A machine learning approach for the rapid identification of measurable residual disease in acute myeloid leukemia

Measurable residual disease (MRD) in acute myeloid leukemia (AML), defined as the presence of a quantifiable number of leukemic cells after therapy, is an independent prognostic factor for relapse-free survival and informative in guiding post-remission therapy.[1-3] MRD is monitored with high-sensitivity methods such as molecular techniques (e.g., detection of core-binding factor rearrangements or *NPM1* mutations) or multiparameter flow cytometry (MFC).[1,2,4] The advantages of MFC MRD analysis include its high applicability, sensitivity, and short turnaround time.[5] However, a major disadvantage of using MFC for MRD detection is the need for manual analysis and interpretation of data, which requires extensive knowledge and expertise, and may not be entirely reproducible.[3,5] Machine learning (ML) has shown great promise in the medical field by providing novel methodologies for diagnosis, prognosis, and treatment.[6] ML is broadly defined as the creation of mathematical models to find patterns and relationships in data. For AML MRD detection, ML models have the potential advantages of being objective, reproducible, and fast. Although previous studies have shown the feasibility of both unsupervised and supervised ML in AML MRD analysis,[7-11] their limitations include few fluorochromes in the flow cytometry panel or events analyzed, or a focus on specific leukemia-associated immunophenotypes that are present in only a fraction of cases such as CD7 positivity. Recently, a study explored the addition of an ML model to the MRD workflow as a complementary tool with promising results; however, manual analysis was still required.[12]

We hypothesized that a fully automated ML approach for MRD detection would address these previous limitations and result in a performance that is at least equivalent to that of manual MRD gating, while reducing the cost and time of analysis. For our single-center study, we compared the performance of three ML models: support vector machine (SVM), light gradient-boosting machine (LGBM), and random forest classifier (RFC). These models were compared for their ability to classify each individual cell within patients' samples in a training dataset using a nested cross-validation approach.[13] This was followed by evaluation of the best performing model (RFC) in two independent cohorts of patients representative of real-world clinical cases. Specifically, we tested: (i) whether the immunophenotype of the leukemic cells can be predicted correctly; (ii) whether there is agreement between the percentage of predicted MRD and MRD analyzed manually by experts, and (iii) whether the model's prediction is robust enough to allow us to determine MRD status based on the current clinically accepted cut-off point of 0.1%.[1]

For our study, a total of 212 non-acute promyelocytic leukemia, post-therapy bone marrow aspiration specimens (*Online Supplementary Table S1*), analyzed in an ISO 15189-accredited laboratory according to European LeukemiaNet and EuroFlow guidelines,[1,2,5,14] were selected. All patients or their guardians provided written informed consent according to the Declaration of Helsinki. The study was conducted in accordance with all relevant national ethical regulations and guidelines. Data analysis was blinded regarding the patients' demographics and treatment protocols. Patient inclusion criteria were (i) identical antibody panels, (ii) sufficient cellularity, and (iii) stable fluidics during acquisition. Samples were processed by bulk lysis and subsequently stained with a cocktail of nine fluorochrome-conjugated antibodies: HLA-DR Pacific Blue (Biolegend, cat. #307624), CD45 OC515 (Cytognos, #CYT-45OC), CD38 FITC (BC Life Sciences, #A07778), CD13 PE (BD Life Sciences, #347406), CD34 PerCP-Cy5.5 (Biolegend, #343522), CD117 PE-Cy7 (BC Life Sciences, #B49221), CD33 APC (BD Life Sciences, #345800), CD56 APC-R700 (BD Life Sciences, #565139), and CD19 APC-C750 (Cytognos, #CYT-19AC750-2). For every sample, 1x10^6 nucleated cells were acquired on a BD FACSLyric™ cytometer (BD Life Sciences). Manual analysis, defined as the gold-standard, was performed using the combination of leukemia-associated immunophenotype and different-from-normal approaches[15] with Infinicyt™ software (BD Life Sciences).

