WG4 Report
A summary of 2019/2023 achievements and work in progress

08 June 2023, Cork University College

COST Action CA18131 - MC&WGs Meeting, Cork, 7-9 June 2023
WG4 Leader: **Domenica D'Elia**
Vice WG-L: **Aldert Zomer**

WG4 Objectives

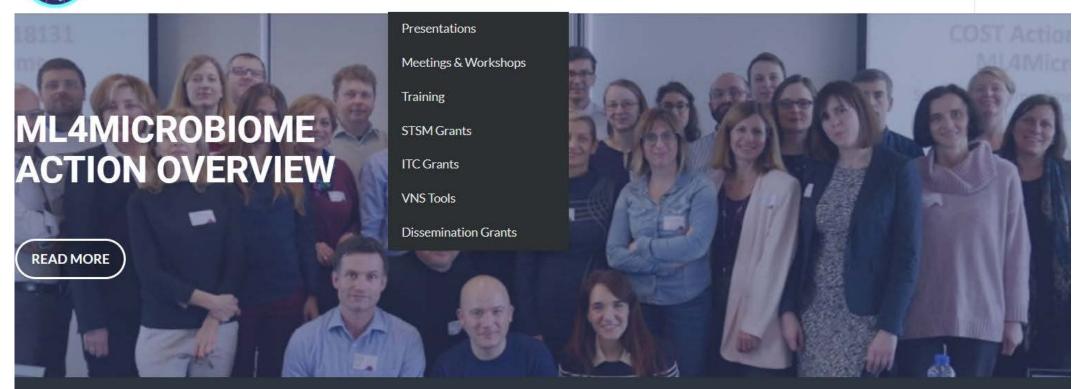**Disseminate the results of the COST Action in**
- peer-reviewed journals
- on the web-portal
- at international conferences
- on social networks (LinkedIn and Twitter)
- through end-user workshops & training schools in microbiome-related ML methodologies recommended by Action members

Tasks
T4.1: Disseminate evaluation, optimisation and standardisation progress
T4.2: Organise training courses (e.g. Summer Schools) and workshops

ML4Microbiome website

Activities:

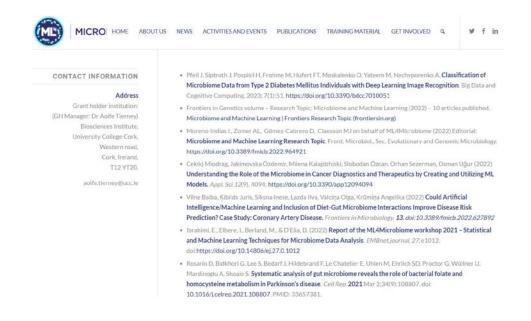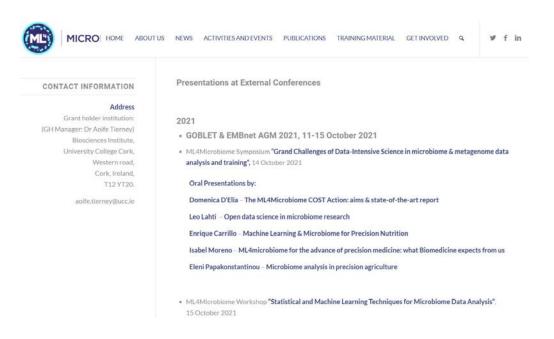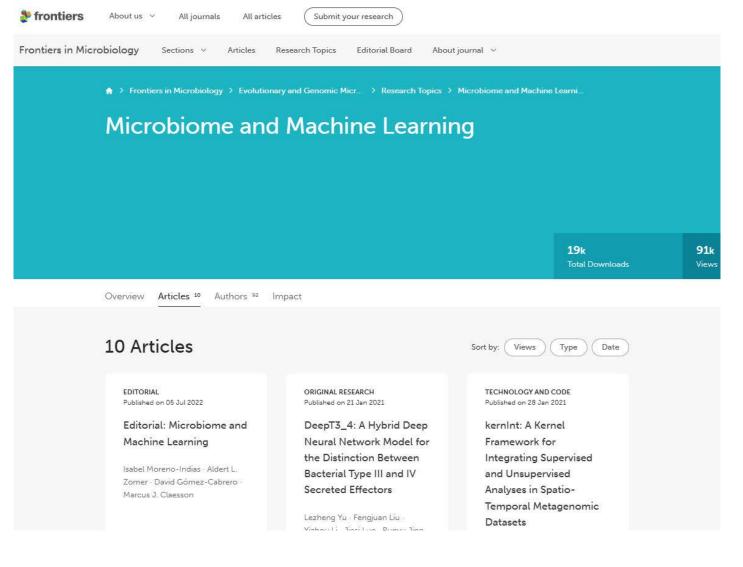**A4.1**: dissemination - **COST progress and outcomes shared within and outside the Action at the Web-portal**

