

Human Teaching, Machine Learning –

developing a sharable syllabus from the recommendations
generated through ML4Microbiome (EU-CA18131)

[Jamie.FitzGerald @ ucc.ie](mailto:Jamie.FitzGerald@ucc.ie)

APC 4.11 Claesson Group
8th of June, 2023



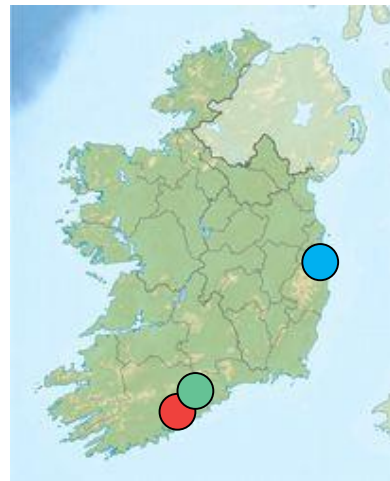
background

Jamie FitzGerald (see left of screen)

microbiome / microbial ecology
Research scientist, FHI

ML4Microbiome STSM:
Utrecht University, the Netherlands

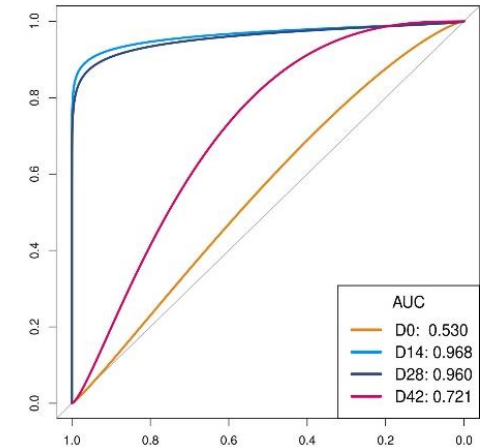
August 20th – August 27th, 2022



**Utrecht
University**

background

- Frequently focused on **discrimination** within a dataset
 - *groups, classes, treatments...*
- Diet intervention: ML identified persistent effect of probiotics in rare microbiome
 - **not detected with multivariate approaches**
- **Roosmarijn Luiken, UU** (PI: Aldert Zomer @UU, ML4Microbiome)
 - Microbiology, health, and disease
 - similar interest in exploring microbiome through ML



Utrecht
University

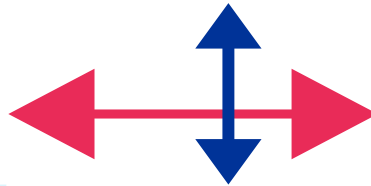


MICROBIOME

motivation

→ Both Roosmarijn & I interested in ML,
but lack background in mathematics &/ machine learning

Somewhat typical:



Microbiome Analysis:

- largely self-taught
- low level of standardisation
- relatively specialised approaches

Machine learning:

- **is a huge field in itself**
- **is very complex** (especially for the self-directed learner)
- often, no clear “best” method

motivation

Enter **ML4Microbiome**:

- Perspective/state of the art, appropriate methods, new and current applications, etc.
- In-person and online workshops
- Extensive online resources for training

→ **Suitable syllabus**

→ **Expert guidance**

→ **No need to re-invent any wheels**



MICROBIOME



cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

STSM proposal

1

condense the ML4Microbiome syllabus, for dissemination to microbiome researchers

2

create a set of learning exercises to allow training in ML-microbiome

3

focus on being:

- clear (for microbiome researchers)
- generalisable (for microbiome studies)

STSM in action

1

condense the ML4Microbiome syllabus, for dissemination to microbiome researchers

2

create a set of learning exercises to training in ML-microbiome

3

focus on being:

- clear (for microbiome researchers)
- generalisable (for microbiome studies)

+

+

+

Conversations @ UU:

quickly refined the background, the scope, and the format

issue: “one-size-fits-all” approach not appropriate for a workshop scenario (*even within scope*)

solution: run one method (RF, XGB, SVM); compare outputs to that of the other two methods:

- highlight where certain approaches succeed, and why (as possible)

1

condensed syllabus

- **major methods of classification/prediction:** Extreme Gradient Boosting, Support Vector Machines, and Random Forest
- **Connecting ML and the microbiome** – appropriate hypothesis testing and study design, and relating this to the underlying classification approaches
 - **data pre-processing** - making different microbiome data robust for classification approaches
 - **Understanding and communicating** microbiome-ML outcomes (post-processing!)

2 learning exercises

- Identifying / generating hypotheses
- feature engineering and normalisation
- feature selection example

- sample classification example

- performance evaluation
- communication (visualisation & interpretation)

candidate data:

- ABX treatment (UU dataset)
- environmental example (e.g. TARA Oceans)
- IBD data – human health outcome

3

clear & generalisable

Method of analysis: `print("In R we trust")`

Method of delivery (independent of language): **HTML; RMD v. IPYNB**

Duration: 3x (2 hours theory / 4 hours practical)

1. ML approaches for the microbiome
2. Preparing microbiome for ML
3. Applying ML to microbiome
4. Evaluating and communicating outcomes

Zoning allows rearrangement / alternate modes of delivery

outcomes



scope : range of requirements feasible within short teaching timeframe



framework : foundational set of connected topics from ML4Microbiome's outputs, to allow informed implementation of ML



practical exercises : applications of framework, using publicly available data



mode : allows for rapid sharing/dissemination, as well as future improvements /



personal: better understanding of machine learning, better appreciation for complexity and role in microbiome analysis



Takeaway: Check COST.eu, apply for your own STSM!

Get funded

→ to go somewhere you want

→ to talk to your peers

→ about something you're interested in

→ and then put it on your CV

Applying to EU COST Action:

- ✓ [COST.eu](https://www.cost.eu) – straightforward portal. Many, many Actions, on all topics
- ✓ 2 straightforward documents – asks you to understand what you are applying for, and to structure your application accordingly.
- ✓ COST staff / administration – very helpful.
- ✓ Excellent experience in starting applications for funding



Takeaway: Check COST.eu, apply for your own STSM!

Recommendations:

- Do serious work, apply for serious funding
- Reporting and submitting claims – very straightforward: use report template as a diary
- Getting the most out of your trip – plan ahead, clear your schedule!

With many *thank you's* to:



Claesson group



Vision 1 group

Roosmarijn Luiken & Aldert Zomer

ML4Microbiome, esp. workshop authors:

Matti Ruuskanen
Dimitrios Vlachakis
Aleksandra Gruca

& Marcus Claesson

EU-COST admin:

Aoife Tierney
Tatjana Turukalo
Anna Power



**Utrecht
University**



MICROBIOME



cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY