Smart deep learning application for single cell classification

David Dannhauser*[a], Paolo A. Netti[a,b], Filippo Causa[a]
[a]Interdisciplinary Research Centre on Biomaterials (CRIB) and Dipartimento di Ingegneria Chimica, dei Materiali e della Produzione Industriale, University of Naples "Federico II", 80125 Naples, Italy; [b]Center for Advanced Biomaterials for Healthcare@CRIB, Istituto Italiano di Tecnologia, 80125 Naples, Italy; *david.dannhauser@unina.it

ABSTRACT

Identifying specific cell types within diverse populations is challenging, especially when labels are limited or unknown. We present a self-supervised deep learning approach to classify single-cell scattering patterns captured via microfluidics, enabling the detection of rare cells such as circulating tumor cells. Unlike traditional methods like flow cytometry, our technique is faster, simpler, and preserves native cells for downstream analysis. Conventional classifiers are limited by their reliance on labeled data and closed-set assumptions, leading to misclassification of unseen cell types. To overcome this, we developed an open-set classifier that dynamically updates known classes and uses an auxiliary domain to detect out-of-distribution cells. Through instance weighting and feature extraction, the model learns to identify novel classes without prior labels. This method is adaptable to other imaging modalities, such as bright-field and holography. Our work encourages researchers to explore their model's feature space to better understand classification boundaries and uncertainties, ultimately improving performance in biological sample analysis.

Keywords: Single Cell, Deep Learning, Light Scattering, Microfluidics, Smart, Circulating Tumor Cells

1. INTRODUCTION

Accurate identification of specific cell types within heterogeneous populations remains a central challenge in biomedical research, particularly when labeling is limited by specificity or when prior knowledge of the target class is unavailable. This issue is especially critical for rare or poorly characterized cell types, such as circulating tumor cells (CTCs), which are important biomarkers for cancer diagnosis and prognosis, yet remain difficult to detect using conventional techniques [1]. Traditional methods such as flow cytometry and immunostaining require extensive labeling protocols and often fail to preserve native cells for downstream analysis. Deep learning-based image classification has emerged as a powerful tool for single-cell analysis [2], but most existing models rely on supervised learning paradigms and closed-set assumptions, restricting classification to only those cell types included in the training data. As a result, these models are prone to misclassifying novel or out-of-distribution (OOD) cell types, limiting their utility in open-ended biological scenarios [3]. To overcome these limitations, we propose a novel, self-supervised deep learning framework for classifying single-cell scattering patterns acquired through microfluidic imaging. Our approach enables identification of unknown cell types—including rare populations like CTCs—without requiring labeled training data for those target classes. By integrating high-content scattering information with an open-set classification strategy, our method dynamically updates its internal representation of known classes and employs an auxiliary domain to detect OOD cells [4,5]. The model further uses an instance-weighting mechanism to emphasize in-distribution features, allowing for more accurate identification of previously unseen classes. In addition to scattering-based imaging, this framework can be adaptable to bright-field or holographic modalities, offering a versatile solution for single-cell classification across diverse experimental contexts [6,7].

2. MATERIAL AND METHODS

Living cells were isolated using a standard density gradient centrifugation technique, followed by cell-type-specific enrichment through a negative selection protocol employing antibody-conjugated magnetic beads, as previously described [8]. The isolated cells were then suspended in a viscoelastic alignment medium at a concentration of

Multimodal Sensing and Artificial Intelligence for Sustainable Future, edited by Francesco Soldovieri, Pascal Picart, Vittorio Bianco, Claas Falldorf, Proc. of SPIE Vol. 13570, 135700J
2025 Published by SPIE · 0277-786X · doi: 10.1117/12.3062288

approximately 1×10^5 cells/mL[9]. These samples were analyzed using a custom-built small-angle light scattering (SALS) system integrated with a microfluidic device comprising an alignment section and an observation channel [8,9]. As individual cells passed through the observation channel, they sequentially intersected a collimated incident laser beam. This interaction produced distinct scattering patterns, which were recorded over a continuous angular range of 3° to 33°, with an angular resolution of approximately 0.1° [8–10]. To prepare the data for convolutional neural network (CNN) analysis, the scattering images were pre-processed to standardize and enhance pattern quality. This involved centroid detection and the application of a binary mask that excluded angular regions outside the 3°–33° range. The mask was generated through a pipeline combining iterative 2D Gaussian low-pass filtering, low-light image inversion for spot enhancement, haze reduction, and global thresholding. The resulting segmented images were then centered, cropped to 650×650 pixels, and resized to 224×224 pixels for the CNN dataset. The neural network architecture consisted of an input layer for processing the scattering patterns, followed by multiple convolutional layers for feature extraction. A final classification layer was used to minimize a predefined loss function and perform label prediction. This conventional learning method, known as 'closed-set' classification (see Fig. 1), assumes that test data is drawn from the same distribution as the training data.