✓ [Publications - ML4 Microbiome](#) (12 and many others in progress including one perspective article)

✓ [Presentations - ML4 Microbiome](#) (International Conferences – more than 15)

Microbiome and Machine Learning | Frontiers Research Topic (frontiersin.org)
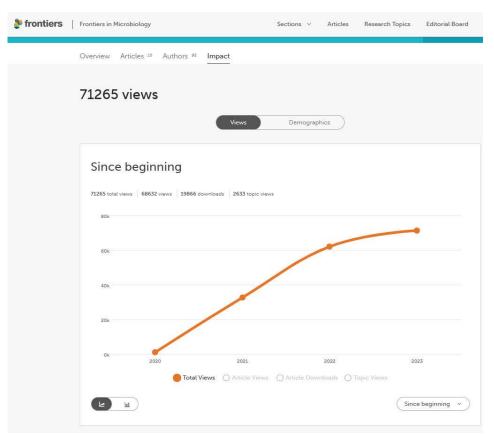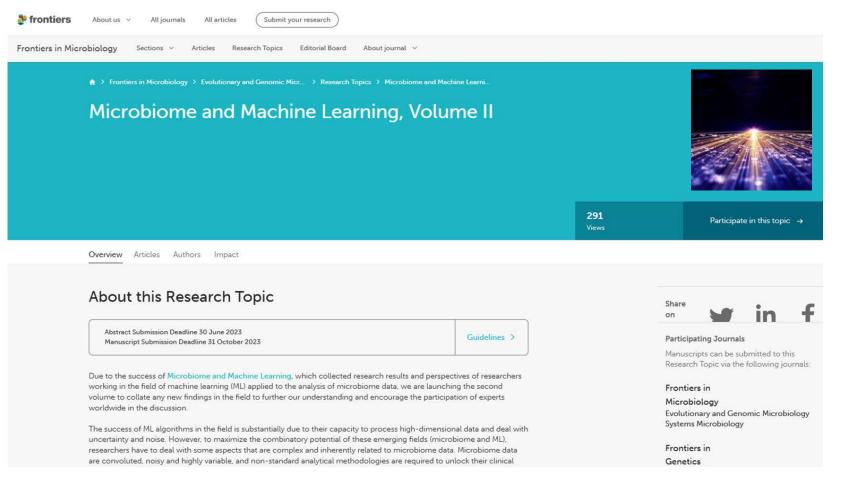
Topic Editors:
Isabel Moreno Indias, Marcus Claesson, Aldert Zomer, David Gomez-Cabrero

Microbiome and Machine Learning | Frontiers Research Topic (frontiersin.org) - Impact

[Microbiome and Machine Learning | Frontiers Research Topic (frontiersin.org) – Volume II](#) (2023)

The collection will include:
- Original Research
- Methods
- Reviews
- Technology and Code
- Perspective

Describing machine learning novel tools or approaches for analysing and classifying microbiome data, improving the interpretability of ML application results, or employing Explainable AI

Topic Editors:

Domenica D'Elia, Isabel Moreno Indias, Marcus Claesson, Aldert Zomer, Erik Bongcam-Rudloff, Randi Jacobsen Bertelsen

The table below displays the general rationale of Frontiers' fee policy. Journal-specific APCs are listed on each journal's page, which also contains journal-specific article types. Please note that APCs are subject to periodic revision.

