

Original Article

Microfluidic droplets content classification and analysis through convolutional neural networks in a liquid biopsy workflow

Gabriele Soldati^{1*}, Fabio Del Ben^{2*}, Giulia Brisotto², Eva Biscontin², Michela Bulfoni³, Aigars Piruska⁴, Agostino Steffan², Matteo Turetta^{3*}, Vincenzo Della Mea^{1*}

¹Department of Mathematics, Computer Science and Physics, University of Udine, Italy; ²Immunopathology and Cancer Biomarkers, CRO Aviano National Cancer Institute, Aviano, Italy; ³M.T., M.B., Department of Medicine, University of Udine, Italy; ⁴A.P. Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands. *Equal contributors.

Received August 2, 2018; Accepted November 7, 2018; Epub December 15, 2018; Published December 30, 2018

Abstract: In a recent paper we presented an innovative method of liquid biopsy, for the detection of circulating tumor cells (CTC) in the peripheral blood. Using microfluidics, CTC are individually encapsulated in water-in-oil droplets and selected by their increased rate of extracellular acidification (ECAR). During the analysis, empty or debris-containing droplets are discarded manually by screening images of positive droplets, increasing the operator-dependency and time-consumption of the assay. In this work, we addressed the limitations of the current method integrating computer vision techniques in the analysis. We implemented an automatic classification of droplets using convolutional neural networks, correctly classifying more than 96% of droplets. A second limitation of the technique is that ECAR is computed using an average droplet volume, without considering small variations in extracellular volume which can occur due to the normal variability in the size of the droplets or cells. Here, with the use of neural networks for object detection, we segmented the images of droplets and cells to measure their relative volumes, correcting over- or under-estimation of ECAR, which was present up to 20%. Finally, we evaluated whether droplet images contained additional information. We preliminarily gave a proof-of-concept demonstration showing that white blood cells expression of CD45 can be predicted with 82.9% accuracy, based on bright-field cell images alone. Then, we applied the method to classify acid droplets as coming from metastatic breast cancer patients or healthy donors, obtaining an accuracy of 90.2%.

Keywords: Cancer, liquid biopsy, circulating tumor cells, neural networks, machine learning

Introduction

As of today, the choice of a cancer treatment is usually based on histopathological and molecular features of the primary tumor, obtained analyzing bioptic specimens. Core biopsy is an invasive surgical procedure, it is rarely repeatable, and sometimes has limitations on accessing tumors in delicate or hard-to-reach organs (e.g.: brain, lung). Tumor cells shed from primary tumors or metastatic sites in the peripheral blood (called circulating tumor cells (CTC)) can be sampled by a “liquid biopsy” with a minimally invasive venous puncture. This minimal invasivity enables clinical applications otherwise impossible with core biopsy, such as serial

monitoring of the disease status and frequent re-assessment of the mutational profile of cancer cells. There have been several studies proving the prognostic value of the number of CTC in metastatic breast [1, 2], colorectal [3] and prostate cancer [4], which eventually led to FDA approval of CellSearch[®] as an in-vitro diagnostic device. More recently, CTC isolation and molecular characterization proved to be able to drive therapy selection in metastatic prostate cancer [5]. The actual trend in oncology is focused on the personalization of targeted therapies to tailor the best treatment for the individual patient; isolated CTC are currently seen as an excellent biological sample for the molecular characterization of the disease, the study

Neural networks for microfluidic droplet classification

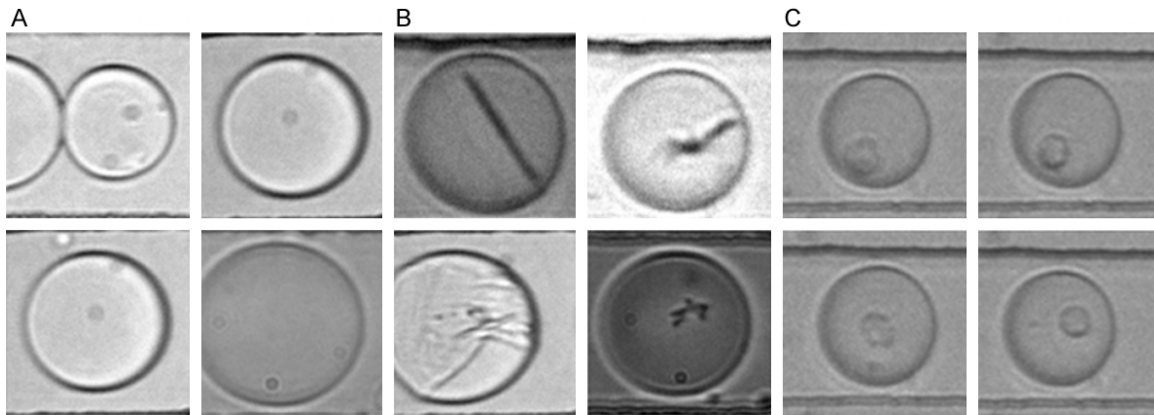


Figure 1. Overview of different types of droplets. Representative images for each class of events, including (A) empty drops (false positive); (B) Debris-containing drops (false positive); (C) Cell-containing drops (true positive).

of tumor heterogeneity and the monitoring of disease evolution. Although there is enough evidence supporting CTC value in cancer care, their clinical utility, especially in terms of survival benefit, has yet to be strongly demonstrated, thus preventing their widespread use in clinical practice. This is mainly due to limitations of current devices for CTC detection: the extreme rarity of CTC (one in a billion blood cells) constitutes a big technological challenge, and devices for sensitive and pure CTC isolation are needed.

We recently published a method for the detection of CTC based on the abnormal tumor metabolism, as an alternative to current size-based and immunolabeling-based methods [6]. The abnormal metabolism is a widespread characteristic of tumors, and it is part of a recent list of the hallmarks of cancer [7]. Our method exploits the increased extracellular acidification rate (ECAR) of cancer cells with respect to white blood cells. The measurement of ECAR is performed at the single-cell level, using a droplet screening system similar to the one published by Mazutis et al. [8]: cells from a blood sample are individually encapsulated in pico-liter drops containing culture medium and a pH-responsive fluorescent dye. After a short incubation, droplets are re-injected in a microfluidic device, and the pH of each drop is measured optically, looking for the droplets with lower pH. Currently, this workflow has the following limitations:

1) During a typical sample analysis, we observe “false positives” in a range of 1-100 per million droplets, which have been identified by means

of a triggered camera as empty droplets and droplets containing debris (**Figure 1**). To filter them out, a human operator must typically run through the acquired pictures, discarding *these* false positives. This affects the overall process, adding a time-consuming and operator-dependent image-screening step.

