Received: 21 January 2025 | Accepted: 8 April 2025

DOI: 10.1002/hem3.70138

HemaSphere eha

REVIEW ARTICLE

Applications of machine learning for immunophenotypic measurable residual disease assessment in acute myeloid leukemia

Tim R. Mocking[1,2] | **Arjan A. van de Loosdrecht**[1,2] | **Jacqueline Cloos**[1,2] | **Costa Bachas**[1,2]

Correspondence: Costa Bachas (c.bachas@amsterdamumc.nl)

Abstract

Immunophenotypic detection and quantification of residual leukemic cells by multiparameter flow cytometry is increasingly adopted in the clinical practice of acute myeloid leukemia (AML) to assess measurable residual disease (MRD). However, MRD levels quantified by manual gating analysis can differ based on differences in gating strategy between trained operators and clinical centers. Manual gating requires extensive training, is time-consuming in daily practice, and faces a significant hurdle in analyzing data from next-generation cytometry platforms. To address these challenges, several computational approaches involving machine learning and artificial intelligence algorithms have been proposed to automate or aid the assessment of MRD. However, the immunophenotypic variability between patients and the relatively low proportions of residual leukemic cells in AML challenge most algorithms and require innovative approaches. This review provides an overview of recent efforts in using computational methods for immunophenotypic AML-MRD assessment. We first explain the technical and conceptual background of the different algorithms that have been explored. Next, we discuss their strengths and limitations in the disease-specific context of AML. Finally, we highlight how computational approaches offer a unique opportunity to standardize or even outperform current manual gating analyses, and ultimately, improve the treatment of AML patients.

INTRODUCTION

In hematological malignancies, measurable residual disease (MRD) assessment allows for detecting residual leukemic cells after therapy below the detection limits of microscopy-based morphology assessment. In acute myeloid leukemia (AML), MRD has emerged as an important prognostic factor[1] and is increasingly used to evaluate the depth of remission and disease kinetics and as a surrogate endpoint in clinical trials.[2] In the ELN-intermediate risk group, several centers also estimate the risk of relapse based on MRD to decide between intensive and less intensive consolidation therapy.[2] Immunophenotypic MRD assessment via multiparameter flow cytometry (MFC)[3,4] has proven particularly useful based on its broad applicability (85%–90% of AML patients), high throughput (millions of cells), and relatively low cost.

Immunophenotypic MRD assessment relies on the identification of leukemia-associated immunophenotypes (LAIPs) defined by specific aberrant expression patterns of cluster of differentiation markers (CD),[3,4] which are not, or at low frequency, present on healthy cells. This is done by manual gating, in which experts evaluate these patterns based on a sequential analysis of two-dimensional CD marker plots. In most leukemias, LAIPs are relatively well-defined and are evaluated in clinical practice using limited numbers of markers. For example, CD10 is aberrantly expressed in more than 70% of B-cell precursor acute lymphoblastic leukemia patients.[5] In chronic lymphocytic leukemia, six markers (CD19, CD20, CD5, CD43, CD79b, and CD81) are sufficient for MRD detection in >95% of patients.[6] In multiple myeloma, a single 10-marker assay can be used for MRD assessment in 99% of cases.[7] In contrast, MRD assessment in AML has proven uniquely challenging due to the heterogeneity of LAIPs. Figure 1 exemplifies this heterogeneity for LAIPs identified by our four-tube, eight-color assay in bone marrow samples of AML patients who were in complete remission after two cycles of therapy (Supporting Information S1). The inter- and intra-patient heterogeneity of LAIPs instigates that immunophenotypic AML-MRD assessment by

[1]Department of Hematology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2]Cancer Center Amsterdam, Imaging and Biomarkers, Amsterdam, The Netherlands

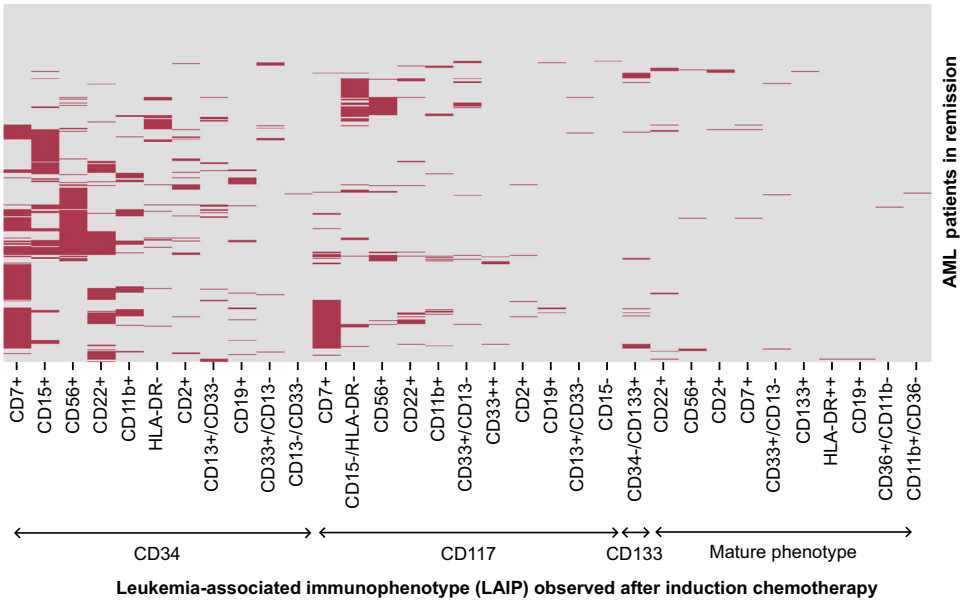

FIGURE 1 **Heterogeneity of leukemia-associated immunophenotypes (LAIPs) identified after two cycles of chemotherapy in AML patients (n = 455) in the HOVON-SAKK-132 trial.** A binary representation of observed (red) or not-observed (gray) LAIPs was used for simplicity. Patients (rows) were hierarchically clustered based on the common LAIPs between rows (Jaccard index). LAIPs (columns) were organized according to the primitive marker (CD34, CD117, CD133, or mature) first and by frequency second.

manual gating is time-consuming in routine diagnostics, requiring flow cytometry experts with extensive knowledge about healthy, regenerating, and malignant differentiation patterns. Another hurdle to the broad implementation of immunophenotypic MRD assessment is the subjective aspect of manual gating. Although standardization efforts can enable consistent MRD percentages between operators,[8] variations in the gating strategy and gate placement can still result in different MRD results between clinical centers. These challenges are further exacerbated for the analysis of novel high-dimensional panels developed for spectral cytometry platforms,[9–11] as the number of possible two-dimensional plots increases exponentially with the number of markers measured. Together, the limitations of manual gating limit the reproducibility of AML-MRD results and hamper broad implementation beyond specialized clinical centers.

Over the last decade, machine learning and artificial intelligence have proven their use in providing fast, objective, and reproducible MFC data analysis,[8,9] including those relevant to hematological malignancies.[12] In contrast to manual gating, computational analysis considers MFC data in tabular format with rows (cells/events) and columns (markers). Consequently, these algorithms can consider the multi-dimensional nature of MFC data to assess single cell or population characteristics, either with (supervised) or without (unsupervised) accompanying human input. This has already been proven effective for AML: in 2011, 12 out of 25 tested approaches at the FlowCAP-II Summit could distinguish between AML diagnosis and non-AML samples with perfect accuracy.[13] However, the quantitative and phenotypic rarity of leukemic cells after chemotherapy in the bone marrow of AML patients provides challenges for most algorithms and requires innovative approaches.

The above-illustrated biological and clinical characteristics of AML, combined with the limitations of current analysis procedures and technical advancements in cytometry, urge a focused discussion on the current state of computational AML-MRD assessment. In this review, we first introduce the theoretical background of computational AML-MRD, illustrating how leukemic cells can be classified by supervised and unsupervised machine learning. Next, we outline how different studies (Table 1) have tried to tackle two critical challenges specific to AML-MRD: the rarity of leukemic cells and the heterogeneity of LAIPs. Lastly, we discuss the trade-offs between computer-assisted and more automated approaches and how computational analysis can objectify or augment the current clinical analysis of AML.

THEORETICAL BACKGROUND OF COMPUTATIONAL AML-MRD

To illustrate how different machine learning approaches can be used to identify leukemic cell populations in AML, four different samples of normal bone marrow (NBM), AML1, AML2, and AML3, with two hypothetical LAIP markers (LAIP1 and LAIP2) are simulated (Figure 2A). In this simplified setting, the differences between healthy and leukemic cells are assumed to be limited to only the expression of these two LAIP markers. The NBM sample contains a single double-negative non-leukemic cluster. In AML1 and AML2, a subset of cells belongs to a leukemic cluster (LAIP1+ cells) with different proportions (10% and 1%). In AML3, a different phenotype is introduced (LAIP2+ cells, 10% of cells).

Supervised classification of leukemic cells

Supervised models are trained on labeled training data (e.g., individual cells gated as LAIP+/LAIP−) to classify new unlabeled test data. To exemplify this, a basic supervised algorithm based on the k-nearest neighbors (kNN) algorithm using AML1 as training data are used to classify new samples (Figure 2B). In AML test samples (AML2 and AML3), cells are classified as leukemic if three out of five of its most similar (i.e., closest or "nearest neighbor" cells) cells

TABLE 1 Studies applying machine learning for identifying leukemic features in cytometry data from AML patients.

Study	Algorithm(s)	Human element	Description
Ni et al. (2016)[14]	SVM	Gating of diagnosis sample	Supervised algorithm trained on diagnosis gates for automated analysis at follow-up
Rajwa et al. (2017)[15]	ASPIRE (DPM)	None	AML vs. normal model used for prediction of disease progression
Arvaniti and Claassen (2017)[16]	CellCnn	None	Neural network to identify cells associated with sample-level labels
Ko et al. (2018)[17]	GMM + SVM	None	AML vs. normal model applied in AML follow-up setting
Licandro et al. (2018)[18]	WGAN + NN	None	Dimensionality reduction followed by supervised modeling for blast identification at diagnosis
Licandro et al. (2018)[19]	RF, SVM, GMM	None	Benchmark of supervised (RF, SVM) and novelty detection (GMM) approaches at diagnosis
Weber et al. (2019)[20]	diffcyt	None	Differential analysis (diffcyt) to identify rare spike-in AML cells
Lacombe et al. (2019)[21]	FlowSOM	Cluster selection, quality control	Cluster-with-normal pipeline to highlight suspected leukemic clusters
Jacqmin et al. (2020)[22]	kNN clustering, KDE	Gating of diagnosis sample	Computer-assisted gating of diagnosis sample leveraged for automated analysis at follow-up
Vial et al. (2021)[23]	FlowSOM	Cluster selection, quality control	Cluster-with-normal pipeline to highlight suspected leukemic clusters
Craddock et al. (2021)[24]	Phenograph, FlowSOM	Pre-gating	Cluster-with-normal pipeline to highlight suspected leukemic clusters
Weijler et al. (2022)[25]	Set-Transformer, UMAP, HDBSCAN	None	Fully automated cluster-with-normal pipeline
Fokken et al. (2024)[9]	Cen-se′	Pre-gating	Semi-automated pipeline for spectral cytometry to highlight suspected leukemic clusters
Matthes (2024)[11]	t-SNE, FlowSOM	Pre-gating	Cluster-with-normal pipeline for spectral cytometry data to highlight suspected leukemic clusters.
Weijler et al. (2024)[26]	FATE	None	Transformer neural network pre-trained on ALL for MRD detection in AML
Shopsowitz et al. (2024)[27]	XGBoost, UMAP	Pre-gating	Supervised predictions shown in UMAP space to highlight leukemic cells
McCarthy et al. (2024)[28]	FlowSOM	Pre-gating	Cluster-with-normal pipeline to highlight suspected leukemic clusters
Mocking et al. (2024)[29]	GMM	None	Fully automated pipeline
Gachon et al. (2025)[30]	Optimal transport	None	Sample-level classification using optimal transport
Vahedi et al. (2025)[31]	Random forests	None	Supervised multi-class model to reproduce extensive gating hierarchy including leukemic blasts

