Advances in Applied Mathematics 应用数学进展, 2026, 15(3), 1-9
Published Online March 2026 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2026.153082

Hans汉斯

基于深度学习的微生物群落数据分析与预测

张冬艳

青岛大学数学与统计学院，山东 青岛

收稿日期：2026年1月25日；录用日期：2026年2月19日；发布日期：2026年2月28日

摘　要

微生物群落与代谢组相互作用共同影响着宿主健康与环境稳定，随着多组学整合研究逐步发展，之前的方法在捕捉微生物与代谢物之间复杂的非线性关系和异质性方面存在局限。本文提出了一种新的深度学习框架，基于双编码器－解码器架构，通过共享潜在空间实现微生物组和代谢组的双向跨模态预测，并结合数据增强策略提升了模型的稳健性，同时在框架中嵌入诊断分类器以预测疾病的诊断结果。试验结果表明我们的架构在多个真实数据集中的预测精度优于现有方法(如SparseNED、MiMeNet等)，模型在三个数据集上预测代谢物的平均斯皮尔曼相关系数均得到有效提升，同时数据扩增策略使得模型诊断预测的AUC值从0.8425提升至0.8895。

关键词

代谢组学，深度学习，数据增强

Microbial Community Data Analysis and Prediction Using Deep Learning

Dongyan Zhang

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: January 25, 2026; accepted: February 19, 2026; published: February 28, 2026

Abstract

Microbial communities and metabolomes interactively influence host health and environmental stability. With the progressive development of multi-omics integration studies, conventional methods exhibit limitations in capturing the complex nonlinear relationships and heterogeneity between microorganisms and metabolites. This study proposes a novel deep learning framework based on a dual encoder-decoder architecture, which enables bidirectional cross-modal prediction between microbiome and metabolome data through a shared latent space. The framework

文章引用: 张冬艳. 基于深度学习的微生物群落数据分析与预测[J]. 应用数学进展, 2026, 15(3): 1-9.
DOI: 10.12677/aam.2026.153082

incorporates a data augmentation strategy to enhance model robustness and embeds a diagnostic classifier to predict disease outcomes. Experimental results demonstrate that our architecture outperforms existing methods (e.g., SparseNED, MiMeNet) in prediction accuracy across multiple real-world datasets. The average Spearman correlation coefficient for metabolite prediction improved significantly across all three datasets, while the data augmentation strategy increased the AUC value of diagnostic prediction from 0.8425 to 0.8895.

Keywords

Metabolomics, Deep Leaning, Data Augment

1. 引言

微生物群落广泛存在于人类肠道、土壤、海洋等环境里[1] [2]，其多样性和功能复杂性直接影响宿主的健康与疾病状态[3] (如炎症性肠病)，还影响环境物质循环与生态系统的稳定性[4]-[6]。代谢组学可以反映宿主与微生物之间的相互作用、环境因素对代谢过程的影响[7] [8]，可是直接测量的代谢组学实验既困难又昂贵[9]-[11]，而定量微生物群落组成的测序方法已经成熟且成本较低[12] [13]，所以通过计算方法从微生物组成预测代谢谱可以减少原有的工作量。早期的研究主要集中在微生物群落与代谢物的相关性验证方面，利用动物模型和人群队列寻找特定菌群和代谢特征之间的联系[14]。之后随着机器学习的快速发展，微生物组－代谢组研究转向了预测建模。例如构建了分类模型识别短链脂肪酸水平相关的肠道微生物标志物，为炎症性肠病(IBD)诊断提供潜在的研究方向[15]，但机器学习方法还是很难捕捉高维微生物组数据的深层非线性关系，也存在模型的普适性问题。

近年来，深度学习技术在多组学整合研究中展现出显著优势，例如，mmvec [16]用于估计在特定微生物存在时每个代谢产物存在的条件概率，推断微生物与代谢物间的相互作用；SparseNED [17]基于稀疏编码优化特征选择提高了代谢物的可解释性；MiMeNet [18]基于多层感知机(MLPNN [19])建模预测微生物组之间的复杂关系推断识别代谢组的功能关联。这些模型在预测精度与代谢物覆盖范围上比传统方法有显著提升，本文使用新提出的 DEDD 方法与该方法比较。通过双编码器－解码器以及共享潜在空间架构，同步实现优化特征提取和跨模态转换预测，结合数据扩增策略提升模型的稳健性。

