

ML_microbiota

Nitin Agrawal

2025-07-23

```
library(phyloseq)
library(randomForest)

## Warning: package 'randomForest' was built under R version 4.5.1
## randomForest 4.7-1.2
## Type rfNews() to see new features/changes/bug fixes.

library(pROC)

## Warning: package 'pROC' was built under R version 4.5.1
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var

library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:randomForest':
##
##     margin

#Loading data

data("GlobalPatterns")
gp <- GlobalPatterns
sample_data(gp)$Group <- ifelse(sample_data(gp)$SampleType == "Feces",
"Disease", "Healthy")

#Preprocessing

otu <- as.data.frame(otu_table(gp))
otu <- t(otu)
labels <- sample_data(gp)$Group

# Random Forest
```

```

model <- randomForest(x = otu, y = as.factor(labels))
print(model)

##
## Call:
## randomForest(x = otu, y = as.factor(labels))
##               Type of random forest: classification
##               Number of trees: 500
## No. of variables tried at each split: 138
##
## OOB estimate of error rate: 3.85%
## Confusion matrix:
##      Disease Healthy class.error
## Disease      3      1      0.25
## Healthy      0     22      0.00

```

#ROC curve

```

pred <- predict(model, type = "prob")[, "Disease"]
roc_obj <- roc(labels, pred)

## Setting levels: control = Disease, case = Healthy
## Setting direction: controls > cases

plot(roc_obj)

```

