Simulated Detection of Circulating Tumor Cells Using Imaging Flow Cytometry and Machine Learning

Rodrigo Pestana Lopes, Supriya Ramachandran, Tri Le, Woodrow Lomas and Aaron Middlebrook
BD Biosciences, 155 N. McCarthy Blvd, Milpitas, CA 95035

Abstract

Background: Circulating tumor cells (CTCs) are rare cells shed from primary tumors into the bloodstream and are of growing interest in cancer research. However, their low abundance and phenotypic diversity present significant detection challenges. Conventional methods such as next-generation sequencing (NGS) are highly sensitive but costly, while flow cytometry is limited by the absence of universal surface markers.
Objective: To assess whether a novel cell analysis technology that combines spectral flow cytometry with imaging capabilities could enable the detection of cancer cells based on image-derived parameters alone in samples with >99% normal cells.
Method: To simulate CTC detection, whole blood was spiked with cancer cell lines—MCF-7, OVCAR, HT-29, and Jurkat—at frequencies as low as 0.1%. A label-free detection approach was established using intrinsic light scattering and morphological features captured via high-speed imaging flow cytometry on a BD FACSDiscover™ instrument. A neural network classifier was trained to automatically distinguish CTC-like cells from normal cells using only image-derived parameters at a throughput of 10,000 cells/second. Ground truth validation was performed using surface markers and fluorescence detection.
Results: The label-free approach achieved >99% accuracy across all datasets (confirmed by fluorescent flow cytometry), with precision scores of 92.8% (OVCAR), 99.7% (MCF-7), 99.6% (HT-29), and 94.1% (Jurkat).
Conclusion: Unlike traditional flow cytometry, which relies on limited light scatter data (FSC/SSC) and antibody-based staining, this approach requires minimal sample preparation and does not rely on known surface markers. The results demonstrate the potential of high-speed image-based flow analysis for automated, label-free cell classification..

Methods

Sample Prep and data Acquisition
Whole blood was spiked with cancer cell lines that exhibited a range of cell sizes, morphologies and phenotypes. These included: MCF-7, OVACAR, HT-29 and Jurkat. Whole blood was spiked with each cancer cell line at a frequency as low as 1%. A ground truth surface marker was employed for validation. Whole blood was lysed using BD Pharm Lyse™ (Catalogue #555899) and washed. Cell lines (either MCF-7, OVCAR, HT-29 or Jurkat) were harvested and washed. Whole blood was stained with CD45 BV605 or CD45 BV421 (in the case of the Jurkat mix). Cell were then washed two times and then counted. Cell line were stained with a ground truth marker that varied by Cell line: OVCAR (Her- Neu FITC), HT-29 and MCF7 (CD326 Epcam BB515) and Jurkat (CD45 BV605). Cell line were then washed twice after staining and then counted. Sample were prepared according to the spiking chart below. Each sample was prepared at a concentration of 2 million cells per mL in 500 uL. Samples were then acquired on BD FACSDiscover™ S8 Cell Sorter and data was analyzed using FlowJo™ Software.

	LWB Alone	50% Cell Line	20% Cell Line	10% Cell Line	1% Cell Line	0.1% Cell Line
Lysed Whole Blood (LWB)	1.00E+06	5.00E+05	8.00E+05	1.00E+05	1.00E+04	1.00E+03
Cell Line	0.00E+00	5.00E+05	2.00E+05	9.00E+05	9.90E+05	9.99E+05
Total Cells	1.00E+06	1.00E+06	1.00E+06	1.00E+06	1.00E+06	1.00E+06

A novel neural network architecture was designed to predict the probability of an event belonging to a Circulating Tumor Cell class based on Imaging Features derived from the BD FACSDiscover™ S8 Instrument. A generalized network architecture was designed and demonstrated to be applicable to detecting 4 Circulating Tumor Cell types including Ovarian, Breast, Colon and T-cell cancer types. Hyper-parameter tuning was performed, and network architecture decisions were applied to a generalized architecture, capturing all 4 CTC types.

Performance evaluation visualized through Confusion matrices demonstrate high Recall and high accuracy of Circulating Tumor Cell identification using label-free morphological features.

Results

1) Quantifying Circulating Tumor Cells Spiked into Whole Blood Samples

A series of standard cleanup gates was applied across each of the four heterogenous sample preps to exclude doublets.

