ABSTRACT

Antimicrobial Resistance (AMR) occurs when microbes like bacteria, viruses, and fungi develop the ability to withstand antimicrobial treatments (drugs used to treat infections), increasing the risk of disease spread, severe illness and death.

Matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry can be used to rapidly identify microbial species [2]. A recent work [4] has shown that machine learning-based method may be applied to MALDI-TOF spectra of bacteria to test antimicrobial susceptibility (or resistance), this allowing more suitable treatment decisions in a clinical context.

Since real data are often not enough to properly train a machine learning model (and thus to make correct predictions), the use of synthetic data is becoming increasingly important.

This work will focus on: (i) evaluating which model - e.g. Conditional variational autoencoder (CVAE) or more specific models [1, 3] - is more suitable for the generation of synthetic data for the study of AMR from a theoretical point of view; (ii) applying the best model to generate synthetic datasets; (iii) testing their goodness w.r.t. the original dataset using consolidated metrics.

**Keywords:** synthetic data, mass spectrometry, MALDI-TOF, bacteria, bacteriology, antimicrobial resistance, AMR.

REFERENCES

#### Bielow C, Aiche S, Andreotti S, Reinert K. MSSimulator: Simulation of Mass Spectrometry Data. *Journal of Proteome Research* 2011 10 (7), 2922-2929. doi: 10.1021/pr200155f.

#### Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis*. 2009 Aug 15;49(4):543-51. doi: 10.1086/600885. PMID: 19583519.

1. Sohn K, Lee H, Yan X. Learning Structured Output Representation using Deep Conditional Generative Models. *Advances in Neural Information Processing Systems* 28 (NIPS 2015). <https://proceedings.neurips.cc/paper_files/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf>.
2. Weis C, Cuénod A, Rieck B et al. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. Nat Med 28, 164–174 (2022). <https://doi.org/10.1038/s41591-021-01619-9.>