To train the ML model, 132 patients were randomly selected from our database (including 104 MRD-negative and 28 MRD-positive samples (Figure 1A). The training dataset was generated using the following approach: for every case, all normal bone marrow populations, as well as residual leukemic cells (if present) were manually gated based on their immunophenotypic profile using Infinicyt™ software (*Online Supplementary Figure S1*). After debris and doublet removal, analyzed files were merged and each population was individually exported as a comma-separated values (CSV) file, containing 12 columns (9 fluorescent and 3 scatter parameters [FSC-A, FSC-H, and SSC-A]) and a variable number of rows (cells). A column was added to annotate the population. A "batch" column was included to divide patients into five batches (stratified according to date of acquisition) for batch effect evaluation and K-fold nested cross-validation. To balance the population classes, a maximum of 1,000,000 cells from abundant populations (e.g., "T- and NK-cells") were randomly selected except for "Residual Leukemic Cells" (to preserve all these cells for training). All files were concatenated, resulting in a single tabular file containing 11,819,872 cells, signal intensities, annotations, and batch numbers (Table 1), which was used for training and nested cross-validation.

A uniform manifold approximation and projection (UMAP) graph was created to explore the training dataset which demonstrated no batch effect (Figure 1B). For every model tested, hyperparameter optimization was performed using StratifiedGroupKFold and GridSearchCV (from the scikit-learn package) on the validation sets (i.e., inner folds). Subsequently, the test sets (i.e., outer folds) were used to assess the performance of the SVM, LGBM, and RFC models, which resulted in average total accuracies of 0.628, 0.877, and 0.914, respectively (N=5) (Table 1, *Online Supplementary Figure S2A*). Among the three models tested, RFC performed superiorly overall with higher accuracy and "Residual Leukemic Cells" F1-score[16] and thus was chosen as the testing model. Final training of the RFC model was performed on the whole dataset, as outlined in the annotated python script provided online (*https://www.github.com/aavhd/AML_MRD_ML*).

Following RFC model training, we tested its performance on two independent cohorts comprising cases not previously "seen" by the model: a first test cohort of 30 samples selected randomly (herein named "retrospective") and a second cohort of 50 consecutively selected patients (named "prospective") to better simulate a real-world setting. The test cohorts constituted actual raw ungated flow cytometry standard (FCS) files selected independently of the training cohort. The retrospective test cohort included 15 MRD-negative and 15 MRD-positive cases with a MRD range of 0.24-84.6% in positive patients. The prospective test cohort comprised 33 MRD-negative and 17 MRD-positive patients with a similar MRD range of 0.18-72.2% in positive cases. Similarly to the training dataset, no batch effect was observed in test cohorts (*Online Supplementary Figure S2B*). The approach to analyze test cases (all automated in our code) is as follows: (i) the raw FCS file is loaded using the FlowIO package, and the spillover matrix is extracted using the FlowUtils package; (ii) event data are compensated, channel numbers are transformed using the FlowCal package, and plain doublets are removed using the FSC-A/FSC-H ratio; (iii) every event is classified by RFC; (iv) white blood cells are selected ("Erythroid Cells" and "Erythroid Precursors" excluded), and population percentages are calculated and saved as a CSV file; (v) finally, a plotting function is called to visualize "Residual Leukemic Cells" based on desired parameters, and the plots are saved as figures (see code for further details).

The RFC model allowed the recognition of aberrant MRD immunophenotypes, indicative of its ability to use relevant structures in data for prediction (Figure 1C). Additionally, to add explainability to the pipeline, we used local interpretable model-agnostic explanations (LIME)[17] to derive the importance given to the most relevant features for a given

Table 1. Cell proportions in the training dataset and classification report of testing for the support vector machine, light gradient-boosting machine and random forest classifier after nested cross-validation.

