Bo-Hyun Moon, Allyson Ayson, James Borneman, et al. 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, 2016.

[12] Minsuk Kim, Emily Vogtmann, David A Ahlquist, Mary E Devens, John B Kisiel, William R Taylor, Bryan A White, Vanessa L Hale, Jaeyun Sung, Nicholas Chia, et al. Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis. *MBio*, 11(1):e03186–19, 2020.

[13] Ruben AT Mars, Yi Yang, Tonya Ward, Mo Houtti, Sambhawa Priya, Heather R Lekatz, Xiaojia Tang, Zhifu Sun, Krishna R Kalari, Tal Korem, et al. Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell*, 182(6):1460–1473, 2020.

[14] Jacques Ravel, Pawel Gajer, Zaid Abdo, G Maria Schneider, Sara SK Koenig, Stacey L McCulle, Shara Karlebach, Reshma Gorle, Jennifer Russell, Carol O Tacket, et al. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences*, 108(supplement_1):4680–4687, 2011.

[15] Xifan Wang, Songtao Yang, Shenghui Li, Liang Zhao, Yanling Hao, Junjie Qin, Lian Zhang, Chengying Zhang, Weijing Bian, LI Zuo, et al. Aberrant gut microbiota alters host metabolome and impacts renal failure in humans and rodents. *Gut*, 69(12):2131–2142, 2020.

[16] Shinichi Yachida, Sayaka Mizutani, Hirotsugu Shiroma, Satoshi Shiba, Takeshi Nakajima, Taku Sakamoto, Hikaru Watanabe, Keigo Masuda, Yuichiro Nishimoto, Masaru Kubo, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*, 25(6):968–976, 2019.

[17] Joann Phan, Divya Nair, Suneer Jain, Thibaut Montagne, Demi Valeria Flores, Andre Nguyen, Summer Dietsche, Saurabh Gombar, and Philip Cotter. Alterations in gut microbiome composition and function in irritable bowel syndrome and increased probiotic abundance with daily supplementation. *Msystems*, 6(6):e01215–21, 2021.