2. 方法

2.1. 数据集与预处理

第一个数据集粪便样本 PRISM + NLIBD 来自一项关于 IBD 患者粪便微生物组和代谢物样本的研究报告[20]，整合了两个独立队列的肠道多组学数据，训练集(IBD-PRISM)：来自麻省总医院(MGH)前瞻性炎症肠病等级研究(PRISM)，包含 121 例病患和 34 名健康对照；外部验证集(IBD-External)：由 20 名健康受试者的 NLIBD 健康队列和 43 例病患组成[21]，样本处理生成了 200 种微生物类群和 8848 种代谢物；第二个数据集来自一项囊性纤维化肺的研究 Cystic fibrosis [22]，172 名患者的痰液样本一共检测到 1119 个微生物特征和 168 种代谢物；第三个数据集是沙漠土壤生物结皮样本 Soil [23]，对土壤进行湿润处理产生的微生物和代谢活动，在生物结四个连续演替阶段中每个阶段五个时间节点取样，生成了 19 个

样本，利用鸟枪宏基因组测序和液相色谱－质谱联用(LC/MS)技术检测到 446 种微生物和 85 种代谢物。

对数据进行预处理时首先使用伪计数法填补零值，然后对微生物组和代谢组丰度数据进行中心对数比(CLR)转换：

$$CLR\left(x_i\right)=\log\left[\frac{\left(x_{i1},\cdots,x_{ip}\right)}{\left(\sum_{j=1}^{p}x_{ij}\right)^{\frac{1}{p}}}\right] \tag{1}$$

其中 x_i 是微生物或代谢物丰度的样本向量，p 是特征数量。通过对数变换和几何平均数的引入，CLR 转换使得样本数据变得更加独立，减少了原始数据中的相互依赖性。

2.2. 数据扩增

第一个数据集还额外包含了每个样本的诊断结果，对数据进行微生物组－代谢组分析后，进一步进行诊断结果的预测，但样本量较少，为了防止模型过拟合，进行数据扩增。如图 1(A)所示数据扩增包括跨模态匹配网络训练和合成生成数据两部分，合成数据标签继承原始标签(诊断结果)，训练正样本为真实配对(同一患者的微生物－代谢组对)，负样本为同标签随机错配(不同患者的微生物－代谢组对)，损失函数为三元组损失(Triplet Loss)

$$L=\sum_{a,p,n}\max\left[0,\cos\left(a,n\right)-\cos\left(a,p\right)+\alpha\right] \tag{2}$$

其中 a 是微生物样本，p 是同标签正代谢组，n 是同标签负代谢组，边界超参数 $\alpha=0.2$。

进行跨模态匹配：对每个微生物样本 m_i，从同标签代谢组池中匹配最相似样本 s_j：

$$s_j^{*}=\underset{s_j|y_j=y_i}{\arg\max}\cos\left[Enc_m\left(m_i\right),Enc_s\left(s_j\right)\right] \tag{3}$$

保留相似度 Top 30%的配对(47 对)；然后对匹配生成的样本对(m_i，s_j)进行多样性增强，添加高斯噪声($\mu=0,\sigma=0.05$)：$m_i'=m_i+\epsilon_m$，$s_j'=s_j+\epsilon_s$，重复 3 次噪声扰动，生成 $47\times3=141$ 对新样本。最后生成配对样本与原始配对样本合并共有 296 对配对样本作为新的训练集。

2.3. 模型框架

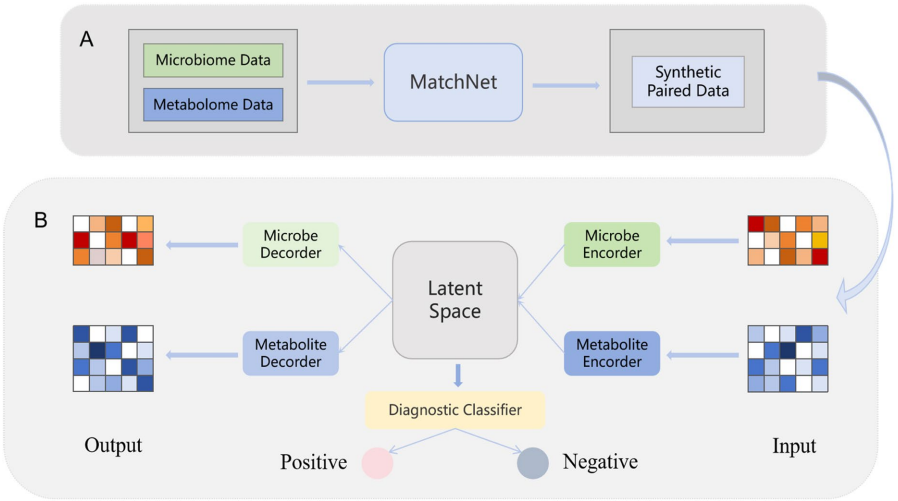

Figure 1. Model workflow diagram
图 1. 模型流程图

DEDD 模型包括四个主要部分：两个编码器网络和两个解码器网络。两个编码器网络对微生物组和代谢组数据进行特征提取。每个自编码器分别负责从输入微生物组数据和代谢组数据中提取潜在特征映射到共享潜在空间中，两个解码器从潜在空间中推断微生物组或代谢组输出，重构原始输入数据。具体流程如上图 1(B)所示。