in the training data (AML1) are leukemic. Because this classification is applied to each cell independently, it is not affected by the rarity of leukemic cells in the test cases, as evidenced by the detection of rare leukemic cells in AML2. However, it is constrained to the phenotypes in the training data: the LAIP2+ cluster was not observed and was not detected.

Shopsowitz et al.[24] recently demonstrated supervised cell classification by manually gating 86 AML diagnosis samples and 30 MRD-negative control samples to train a cell-level classification model. While this model was able to highlight leukemic cells, the authors also reported poor performance in identifying two abnormal clusters with unique phenotypes compared to the training data, similar to the output for LAIP2 in our simulated example. To cover all possible LAIPs (Figure 1), training algorithms for supervised classification of LAIP+ cells will demand extensive training data, further complicated by batch effects, evolving antibody panels and cytometry platforms.

Detecting leukemic cells using unsupervised clustering

In unsupervised approaches, prior knowledge of leukemic features is not used to train algorithms. Instead, cells are grouped (i.e., clustered) in a data-driven fashion based on their immunophenotypic similarity. A common method to identify leukemic clusters is the "cluster-with-normal" approach in which a sample-level, rather than cell-level labeling, is used. In this "semi-supervised" setting, cells from AML samples are clustered with a set of non-leukemic reference or control samples, assuming that the leukemic characteristics create clusters enriched for cells from the AML test sample. Because most popular clustering algorithms, such as FlowSOM, require a pre-specified number of clusters, which is hard to determine objectively, studies often use over-clustering to capture as many different cell populations as possible. To demonstrate this approach, cells from the NBM sample were clustered with cells from AML1, AML2, and AML3 (Figure 2C). Cells were divided into six clusters using the k-means clustering algorithm. Next, each