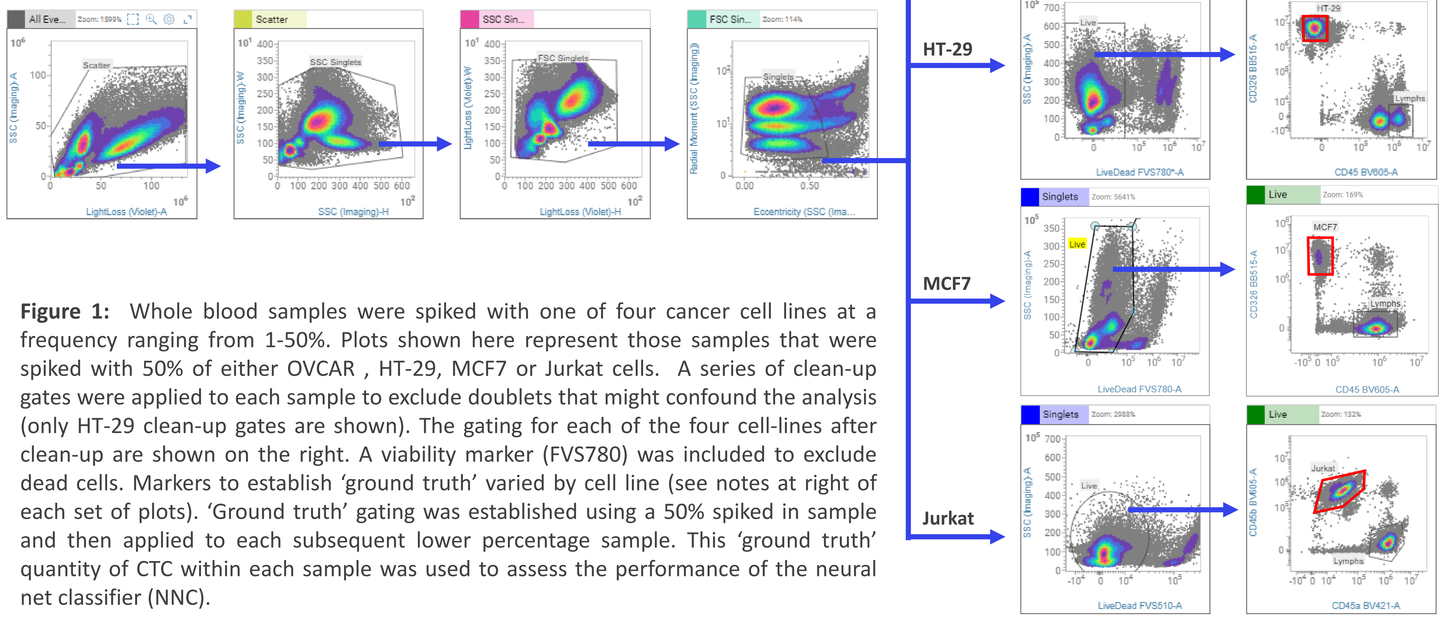


Figure 1: Whole blood samples were spiked with one of four cancer cell lines at a frequency ranging from 1-50%. Plots shown here represent those samples that were spiked with 50% of either OVCAR , HT-29, MCF7 or Jurkat cells. A series of clean-up gates were applied to each sample to exclude doublets that might confound the analysis (only HT-29 clean-up gates are shown). The gating for each of the four cell-lines after clean-up are shown on the right. A viability marker (FVS780) was included to exclude dead cells. Markers to establish 'ground truth' varied by cell line (see notes at right of each set of plots). 'Ground truth' gating was established using a 50% spiked in sample and then applied to each subsequent lower percentage sample. This 'ground truth' quantity of CTC within each sample was used to assess the performance of the neural net classifier (NNC).