损失函数我们使用均方误差(MSE)损失函数：$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$，设计每个编码器都能与任意一个解码器组合，使所有网络兼容，则总体损失 L 表示为

$$L = L_{MSE}(x, x_{micro}) + L_{MSE}(y, y_{meta}) + L_{MSE}(x, x_{meta}) + L_{MSE}(y, y_{micro}) + \lambda\sum_{l\in L}\|W_l\|_2 \quad (4)$$

x 和 y 分别表示测量的微生物组和代谢组丰度，下标表示用于推断微生物或代谢物的源模态，前两项反映模型的“重建”能力，后两项反映模型在跨模态预测方面的表现，最后一项表示由参数 λ 控制的 L2 正则化惩罚。

在分析诊断分类任务时额外加上分类损失，即二元交叉熵损失：

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right] \quad (5)$$

其中 N 是样本数量，y_i 是第 i 个样本的真实标签，p_i 是模型预测的第 i 个样本属于类别 1 的概率，使用 sigmoid 激活函数将模型输出转化为概率值。通过这种损失，模型能够在同一个训练过程中综合考虑多个目标，并且通过优化总损失同时提升多个任务的性能。

我们利用 ADAM 优化器训练模型，根据训练集上的五折交叉验证结果选择了最优超参数集，在交叉验证时每个数据集被分为两个子集：80%用于训练，20%用于测试。对每个训练区分的 80%数据进一步分为 80%用于模型训练和 20%用于验证。为了防止过拟合，模型采用了早停策略，当验证集的损失在连续五个时期内没有明显改善，训练就会提前停止。对于第一个 IBD 数据集，最终模型是在完整的 IBD (PRISM) 扩增数据集上训练的，在 IBD 外部数据集上进行评估。

3. 结果

3.1. 在预测代谢组方面 DEDD 表现出更好的效果

在进行预测效果评估对比时，我们用到了三个评价指标，平均斯皮尔曼相关系数 $\bar{\rho}$、前 50 名的平均斯皮尔曼相关系数 $\bar{\rho}_{50}$ 和 SCC 值大于 0.5 的代谢物数量 N。

斯皮尔曼相关系数(Spearman’s rank correlation coefficient)衡量的是模型预测值和真实值之间的关系强度，$\bar{\rho}$ 即为对每个代谢物计算其预测效果排名与真实效果排名之间的斯皮尔曼相关系数，然后对所有代谢物取平均值。它可以反映模型所有代谢物上的整体排序的准确性，$\bar{\rho}$ 越接近 1 说明模型能较好地把有效代谢物排在无效的代谢物之前，它是评估模型全局性能的一个核心指标。

$\bar{\rho}_{50}$ 就是只计算模型预测结果中排名前 50 的代谢物的斯皮尔曼相关系数平均值，在讨论微生物和代谢物作用关系预测代谢物浓度时我们更关注排名靠前的最可能有效的代谢物，$\bar{\rho}_{50}$ 用来评估模型在最有效代谢物上的排序准确性，避免整体指标被大量无关代谢物影响。

SCC 值大于 0.5 的代谢物数量 N 就是统计所有预测代谢物中斯皮尔曼相关系数大于 0.5 的代谢物数量。它可以衡量模型预测结果可靠的比例，SCC > 0.5 表示预测和真实值之间有较强的关联，也就是关注可解释性强且预测置信度高的代谢物，如果 N 较大就说明模型可用性比较广。综合使用这三个评价指标可以帮助我们判断模型整体预测的性能，能否可靠地优先选择高潜力代谢物以及能否在不同代谢物上保持稳健性。

将 DEDD 与其他方法进行对比测试，包括 mmvec、线性模型、SparseNED 和 MiMeNet，这几种方法都是用于识别微生物组－代谢组之间的复杂关系，并从微生物组成预测代谢物浓度。将这些方法都用于配对的微生物组－代谢组真实数据中，结果表明，DEDD 在预测代谢组方面优于其他方法。

