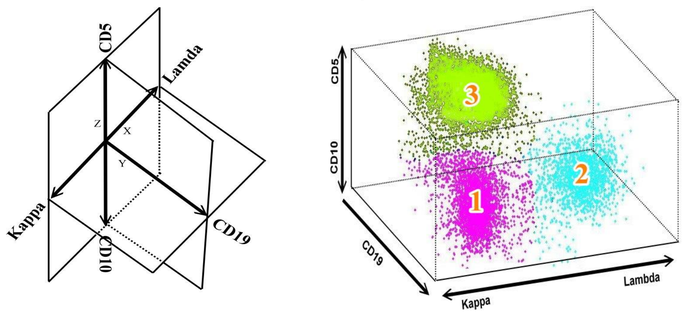
Automatic B Cell Lymphoma Detection Using Flow Cytometry Data

Ming-Chih Shih[1] , Stephen Huang[1], Youli Zu[2]

[1] Department of Computer Science, University of Houston
[2] Department of Pathology, Houston Methodist Hospital

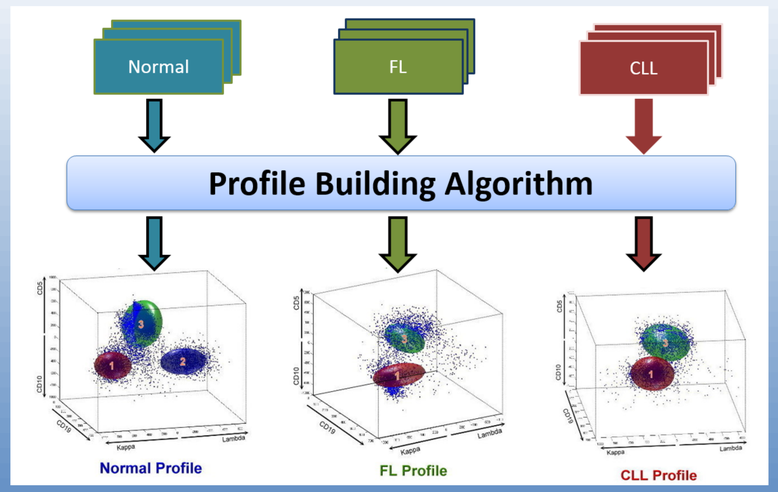
Abstract

Flow cytometry has been widely used for the diagnosis of various hematopoietic diseases. The current diagnosis is still based on manual gating, which is labor-intensive, time-consuming, and subject to human error. The 3D 5-parameter computational model is designed to learn from the domain expert's knowledge and to select and classify the cells automatically. By examining clustering of the cells, the proposed solution is able to distinguish data samples of lymphoma subjects from healthy individuals.

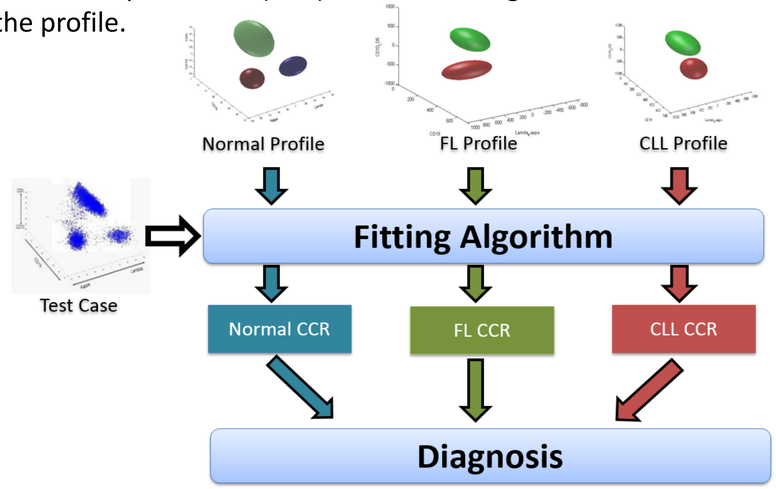


The Methodology

Profiling: After clustering the cells into several clusters, we use ellipsoids that contains most of the cells in the center of the cluster to represent each cluster. There are three profile built in this study: (i) Normal Profile (healthy subjects), (ii) Chronic Lymphocytic Lymphoma (CLL), (iii) Follicular Lymphoma (FL).