2) The droplets usually have a negligible variation in size, thanks to microfluidics generating a monodisperse droplet population. However, a significant variation in droplet volume might occur, though rarely. On the other hand, cells often display significant variations in size. *The variation in the size* of both droplets and cells affect *the* total extracellular volume, leading to a decrease in the accuracy of ECAR calculation, which is instead based on an average extracellular volume.

3) Pictures are *acquired* for each droplet with increased ECAR, but no further analysis is performed on the image of the cell, potentially missing significant information.

These problems belong to the class of image classification (1), object detection, and segmentation (2-3). Machine learning methods have been recently used to solve similar problems. Among all machine learning methods, neural networks, and more specifically convolutional neural networks (CNN) recently showed unprecedented performances in tasks such as human-grade image classification [9], object detection and segmentation in complex images [10, 11].

In this work, we describe how we applied CNN to meet the following objectives:

Neural networks for microfluidic droplet classification

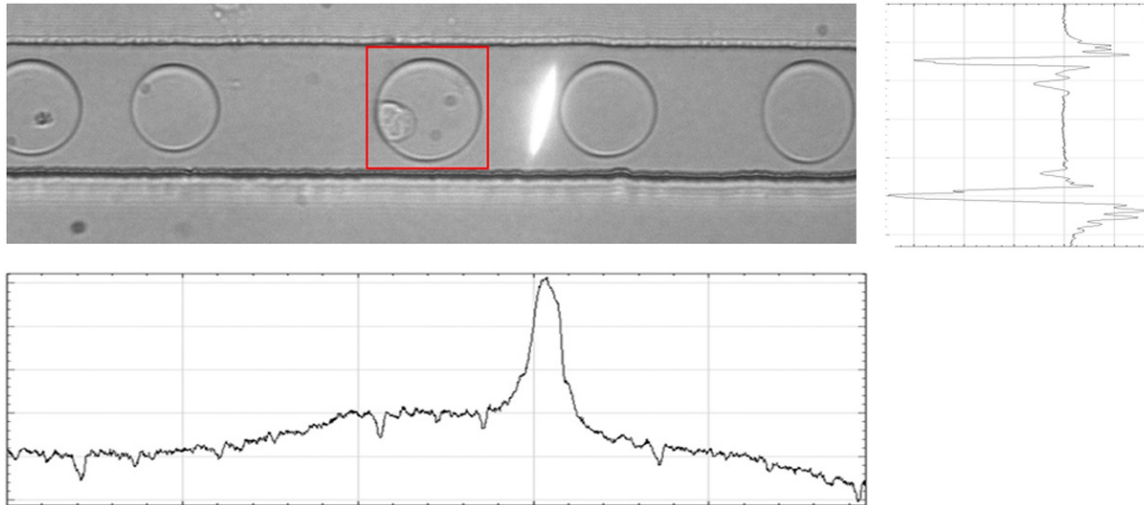


Figure 2. Picture of a droplet flowing in the microfluidic channel. The white laser slit is visible next to the drop of interest, which is surrounded by a red square, indicating the crop applied to the droplet. Below the image, the vertical projection of intensity is shown, which has a peak corresponding to the laser beam. At the right, the horizontal projection of the same image is shown. The two minima values corresponding to the walls of the capillary.

Classify automatically the content of droplets; Correct automatically the measurement of pH taking into account both droplet and cellular volume; Segment cellular image, on which we performed exploratory data mining using unsupervised machine learning methods.

Materials and methods

Image acquisition

Pictures (8-bit, 1216×250) were acquired during the metabolic assay described in a previous work [6] with an USB 3.0 monochromatic CMOS camera (PointGrey Grasshopper) synchronized with a monochromatic LED (Roithner LaserTechnik). Light was captured by a $40 \times$ objective (Olympus). Events with lower pH, according to a user-defined threshold, activate the trigger. To ensure real-time operation, we used an embedded Linux Real-Time controller (NI CompactRIO) with an on-board field-programmable gate array (FPGA) delivering TTL-trigger to both the camera and the LED. The controller was operated by customized software (NI LabVIEW). Collection of blood samples was approved by the institutional review board with n. IRB-12-2014.

Pre-processing

The main sources of image variability were the spatial positioning of the microfluidic channel, and the variations in brightness and contrast

of the picture. Furthermore, pictures acquired by the system were composed of multiple droplets in the channel, with the droplet of interest next to the laser beam used to excite droplet fluorescence and measure pH. This constituted an issue for the training of the neural networks, in fact, a first training on an unprocessed data set led to overfitting and failure of the model on the test set. We pre-processed the datasets performing normalization, contrast-enhancement, using homonymous functions in ImageJ, and cropping the images on the droplet of interest. To crop the image, a horizontal projection of the image was used to find the coordinates of the microfluidic channel, while a vertical projection was used to find the position of the laser beam. The projection was analyzed to find the maxima and minima values, corresponding to the light slit in the vertical projection and the borders of the microfluidic channel in the horizontal projection (**Figure 2**). Once the coordinates of the capillary walls and the laser beam were found, it was possible to create a box around the drop of interest, highlighted by the red square in **Figure 2**. The distance between the laser beam and the drop was consistently the same, because flows were steady and the trigger actioned by a nanosecond-delay FPGA controller, so it was sufficient to manually set the value of the rectangle and to apply it to every image in a single experiment batch. For image processing not

Neural networks for microfluidic droplet classification

Table 1. Comparison of different models of neural networks trained for 100 epochs

