Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5ay01064k

Enabling multiple cellular enumeration applications of a bioparticle sensing platform using machine learning

Muhammad Nabeel Tahir, [a] Brandon K. Ashley, [ab] Jianye Sui,[ac] Mehdi Javanmard[ac] and Umer Hassan *[ad]

Cellular surface receptors are commonly used as diagnostic and prognostic biomarkers for many infectious diseases. Benchtop clinical instruments, *e.g.*, flow cytometers and fluorescence microscopes are commonly employed for identification and quantification of these biomarkers. These diagnostic techniques suffer from high instrument cost, laborious protocols and limited multiplexing ability. Recently, we have demonstrated the surface receptor detection of blood cells using novel electrically sensitive microparticles. Here, we employ computational pruning and machine learning techniques to improve the detection of cell surface receptors as biomarkers. The data collected from microfluidic impedance flow cytometry consisting of a multifrequency response of metal oxide-coated microparticles conjugated to blood cells (granulocytes) with either CD11b or CD66b surface receptors and video recordings were transformed to numerical values and manually annotated to train machine learning models. Classification accuracies of ~95% and 97% before and after applying outlier removal techniques were observed when differentiating between cells and cells with 10 nm-Al_2O_3 anti-CD11b conjugated particles. Classification accuracies of ~94% and ~96% were observed when differentiating between cells and cells with 20 nm-Al_2O_3 anti-CD66b conjugated particles. Similarly, an improvement in the classification of neural networks was observed for different applications of the impedance spectrometry platform data. This provides preliminary results that impedance cytometer data collected for multiple biomarkers, along with the trained machine-learning models and noise reduction techniques, can efficiently quantify the biomarkers. This study will help in enabling next-generation cytometry technology with a potential for improved disease diagnosis in the future.

Received 26th June 2025
Accepted 13th October 2025

DOI: 10.1039/d5ay01064k

rsc.li/methods

Introduction

Several multi-faceted diseases, *e.g.*, sepsis, cancer, *etc.* affecting several millions of people around the world, are very difficult to diagnose as they show non-specific symptoms and because of the inability of current diagnostic tools to identify host response to infection or tumors. Significant efforts have been made to study the different biomarkers for these complex diseases and factors limiting the use of one specific biomarker as a diagnosis tool for sepsis, [1–4] for cancer, [5–7] and in other acute diseases, *e.g.*, kidney injury,[8,9] liver diseases,[10] *etc.* Although the biomarkers reported in the previous study lack the ability to directly correlate with the diseases, a new class of cell surface-based biomarkers is emerging as a possible alternative due to their highly correlated expression response under inflammatory or infectious conditions.[11] Researchers have evaluated various cell surface receptors to identify the diseases, *e.g.*, CD4, CD8,[12] CD18,[13] and CD28 [14] on T-cells, and CD64,[15] CD11b,[16] and CD66b.[17] Quantifying these cell surface receptors poses a problem as it requires expensive equipment and complicated sample processing protocols to collect the data from multiple biomarkers to draw a conclusion on the status of the disease. Therefore, there's a clinical need for inexpensive multiplexed biomarker detection systems that can not only detect the presence of certain biomarkers present in the blood sample but also quantify the expression of cell surface receptors.

To address these challenges, point-of-care devices may provide a possible solution with the ability to analyze multiple biomarkers at a time. These point-of-care devices offer various advantages, from providing quicker results to processing smaller sample volumes to faster analysis, miniaturized design, and economical affordability.[18–20] One class of point-of-care devices that holds significant promise in delivering these features in the current disease diagnosis landscape is impedance flow cytometry-based point-of-care devices. These

[a]*Department of Electrical and Computer Engineering at Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA*
[b]*Siemens Healthineers, USA*
[c]*Riz Health, Inc., USA*
[d]*Global Health Institute at Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA. E-mail: umer.hassan@rutgers.edu*

impedance-based devices can not only profile cells but also detect protein or DNA levels without destroying the structures and provide electrical data rapidly.[21,22] Although the impedance-based bioparticle profiling techniques are promising, they are limited in identifying multiple species simultaneously. Recently, our team presented a novel approach[23–26] where metal oxide-coated particles with varying thicknesses and materials of oxide layers can be used in a multiplexing modality to electrically identify the bioparticles in a multifrequency electric field. Moreover, the studies presented a heterogeneous data collection protocol along with the functionalization of metal oxide-coated Janus particles (MOJPs) with antibodies to target multiple cell surface receptors, hence providing a solution to the multiplexing challenge faced by impedance flow cytometry-based point-of-care devices.

Herein, enhanced classification accuracy of the multiplexed impedance flow cytometer is demonstrated with the application of computational and machine learning techniques. The data collected from electrical impedance values and video recording are transformed into a numerical set of features and are used to train the custom-designed neural network. To further enhance the detection accuracy, three computational outlier detection and removal techniques, isolation forest, local outlier factor, and Mahalanobis distance, were employed to prune the data from unwanted or noise signatures, and the accuracies are compared with those in the previously reported study.

