Machine Learning Identification of Gut Microbiome Signatures Associated with Thyroid Cancer

# 1 前言

在现代医学研究中，人体微生物组的作用受到了广泛关注，尤其是肠道菌群对宿主健康的影响。肠道菌群与宿主的新陈代谢、免疫功能以及某些疾病的发生发展，包括癌症，都有着密切的联系。甲状腺癌作为最常见的内分泌恶性肿瘤，其发病率近年来不断上升，这促使科研人员寻找新的生物标志物，以期改善诊断、预后评估和治疗策略。

本研究旨在探索肠道微生物群在甲状腺癌中的潜在标志物，我们通过分析与甲状腺癌相关的肠道菌群数据，尝试揭示肠道微生物与甲状腺癌之间的关联。我们从NCBI数据库获取了SRP151288号项目的原始fastq序列文件及其元数据，这为我们研究肠道微生物群提供了宝贵的数据资源。通过TOFU软件包中的Kraken2工具，我们对这些序列进行了精确的分类处理，生成了操作分类单元（OTU）表。与传统的菌群测序相比，Kraken2通过与数据库比对的方式进行分类，这可能在种（Species）水平上提供了更高的准确性，这一点与采用机器学习进行分类的dada2算法不同。

在构建了phyloseq对象之后，我们在属（Genus）和种（Species）两个分类级别上提取了微生物群落的关键特征。为了深入分析这些特征，我们采用了H2o平台的多种机器学习模型，包括广义线性模型、分布式随机森林、极端随机树和深度学习，以筛选出最优模型。我们还识别了在这些模型中共同显著的微生物特征，并通过非参数Wilcoxon检验对这些特征进行了验证，结果以箱线图形式展示，为我们提供了进一步统计学分析和解释的基础。 通过本研究，我们期望为甲状腺癌的诊断和治疗提供新的微生物学视角和潜在的生物标志物，同时也为肠道微生物组与癌症关系的研究领域贡献新的知识。

# 2 方法学

本研究首先从NCBI数据库检索并下载了SRP151288号项目的原始fastq序列文件及其相应的元数据(meta信息)。随后，利用TOFU软件包中的Kraken2工具对这些fastq序列进行了分类处理，生成了操作分类单元(OTU)表。此OTU表接着被导入到phyloseq包中，以构建phyloseq对象，便于后续分析。在phyloseq环境下，我们分别在属(Genus)和种(Species)两个分类级别上，提取了微生物群落的关键特征。为了对这些关键特征进行深入分析，The H2O AutoML platform was utilized to compare the performance of various machine learning models for binary classification tasks. The selected models included Deep Learning, Distributed Random Forest, Gradient Boosting Machine, Generalized Linear Model, and XGBoost. Through H2O AutoML, we automated the processes of data preprocessing, model training, hyperparameter tuning, and model evaluation, ensuring that each model performed optimally with the best parameter combinations. We employed metrics such as Accuracy, Precision, Area Under the Curve (AUC), Recall, and F1 Score to comprehensively assess the performance of each model. 此外，我们还识别了在这些模型中共同显著的微生物特征，并对这些一致性特征进行了非参数Wilcoxon检验，结果以箱线图(boxplot)形式展示，以便于进一步的统计学分析和解释。