Model Name	Classification Accuracy	Recall	Specificity	Precision
ResNeXt				
<i>Augmented</i>	89.38% (+4.32%)	0.80 (+0.09)	0.99 (-0.01)	0.99 (=)
<i>Original</i>	85.06%	0.71	1.00	0.99
<i>ConvNet</i>	87.50% (+2.53%)	0.84 (-0.03)	0.91 (+0.08)	0.90 (+0.07)
	84.97%	0.87	0.83	0.83
<i>CNN8</i>	85.16% (+0.38%)	0.82 (+0.03)	0.88 (-0.02)	0.88 (-0.01)
	84.78%	0.79	0.90	0.89
<i>ResNet-v2</i>	82.78% (-1.35%)	0.76 (-0.02)	0.9 (=)	0.88 (-0.01)
	84.13%	0.78	0.90	0.89
<i>CNN6</i>	82.00% (+1.50%)	0.82 (+0.12)	0.821 (-0.09)	0.82 (-0.06)
	80.50%	0.70	0.9	0.88
<i>GoogLeNet</i>	81.38% (+1.35%)	0.70 (-0.15)	0.92 (+0.17)	0.90 (+0.13)
	80.03%	0.85	0.75	0.77
<i>AlexNet</i>	80.66% (+1.28%)	0.69 (-0.13)	0.93 (+0.16)	0.90 (+0.13)
	79.38%	0.82	0.77	0.78
<i>CNN2</i>	68.22% (-11.03%)	0.82 (+0.07)	0.55 (-0.29)	0.64 (-0.18)
	79.25%	0.75	0.84	0.82
Extended training - number of epochs ⁿ - Augmented dataset				
<i>ResNeXt¹⁵⁰</i>	90.03%	0.81	0.99	0.99
<i>ConvNet⁷⁵</i>	90.00%	0.84	0.96	0.96
<i>Inception v3²⁰⁰</i>	88.75%	0.78	0.99	0.99
<i>ConvNet¹⁵⁰</i>	87.66%	0.77	0.99	0.98
Object detection algorithm - Augmented dataset				
<i>Combined</i>	96.00%	0.94	0.98	0.98
<i>MobileNet</i>	89.84%	0.81	0.99	0.99
<i>Inception-v2</i>	86.41%	0.77	0.96	0.95

Results using Augmented and Original dataset are shown in two lines, (variation in performances shown in parenthesis). ResNeXt and ConvNet were also trained for “n” different number of epochs, indicated as e.g.: ResNeXtⁿ. Inception v3 indicates the attempt of retraining a publicly available network. “Combined” indicates a combination of MobileNet and Inception-v2.

related to the classification task we used NIH ImageJ [12].

Convolutional neural networks

We decided to classify the images in two classes: “Empty/Debris”, if the droplet was empty or containing junk and “Cell”, if the droplet contained one or more cells. In a first attempt, we tried to re-train a publicly-available pre-trained model. Then, we tried to train a series of neural networks from the start. Lastly, we compared two object-detection networks to detect the presence of cells inside the image.

The publicly-available model used was GoogLeNet Inception v3, trained for the ImageNet Large Visual Recognition Challenge of 2012 [13]. The framework used to re-train the model was Google TensorFlow. The implementation was made using Tflearn, a high-level/abstraction layer for TensorFlow. Tflearn offers a series of ready-to-use models, which had been modified to accept the greyscale images of the dataset.

The networks implemented from the start were Alexnet [14], GoogLeNet (Inception-v1) [15], Inception-ResNet-v2 [16], ResNeXt [17], ConvNet [18] and three other different custom CNNs. These custom networks consisted in a series of convolutional layers followed by a pooling layer and a fully connected layer on top. The first had 2 convolutional layers (CNN2), the second had 6 (CNN6), and the last one had 8 (CNN8).

The training of the models was controlled by a

Python script where we could set all the variables to control the learning process. We set the learning rate to 1e-3, and the dimension of the images to 50 × 50 pixels.

In order to perform image segmentation and subsequent analysis, we trained two different object detection networks: MobileNet [19] and Inception-v2 [20]. We manually labelled 500 images, taken randomly from the training set, using a program called Labellmg. The two models were trained on the same dataset for approximately the same amount of time. The training-set consisted in 90% of labelled images, while the remaining 10% was used as a

Table 2. Confusion matrix of results obtained with ResNeXt model

		Predicted Class		
		Cell	Empty/debris	Total
Actual Class	Cell	1129	471	1600
	Empty/debris	7	1593	1600
Total		1136	2064	3200

validation-set. In addition to the image segmentation, we evaluated the performance of the object-detection algorithm in classifying droplets as “Empty/Debris” and “Cell”, comparing it with previously trained networks.

Datasets

The “training set” was composed by 10698 pictures taken from 65 different acquisitions, composed of 1575 pictures from MDA-MB-231 cancer cell-line, 1485 from healthy controls, and 7638 from metastatic breast cancer patients. The “validation set” was set as 5% of the “training set”. The “test set” was a completely different set of images, taken from 13 separate acquisitions of metastatic breast cancer samples, manually labeled, and not used in the training set. The “test set” was composed using a series of 3200 images, half of which were empty/debris and half containing cells.

Data augmentation

In order to increment the number of images of the dataset and try to reduce the problem of the overfitting, we applied a data-augmentation process to obtain an “augmented” training set. We used Python libraries and built-in TensorFlow functions to flip both vertically and horizontally images, and perform random transformations, like modifying contrast, blur, and rotations. The “augmented” training set was eventually composed of 42792 images.

Increasing the accuracy of metabolism measurement

To provide a more accurate measurement of proton production by the cell, we evaluated the possibility of automatically determine the extracellular volume of droplets. Object segmentation, provided by object-detection algorithm, enabled easy cropping and measurement of

the picture area by ImageJ. The square root of such area corresponded to the diameter of the drop/cell. Using the diameter, we computed the object volume assuming sphericity. Extracellular volume was then computed as the difference between the droplet volume and the cellular volume. With the extracellular volume and pH value we could compute the number of protons in solution.

Characterization of white blood cells

Blood samples were obtained from blood donors with informed consent. 1 mL of whole blood was processed by lysing red blood cells and staining for CD45-BV421 (BD-Biosciences). White blood cells (WBC) were re-suspended in 50 uL of Joklik’s EMEM, 0.1% Bovin Serum Albumin and 4 uM SNARF-5F, emulsified and re-injected for droplet analysis and image collection according to the method described in [6]. Flow cytometry data were collected using BD FACS Canto™.

Characterization of droplets content

The pictures of cells segmented with the object-detection algorithm were collected and normalized, dividing the gray value of each pixel by the mean gray value of the respective image.