Methods and materials

The following materials were used in sample preparation and processing: Ficoll–Paque density gradient, streptavidin, phosphate saline buffer (PBS, 1X, PH = 7.2), triethoxysilane, Roswell Park Memorial Medium (RPMI 1640), anti-CD11b monoclonal mouse antibody with biotin, and anti-CD66b antibody with biotin.[26] The following equipment was used to perform the experiments: HF2TA, a current amplifier and HF2LI a lock-in amplifier from Zurich instruments to record the impedance cytometry readings; an NE-300 syringe pump was used to control the flow of liquid in the microfluidic channel; an inverted microscope was used to view the channel and record the sample video using a Chronos 1.4 high-speed camera.[26] The deidentified human blood samples were acquired from Robert Wood Johnson Hospital (RWJH) following the institutional review board guidelines with IRB application number Pro2018002356.

Fabrication of microfluidic channels, electrodes, and barcoded particles

The fabrication process of microfluidic channels and electrodes follows the process discussed in a previous publication.[26,27] A photolithography process was employed where a silicon wafer coated with a negative photoresist is first baked, exposed to UV light, washed, and hard baked to etch the microfluidic channel structures on the wafer. A polydimethylsiloxane (PDMS) device containing a microchannel is fabricated by pouring the PDMS solution onto a Si-wafer mold, followed by a vacuum and baking step to remove the air bubbles and solidify the device. Finally, the device is removed from the wafer and attached to the electrode glass substrate containing electrodes. The fabrication of gold-plated electrodes also follows a similar photolithography process. A borosilicate wafer covered with a positive photoresist is exposed to UV light passing through the mask, making the photoresist soluble in areas where the electrode layers will be deposited. The electrodes are fabricated by sputtering a 250 nm layer of chromium followed by 750 nm of gold. The metal oxide-coated Janus particles (MOJPs) are fabricated following the technique reported in a previous study.[28,29] A single layer of 3 μm polystyrene beads is formed on the surface of a glass substrate by drop-casting. The electron beam deposition method deposited a 20 nm gold layer on the beads. Finally, a 20 nm or 10 nm aluminum oxide layer is deposited using the electron beam deposition method, thus creating the aluminum oxide Janus microparticles (MOJPs).

Functionalizing MOJPs

Microparticles need to be functionalized with receptor-targeting antibodies to bind with the neutrophils. The details of the functionalizing receptor-targeting antibodies have been published in previous studies.[24,26] First, streptavidin (SAv) is adsorbed on the surface of the 20 nm (SAv|20 nm) or 10 nm (SAv|10 nm) MOJPs by adding 2 μL of SAv into 200 μL of either particle solution containing approximately 6.0×10^7 particles per mL. A wash step is performed to remove the unbounded SAv. In the next step, 10 μL of biotinylated anti-CD11b is added to the (SAv|10 nm) solution, and 10 μL of biotinylated anti-CD66b is added to the (SAv|20 nm) solution to achieve the functionalization of the antibodies.

Sample preparation and neutrophil isolation

The deidentified patient blood samples acquired from RWJH were used to isolate the neutrophils. The blood sample was mixed with 1X PBS solution with equal ratios of 1 : 1, and a Ficoll-density gradient was added to the mixture at a ratio of 2 : 3. The solution was centrifuged to remove red blood cells, platelets, plasma, and non-neutrophil cells by lysing and aspiration through adding the deionized water for a short duration and restoring tonicity by adding 10X PBS. The step was repeated until a pellet of neutrophils was formed, and the supernatant was pipetted out. The neutrophil pellet was then suspended in RPMI 1640 media at a ratio of 1 : 10. After the isolation step, three solutions were prepared: 10 nm CD11b particle solution, 20 nm CD66b particle solution by mixing them with diluted neutrophil solution at a ratio of 10 : 1, and the third solution without MOJPs consisting only of the neutrophils alone. The solutions were incubated for one hour to allow conjugation.

Heterogeneous data acquisition (impedance and video)

Data for used in this study is collected using impedance flow cytometry and simultaneous video recording. A configuration consisting of three electrodes, a lock-in amplifier, and a current amplifier was used to collect the impedance flow cytometry data. Fig. 1 shows the configuration described for the

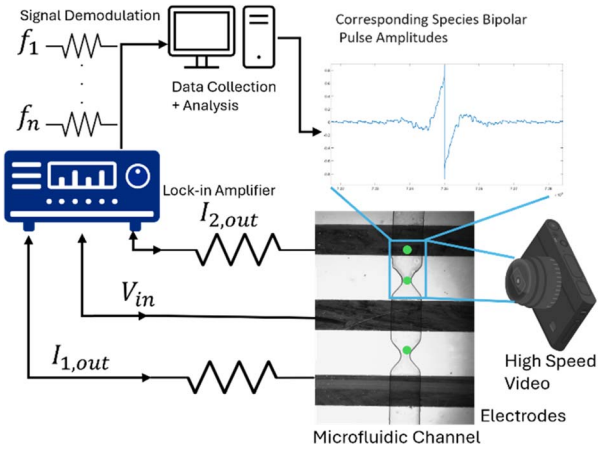

Fig. 1 System overview of the multi-modal data collection protocol.