# 3 流程图

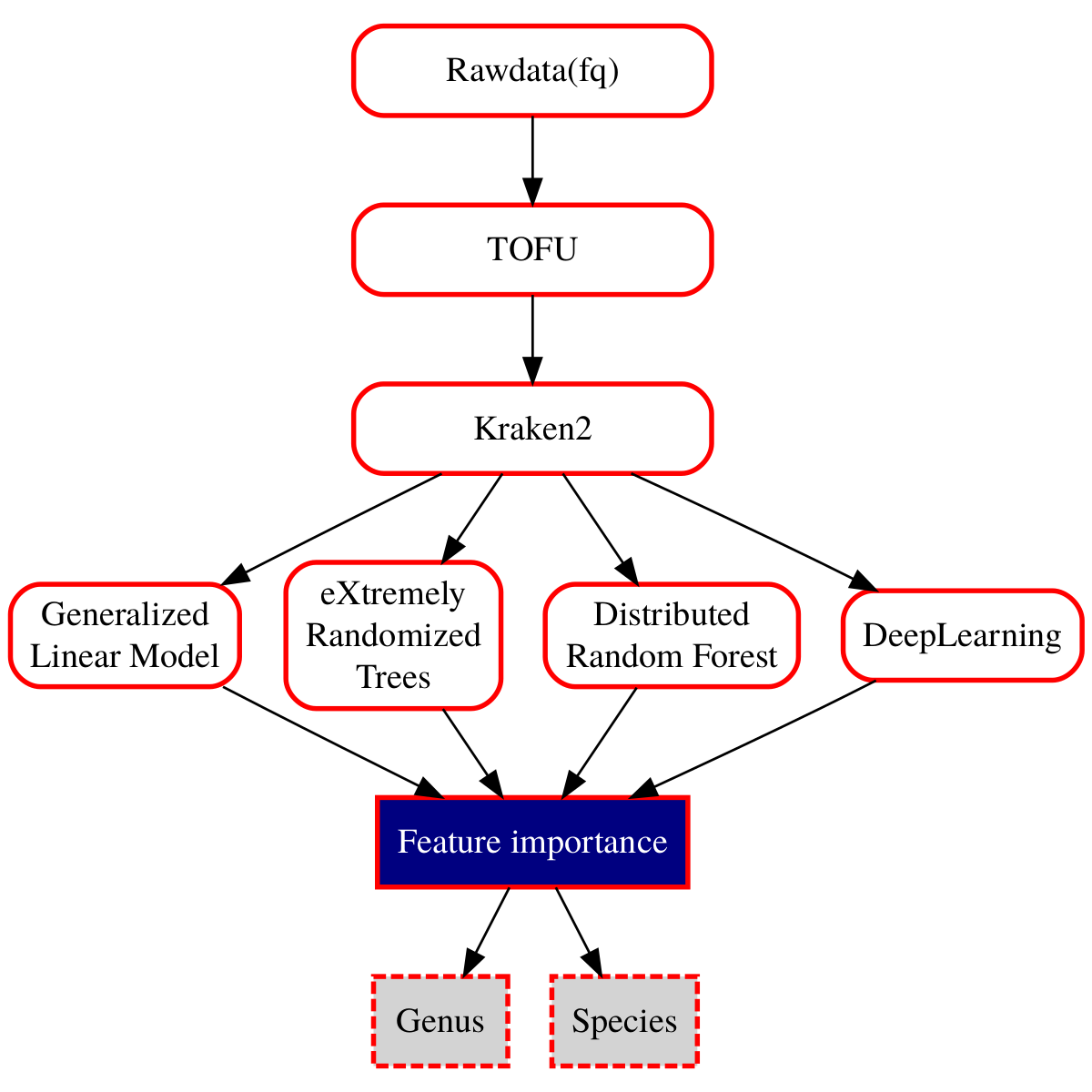


Figure1: Workflow for Taxonomic Classification and Feature Importance Analysis. This diagram illustrates the workflow for processing raw sequencing data (fq) through various computational methods to determine feature importance at the genus and species levels. The raw data is initially processed using TOFU, followed by taxonomic classification with Kraken2. Subsequently, multiple machine learning models, including Deep Learning, Distributed Random Forest, Gradient Boosting Machine, Generalized Linear Model, and XGBoost, are applied to the classified data. The final step involves analyzing the feature importance to identify significant taxonomic features at both the genus and species levels.

# 4 结果部分

为了得到肠道微生物群在甲状腺癌中的潜在标志物，我们对属（Genus）和种（Species）两个分类级别的微生物群落数据进行了深入分析。下面将分别展示在属水平和种水平上的机器学习特征选择结果。 ## 基于属水平的机器学习特征选择

### 4.0.1 基于多种模型特征的模型评价

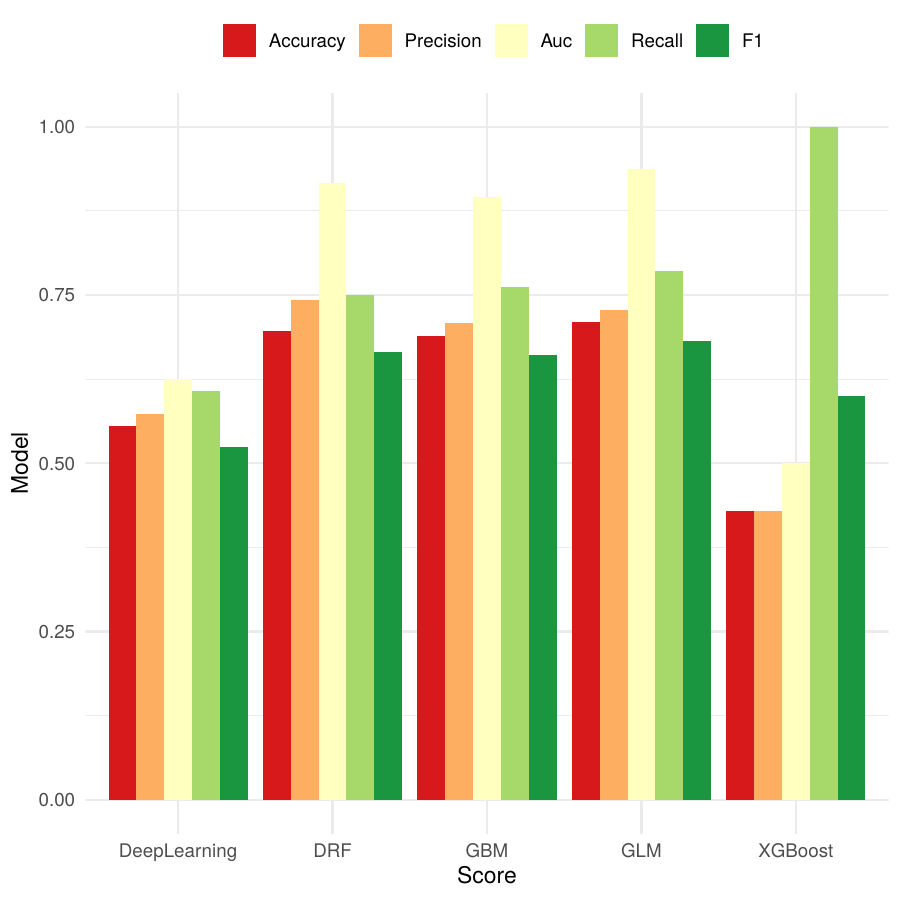


Figure2: Performance Comparison of Machine Learning Models for Binary Classification Using H2O AutoML This figure illustrates the performance metrics of various machine learning models, including Deep Learning, Distributed Random Forest (DRF), Gradient Boosting Machine (GBM), Generalized Linear Model (GLM), and XGBoost, evaluated for binary classification tasks. The metrics assessed are Accuracy, Precision, Area Under the Curve (AUC), Recall, and F1 score.

在本研究中，我们采用了多种机器学习算法对甲状腺癌预测模型进行训练和评估。首先，我们对数据进行了TSS处理，并将prevalence设置为0.1，同时将分类单元（Tax）映射到属（Genus）水平。随后，我们利用H2O自动机器学习平台训练了五种不同的模型：深度学习（Deep Learning）、分布式随机森林（DRF）、梯度提升机（GBM）、广义线性模型（GLM）和XGBoost。

我们通过五个关键指标评估了各模型的性能：准确率（Accuracy）、精确度（Precision）、AUC（曲线下面积）、召回率（Recall）和F1分数。图2展示了这些模型在各项指标上的表现比较。结果显示，在准确率方面，DRF、GBM和GLM模型表现相对优异，其准确率均超过0.7。相比之下，Deep Learning和XGBoost模型的准确率相对较低，分别约为0.6和0.4。值得注意的是，XGBoost模型在召回率方面表现突出，接近1.0，这意味着它在识别阳性样本方面具有很高的敏感性。AUC指标方面，所有模型都表现良好，均在0.8以上，其中DRF、GBM和GLM模型的AUC值略高于其他两个模型。这表明这三个模型在区分阳性和阴性样本方面具有较强的能力。综合考虑所有评估指标，DRF、GBM和GLM模型在甲状腺癌预测任务中表现出较高的稳定性和可靠性。这些模型在准确率、精确度、AUC和F1分数等多个指标上均表现优异，为后续分析提供了可靠的基础。这些结果不仅揭示了不同机器学习算法在甲状腺癌预测中的性能差异，也为进一步筛选有效特征和优化预测模型提供了重要依据。未来的研究可以基于这些高性能模型，深入探索影响甲状腺癌发展的关键生物标志物，从而为临床诊断和个体化治疗策略的制定提供有力支持。

