

# 集成学习模型架构图

## 特征类型:

- 微生物组 (424维)
- 代谢组 (212维)
- 联合特征 (636维)

## Level 0 特点:

- 使用原始特征训练
- 5折交叉验证
- 生成out-of-fold预测

## Level 1 特点:

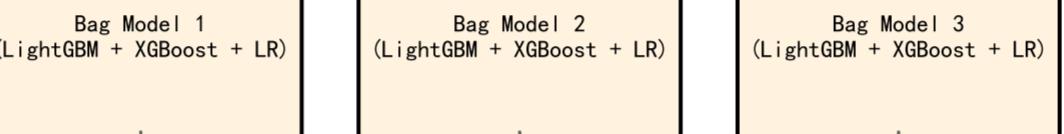
- 使用Level 0预测+原始特征
- 5折交叉验证
- 生成out-of-fold预测

## Level 2 特点:

- 使用Level 1预测训练
- 神经网络架构
- 超参数优化

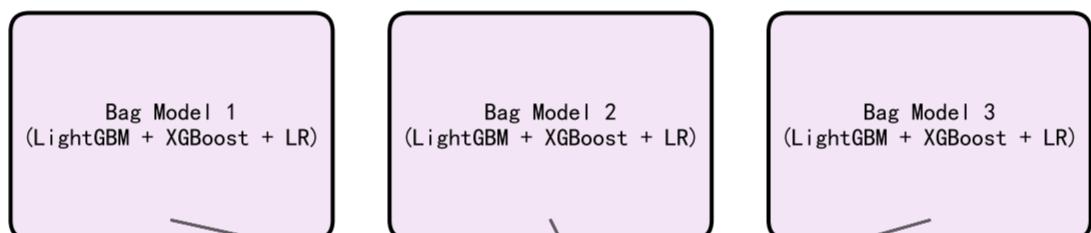
输入数据  
(微生物组 + 代谢组 + 联合特征)

Level 0: 多个Bag Models (使用原始特征)



Level 0 预测结果

Level 1: 多个Bag Models (Level 0预测 + 原始特征)



Level 1 预测结果

Level 2: Meta Learner (神经网络)

Meta Learner  
(多层感知机)

最终预测  
(概率 + 标签)

## 外部预测集:

- 19个样本
- 健康对照: 10个
- 口腔鳞癌: 9个
- 完全独立于训练集

## 评估指标:

- F1分数 (主要指标)
- AUC
- 准确率
- 精确率/召回率
- 特异性

## 技术细节:

- 分层交叉验证
- 超参数优化 (Optuna)
- 早停机制
- 模型持久化

## 模型性能:

- 训练集F1: 0.95
- 训练集AUC: 0.99
- 外部预测F1: 0.89
- 外部预测AUC: 0.94

## 数据统计:

- 训练样本: 231个
- 特征总数: 636维
- 交叉验证: 5折
- 模型数量: 9个