

Graph-affiliated Clustering Help Neural Network

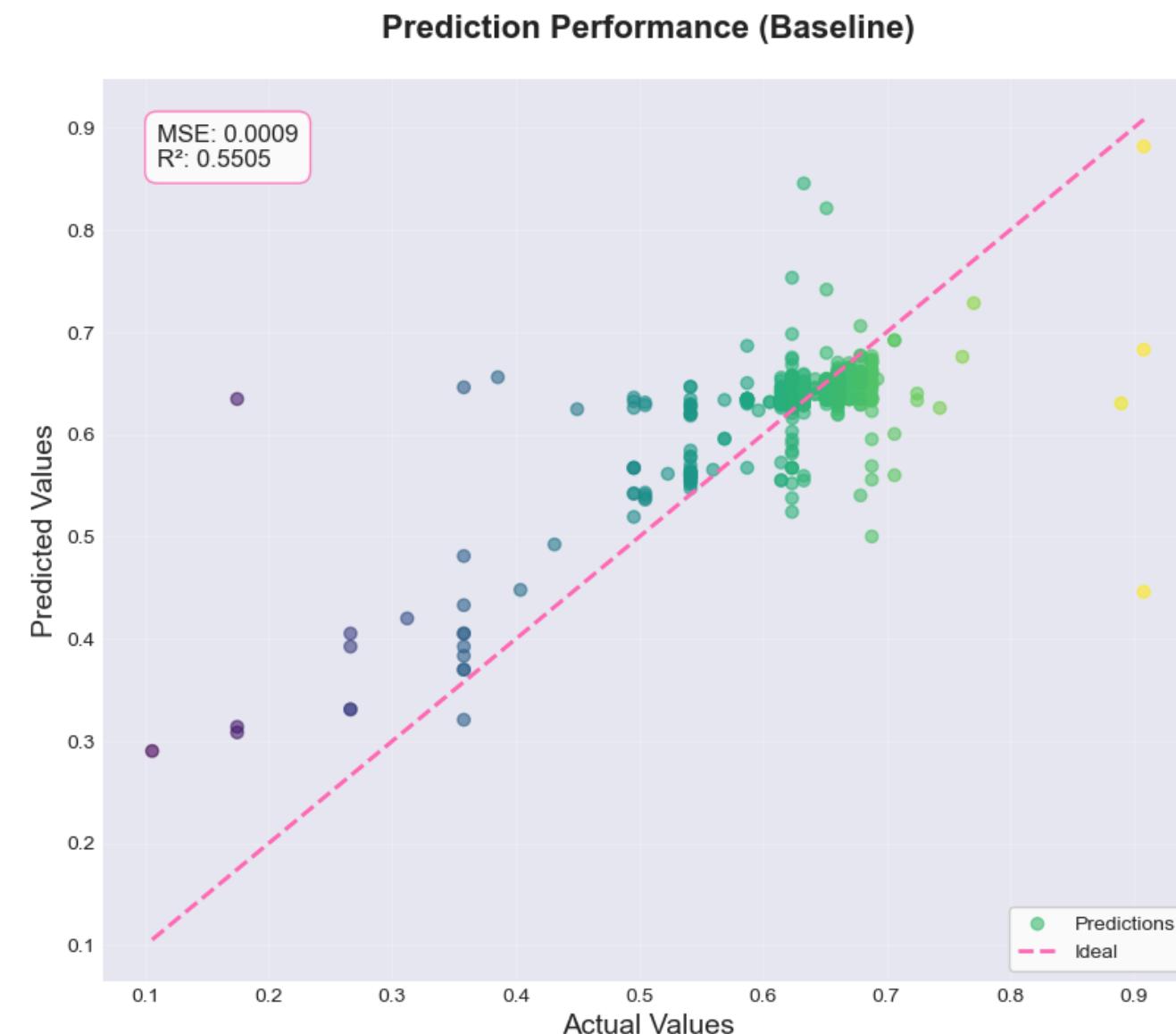
16S rRNA sequencing for
cultivation media prediction

Juntang Wang,
Yihan Wang,
(Runkun Guo).

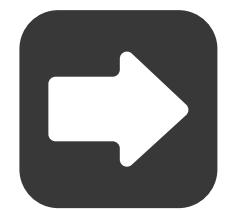
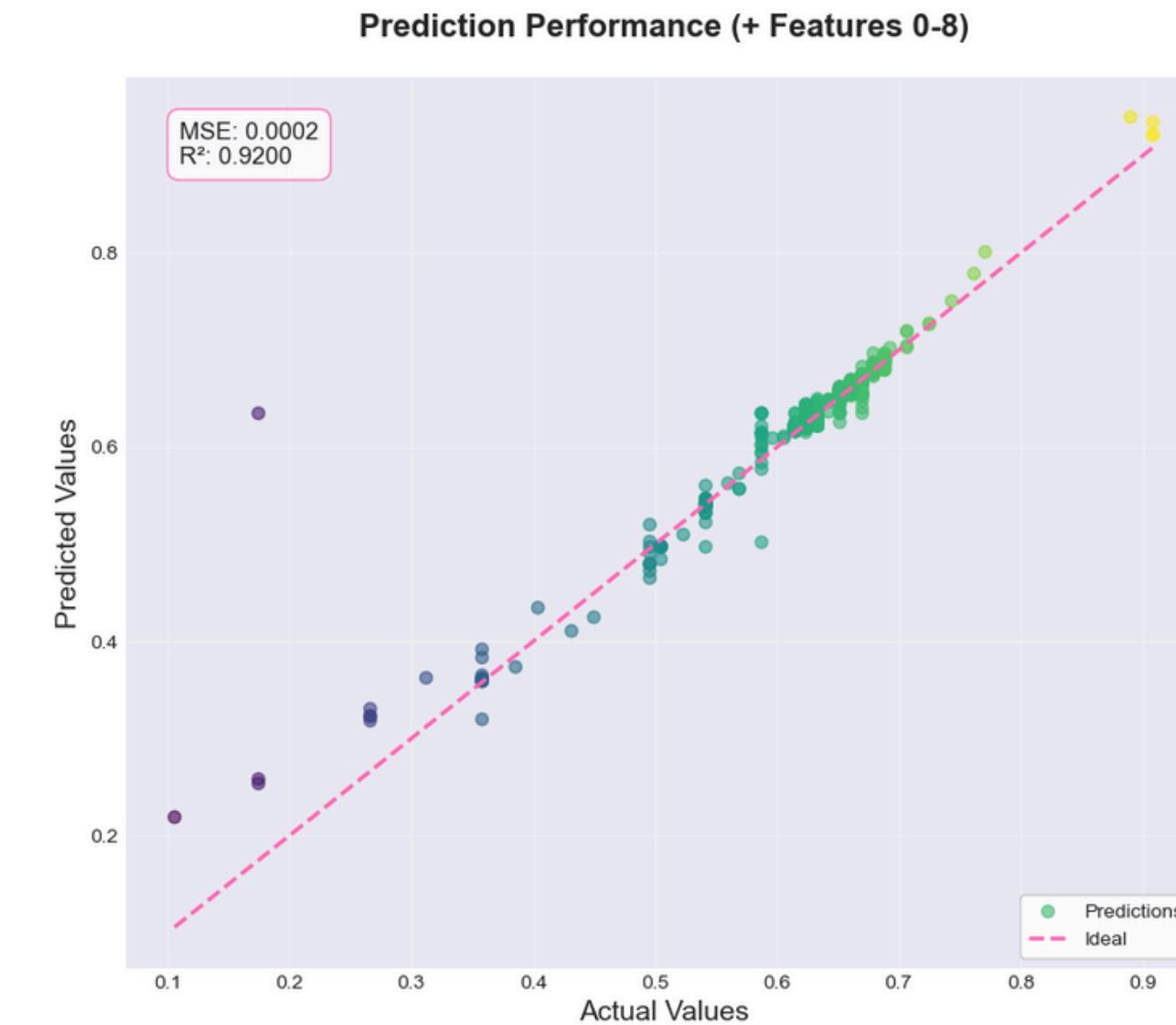
A Glance at Result