Population	N of cells in training dataset	Average precision			Average recall			Average F1-score		
		SVM	LGBM	RFC	SVM	LGBM	RFC	SVM	LGBM	RFC
B-cell precursors	807,460	0.716	0.957	0.961	0.860	0.949	0.951	0.754	0.953	0.956
Basophils	153,289	0.738	0.879	0.911	0.040	0.906	0.942	0.073	0.890	0.924
Eosinophils	1,000,000	0.895	0.979	0.989	0.891	0.978	0.987	0.888	0.979	0.988
Erythroid cells	1,000,000	0.988	0.993	0.995	0.766	0.996	0.996	0.802	0.994	0.996
Erythroid precursors	261,226	0.847	0.875	0.908	0.462	0.863	0.898	0.571	0.866	0.901
Mast cells	30,422	0.669	0.595	0.630	0.353	0.438	0.476	0.442	0.480	0.488
Mature B cells	445,746	0.897	0.929	0.925	0.671	0.939	0.919	0.750	0.933	0.921
Mature monocytes	1,000,000	0.625	0.848	0.896	0.824	0.875	0.913	0.661	0.861	0.904
Monoblasts and promonocytes	1,000,000	0.817	0.793	0.813	0.280	0.800	0.882	0.369	0.795	0.843
Myelocytes, metamyelocytes and band cells	1,000,000	0.485	0.879	0.915	0.402	0.895	0.921	0.308	0.886	0.918
Normal myeloid precursors	956,754	0.673	0.714	0.809	0.284	0.863	1.000	0.306	0.779	0.893
Plasma cells	134,750	0.774	0.966	0.970	0.914	0.946	0.957	0.831	0.955	0.964
Plasmacytoid dendritic cells	240,712	0.237	0.736	0.820	0.180	0.768	0.801	0.138	0.748	0.810
Promyelocytes	739,529	0.675	0.892	0.911	0.697	0.897	0.921	0.578	0.893	0.914
Residual leukemic cells	1,049,984	0.388	0.726	0.925	0.641	0.503	0.588	0.392	0.591	0.710
Segmented neutrophils	1,000,000	0.937	0.922	0.937	0.586	0.944	0.959	0.686	0.933	0.948
T and NK cells	1,000,000	0.976	0.987	0.991	0.919	0.992	0.997	0.944	0.990	0.994
Total N of cells	11,819,872	-			Average total accuracy			0.628	0.877	0.914

SVM: support vector machine; LGBM: light gradient-boosting machine; RFC: random forest classifier; NK: natural killer.

prediction; e.g., $CD19^+$ and $CD45^-$ being most important to predict “Mature B-Cells” and “Erythroid Cells”, respectively (Figure 1D). The model’s performance can be interpreted by LIME denoting underlying biological explanations. To evaluate the classification report on a case-by-case basis, ten cases from retrospective and prospective test cohorts (5 each) were selected randomly (*Online Supplementary Figure S2C*). “Erythroid Cells” and “T and NK Cells” showed the best predictions (average F1-scores of 0.993 and 0.990, respectively). “Normal Myeloid Precursors” showed an acceptable average F1-score of 0.618, while “Residual Leukemic Cells” showed inferior performance on the account of being detected in MRD-negative cases (average F1-score of 0.426; range in MRD-positive cases, 0.213-0.929). To further evaluate the strength of agreement in MRD percentage between manual analysis and RFC, a correlation analysis for all cases (N=80, including 64 remission cases) was performed (Figure 1E). The analysis showed good correlation with manual gating for leukemic cell percentage in all cases with a Spearman ρ of 0.74 (0.84 and 0.71 for retrospective and prospective cohorts, respectively, $P<0.0001$). However, the same analysis for only remission cases demonstrated a weak correlation due to detection of residual leukemic cells in cases identified as MRD-negative by manual analysis (Spearman $\rho=0.45$ [0.57 and 0.46 for retrospective and prospective cohorts, respectively, $P=0.0001$]). We think that this is likely due to the misprediction of “Normal Myeloid Precursors” as MRD since the model showed reliable performance in identifying all myeloid precursors (normal and abnormal) in all cases (Spearman $\rho=0.90$ [0.94 and 0.88 for retrospective and prospective cohorts, respectively, $P<0.0001$]), as well as other normal populations (*Online Supplementary Figure S2D*). We think that the difficulty of distinguishing malignant from normal myeloid precursors is related to the lack of specific surface protein markers delineating these two populations in our current flow cytometry panel, as well as subtle immunophenotypic differences in myeloid marker expression patterns, suggesting that the model’s performance could be enhanced with the inclusion of more specific markers such as CD123, TIM3, and CLEC12A.[2,18] The degree to which these markers could improve the results must be studied with new panels.

