On Its Way to Primetime: Artificial Intelligence in Flow Cytometry Diagnostics

Stefan W. Krause*

In this issue of Cytometry A, Zhao et al. (page 1073–1080) report on their work to diagnose leukemic B cell non-Hodgkin’s Lymphoma from flow cytometry (FCM) raw data of blood and bone marrow samples using a dedicated computer approach, which would assign one of eight B-cell lymphoma diagnoses or “normal” to a sample. A remarkable level of classification performance could be achieved in the validation set. For the “true” classification of B-cell lymphomas, conventional diagnostics had incorporated morphology, FCM and additional information from histology and genetics if needed, whereas computer diagnosis was derived from FCM data alone. In this context, uncertainty to delineate, for example, monoclonal B-cell lymphocytosis from chronic lymphocytic leukemia or to subclassify a B-cell malignancy as either mantle cell lymphoma or prolymphocytic leukemia is not an outright error, but is rather based on the limitations of FCM itself. Furthermore, cell populations tagged as abnormal by the algorithm and color-coded accordingly in conventional plots can help human diagnosticians to review and fine-tune the diagnosis. However, some lymphomas (most prominent in follicular lymphoma) were classified as normal by the algorithm. Vice versa, only few samples classified as “normal” by human diagnosticians were classified as lymphoma by the algorithm. Thus, a deficit in sensitivity exists, which is clinically relevant.

Computer support is instrumental for the analysis of FCM data, because nobody is able to draw conclusions from raw list mode files. However, conventional FCM computer programs execute relative simple tasks to support the workflow of a human researcher or diagnostician. In a typical workflow, several sequential steps have to be performed (Fig. 1, left side). Fluorescence spillover compensation is calculated from control samples. One-dimensional transformation of raw data (logarithmic, logical, possibly a shift of zero and negative values to some defined minimum, etc.) is routinely performed on fluorescence channels. Data are displayed in histograms or two-dimensional plots. Starting gates are used to look for artifacts and to remove debris and cells not of interest. A considerable number of plots are necessary, if several fluorochromes are used and several populations are of interest. Data from several samples with identical panel may be displayed in parallel in an overlay. Cells are tagged according to gates in these plots and may then be displayed separately and/or color-coded. Hierarchical and/or Boolean gating strategies are used for the definition of cell populations and subpopulations of interest. Cell numbers and antigen expression of these cell populations of interest constitute the readout of a single tube. A final result or diagnosis is derived assessing this readout or the synopsis of the readout of several tubes.

All of the calculations in such a manual workflow are based on straight “if A then B” logic, performing calculations on a maximum of two parameters concurrently. Conventional FCM computer support aims at displaying data in a clear manner to the human operator, especially effects of manipulation in two-parameter plots upon plots of other parameters, but not at automation. The most advanced process in standard applications is the calculation of fluorescence spillover compensation, which nowadays usually is performed in some (semi-) automated fashion. However, although every single step in this procedure is quite straightforward, due to the...
In the recent decades, many attempts have been reported to introduce more advanced computation methods into histology, cytopathology, image cytometry and conventional FCM analysis (1,2). These algorithms will be called artificial intelligence (AI) from here, although some of them do not deserve this name in its strict sense. Two strategies, sometimes overlapping, are applied in these attempts: firstly, AI may be used to automatize conventional data processing and analysis as described above in order to reduce the workload for the investigator, reduce bias using standardized procedures, and speed up analyses. To this end, regarding FCM, algorithms search for minimal values in distributions to define optimal positions for gates to divide populations or search for appropriate cut off values to gate out debris. Furthermore, normalization algorithms can be applied to level out differences due to instrument settings or biological variations in sets of multiple similar data. Many of these algorithms are available in the Bioconductor “flow Core” FCM package implemented in R (3). Secondly, new methods were introduced that go beyond the sequential analysis of two-dimensional plots and base calculations on more parameters of the higher-dimensional space in parallel, which is a crucial need nowadays, when standard cytometers report 10 to 14 parameters per cell and dedicated research instruments up to over 100 parameters. Such algorithms can either substitute conventional strategies, for example, to gate cell populations and read out antigen expression levels or they can be used to extract information from the raw data that is not accessible by conventional gating (4). One of the prominent tasks within an FCM workflow is to define cell populations within a mixture of different cells (“clustering”) that may be of interest for research or diagnosis. AI can directly use higher dimensional data as input for cell clustering or it can perform dimensionality reduction and data visualization, for example, by tSNE or one of its variants (5, 6) or SOM (7), the latter already including some clustering of the data. After dimensionality reduction, population clustering can be added by separate AI algorithms or a human
operator can take over for this task, integrating the output of the dimensionality reduction and conventional gating. Many different algorithms are able to solve the task of clustering in an automated fashion either performing a two-step procedure integrating dimension reduction and subsequential clustering or direct clustering of higher dimensional data; however, as shown in the FlowCAP challenges, results are not unequivocal, especially, if the number of clusters is not defined a priori, and differences remain between different algorithms and human experts. Up to now, no perfect automatic solution for cell clustering exists, although many solutions perform quite well (7). Furthermore, clustering revealing further information on relatedness between populations has been suggested for a multitude of different research questions, for example, cellular developmental trajectories, and has been optimized according to these special tasks (further Ref. in 4). Furthermore, metadata extracted from raw FCM data may also be clustered, for example, in order to define diagnostic or prognostic subgroups (8).