0.55 → 0.92 on a simple MLP

R²=0.55

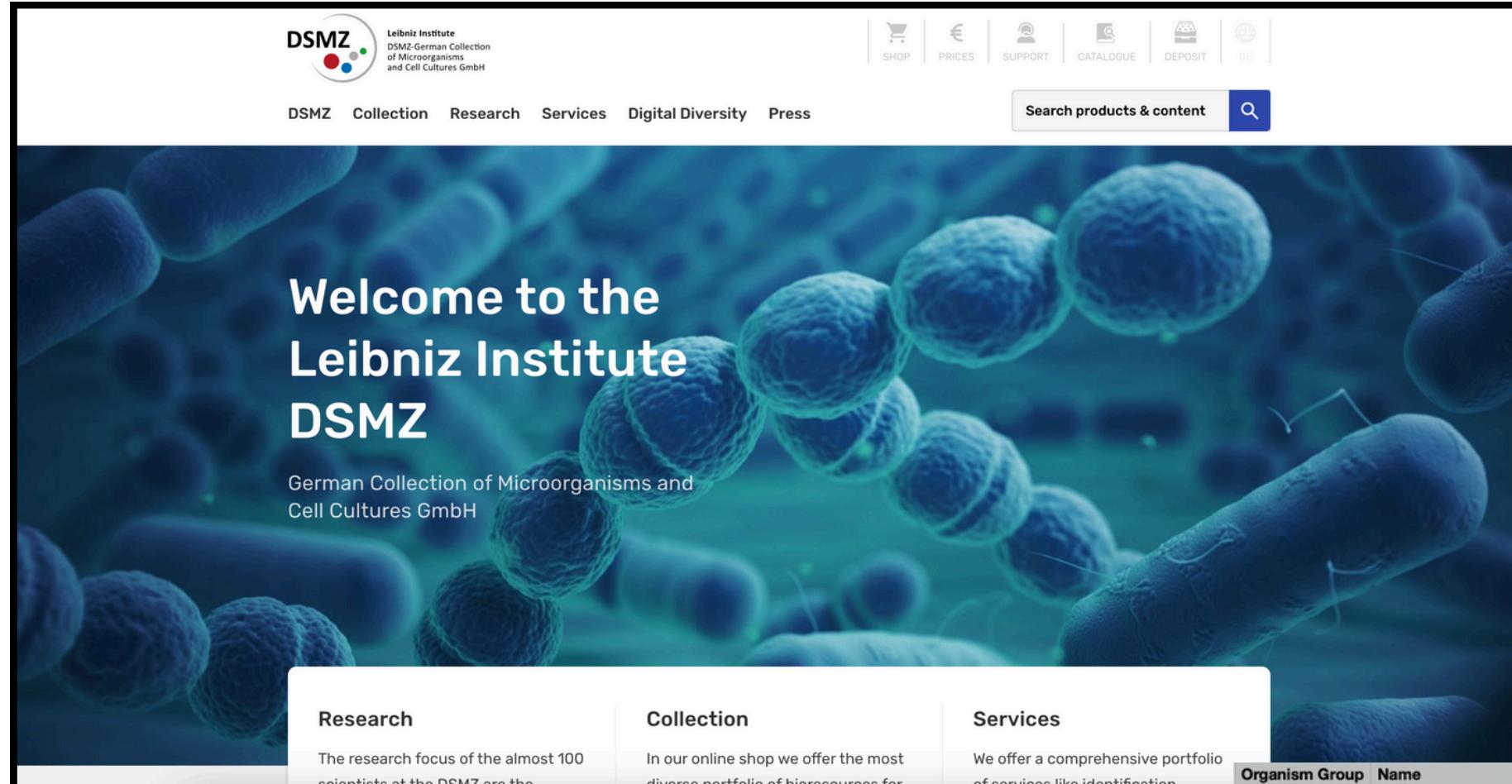


R²=0.92



Regression

Dataset



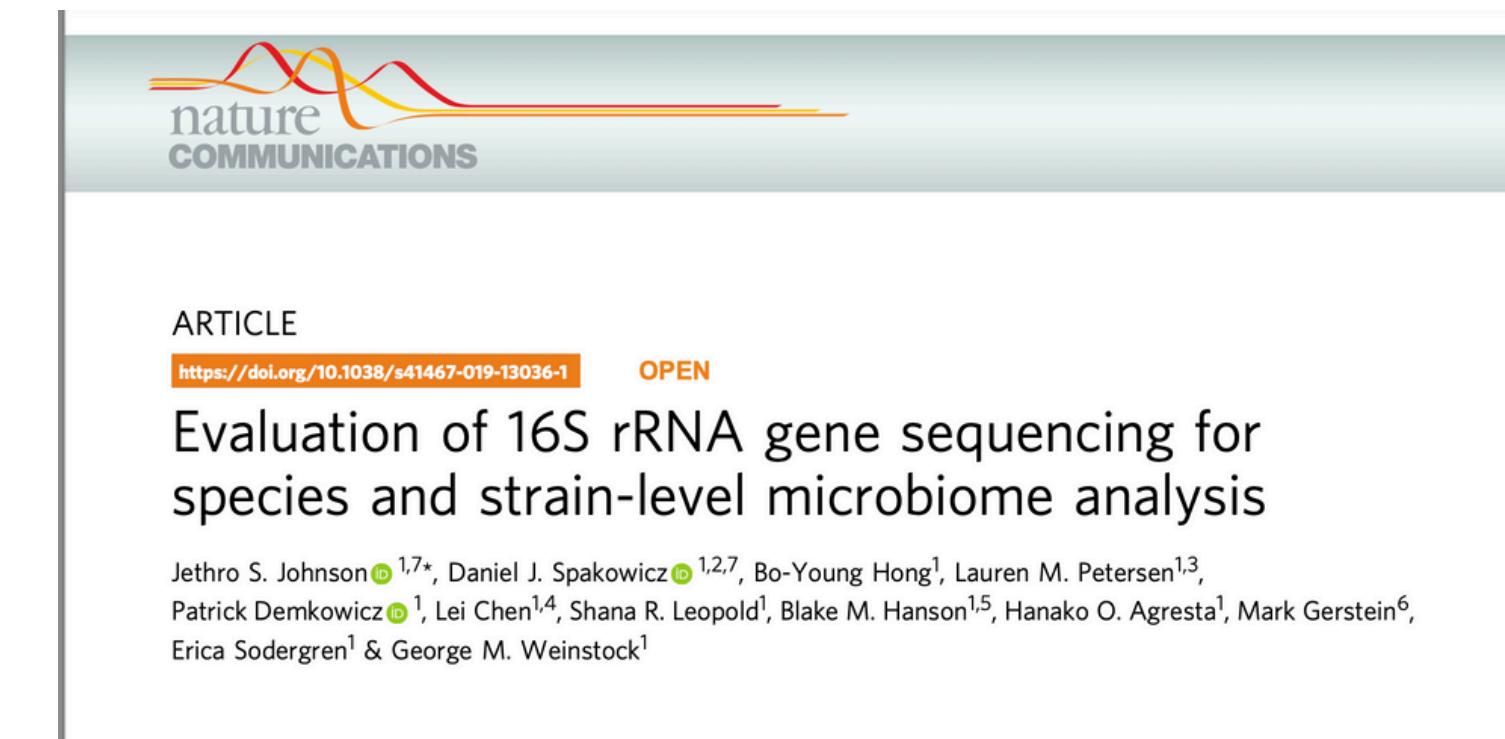
<https://www.dsmz.de>

X: rRNA sequence → Kmer
y: pH, Carbon Source, N Source...
(media and strain detail for cultivation)

Kmer Data

SOTA for 16-S rRNA analysis