Our final aim was to evaluate the model’s performance in classifying patients into MRD-positive and MRD-negative groups with the commonly accepted clinical cut-off of 0.1%.[1,2] At this prespecified cut-off, the model predicted most cases to be positive due to assigning “Normal Myeloid Precursors” as “Residual Leukemic Cells” and only two cases were true negatives (Table 2). Upon evaluating the best cut-off point in terms of performance, cut-offs of 0.6% and 0.8% were found to be optimal with areas under the curves of 0.90 and 0.85 in retrospective and prospective test cohorts, respectively (Figure 1F, Table 2). The assessment of agreement in MRD percentage between RFC and manual analysis demonstrated a strong correlation for the proposed cut-off of 0.8% (Spearman $\rho=0.93$, N=26, $P<0.0001$).

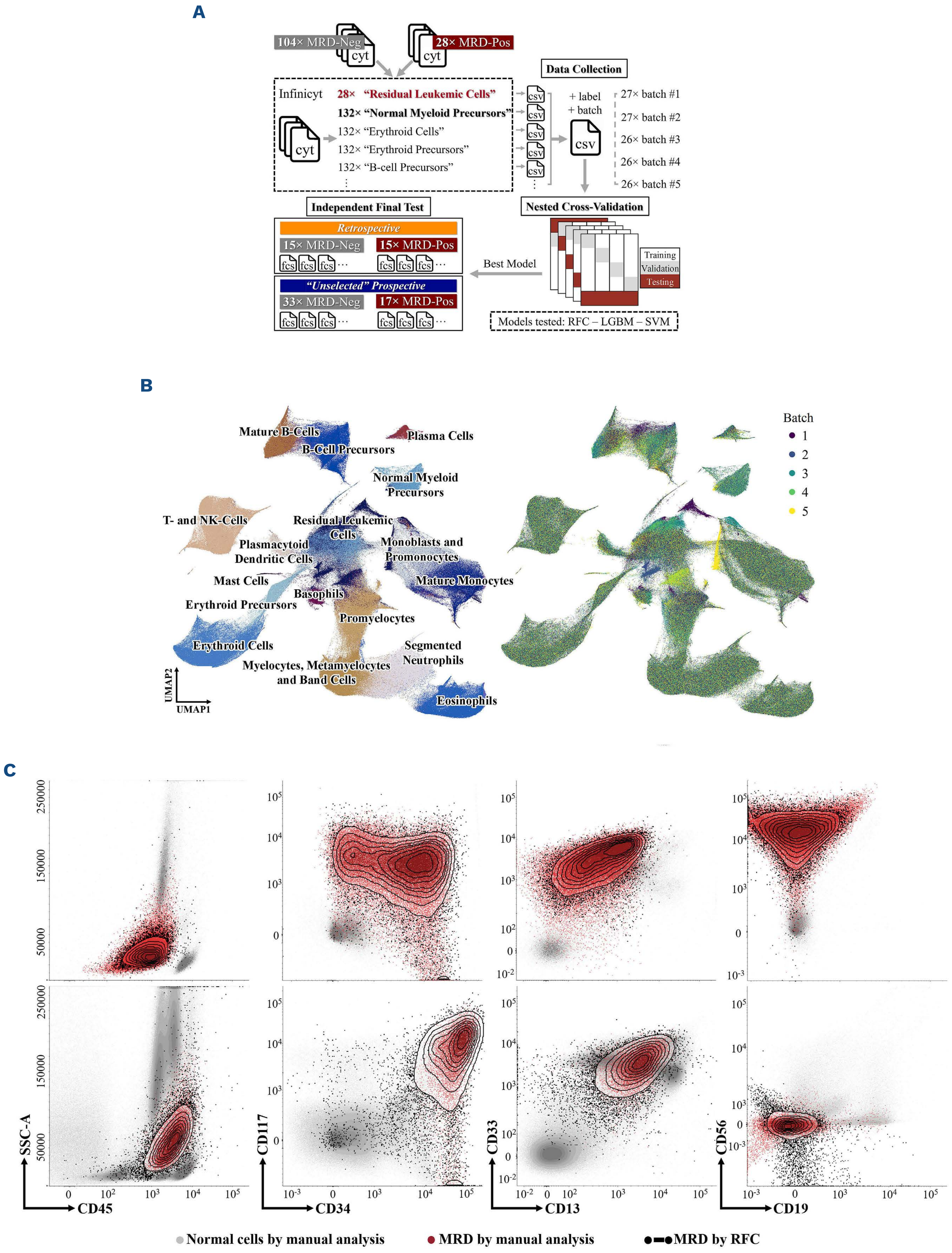
To estimate the clinical benefit of this ML pipeline, we evaluated the frequency of cases in this category at our center. Among 256 consecutive cases that were referred to the laboratory for MRD monitoring in 1 year, 75 (29%) were MRD-positive above the 0.8% level and 26 (10%) had MRD levels between 0.1-0.8%, while the rest were MRD-negative (61%). Regarding the runtime performance for every patient in both cohorts, an average runtime of 3.6 seconds (range, 0.8-4.9) was achieved on a personal laptop with 8 CPU cores. We, therefore, estimate that this model allows reliable triage of ~30% of cases (especially relapsed patients) in seconds, representing a significant time saving.