impedance data collection. The central electrode was provided with an input signal of amplitude 10 V containing frequency components of 500 kHz, 7.5 MHz, 8.3 MHz, and 9 MHz. The side electrodes, grounded with a 10 kΩ resistance, provided the current readings. A differential amplifier was used along with a current amplifier and a current-to-voltage converter circuit to generate the bipolar amplitude pulse. The signal was demodulated and sampled at 250 kHz using an NI PCIe-6361 DAQ. Following the data acquisition, digital filtering was performed using MATLAB, where the signal was passed through a high-pass 4th-order Butterworth filter with a cutoff at 20 Hz, a 4th-order low-pass Butterworth filter with a cutoff at 100 Hz, and a 1st-order band-stop filter with cutoffs at 60 Hz and 120 Hz. The response was recorded for all four frequency components later to assess the conjugation of particles.

To simultaneously record videos of cells passing through the channel at a flow rate of 15 μL min^{-1}, a Chronos 1.4 high-speed camera was used, focusing on the narrow regions of the channel as depicted in Fig. 1. The camera recorded the video at a frame rate of 4500 fps, with a shutter speed of 25 μs, and a resolution of 1280 × 360. Neutrophils with or without MOJPs were manually annotated by matching the time of occurrence of the corresponding electrical pulse (higher in magnitude than red blood cells or residual) and the timestamp of each frame. The number of particles conjugated with each cell was recorded, and corresponding pulse data were saved for each frequency component. The data are then used to train the machine-learning models to classify cells alone or cells with N particles conjugated. The details of the data acquisition and labelling of the cells with conjugated particles are discussed in our previous study[26] and provided in the SI.

Outlier removal and deep learning

The data collected using the protocol explained in the previous section consist of 20 features providing information regarding the noise, amplitudes, and full-width half-maximum (FWHM) of all four frequencies. The cytometry data are subject to outliers or noise spikes for various reasons, *e.g.*, sensor malfunction, unexpected particles (clump of cells), or correlated features. This study evaluated three outlier detection and removal techniques, including isolation forest (ISF), local outlier factor (LOF), and Mahalanobis distance (MD). The deep learning models were trained on the datasets before and after applying outlier removal methods, and the impact of each technique was evaluated.

The deep neural network containing two hidden batch normalization and activation layers was employed to classify whether the cell is alone or has 1, 2, or 3+ particles conjugated. The input layer contains 20 units corresponding to the number of features, followed by a hidden layer with 256 units, a batch normalization layer, and an activation layer. The second hidden layer consists of 128 hidden units, and finally, the output layer consists of variable units, followed by the SoftMax layer for classification. Fig. 2 shows the architecture of the neural network used in the study, along with details of each layer. While training the neural network, hyperparameters were kept constant to make a fair comparison regarding the effectiveness of the outlier removal techniques. Stochastic gradient descent with momentum was used as an optimizer, with a learning rate of 0.001, a batch size of 16, and a number of epochs of 250. The datasets were divided into three parts: training, test, and validation, with percentages of 70%, 15%, and 15%, respectively. We performed the power analysis to calculate the minimum number of the training examples required. Targeting a confidence interval of 95% (z − score = 1.96), a model accuracy of 85% (p = 0.85), and a margin of error E = 5%, the number of training examples required can be calculated as follows:

$$N = \frac{Z^2 \times p(1-p)}{E^2} = \frac{1.96^2 \times 0.85(1-0.85)}{0.05^2} = 196$$

Which gives us the lower limit on the training examples, and for the case where we have the lowest training examples of 397 (Sample C cells *vs.* Cells with 10 nm CD11b after application of Mahalanobis) the data split will result in the training examples of 277. This number is still higher than the calculated training examples. Table S1 shows the distribution of the data for each class used in the study. The models were trained and evaluated on test and validation data. The resulting classification accuracy plots were generated. We also evaluated other performance metrics including the precision, recall, and F1-score for all the applications. The details of these metrics are reported in Tables S2–S5.

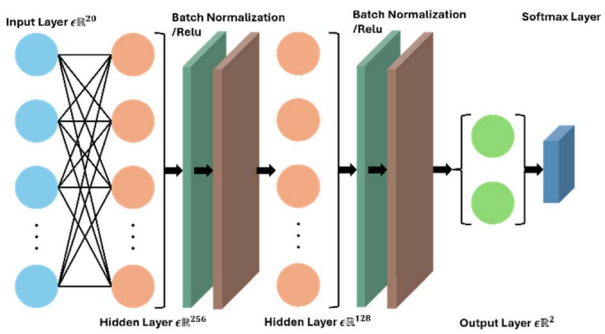

Fig. 2 Architecture of the neural network evaluated in this study.

