RAMP: drug spectra

setup by:

Camille Marini (LTCI/CNRS)

Alex Gramfort (LTCI/Télécom ParisTech)

Sana Tfaili (Lip(Sys)^2/UPSud)

Laetitia Le (Lip(Sys)^2/UPSud)

Mehdi Cherti (LAL/CNRS)

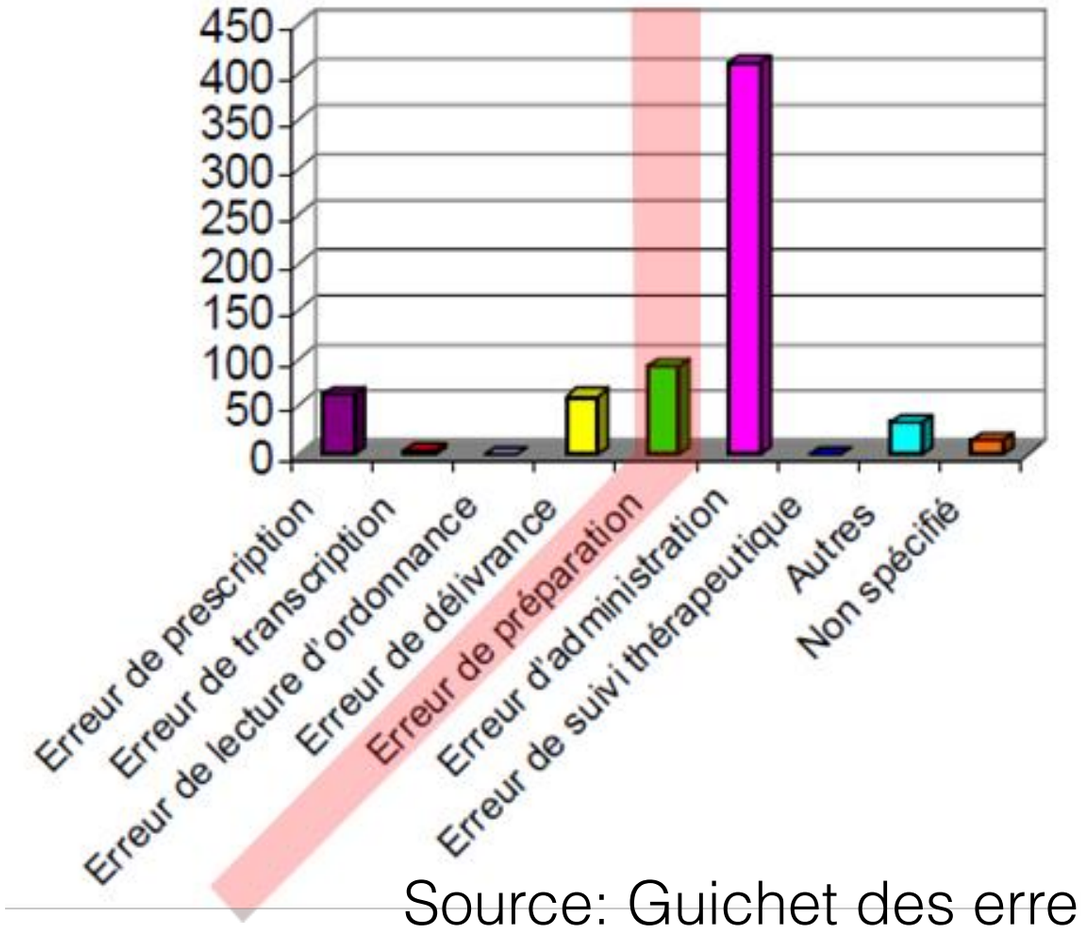
Balázs Kégl (LAL/CNRS)

RAMP: drug spectra

- Chemotherapy is one of the most used treatment against cancer.
- It uses chemical substances (chemotherapeutic agents) which kill cells that divide too quickly.
- These chemical substances are often diluted in a particular solution and packaged in bags, diffusers, or syringes, before being administered.

RAMP: drug spectra

- Wrong medication (wrong chemotherapeutic agent or wrong concentration) can have major impacts for patients.



