Supervised machine learning in microfluidic impedance flow cytometry for improved particle size determination

de Bruijn, Douwe S.; ten Eikelder, Henricus R.A.; Papadimitriou, Vasileios A.; Olthuis, Wouter; van den Berg, Albert

DOI
10.1002/cyto.a.24679

Publication date
2022

Document Version
Final published version

Published in
Cytometry Part A

Received: 24 February 2022 | Revised: 6 May 2022 | Accepted: 29 July 2022

DOI: 10.1002/cyto.a.24679

ORIGINAL ARTICLE

ISAC CYTOMETRY Journal of Quantitative Cell Science PART A

Supervised machine learning in microfluidic impedance flow cytometry for improved particle size determination

Douwe S. de Bruijn[1] | Henricus R. A. ten Eikelder[1] | Vasileios A. Papadimitriou[2] | Wouter Olthuis[1] | Albert van den Berg[1]

[1]BIOS Lab-on-a-Chip Group, MESA+ Institute for Nanotechnology, Max Planck – University of Twente Center for Complex Fluid Dynamics, University of Twente, The Netherlands

[2]Chemical Engineering TU Delft, Delft, The Netherlands

Correspondence
Douwe S. de Bruijn and Albert van den Berg, University of Twente, P.O. Box 217 7500 AE Enschede, The Netherlands.
Email: d.s.debruijn@utwente.nl and a.vandenberg@utwente.nl

Funding information
Stichting voor Fundamenteel Onderzoek der Materie; Foundation for Fundamental Research on Matter

Abstract

The assessment of particle and cell size in electrical microfluidic flow cytometers has become common practice. Nevertheless, in flow cytometers with coplanar electrodes accurate determination of particle size is difficult, owing to the inhomogeneous electric field. Pre-defined signal templates and compensation methods have been introduced to correct for this positional dependence, but are cumbersome when dealing with irregular signal shapes. We introduce a simple and accurate post-processing method without the use of pre-defined signal templates and compensation functions using supervised machine learning. We implemented a multiple linear regression model and show an average reduction of the particle diameter variation by 37% with respect to an earlier processing method based on a feature extraction algorithm and compensation function. Furthermore, we demonstrate its application in flow cytometry by determining the size distribution of a population of small (4.6 ± 0.9 μm) and large (5.9 ± 0.8 μm) yeast cells. The improved performance of this coplanar, two electrode chip enables precise cell size determination in easy to fabricate impedance flow cytometers.

KEYWORDS
impedance flow cytometry, machine learning, multiple linear regression, neural network, particle size

1 | INTRODUCTION

Electrical microfluidic flow cytometers facilitate non-invasive, label-free and high-throughput single-cell analysis, enabling a large variety of biological cell studies based on e.g. their size and dielectric cell properties.[1–3] Especially in simple and easy to fabricate devices with coplanar electrodes (side by side), one of the main challenges is the positional dependence of the impedance signal in the sensing region. Here, the electric field is highly inhomogeneous in the sensing region (Figure 1). An oppositely placed electrode configuration (facing each other) can increase the sensitivity and reduce the particle position dependency to a certain extent, but strongly increases the complexity of fabrication.[2] Another possible work around, particle focusing,[4] but adds operational complexity to the system.

Meanwhile signal processing solutions have been studied to assess particle sizes accurately in all sorts of flow cytometers without particle focusing.[5–10] Most studies focus on pre-defined template and compensation functions to extract and process features from impedance signals: for example, particle size and position,[9, 10] particle size and velocity[5–8] and particle size, velocity and cross-sectional

This paper is an extended version of a conference contribution.