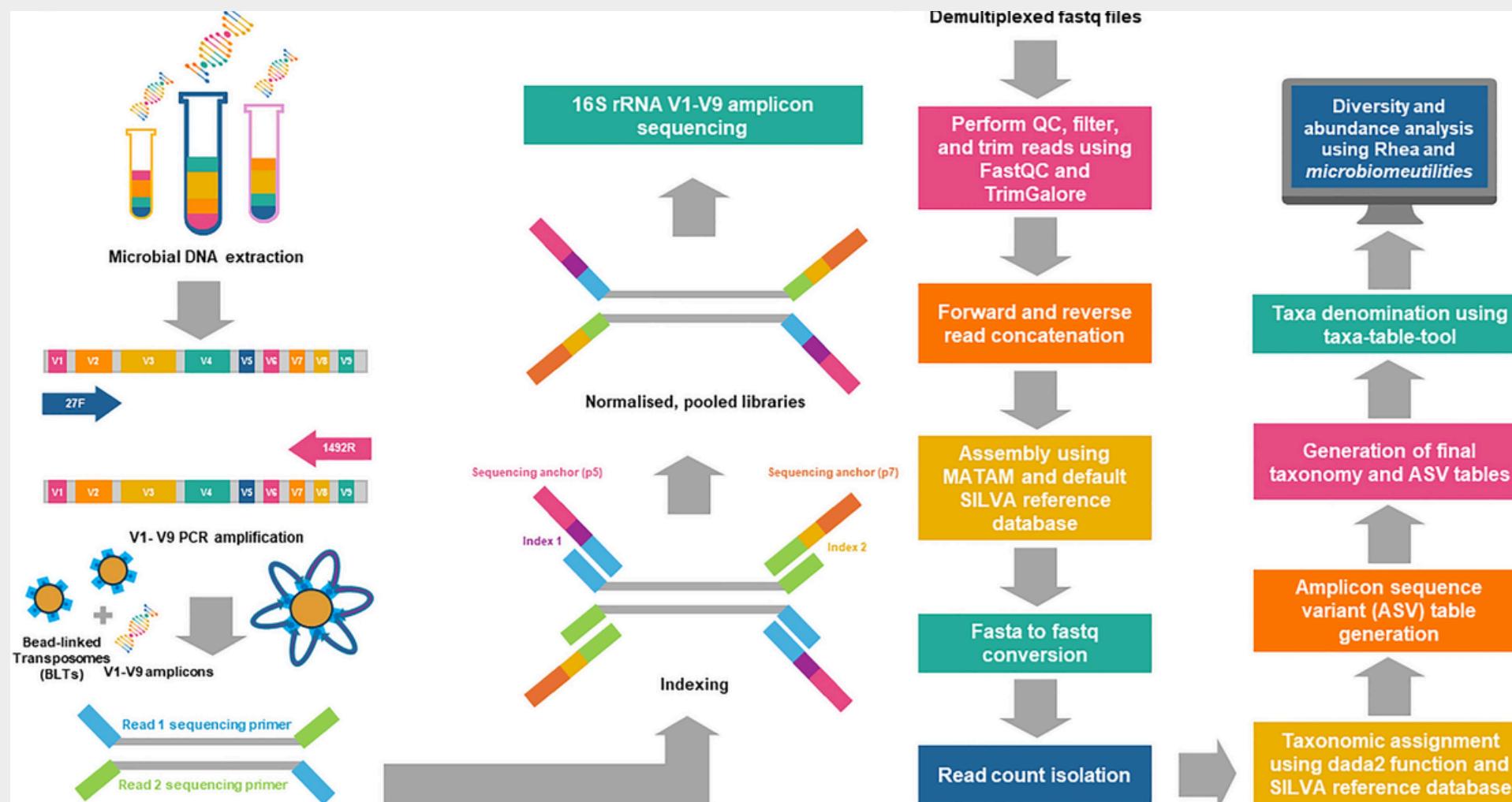
- **Full-Length 16S Sequencing:** Achieves ~90% species-level accuracy vs. 56% for V4 region (PacBio vs. Illumina).
- **Intragenomic Variation:** Resolves strain-level differences in 205 unique SNP profiles across 381 isolates.
- **Nothing really about prediction using rRNA sequencing...**