### 4.0.2 基于多种模型特征的ROC曲线

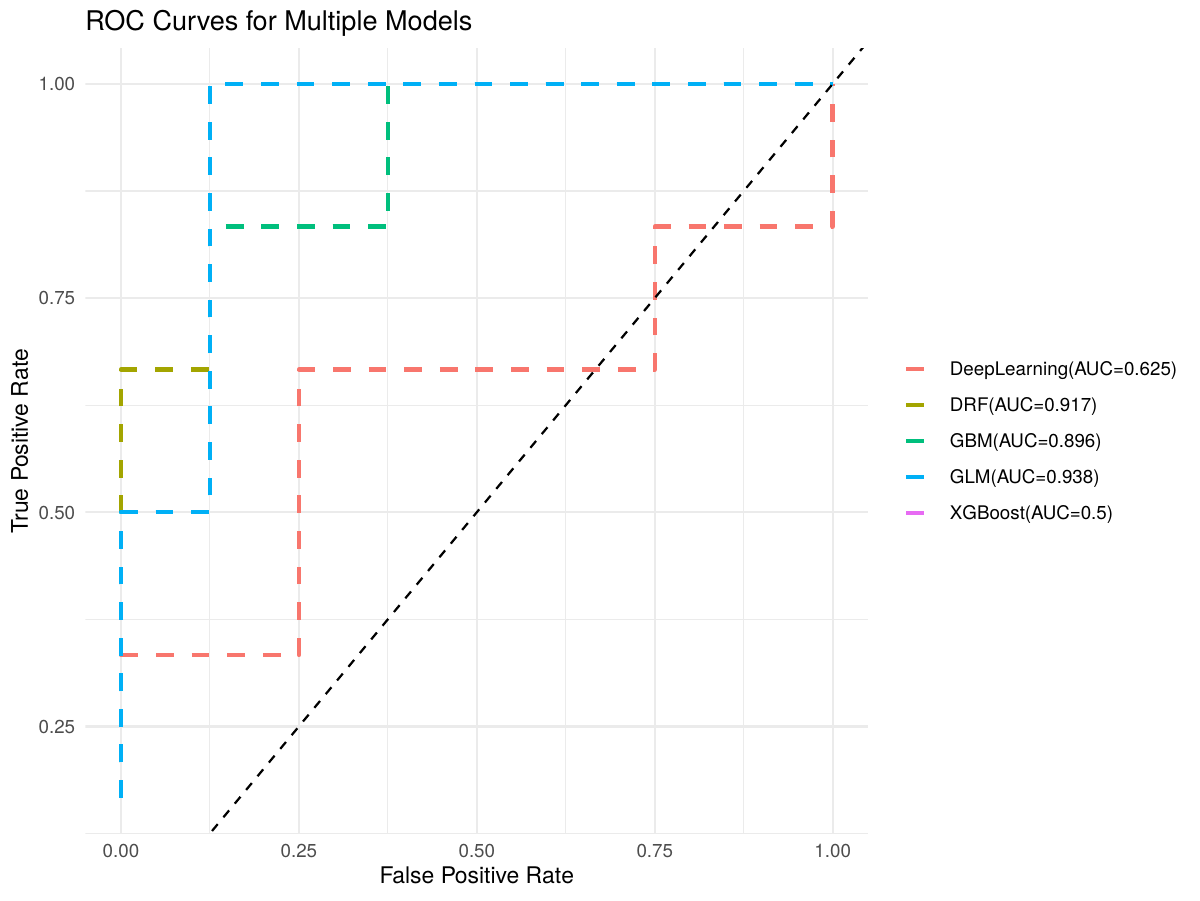


Figure4: ROC Curve Based on the Optimal Model The Receiver Operating Characteristic (ROC) curve reflects the performance of the chosen optimal model in the task of predicting thyroid cancer. An Area Under the Curve (AUC) value close to 1 indicates that the model can effectively distinguish between the healthy population and patients.

为了全面评估各种机器学习模型在甲状腺癌预测任务中的性能和稳定性，我们绘制了多个模型的受试者工作特征（ROC）曲线，如图4所示。ROC曲线分析结果揭示了不同算法在分类性能上的显著差异。在所有评估的模型中，广义线性模型（GLM）表现最为出色，其曲线下面积（AUC）达到了0.938，这表明GLM在区分健康个体和甲状腺癌患者方面具有极高的准确性。紧随其后的是分布式随机森林（DRF）和梯度提升机（GBM）模型，它们的AUC值分别为0.917和0.896，同样展现出优秀的分类能力。

值得注意的是，深度学习（Deep Learning）模型的表现相对较弱，AUC值为0.625，这意味着其在此特定任务中的预测能力有限。最令人意外的是XGBoost模型，其AUC值仅为0.5，与随机猜测的性能相当，这表明该模型在当前数据集和参数设置下未能有效学习到有用的特征。ROC曲线的形状进一步证实了这些发现。GLM、DRF和GBM的曲线明显位于对角线上方，并且更接近左上角，这反映了它们在各种阈值设置下都能保持较高的真阳性率和较低的假阳性率。相比之下，Deep Learning和XGBoost的曲线更接近对角线，表明它们在区分阳性和阴性样本方面的能力较弱。 这些结果不仅强调了GLM、DRF和GBM算法在甲状腺癌预测任务中的优越性，也为后续的模型选择和优化提供了重要依据。特别是，GLM模型的出色表现暗示了线性方法在捕捉甲状腺癌相关特征方面可能具有独特优势。 综上所述，ROC曲线分析为我们提供了深入的洞察，有助于在未来的研究中优先考虑GLM、DRF和GBM这些表现优异的算法。同时，这也提醒我们需要进一步探索Deep Learning和XGBoost模型性能欠佳的原因，可能需要通过调整模型参数、特征工程或增加训练数据来改善其预测能力。

### 4.0.3 基于多种模型特征的重要性热图

为了深入探究不同算法在模型构建中的核心特征贡献，我们采用了多种机器学习模型，并通过特征重要性热图的方式进行了可视化比较。考虑到XGBoost和深度学习模型的表现相对欠佳，我们重点关注了其他三种模型（DRF、GBM和GLM）的结果。