Results and discussion

The dataset collected, as discussed in the previous section, can be utilized in various applications to draw a comparison between cells alone (without any conjugated particle) and cells conjugated with *N* number of particles, indicating the expression of certain proteins on the surface of the cells. This study aims to focus on CD11b and CD66b expressions on the cells and collect the respective data. We divided the data into five groups: Cells *vs.* Cells with 10 nm CD11b particles as one class, Cells *vs.* Cells with 20 nm CD66b particles as one class, and Cells *vs.* Cells with 10 nm CD11b and 20 nm CD66b as one class each. With this combination, we aimed to classify the cells alone and the cells with conjugated particles. Models trained on these group data will be able to differentiate if a cell is alone or has 10 nm or 20 nm particles conjugated, hence representing the presence of a particular surface receptor. It can be used in applications where it's required to differentiate the cells from other cells expressing certain proteins. Similarly, we created two more groups, but instead of classifying between cells alone and 10 nm CD11b or 20 nm CD66b conjugated cells as one class, we further stratified the data and created multiple classes, *i.e.*, Cells with (1, 2, 3+) 10 nm CD11b or 20 nm CD66b particles. Hence, the groups now contain four classes, and by training the models, the presence of proteins on the surface of the cells can be roughly quantified as well. The dataset was collected from three subjects and contains three samples. The results reported in the coming sections are on individual samples and the combined dataset.

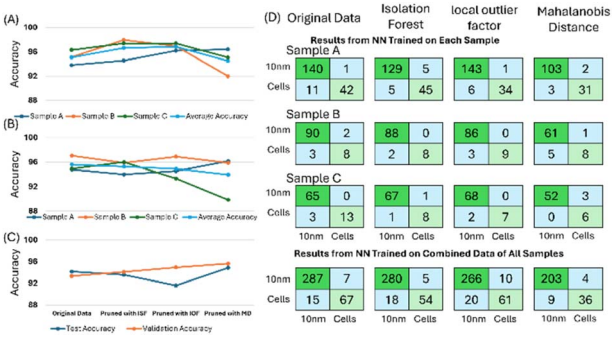

Fig. 3 Cells *vs.* (cells with all 10 nm CD11b as one class for each sample): (A) Test accuracies for averaged and individual samples. (B) Validation accuracies for average and individual samples. (C) Cells *vs.* (cells with all 10 nm CD11b as one class combined data) test/validation accuracy. (D) Confusion matrices for each sample plotted before and after application of the outlier removal methods.

Application 1: cells alone *vs.* cells with 10 nm CD11b (indicating single cell enumeration)

The dataset containing the cells alone and cells conjugated with 10 nm CD11b particles was prepared as a class problem as one of the applications of the study, where all the 10 nm particles attached to a neutrophil were considered as one class. The classification accuracy of a custom neural network (NN) on original data and after application of pruning methods was generated on test and validation data and is reported in Fig. 3(A and B) for individual samples and the entire dataset combined in Fig. 3(C). Even without removing outliers, the average classification accuracy of the model on all sample test data is ∼95% and average F1-score is ∼96.9%, which is higher than the reported test accuracy of ∼92% in the previous study.[26] It can be observed from Fig. 3(A and B) that with the application of both isolation forest and local outlier detection algorithms, the overall average sample test accuracy has been increased to ∼97% with an F1-score of ∼98%. One possible explanation for the higher performance of ISF and LOF compared to the slightly lower performance of the MD can be either the presence of local outliers due to noise spikes or the clumping of different-sized particles, resulting in the varied local density of the individual reading as compared to its neighbours. Since the ISF and LOF can pick up these patterns, efficiency increases. When combining the data of all three samples and training the same model, a slightly different trend in the test accuracy is observed, where the test accuracies on the original data remain the same while the test accuracy with LOF experiences a decrease in value. This can be justified by the fact that by combining the data of three samples, the data points that were local outliers are now part of a bigger cluster (as LOF compares the density of a point to its local neighbours in a cluster) and no longer exhibit lower density. Table 1 shows the effect of outlier removal techniques on the number of training examples, and it can be observed that significant numbers of outliers have been removed from the original dataset while the overall test accuracy of the model has been improved.

Application 2: cells alone *vs.* cells with 20 nm CD66b (indicating single cell enumeration)

The second application of this study comprises differentiating between cells alone *vs.* cells conjugated with 20 nm CD66b particles. Like the results discussed in the previous application, even without removing outliers, the NN performs better at classifying the test examples with an accuracy of ∼94% as compared to the test accuracy of 91% reported in ref. 26. An interesting pattern observed from the analysis, as reported in Fig. 4(A and B), is that the Mahalanobis distance and LOF perform better in terms of improving the accuracy of the classification model. In contrast, the ISF performs slightly poorly. This trend could be attributed to the fact that the data collected for neutrophils conjugated with 20 nm CD66b exhibit feature values that are closely related and contain a lower number of local outliers, so ISF and MD performed better. Applying MD increased the average sample test accuracy to 96%, which is a significant improvement from the test accuracy reported in ref. 26. Training the NN on combined data of 20 nm CD66b samples resulted in better performance, as shown in Fig. 4(C), with a test accuracy of ∼92%, an F1-score of 95.0% and a maximum accuracy observed with an LOF of ∼95% and F1-score of 97%. It can also be observed from Fig. 4(C) that all the outlier removal techniques have improved the overall test accuracy, indicating that erroneous values were present in the data, and a more generalized model has been trained with

