# http://ifisc.uib-csic.es

*@ifisc\_mallorca* 

#### www.facebook.com/ifisc



# Improving prediction of childhood obesity using microbiome data

#### Francesc Serra, Claudio R. Mirasso, Silvia Ortín

IFISC (CSIC-UIB) Palma de Mallorca – Spain. Affiliation 2

fserra@ifisc.uib-csic.es







#### Introduction

- Prevalence rates of obesity have increased dramatically.
- Obesity is a major factor risk for health.
- Addressing obesity in childhood would help alleviate the problem.
- The human body hosts a highly complex and densely populated ecosystem known as human microbiome.



Data

The microbiome data includes:

- Intestinal microbiome is connected to obesity.
- Our goal is to build a model able to predict whether a child whill develop obesity using microbiome data together with information about dietary habits, clinical and physical examinations and physical activity using machine learning and deep learning techniques.

-Relative abundances up to genus level.

-Measures of alpha diversity.

-Microbiota groups (C1-C4).



#### Methodology

We have tested several models. To evaluate the models, we followed a stratified 7-fold cross validation procedure.



#### **Results**

- Best model: CatBoost (gradient boosting algorithm) using a representation of the relative abundances obtained with a Variational Autoencoder (VAE). Variational Autoencoders are a kind of Autoencoders able to learn probabilistic representations given an input and then use these representations to reconstruct the input.
- To observe the effect of the microbiome data we also present the best results obtained without microbiome data.



For each iteration, using only the training set, we performed a 5fold cross-validation to optimize the model hyperparameters using a grid search.

## **Feature importance**

The SHAP values method calculated the most important features of the model.



	AUC	ACC
All data	0.75 ± 0.17	0.67 ± 0.17
Without microbiome data	0.68 ± 0.13	0.64 ± 0.09

This results point in the direction that including microbiome data increases the performance of the model.

### Conclusions

- Microbiome data can improve the prediction of childhood obesity.
- The variables related to the microbiota that have a greater impact on the model are: the Shannon alpha diversity measure, the microbiota group, and the Variables 1 and 14 from the VAE.
- Interpretability of the model is lost using VAE.
- Taking into account the high variations of the microbiome, our results seem promising.
- We will be able to validate or even improve our model with 1000 new samples.

### **References and acknowledgments**

[1] Simone Rampelli, Kathrin Guenther, Silvia Turroni, Maike Wolters, Toomas Veidebaum, Yiannis Kourides, Dénes Molnár, Lauren Lissner, Alfonso Benitez-Paez, Yolanda Sanz, Arno Fraterman, Nathalie Michels, Patrizia Brigidi, Marco Candela, and Wolfgang Ahrens. Pre- obese children's dysbiotic gut microbiome and unhealthy diets may predict the development of obesity. Communications Biology 2018 1:1, 1:1–11, 12 2018.

[2] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: Unbiased boosting with categorical features. volume 2018- December, 2018.

[3] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions.volume 2017-December, 2017.

This research has been carried out within the framework of the Climb-Out project.