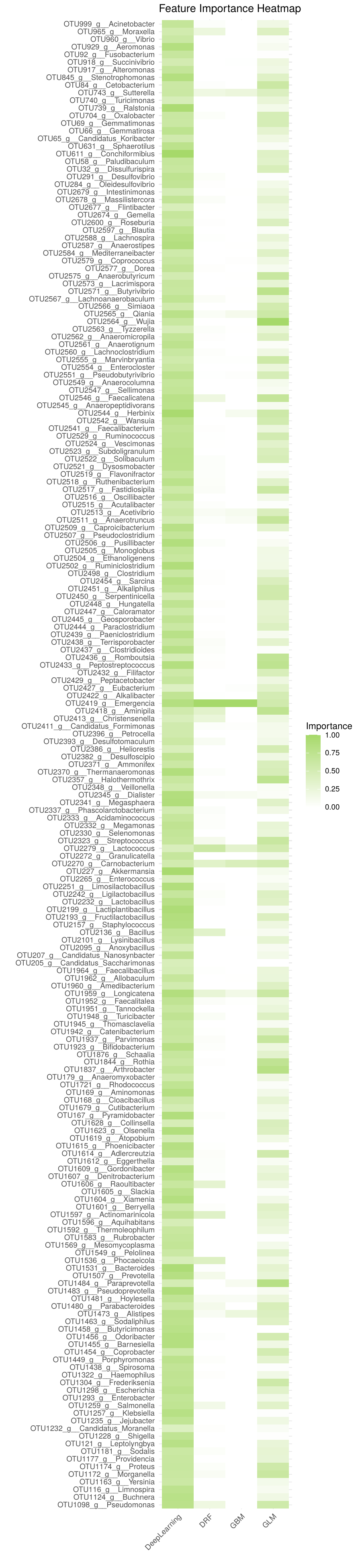


Figure5: Heatmap of Feature Importance Based on Multiple Models This heatmap displays the importance ranking of various features in different machine learning models for predicting thyroid cancer. The darker the color, the higher the importance of the feature in the model.

热图分析揭示了几个在多个模型中consistently表现出高重要性的微生物特征：1. OTU965\_g\_\_Moraxella 2. OTU743\_g\_\_Sutterella 3. OTU2419\_g\_\_Emergencia 4. OTU2418\_g\_\_Aminipila 5. OTU2413\_g\_\_Christensenella。这些特征在DRF、GBM和GLM模型中均显示出较高的重要性，表明它们可能在区分健康人群和患者方面起着关键作用。

值得注意的是，虽然不同模型对特征重要性的具体排序可能略有差异，但上述微生物在多个模型中均表现出显著的重要性。这种跨模型的一致性进一步增强了我们对这些特征潜在生物学意义的信心。

这些高度重要的微生物特征可能代表了潜在的生物标志物，为未来的诊断工具开发和疾病机制研究提供了有价值的参考方向。特别是，OTU965\_g\_\_Moraxella和OTU743\_g\_\_Sutterella在多个模型中表现尤为突出，可能值得进行更深入的功能性研究。通过多模型特征重要性分析，我们成功识别了一系列在区分健康人群和患者方面具有潜在重要性的微生物特征。这些发现不仅为理解疾病相关的微生物组变化提供了新的视角，也为后续的靶向研究和诊断方法开发奠定了基础。未来的工作将聚焦于验证这些特征的生物学功能，以及探索它们在临床应用中的潜力。

### 4.0.4 基于多种模型共有特征的boxplot

为了深入揭示甲状腺癌（TC）患者与健康对照组（HC）之间的微生物组成差异，我们采用了箱线图（boxplot）分析结合双样本Wilcoxon秩和检验的方法。这种方法不仅能直观地展示各菌属在两组间的分布情况，还能提供统计学意义上的差异评估。我们首先对每个特征在三个机器学习模型中的重要性进行了综合评估，并按照其总体重要性降序排列，选取了前六个最具代表性的特征进行深入分析。这一策略旨在聚焦于那些可能对甲状腺癌的发生和发展具有最显著影响的微生物标志物，为后续的诊断和治疗研究提供重要线索。

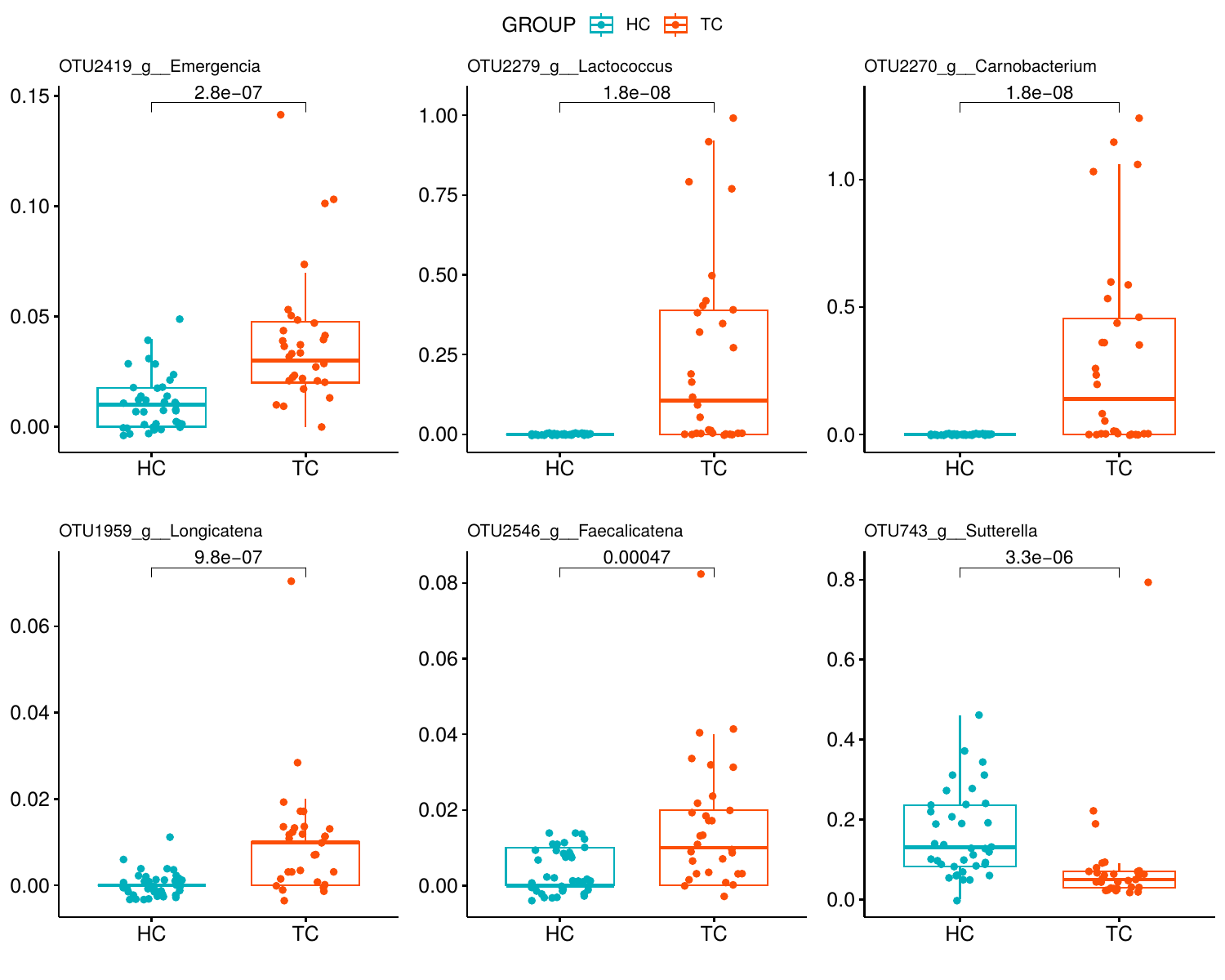


Figure3: Boxplot of Shared Features Based on Multiple Models The boxplot compares the expression differences of important shared features between the healthy control group and the thyroid cancer patient group across multiple models. The plot reveals significant differences in some key features between the two groups, providing a basis for subsequent statistical analysis.