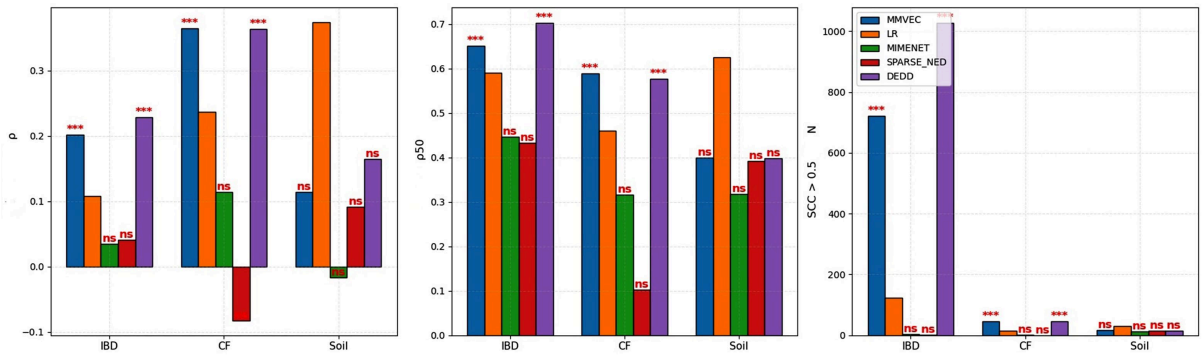

Figure 2. Bar chart comparing DEDD with other methods
图 2. DEDD 与其他方法对比柱状图

我们最开始先使用了 PRISM 和 NLIBD 数据集，这个数据集手机里克罗恩病患者、溃疡性结肠炎患者以及健康个体的粪便样本。包含两个独立的 IBD 队列：一个由 155 名成员组成的队列在马萨诸塞州总医院收集(PRISM)，另一个在荷兰收集的由 62 名成员组成的验证队列(NLIBD/LLDeep)。对于这个数据集，我们先根据 PRISM 扩增数据的五折交叉验证选择了 DEDD 的超参数，在整个 PRISM 扩增数据上使用这个超参数集训练 DEDD，得到了外部队列代谢组的预测值。为了验证 DEDD 在其他不同环境微生物群落的表现的优越性，又在 172 名囊性纤维化患者的肺痰样本和 19 个连续五次湿润后沙漠土壤生物结皮样本这两个配对数据集上进行了测试。结果如图 2 所示，图中展示了五种方法的平均斯皮尔曼相关系数、前 50 名的平均斯皮尔曼相关系数以及 SCC 值大于 0.5 的代谢物数量 N。对于土壤生物结皮数据可能由于样本量过于少，导致进行统计显著性检验时模型不显著，对比的深度学习模型不适合小样本数据，这时线性回归就显示其优势。最后综合三个评价指标来说我们的模型在数据集上的预测效果相对较好。

3.2. 各组件对模型性能的影响

因为 PRISM 和 NLIBD 数据集不止包含配对的微生物组和代谢组丰度数据，每个样本的诊断结果也包含在里面，想在 DEDD 的基础上进行诊断预测，再进一步评估其在临床诊断中的潜力。而 PRISM 队列只有 155 个样本信息，这样不做其他处理训练 DEDD 容易影响模型诊断结果预测的准确性，所以我们采用了数据扩增技术，具体过程如图 1(A)所示，把配对数据拆分为微生物池和代谢组池，在预训练好的跨模态匹配网络中进行匹配：每个微生物样本在同标签代谢组池中匹配最相似的样本，只保留相似度前30%的配对，然后进行多样性增强合成更多的数据，把合成配对数据与原始配对数据合并形成一个增强的训练集。

另外为了验证模型框架各部分对最终性能的独立贡献，进行了消融实验，结果如图 3 和图 4 所示。将双编码器改为单编码器验证其效果，从以上四个评估标准均能发现，双编码器架构相较于单编码器提高了模型预测代谢组丰度的性能；同时我们设计的数据扩增方式，显著增加了样本量，提高了模型对复杂关系的学习能力，从而相较于原始数据集提升了预测性能。

模型在进行诊断预测时，我们使用微生物丰度的潜在表示来预测疾病状态，正如图 5 和图 6 所示，训练集和测试集的三种潜在表示中使用微生物丰度的潜在表示对诊断预测的效果更有利，测试集的效果

