Unsupervised (machine) learning to improve analysis of flow cytometry data of extracellular vesicles

An internship position for Master students (Physics, Computer Science, Medical Informatics) is available at the Biomedical Engineering and Physics department of the Amsterdam University Medical Centers. In our group, new treatment and diagnostic procedures based on innovative physical techniques are developed. Research is performed by a multidisciplinary team that includes physicists, engineers, mathematicians, medical doctors, biologists, and chemists.

Background. Human body fluids are "liquid biopsies". Analysis of the composition of "liquid biopsies" provides detailed insight in the health or disease status of an individual. A key component of body fluids are extracellular vesicles (EVs), which can be seen as biological nanoparticles. Because the concentration, composition and function of EVs chance in disease, accurate EV measurements together with sophisticated data analyses will likely provide new clinical information.

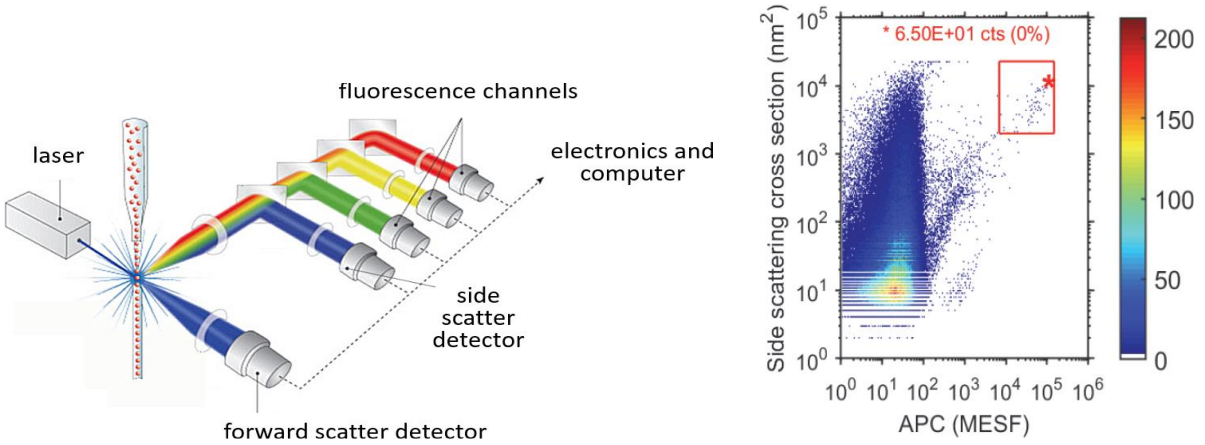

Figure 1. Left: schematic representation of a flow cytometer. Flow cytometry detects light scattering and fluorescence signals of hundreds of thousands of extracellular vesicles (EVs) per sample. Right: snap-shot of a typical flow cytometry dataset of EVs.

Problem. Worldwide, our lab is at the forefront of EV detection by flow cytometry. Currently, we are analyzing hundreds of patient samples per month to investigate the clinical value of EVs. In one patient sample, flow cytometry (Fig. 1, left) typically detects light scattering and fluorescence signals of hundreds of thousands of individual EVs as small as 150 nm. However, it is unclear how to differentiate between relevant data and background noise. Hitherto clustering of potentially relevant data is done manually, which is a slow and subjective procedure.

Solution. Unsupervised (machine) learning offers a promising tool to differentiate between relevant data and background noise.

Tasks. In this project, you will develop an automated, quantitative technique to cluster flow cytometry data and identify clinically relevant clusters of EVs. To realize this, you will use unsupervised machine learning techniques (c.q. clustering algorithms). Clinical datasets are available. Challenges involve (1) data calibration, (2) selecting, optimizing and comparing clustering algorithms, and (3) writing computationally efficient algorithms.

Learning points: At the end of the project you have learned the essential considerations involved in preprocessing of data for machine learning. You are familiar with clustering data without a label by using unsupervised learning in a computationally efficient framework. In addition, you have insights in the plethora of available clustering algorithms.

Requirements: ample experience with programming. Most preferably, you are already familiar with the Python packages Numpy and Sklearn.

Contact

Henk van Voorst, M.Sc. (h.vanvoorst@amc.uva.nl)
Dr. Edwin van der Pol (e.vanderpol@amc.uva.nl)
Dr. Henk Marquering (h.a.marquering@amc.uva.nl)
Websites: https://www.amc.nl/bmep and https://www.edwinvanderpol.com