

# Microorganisms identification at strain-level: statistical and deep learning methods for analysis of MALDI-TOF mass spectra

P.M. Grollemund<sup>1,2,\*</sup>, M. Bourlioux, C. Arous<sup>1</sup>, A. Geneste<sup>1</sup>, A. Grizon<sup>1</sup>

<sup>1</sup> UMR 0545 Fromage, Université Clermont Auvergne, INRAE, VetAgro Sup, Aurillac, France

<sup>2</sup> LMBP Laboratoire de Mathématiques Blaise Pascal, Campus universitaire des Cézeaux, Aubière, France

\*Corresponding author e-mail: [paul\\_marie.grollemund@uca.fr](mailto:paul_marie.grollemund@uca.fr)

## Abstract

The use of Matrix Assisted Laser Desorption Ionization - Time of Flight Mass Spectrometry (MALDI-TOF MS) represents a promising alternative to conventional molecular, phenotypic or biochemical microorganism identification methods [1]. Nevertheless, some significant limitations remain that must be addressed before the approach can be considered fully reliable: while the mass spectrometer provides robust identification at the species level, its discriminatory power remains insufficient for consistent strain-level resolution [2]. Besides, the analytical performance of the instrument is significantly challenged when there are multiple strains mixed within a sample [3].

Therefore, an in-depth analysis of the spectra is needed to discriminate microorganisms at the strain-level and to take full advantage of MALDI-TOF mass spectrometry for routine microbial identification. The performance of statistical approaches, including Functional Partial Least Square Discriminant Analysis (FPLS-DA), and deep learning architectures, such as Functional Neural Networks [4], using the functional structure of spectra, is evaluated. During the initial phase of analysis, an average classification accuracy of 90% is achieved for strain-level assignment, using a training set comprising seven biological replicates for each of six *Streptococcus thermophilus* strains studied.

This work constitutes a proof of concept and suggests that a compromise can be reached between limited data acquisition and robust strain-discrimination performance in both mono- and multi-strain samples. The results reported here provide a base for the development of an automated microorganism-identification workflow by combining standardized sample processing, reproducible MALDI-TOF spectral preprocessing and functional classification methods.

## Keywords

MALDI-TOF MS, strain-level identification, *Streptococcus thermophilus*, mass spectra data, statistical models

## References

- [1] S.K. Urwyler, J. Glaubitz, Advantage of MALDI-TOF-MS over biochemical-based phenotyping for microbial identification illustrated on industrial applications, *Letters in Applied Microbiology*, Volume 62, Issue 2, 1 February 2016, Pages 130–137, <https://doi.org/10.1111/lam.12526>
- [2] B. Mehramouz, E. Khodadadi, E. Khodadadi, S. Taghizadeh, K. Ghanbarov, Ş. Köse & H. S. Kafil (2022). The Use of MALDI-TOF Mass Spectrometry Technology in Molecular Analysis of Microbial Pathogenesis. <https://doi.org/10.33263/BRIAC135.450>
- [3] M. A. Reeve, D. Bachmann, MALDI-TOF MS protein fingerprinting of mixed samples, *Biology Methods and Protocols*, Volume 4, Issue 1, 2019, bpz013, <https://doi.org/10.1093/biomethods/bpz013>
- [4] F. Heinrichs, M. Heim & C. Weber (July 2023). Functional neural networks: Shift invariant models for functional data with applications to EEG classification. *In International Conference on Machine Learning* (pp. 12866-12881). PMLR.

## Acknowledgments

Research supported by The University Institute of Technology IUT Clermont-Auvergne (<https://iut.uca.fr/>)