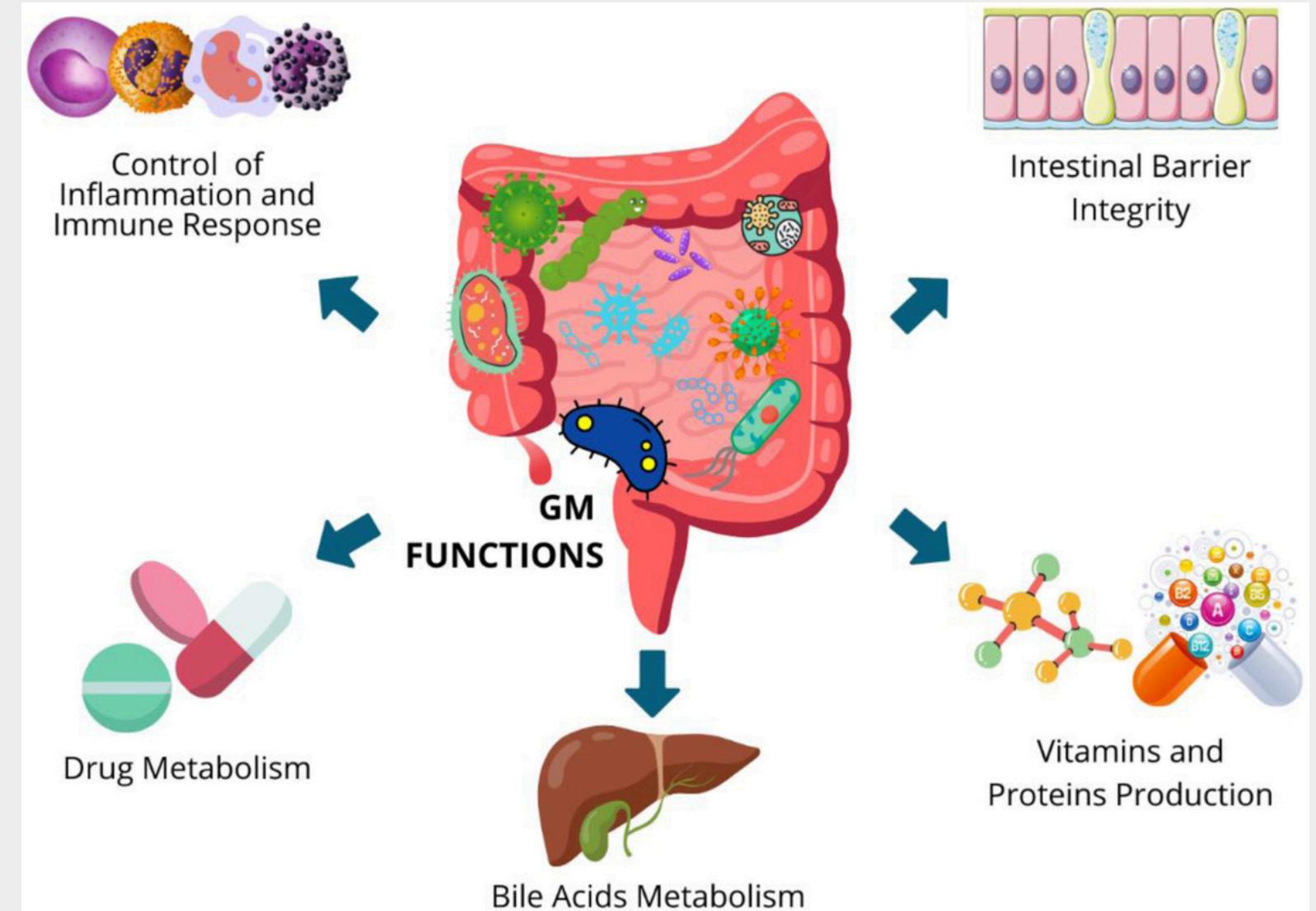
Dataset Further



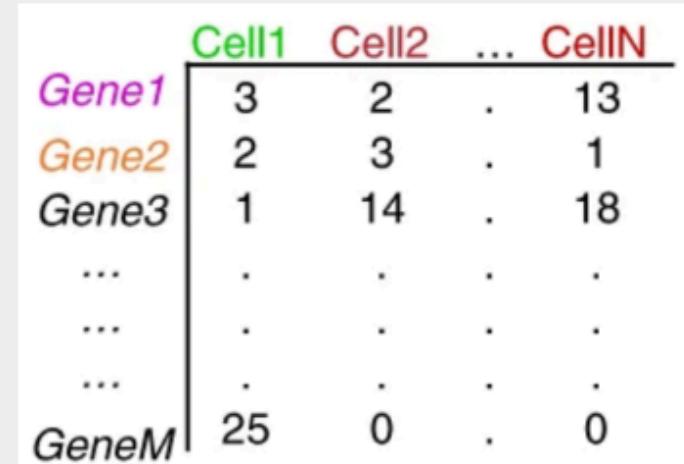
Bacteria rRNA Sequencing



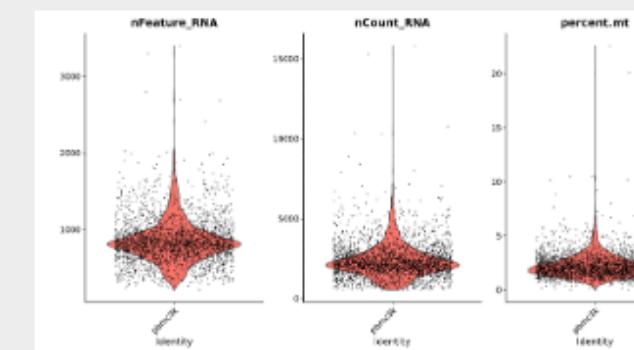
Intestinal Flora Therapy



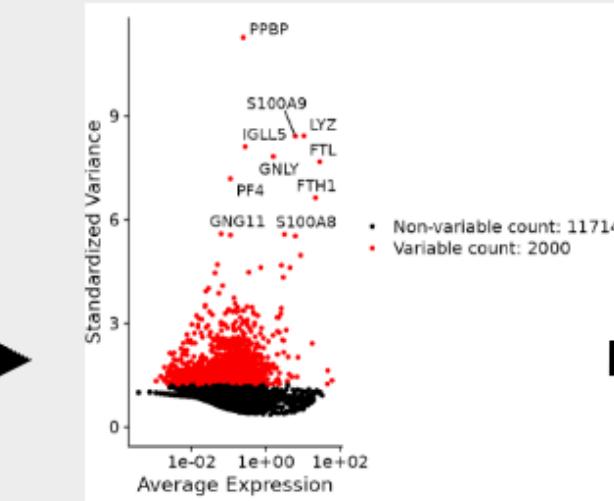
Pipline