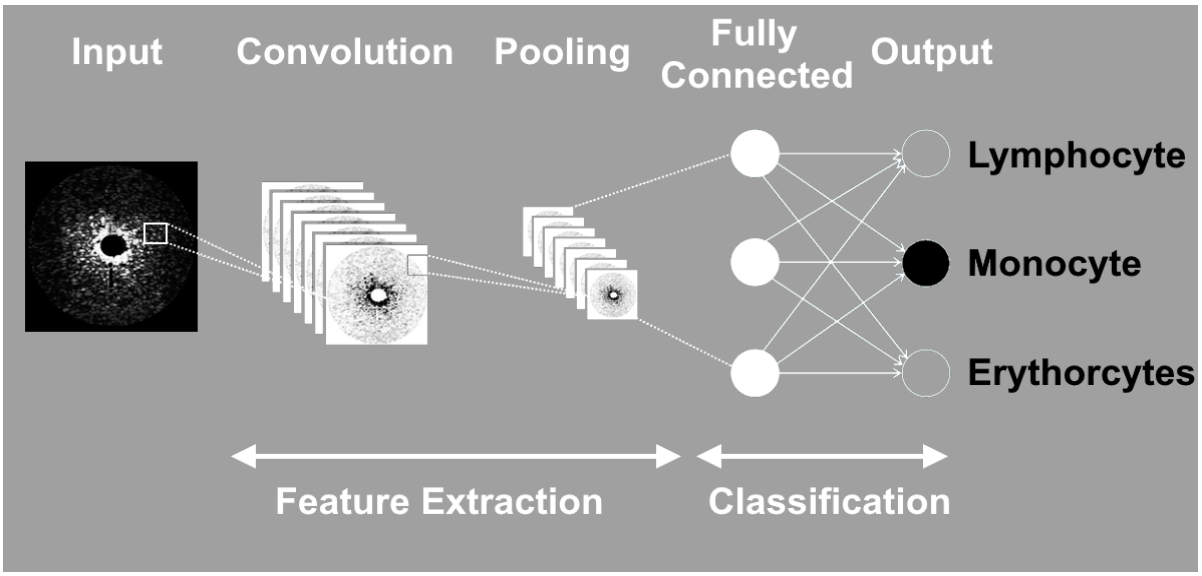

Figure 1. Schematic CNN workflow for the closed-set cell type prediction.

To enable open-set recognition -smart deep learning- in a self-supervised manner, the Auxiliary Open-Set Risk (AOSR) framework was employed[11,12]. This method identifies previously unseen cell classes by aligning the feature space of the CNN with that of auxiliary samples. The feature vectors extracted from the closed-set CNN served as input to train the AOSR algorithm, which initializes its auxiliary domain with randomly generated samples and iteratively estimates class weights to detect and classify unknown cell types (see Fig. 2).

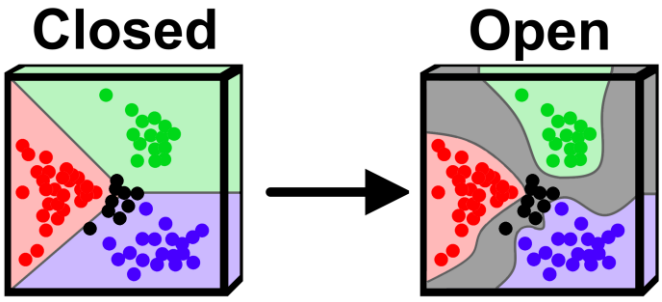

Figure 2. Feature space representation from closed-set to open-set with AOSR. Black dots indicate unknown cells, while the gray area in the open-set indicates the newly created feature space during AOSR.

3. RESULTS

We initially trained the closed-set CNN using a dataset obtained from independent measurements of living, unlabeled erythrocytes [8], lymphoblasts [9], and monoblasts [10]. Once the model achieved a cell classification accuracy exceeding 90% for the known classes, we proceeded to optimize the OOD detection using the AOSR method. As illustrated in Figure 2, the AOSR framework effectively partitions the feature space, clustering known cell types into distinct regions while allocating a separate region for potential unknown classes. This spatial separation enables robust unknown cell detection without significantly compromising the classification accuracy of known cell types. To achieve this balance, we adopted a two-phase optimization strategy: the network was first tuned for accurate classification of known cells, followed by a fine-tuning phase focused on improving the model's ability to detect unknown classes. During this process, we deliberately limited the known cell prediction accuracy to remain below 95%, as excessively high accuracy can hinder the network's sensitivity to unfamiliar cell types in the feature space (see Figure 3). In other words, a feature space tightly optimized for perfect known class separation reduces the capacity to accommodate novel class clusters.