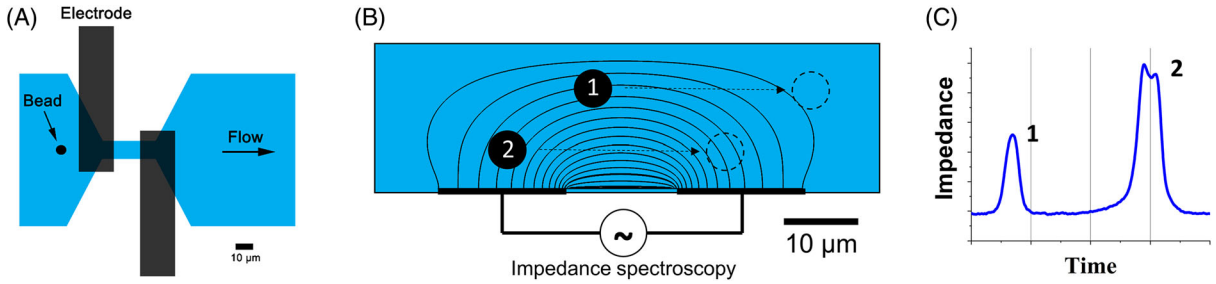

FIGURE 1 (A) Microfluidic chip with coplanar electrodes, (B) resulting in an inhomogeneous electric field. (C) The measured irregular signal shape and apparent particle size is position dependent as a result of the inhomogeneous field. The M-shaped peak (peak 2) is the result of a particle passing close to the electrodes, owing to a high field density near the electrodes[12] [Color figure can be viewed at wileyonlinelibrary.com]

position.[11] An elaborate discussion of feature extraction and feature processing techniques can be found in the recent review of Reference [1].

In this work, we focus on a new, accurate and simple method to extract the particle size from an irregular signal response (Figure 1C). The irregular signal response is the result of the particle position: an M-shaped signal for particles passing close the electrodes,[12] owing to the high field density near the electrodes and a single peak for particles further away from the electrodes (for more details see SI Reference [9]). We demonstrate a reduction in particle size variation and the ease of data processing, with respect to this earlier work,[9] using a multiple linear regression model in combination with a machine learning algorithm.[13]

Recent advances in machine learning and (impedance) flow cytometry have already shown promising results regarding: quantification of algal lipid accumulation using feature selection and regression analysis,[14] real-time intrinsic characterization of cancer cells and lymphocyte cells using a neural network,[15] classification of pollen using a combination of impedance cytometry and optical image processing,[16] a recurrent neural network to find the size, velocity and cross-sectional position of beads, red blood cells and yeast[17] and intrinsic characterization of red blood cells and detecting coinciding particles with a neural network.[18] Most of these papers rely on template fitting of the impedance signal to obtain the training labels and electrical features.

In short, we demonstrate a simple post-processing method based on supervised machine learning, which improves our particle size determination without the use of pre-defined signal templates and compensation functions.

2 | METHOD

The microfluidic device and experimental parameters that were used to acquire the multi-frequency impedance signals have been discussed elaborately in previous work.[9] In short, impedance data was recorded at 0.5, 1, and 12 MHz simultaneously using a lock-in amplifier and preamplifier (HF2LI and HF2TA, Zurich Instruments). The sample rate was 28.8 kSa/s and the signal was set at 1 $V_{peak\text{-}to\text{-}peak}$. Samples were pulled through the chip with a constant flow rate of 0.05 µl/min using a syringe pump (neMESYS, Cetoni). Before each experiments the microfluidic chip was coated with a monolayer surface coating (0.1 mg/ml PLL-g-PEG, SuSoS). 5, 6, and 7 µm polystyrene beads (Sigma-Aldrich and PolySciences) were diluted in seawater (conductivity = 4 S/m) to $\sim 5 \times 10^6$ beads/ml. A surfactant (Tween 20) was added and the samples were sonificated to prevent clumping. Additionally, a cell experiment was performed with fresh baker's yeast (*Saccharomyces cerevisiae*). Yeast was diluted in phosphate buffered saline (PBS, conductivity = 1.2 S/m) to $\sim 1 \times 10^7$ cells/ml and sucrose was added to make the yeast cells buoyant ($\sim$1.1 g/cm^3[19]). The same procedure was used to create training samples with 5 and 7 µm polystyrene beads. The yeast sample was spiked with 5 µm beads. The experiment with yeast and its training data was recorded at 0.12, 0.75, and 12 MHz, as will be explained later.