Whereas unsupervised clustering can be helpful for many exploratory research questions to identify cell populations and subpopulations, for medical diagnostic purposes supervised AI methods have been described, that use external information such as diagnoses or outcome to train the AI, for example, using support vector machines or neural networks. All of these strategies rely on a large dataset for training and may incorporate more or less steps from a conventional workflow (4,9,10). Manual gating and tagging of cell populations may be used for training of the AI (11) or AI may be trained using only the final results, that is, diagnosis, as described, for example, in Ref. (12) or in the work by Zhao et al. discussed here. Several AI strategies have been able to discern overt acute myeloid leukemia (AML) from normal samples with a high success rate in the second FlowCap challenge (7), however, this can be a considered a quite simple task, since overt AML is easily characterized by a large abnormal population of blast or sometimes monocytic cells. In contrast, separation of AML from myelodysplastic syndromes or from acute lymphoblastic leukemia, everyday questions in diagnostics, is less trivial.

In contrast to simplified “yes or no” tasks, Zhao et al. tackled a much more realistic question: to deduce a specific diagnosis from FCM panels as they are used in conventional diagnostics. They achieved this goal without an attempt to mimic a conventional human FCM workflow. They transformed the FCM data by self-organizing maps (SOM) and classified these representations by a convolutional neural network (CNN), dealing with each tube separately first and finally with data from all three tubes. The researchers took advantage of a very large database of patient sample FCM data. Data from more than 18,000 samples analyzed in a uniform fashion with identical antibody combinations and more than 200 samples of the rarest subtype of lymphoma could be used to train the CNN. In order to get some insight into the CNN “black box,” they checked, which markers were of most importance for the AI to classify a specific diagnosis correctly and they had cell populations tagged that were detected to be abnormal and discriminative by the algorithm for the respective disease in a way to understand the AI’s decision (and to use this assignment for a possible refinement by a human diagnostician in practical diagnostic use in the future).

As described above, the results of their approach are remarkable, but a problem in sensitivity to detect all true lymphoma cases remains, which is most prominent for follicular lymphoma. Maybe the CNN could be trained in a way, that the correct distinction B-NHL of any type versus normal is assigned a higher weight compared to B-NHL subtyping. If we inspect the importance of single markers for AI performance in Supporting Figure 5, we note that some diagnosis assignments rely heavily on a few markers, whereas other diagnoses seem to rather depend on the distribution of many markers. Interestingly, the latter diagnoses without dependence on dominant markers have the highest rate of falsely being categorized as normal (follicular lymphoma, marginal zone lymphoma, lymphoplasmatic lymphoma). Furthermore, for a human diagnostician, an imbalance of kappa versus lambda light chain expression on B cells is a very important clue for a diagnosis of B-cell lymphoma, whereas the CNN of Zhao et al. does not seem to rely heavily on this information. In a different approach, to detect minimal residual disease in childhood acute leukemia, conventional gating was used to train a machine learning algorithm based on Gaussian mixture models (11). Thus, for the non-AI expert the idea comes up, if some information of a conventional workflow, collected by an automated application, could be “injected” into a CNN algorithm.