Fees

Category	Description	Article type A	Article type B	Article type C	Article type D
1	Journals in fields that have lower research budgets	$1,260	$765	$490	0
2	Journals in fields that support open access	$2,080	$955	$490	0
3	Mature and impactful journals that are well-established and in fields that support open access	$2,720	$1,260	$490	0
4	Sustainable and impactful journals in fields that support open access	$3,225	$2,020	$490	0

Table 1: *Frontiers' APCs are structured in four categories, depending on the journal's maturity and the level of research funding and/or open access funding available in various research communities. They also support article types with lower prices or articles that are free of charge. Frontiers' APCs are charged in US$.*

For example, for Original Research articles (article type A), Frontiers' APCs range from **US$ 3,225** *in our most mature journals with well-established open access (OA) support, to* **US$ 1,260** *in newly launched journals and/or fields that have lower research budgets available and/or where OA is not yet well-supported (e.g. humanities and social sciences fields). This means that mature journals in well-funded fields have higher APCs and partially subsidize journals with lower APCs that are younger or cover communities with less funding. This allows all research communities to benefit from open access.*

Fee discounts

Our fee support program ensures that all articles that pass peer review can benefit from open access – regardless of the author's field or funding situation.

Authors and institutions with insufficient funding will be eligible for discounts to their APC. This includes authors in countries classified by the World Bank as low and low-middle income economies.

To apply for fee support, please complete our fee support application form.

Your request for a fee support will in no way influence whether your submission is accepted or rejected for publication.

ML4Microbiome deadline for article submission:

15-20 June 2023

✓ **YouTube Playlists: 97 subscribers**

(560) ML4Microbiome Workshop 2021 – YouTube

(560) ML4Microbiome Symposium 2021 – YouTube

✓ **Accessible also from the Action website**

Training Material - ML4 Microbiome

Communication

✓ [LinkedIn ML4Microbiome Channel](): **216 Followers**

- Visitor highlights June 2022-23: 357
- Page views: 139
- Unique visitors:32

✓ [Twitter ML4Microbiome Channel]():

546 Following; 366 Followers

Internal Communication & Document Sharing

✓ Slack Channel: 160 members

✓ GoogleDrive

✓ Gmail and COST mailing service

A4.2: Design, organisation and implementation of regular training courses and workshops

Milestones:
M4.1: Initial workshop (Y1Q4), training courses (Y2Q3/Y3Q3) and YouTube tutorials (Y4Q3)

Achievements:
✓ **Training Schools** (2019; 2021; 2022; 2023) – Programme and training material available on the **Action website, YouTube and GoogleDrive ML4Microbime accounts**

<div align="center">(Activities and Events - ML4 Microbiome)</div>

✓ **Workshops** (August 2022, Turku – May 2022, Tirana – October 2021, Bari, virtual)

✓ **Symposium** (October 2021, Bari, virtual)

Major Deliverables:
D4.1-3: Reports on the workshops & training courses of best practices in ML/Stats for microbiome studies
D4.4: Educational material shared: tutorials on paper and video (including a YouTube channel)

WG4 Report: March-June 2023

- Periodic virtual meetings (every 2-3 weeks)
- Website, LinkedIn and Twitter accounts - timely updated
- [Frontiers in Microbiology: The Second Volume launched](#)
- [ML4Microbiome workshop in BITS 2023 (21-06-2023)](#)
- Perspective paper drafting ongoing….to be continued and hopefully concluded on June 9th in Cork (WG4 writing meeting)

MICROBIOME

Bioinformatics Italian Society

BITS 2023 BARI June 21-23
BITS Annual meeting

Workshop

Advanced machine-learning approaches for the analysis of microbiome data
Workshop

The analysis of the human microbiome has recently attracted the attention of several research communities, due to its potential diagnostics, prognostics, and therapeutics role for several diseases, including diabetes, liver cirrhosis, some types of cancer (e.g., colorectal cancer) as well as for disorders like the autism spectrum disorder. The adoption of statistical and Machine Learning approaches appears very promising to elucidate existing (or identify novel) relationships between microbiome conditions and diseases or to build descriptive and predictive models that can be adopted to improve existing therapeutic procedures. The workshop "Advanced machine learning approaches for the analysis of microbiome data" will focus on advanced machine learning approaches and their (potential or actual) application to microbiome data, including semi-supervised learning methods, multi-view learning methods, and transfer learning approaches.