Source: Guichet des erreurs médicamenteuses 2009, publié en Juillet 2010

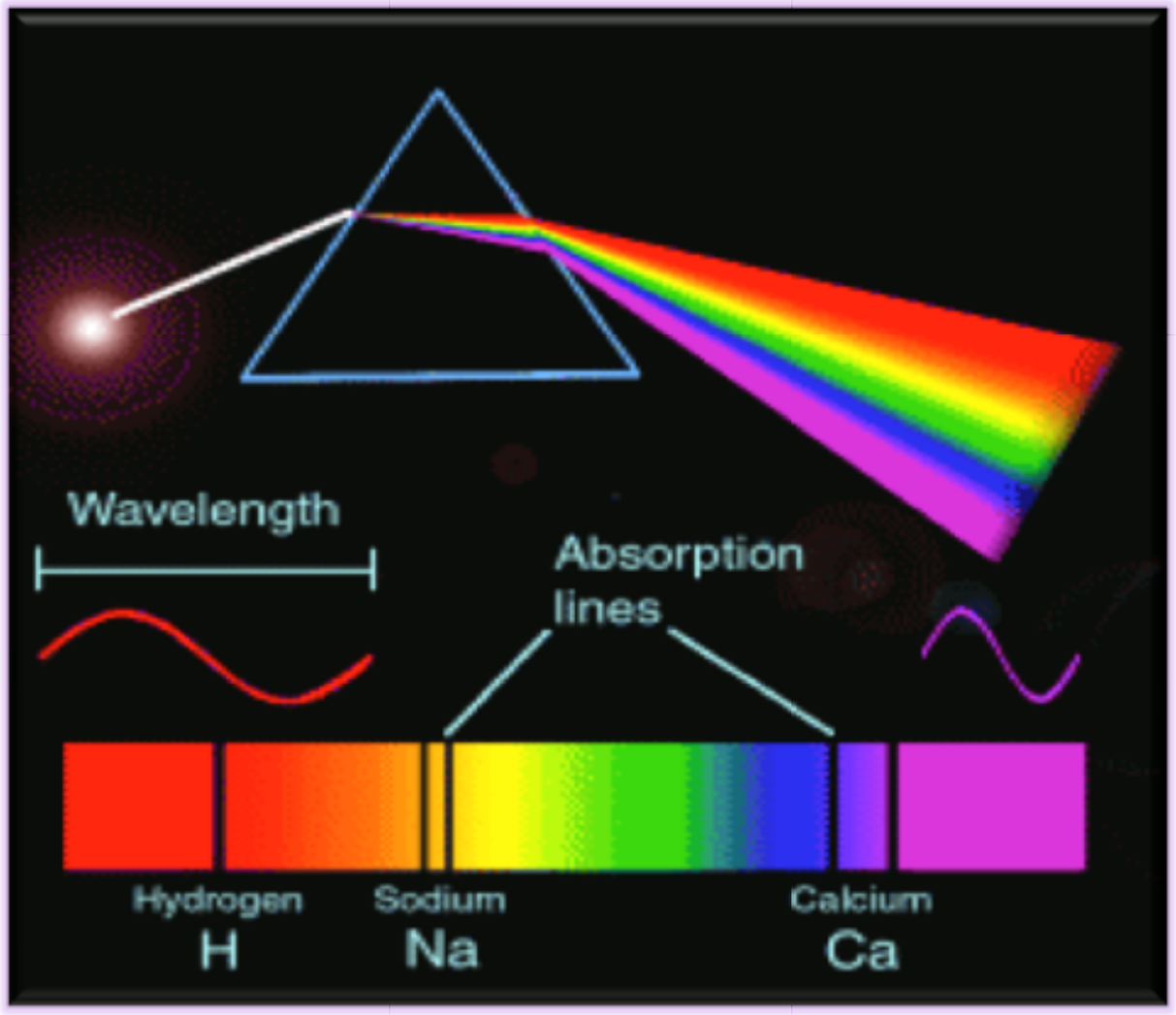
- **Objective:** Use data science and machine learning to prevent wrong medication.