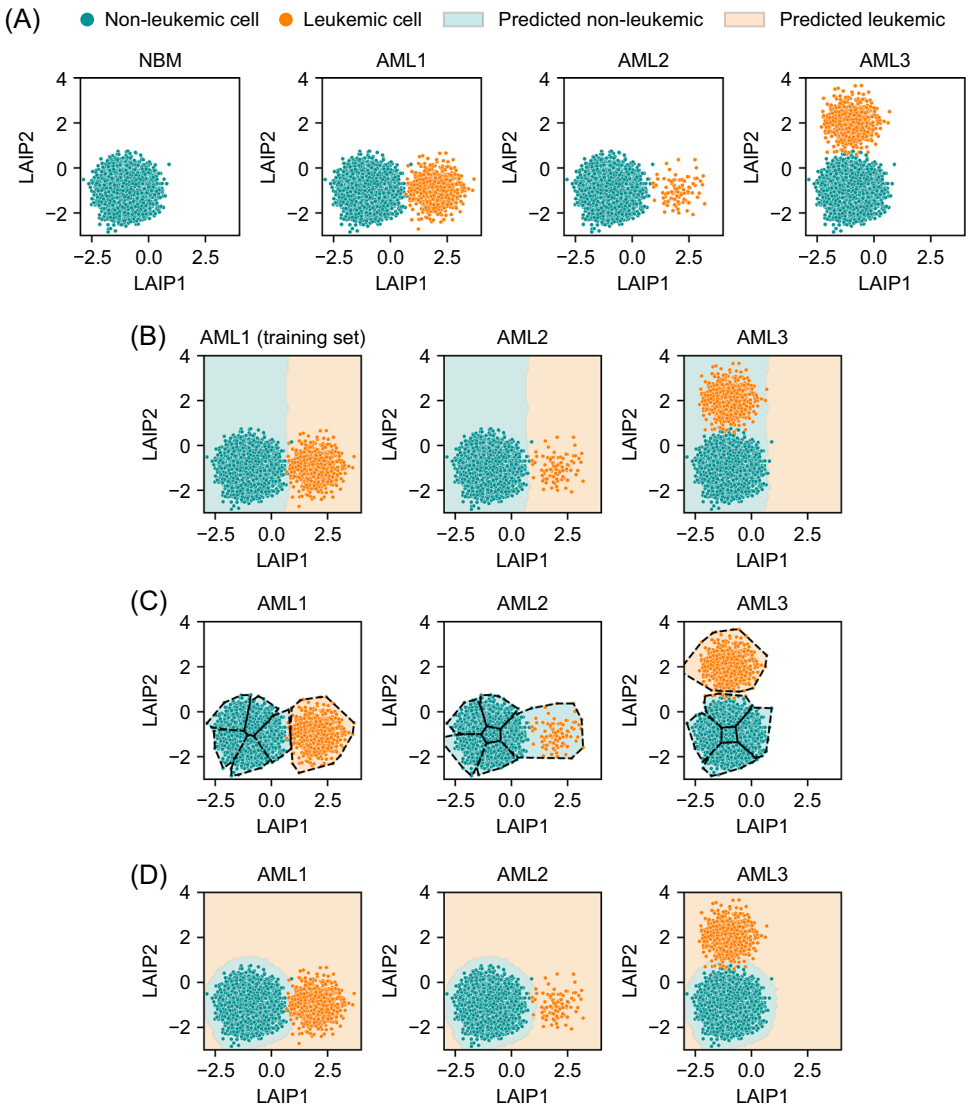

FIGURE 2 **Theoretical overview of computational approaches for AML-MRD assessment. (A)** Four samples are simulated for two hypothetical leukemic markers (LAIP1 and LAIP2). **(B)** Supervised classification of the three AML samples using AML1 as training data. Cells are classified as leukemic if the majority of its five closest (i.e., "nearest neighbor") cells in AML1 are leukemic. **(C)** Semi-supervised "cluster-with-normal" approach using $K = 5$. K means clusters. Cells in a cluster are classified as leukemic if more than 95% of cells belong to the test sample. **(D)** Semi-supervised novelty detection approach. Cells are classified as leukemic if the median distance to its five closest (i.e., "nearest neighbor") cells exceeds a certain distance (0.5). An interactive environment online to explore this simulation is available at: https://github.com/AUMC-HEMA/cMRD-review.

cluster was classified as leukemic if the cells from test samples comprised more than 95% of cells in a cluster. Because the cluster-with-normal approach does not leverage leukemic training data, it can detect the different phenotypes of AML1 and AML2. However, it is dependent on the clustering resolution. Failing to achieve a good clustering may fail to detect rare subsets, exemplified by the poor performance in AML2.

Unsupervised clustering-based methods for MRD detection feature most prominently in computer-assisted approaches (Table 1), which often use FlowSOM to highlight aberrant clusters.[21,23,24,28] However, other algorithms such as HDBSCAN,[25] PhenoGraph,[24] and KNN-based clustering[22] have also been evaluated.

Detecting leukemic cells using novelty detection

Lastly, leukemic cells can be classified based on their dissimilarity to NBM without clustering in an approach called novelty detection. In Figure 2D, each of the cells in AML1-3 was classified as leukemic if its five closest cells in NBM were above a certain distance. Like the supervised classifier, each cell is evaluated independently, resulting in good performance in rare cell detection (AML2). It also lacks prior knowledge of leukemic phenotypes, detecting the different LAIPs of AML1 and AML2. However, this approach will classify all outliers in test data as leukemic, which may include inherent levels of "noise" or background, such as autofluorescence, potentially resulting in a high false-positive rate.

Despite its common use in the broader anomaly detection field, novelty detection remains relatively unexplored in the (AML-)MRD setting. Novelty detection based on Gaussian mixture models (GMMs) can yield accurate MRD results.[29] Although this can mitigate some of the downsides of ad hoc clustering, future model refinements are still required to address potential false positives.

The above simulations demonstrate the theoretical strengths and weaknesses of different machine learning approaches in a simplified

setting. However, this experiment cannot draw strong conclusions, as its output depends on many parameters (e.g., inter-cluster distance, cell rarity, and model parameters). We provide an interactive environment online to explore these different approaches: https://github.com/AUMC-HEMA/cMRD-review.

IMPROVING UNSUPERVISED RARE CELL DETECTION IN AML

A key challenge in identifying residual leukemic cells is their rarity. At the recommended clinical cut-off of 0.1%,[32] a cluster of only 500 leukemic cells among 500,000 WBCs is already sufficient for MRD positivity. MRD levels below the 0.1% cut-off can also be prognostic, demonstrated by patient outcomes based on the limit of detection (LOD) or limit of quantification (LOQ) established at 20 or 50 clustering events, respectively.[33] Hence, to maintain proper sensitivity, computational methods should be able to process at least 500,000–1,000,000 WBCs.[34] While computational resources have been mentioned as constraints in older literature,[14] most approaches are nowadays feasible by dedicated hardware or by through more efficient implementations and alternatives of existing algorithms.[35] However, rare cells can nevertheless be challenging to detect by computational means. While trained supervised algorithms evaluate cells independently in test data, unsupervised algorithms typically rely on relative density to define clustered points.[36] This approach can fail to distinguish rare cell populations from background noise or larger non-leukemic clusters without prior knowledge, as demonstrated theoretically in Figure 2C. Also, leukemic cells with subtle differences from healthy regenerating progenitors may be erroneously grouped, hindering detection. Several solutions have been proposed to mitigate or overcome the issue of rare cells in unsupervised analyses.

