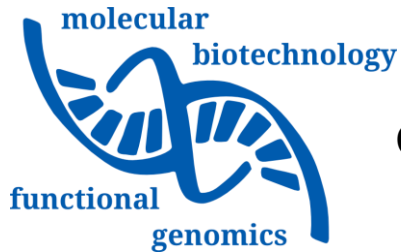# Follow up COST action

Alina Nechyporenko & Marcus Frohme,
COST Action CA18131 "Statistical and machine learning techniques in human microbiome studies",
Cork, 8 June 2023

Idea 1
Boilerplate framework for Machine Learning microbiome analysis

Boilerplate

Boilerplate framework can be used to provide standardization and consistency across projects, making it easier to maintain and update code

- start your next ML project for microbiome analysis in seconds
- highly scalable
- best DX
- focus on data analysis
- best practices

What is Boilerplate code

- part of code that can be reused over and over again without modification
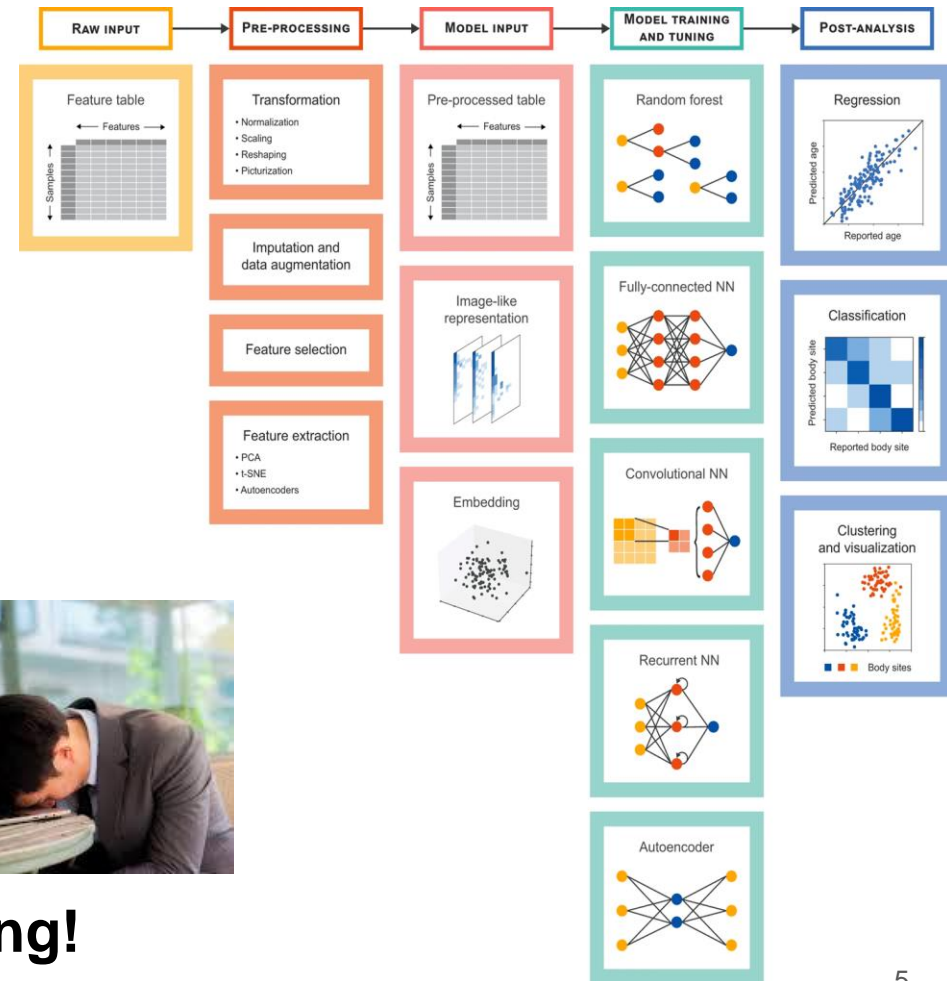- something that you copy and paste
- frame for your project