Our new data processing method is the following. Firstly, the baseline of the acquired signals was removed by subtracting a polynomial fit or a moving mean (also required for the previous methods), where after the passage of a bead was registered by a simple peak find algorithm ('findpeaks') in MATLAB (R2020a, MathWorks). Each event was saved at two frequencies (f_1 and f_2) with a window size N centered around the peak of signal f_1 (Figure 2A). Two frequencies were fed to the model to exploit the correlation between the particle height and the electrical opacity,[9] where the electrical opacity is defined as the impedance ratio at high and low frequency.[20] The individual events were later introduced to a single-layer neural network with linear activation function, which acts as a multiple linear regression model to predict the particle volume V, which scales linearly with the impedance magnitude $|Z|$ (Figure 2B). Subsequently, the particle diameter D can be calculated (Figure 2B). Investigations of models with hidden layers (deep learning) did not show any improvements, therefore we focus on the simpler and faster multiple linear regression model.

The multiple linear regression model with gradient descent algorithm was implemented in Python 3.7 using TensorFlow 2.4.1. The model was optimized using the adaptive moment estimation algorithm (Adam[21]) using the mean absolute error as a cost function. Two datasets, each acquired by a different chip (i.e., different alignment of the electrodes and fluidic channel) but with the same experimental

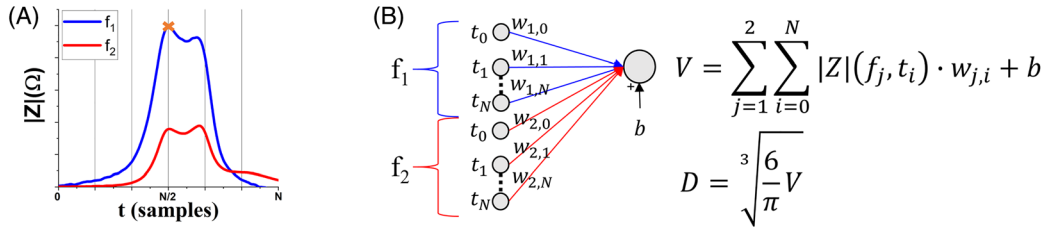

FIGURE 2 (A) The impedance response (|Z|) of a passing bead measured at two frequencies (f_1 and f_2) simultaneously. (B) The signals are fed to a multiple linear regression model, here visualized with N input nodes and a single output node. The weights (w) and bias (b) of the model are optimized using a gradient descent algorithm to predict the correct particle volume (V), where after the electrical diameter D is calculated [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Sample size of the datasets used for training (labeled) and testing (unlabeled) the model

	Training			Testing
Dataset	5 μm beads (# of samples)	6 μm beads (# of samples)	7 μm beads (# of samples)	Mixture of beads (# of samples)
1	880	463	292	775
2	906	473	536	1117

Note: The column "Training" represented all training data and is split into 85% training and 15% validation data.