与训练集的预测效果也相吻合。从潜在空间的可视化图来看，对于不同疾病状态下的样本分布，我们的模型确实学习到了有区分度的特征表示。

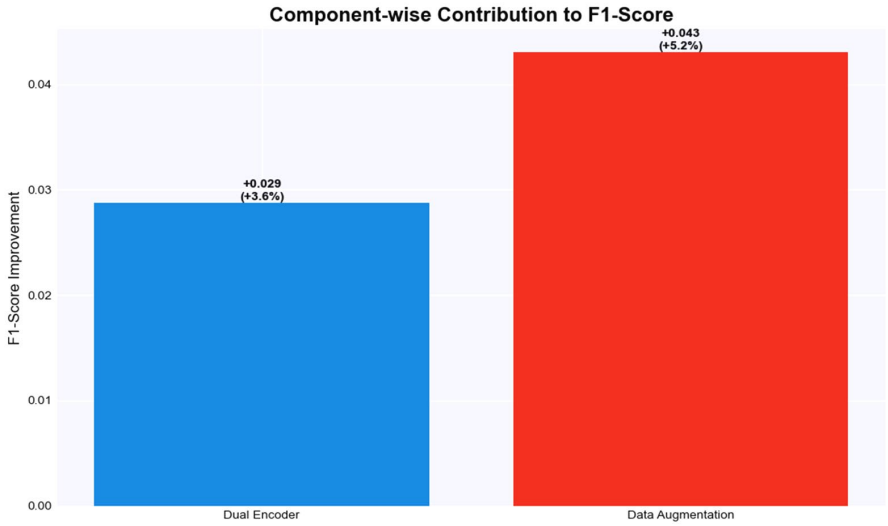

Figure 3. Component-wise contribution to the F1 score
图 3. 各组件对 F1 分数的贡献

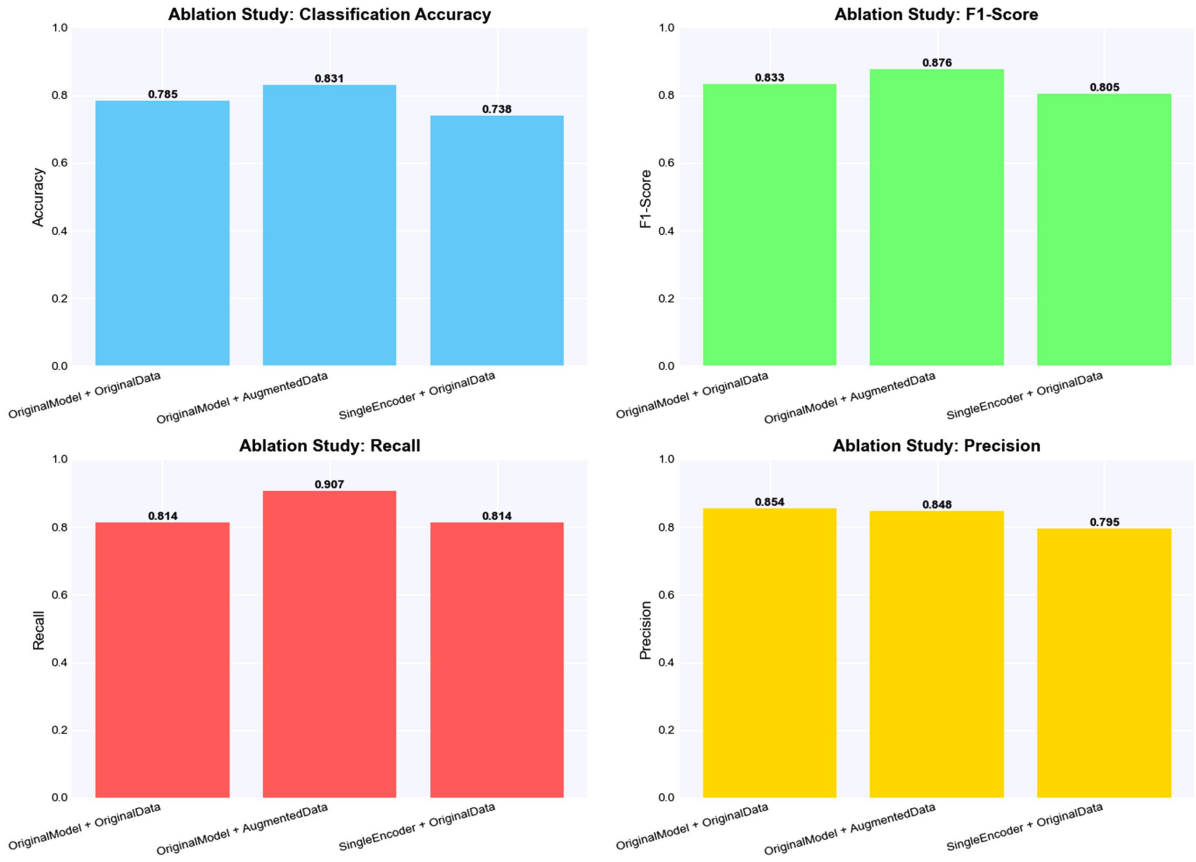

Figure 4. Ablation study
图 4. 消融实验

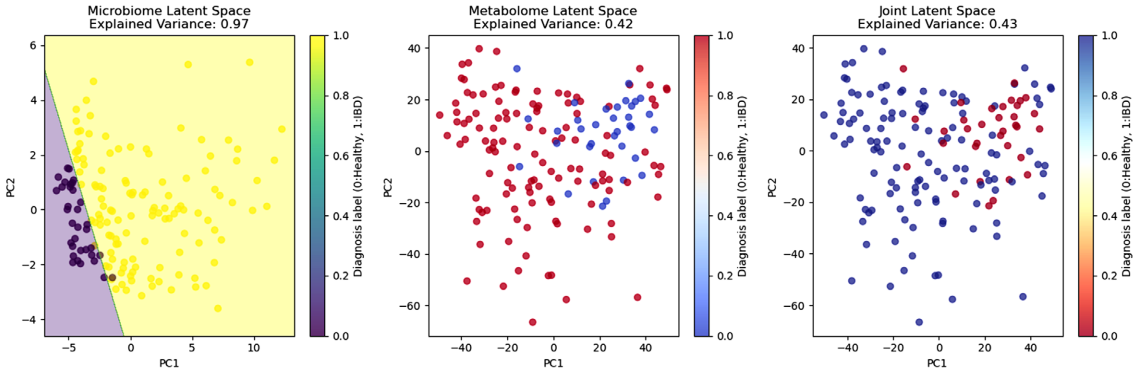

Figure 5. Latent space visualization by disease status on the train set
图 5. 训练集按疾病状态可视化的潜在空间

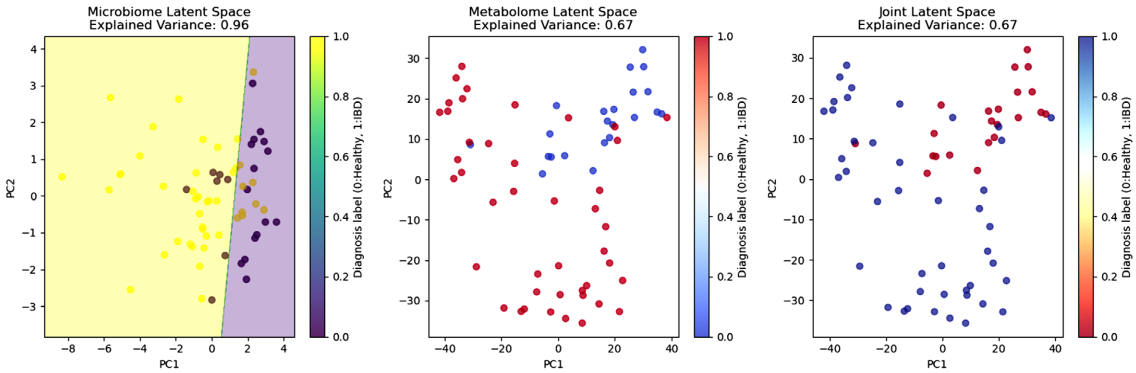

Figure 6. Latent space visualization by disease status on the test set