We used NIH ImageJ to perform the following measurements on each cell picture: area, perimeter, gray values (mean, modal, median, min, max, standard deviation), centroid, center of mass, fit ellipse (major and minor axis, angle), integrated density, skewness and kurtosis. We performed texture analysis with gray-level co-occurrence matrix using the GLCM plugin for ImageJ, obtaining the following parameters: angular second moment, contrast, correlation, inverse difference moment, entropy.

For data analysis we used *Orange*, an open-source data visualization, machine learning, and data mining toolkit [21]. *Orange* is based on open-source Python libraries for scientific computing such as *NumPy* [22], *SciPy* [23], and *Scikit-learn* [24]. Classification accuracy is an average of 10-fold cross-validation. Naïve Bayes, logistic regression, and multilayer perceptron algorithms were employed for the classification of cell images, as described in the *Orange Canvas* documentation, with default settings.

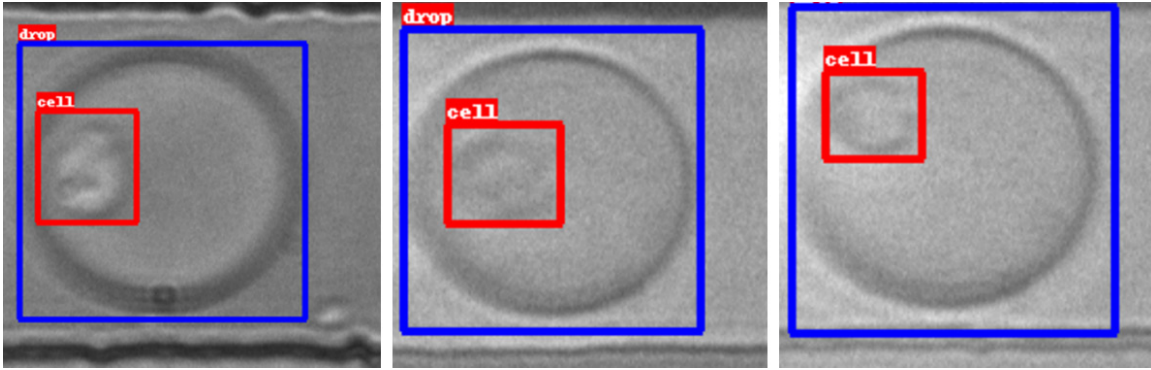


Figure 3. Object detection. Representative pictures of droplets (blue squares) and cells (red squares) detected and segmented with the object-detection neural network.

Table 3. Confusion matrix of results obtained with MobileNet + Inception-v2 model

		Predicted Class		
		Cell	Empty/debris	Total
Actual Class	Cell	1511	89	1600
	Empty/debris	36	1564	1600
Total		1547	1653	3200

Results

Classification of droplets content

We made a first attempt of classification by re-training Inception v3 model, in order to test the approach with the lesser degree of customization. The model displayed a classification accuracy of 88.75% with 200 epochs of training.

Successively, we tried training eight different networks from the start in the attempt to improve classification accuracy. We trained all models for 100 epochs with an initial learning rate of 1e-3, using the non-augmented dataset.

ResNeXt showed the best accuracy (85.06%), but showed a trend towards the classification of droplets as “Empty/Debris”: only 0.4% of truly empty/debris drops were misclassified, while a significant number of cell-containing droplets (29.4%) was misclassified, as shown in **Table 1**. Data augmentation increased considerably the accuracy of ResNeXt and other networks, up to 89.38% (**Table 2**). A third training was done on networks yielding the best results, ResNeXt and Convnet, using a different augmented dataset, in which all the images were flipped on both the horizontal and the vertical axes, while we applied the other augmenta-

tions at random before training the networks. This increased by four times the size of the training set. After extending the training duration to 150 epochs, ResNeXt reached an overall accuracy of 90.03%, while Convnet raised to 87.66%. Misclassified cell-containing droplets were reduced to 19%. An interesting observation was that the trend of the validation loss of the Convnet model at around 50k iterations was slightly lower than the loss of ResNeXt. So we decided to try “early stopping” Convnet training to avoid overfitting. We implemented it setting the training epochs to 75 instead of 150, leaving all other parameters unchanged. The model performed slightly better, reaching an accuracy of 90.00%. Since CTC are rare events, usually found in a number of 1-10 per mL of blood, the loss of around 19% of cell-containing droplets due to misclassification was not acceptable. In the attempt of reducing such loss, we tried an alternative approach, using two object-detection algorithms, MobileNet and Inception-v2. The models were trained to identify the presence of cells and droplets in an image (**Figure 3**). If at least one cell was detected with a percentage of certainty above 50%, then the image was classified as cell-containing. These methods gave results comparable to ResNeXt. By examining the confusion matrix, we noticed that the images misclassified by the two models belonged to different acquisitions, so we hypothesized that combining the two networks could improve the sensitivity. The combination of the two models, in fact, gave significantly better results, with the recall increased from 0.80 to 0.94 and the overall accuracy raised to 96%. The most important achievement was the strong reduction of misclassified cell-containing drops

Table 4. Variations in the extracellular volume as computed for the three representative droplets examined

	Total drop volume (pL)	Variations wrt average	Extracellular volume	Variations wrt average
Drop 1	52.0	-16%	48.7	-21%
Drop 2	60.1	-2%	57.4	-7%
Drop 3	72.3	+17%	70.6	+14%
Average	61.7		58.9	

