**AHEAD Medicine Interview Questions**

Programming Task (1-2 hours on a laptop)

You have familiarized yourself with the FCS data files and relevant documents we shared with you prior to the interview. As well as the Python library FlowCal to read and manipulate the FCS data. Please put your answers on GitHub and share the link with us.

TASK: There are a group of patients who were diagnosed either COVID-19 positive (sick) )or negative (healthy). Each FCS file represents the specimen collected from one patient. Build an automatic predictor using a ML model of your selection, the labels provided in the “EU\_label.xlsx” as ground truth, and marker-channels with “use” = 1 in “”EU\_marker\_channel\_mapping.xlsx” as ” as data features.

NOTE: This dataset is NOT PERFECT. Do not spend too much time tuning the model for performance. We focus more on how you build the workflow and get your results.

Answers include three .py files. “copy\_files.py” and “preprocess\_npy.py” are for data preparations. ML.py is the main workflow of the ML.

Steps in ML.py weren’t being optimized but they present the main workflow.

Please let me know any questions about the workflow.

**Bonus Question**

1. Please explain the fundamental principles of flow cytometry and walk through the step-by-step process of how it works? Additionally, highlight some common applications of flow cytometry in scientific research and clinical settings.

All answers can be found in the paper. “McKinnon KM. Flow Cytometry: An Overview. Curr Protoc Immunol. 2018 Feb 21;120:5.1.1-5.1.11. doi: 10.1002/cpim.40. PMID: 29512141; PMCID: PMC5939936.”

The fundamental principles of flow cytometry can be found in the “instrumentation” section in the paper. Flow cytometry has been applied in many disciplines such as immunology, virology, molecular biology, cancer biology and infectious disease monitoring. For example, it is very effective for the study of the immune system and its response to infectious diseases and cancer. This is also a highlight from the paper.

1. Below are plots of selected cell surface biomarkers of blood cell samples. Researchers are interested in picking out cells marked in yellow (; a high-density chunk at the bottom-right) for further analysis. How would you suggest a method to automatically identify these cells?

Don’t really understand this question. It seems the yellow-marked region is clearly defined, gating might be a good approach?





