

Towards the optimisation and standardisation of Machine Learning techniques for human microbiome research: the ML4Microbiome COST Action (CA 18131)

Tatjana Loncar-Turukalo¹, Marcus J. Claesson², Randi J. Bertelsen³, Aldert Zomer⁴, Domenica D'Elia⁵✉

¹Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia

²APC Microbiome Ireland, University College Cork, Cork, Ireland

³University of Bergen, Bergen, Norway

⁴Utrecht University, Utrecht, Netherlands

⁵Institute for Biomedical Technologies, National Research Council, Bari, Italy

Competing interests: TLT none; MJC none; RJB none; AZ none; DD none

The analysis of data generated by metagenome projects in different human body sites has unveiled relevant differences in the microbiome in health and disease. For example, it has been demonstrated that the gut microbiome is crucial in many intestinal and non-intestinal diseases or pathophysiological conditions such as obesity, diabetes, development and functionality of the immune system, cardiovascular diseases, etc. (Lynch and Pedersen, 2016).

The ML4Microbiome COST Action¹ (CA18131) is a network of bioinformaticians, computer scientists and biologists, from 34 different European countries, working together to evaluate and optimise application of machine learning (ML) algorithms on microbiome data and develop standardised data processing pipelines for the analysis and interpretation of microbiome sequencing data. The need to establish this kind of network is related to the complexity of metagenomics data. Microbiota are dynamic ecosystems with active host regulation (Schnorr, 2018). Metagenome data and their influencing factors are complex to analyse and interpret. They are indeed inherently convoluted, noisy and highly variable. Due to the data's compositional nature, non-standard statistical methodologies and ML methods are required to unlock its clinical and scientific potential. While a range of statistical modelling and ML methods are now available, sub-optimal implementation often leads to errors, over-fitting and misleading results due to a lack of suitable analytical practices and ML expertise in the microbiome community. For these reasons, the field requires innovative approaches, specifically adapted to the properties of microbiome data and training to create ML expertise in the microbiome scientific community.

The ML4Microbiome working plan relays on the coordinated and integrated efforts of four working groups

(WGs). WG1 aims to evaluate and constantly update the state-of-the-art of ML technologies and methods applied in the field to define priority areas of intervention. WG2 is committed to establishing benchmark datasets for testing ML methods. In particular, based on available public and private data, WG2 selects the data types and creates public benchmark repositories to propose a DREAM Challenge to foster ML analysis of microbiome data. WG3 applies information and data made available by WG1 and 2, to optimise and standardise the use of ML existing methods to microbiome data, also investigating automation opportunities (*i.e.*, pipelines). Finally, WG4 is devoted to disseminating results of the Action, organising training courses and maintaining the website, newsletters and social media.

At this conference, we will present the ML4Microbiome aims and working plans focusing on upcoming training activities and COST tools the Action makes available to the scientific community, such as the possibility to apply for Short Term Scientific Missions². STSMs are exchange visits aimed at supporting individual mobility, strengthening existing networks and fostering collaboration between researchers. We also provide the possibility to apply for grants dedicated to Inclusiveness Target Countries³ (ITC). ITC grants support early career investigators (ECI) and PhD researchers from ITC to present their work at international conferences on the COST Action topic.

Availability

Information on how to join and collaborate is available at the ML4Microbiome website: <https://www.ml4microbiome.eu/>

²<https://www.ml4microbiome.eu/stsm/>

³<https://whocaresineurope.eu/inclusiveness-target-countries-itc>

¹<https://www.cost.eu/actions/CA18131>

Acknowledgements

ML4Microbiome is funded by COST (European Cooperation in Science and Technology) under the Grant Agreement CA18131. www.cost.eu.

The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

References

1. Lynch SV, Pedersen O (2016) The human intestinal microbiome in health and disease. *N. Engl. J. Med.* **375**:2369–2379. <http://dx.doi.org/10.1056/NEJMra1600266>
2. Schnorr SL (2018) Meanings, measurements, and musings on the significance of patterns in human microbiome variation. *Current Opinion in Genetics & Development* **53**:43-52. <http://dx.doi.org/10.1016/j.gde.2018.06.014>