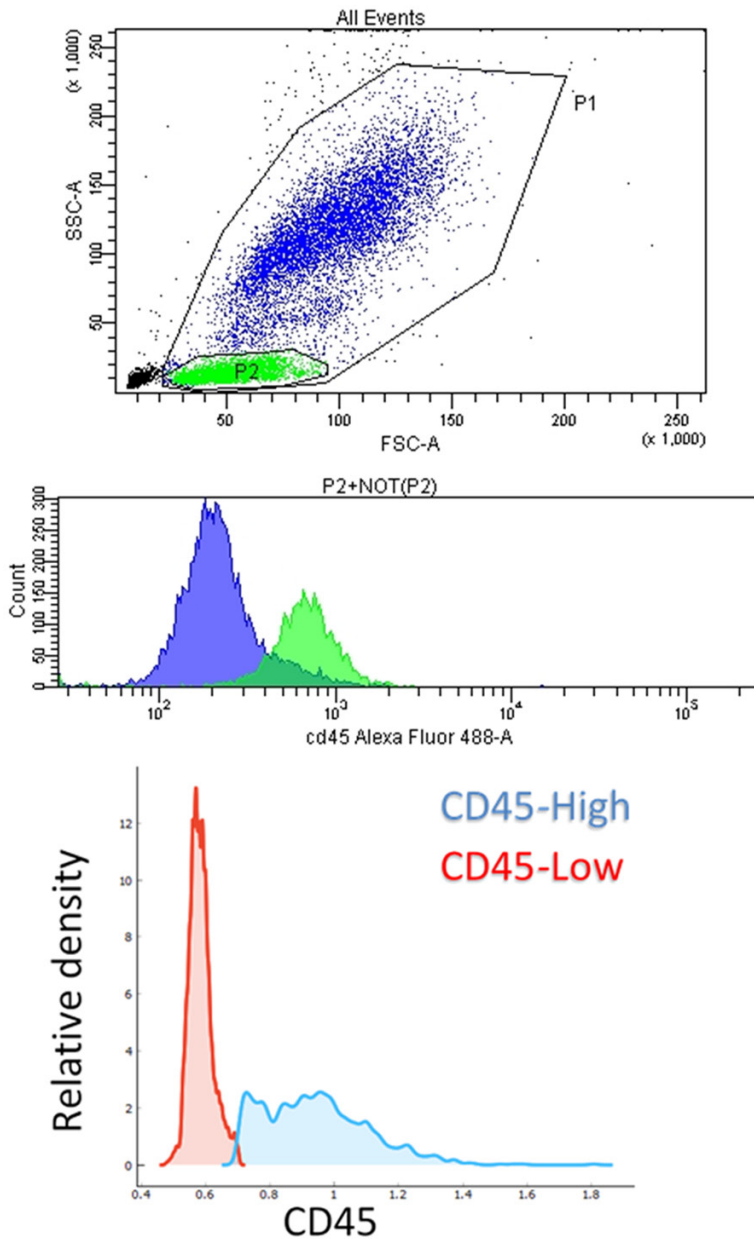


Figure 4. Flow cytometry. Top: dot-plot of a typical healthy blood lysate obtained with flow cytometry, in which P1 population (WBC), P2 (lymphocytes) in green, and NOT(P2) (granulocytes and monocytes) in blue, are visible. Middle: histogram of fluorescence intensity corresponding to CD45 expression for the described populations. Bottom: relative density of cell populations with high- and low-CD45 expression in a healthy sample.

down to 5.5%. Results given by combined networks are shown in **Table 3**.

Increasing accuracy of metabolism measurement

As explained in the introduction, the estimation of single-cell ECAR is obtained by measuring the pH of the whole droplet. This does not take into account small variations in volumes due to different size of droplets or cells. The segmentation of droplets and cells with object-detection networks enabled the measurement of their volumes, thus allowing for correction of the ECAR estimation. As a quantitative example, we report the case of three droplets with nearly identical pH (6.75 ± 0.02), coming from three different experiments (**Figure 3**).

In **Table 4** we report the volumes of the droplets displayed in **Figure 3**, their total volume, and their extracellular volume, normalized with respect to the average volumes, respectively.

Table 4 shows that between these three droplets, the extracellular volume can vary up from -21% to +14% of the average, leading to a proportional over- or under-estimation of ECAR if based only on pH measurement. The combination of the two measurements (pH and extracellular volume), instead, can significantly increase the accuracy of the assay.

Content characterization-proof of concept on white blood cells

We used the segmented images of cells obtained with

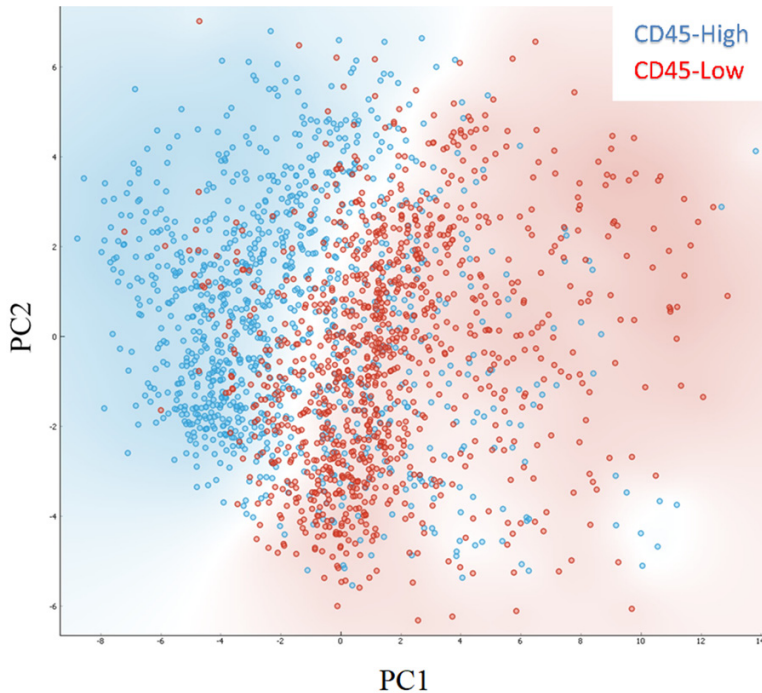


Figure 5. PCA analysis of WBC images. Dot plots showing principal components (PC1, PC2) obtained performing PCA on white blood cell images. In red, cells with CD45-low expression. In blue, cells with CD45-high expression. Background is coloured according to class density.

Table 5. Comparison of different models in the classification of CD45 expression using pictures of cells

Method	Classification Accuracy	Area under ROC curve
Neural Network	82.9%	0.871
Naïve Bayes	77.0%	0.826
Logistic Regression	77.7%	0.812

the object-detection algorithm to further characterize the content of droplets. A proper characterization was complicated by the fact that the images of the cells, acquired in bright-field microscopy, were often out-of-focus, as cells freely float inside the droplets, which in turn are pushed through the channel in setting similar to flow cytometry. For these reasons, images are slightly blurred even with optimal light intensity and exposure time. We tested the hypothesis that it was possible to discriminate between different cellular types using the collected images, by ideating a proof of concept experiment, in which we used WBC from the blood of a healthy donor. We stained cells for CD45 and subsequently collected pic-

tures of CD45-high and CD45-low populations, corresponding to neutrophils and lympho-monocyte, respectively (Figure 4).