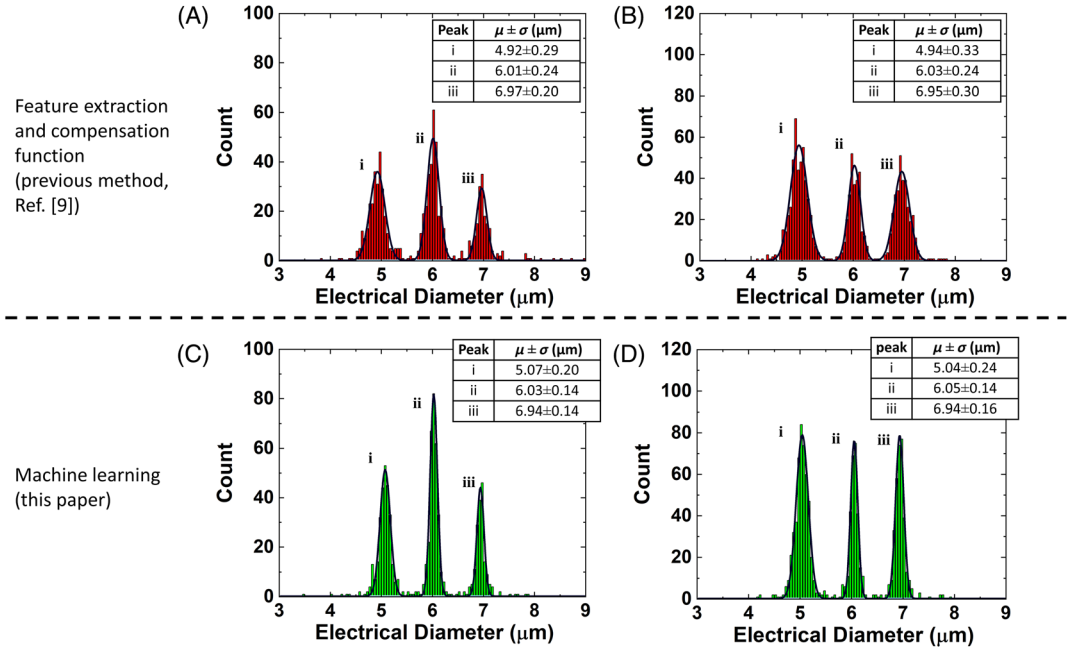

FIGURE 3 Histograms of the two testing datasets for both processing methods. The insets indicate the mean diameter (μ) and standard deviation (σ) of each peak [Color figure can be viewed at wileyonlinelibrary.com]

parameters, were used to train and test the model (Table 1). Each dataset contains training data of separate runs of 5, 6, and 7 μm monodisperse polystyrene beads (85% training and 15% validation) and test data with a mixture of these beads. The model was separately trained for dataset 1 and dataset 2 using a 0.5 and 12 MHz signal unless stated differently. The window size N for each event was set arbitrarily at 200 (for reference: Figure 2A displays 864 samples). A smaller window size (e.g., 100) reduced the performance of the model and a larger window sizes (e.g., 600) gave similar results.

The performance of the model was assessed by fitting a Gaussian distribution in Origin (2019b, OriginLab Corporation) on the unseen test data (mixture of 5, 6, and 7 μm beads). The standard deviation of each particle size distribution was compared to the previous method and the manufacturer's specifications (5.05 ± 0.16 μm (Sigma-Aldrich), 6.017 ± 0.168 μm (PolySciences) and 7.00 ± 0.11 μm (Sigma-Aldrich)). The mean diameter given by the manufacturer was converted to the particle volume and used as label for the training data.

TABLE 2 Performance of the processing methods expressed in the coefficient of variation (CV)

	Manufacturer	Dataset 1		Dataset 2	
Particle size (μm)	Specification CV (%)	Previous method, Reference [9] CV (%)	Machine learning CV (%)	Previous method, Reference [9] CV (%)	Machine learning CV (%)
5	3.1	5.8	4.0 (−31%)	6.6	4.8 (−27%)
6	2.8	4.0	2.3 (−42%)	4.0	2.3 (−42%)
7	1.5	2.9	2.0 (−31%)	4.3	2.3 (−47%)