Altogether, we created an automated ML model to identify and quantify residual leukemia in AML. Our model could reliably detect MRD above the cut-off of 0.8% (sensitivity 82% and specificity 88%) in two independent test cohorts. These values, while encouraging, are above the common

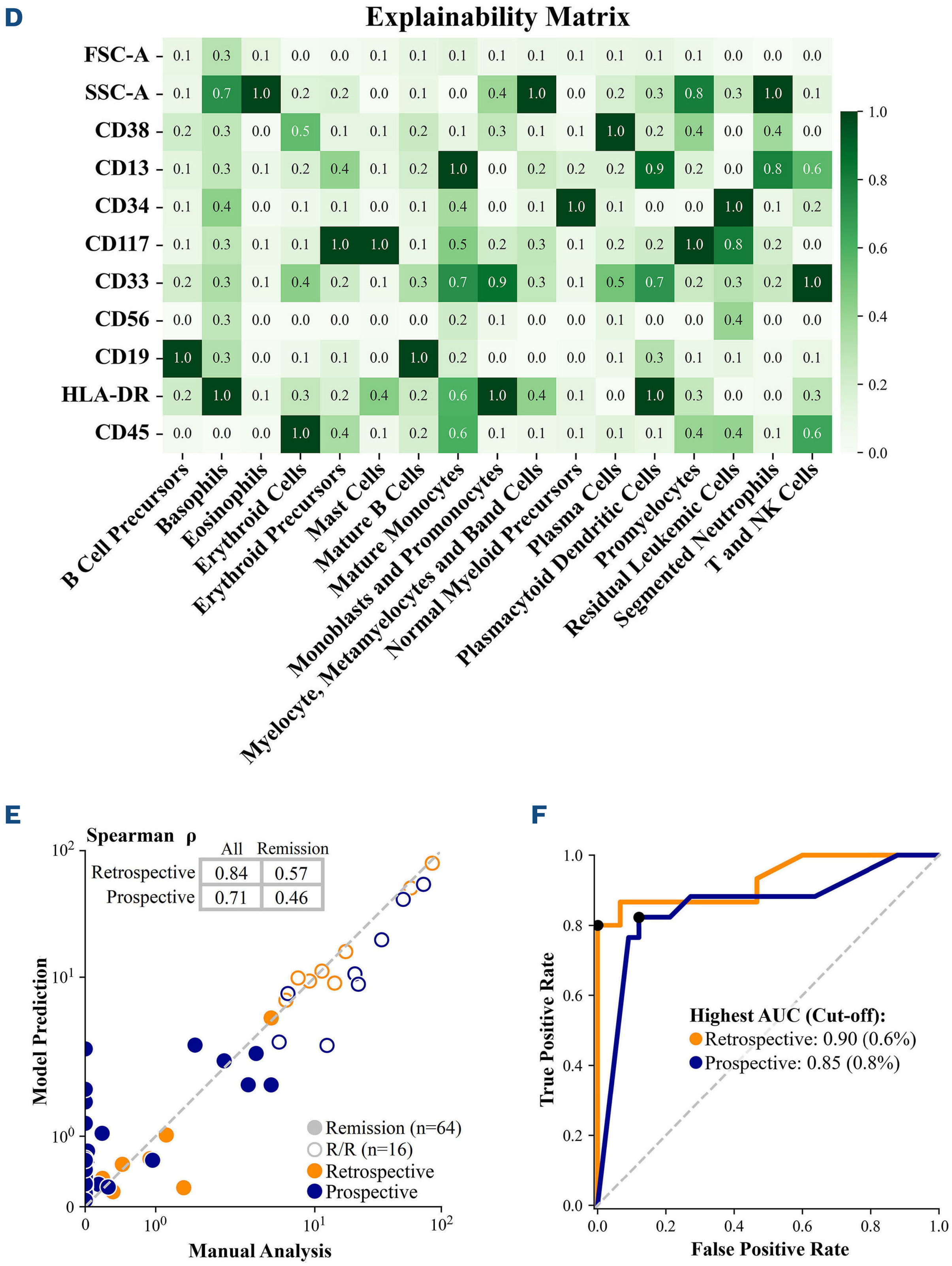
Table 2. Confusion matrix demonstrating the performance of the model with different cut-offs.