Objective

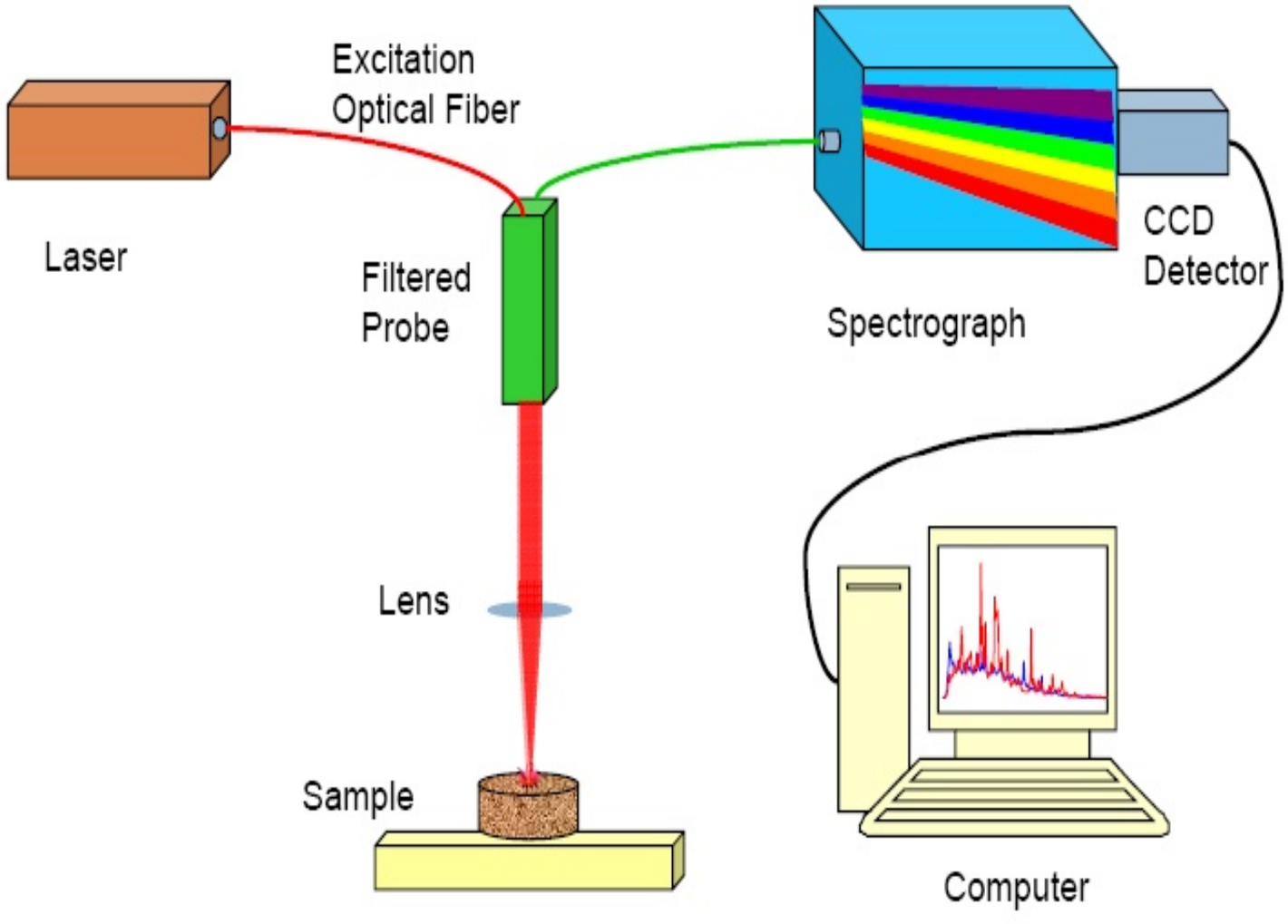
- Check that medication contain:
  - The **good chemotherapeutic agent** / molecule
  - The **good dosage** / concentration

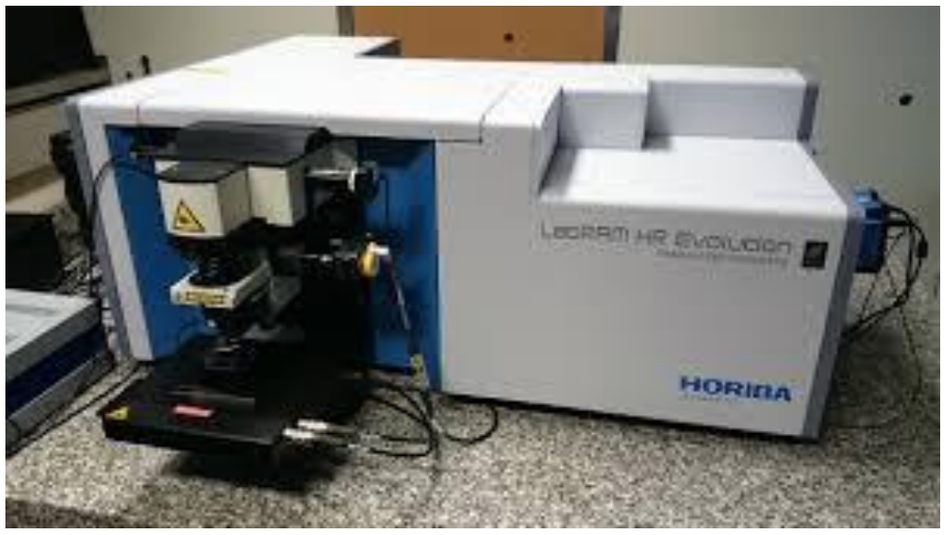
Data: Raman spectroscopy

- Spectroscopy is the study of molecular structure and dynamics through the absorption, emission and scattering of light