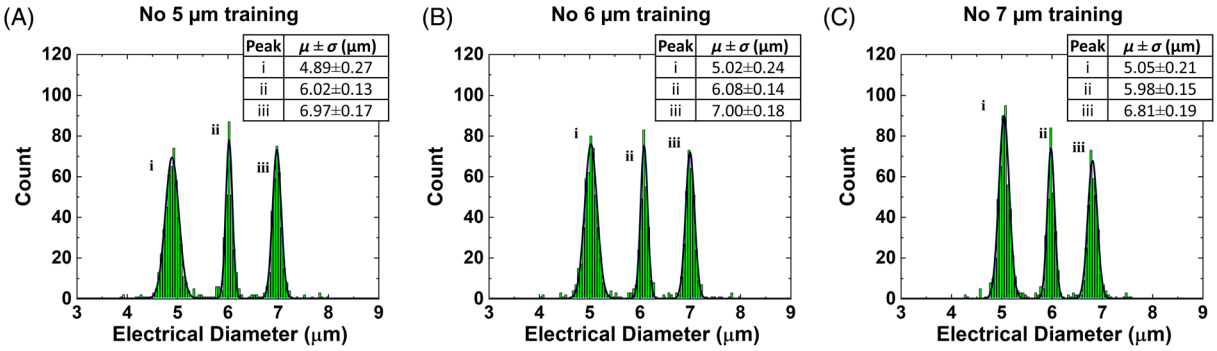

FIGURE 4 Histograms of inter- and extrapolation (dataset 2). (A,C) The extrapolation of 5 and 7 μm beads, respectively. (B) The interpolation of 6 μm beads. The insets indicate the mean diameter (μ) and standard deviation (σ) of each peak [Color figure can be viewed at wileyonlinelibrary.com]

To conclude, our new approach does not require any pre-defined fitting template and compensation function, but allows for feature extraction and processing in one step.

3 | RESULTS AND DISCUSSION

Two datasets have been processed using two different methods: the previous signal processing method and the machine learning approach (Figure 3). The machine learning method shows a significant improvement on the unseen particle mixtures for both datasets. We observe an average reduction of 37% in the particle size variation (Table 2). The coefficient of variation (CV) of the 6 μm beads meet the manufacturer's specifications, whereas the CV of the 5 and 7 μm beads is still larger than specified by the manufacturer.

We demonstrate a wider application of the machine learning approach by inter- and extrapolation of particle sizes at which the model has not been trained (Figure 4). The results are based on training of the model using only two particle sizes. One outstanding difference is the deviation in mean diameter ($\mu = 4.89$) for the 5 μm beads (Figure 4A) and the deviation in mean diameter ($\mu = 6.81$) for the 7 μm beads (Figure 4c). A two-point calibration with the calibration points close together is more prone to deviations. The interpolation of 6 μm beads (Figure 4B) performs comparable to the initial results with all training data. Summarizing, the model can be trained using just two particle sizes (preferably 5 and 7 μm beads), enabling the application in biological samples with a wider size distributions.

So far, all results were based on processing a 0.5 and 12 MHz signal for which the previous method worked best.[9] These frequencies were chosen rather far apart to enhance the opacity effect. For size determination in flow cytometry it is required to operate below the β-dispersion of cells (<several MHz[9, 22]). Therefore, we also have tested the model with a 0.5 and 1 MHz signal response (Figure 5) and we have observed a significant improvement for all particles compared to the previous method. The accuracy (machine learning with 0.5 and 1 MHz signal) is also close the machine learning approach with the 0.5 and 12 MHz signal. Altogether this machine learning approach makes a promising method to measure cell diameter accurately. This is further investigated by studying the size distribution of yeast cells.

As mentioned before, impedance data should be acquired below the β-dispersion of yeast cells to guarantee the same insulating properties of the cells as the calibration beads. Earlier research shows small differences between beads and yeast cells at ~1.1 MHz[23] and we observe no differences in opacity between beads and cells at 750/120 kHz (Figure S1). Thus, we have used a training set of 5 and 7 μm beads at 120 and 750 kHz to train our model, where after the model was used to predict the size of yeast cells spiked with 5 μm beads (Figure 6). The high frequency signal at 12 MHz was used to differentiate the beads and cells (Figure S1). The yeast cells show a distribution of small (4.6 ± 0.9 μm) and large (5.9 ± 0.8 μm) cells as reported earlier.[6, 7, 9, 23, 24] Optical inspection also shows small and large yeast cells and no budding events (Figure S2).