One approach, mirroring pre-selection steps in manual gating, focuses on using subsets enriched for leukemic cells as input for unsupervised analysis. This process already starts in a pre-analytical phase, where outlier events such as aggregates, doublets, debris, and artifacts due to flow turbulence should be removed. Computational approaches can also be implemented to automate and objectify this process[37–39] and have been discussed in more depth elsewhere.[40] To enrich for cell populations of interest, Weijler et al.,[25] Arvaniti and Claassen,[16] and Weber et al.[20] all utilized pre-selected CD34+ myeloid progenitors as input for unsupervised analyses. Limiting the analyses to this subset (~2% of white blood cells in bone marrow aspirates) can improve the performance of rare cell detection. Weijler et al.[25] quantified this benefit by comparing model performance between raw and pre-filtered input cells. They also developed an automated approach for performing this CD34+ cell selection, eliminating the need for potentially subjective manual pre-selection. However, these approaches must account for heterogeneity across AML subtypes (e.g., CD34− phenotypes frequently observed for NPM1+ AML patients) to avoid leaving out leukemic cells during this step. Arvaniti and Claassen[16] aimed for an alternative, data-driven enrichment for rare leukemic cells and found increased performance when putting more weight on outlier cells during model training. Additionally, a recent algorithm called SuperCellCyto[41] has shown improved detection of rare cell populations in unsupervised settings by aggregating cells with similar phenotypic characteristics into "supercells" that lower the size of the dataset while retaining the biological signal. This approach proved beneficial compared to random sampling, often required when running computationally intensive algorithms, but has not been evaluated specifically for (AML-)MRD.

Another promising strategy to improve performance with unsupervised algorithms involves using prior knowledge from the diagnosis timepoint, analogous to the so-called LAIP approach for AML-MRD gating, in which features of leukemic cells at diagnosis are used to track cells with similar features in follow-up samples.[3] Jacqmin et al.[22] described a computer-assisted method for identifying leukemic populations at diagnosis, using personalized gates based on clustering of a diagnosis sample to detect small leukemic populations at follow-up. While this approach circumvents the issue of clustering rare cells at follow-up, its ability to adapt to changes in CD marker expression (e.g., due to clonal evolution) remains unclear. A more popular approach is to cluster cells from the diagnosis and follow-up point together, giving leukemic clusters more "weight" based on the abundance of leukemic cells at diagnosis and allowing for leukemic marker expression changes at follow-up. This approach has been most prominently featured in several computer-assisted approaches utilizing the FlowSOM algorithm.[11,21,23,24,28]

In terms of modeling, over-clustering is often recommended as a popular technique for improving sensitivity for rare cells.[36,42] Using many clusters, this approach ensures the detection of as many cell populations as possible. Weber et al.[20] demonstrated this method using 400 clusters to identify a 1% spiked-in AML cluster among control cells. However, the need for such a high number of clusters to detect leukemic cells underscores the challenges of using unsupervised approaches, especially when compared to a limited set of gates. In particular, over-clustering may also present difficulties in computer-assisted settings due to the time-consuming nature of evaluating each cluster for aberrant expression patterns.

MODELING OF HETEROGENEOUS AML PHENOTYPES

AML is notorious for its heterogeneity of LAIPs (Figure 1). This poses significant challenges for supervised learning of leukemic characteristics, particularly given the limited availability of high-quality AML(-MRD) cytometry data. Several studies have proposed methods to navigate or adapt to the inherent heterogeneity of expression patterns in AML data to address this challenge.

One promising approach focuses on classifying leukemic cells based on the absence of normal features rather than the presence of predefined leukemic features. This is similar to the rationale underlying MFC assays, which classify patterns as leukemic or non-leukemic based on observations from normal and post-therapy hematopoiesis.[3,4] Computational MRD approaches offer a unique opportunity to objectively assess and quantify these patterns in a data-driven fashion. Unsupervised clustering has emerged as the most frequently explored method (Table 1). In Figure 2C, we illustrate how pooling an AML test sample with non-leukemic controls can highlight clusters enriched for leukemic cells. A computer-assisted setting can evaluate individual suspected clusters in conventional two-dimensional plots to classify leukemic clusters. Other studies have further objectified this classification, for example, by calculating a ratio[22,23] or percentage[25] of patient to control cells within each cluster. Alternatively, Craddock et al.[24] calculated a reference range based on the control samples to identify candidate leukemic clusters using confidence intervals defined by the control cluster abundances and marker distributions. Additional constraints on candidate aberrant clusters have also been proposed, such as a minimum cell count,[28] minimum percentage of immature (healthy) myeloid progenitors,[23,24] or a maximum frequency in control samples.[23,24] However, cluster enrichment scores ultimately depend on the mixture of patient to control cells in the pooled sample setting frequently used for clustering. For clustering these pooled samples, it is also essential to ensure that individual samples with higher cell counts do

not dominate the clustering. To prevent this, random sampling is often used to create a pooled sample containing an equal contribution of cells from all samples.

Leukemic cells can also be identified accurately based on control samples without clustering these samples together with a test sample[29] in an approach called novelty detection, also shown in Figure 2D. This "semi-supervised" approach does not suffer from the requirement of maintaining sample balances during clustering and evaluates cells independently, similar to supervised approaches, ensuring high sensitivity for rare cells. The fact that a limited ($n = 18$) control set already showed high concordance (97%) with MRD status based on manual gating demonstrates that a large AML training cohort capturing all phenotypes is not required for computational AML-MRD assessment. This can benefit laboratories with lower patient admissions, which would otherwise require unrealistically large AML cohorts to train in-house supervised algorithms.