Data: Raman spectroscopy

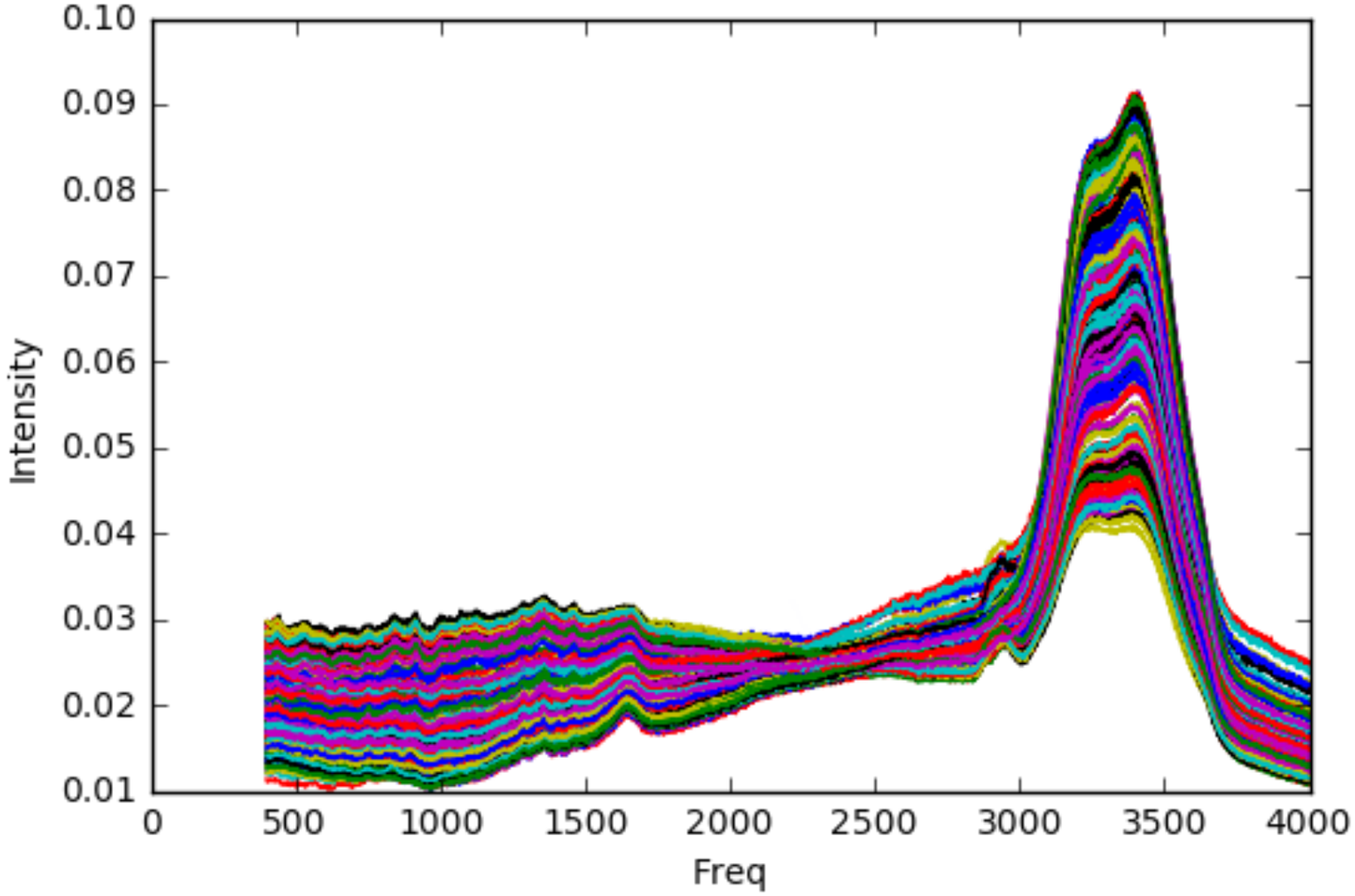




- **Spectrum** is a graph of intensity of absorbed or emitted radiation by sample verses frequency or wavelength

Task

- The goal is to develop **prediction models** able to:
  - **identify** and **quantify** chemotherapeutic agents from their Raman spectra.
- Samples are spectra !



Data description

- 4 types of chemotherapeutic agents (called molecule)
- in 3 different packages (called vial)
- diluted in 9 different solutions (called solute gammes)
- and having different concentrations.

- Total of 360 spectra were measured for each agent, except for one (348 spectra).
- molecule: A for infliximab, B for bévacizumab, Q for ramucirumab, R for rituximab.
- vial: Vial type. Three possible values: 1, 2, 3.
- solute: Solute group. Fourteen possible values: 1, 2, ..., 14.
- concentration: Concentration of the molecule.
- spectra: Intensity of Raman spectrum. Dimension: (n_samples, 1866)

Objective and metrics

- There are **2 objectives**:
  - **classification**: predict which molecule it corresponds to given the spectrum.
  - **regression**: predict the concentration of a molecule. The prediction should not depend on the vial or the solute group.
- The **error metrics** are:
  - For classification: The 0-1 loss.
  - For regression : The mean absolute relative error (mare):

$$\frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Questions?

Let’s have a look at the starting kit