Both processing methods need calibration of specific chips and parameters, so training the model with dataset 1 yields poor results with the test data of dataset 2 and vice versa, because of small deviations in the alignment of electrodes and microfluidic channel, hence small deviations in the signal response. In general, parameters that influence the frequency response, like the buffer (different conductivity

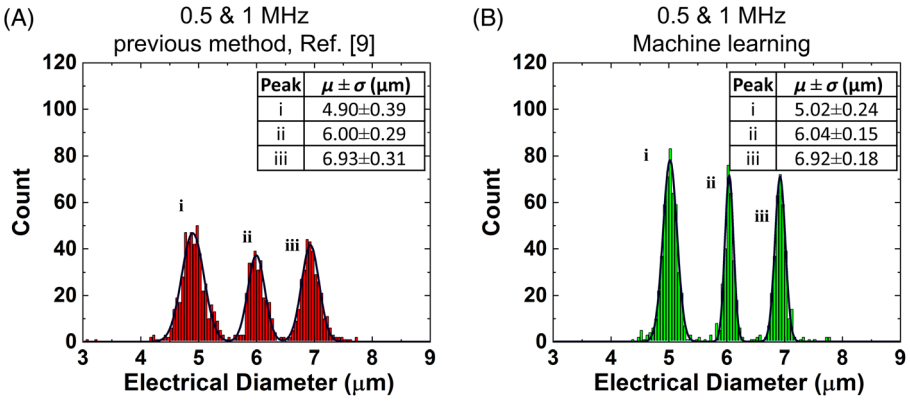

FIGURE 5 Size distribution of 5, 6, and 7 μm beads using 0.5 and 1 MHz signal responses using (A) the previous method, Reference [9] (B) machine learning. The insets indicate the mean diameter (μ) and standard deviation (σ) of each peak [Color figure can be viewed at wileyonlinelibrary.com]

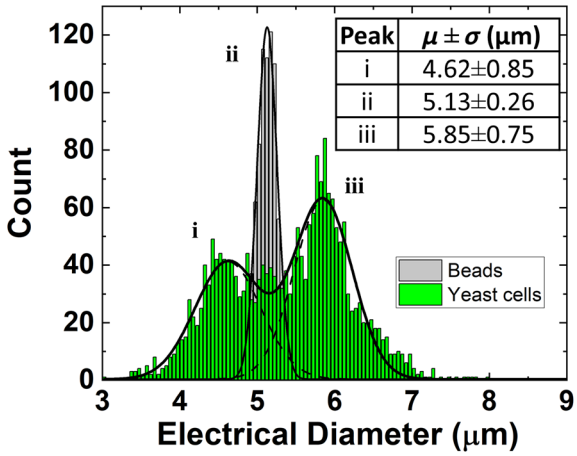

FIGURE 6 Size distribution of yeast cells and 5 μm beads. The electrical diameter was determined using a 120 and 750 kHz signal. The beads ($n = 837$) and yeast cells ($n = 2158$) were separated based on their opacity (12 MHz/750 kHz, see Figure S1).The insets indicate the mean diameter (μ) and standard deviation (σ) of each peak [Color figure can be viewed at wileyonlinelibrary.com]

and/or pH) or the chip alignment will influence the signal response. Switching the flow rate after calibration can most likely be accounted for using intra- and extrapolation of the recorded impedance signal. In cell experiments we can calibrate each measurement by mixing the sample with both 5 and 7 μm beads, classifying the beads and cells (using the high frequency response), then calibrating the model with the beads. Larger datasets in combination with deep learning might also yield better results in accounting for these before mentioned deviations, however that needs further investigation.