A common denominator between studies that have attempted to overcome the heterogeneity of AML is the use of (non-leukemic) control data against which to "contrast" leukemic cells. For this purpose, different sample types such as MRD negative AML,[17,25–27,29] non-AML control samples including acute lymphoblastic leukemia[18,19,25] lymphoma,[15] and samples from healthy donors[14–17,20–25] have been used. The definition of appropriate control data is ultimately contingent upon the clinical context, given that non-leukemic expression patterns may vary between samples obtained directly after induction chemotherapy, shortly before consolidation treatment, and in more longitudinal settings. Diagnostic screening should rule out hematological malignancies to exclude undiagnosed malignancies in relevant control data. This is especially important for NBM material collected during thoracic surgeries, as clonal hematopoiesis, which has been identified as a precursor for both myeloid malignancies and cardiovascular disease,[43] may affect the background levels of LAIPs.[44,45] However, little is known on which LAIPs are affected by clonal hematopoiesis and to what extent.

AUTOMATION OF ROUTINE DIAGNOSTICS

Computational approaches offer the potential for more objective and rapid identification of MRD in AML. However, implementing these algorithms as medical decision tools in routine diagnostics still faces significant barriers. A recent survey[46] identified a lack of trust and understanding as the primary barrier to adopting automated cytometry analyses. This skepticism predominantly stems from potential discrepancies between the output of computational analyses and traditional manual gating. To bridge this gap, a pragmatic approach may limit computational analysis to objectifying the gating process or supplementing manual analysis with additional leukemic populations identified by models. Consequently, implementing computational approaches may not deviate significantly from manual analysis and could maintain existing clinical cut-offs for MRD positivity in AML. However, this hybrid approach requires software that facilitates this interaction and the possibility to run computationally demanding algorithms. Although computer-assisted methods facilitate a gradual transition toward more comprehensive automated approaches, it remains to be seen whether these approaches can deliver on the promise of reduced analysis time.

Beyond human assistance, addressing algorithmic complexity through interpretability and explainability is another pathway toward increased adoption of computational methods. Within theoretical machine learning frameworks, interpretability refers to understanding the inner workings of models (the "how"), while explainability addresses how decisions are made (the "why"). Most approaches (Table 1) classify cells at the single cell level, allowing for expert-based evaluation of model results in two-dimensional marker plots. For supervised models, Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) can be implemented to interpret model decisions. Both methods have been successfully applied in AML-MRD contexts to interpret which markers were likely aberrantly expressed on predicted leukemic cells.[27,31] However, these methods are not readily applicable to popular unsupervised algorithms such as FlowSOM (Table 1). In this context, evaluating model decisions relies on comprehensive inspection of clusters in two-dimensional plots. Alternatively, some studies have implemented UMAP or t-SNE algorithms to plot cells in a single two-dimensional plot where cells with similar marker expression cluster together.[9,11,25,27,31] However, these approaches perform complex data transformations that remain black-box and potentially arbitrary,[47,48] raising questions about their utility in clinical practice. As an alternative, established statistical methods such as GMMs have been utilized for clustering in AML-MRD analysis.[17,19,29] These models, comprising sets of multi-dimensional Gaussian ("normal") distributions, can be visualized using two-dimensional ellipsoids, facilitating more direct model interpretation.[29] However, the brief history of computational approaches for AML-MRD has already demonstrated significant challenges compared to other malignancies. Therefore, it remains to be seen whether such models are sufficient or whether more complex models (i.e., deep learning) widely adopted for other medical data, such as image recognition, are required.

Lastly, clinical implementation necessitates approval from relevant regulatory authorities such as the US Food and Drug Administration (FDA) or the European Medicines Agency (EMA). The regulatory pathway for such computational tools is complex and multifaceted, involving considerations such as external retrospective and prospective validation studies, intellectual property rights, addressing model biases, and post-market surveillance. A comprehensive discussion of these regulatory requirements is beyond this review's scope and has been discussed elsewhere in literature.[49,50]

FUTURE PERSPECTIVES

Algorithms' capabilities to identify clinically relevant cell populations without human intervention inevitably raise the question of whether they can improve upon the established prognostic value of MRD based on manual gating. Additional clinical information, such as relapse status or survival, can be leveraged in this setting as a gold standard rather than manual gates. Here, the focus shifts from classifying known leukemic populations to predicting patient response, opening new possibilities and challenges.

The prediction of AML relapse may include additional features not currently used in manual gating. For instance, non-leukemic populations with prognostic information for relapse can be identified in MRD assays.[51] Additionally, models can integrate information across different tubes in an MFC panel without data imputation.[17] These added dimensions could allow for enhanced identification of leukemic cells and characterize the immune environment's potential in limiting their proliferation.

Lastly, future computational MRD approaches will analyze data generated on next-generation cytometry platforms. For AML-MRD, 19 and 24-color panels for AML-MRD assessment have already been developed.[9,10] This number will likely expand soon with the inclusion of novel LAIP markers and increased attention to immune monitoring. Given that most people list high dimensional panels as the main reason for adopting computational analyses,[46] the demand for computational analysis will also increase. Future methods should consider this development, ensuring that models can operate with large numbers of markers. Dimensionality reduction algorithms that can address this "curse of dimensionality," such as UMAP, have already

been used in the context of AML-MRD.[22,24] However, future studies are still needed to validate the reliability of such data transformations for diagnostic purposes. Moreover, because most algorithms require identical marker features in training and test settings, implementing new platforms will likely limit the usability of previous datasets and models. An interesting solution for this problem is learning generalized representations of MFC data, such as FATE,[23] that can operate across incongruent panel compositions. However, the explainability and interpretability of such models still require more investigation.