Table 1 Number of outliers removed for each sample with the application of outlier removal techniques

	Outlier removal results on individual samples				Outlier removal results on combined dataset			
	Original	Isolation forest	Local outlier factor	Mahalanobis distance	Original	Isolation forest	Local outlier factor	Mahalanobis distance
Cells *vs.* (cells with all 10 nm CD11 as one class)	1285 681 531	1221 647 504	1221 647 504	918 493 397	2497	2372	2372	1675
Cells *vs.* (cells with all 10 nm CD11 as one class)	1370 961 690	1302 914 656	1302 913 656	941 658 485	3021	2870	2870	1976
Cells *vs.* (cells with all 10 nm CD11 class + cells with all 20 nm CD66 as one)	2303 1550 1124	2188 1473 1068	2188 1473 1068	1583 1029 787	4977	4728	4728	3248
Cells *vs.* (cells with all 10 nm CD11 as three classes)	1285 681 531	1221 647 504	1221 647 504	918 493 397	2497	2372	2372	1678
Cells *vs.* (cells with all 20 nm CD66 as three classes)	1370 961 690	1302 913 657	1302 913 657	940 657 485	3021	2871	2870	1974

similar differentiation accuracy to that achieved on individual samples.

Application 3: cells alone *vs.* cells with 10 nm CD11b and cells with 20 nm CD66b (indicating multiplexing)

The third application of the study focuses on differentiating the cells alone from cells with either 10 nm CD11b or cells with 20 nm CD66b, making it a three-class problem. The same NN was employed with minor changes to the output layer, which contains three output neurons. Fig. 5 shows the test and validation accuracy of models trained on individual samples and combined data. The average test data accuracy of the individual samples remains consistent with or without applying the outlier removal techniques, *i.e.*, ISF and LOF, at ~78%, while MD reduces the classification to 75%, as shown in Fig. 5(A and B). Similarly, the combined data depict a decrease in the classification test accuracy trend with all outlier detection methods, with the maximum accuracy reported for the original data of 77% and F1-score of 77%, as shown in Fig. 5(C). Since NN relies on the variance and differences in the data to learn to differentiate between classes, the results suggest that outlier removal techniques remove the outliers in the dataset, which is necessary to differentiate between different populations of the cells. Removing the outliers also reduced the diversity in the data and unique values of each class label, resulting in the model overfitting the training data and, hence, performing poorly on test and validation data.

Application 4: cells alone *vs.* cells with 10 nm CD11b (as three classes indicating receptor expression)

One possible application of the study is in identifying if the cells are alone or have 1, 2, or 3 + 10 nm CD11b particles conjugated.

Fig. 4 Cells *vs.* (cells with all 20 nm CD66b as one class for each sample). (A) Test accuracies for averaged and individual samples. (B) Validation accuracies for averaged and individual samples. (C) Cells *vs.* (cells with all 20 nm CD66b as one class combined data) test/validation accuracy. (D) Confusion matrices for each sample plotted before and after application of the outlier removal methods.

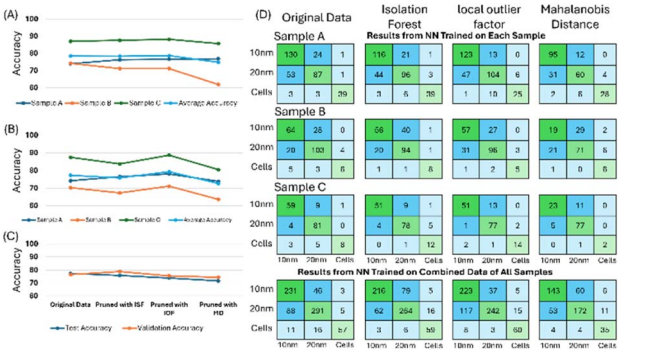

Fig. 5 Cells *vs.* (cells with all 10 nm CD11b as one class + cells with all 20 nm CD66 as one class for each sample). (A) Test accuracies for averaged and individual samples. (B) Validation accuracies for averaged and individual samples. (C) Cells *vs.* (cells with all 10 nm CD11b as one class + cells with all 20 nm CD66 as one class combined data) test/validation accuracy. (D) Confusion matrices for each sample plotted before and after application of the outlier removal methods.