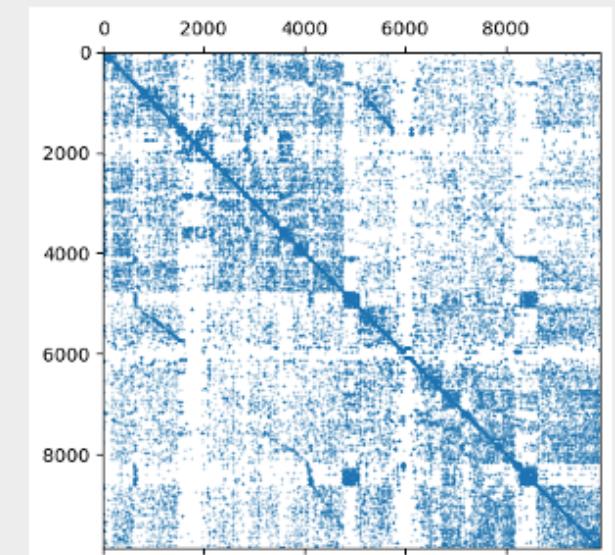
Input



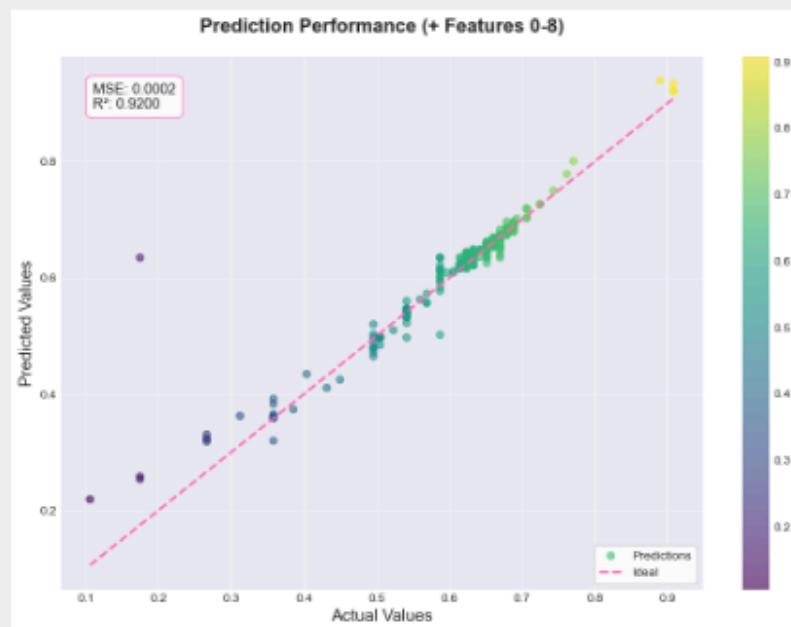
Quality Control



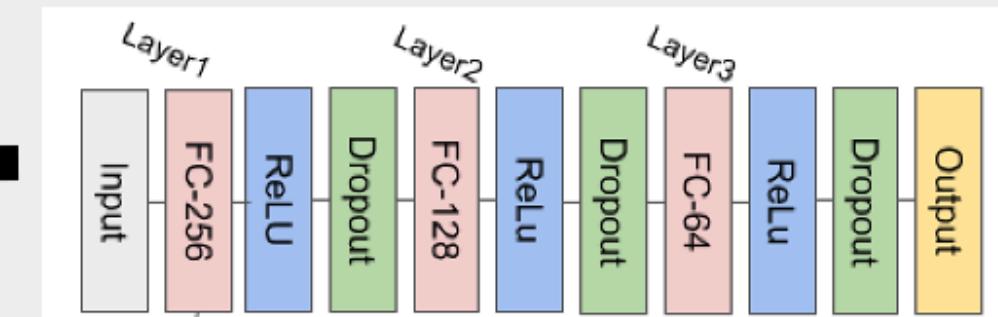
Feature selection



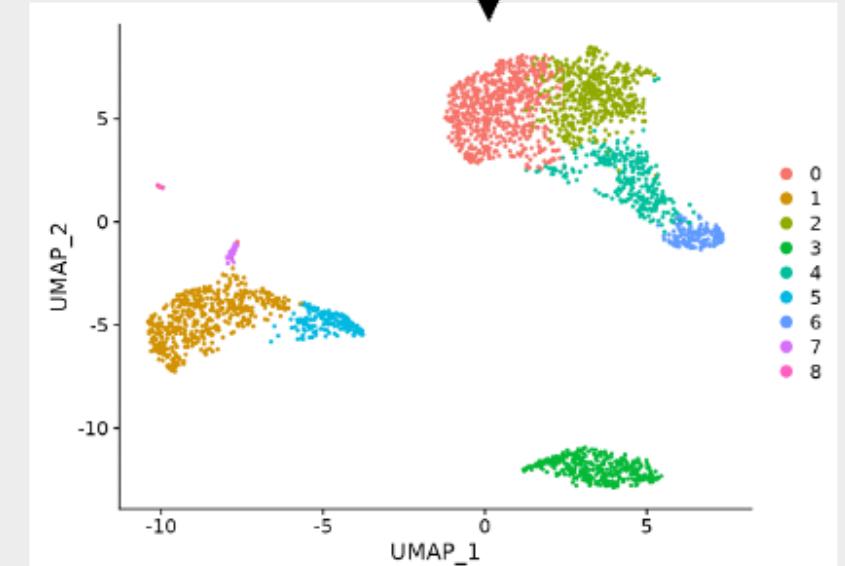
Graph construction



Results



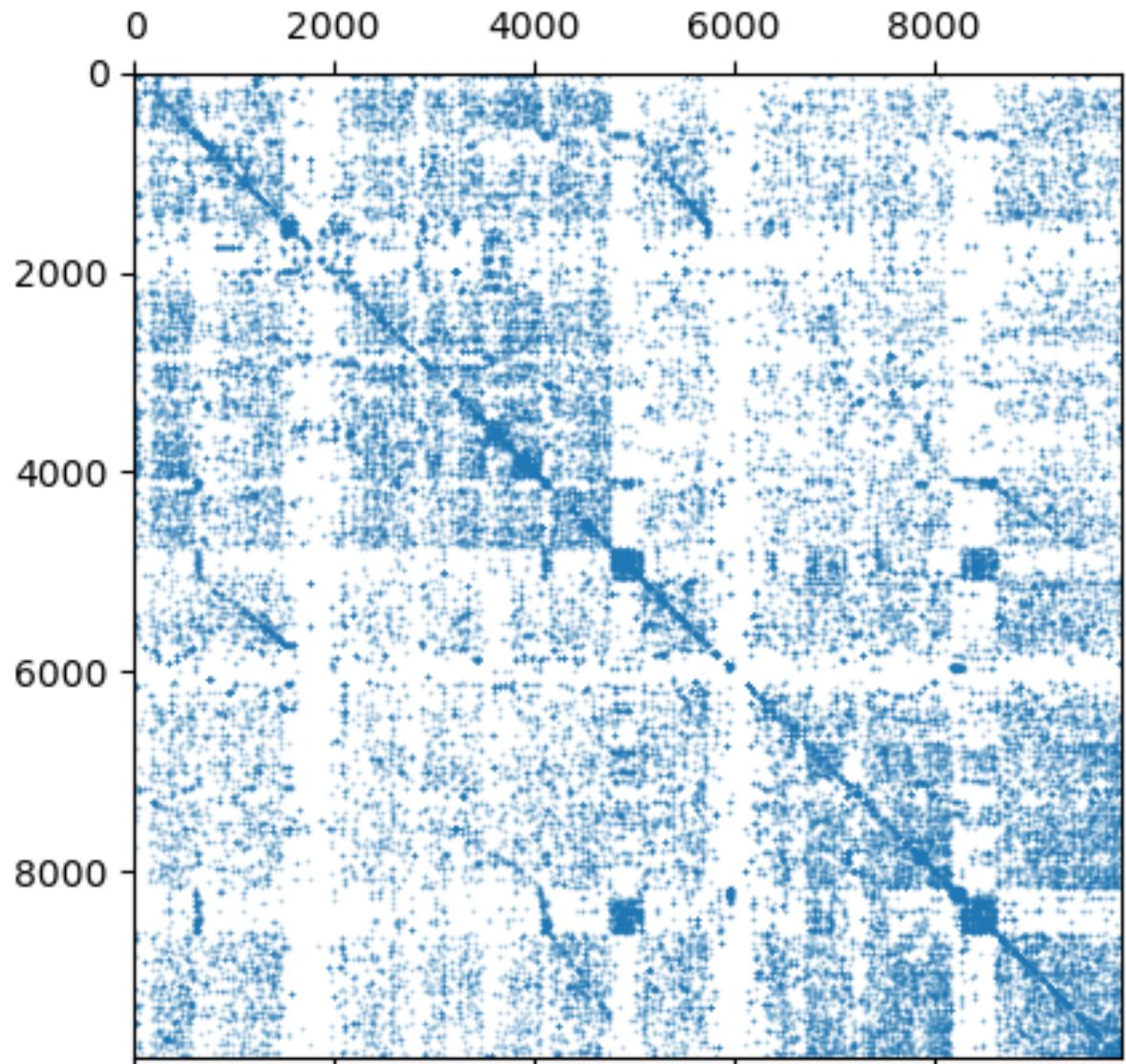