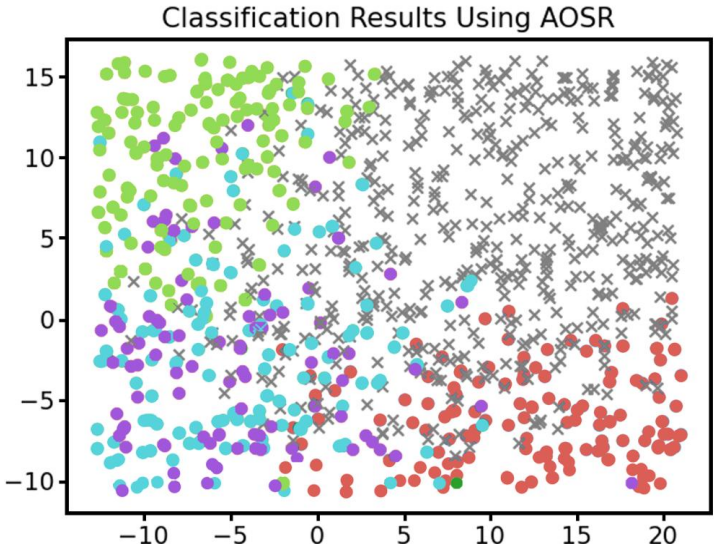

Figure 3. Feature space information from the AOSR classifier. Gray crosses show unknown cells (cells that have not been used for the training of the network), while dots indicate different known cell classes (also called known cell classes).

Optimization involved adjusting both architectural parameters -such as convolutional filter configurations- and hyperparameters that govern feature space geometry. A key parameter in this context was β, which controls the degree of overlap between known and unknown feature distributions. Increasing the beta value promotes separation between known and unknown classes but may also increase the rate at which known cells are misclassified as unknown. Therefore, beta must be carefully tuned for each newly introduced cell type to ensure optimal open-set performance.

4. CONCLUSION

In contrast to conventional flow cytometry, our method provides a simple and cost-effective alternative for classifying cell subtypes, eliminating the need for large sample volumes or labor-intensive labeling procedures. A key advantage of our approach is its ability to operate on living cells in suspension, enabling subsequent collection and potential reuse in diagnostic or therapeutic contexts. Notably, our AOSR framework allows for the prediction of previously unseen cell types without requiring a dedicated training dataset, which is particularly valuable for detecting and analyzing rare cell populations. The optimized system demonstrates robust performance in classifying scattering patterns across a wide range of cell types and holds strong potential for broader application to other modalities of single-cell imaging data. Furthermore, the integration of uncertainty analysis enhances the reliability of classification results, which is especially critical for the identification and study of rare or clinically significant populations, such as circulating tumor cells.

ACKNOWLEDGEMENTS

We thank Gaia Cioffi for her valuable insights and support in the development of the open-set recognition architecture.

REFERENCES

[1] Alix-Panabières, C., & Pantel, K. "Challenges in circulating tumour cell research," Nature Reviews Cancer, 14(9), 623–631 (2014). https://doi.org/10.1038/nrc3820
[2] Esteva, A., *et al.* "A guide to deep learning in healthcare," Nature Medicine, 25, 24–29 (2019). https://doi.org/10.1038/s41591-018-0316-z
[3] Bendale, A., & Boult, T. E. "Towards open set deep networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1563–1572 (2016). https://doi.org/10.1109/CVPR.2016.173
[4] Sun, J., *et al.* "Scattering-based label-free cell classification using deep learning," Nature Communications, 11, 1246 (2020). https://doi.org/10.1038/s41467-020-14957-4
[5] Perera, P., and Patel, V. M. "Generative-discriminative feature representations for open-set recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 48(2), 441–453 (2020). https://doi.org/10.1109/TPAMI.2019.2913377
[6] Jo, Y., *et al.* "Holographic deep learning for rapid optical screening of anthrax spores," Science Advances, 3(8), e1700606 (2019). https://doi.org/10.1126/sciadv.1700606
[7] Zhang, L., *et al.* "Deep learning-based single-cell classification using 3D refractive index tomograms," Nature Communications, 11, 3101 (2020). https://doi.org/10.1038/s41467-020-16920-2
[8] D. Dannhauser, D. Rossi, F. Causa, P. Memmolo, A. Finizio, T. Wriedt, J. Hellmers, Y. Eremin, P. Ferraro and P. A. Netti, "Optical signature of erythrocytes by light scattering in microfluidic flows," Lab on aChip, **15**(16), 3278-3285 (2015). https://doi.org/10.1039/C5LC00525F
[9] D. Dannhauser, D. Rossi, A. T. Palatucci, V. Rubino, F. Carriero, G. Ruggiero, M. Ripaldi, M. Toriello, G. Maisto, P. A. Netti, G. Terrazzano and F. Causa, "Non-invasive and label-free identification of human natural killer cell subclasses by biophysical single-cell features in microfluidic flow," Lab on a Chip, **21**(21), 4144-4154 (2021). https://doi.org/10.1039/D1LC00651G
[10] D. Dannhauser, D. Rossi, V. De Gregorio, P. A. Netti, G. Terrazzano and F. Causa, "Single cell classification of macrophage subtypes by label-free cell signatures and machine learning. Royal Society Open Science, 9(9), 220270 (2022). https://doi.org/10.1098/rsos.220270
[11] Cioffi, G., Dannhauser, D., Rossi, D., Netti, P. A., & Causa, F. (2023). Unknown cell class distinction via neural network based scattering snapshot recognition. Biomedical Optics Express, *14*(10), 5060-5074. https://doi.org/10.1364/BOE.492028
[12] Dannhauser, D. (2025, June 1). Github page for unknown cell class distinction. https://github.com/David-UniNA/Unknown-Cell-Class-Distinction