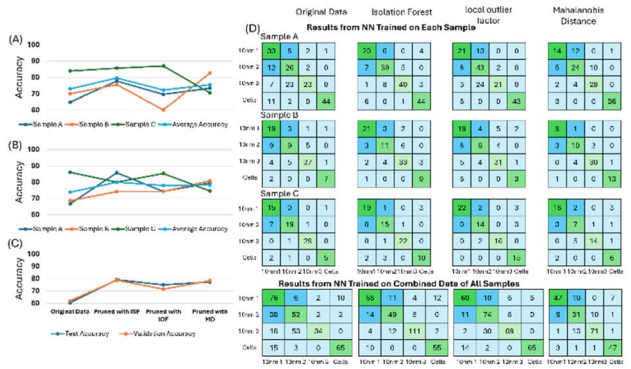

Fig. 6 Cells *vs.* (cells with all 10 nm CD11b as three classes for each sample). (A) Test accuracies for averaged and individual samples. (B) Validation accuracies for averaged and individual samples. (C) Cells *vs.* (cells with all 10 nm CD11b as three classes combined) test/validation accuracy. (D) Confusion matrices for each sample plotted before and after application of the outlier removal methods.

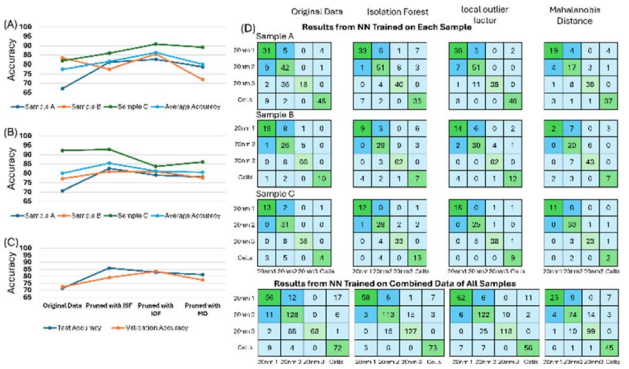

Fig. 7 Cells *vs.* (cells with all 20 nm CD66b as three classes for each sample). (A) Test accuracies for averaged and individual samples. (B) Validation accuracies for averaged and individual samples. (C) Cells *vs.* (cells with all 10 nm CD11b as three classes combined) test/validation accuracy. (D) Confusion matrices for each sample plotted before and after application of the outlier removal methods.

One can estimate the overall expression of respective surface proteins on the cells with individual particle conjugation identification. Quantifying the expression of the cell surface receptors can be used in targeted drug discovery, understanding the cell–surface interactions, and receptor binding studies. Fig. 6(A and B) show the validation and test classification accuracies generated for each sample and averaged with and without outlier removal. It was observed that without outlier removal, the model had an average test accuracy of ∼73%, increasing to ∼79% with ISF. Although the overall accuracies are lower compared to the results discussed in the previous applications, it's worth noting that outlier removal techniques can improve the accuracies, and increasing the size of the dataset can further improve the classification accuracies. The ISF outlier removal method not only improved the classification for individual sample data but also significantly improved the test accuracy of the combined data. It can be observed in Fig. 6(C) that the application of ISF improved the test accuracy from ∼60% to ∼79% with an F1-score of 79%, which is the same as the average test accuracy of individual sample data. The overall lower classification accuracies can be attributed to the imbalance of the class datasets, as was observed while training the models, or minor differences between the feature values of each class label. This problem can be addressed by increasing the size of the dataset, as discussed earlier, or applying data augmentation and feature engineering techniques.

Application 5: cells alone *vs.* cells with 20 nm CD66b (as three classes indicating receptor expression)

Similar to the previous section, an application of the study in differentiating the cells or cells with 1, 2, 3 + 20 nm CD66b particles was evaluated. A lower average test accuracy of ∼77% was observed for the individual sample dataset, and outlier removal further improved the average test accuracy to ∼86% with LOF, as shown in Fig. 7(A and B). However, the other outlier removal methods improved the test accuracy, and significant improvements were observed with LOF. On the combined dataset, the outlier removal also improved the test accuracy from 71% to ∼85% with ISF with an F1-score of 86%. ISF significantly improved the test accuracy, suggesting that the dataset consisted of enough anomalies to make the model less generalized or let the model overfit the data.

Discussion

Training a custom neural network and data pruning methods can help improve classification accuracy, resulting in various applications of the microfluidic impedance cytometer. The dataset from the impedance flow cytometer is subject to variance due to manual processing of blood samples, equipment, and manual annotations. Training a new model on each sample is not a practical solution, as is the case in ref. 26. Therefore, there is a need to develop a more generalized model that can provide similar or improved accuracies to datasets of varied spreads. The results presented in this study suggest that this aim can not only be achieved but also further improved with more data collection and complex NN design. Moreover, it can also be observed that significant improvements in the test accuracies were observed when classifying the expression of the protein on the surface of the cells in applications 5 and 6.

Conclusion

Three outlier detection and removal techniques, followed by a custom NN training pipeline, have been presented to enhance the classification accuracy of the data collected from the impedance flow cytometer. Multiple potential applications of the current system have been evaluated, targeting the differentiation of the cells alone against the cells expressing certain proteins. The applications are not only limited to differentiating if a cell expresses certain proteins but also encompass the quantification of the expression of proteins on the surface of the cells. Two cell surface receptors, CD11b and CD66b, expressed on the surface of the neutrophils, were targeted in the study,

enabling multiple surface receptor identification. Metal oxide-coated Janus particles with varying thicknesses of oxide coating and functionalized with anti-CD11b or anti-CD66b were incubated with neutrophils, and data were collected from an impedance flow cytometer and video recordings. The results discussed in the previous sections suggest that outlier removal techniques are effective in improving the classification accuracy of custom NN, which already performs better on average as compared to previous studies. Although different outlier removal techniques performed better in different applications on average, the ISF appeared to improve the accuracy in all the applications. Hence, an outlier removal technique, followed by a generalized neural network, can be used as an effective tool for differentiating the cells with or without certain proteins expressed on the surface. In the future, we aim to extensively increase the size of the dataset and include various cell surface receptors to not only provide multiple biomarker identification but also improve the classification accuracy of the machine learning models.