CONCLUDING REMARKS

The application of computational analysis to identify MRD in AML patients holds significant promise for overcoming the limitations of manual analysis of cytometry data. The rarity of leukemic cells and the heterogeneity of LAIPs remain critical challenges, but several studies have already proposed innovative solutions. Implementing these approaches in clinical diagnostics can significantly reduce analysis time and provide more objective results across different clinical settings. However, the transition to fully automated MRD assessment still requires careful consideration of regulatory requirements such as validation studies and the development of user-friendly tools to explore and understand model decisions. Building trust and understanding through explainable modeling will be crucial for adopting these technologies. Looking forward, advancing computational methods can shift the focus from merely replicating manual gating to leveraging high-dimensional data for predicting AML relapse. This includes exploring the prognostic value of the immune environment and integrating data from spectral cytometry platforms, which offer unprecedented levels of detail. Consequently, we can move toward more precise, efficient, and predictive models for monitoring AML patients.

AUTHOR CONTRIBUTIONS

Tim R. Mocking: Writing—original draft; writing—review and editing; software; visualization. **Arjan A. van de Loosdrecht**: Writing—review and editing. **Jacqueline Cloos**: Writing—review and editing. **Costa Bachas**: Writing—original draft; writing—review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

FUNDING

This project was in part funded by an AMC-VUMC Cancer Center Amsterdam grant (#CCA_2020-9-68).

ORCID

Tim R. Mocking http://orcid.org/0000-0001-9170-4131
Arjan A. van de Loosdrecht https://orcid.org/0000-0001-8311-983X
Jacqueline Cloos https://orcid.org/0000-0001-9150-8026
Costa Bachas https://orcid.org/0000-0001-5983-6193

SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article.

REFERENCES

1. Short NJ, Zhou S, Fu C, et al. Association of measurable residual disease with survival outcomes in patients with acute myeloid leukemia: a systematic review and meta-analysis. *JAMA Oncol*. 2020;6(12):1890-1899.
2. Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood*. 2022;140(12):1345-1377.
3. Zeijlemaker W, Kelder A, Cloos J, Schuurhuis GJ. Immunophenotypic detection of measurable residual (stem cell) disease using LAIP approach in acute myeloid leukemia. *Curr Protoc Cytom*. 2019;91(1):e66.
4. Wood BL. Acute myeloid leukemia minimal residual disease detection: the difference from normal approach. *Curr Protoc Cytom*. 2020;93(1):e73.
5. Verbeek MWC, van der Velden VHJ. The evolving landscape of flowcytometric minimal residual disease monitoring in B-cell precursor acute lymphoblastic leukemia. *Int J Mol Sci*. 2024;25(9):4881.
6. Rawstron AC, Fazi C, Agathangelidis A, et al. A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. *Leukemia*. 2016;30(4):929-936.
7. Flores-Montero J, Sanoja-Flores L, Paiva B, et al. Next Generation Flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. *Leukemia*. 2017;31(10):2094-2103.
8. Tettero JM, Dakappagari N, Heidinga ME, et al. Analytical assay validation for acute myeloid leukemia measurable residual disease assessment by multiparametric flow cytometry. *Cytom B Clin Cytom*. 2023;104(6):426-439.
9. Fokken H, Waclawski J, Kattre N, et al. A 19-color single-tube full spectrum flow cytometry assay for the detection of measurable residual disease in acute myeloid leukemia. *Cytometry A*. 2024;105(3):181-195.
10. Chen M, Fu M, Gong M, et al. Twenty-four-color full spectrum flow cytometry panel for minimal residual disease detection in acute myeloid leukemia. *Open Med*. 2023;18(1):20230745.
11. Matthes T. Phenotypic analysis of hematopoietic stem and progenitor cell populations in acute myeloid leukemia based on spectral flow cytometry, a 20-color panel, and unsupervised learning algorithms. *Int J Mol Sci*. 2024;25(5):2847.
12. Duetz C, Bachas C, Westers TM, van de Loosdrecht AA. Computational analysis of flow cytometry data in hematological malignancies: future clinical practice? *Curr Opin Oncol*. 2020;32(2):162-169.
13. Aghaeepour N, Finak G, Hoos H, et al. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods*. 2013;10(3):228-238.
14. 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.
15. Rajwa B, Wallace PK, Griffiths EA, Dundar M. Automated assessment of disease progression in acute myeloid leukemia by probabilistic analysis of flow cytometry data. *IEEE Trans Biomed Eng*. 2017;64(5):1089-1098.
16. Arvaniti E, Claassen M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nat Commun*. 2017;8(1):14825.
17. Ko B-S, Wang Y-F, Li J-L, 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.
18. Licandro R, Schlegl T, Reiter M, et al. WGAN latent space embeddings for blast identification in childhood acute myeloid leukaemiaeds., *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE; 2018.