		Retrospective test cohort, N=30								Prospective test cohort, N=50			
Cut-off		0.1%*				0.6%				0.8%			
		Manual analysis (gold standard)											
		Negative N=15		Positive N=15		Negative N=15		Positive N=15		Negative N=33		Positive N=17	
RFC	Negative	TN 2		FN 0		TN 15		FN 3		TN 29		FN 3	
	Positive	FP 13		TP 15		FP 0		TP 12		FP 4		TP 14	
		SEN 100%	FPR 87%	FNR 0%	SPC 13%	SEN 80%	FPR 0%	FNR 20%	SPC 100%	SEN 82%	FPR 12%	FNR 18%	SPC 88%
		ACC		57%		ACC		90%		ACC		86%	

*The 0.1% cut-off is only shown for the retrospective test cohort as a similar result (high false-positive rate) was achieved for the prospective patients. N: number; RFC: random forest classifier; TN: true negative; FN: false negative; FP: false positive; TP: true positive; SEN: sensitivity; FPR: false positive rate; FNR: false negative rate; SPC: specificity; ACC: accuracy.



Continued on following page.



	B Cell Precursors	Basophils	Eosinophils	Erythroid Cells	Erythroid Precursors	Mast Cells	Mature B Cells	Mature Monocytes	Monoblasts and Promonocytes	Myelocyte, Metamyelocytes and Band Cells	Normal Myeloid Precursors	Plasma Cells	Plasmacytoid Dendritic Cells	Promyelocytes	Residual Leukemic Cells	Segmented Neutrophils	T and NK Cells
FSC-A	0.1	0.3	0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0
SSC-A	0.1	0.7	1.0	0.2	0.2	0.0	0.1	0.0	0.4	1.0	0.0	0.2	0.3	0.8	0.3	1.0	0.1
CD38	0.2	0.3	0.0	0.5	0.1	0.1	0.2	0.1	0.3	0.1	0.1	1.0	0.2	0.4	0.0	0.4	0.0
CD13	0.1	0.3	0.1	0.2	0.4	0.1	0.2	1.0	0.0	0.2	0.2	0.2	0.9	0.2	0.0	0.8	0.6
CD34	0.1	0.4	0.0	0.1	0.1	0.0	0.1	0.4	0.0	0.1	1.0	0.1	0.0	0.0	1.0	0.1	0.2
CD117	0.1	0.3	0.1	0.1	1.0	1.0	0.1	0.5	0.2	0.3	0.1	0.2	0.2	1.0	0.8	0.2	0.0
CD33	0.2	0.3	0.1	0.4	0.2	0.1	0.3	0.7	0.9	0.3	0.1	0.5	0.7	0.2	0.3	0.2	1.0
CD56	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.1	0.0	0.0	0.4	0.0	0.0
CD19	1.0	0.3	0.0	0.1	0.1	0.0	1.0	0.2	0.0	0.0	0.0	0.1	0.3	0.1	0.1	0.0	0.1
HLA-DR	0.2	1.0	0.1	0.3	0.2	0.4	0.2	0.6	1.0	0.4	0.1	0.0	1.0	0.3	0.0	0.0	0.3
CD45	0.0	0.0	0.0	1.0	0.4	0.1	0.2	0.6	0.1	0.1	0.1	0.1	0.1	0.4	0.4	0.1	0.6