Neural Network



Graph affiliated Clustering

Graph Affiliated Clustering

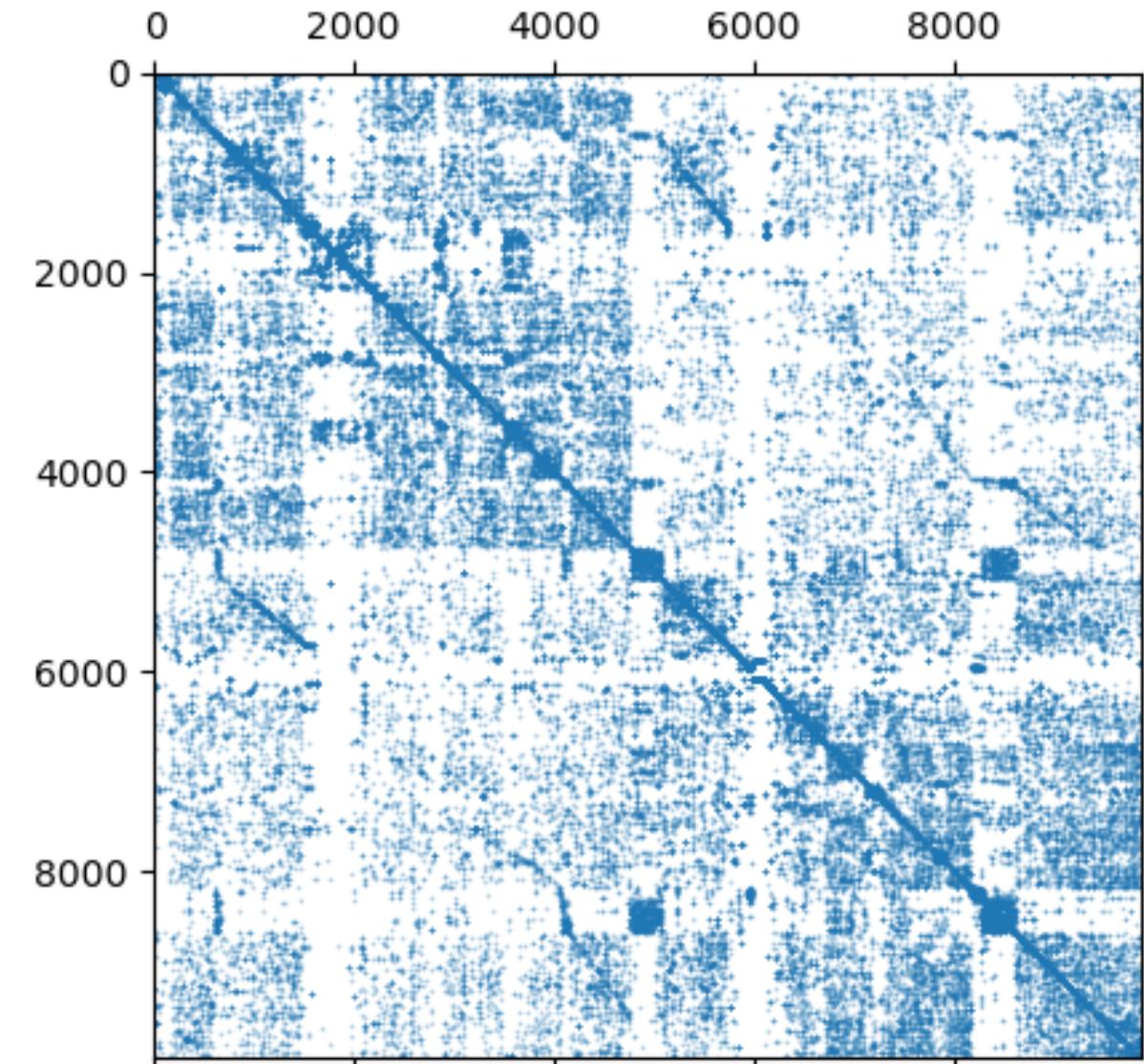
kNN + SGtSNE π λ eq.



D from kNN

$$w(x, y) = \frac{1}{\lambda} \exp \left(-d^2(x, y)/(2\sigma_x^2) \right), \quad x, y \in X, \quad (x, y) \in E(G)$$