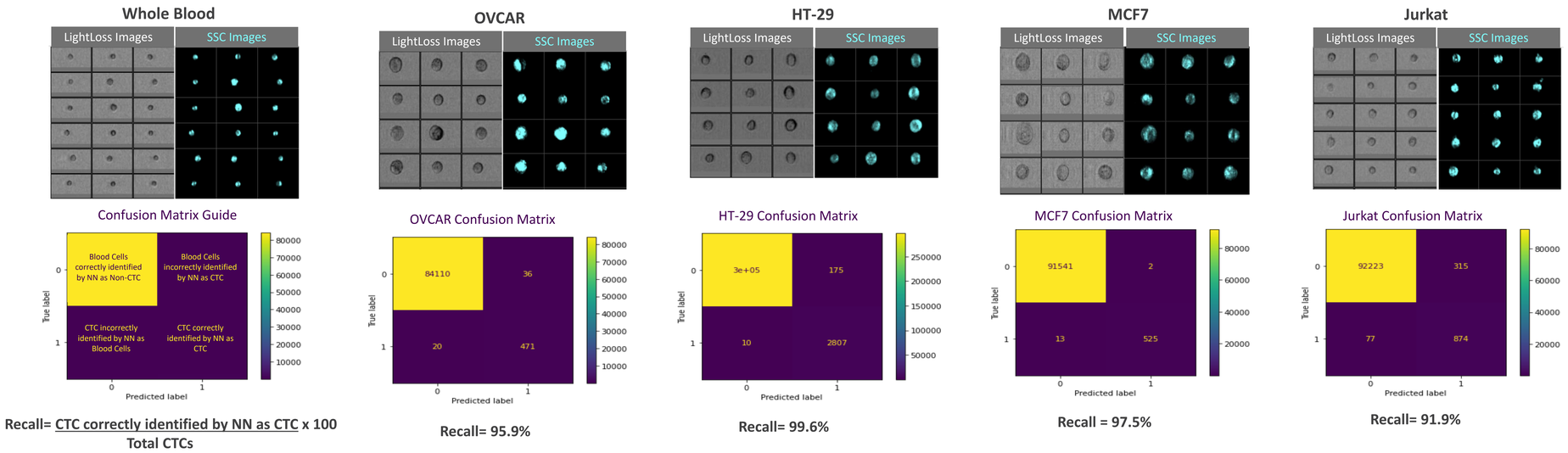
OVCAR Cells, an ovarian adenocarcinoma cell line, were spiked into a whole blood sample at a ratio of approximately 1:1. **OVCAR** cells express Her2-Neu. A plot of CD45 (expressed by blood cells) vs Her2-Neu (Expressed by the **OVCAR** cells) was used to identify and quantify **OVCAR** cell content in the sample. This **gate** established the 'ground truth' to be compared against the neural net classifier (NNC).

HT-29 Cells, a colorectal adenocarcinoma cell line, spiked into a whole blood sample at a ratio of approximately 1:1. **HT-29** cells express CD326 (Epcam). A plot of CD45 (expressed by blood cells) vs CD326 (Expressed by the **HT-29** cells) was used to identify and quantify **HT-29** cell content in the sample. This **gate** established the 'ground truth' to be compared against the NNC.

MCF7 Cells, a mammary adenocarcinoma cell line, were spiked into a whole blood sample prep at a ratio of approximately 1:1. MCF7 cells express CD326 (Epcam). A plot of CD45 (expressed by blood cells) vs CD326 (Expressed by the **MCF7** cells) was used to identify and quantify **MCF7** cell content in the sample. This **gate** established the 'ground truth' to be compared against the NNC.

Jurkat Cells, a T-cell leukemia cell line, were spiked into a whole blood sample at a ratio of approximately 1:1. Both Jurkat and blood cells express CD45. To identify and quantify spiked Jurkat cells, they were stained with CD45 BV605 while blood cells were stained with CD45 BV421 prior to mixing the cells. A plot of CD45a (BV421) vs CD45b (BV605) was used to identify and quantify **Jurkat** cell content in the sample. This **gate** established the 'ground truth' to be compared against the NNC.

2) Circulating Tumor Cell Types detected by the novel Generalized Neural Network Architecture



Recall= CTC correctly identified by NN as CTC x 100 / Total CTCs

Recall= 95.9%

Recall= 99.6%

Recall = 97.5%

Recall= 91.9%

Figure 2: Top panel, from left, clockwise, shows LightLoss and Side scatter images of Whole Blood and each CTC type OVCAR (Ovarian Cancer), MCF-7 (Breast Cancer), HT-29 (Colon Cancer) and Jurkat (T-Cell cancer) respectively. Bottom panel, shows corresponding Confusion Matrices computed based on the performance of a Novel Neural Network Architecture applied towards identifying each CTC type spiked into a whole blood sample at 1% class frequency. The confusion matrices demonstrate that a generalized neural network architecture can detect 4 types of CTCs with Recall scores of of 95.9% (OVCAR), 99.6% (MCF-7). 97.9% (HT-29) and 91.9% (Jurkat) respectively.

Conclusions

1) Label-free detection method: The study introduces a novel label-free detection method leveraging intrinsic light scattering properties of CTCs, enabling both enumeration and isolation without the need for known surface marker phenotypes.
2) High accuracy and recall: The neural-net based classifier demonstrated high accuracy (>99%) and high recall scores for various cancer cell lines, indicating its effectiveness in distinguishing CTCs from normal cells in heterogeneous samples.
3) Broad applicability: This detection method requires minimal sample preparation and is applicable to a wide range of cancer types, making it a versatile tool for CTC detection in cancer research.