Why to use the boilerplate code

- reduce development time
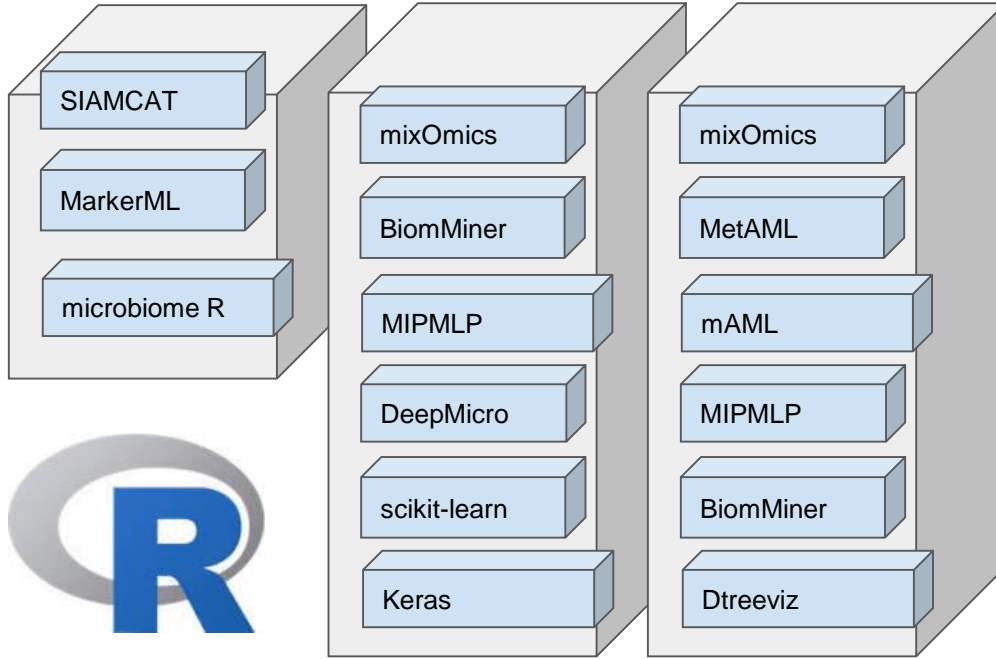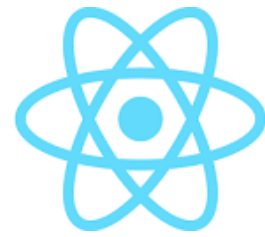- less complexity
- short learning curve

Microbiome data analysis.
Why it is challenging?

- Microbiome data is complex
- Plenty of tasks
- Plenty of tools, but no universal one
- Different Operating Systems
- Compatibility issues
  (dif. versions of TensorFlow..)
- Lack of best practices
- Reproducibility issues

Do research instead of engineering!
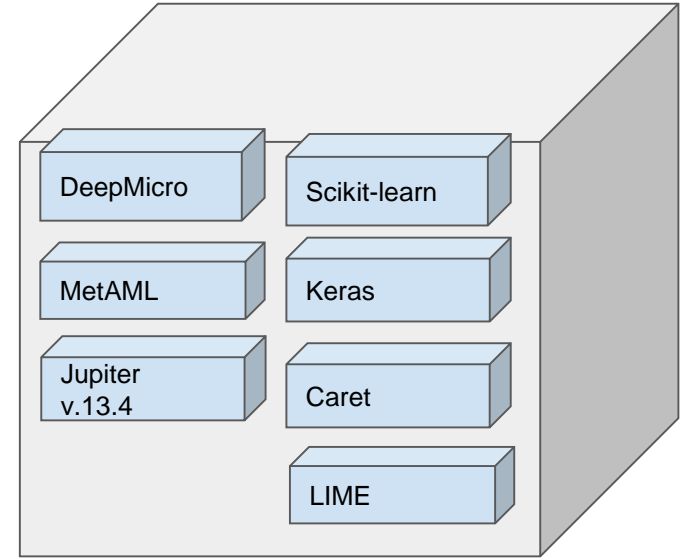
How boilerplate will help to resolve it

SIAMCAT

MarkerML

microbiome R

Normalization

…

mixOmics

BiomMiner

MIPMLP

DeepMicro

scikit-learn

Keras

Dimension reduction

mixOmics

MetAML

mAML

MIPMLP

BiomMiner

Dtreeviz

Feature selection

Preprocessing of microbiome data

DeepMicro

MetAML

Jupiter v.13.4

Scikit-learn

Keras

Caret

LIME

Prediction

Treatment

Biomarker Identification

…

Idea 2 ....