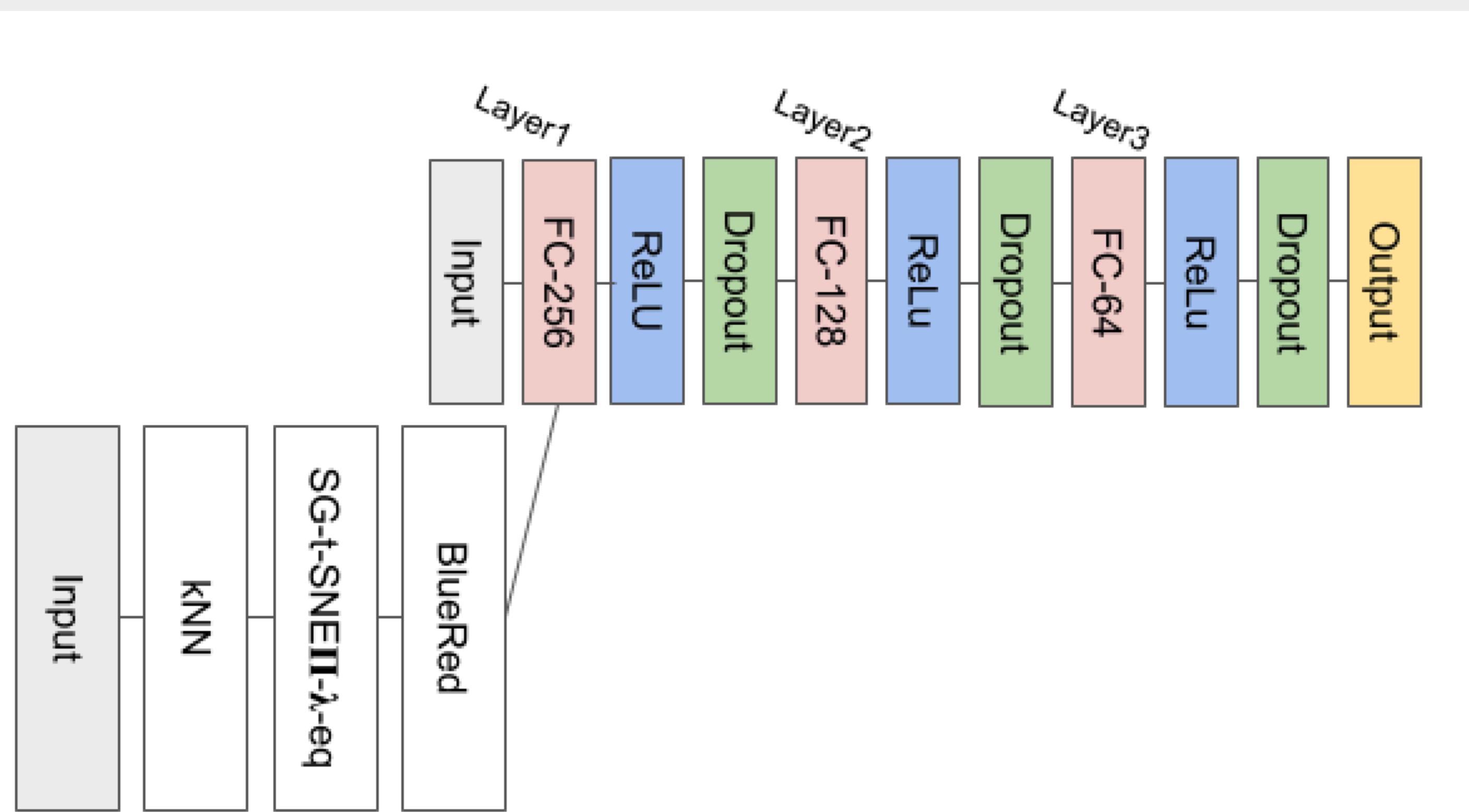
$$\sum_{y:(x,y) \in E} \exp \left(-d^2(x, y)/(2\sigma_x^2) \right) = \lambda, \quad x \in X$$



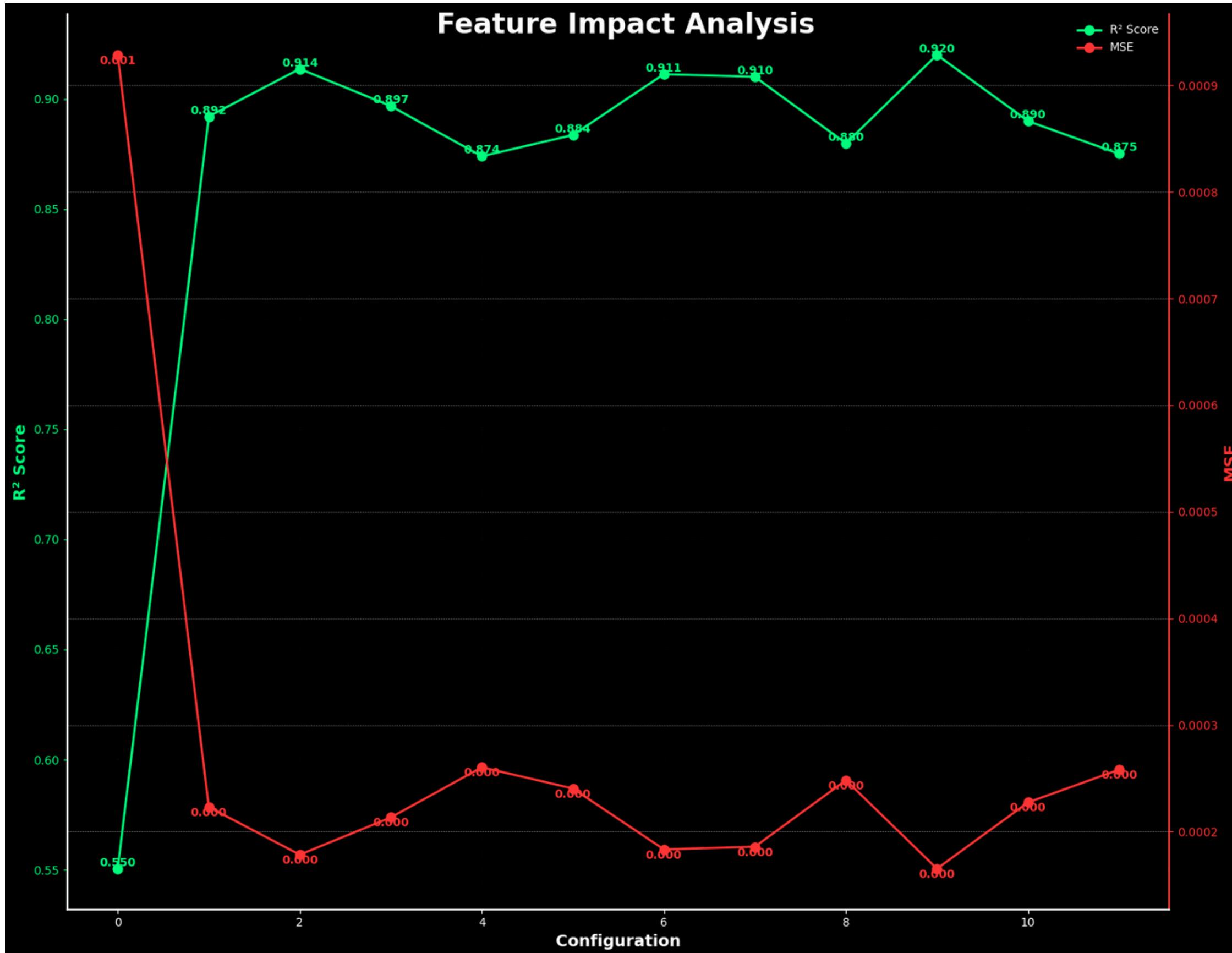
P after SG-t-SNE-Pi-lambda eq

Then BlueRed Clustering...

Neural Network Training



Results



	Features Added	R ²	MSE	R ² Δ	MSE Δ
0	Baseline	0.5505	0.0009	0.0000	0.0000
1	0-0	0.8920	0.0002	0.3416	-0.0007
2	0-1	0.9137	0.0002	0.3632	-0.0008
3	0-2	0.8967	0.0002	0.3463	-0.0007
4	0-3	0.8740	0.0003	0.3235	-0.0007
5	0-4	0.8836	0.0002	0.3332	-0.0007
6	0-5	0.9112	0.0002	0.3608	-0.0007
7	0-6	0.9100	0.0002	0.3596	-0.0007
8	0-7	0.8798	0.0002	0.3294	-0.0007
9	0-8	0.9200	0.0002	0.3696	-0.0008
10	0-9	0.8899	0.0002	0.3395	-0.0007
11	0-10	0.8751	0.0003	0.3246	-0.0007
12	0-11	-0.0000	0.0021	-0.5505	0.0011

How we can go further?