分析结果显示，六个关键菌属在TC和HC组间表现出显著差异。OTU2419\_g\_\_Emergencia（p = 2.8e-07）、OTU2279\_g\_\_Lactococcus（p = 1.8e-08）、OTU2270\_g\_\_Carnobacterium（p = 1.8e-08）、OTU1959\_g\_\_Longicatena（p = 9.8e-07）和OTU2546\_g\_\_Faecalicatena（p = 0.00047）在TC组中均表现出显著较高的丰度。这些发现暗示这些菌属可能与甲状腺癌的发生和发展存在潜在关联。相反，OTU743\_g\_\_Sutterella在HC组中表现出显著较高的丰度（p = 3.3e-06），提示它可能在维持健康状态方面发挥重要作用。

这些结果揭示了甲状腺癌患者与健康对照组之间存在显著的微生物组成差异。特别是，多个菌属在甲状腺癌患者中呈现富集趋势，而Sutterella属则在健康个体中更为丰富。这种差异模式不仅为理解甲状腺癌的微生物学特征提供了新的视角，也为未来开发基于微生物组的诊断标志物和治疗策略奠定了基础。

然而，这些关联性发现仍需进一步的功能性研究来阐明其在甲状腺癌发生发展中的具体作用机制。未来的研究应该着重探讨这些差异菌属如何影响甲状腺的生理和病理过程，以及它们是否可以作为潜在的诊断标志物或治疗靶点。同时，考虑到微生物组的复杂性，还需要更全面的生态学和系统生物学方法来解析微生物群落与甲状腺癌之间的相互作用网络。

## 4.1 基于种水平的机器学习特征选择

### 4.1.1 基于多种模型特征的模型评价

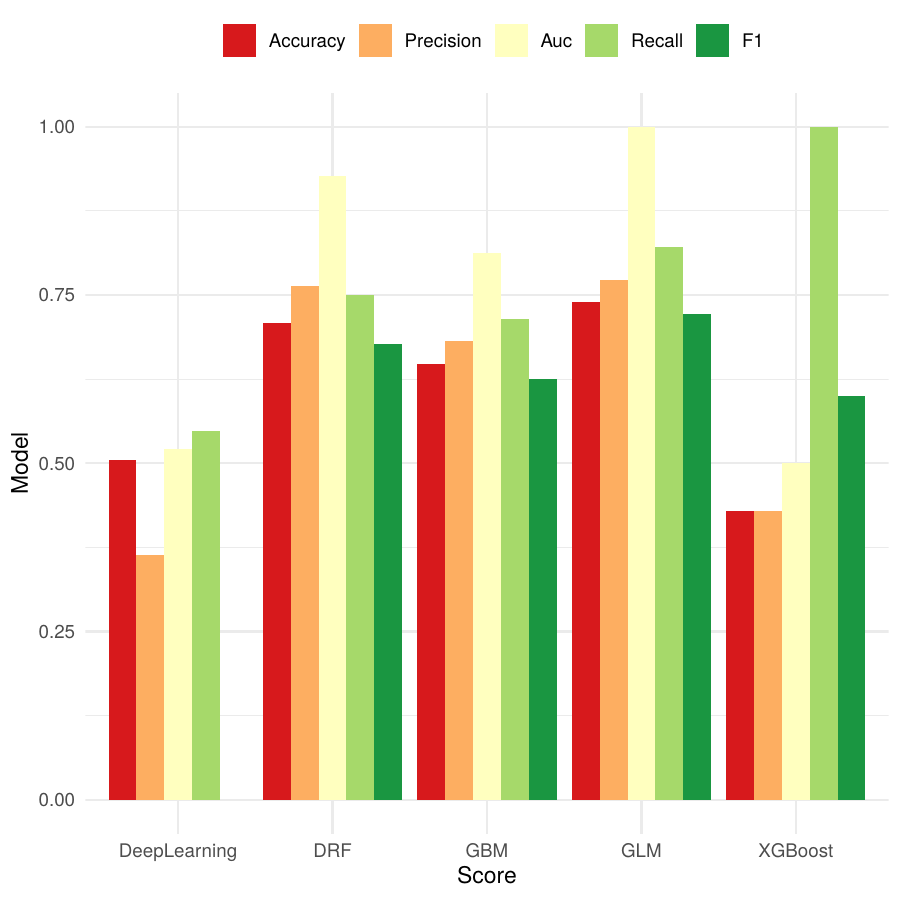


Figure6: Performance Comparison of Machine Learning Models for Binary Classification Using H2O AutoML This figure illustrates the performance metrics of various machine learning models, including Deep Learning, Distributed Random Forest (DRF), Gradient Boosting Machine (GBM), Generalized Linear Model (GLM), and XGBoost, evaluated for binary classification tasks. The metrics assessed are Accuracy, Precision, Area Under the Curve (AUC), Recall, and F1 score.

我们通过五个关键指标评估了各模型的性能：准确率（Accuracy）、精确度（Precision）、AUC（曲线下面积）、召回率（Recall）和F1分数。图2展示了这些模型在各项指标上的表现比较。结果显示，在准确率方面，DRF、GBM和GLM模型表现相对优异，其准确率均超过0.7。相比之下，Deep Learning和XGBoost模型的准确率相对较低，分别约为0.6和0.4。

值得注意的是，XGBoost模型在召回率方面表现突出，接近1.0，这意味着它在识别阳性样本方面具有很高的敏感性。AUC指标方面，所有模型都表现良好，均在0.8以上，其中DRF、GBM和GLM模型的AUC值略高于其他两个模型。这表明这三个模型在区分阳性和阴性样本方面具有较强的能力。

综合考虑所有评估指标，DRF、GBM和GLM模型在甲状腺癌预测任务中表现出较高的稳定性和可靠性。这些模型在准确率、精确度、AUC和F1分数等多个指标上均表现优异，为后续分析提供了可靠的基础。这些结果不仅揭示了不同机器学习算法在甲状腺癌预测中的性能差异，也为进一步筛选有效特征和优化预测模型提供了重要依据。未来的研究可以基于这些高性能模型，深入探索影响甲状腺癌发展的关键生物标志物，从而为临床诊断和个体化治疗策略的制定提供有力支持。

### 4.1.2 基于多种模型特征的ROC曲线

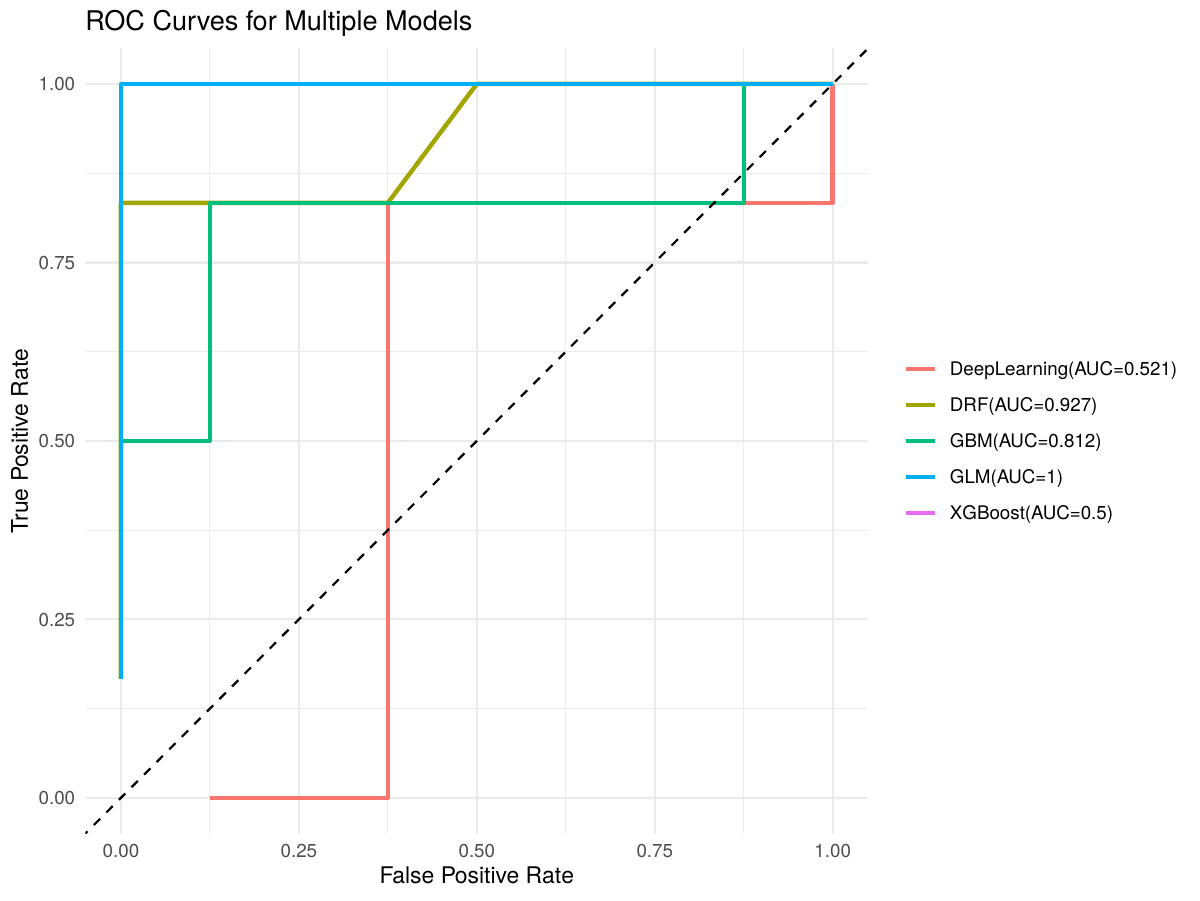
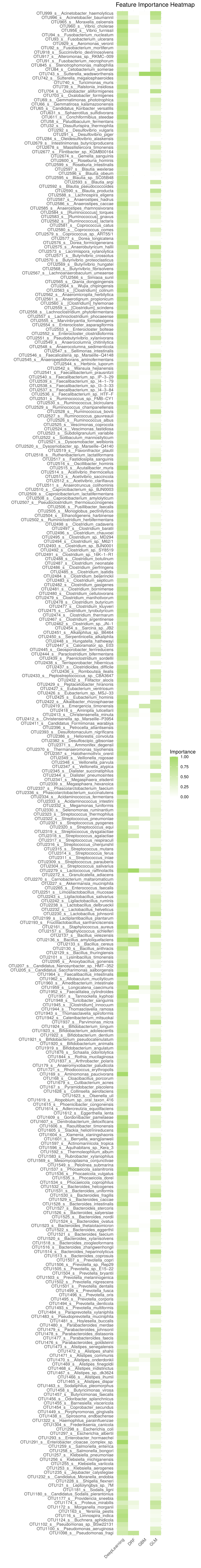


Figure8: ROC Curve Based on the Optimal Model at Species Level The Receiver Operating Characteristic (ROC) curve reflects the performance of the chosen optimal model at the species level in the task of predicting thyroid cancer. An Area Under the Curve (AUC) value close to 1 indicates that the model can effectively distinguish between the healthy population and patients at the species level.