Status quo (… at least what we believe)

- Many tools in/for microbiome analyses
- Low level of standardization / few guidelines
- Mostly connected to medical problems
- Little clinical relevance so far
- High potential for personalized medicine

- Lack of reproducibility / Quality control problems
  (*inconsistent experimental protocols, lack of access to relevant metadata*…)

- Maintaining privacy / Patient safety

- Explainability
- Lack of trust from the physician's side

Your ideas from questionnaire **(question #4, 35 answers)**

- 1/3 … a bit useless (too little content)
- 1/3 adressing ML issues with more details
- 1/3 adressing applicational issues

ML issues with more details

- LLMs (large language models) or other neural networks for metagenomics
- inference and interpretation in ML
- multi-omics microbiome data
- e-learning platform how to analyze microbiome data.
- Computational and data science techniques
- deep learning models.
- AI
- invest into pretraining models
- bioinformatics, biostatistical analysis
- **explainable AI**

Applicational issues

- Integration of **other physiological data** .
- **Medical (microbiome) concepts/aspects**
- **tissue microbiome**
- **clinical practice**
- **inviting physicians**
- **medical questions /** microbiome interactions with health
- **One Health approach**
- **food and nutrition** focused on human **health and diseases**.
- **health care and agricultural setting**
- integration of the **entire ecosystem (host, environment, etc.)**

- **standardisation** issues.
- Stronger recommendations, clearer guidelines

- **Products / production**
- transparent disclosure from industry
- Dialogue with healthcare providers

Synthesis for a new COST action ?

stronger focus on medical questions
integration of clinicians / clinical microbiologists

definitively integration previous results
potentially boilerplate, explainable AI

integration of standardisation issues

Idea 2

Standardising microbiome analyses for medical practice

Current EU-projects and standardisation initiatives

- Running/recent EU projects in Cordis
  Keywords: personalised medicine, standardisation, microbiome:
  14 mostly on special medical issues

- Standardisations activities
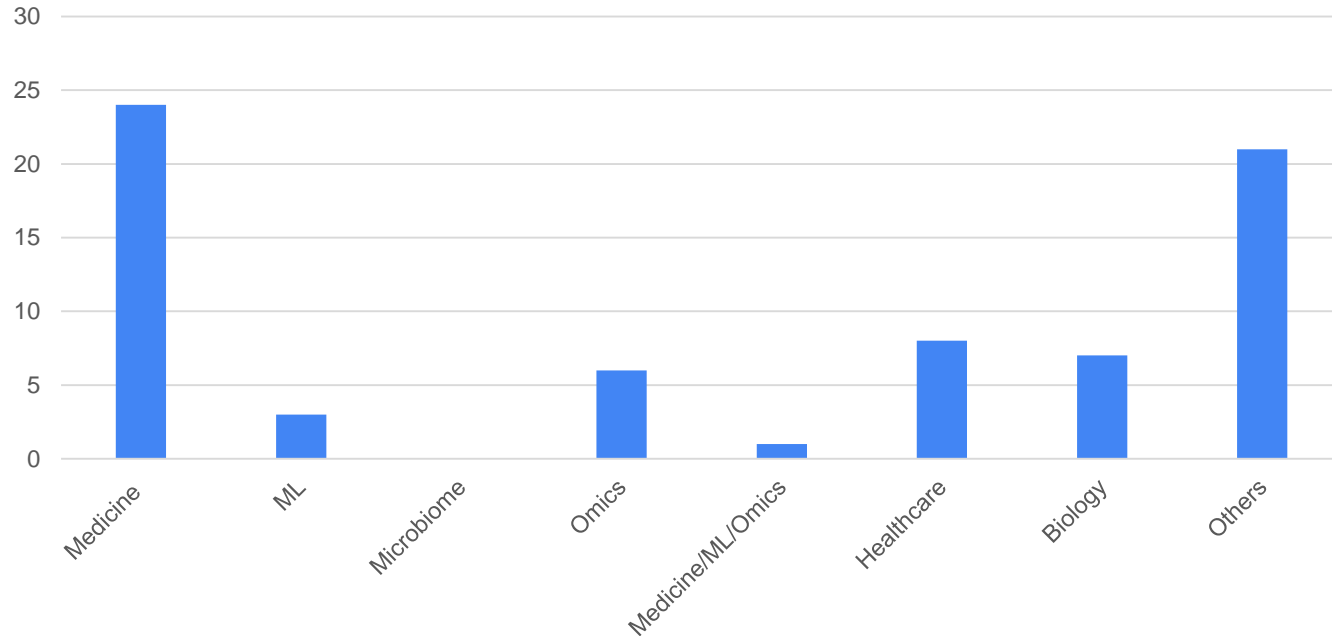ISO 9491 Predictive computational models in personalized medicine research
Part 1: Constructing, verifying and validating models
Part 2: Guidelines for implementing computational models
        in clinical integrated decision support systems

  ISO – TC 276 Biotechnology
  Working Group 5: Data processing & integration

Ongoing COST actions (2022-2026)

*CA21153 - Network for implementing multiomics approaches in atherosclerotic cardiovascular disease prevention and research (AtheroNET)

To be considered: new COST actions in 2024

Thank You for Your Attention !
Discussion now !