19. 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. *Cancer Cells*. 2018;1012:1010.
20. Weber LM, Nowicka M, Soneson C, Robinson MD. diffcyt: differential discovery in high-dimensional cytometry via high-resolution clustering. *Commun Biol*. 2019;2(1):183.
21. Lacombe F, Lechevalier N, Vial JP, Béné MC. An R-derived FlowSOM process to analyze unsupervised clustering of normal and malignant human bone marrow classical flow cytometry data. *Cytometry A*. 2019;95(11):1191-1197.
22. Jacqmin H, Chatelain B, Louveaux Q, et al. Clustering and kernel density estimation for assessment of measurable residual disease by flow cytometry. *Diagnostics*. 2020;10(5):317.
23. 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*. 2021;13(4):629.
24. Craddock C, Jackson A, Loke J, et al. Augmented reduced-intensity regimen does not improve postallogeneic transplant outcomes in acute myeloid leukemia. *J Clin Oncol*. 2021;39(7):768-778.
25. Weijler L, Kowarsch F, Wödlinger M, et al. UMAP based anomaly detection for minimal residual disease quantification within acute myeloid leukemia. *Cancers*. 2022;14(4):898.
26. Weijler L, Kowarsch F, Reiter M, et al. FATE: Feature-Agnostic Transformer-based Encoder for learning generalized embedding spaces in flow cytometry data. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*; 2024: 7956-7964.
27. Shopsowitz K, Lofroth J, Chan G, et al. MAGIC-DR: an interpretable machine-learning guided approach for acute myeloid leukemia measurable residual disease analysis. *Cytom B Clin Cytom*. 2024;106:239-251.
28. McCarthy N, Gui G, Dumezy F, et al. Pre-emptive detection and evolution of relapse in acute myeloid leukemia by flow cytometric measurable residual disease surveillance. *Leukemia*. 2024;38:1667-1673.
29. Mocking TR, Kelder A, Reuvekamp T, et al. Computational assessment of measurable residual disease in acute myeloid leukemia using mixture models. *Commun Med*. 2024;4(1):271.
30. Gachon E, Bigot J, Cazelles E, et al. Low dimensional representation of multi-patient flow cytometry datasets using optimal transport for measurable residual disease detection in leukemia. *Cytometry A*. 2025;107(2):126-139.
31. Vahedi A, Royaei M, Madani T, Mercier F, Poopak B. A machine learning approach for the rapid identification of measurable residual disease in acute myeloid leukemia. *Haematologica*. Published online March 6, 2025. doi:10.3324/haematol.2024.286019
32. 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.
33. Buccisano F, Palmieri R, Piciocchi A, et al. Clinical relevance of an objective flow cytometry approach based on limit of detection and limit of quantification for measurable residual disease assessment in acute myeloid leukemia. A post-hoc analysis of the GIMEMA AML1310 trial. *Haematologica*. 2022;107(12):2823-2833.
34. 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.
35. Bagwell CB, Bray CM, Herbert DJ, et al. Improving the t-SNE algorithms for cytometry and other technologies: Cen-se′ mapping. *J Biom Biostat*. 2019;10(430):2.
36. Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytometry A*. 2016;89(12):1084-1096.
37. Monaco G, Chen H, Poidinger M, Chen J, de Magalhães JP, Larbi A. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics*. 2016;32(16):2473-2480.
38. Emmaneel A, Quintelier K, Sichien D, et al. PeacoQC: peak-based selection of high quality cytometry data. *Cytometry A*. 2022;101(4): 325-338.
39. Meskas J, Yokosawa D, Wang S, Segat GC, Brinkman RR. flowCut: an R package for automated removal of outlier events and flagging of files based on time versus fluorescence analysis. *Cytometry A*. 2023;103(1):71-81.
40. Montante S, Brinkman RR. Flow cytometry data analysis: recent tools and algorithms. *Int J Lab Hematol*. 2019;41:56-62.
41. Putri GH, Howitt G, Marsh-Wakefield F, Ashhurst TM, Phipson B. SuperCellCyto: enabling efficient analysis of large scale cytometry datasets. *Genome Biol*. 2024;25(1):89.
42. Saeys Y, Van Gassen S, Lambrecht B. Response to Orlova et al. "Science not art: statistically sound methods for identifying subsets in multi-dimensional flow and mass cytometry data sets. *Nat Rev Immunol*. 2018;18(1):78.
43. Calvillo-Argüelles O, Jaiswal S, Shlush LI, et al. Connections between clonal hematopoiesis, cardiovascular disease, and cancer: a review. *JAMA Cardiol*. 2019;4(4):380-387.
44. Camburn AE, Petrasich M, Ruskova A, Chan G. Myeloblasts in normal bone marrows expressing leukaemia-associated immunophenotypes. *Pathology*. 2019;51(5):502-506.
45. Loghavi S, DiNardo CD, Furudate K, et al. Flow cytometric immunophenotypic alterations of persistent clonal haematopoiesis in remission bone marrows of patients with NPM1-mutated acute myeloid leukaemia. *Br J Haematol*. 2021;192(6):1054-1063.
46. Czechowska K, Lannigan J, Aghaeepour N, et al. Cyt-Geist: current and future challenges in cytometry: reports of the CYTO 2019 conference workshops. *Cytometry A*. 2019;95(12):1236-1274.
47. Chari T, Pachter L. The specious art of single-cell genomics. *PLoS Comput Biol*. 2023;19(8):e1011288.
48. Wang S, Sontag ED, Lauffenburger DA. What cannot be seen correctly in 2D visualizations of single-cell 'omics data? *Cell Syst*. 2023;14(9):723-731.
49. Harvey HB, Gowda V. How the FDA regulates AI. *Acad Radiol*. 2020;27(1):58-61.
50. Couckuyt A, Seurinck R, Emmaneel A, et al. Challenges in translational machine learning. *Hum Genet*. 2022;141(9):1451-1466.
51. Pessach I, Spyropoulos T, Lamprianidou E, Kotsianidis I. MRD monitoring by multiparametric flow cytometry in AML: is it time to incorporate immune parameters? *Cancers*. 2022;14(17):4294.