	All	Remission
Retrospective	0.84	0.57
Prospective	0.71	0.46

Figure 1. Performance evaluation of a machine learning approach for identification and quantitation of measurable residual disease by multiparameter flow cytometry in patients with acute myeloid leukemia. (A) A summary of the data collection procedure, schematic nested-cross validation with inner and outer folds, and the composition of the training and the two test sets. (B) An exploratory analysis of the training dataset before the training step via a down-sampled uniform manifold approximation and projection (UMAP) (10^6 cells) constructed with nine fluorescence and two main scatter parameters, demonstrating all cellular subsets present in bone marrow (left). A simplified hematopoietic maturation is inferable in myeloid lineages indicative of the dataset's biological explainability. Unlike other cell populations demonstrating uniform colors, "Residual Leukemic Cells" (dark blue) exhibit heterogeneous immunophenotypes due to their similarity with and/or differentiation toward different cell populations including common myeloid progenitors, neutrophilic precursors, and monocytic precursors. Furthermore, analysis of batch distribution (acquisition date) excludes the possibility of batch effects in the dataset (right). The UMAP was created using the umap-learn package with number of components = 2, number of neighbors = 15, and minimum distance = 0.1. (C) The comparison between phenotypes defined by manual analysis and predicted by the random forest classifier (RFC) model. Each row represents one measurable residual disease (MRD)-positive patient from the retrospective test cohort. Normal cells and MRD identified by manual analysis are shown in gray and maroon, respectively. Briefly, the immunophenotype defined for the first case included heterogeneous CD34, decreased CD13, increased CD33 and aberrant CD56 expression. The immunophenotype in the second case was defined as increased CD45, increased side scatter, and bright CD34 expression. Black contour lines and dots represent the RFC prediction. (D) A heatmap representing the explainability matrix for the predicted populations using the initially trained mod-

Continued on following page.

el for the patient shown in (C) upper row as an example. Darker color depicts higher importance. Feature importance can be in both directions e.g., CD19-positivity for "Mature B-Cells" or CD45-negativity for "Erythroid Cells", as well as ranges of signal values. (E) The Spearman correlation analysis of MRD percentage between manual analysis and RFC for all cases, as well as only remission cases of the retrospective and prospective test cohorts (*P*<0.0001). Filled circles represent the remission cases. The light line shows perfect correlation. (F) Receiver operating characteristic curve of different MRD cut-offs in the retrospective (N=30) and prospective (N=50) test cohorts. The 0.6% and 0.8% cut-offs (black circles) show the greatest areas under the curve for the retrospective and prospective test cohorts, respectively. Statistical analyses were performed with SciPy and scikit-learn packages. Figures were created with seaborn and Matplotlib packages in python, Infinicyt™, and FlowJo™ (BD Life Sciences). Neg: negative; Pos: positive; cyt: Infinicyt analyzed file; fcs: flow cytometry standard file; csv: comma-separated values file; RFC: random forest classifier; LGBM: light gradient-boosting machine; SVM: support vector machine; NK: natural killer; SSC: side scatter; CD: cluster of differentiation; FSC: forward scatter; R/R: relapsed/refractory; AUC: area under the curve.

0.1% clinical cut-off, meaning that cases with MRD reported below 0.8% using this approach still need to be analyzed manually. This ML pipeline (publicly available at *https://www.github.com/aavhd/AML_MRD_ML*) is written in a way such that it can be used for training different panels with various numbers of fluorochromes, even for other purposes, such as assessing B- or T-acute lymphoblastic leukemia MRD. Unlike other AML MRD detection algorithms,[10-12] this automated approach does not require clustering or dimensionality reduction steps, while achieving promising performance in two independent test cohorts. As for every ML approach, the quality of the acquisition process, such as fluidics stability, must be optimal for accurate model results. In addition, every laboratory requires a validated workflow and quality-controlled flow cytometers to collect a dataset specific to their panel and train their model accordingly.

Overall, our work constitutes an important contribution to the development of a fully automated ML approach in AML MFC residual leukemia monitoring. The model's performance might be improved by increasing the number of training cases, exploring larger hyperparameter spaces, incorporating diagnostic information into the training data, and adding markers that would distinguish normal myeloid progenitors and residual leukemia. Future prospective studies with more patients and multicenter validation will be informative in this regard.