图 6. 测试集按疾病状态可视化的潜在空间

3.3. DEDD 实现代谢组反向预测微生物组

如模型流程图 1(A)所示，我们的模型采用了两个编码器和两个解码器的结构，且每一对编码器和解码器相互兼容，共同协作以实现代谢组与微生物组之间的有效互预测。与之前相关的深度学习模型框架不同的是，代谢组解码器不仅能够成功解码代谢组信息，还能够通过与微生物组编码器的相互作用，反向推测出微生物组丰度，这种多模块相互作用的设计使得代谢组和微生物组之间的关系可以在模型中被更准确地捕捉和预测到，显著超越了过去方法的限制。

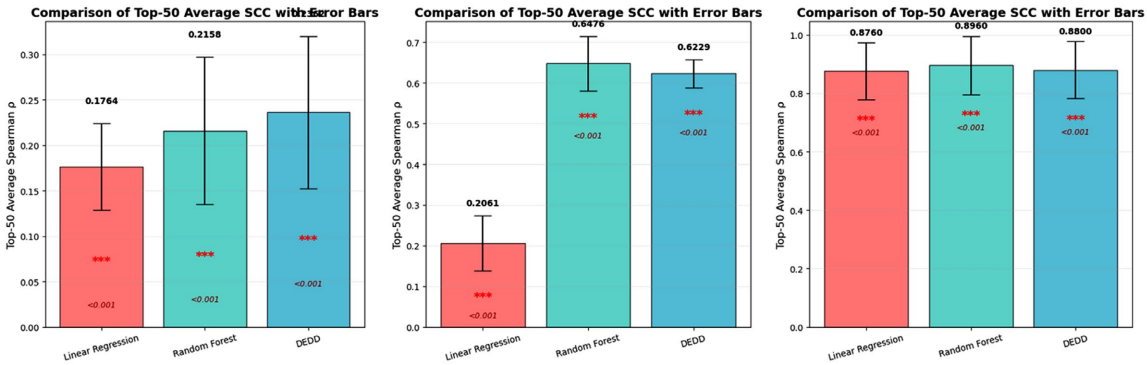

Figure 7. The comparison of reverse prediction performance among IBD, Soil and Cystic fibrosis three datasets
图 7. IBD、土壤和囊性纤维化肺三个数据集的反向预测效果对比