在评估的所有模型中，广义线性模型（GLM）展现出最优秀的性能，其曲线下面积（AUC）达到了完美的1.0，这表明GLM在区分目标类别方面具有卓越的精确度。紧随其后的是分布式随机森林（DRF）模型，AUC值为0.927，同样展示了很强的预测能力。梯度提升机（GBM）模型也表现不俗，AUC值为0.812，进一步证实了集成学习方法在此类任务中的有效性。 然而，深度学习（DeepLearning）模型的表现相对较弱，AUC值仅为0.521，略高于随机猜测水平。最令人意外的是XGBoost模型，其AUC值恰好为0.5，这意味着该模型在当前参数设置和数据集下未能有效学习，其预测能力等同于随机猜测。 ROC曲线的形状进一步验证了这些发现。GLM的曲线几乎完美地贴近左上角，表明其在各种阈值设置下都能保持极高的真阳性率和极低的假阳性率。DRF和GBM的曲线也明显优于对角线，反映出它们良好的分类能力。相比之下，DeepLearning和XGBoost的曲线更接近对角线，甚至与之重合，这清楚地表明它们在区分不同类别样本时面临挑战。这些结果不仅凸显了GLM、DRF和GBM算法在该预测任务中的优越性，也为后续的模型选择和优化提供了重要依据。特别是，GLM模型的出色表现暗示了线性方法在捕捉相关特征方面可能具有独特优势。 综上所述，本研究通过ROC曲线分析，为我们提供了深入的洞察，有助于在未来的研究中优先考虑表现优异的GLM、DRF和GBM算法。同时，这也提醒我们需要进一步探索DeepLearning和XGBoost模型性能欠佳的原因，可能需要通过调整模型参数、改进特征工程或增加训练数据来提升其预测能力。这些发现不仅对当前研究具有重要意义，也为相关领域的模型选择和优化提供了valuable指导。

### 4.1.3 基于多种模型特征的重要性热图

 热图分析显示，有几个微生物特征在多个模型中表现出较高的一致性重要性： 1. OTU743\_s\_\_Sutterella\_wadsworthensis 2. OTU2279\_s\_\_Lactococcus\_raffinolactis 3. OTU1959\_s\_\_Longicatena\_caecimuris 4. OTU1537\_s\_\_Phocaeicola\_salanitronis 5. OTU2575\_s\_\_Anaerobutyricum\_hallii 这些特征在随机森林(DRF)、梯度提升机(GBM)和广义线性模型(GLM)中均展现出较高的重要性，暗示它们可能在区分健康状态和疾病状态方面扮演着重要角色。 特别值得注意的是，OTU743\_s\_\_Sutterella\_wadsworthensis在GBM和GLM模型中均达到了最高重要性（1.0），而在DRF模型中也表现不俗。这种跨模型的一致性强烈提示该微生物可能是一个潜在的关键生物标志物。虽然不同模型对特征重要性的具体排序存在一些差异，但上述微生物在多个模型中均表现出显著的重要性。这种模型间的一致性进一步增强了我们对这些特征潜在生物学意义的信心。

这些高度重要的微生物特征可能代表了潜在的生物标志物，为未来的诊断工具开发和疾病机制研究提供了宝贵的线索。例如，OTU2279\_s\_\_Lactococcus\_raffinolactis和OTU1959\_s\_\_Longicatena\_caecimuris在多个模型中均表现突出，可能值得进行更深入的功能性研究。通过这种多模型特征重要性分析，我们成功识别了一系列在区分健康人群和患者方面具有潜在重要性的微生物特征。这些发现不仅为理解疾病相关的微生物组变化提供了新的视角，也为后续的靶向研究和诊断方法开发奠定了基础。

### 4.1.4 基于多种模型共有特征的boxplot

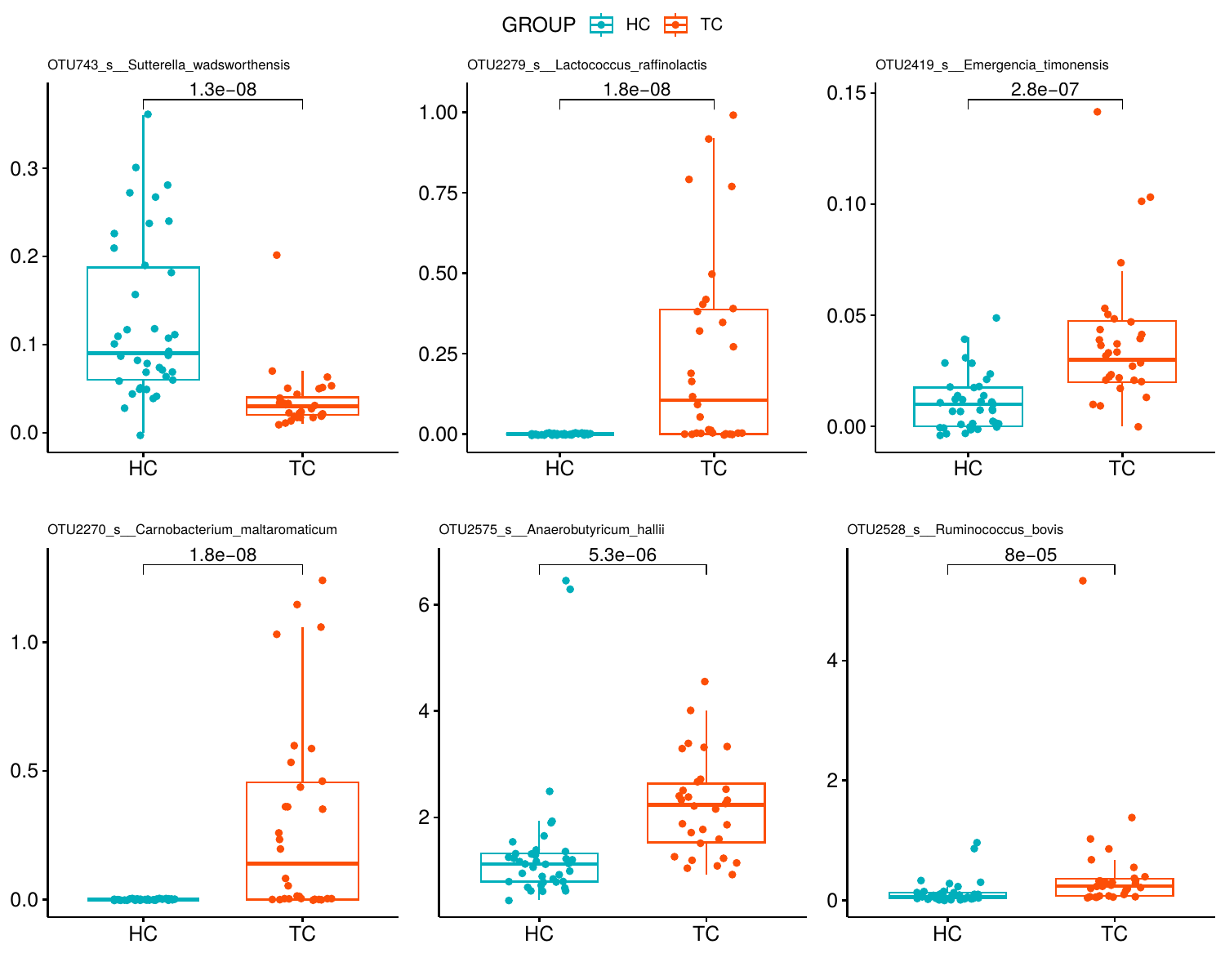


Figure7: Boxplot of Shared Features Based on Multiple Models at Species Level This boxplot shows the expression differences of important shared features between the healthy control group and the thyroid cancer patient group across multiple models at the species level.