Testing/Fitting: Given a profile (i. e., ellipsoids) and a test case, we first align the ellipsoids and the cell clusters together and then compute the percentage of cells captured in the respective ellipsoids. The Cell Capture Rate (CCR) indicates how good the test case matches the profile.



Diagnosis: The process of making of decision.

Input: Cell Capture Rates C_i for i = Normal, FL, CLL.

$$d_i = \sqrt{\sum_{j \neq i} C_j^2}$$ Diagnosis = arg min { d_i | i = Normal, FL, CLL}

Confidence Level: To provide more information of the diagnosis to the physicians, a confidence level is defined to complement the binary (yes/no) result. The confidence level of making one diagnosis of certain profile is defined as

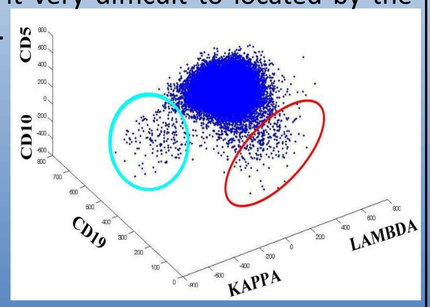
$$Confidence_j = CCR_j / \sum_{j \in D} CCR_j \ .$$

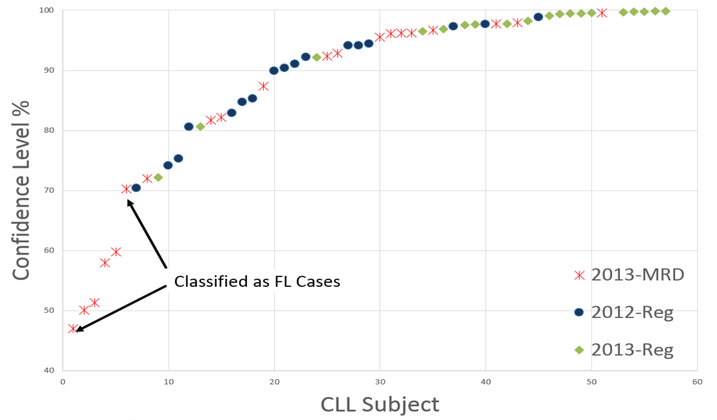
Where j = CLL, FL or Normal. Thus the confidence level of the final diagnosis is defined as

$$Confidence \ = CCR_{Diag} / \sum_{j \in D} CCR_j .$$

Minimal Residual Disease (MRD) is a post-treatment condition, which refers to small numbers of cancerous cells remaining in the body's tissues. The small population (lymphoma cells) is overlapping by its adjacent populations, which makes it very difficult to located by the automated computational approach.

The small amount of lymphoma cells is marked in a red ellipse, and there is also a small cluster of cells marked in a blue circle. The algorithm can correctly identify the small amount of lymphoma (cells in blue circle).

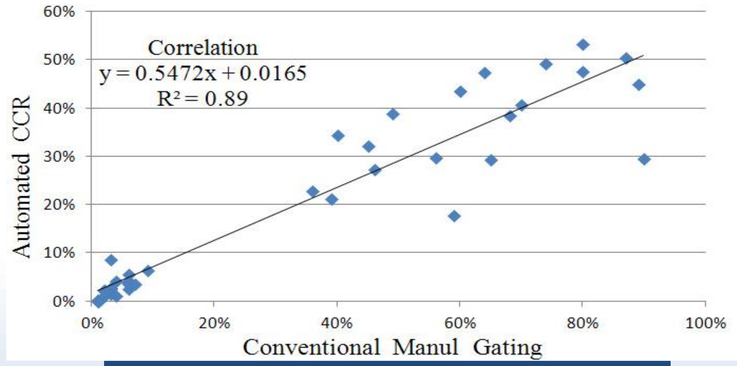




A low confidence level will provide a warning to the physicians about a possible incorrect diagnosis.

Validation

The result shows the lymphoma cell population percentage obtained by the manual method and the Cell Capture Rate (CCR) obtained by the computational method are highly correlated in both the regular and MRD cases.



Summary

There are only 4 error diagnosis among the 107 cases, and 2 of the error cases have a low confidence level.

			Diagnose		
			Normal	B cell lymphoma	
				CLL	FL
Ground Truth	Healthy (37 cases)		36	0	**1**
	B cell lymphoma	CLL Regular (35 cases)	0	35	0
		CLL MRD (20 cases)	0	18	**2**
		FL (15cases)	**1**	0	14