**Artificial Intelligence-Driven Automation of Flow Cytometry Gating**

Gabriella Rivera

Applied Data Science Master’s Program

Shiley Marcos School of Engineering / University of San Diego  
 [email@sandiego.edu](mailto:email@sandiego.edu)

John Vincent Deniega

Applied Data Science Master’s Program  
 Shiley Marcos School of Engineering / University of San Diego  
 [email@sandiego.edu](mailto:email@sandiego.edu)

**ABSTRACT**

Flow cytometry is a biochemical analytical process that involves identifying and classifying populations of peripheral blood mononuclear cells (PBMC) - such as lymphocytes and monocytes - using chemical substrates, imaging, and highly-capital intensive medical research equipment. Clustering algorithms aim to objectively simplify the classification accuracy of a given population's cellular type to further immunological and clinical diagnostic purposes. By exploring algorithms like K-means, Agglomerative, and Gaussian Mixture Modeling, the dataset containing metrics on front and side scatter areas of cell scans, fluorescence on markers, and other dimensions can potentially lead to automatic gating and bring forth insights on cell population characteristics, resulting in increased analytical throughput and increased rate of medicinal and pharmacological breakthroughs.

**1 Introduction**

Automating flow cytometry gating with artificial intelligence has significant benefits to the biochemical and larger bioscience community with respect to cellular identification and pharmacological discovery and development. Using mathematically-driven principles inherent to artificial intellgence, biochemists may be able to reduce the analytical inputs required to perform routine cellular classification and clustering of PBMCs. This labor-intensive process ultimately serves to evaluate the efficacy of experimental groups related to clinical trials that are required in the development of new life-saving medicines.

**2 Background**

Flow cytometry can be a capital-intensive process that requires significant investments in laboratory-grade biomedical equipment and proprietary analytical software licenses. With Flow Cytometry Standard (FCS) files readily available on public repositories and by leveraging open-source and permissive license packages such as Scikit-Learn and Matplotlib, this project aims to be able to solve this problem with existing no-cost alternatives, which implies a significant reduction in the barriers to entry related to medical pharmacological development.

2.1**Problem Identification and Motivation**

As of this publication, flow cytometry gating is a manual process that involves a highly-trained biochemist to process and analyze the results of optical scans of cellular assays that may be further augmented by fluorescent substrates. Because of the complex and highly-dimensional nature of the data, these scientists rely on a best-practices approach based on their own respective processes and frameworks. Because of the potential variability of these processes and frameworks, the resulting findings from interpreting scan results is dependent on both the breadth and depth of methods of a given supervising scientist, thus resulting in both an increase in cost of analysis due to human error and omission as well as a reduction in consistency of results.

2.2**Definition of Objectives**

The research team aims to utilize open-source and publicly-available resources from Python libraries Scikit-Learn and Matplotlib as well as FCS data hosted by FlowRepository (2020). Once data is cleaned for noise from scan data, the team aims to train models or machine-learning applications that have potential for value-added analysis relative to that of a typical human biochemist. Upon evaluation, success is generally defined when automated analysis reaches parity with a human analyst of at least 90% classification accuracy of PBMCs toward their respective dendritic cellular type on an unseen test set containing FCS scan data. In the event that this evaluation criterion is not met, further justification would have to be provided whether the measured degree of accuracy is acceptable relative to the speed of analyses.

**3 Literature Review (related works)**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.1 **FlowAI: Automatic and interactive anomaly discerning tools for flow cytometry data**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.2**An open-source solution for advanced imaging flow cytometry data analysis using machine learning**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.3**Comprehensive Phenotyping of Human Dendritic Cells and Monocytes**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.4**Application of machine learning for cytometry data**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

3.5**Recommendations for using artificial intelligence in clinical flow cytometry**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

**References**

Beckman Coulter. (2022). *Automatic gating - Beckman Coulter*. Flow Cytometry. <https://www.beckman.com/flow-cytometry/software/cytobank-premium/learning-center/automatic-gating>

FlowRepository. (2020). FlowRepository ID FR-FCM-Z32U. Retrieved from <http://flowrepository.org/experiments/3166/download_ziped_files>

Hennig, H., Rees, P., Blasi, T., Kamentsky, L., Hung, J., Dao, D., Carpenter, A. E., & Filby, A. (2016). An open-source solution for advanced imaging flow cytometry data analysis using machine learning. *Methods*, 112(2017), 201–210. <http://dx.doi.org/10.1016/j.ymeth.2016.08.018>

Hu, Z., Bhattacharya, S., & Butte, A. J. (2022). Application of machine learning for cytometry data. *Frontiers in immunology,* 12, 787574. <https://doi.org/10.3389/fimmu.2021.787574>

Mair, F., & Leichti, T. (2020). Comprehensive Phenotyping of Human Dendritic Cells and Monocytes. *Journal of Quantitative Cell Science*, 99(3), 231–242. <https://doi.org/10.1002/cyto.a.24269>

Monaco, G., Chen, H., Poidinger, M., Chen, J., Magalhaes, J., & Larbi, A. (2016). FlowAI: Automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics*, 32(16), 2473–2480. <https://doi.org/10.1093/bioinformatics/btw191>

Ng, D. P., Simonson, P. D., Tarnok, A., Lucas, F., Kern, W., Rolf, N., Bogdanoski, G., Green, C., Brinkman, R. R., & Czechowska, K. (2024). Recommendations for using artificial intelligence in clinical flow cytometry. *Cytometry part b: Clinical cytometry*, 106(4), 228–238. <https://doi.org/10.1002/cyto.b.22166>