The machine learning method requires two sets of monodisperse beads for calibration in comparison to the previous method, which required one set. On the other hand, the machine learning method simplified the data processing, which was cumbersome in the previous method. I.e. it is no longer needed to differentiate between a single and M-shaped impedance response (Figure 1c) and it is no longer needed to express and evaluate a compensation function. The aforementioned steps needed proper interpretation and processing of the data. For example, in the previous method the M-shaped peak had to be differentiated from the single peak by correctly detecting the double peaks of the 'M-shape', where after the local minima had to be found (see SI and Reference [9]). This is much easier with the machine learning algorithm, which only needs to register the highest peak of the passing particle (Figure 2A) and then processes the data in one go. Please note, the algorithm does not account for coinciding particles, which might also give an M-shaped response, but it will give the apparent size of the coinciding particles.

Other studies may also benefit from this machine learning approach by replacing pre-defined templates and compensation functions. This requires the availability of labels for the training set, such as the mean particle diameter given by the manufacturer, or e.g. the particle position or velocity via optical inspection.

4 | CONCLUSION

We have demonstrated a simplified and more accurate processing method of irregular impedance signals based on supervised machine learning. Irregular signal shapes are easily processed using a multiple linear regression model, which does not require any pre-defined template fitting or correction function. This new method shows an average improvement of 37% in the particle size variation compared to the previous processing method (based on a feature extraction algorithm and compensation function[9]). Additionally, the performance with biological cells at low frequency has been demonstrated by determining the size distribution of yeast cells.

The improved performance of the system using input signals at low frequencies (0.5 and 1 MHz) is a promising development for the characterization of cell size in a very simple impedance flow cytometer with only two electrodes. Furthermore, requirements for the chip alignment are low, each chip (or different channel/electrode geometry) needs simple calibration with two sets of monodisperse beads.

AUTHOR CONTRIBUTIONS

Douwe S. de Bruijn: Conceptualization (lead); formal analysis (equal); investigation (equal); resources (lead); validation (lead); visualization (lead); writing – original draft (lead). **Henricus R. A. ten Eikelder:** Formal analysis (equal); investigation (equal); methodology (equal);

software (lead); writing – review and editing (supporting). **Vasileios A. Papadimitriou:** Conceptualization (supporting); resources (supporting); writing – review and editing (equal). **Wouter Olthuis:** Supervision (lead); validation (equal); writing – review and editing (equal). **Albert van den Berg:** Project administration (equal); supervision (supporting); writing – review and editing (equal).

ACKNOWLEDGMENTS

This work is part of the research program of the Foundation for Fundamental Research on Matter (FOM), which is part of the Dutch Research Council (NWO) and we thank the Max Planck, Center for Complex Fluid Dynamics.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cyto.a.24679.