Authors

Amirali Vahedi,[1,2] Mohammadreza Royaei,[3] Tahereh Madani,[3] François Mercier[1,2#] and Behzad Poopak[3,4#]

[1]Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, Canada; [2]Lady Davis Institute for Medical Research, Montreal, Canada; [3]Payvand Clinical, Specialty, Pathology, Medical Genetics and Molecular Laboratory, Tehran, Iran and [4]Hematology Department, School of Paramedics, Islamic Azad University, Tehran Medical Sciences, Tehran, Iran

[#]FM and BP contributed equally as senior authors.

Correspondence:
F. MERCIER - francois.mercier@mcgill.ca
B, POOPAK - bpoopak@gmail.com

https://doi.org/10.3324/haematol.2024.286019

Received: June 19, 2024.
Accepted: February 24, 2025.
Early view: March 6, 2025.

Disclosures

No conflicts of interest to disclose.

Contributions

BP, FM and AV designed the study. AV analyzed and interpreted the data and performed the statistical analyses. FM and AV wrote the manuscript. MR collected the data. TM critically reviewed the manuscript. All authors approved the last version of the manuscript.

Acknowledgments

Computational analyses were supported by the Digital Research Alliance of Canada. The authors would like to thank the flow cytometry department personnel at Payvand Clinical and Specialty Laboratory.

Funding

Work in FM's laboratory was supported by the Cole Foundation and the Sir Mortimer B. Davis Foundation of the Jewish General Hospital. FM is a Clinical Research Scholar – Junior 2 of the Fonds de Recherche du Québec – Santé.

Data-sharing statement

The entire pipeline used in this study accompanied by instructions for use can be found at *https://www.github.com/aavhd/AML_MRD_ML*. An example training dataset and an anonymized test patient are also provided.

References

1. Heuser M, Freeman SD, Ossenkoppele GJ, et al. 2021 Update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2021;138(26):2753-2767.
2. Schuurhuis GJ, Heuser M, Freeman S, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2018;131(12):1275-1291.
3. Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukaemia. Leukemia. 2017;31(7):1482-1490.
4. WHO. Classification of Tumours Editorial Board. Haematolymphoid tumours. 5th ed. Lyon (France): International Agency for Research on Cancer. 2024.
5. Tettero JM, Freeman S, Buecklein V, et al. Technical aspects of flow cytometry-based measurable residual disease quantification in acute myeloid leukemia: experience of the European LeukemiaNet MRD Working Party. Hemasphere. 2022;6(1):e676.
6. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347-1358.
7. Licandro R, Reiter M, Diem M, Dworzak M, Schumich A, Kampel M. Application of machine learning for automatic MRD assessment in paediatric acute myeloid leukaemia. Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods ICPRAM. 2018;1:401-408.
8. Ni W, Hu B, Zheng C, et al. Automated analysis of acute myeloid leukemia minimal residual disease using a support vector machine. Oncotarget. 2016;7(44):71915-71921.
9. Ko BS, Wang YF, Li JL, 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;37:91-100.
10. Vial JP, Lechevalier N, Lacombe F, et al. Unsupervised flow cytometry analysis allows for an accurate identification of minimal residual disease assessment in acute myeloid leukemia. Cancers (Basel). 2021;13(4):629.
11. Weijler L, Kowarsch F, Wodlinger M, et al. UMAP based anomaly detection for minimal residual disease quantification within acute myeloid leukemia. Cancers (Basel). 2022;14(4):898.
12. Shopsowitz K, Lofroth J, Chan G, et al. MAGIC-DR: an interpretable machine-learning guided approach for acute myeloid leukemia measurable residual disease analysis. Cytometry B Clin Cytom. 2024;106(4):239-351.
13. Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nat Methods. 2016;13(9):703-704.
14. Kalina T, Flores-Montero J, van der Velden VH, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia. 2012;26(9):1986-2010.
15. Wood BL. Acute myeloid leukemia minimal residual disease detection: the difference from normal approach. Curr Protoc Cytom. 2020;93(1):e73.
16. Nazha A, Elemento O, McWeeney S, Miles M, Haferlach T. How I read an article that uses machine learning methods. Blood Adv. 2023;7(16):4550-4554.
17. Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?”: explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA. Association for Computing Machinery. 2016:1135-1144.
18. Haubner S, Perna F, Kohnke T, et al. Coexpression profile of leukemic stem cell markers for combinatorial targeted therapy in AML. Leukemia. 2019;33(1):64-74.