代谢组反向预测微生物组是我们模型的创新应用，为了全面评估模型的表现，我们将该反向预测应用到之前使用的三个不同的数据集，且与线性回归和随机森林两个模型做了对比，并针对每个数据集计算了原始与预测微生物丰度的前 50 名的平均斯皮尔曼相关系数 $\bar{\rho}_{50}$。

如图 7 所示，从左到右依次为 IBD、土壤生物结皮和囊性纤维化肺数据，我们的模型在不同环境和数据集中的适应性表现优异，能够有效地进行代谢组到微生物组的反向预测。从结果来看，模型的预测性能在不同数据集上有显著差异，与线性回归和随机森林两个传统模型相比，我们的模型 DEDD 仅在 IBD 数据集上表现出了明显的优势，说明 DEDD 在处理高维数据时反向预测微生物组数据效果较好。

4. 总结

我们提出的 DEDD 框架在微生物组–代谢组数据的跨模态预测方面取得了较为显著的进展，通过设计的互相可兼容的双编码器——解码器以及共享潜在空间的架构，模型有效地连接了微生物组成与代谢组谱，实现了两者的双向预测，还具有预测患病结果的能力。DEDD 在多个数据集上的表现优于现有的方法，数据扩增也使得诊断预测的效果得到了提升。

参考文献

[1] Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L. and Gordon, J.I. (2011) Human Nutrition, the Gut Microbiome and the Immune System. *Nature*, **474**, 327-336. https://doi.org/10.1038/nature10213

[2] Ghaisas, S., Maher, J. and Kanthasamy, A. (2016) Gut Microbiome in Health and Disease: Linking the Microbiome–gut-Brain Axis and Environmental Factors in the Pathogenesis of Systemic and Neurodegenerative Diseases. *Pharmacology & Therapeutics*, **158**, 52-62. https://doi.org/10.1016/j.pharmthera.2015.11.012

[3] McHardy, I.H., Goudarzi, M., Tong, M., Ruegger, P.M., Schwager, E., Weger, J.R., *et al*. (2013) Integrative Analysis of the Microbiome and Metabolome of the Human Intestinal Mucosal Surface Reveals Exquisite Inter-Relationships. *Microbiome*, **1**, Article No. 17. https://doi.org/10.1186/2049-2618-1-17

[4] Pérez-Cobas, A.E., Gosalbes, M.J., Friedrichs, A., Knecht, H., Artacho, A., Eismann, K., *et al*. (2012) Gut Microbiota Disturbance during Antibiotic Therapy: A Multi-Omic Approach. *Gut*, **62**, 1591-1601. https://doi.org/10.1136/gutjnl-2012-303184

[5] Barton, W., Penney, N.C., Cronin, O., Garcia-Perez, I., Molloy, M.G., Holmes, E., *et al*. (2018) The Microbiome of Professional Athletes Differs from That of More Sedentary Subjects in Composition and Particularly at the Functional Metabolic Level. *Gut*, **67**, 625-633. https://doi.org/10.1136/gutjnl-2016-313627

[6] Antharam, V.C., McEwen, D.C., Garrett, T.J., Dossey, A.T., Li, E.C., Kozlov, A.N., *et al*. (2016) An Integrated Metabolomic and Microbiome Analysis Identified Specific Gut Microbiota Associated with Fecal Cholesterol and Coprostanol in Clostridium Difficile Infection. *PLOS ONE*, **11**, e0148824. https://doi.org/10.1371/journal.pone.0148824

[7] Parker, A., Lawson, M.A.E., Vaux, L. and Pin, C. (2018) Host-Microbe Interaction in the Gastrointestinal Tract. *Environmental Microbiology*, **20**, 2337-2353. https://doi.org/10.1111/1462-2920.13926

[8] Martin, A.M., Sun, E.W., Rogers, G.B. and Keating, D.J. (2019) The Influence of the Gut Microbiome on Host Metabolism through the Regulation of Gut Hormone Release. *Frontiers in Physiology*, **10**, Article 428. https://doi.org/10.3389/fphys.2019.00428

[9] Yang, Q., Zhang, A., Miao, J., Sun, H., Han, Y., Yan, G., *et al*. (2019) Metabolomics Biotechnology, Applications, and Future Trends: A Systematic Review. *RSC Advances*, **9**, 37245-37257. https://doi.org/10.1039/c9ra06697g