ORCID

Douwe S. de Bruijn https://orcid.org/0000-0002-9547-3411

REFERENCES

1. Honrado C, Bisegna P, Swami NS, Caselli F. Single-cell microfluidic impedance cytometry: from raw signals to cell phenotypes using data analytics. Lab Chip. 2021;21(1):22–54.
2. Petchakup C, Li H, Hou HW. Advances in single cell impedance cytometry for biomedical applications. Micromachines. 2017;8(3):87.
3. Chen J, Xue C, Zhao Y, Chen D, Wu MH, Wang J. Microfluidic impedance flow cytometry enabling high-throughput single-cell electrical property characterization. Int J Mol Sci. 2015;16(5):9804–30.
4. Yang RJ, Fu LM, Hou HH. Review and perspectives on microfluidic flow cytometers. Sens Actuators B: Chem. 2018;266:26–45.
5. Spencer D, Caselli F, Bisegna P, Morgan H. High accuracy particle analysis using sheathless microfluidic impedance cytometry. Lab Chip. 2016;16(13):2467–73.
6. Errico V, De Ninno A, Bertani FR, Businaro L, Bisegna P, Caselli F. Mitigating positional dependence in coplanar electrode coulter-type microfluidic devices. Sens Actuators B: Chem. 2017;247:580–6.
7. De Ninno A, Errico V, Bertani FR, Businaro L, Bisegna P, Caselli F. Coplanar electrode microfluidic chip enabling accurate sheathless impedance cytometry. Lab Chip. 2017;17(6):1158–66.
8. Caselli F, De Ninno A, Reale R, Businaro L, Bisegna P. A novel wiring scheme for standard chips enabling high-accuracy impedance cytometry. Sens Actuators B: Chem. 2018;256:580–9.
9. de Bruijn DS, Jorissen KFA, Olthuis W, van den Berg A. Determining particle size and position in a coplanar electrode setup using measured opacity for microfluidic cytometry. Biosensors. 2021 Sep;11(10):353.
10. Yang D, Ai Y. Microfluidic impedance cytometry device with N-shaped electrodes for lateral position measurement of single cells/particles. Lab Chip. 2019;19(21):3609–17.
11. Reale R, De Ninno A, Businaro L, Bisegna P, Caselli F. High-throughput electrical position detection of single flowing particles/cells with non-spherical shape. Lab Chip. 2019;19(10):1818–27.
12. Cottet J, Kehren A, van Lintel H, Buret F, Frénéa-Robin M, Renaud P. How to improve the sensitivity of coplanar electrodes and micro channel design in electrical impedance flow cytometry: a study. Microfluid Nanofluid. 2019;23(1):1–11.
13. D. S. de Bruijn, H. R. A. Eikelder, V. A. Papadimitriou, W. Olthuis, and A. van den Berg, Particle size determination via supervised machine learning in microfluidic impedance spectroscopy, in *The 25th International Conference on Miniaturized Systems for Chemistry and Life Sciences (MicroTAS)*, 2021, Oct; pp. 1653–1654.
14. Tanhaemami M, Alizadeh E, Sanders CK, Marrone BL, Munsky B. Using flow cytometry and multistage machine learning to discover label-free signatures of algal lipid accumulation. Phys Biol. 2019;16(5):055001.
15. Feng Y, Cheng Z, Chai H, He W, Huang L, Wang W. Neural network-enhanced real-time impedance flow cytometry for single-cell intrinsic characterization. Lab Chip. 2021;22:1037–43.
16. DaOrazio M, Reale R, De Ninno A, Brighetti MA, Mencattini A, Businaro L, et al. Electro-optical classification of pollen grains via microfluidics and machine learning. IEEE Trans Biomed Eng. 2022 Feb;69(2):921–31.
17. Honrado C, McGrath JS, Reale R, Bisegna P, Swami NS, Caselli F. A neural network approach for real-time particle/cell characterization in microfluidic impedance cytometry. Anal Bioanal Chem. 2020;412(16):3835–45.
18. Caselli F, Reale R, De Ninno A, Spencer D, Morgan H, Bisegna P. Deciphering impedance cytometry signals with neural networks. Lab Chip. 2022;22:1714–1722.
19. Baldwin WW, Kubitschek HE. Buoyant density variation during the cell cycle of Saccharomyces cerevisiae. J Bacteriol. 1984;158(2):701–4.
20. Hoffman RA, Johnson TS, Britt WB. Flow cytometric electronic direct current volume and radiofrequency impedance measurements of single cells and particles. Cytometry. 1981;1(6):377–84.
21. D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," *3rd International* Conference on Learning Representations ICLR 2015, pp. 1–15, 2015.
22. Schwan HP. Electrical properties of tissue and cell suspensions. Vol 5. Cambridge: Academic Press; 1957.
23. Haandbæk N, Bürgel SC, Heer F, Hierlemann A. Characterization of subcellular morphology of single yeast cells using high frequency microfluidic impedance cytometer. Lab Chip. 2014;14(2):369–77.
24. Sherman F. Getting started with yeast. Methods Enzymol. 2002;350(2002):3–41.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: de Bruijn DS, ten Eikelder HRA, Papadimitriou VA, Olthuis W, van den Berg A. Supervised machine learning in microfluidic impedance flow cytometry for improved particle size determination. Cytometry. 2022. https://doi.org/10.1002/cyto.a.24679