**Machine learning model development for the classification of bacterial strains using their Raman chemical fingerprint**

Responsible Scientific Assistant: Marwan El Chazli (marwan.elchazli@epfl.ch)
Professor: Christophe Galland (chris.galland@epfl.ch)
Project type: Masters project

**Context**

Bacterial contamination of food products represents an ongoing public health hazard, results in a massive waste of food and costs hundreds of millions of francs to food companies every year.

At the heart of it is the fact that current bacterial detection and identification methods used in the food industry (and many other industries) are slow and can require up to a week to provide results.

At the Laboratory of Quantum and Nano-Optics (LQNO), we are looking to build the next generation of bacterial sensing technology, using Raman spectroscopy and its variants, combined with a solid understanding of biochemistry, microfluidics know-how and powerful machine learning techniques.

The project is extremely interdisciplinary and gives students the opportunity to familiarize themselves with a wide range of technologies, as well as gain experience in translating basic sciences to applied research.

**Project overview**

The working principle of our method is illustrated in Figure 1. In order to detect and identify bacterial cells, their "chemical fingerprint" is acquired through Surface-Enhanced Raman Spectroscopy and run through a machine learning algorithm to determine the specific strain under observation. Available methods for this kind of application include PCA, LDA, Neural Networks, etc.

The goal of the project will be to test different machine learning approaches, particularly neural networks-based techniques, to classify bacterial cells using their Raman scattering spectra.

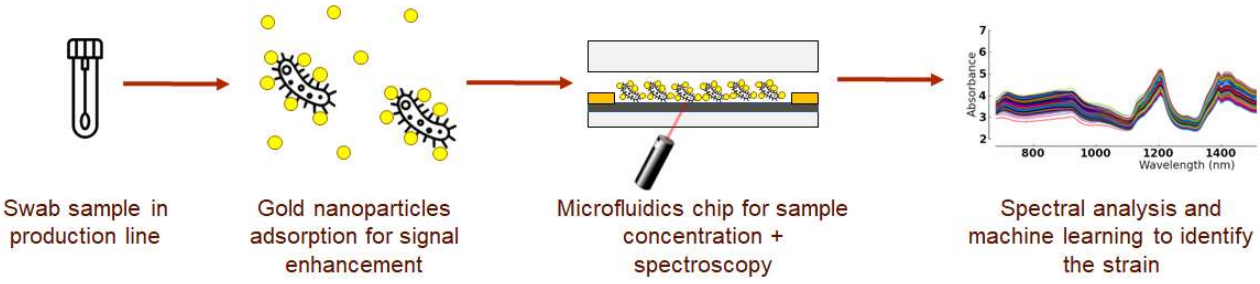


*Figure 1. Current process flow: a liquid sample containing bacteria is enriched with gold nanoparticles and immobilized in a microfluidic chip. Then, the spectral signature of the bacteria-gold complexes is acquired and analysed with machine learning to identify the strain of bacteria.*

**What the student will do**

Throughout the project, the student will use pre-processing and machine learning methods in Python to work with large amounts of spectral data and classify them. They will be training and testing models and evaluating their effectiveness at classifying bacterial strains through a variety of metrics. Depending on the progress of the project, opportunities for publications might arise, which the student will be encouraged to participate in if they wish.

**Who we are looking for**

We are looking for students from any degree with previous experience in Python coding and machine learning applications. That experience need not be extensive, so long as the student is willing and interested to learn. Most importantly, we are looking for students who are motivated and able to work autonomously.

**Interested? Get in touch to apply or learn more ->** marwan.elchazli@epfl.ch

**Some useful references**

Dib, O. H., Assaf, A., Grangé, E., Morin, J. F., Cordella, C. B. Y., & Thouand, G. (2023). Automatic recognition of food bacteria using Raman spectroscopy and chemometrics: A comparative study of multivariate models. In Vibrational Spectroscopy (Vol. 126, p. 103535). Elsevier BV. https://doi.org/10.1016/j.vibspec.2023.103535

Ho, CS., Jean, N., Hogan, C.A. *et al.* Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat Commun* **10**, 4927 (2019). https://doi.org/10.1038/s41467-019-12898-9

Thomsen, B.L., Christensen, J.B., Rodenko, O. *et al.* Accurate and fast identification of minimally prepared bacteria phenotypes using Raman spectroscopy assisted by machine learning. *Sci Rep* **12**, 16436 (2022). https://doi.org/10.1038/s41598-022-20850-z