We performed a Principal Component Analysis (PCA) with 2 components, covering approximately two thirds of variance, on the parameters obtained with image analysis. Results showed that the two populations (CD45-high and -low) trended towards the formation of two separate clusters (Figure 5).

Supervised analysis using naïve Bayes and logistic regression yielded 77.0% and 77.7% classification accuracy, respectively (Table 5). The employment of a neural network, a multi-layer perceptron algorithm capable of learning both linear and non-linear models, increased classifica-

tion accuracy to 82.9%. The multilayer perceptron was composed by three layers of 20, 10 and 2 nodes and iterated for 200 times. Results were subject to 10-folds cross-validation.

The analysis of frequency distributions showed that the variables related to the gray levels were identical between the two groups, while variables related to the cell size and texture were different. Specifically, CD45-high cells had lower area and inverse difference moment, and minor variations in the distributions of contrast, correlation, entropy and angular second moment.

Droplet content characterization-exploratory analysis on metastatic breast cancer patients vs healthy donors

Finally, we used the same method to analyze droplet content of metastatic breast cancer patients and healthy donors. We selected CD45-negative cells, in droplets with pH lower than 7.0. We performed a preliminary normalization of images, because the pictures were coming from different acquisitions in different days. Normalization was implemented by divid-

Neural networks for microfluidic droplet classification

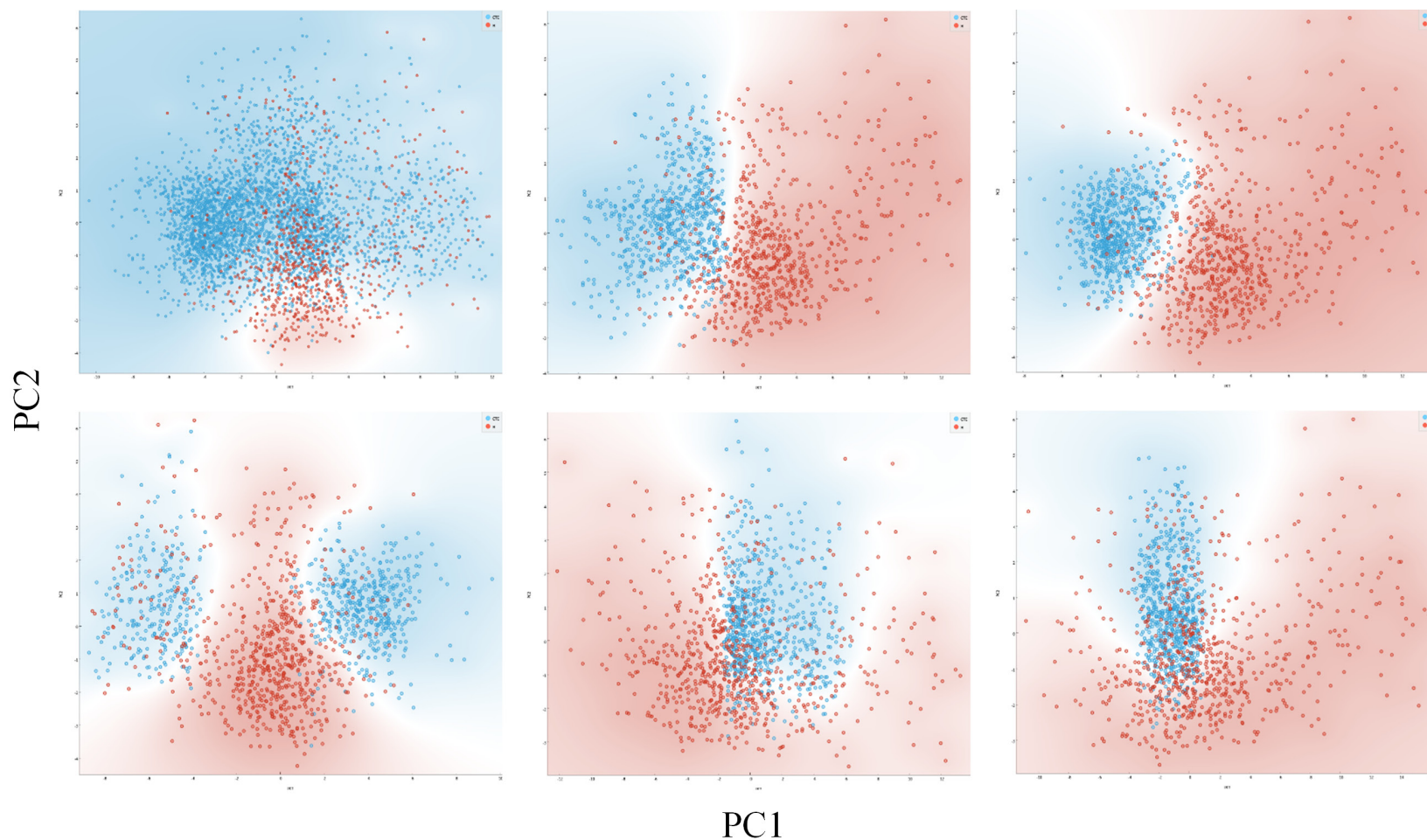


Figure 6. PCA analysis of cancer cells and WBC. Dot plots showing principal components of cells from metastatic cancer patients (blue) or healthy controls (red). Top-right graph shows the whole dataset. Images from cancer patients are much more than false positives from healthy donors, thus to generate a more balanced visualization we obtained five equal-size datasets by sampling the same number of images from cancer patients or healthy controls. The other graphs show these five balanced datasets. Background is colored according to class density.