Workshop Chairs

- **Michelangelo Ceci**, University of Bari, Italy
- **Gianvito Pio**, University of Bari, Italy
- **Domenica D'Elia**, CNR-Institute for biomedical Technologies, Bari, Italy

BITS 2023 BARI June 21-23

WELCOME TO BITS 2023

CALLS

WORKSHOP

Andrea Simeon, Mathematician and Data Science researcher at BioSense Institute, Novi Sad, Serbia. Focused on applying Machine Learning and Deep Learning techniques in microbiome studies and exploring different preprocessing pipelines for analysing amplicon and shotgun sequence data. Thanks to the research work carried out during her ML4Microbiome COST Action STSM, Andrea won from the Faculty of Sciences, University of Novi Sad, Serbia, the Aleksandar 'Saša' Popović Award For Best Student Paper – 2022.

Talk title: Multi-class boosting with adversarial multi-arm bandits on incomplete microbiome views

Abstract: Microbiome has been massively associated with different diseases and disorders. To identify individual microorganisms and their abundances across samples, different sampling, sequencing and preprocessing techniques could be considered. This leads to different input feature sets (views) to learn predictive models through machine learning (ML) approaches. ML models aid in finding the associations between microbiome and disease. Standard (single view) ML models are not capable of dealing with multiple views at once, and thus they were upgraded to fit multi-view datasets (e.g. Adaboost and Multi-view Adaboost). Moreover, microbiome data comes from various sources, and incompleteness is often inevitable. Existing classifiers, even multi-view, cannot be directly used because they cannot work with incomplete views and in multi-class settings. To our knowledge, there is no multi-view boosting algorithm for multi-class classification with incomplete views. The proposed algorithm is the extension of an existing multi-view boosting algorithm based on multi-arm bandits, now able to work in multi-class setting and with incomplete views (views with missing sample representation). At each iteration, it proclaims one view as the winning using adversarial multi-arm bandits and uses its predictive information to update the final model weights and prediction in a boosting process. Three data sets were created from several microbiome studies and used to examine the performance of the proposed algorithm. One of the experiments showed a 7% increase in F1 score compared to a single view classifier, while the other showed 54%. The application domain is not restricted to microbiome data. Further work will involve examinations in other domains.

Welcome to BITS 2023

19th Annual Meeting of the Bioinformatics Italian Society
June 21-23, 2023, Bari, Italy

The Bioinformatics Italian Society is glad to invite you to its 19th Annual Conference.

The conference highlights keynote talks by excellent scientists in bioinformatics and its applications, presentations of state-of-the-art research in computational biology, and poster sessions on the latest research progress.

The 2023 edition will be held in **Bari 21-23 June**. The event will take place at the Hotel Excelsior Bari, Via Giulio Petroni, 15, 70124 Bari (IT).

BITS 2023 BARI June 21-23

BITS Meeting 2023 .:. BITS Bioinformatics Italian Society

BITS Meeting 2023: Workshop

Programme - 21 June 2023
[09:40-10:00] **Welcome and workshop introduction (Michelangelo Ceci, Domenica D'Elia)**
[10:00-10:40] **Silvio Tosatto,** DOME: recommendations for supervised machine learning validation in biology
[10:40-11:10] Coffee break
[11:10-11:40] **Andrea Simeon,** Multi-class boosting with adversarial multi-arm bandits on incomplete microbiome views
[11:40-12:00] **Adriano Zaghi,** Machine learning models for the detection of antimicrobial resistance using
synthetic data
[12:00-12:20] **Meritxell Pujolassos,** coda4microbiome: compositional data analysis for microbiome crosssectional and longitu
[11:50-12:10] **Donato Romano,** Explainable Artificial Intelligence (XAI) for Microbiome Data Analysis in Autistic Spectrum Disorder

Perspective paper: [PAPER-DRAFT-revised-DD-05-04-2023 - Google Docs](#)

Frontiers in Microbiology, second volume edition:

PERSPECTIVE Paper
Perspective articles present a viewpoint on a specific area of investigation.
They should provide the following:
1) Discuss current advances and future directions
2) Clear presentation of the authors' perspective
3) Accurate presentation and citations of other authors' work
4) May include original data as well as personal insights and opinions

Perspective articles are peer-reviewed, have abstract length 250 words, a maximum word count of 3,000 and may contain no more than 2 Figures/Tables.