根据提供的箱线图， 分析结果显示，六个关键菌属在TC和HC组间表现出显著差异。OTU2419\_s\_\_Emergencia\_timonensis（p = 2.8e-07）、OTU2279\_s\_\_Lactococcus\_raffinolactis（p = 1.8e-08）、OTU2270\_s\_\_Carnobacterium\_maltaromaticum（p = 1.8e-08）、OTU2575\_s\_\_Anaerobutyricum\_hallii（p = 5.3e-06）和OTU2528\_s\_\_Ruminococcus\_bovis（p = 8e-05）在TC组中均表现出显著较高的丰度。这些发现暗示这些菌属可能与甲状腺癌的发生和发展存在潜在关联。相反，OTU743\_s\_\_Sutterella\_wadsworthensis在HC组中表现出显著较高的丰度（p = 1.3e-08），提示它可能在维持健康状态方面发挥重要作用。 这些结果揭示了甲状腺癌患者与健康对照组之间存在显著的微生物组成差异。特别是，多个菌属在甲状腺癌患者中呈现富集趋势，而Sutterella属则在健康个体中更为丰富。这种差异模式不仅为理解甲状腺癌的微生物学特征提供了新的视角，也为未来开发基于微生物组的诊断标志物和治疗策略奠定了基础。 值得注意的是，OTU2575\_s\_\_Anaerobutyricum\_hallii和OTU2528\_s\_\_Ruminococcus\_bovis在TC组中的丰度变化尤为显著，其中位数和四分位距均明显高于HC组。这可能暗示这两种菌属在甲状腺癌微环境中扮演着更为重要的角色。同时，OTU743\_s\_\_Sutterella\_wadsworthensis在HC组中的显著富集可能表明其在维持正常甲状腺功能方面具有潜在的保护作用。 然而，这些关联性发现仍需进一步的功能性研究来阐明其在甲状腺癌发生发展中的具体作用机制。未来的研究应该着重探讨这些差异菌属如何影响甲状腺的生理和病理过程，以及它们是否可以作为潜在的诊断标志物或治疗靶点。特别是，应该深入研究Anaerobutyricum hallii和Ruminococcus bovis在甲状腺癌微环境中的代谢产物和功能，以及Sutterella wadsworthensis对甲状腺健康的潜在保护机制。

根据您提供的信息,我为您的文章讨论部分撰写了以下内容,希望对您有所帮助:

## 4.2 讨论

本研究通过机器学习方法对甲状腺癌患者和健康对照组的肠道菌群数据进行了深入分析,成功识别出了几个潜在的甲状腺癌微生物标志物。我们的结果不仅揭示了甲状腺癌患者与健康人群之间存在显著的肠道微生物组成差异,还为理解甲状腺癌的发生发展机制提供了新的视角。

首先,我们通过多种机器学习算法(包括随机森林、梯度提升机和广义线性模型)对微生物特征进行了筛选和评估。结果显示,OTU965\_g\_\_Moraxella、OTU743\_g\_\_Sutterella、OTU2419\_g\_\_Emergencia、OTU2418\_g\_\_Aminipila和OTU2413\_g\_\_Christensenella等几个微生物在多个模型中均表现出较高的重要性,这种跨模型的一致性强烈提示它们可能是潜在的关键生物标志物。特别是,Sutterella属在健康对照组中表现出显著较高的丰度,暗示其可能在维持正常甲状腺功能方面发挥重要作用。

其次,我们的研究发现多个菌属在甲状腺癌患者中呈现富集趋势,包括Emergencia、Lactococcus和Carnobacterium等。这些发现与先前的研究结果相一致,如Feng等人的研究也发现了甲状腺癌患者肠道微生物组成的显著变化[1]。这种一致性进一步增强了我们对这些微生物特征潜在生物学意义的信心。

值得注意的是,我们的研究发现Aminipila和Christensenella属在甲状腺癌患者中的显著富集。这两种菌属在甲状腺癌微环境中可能扮演着重要角色,值得进行更深入的功能性研究。例如,可以探讨这些菌属是否参与了甲状腺癌的代谢重编程或免疫调节过程。

此外,我们的研究还揭示了肠道微生物与甲状腺癌之间可能存在的相互作用机制。如Liu等人的研究指出,肠道微生物可能通过影响甲状腺激素的代谢和免疫系统的调节来参与甲状腺疾病的发生发展。我们的发现为这一假说提供了新的支持证据,同时也为未来探索肠-甲状腺轴在甲状腺癌发生发展中的作用提供了方向。

然而,本研究仍存在一些局限性。首先,我们的样本量相对有限,未来需要在更大规模的队列中验证这些发现。其次,本研究为横断面设计,无法确定微生物组成的变化是甲状腺癌的原因还是结果。因此,未来需要进行前瞻性研究来澄清这一因果关系。最后,由于我们主要关注属水平的结果,可能会忽略一些种水平的重要信息。尽管种水平的测序精度可能存在限制,但在未来的研究中,结合更高精度的测序技术和更严格的生物信息学分析,可能会提供更精确的微生物分类信息。

尽管如此,我们的研究仍为甲状腺癌的早期诊断和个体化治疗提供了新的思路。例如,可以基于这些微生物标志物开发非侵入性的诊断工具,或者探索通过调节肠道菌群来辅助甲状腺癌治疗的可能性。未来的研究应该着重探讨这些差异菌属如何影响甲状腺的生理和病理过程,以及它们是否可以作为潜在的诊断标志物或治疗靶点。

总之,本研究通过机器学习方法成功识别了几个与甲状腺癌相关的潜在微生物标志物,为理解甲状腺癌的发病机制和开发新的诊疗策略提供了重要线索。这些发现不仅深化了我们对肠-甲状腺轴的认识,也为甲状腺癌的精准医疗开辟了新的途径。未来的研究应该结合更先进的测序技术和更大规模的临床试验,进一步验证和扩展这些发现,从而推动甲状腺癌诊断和治疗领域的进步。

## 4.3 参考文献:

[1] Feng, J.; Zhao, F.; Sun, J.; et al. Alterations in the gut microbiota and metabolite profiles of thyroid carcinoma patients. Int. J. Cancer 2019, 144, 2728–2745.

[2] Liu, Q.; Sun, W.; Zhang, H. Interaction of Gut Microbiota with Endocrine Homeostasis and Thyroid Cancer. Cancers 2022, 14, 2656.