Author contributions

M. N. Tahir contributed to study design, data analysis, building machine learning models and preparing the original draft. B. K. Ashley contributed towards the sensor design, development, and data collection. J. Sui contributed towards the fabrication of the metal oxide particles. M. Javanmard contributed to study design, funding acquisition, reviewing, and co-editing. U. Hassan contributed to study design, funding acquisition, supervision, reviewing, and co-editing the manuscript.

Conflicts of interest

There is no conflict of interest.

Data availability

Data supporting this article has been included as part of Supplementary Information (SI). Raw experimental data related to this study will be available upon request and based on the funding agencies and Rutgers, The State University of New Jersey policies. Supplementary information: additional details of experimental setup and performance metrics of machine learning models for different applications discussed in the manuscript. See DOI: **https://doi.org/10.1039/d5ay01064k**.

Acknowledgements

The authors would like to acknowledge the funding support from the National Science Foundation (NSF award number: 2329761 and 2002511). The authors also acknowledge support from the Department of Electrical and Computer Engineering and Global Health Institute at Rutgers, The State University of New Jersey.

References

1 S.-K. Hung, H.-M. Lan, S.-T. Han, C.-C. Wu and K.-F. Chen, Current Evidence and Limitation of Biomarkers for Detecting Sepsis and Systemic Infection, *Biomedicines*, 2020, **8**(11), 494, DOI: **10.3390/biomedicines8110494**.

2 M.-H. Kim and J.-H. Choi, An Update on Sepsis Biomarkers, *Infect. Chemother.*, 2020, **52**(1), 1, DOI: **10.3947/ic.2020.52.1.1**.

3 B. M. Biron, A. Ayala and J. L. Lomas-Neira, Biomarkers for Sepsis: What is and What Might Be?, *Biomarker Insights*, 2015, **10s4**(S29519), BMI, DOI: **10.4137/BMI.S29519**.

4 S. Leonard, *et al.*, Pediatric sepsis inflammatory blood biomarkers that correlate with clinical variables and severity of illness scores, *J. Inflammation*, 2024, **21**(1), 7, DOI: **10.1186/s12950-024-00379-w**.

5 F. Boissière-Michot, W. Jacot, J. Fraisse, S. Gourgou, C. Timaxian and G. Lazennec, Prognostic Value of CXCR2 in Breast Cancer, *Cancers*, 2020, **12**(8), 2076, DOI: **10.3390/cancers12082076**.

6 E. Schmidt, *et al.*, Tumor-Associated Neutrophils Are a Negative Prognostic Factor in Early Luminal Breast Cancers Lacking Immunosuppressive Macrophage Recruitment, *Cancers*, 2024, **16**(18), 3160, DOI: **10.3390/cancers16183160**.

7 Y. Liu, *et al.*, Advances in immunotherapy for triple-negative breast cancer, *Mol. Cancer*, 2023, **22**(1), 145, DOI: **10.1186/s12943-023-01850-7**.

8 Y. Wen and C. R. Parikh, Current concepts and advances in biomarkers of acute kidney injury, *Crit. Rev. Clin. Lab. Sci.*, 2021, **58**(5), 354–368, DOI: **10.1080/10408363.2021.1879000**.

9 F. Fan and P. Xu, Global biomarkers trends in acute kidney injury: a bibliometric analysis, *Renal Failure*, 2023, **45**(2), 2278300, DOI: **10.1080/0886022X.2023.2278300**.

10 K. Patel and G. Sebastiani, Limitations of non-invasive tests for assessment of liver fibrosis, *JHEP Rep.*, 2020, **2**(2), 100067, DOI: **10.1016/j.jhepr.2020.100067**.

11 T. Schmidt, *et al.*, CD66b overexpression and homotypic aggregation of human peripheral blood neutrophils after activation by a gram-positive stimulus, *J. Leukocyte Biol.*, 2012, **91**(5), 791–802, DOI: **10.1189/jlb.0911483**.

12 G. A. Koretzky, Multiple Roles of CD4 and CD8 in T Cell Activation, *J. Immunol.*, 2010, **185**(5), 2643–2644, DOI: **10.4049/jimmunol.1090076**.

13 B. K. A. Abdel-Salam and H. Ebaid, Clinical immunology Expression of CD11b and CD18 on polymorphonuclear neutrophils stimulated with interleukin-2, *Cent. Eur. J. Immunol.*, 2014, **2**, 209–215, DOI: **10.5114/ceji.2014.43725**.