Authors are required to pay a fee (B-type article) to publish a Perspective article.

Perspective articles should have the following format:
1) Abstract
2) Introduction
3) Subsections relevant for the subject
4) Discussion

Perspective paper: [PAPER-DRAFT-revised-DD-05-04-2023 - Google Docs](#)

Titles proposed:
1. **Bridging the gap between microbiome data and machine learning: Insights from the ML4Microbiome COST Action**
2. **Bridging the Microbiome-Machine Learning Divide: Insights from the ML4Microbiome COST Action**
3. **Facing the challenges encountered by Machine Learning methods on microbiome data: Insights from the ML4Microbiome Cost Action**
4. **On the challenges encountered by Machine Learning methods on microbiome data: Insights from the ML4Microbiome Cost Action**
5. **Challenges encountered by Machine Learning methods on microbiome data: Insights from the ML4Microbiome Cost Action**
6. **Machine Learning and microbiome data: Insights to face the challenges from the ML4Microbiome Cost Action**
7. **Machine Learning Meets Microbiome: Unpacking the ML4Microbiome COST Action**
8. **The Intersection of Microbiome and Machine Learning: Key Takeaways from the ML4Microbiome COST Action**
9. **Uniting Microbiome Data and Machine Learning: Lessons Learned from the ML4Microbiome COST Action**
10. **Insights from the ML4Microbiome COST Action: Navigating the Convergence of Microbiome and Machine Learning**
11. **The ML4Microbiome COST Action: A Bridge between Microbiome Data and Machine Learning**

Abstract
The rapid development of machine learning (ML) techniques has opened up the data-dense field of microbiome research for novel therapeutic, diagnostic, and prognostic applications targeting a wide range of disorders, which could substantially improve healthcare practices. However, several challenges must be addressed to fully exploit the benefits of ML in this field. In particular, there is a need to establish "gold standard" protocols for conducting ML analysis experiments and improve interactions between microbiome researchers and ML experts. The Statistical and Machine Learning Techniques in Human Microbiome Studies (ML4Microbiome) COST Action is a European network established in 2019 to promote collaboration between discovery-oriented microbiome researchers and data-driven ML experts, aiming to optimise and standardising ML approaches to the analysis of microbiome datasets. This perspective paper presents the key achievements of ML4Microbiome in optimising, standardising and disseminating the use of ML on human microbiome datasets. These achievements include identifying predictive and discriminatory 'omics' features, improving repeatability and comparability, developing automation procedures, and defining priority areas for the novel development of ML/statistics methods targeting the microbiome. The insights gained from ML4Microbiome will help to maximize the potential of ML in microbiome research and pave the way for new and improved healthcare practices.

Perspective paper: [PAPER-DRAFT-revised-DD-05-04-2023 - Google Docs](#)

Presently completed:
1. Abstract
2. Introduction
3. Subsection: The ML4Microbiome Action Plan and Challenges

Discussion
The discussion should be articulated as follow:
1. Key findings in clear and concise language
2. Place our results within the context of previous studies
3. Discuss potential future research
4. "take-away" statements for the readers

Presently:
1. The current state of ML applications for microbiome data analysis (WG1)
2. Benchmark datasets and online repositories
3. Data preprocessing, transformation and normalization: challenges and solutions (still incomplete)
4. Dissemination, communication and training (not sure it is the case of inclusion unless of training)
5. Define priority areas for the novel development of ML/stats applications targeting the microbiome (still incomplete)
6. New emerging areas: Ethical considerations (contributor Ramona Suharoschi)
7. Intellectual property (IP) rights (contributor Ramona Suharoschi)
8. Explainable AI (contributors Ciprian-Octavian, Marcus Frome, Alina, Elena-Simona)

Dissemination & Communication Rules

- **Publications** – any article where at least two members are involved can be published with acknowledgements to ML4Microbiome, and communication has to be done to WG4 (Rules at: <u>Visual identity – COST</u>)

- Production of reports and their publication, hopefully in journals or web repositories that provide a DOI

THANK YOU SO MUCH

for your attention!

Question?