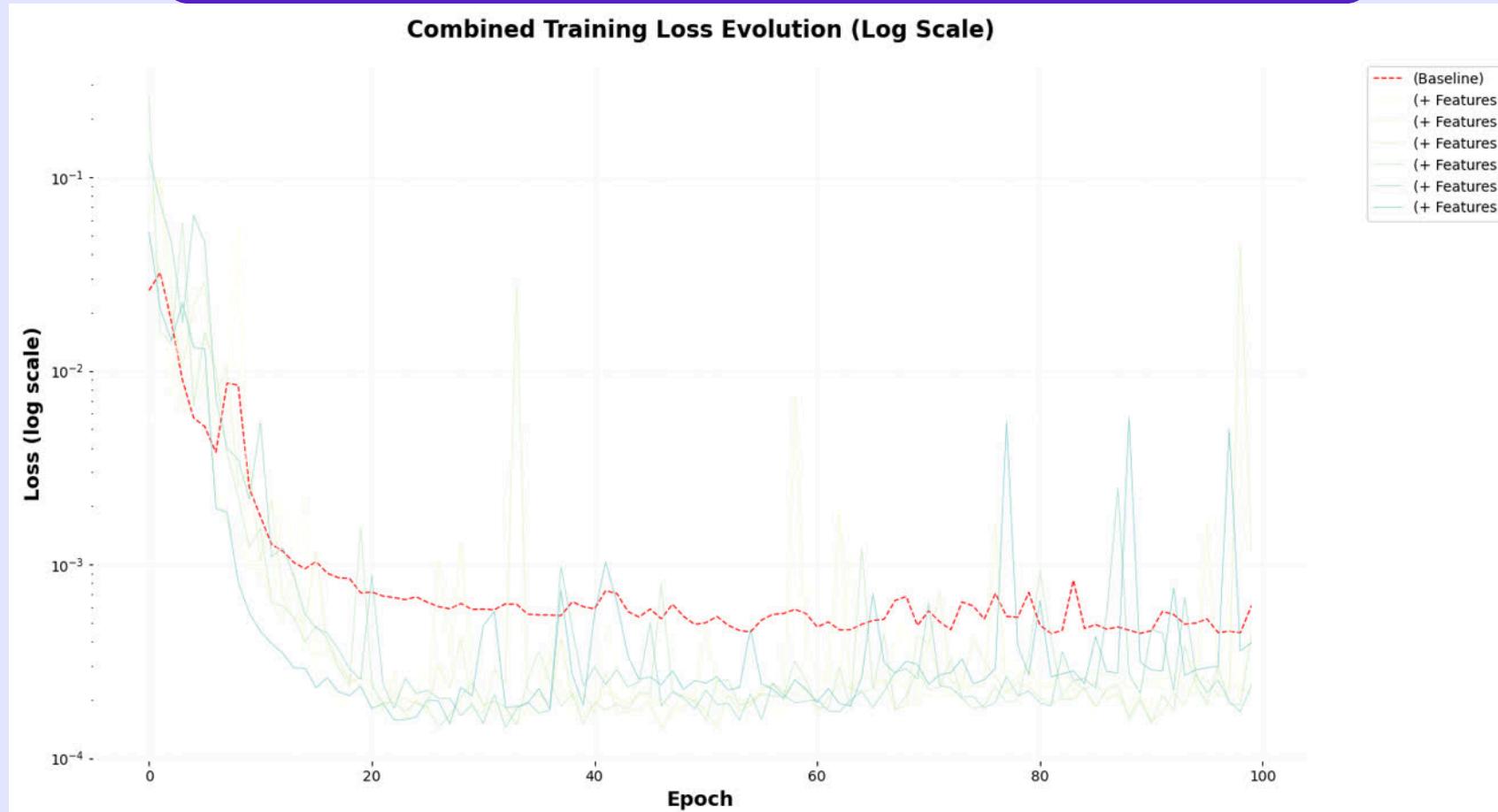
- Other auxiliary learning methods
- Other Neural Networks (go deeper)

Appendix

December 11, 2024

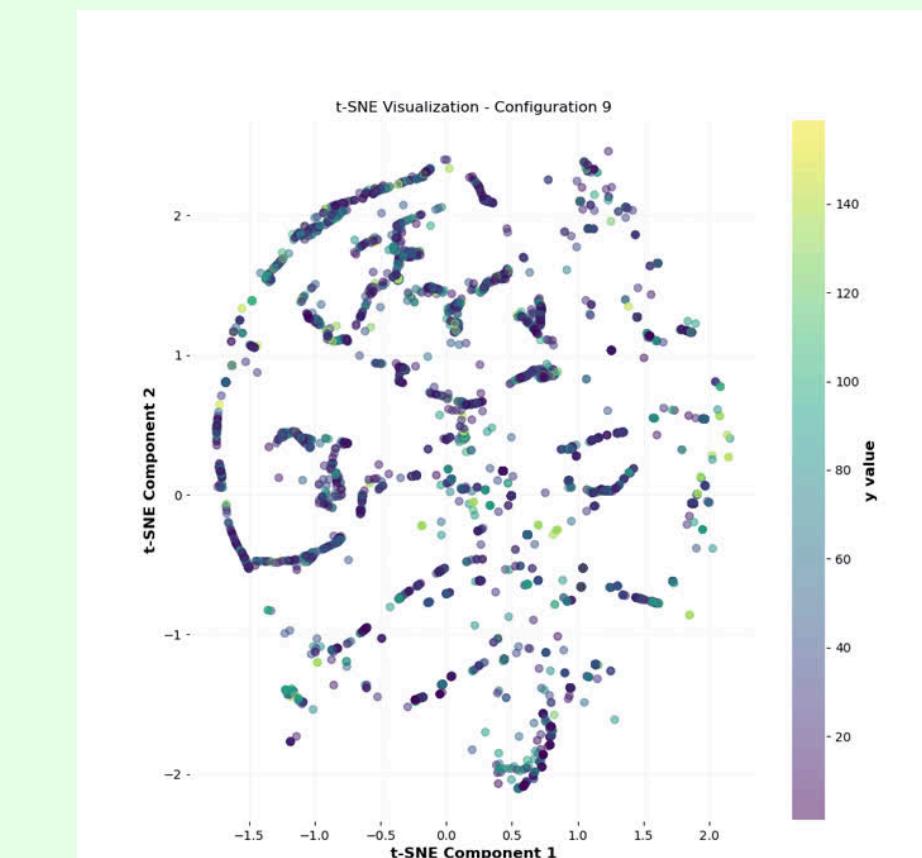
Loss_Evolution

Combined Training Loss Evolution (Log Scale)



tsne_configurations

t-SNE Visualization - Configuration 9



Training Loss Evolution where R²=0.92

Training Loss Evolution (+ Features 0-8)



Training Loss Evolution (Log Scale) (+ Features 0-8)