Neural networks for microfluidic droplet classification

Table 6. Comparison of different models in the classification of cells as coming from cancer patients or healthy donors based on the pictures of cells

Method	Classification Accuracy	Area under ROC curve
<i>Neural Network</i>	90.2%	0.947
<i>Naïve Bayes</i>	83.6%	0.907
<i>Logistic Regression</i>	81.4%	0.859

ing the gray value of each pixel by the average gray value of each image. The dataset of pictures was unbalanced because there were about 5-folds more cells coming from patients than cells coming from healthy controls, so we created 5 balanced batches, in which the size of the two classes were equal. For each batch, we performed a PCA analysis and plotted the principal components with the best cluster separation (**Figure 6**). In this setting the classification task using naïve Bayes, logistic regression and the multilayer perceptron yielded a classification accuracy (averaged between 5-subsets) of 83.6%, 81.4%, and 90.2%, with an area under curve of 0.907, 0.859, and 0.947, respectively (**Table 6**). The analysis of frequency distributions showed that the variables related to the gray levels were identical between the two groups, while variables related to cell size and texture were different.

Specifically, in cells from metastatic patients, angular second moment was higher, contrast was higher and with increased variance, while entropy was lower, and area was lower on average but with higher variance (**Figure 7**). Average areas corresponded to diameters of 11.18 ± 6.45 micrometers in patients versus 12.18 ± 6.18 micrometers in healthy controls.

Discussion

In this study we implemented an accurate automatic classification of droplet content by combining multiple machine learning approaches. CTC are rare events, thus achieving a misclassification rate as low as possible is an important objective. Even if the achieved misclassification of cell-containing events (5%) is acceptable, it might be worth exploring the possibility of decreasing even further such percentage by creating a third category of “noisy” or “doubtful” pictures, containing the droplets in

which the object-detection algorithm detects a cell with suboptimal probability. Such “doubtful” pictures could be then manually checked by an operator, or used to re-train the algorithms. It is important to note that some pictures of the database are noisy and difficult to classify even by an experienced operator. Such pictures are present in a proportion comparable to the 5% of misclassified pictures.

By analyzing cell images, we successfully classified different populations of white blood cells, and specifically cells having high- or low-CD45 expression. It is worth noticing that CD45 is only partially able to divide WBC into morphologically homogeneous categories: CD45-low cells are mostly granulocytes, but to a lesser extent also lymphocytes, while CD45-high cells are mostly lymphocytes, but also some monocytes. Granulocytes are somewhat morphologically homogeneous, while lymphocytes and monocytes differ significantly. The overall value of classification accuracy shows that distinction between granulocytes and lymphomonocytes by bright-field images is possible. However, in this specific case, the performance of image-based classification might be underestimated by the heterogeneous composition of WBC population grouped by the same CD45 expression. A better result might be obtained by employing forward and side scatter to fully separate WBC population and train the algorithm on more homogeneous populations, but such optimization of performance was out of the aim of the presented proof-of-concept. The concept of detecting biological characteristics by applying machine learning algorithm to label-free images of cells has been already proven by other groups [25-27]. However, no reports were published on the discrimination between CTC and WBC. We found that CD45-negative, hyper-metabolic cells coming from patients could be discriminated from the same cells found in healthy controls, with around 90% accuracy. This suggests that such cells are phenotypically different from WBC which are normally found in peripheral blood, compatibly with being CTC or rare CD45-negative WBC with unusual phenotype. These cells might be potential biomarkers, and proper studies should evaluate their prognostic or diagnostic meaning. These data are collected on cells in suspension, analyzed with a droplet-screening apparatus similar to the one described in

Neural networks for microfluidic droplet classification

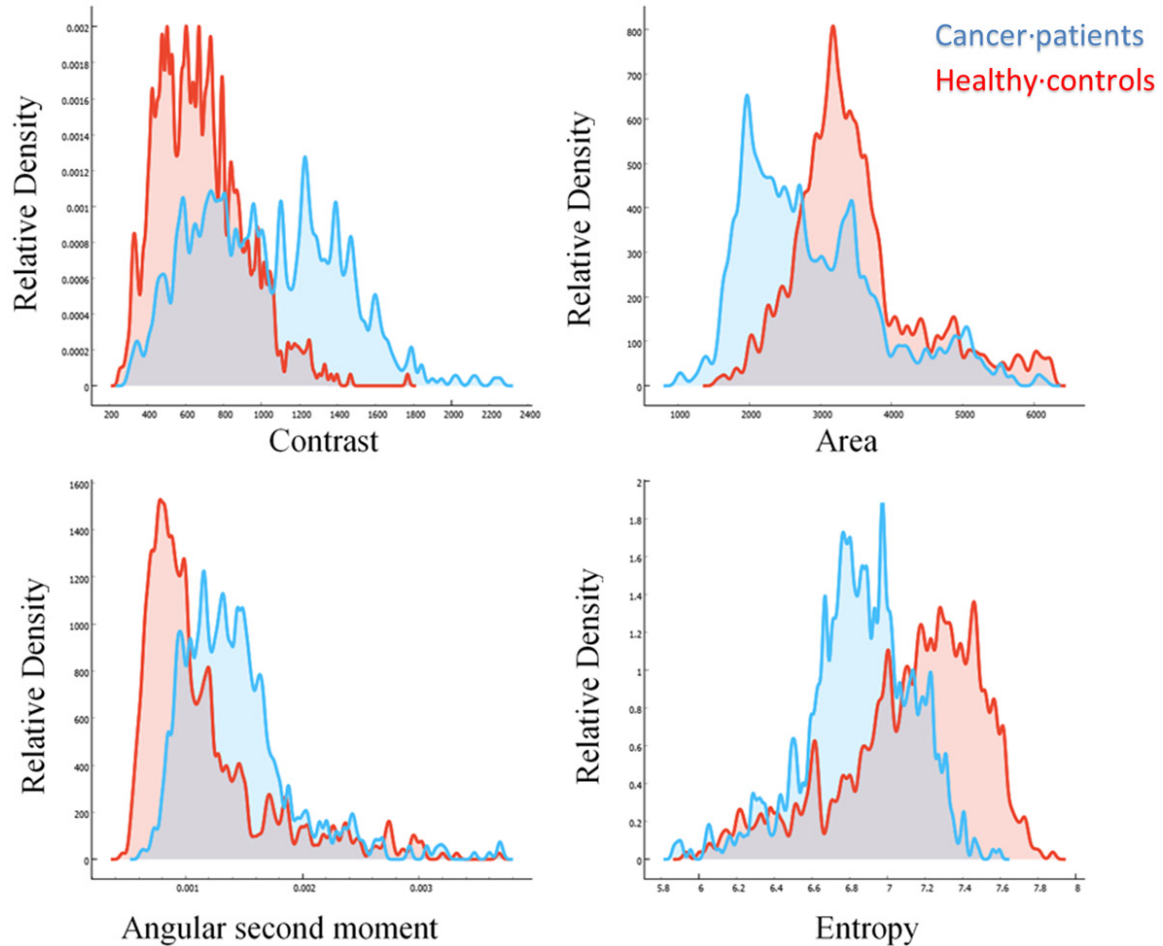


Figure 7. Frequency distributions of image features. Contrast (top-left), Area (top-right), Angular Second Moment (bottom-left), and Entropy (bottom-right) of images of cells coming from metastatic breast cancer patients (blue), or from healthy controls (red).