If we assume that the problem of sensitivity will be tackled by improved versions in the near future, the AI solution of Zhao et al. will in fact be able to perform at “hematologist-level” and may even deliver B-NHL subtyping competence exceeding the results of conventional FCM alone. However, further problems have to be solved for a broader uptake of such a method: different laboratories work with different antibody panels and even antibodies recognizing the same cluster of differentiation antigen behave differently due to different antibody clones, different fluorochromes and different spill-over from other fluorochromes in the panel. Thus, some methods of knowledge transfer are needed, if we want to avoid starting again with a training sample of more than 10,000 cases for every new antibody panel. If researchers will be able to solve these problems, AI for diagnostic FCM may finally leave the “proof of concept” stage and enter routine diagnostics.

ACKNOWLEDGEMENTS

Open access funding enabled and organized by Projekt DEAL.

LITERATURE CITED

1. Acs B, Ramatalonen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med 2020;288(1):62-81.
2. Gupta A, Harrison PJ, Wieslander H, Pielawski N, Kartasalo K, Partel G, Solorzano L, Suveer A, Klemm AH, Spjuth O, et al. Deep learning in image cytometry: A review. Cytometry A 2019;95(4):366–380.
3. Hahne F, LeMeur N, Brinkman RR, Ellis R, Haaland P, Sarkar D, Spidlen J, Strain E, Gentleman R. flowCore: A Bioconductor package for high throughput flow cytometry. BMC Bioinformatics 2009;10:106.

4. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: Helping to make sense of high-dimensional immunology data. Nat Rev Immunol 2016;16(7):449–462.

5. Szubert B, Cole JE, Monaco C, Drozdov I. Structure-preserving visualisation of high dimensional single-cell datasets. Sci Rep 2019;9(1):8914.

6. Belkina AC, Ciccolella CO, Anno R, Halpert R, Spidlen J, Snyder-Cappione JE. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. Nat Commun 2019;10(1):5415.

7. Van Gassen S, Callebaut B, Van Helden MJ, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytometry A 2015;87(7):636–645.

8. Voigt AP, Brodersen LE, Alerozar TA, Gerbring RB, Mennesen AJ, Wilson ER, Kalswach S, Raimondi SC, Hirsch BA, Gamiis AS, et al. Phenotype in combination with genotype improves outcome prediction in acute myeloid leukemia: A report from Children’s oncology group protocol AAML0531. Haematologica 2017;102(12):2058–2068.

9. Aghaeepour N, Finak G. FlowCAP Consortium. Critical assessment of automated flow cytometry data analysis techniques [published correction appears in Nat Methods. 2013 May;10(5):445]. Nat Methods 2013;10(3):228–238.

10. Aghaeepour N, Chattopadhyay P, Chikina M, Dhaene T, van Gassen S, Kursa M, Lambrecht BN, Malek M, McLachlan GJ, Quan Y, et al. A benchmark for evaluation of algorithms for identification of cellular correlates of clinical outcomes. Cytometry A 2016;89(1):16–21.

11. Reiter M, Diem M, Schumich A, Maurer-Granofszky M, Karawajew L, Rossi JG, Ratei R, Groeneveld-Krentz S, Sajaroff EO, Suhendra S, et al. Automated flow cytometric MRD assessment in childhood acute B - lymphoblastic leukemia using supervised machine learning. Cytometry A 2019;95(9):966–975.

12. Ko BS, Wang YF, Li JL, Li CC, Weng PF, Hsu SC, Hou HA, Huang HH, Yao M, Lin CT, et al. Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome. EBioMedicine 2018;37:94–100.