14 L. E. Petersen, *et al.*, Characterization of senescence biomarkers in rheumatoid arthritis: relevance to disease progression, *Clin. Rheumatol.*, 2019, **38**(10), 2909–2915, DOI: **10.1007/s10067-019-04615-0**.

15 A. A. H. Zeitoun, S. S. Gad, F. M. Attia, A. S. Abu Maziad and E. F. Bell, Evaluation of neutrophilic CD64, interleukin 10 and procalcitonin as diagnostic markers of early- and late-

onset neonatal sepsis, *Scand. J. Infect. Dis.*, 2010, **42**(4), 299–305, DOI: **10.3109/00365540903449832**.

16 L. Yang, C. M. Edwards and G. R. Mundy, Gr-1+CD11b+ myeloid-derived suppressor cells: Formidable partners in tumor metastasis, *J. Bone Miner. Res.*, 2010, **25**(8), 1701–1706, DOI: **10.1002/jbmr.154**.

17 T. Schmidt, *et al.*, CD66b Overexpression and Loss of C5a Receptors as Surface Markers for Staphylococcus aureus-Induced Neutrophil Dysfunction, *PLoS One*, 2015, **10**(7), e0132703, DOI: **10.1371/journal.pone.0132703**.

18 L. L. Sun, Y. S. Leo, X. Zhou, W. Ng, T. I. Wong and J. Deng, Localized surface plasmon resonance based point-of-care system for sepsis diagnosis, *Mater. Sci. Energy Technol.*, 2020, **3**, 274–281, DOI: **10.1016/j.mset.2019.10.007**.

19 K. Yuan, *et al.*, Smartphone-based hand-held polarized light microscope for on-site pharmaceutical crystallinity characterization, *Anal. Bioanal. Chem.*, 2023, **415**, 4401–4410, DOI: **10.1007/s00216-023-04582-1**.

20 N. N. Watkins, *et al.*, Microfluidic CD4 + and CD8 + T Lymphocyte Counters for Point-of-Care HIV Diagnostics Using Whole Blood, *Sci. Transl. Med.*, 2013, **5**(214), 214ra170, DOI: **10.1126/scitranslmed.3006870**.

21 B. K. Ashley and U. Hassan, POINT-OF-CRITICAL-CARE diagnostics for sepsis enabled by multiplexed MICRO AND NANO sensing technologies, *Wiley Interdiscip. Rev.:Nanomed. Nanobiotechnol.*, 2021, **13**(5), e1701, DOI: **10.1002/wnan.1701**.

22 C. Petchakup, P. E. Hutchinson, H. M. Tay, S. Y. Leong, K. H. H. Li and H. W. Hou, Label-free quantitative lymphocyte activation profiling using microfluidic impedance cytometry, *Sens. Actuators, B*, 2021, **339**, 129864, DOI: **10.1016/j.snb.2021.129864**.

23 P. Xie, X. Cao, Z. Lin and M. Javanmard, Top-down fabrication meets bottom-up synthesis for nanoelectronic barcoding of microparticles, *Lab Chip*, 2017, **17**(11), 1939–1947, DOI: **10.1039/C7LC00035A**.

24 B. K. Ashley, J. Sui, M. Javanmard and U. Hassan, Antibody-functionalized aluminum oxide-coated particles targeting neutrophil receptors in a multifrequency microfluidic impedance cytometer, *Lab Chip*, 2022, **22**(16), 3055–3066, DOI: **10.1039/D2LC00563H**.

25 B. K. Ashley, J. Sui, M. Javanmard, and U. Hassan, Aluminum Oxide-Coated Particle Differentiation Employing Supervised Machine Learning and Impedance Cytometry, in *2022 IEEE 17th International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)*, Taoyuan, Taiwan, IEEE, Apr. 2022, pp. 211–216, DOI: **10.1109/NEMS54180.2022.9791160**.

26 B. K. Ashley, J. Sui, M. Javanmard and U. Hassan, Multi-modal sensing with integrated machine learning to differentiate specific leukocytes targeted by electrically sensitive hybrid particles, *Biosens. Bioelectron.*, 2023, **241**, 115661, DOI: **10.1016/j.bios.2023.115661**.

27 B. K. Ashley and U. Hassan, Time-domain signal averaging to improve microparticles detection and enumeration accuracy in a microfluidic impedance cytometer, *Biotechnol. Bioeng.*, 2021, **118**(11), 4428–4440, DOI: **10.1002/bit.27910**.

28 J. Sui, P. Xie, Z. Lin and M. Javanmard, Electronic classification of barcoded particles for multiplexed detection using supervised machine learning analysis, *Talanta*, 2020, **215**, 120791, DOI: **10.1016/j.talanta.2020.120791**.

29 B. K. Ashley, J. Sui, M. Javanmard and U. Hassan, Functionalization of hybrid surface microparticles for *in vitro* cellular antigen classification, *Anal. Bioanal. Chem.*, 2021, **413**(2), 555–564, DOI: **10.1007/s00216-020-03026-4**.