reported papers [6, 8], but the concept can be further exploited in other conditions, such as enriched cells collected on a plate or imaged through an image flow cytometer like ImageStream (Merck). With respect to the droplet screening pipeline, implementing this image analysis step to sort out false-positives generated by the analysis significantly increased the specificity, speed and operator-independency of the assay, and could potentially be applied to any other application involving droplet microfluidics.

Acknowledgements

Author contributions according to CASRAI CRediT: Conceptualization, F.D.B., V.D.M; Methodology, G.S., F.D.B., M.T; Investigation and Data Curation, E.B., G.B., M.B.; Writing-Original Draft, G.S., F.D.B.; Writing-Review & Editing,

F.D.B., M.T., V.D.M.; Resources, G.S., F.D.B., M.T., A.P.; Project administration, F.D.B., V.D.M.; Supervision, M.T., V.D.M.; Funding Acquisition, A.S. and V.D.M. We are grateful to Tania Corazza for the organization of the workflow granting access to blood samples, and to Alan Michael Sacilotto for the fabrication of the transimpedance amplifier. This work was supported by the EC Marie Curie Actions, AIDPATH project (Contract No.612471), by CRO Aviano 5X1000_2010_MdS, by AIRC 5X1000_2011_AIRC_12214. The TITAN X used for this research was donated by the NVIDIA Corporation.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Fabio Del Ben, Immunopathology and cancer biomarkers, CRO

Neural networks for microfluidic droplet classification

Aviano National Cancer Institute, Aviano, Italy.
E-mail: fabio.delben@cro.it

References

- [1] Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, Reuben JM, Doyle GV, Allard WJ, Terstappen LW, Hayes DF. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* 2004; 351: 781-791.
- [2] Hayes DF, Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Miller MC, Matera J, Allard WJ, Doyle GV, Terstappen LW. Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin Cancer Res* 2006; 12: 4218-24.
- [3] Cohen SJ, Punt CJ, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, Picus J, Morse M, Mitchell E, Miller MC, Doyle GV, Tissing H, Terstappen LW, Meropol NJ. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J Clin Oncol* 2008; 26: 3213-3221.
- [4] de Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, Doyle GV, Terstappen LW, Pienta KJ, Raghavan D. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res* 2008; 14: 6302-9.
- [5] Scher HI, Lu D, Schreiber NA, Louw J, Graf RP, Vargas HA, Johnson A, Jendrisak A, Bambury R, Danila D, McLaughlin B, Wahl J, Greene SB, Heller G, Marrinucci D, Fleisher M, Dittamore R. Association of AR-V7 on circulating tumor cells as a treatment-specific biomarker with outcomes and survival in castration-resistant prostate cancer. *JAMA Oncology* 2016; 2: 1441-1449.
- [6] Del Ben F, Turetta M, Celetti G, Piruska A, Bulfoni M, Cesselli D, Huck WT, Scoles G. A method for detecting circulating tumor cells based on the measurement of single-cell metabolism in droplet-based microfluidics. *Angew Chem Int Ed Engl* 2016; 55: 8581-8584.
- [7] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; 144: 646-674.
- [8] Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA. Single-cell analysis and sorting using droplet-based microfluidics. *Nat Protoc* 2013; 8: 870-91.
- [9] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542: 115-118.
- [10] Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017; 36: 61-78.
- [11] Sirinukunwattana K, Ahmed Raza SE, Yee-Wah Tsang, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging* 2016; 35: 1196-1206.
- [12] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to imageJ: 25 years of image analysis. *Nat Methods* 2012; 9: 671-675.
- [13] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2015.
- [14] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems* 2012; 1-9.
- [15] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. *IEEE CS Home* 2014; 1-9.
- [16] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. 2016.
- [17] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. *IEEE CS Home* 2016; 5987-5995.
- [18] Krizhevsky A. Learning multiple layers of features from tiny images. *Science Department, University of Toronto Tech* 2009; 1-60.
- [19] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. MobileNets: efficient convolutional neural networks for mobile vision applications. 2017.
- [20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015.
- [21] Demšar J, Curk T, Erjavec A, Hočevar T, Milutinović M, Možina M, Polajnar M, Toplak M, Starič A, Stajdohar M, Umek L, Zagar L, Zbončar J, Zitnik M, Zupan B. Orange: data mining toolbox in python. *J Mach Learn Res* 2013; 14: 2349-2353.
- [22] van der Walt S, Colbert SC, Varoquaux G. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering* 2011; 13: 22-30.
- [23] Jones E, Oliphant T, Peterson P. SciPy: open source scientific tools for python. <http://www.scipy.org/> 2001.
- [24] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011; 12: 2825-2830.

Neural networks for microfluidic droplet classification

- [25] Blasi T, Hennig H, Summers HD, Theis FJ, Cerveira J, Patterson JO, Davies D, Filby A, Carpenter AE, Rees P. Label-free cell cycle analysis for high-throughput imaging flow cytometry. *Nat Commun* 2016; 7: 10256.
- [26] Kobayashi H, Lei C, Wu Y, Mao A, Jiang Y, Guo B, Ozeki Y, Goda K. Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning. *Sci Rep* 2017; 7: 12454.
- [27] Chen CL, Mahjoubfar A, Tai LC, Blaby IK, Huang A, Niazi KR, Jalali B. Deep learning in label-free cell classification. *Sci Rep* 2016; 6: 21471.