[10] Castelli, F.A., Rosati, G., Moguet, C., Fuentes, C., Marrugo-Ramírez, J., Lefebvre, T., *et al*. (2022) Metabolomics for Personalized Medicine: The Input of Analytical Chemistry from Biomarker Discovery to Point-of-Care Tests. *Analytical and Bioanalytical Chemistry*, **414**, 759-789. https://doi.org/10.1007/s00216-021-03586-z

[11] Dias-Audibert, F.L., Navarro, L.C., de Oliveira, D.N., Delafiori, J., Melo, C.F.O.R., Guerreiro, T.M., *et al*. (2020) Combining Machine Learning and Metabolomics to Identify Weight Gain Biomarkers. *Frontiers in Bioengineering and Biotechnology*, **8**, Article 6. https://doi.org/10.3389/fbioe.2020.00006

[12] Ayling, M., Clark, M.D. and Leggett, R.M. (2020) New Approaches for Metagenome Assembly with Short Reads. *Briefings in Bioinformatics*, **21**, 584-594. https://doi.org/10.1093/bib/bbz020

[13] Brumfield, K.D., Huq, A., Colwell, R.R., Olds, J.L. and Leddy, M.B. (2020) Microbial Resolution of Whole Genome Shotgun and 16S Amplicon Metagenomic Sequencing Using Publicly Available NEON Data. *PLOS ONE*, **15**, e0228899.

https://doi.org/10.1371/journal.pone.0228899

[14] Johnson, C.H. and Gonzalez, F.J. (2012) Challenges and Opportunities of Metabolomics. *Journal of Cellular Physiology*, **227**, 2975-2981. https://doi.org/10.1002/jcp.24002

[15] Koh, A., De Vadder, F., Kovatcheva-Datchary, P. and Bäckhed, F. (2016) From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell*, **165**, 1332-1345. https://doi.org/10.1016/j.cell.2016.05.041

[16] Morton, J.T., Aksenov, A.A., Nothias, L.F., Foulds, J.R., Quinn, R.A., Badri, M.H., *et al*. (2019) Learning Representations of Microbe-Metabolite Interactions. *Nature Methods*, **16**, 1306-1314. https://doi.org/10.1038/s41592-019-0616-3

[17] Le, V., Quinn, T.P., Tran, T. and Venkatesh, S. (2020) Deep in the Bowel: Highly Interpretable Neural Encoder-Decoder Networks Predict Gut Metabolites from Gut Microbiome. *BMC Genomics*, **21**, Article No. 256. https://doi.org/10.1186/s12864-020-6652-7

[18] Reiman, D., Layden, B.T. and Dai, Y. (2021) MiMeNet: Exploring Microbiome-Metabolome Relationships Using Neural Networks. *PLOS Computational Biology*, **17**, e1009021. https://doi.org/10.1371/journal.pcbi.1009021

[19] Mallick, H., Franzosa, E.A., Mclver, L.J., Banerjee, S., Sirota-Madi, A., Kostic, A.D., *et al*. (2019) Predictive Metabolomic Profiling of Microbial Communities Using Amplicon or Metagenomic Sequences. *Nature Communications*, **10**, Article No. 3136. https://doi.org/10.1038/s41467-019-10927-1

[20] Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker, S., *et al*. (2018) Gut Microbiome Structure and Metabolic Activity in Inflammatory Bowel Disease. *Nature Microbiology*, **4**, 293-305. https://doi.org/10.1038/s41564-018-0306-4

[21] Tigchelaar, E.F., Zhernakova, A., Dekens, J.A.M., Hermes, G., Baranska, A., Mujagic, Z., *et al*. (2015) Cohort Profile: Lifelines DEEP, a Prospective, General Population Cohort Study in the Northern Netherlands: Study Design and Baseline Characteristics. *BMJ Open*, **5**, e006772. https://doi.org/10.1136/bmjopen-2014-006772

[22] Quinn, R.A., Comstock, W., Zhang, T., Morton, J.T., da Silva, R., Tran, A., *et al*. (2018) Niche Partitioning of a Pathogenic Microbiome Driven by Chemical Gradients. *Science Advances*, **4**, eaau1908. https://doi.org/10.1126/sciadv.aau1908

[23] Swenson, T.L., Karaoz, U., Swenson, J.M., Bowen, B.P. and Northen, T.R. (2018) Linking Soil Biology and Chemistry in Biological Soil Crust Using Isolate Exometabolomics. *Nature Communications*, **9**, Article No. 19. https://doi.org/10.1038/s41467-017-02356-9