Computers in Biology and Medicine 193 (2025) 110394

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A machine learning framework for cross-institute standardized analysis of flow cytometry in differentiating acute myeloid leukemia from non-neoplastic conditions

Yu-Fen Wang [a,b], En-Ping Chu [a], Fong-Ci Lin [a], Huan-Yu Chen [b,f], Tsung-Chih Chen [c], Joseph Hanson [d], Joseph D. Tario [d], Kai Fu [d], Sara A. Monaghan [e], Paul K. Wallace [d], Chi-Chun Lee [b,f], Bor-Sheng Ko [b,g,h,i,*]

[a] AHEAD Intelligence Ltd, Taipei, Taiwan
[b] AHEAD Medicine Corporation, San Jose, CA, United States
[c] Taichung Veterans General Hospital, Taichung, Taiwan
[d] Roswell Park Comprehensive Cancer Center, Buffalo, NY, United States
[e] University of Pittsburgh School of Medicine, Pittsburgh, PA, United States
[f] Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan
[g] Department of Hematological Oncology, National Taiwan University Cancer Center, Taipei, Taiwan
[h] Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan
[i] School of Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

ARTICLE INFO

Keywords:
Flow cytometry
Acute myeloid leukemia
Machine learning
Universal algorithm

ABSTRACT

Flow cytometry (FC) remains a cornerstone diagnostic tool for acute myeloid leukemia (AML), yet standardizing panels across laboratories presents persistent challenges. Our study introduces a validated machine learning framework enabling cross-panel AML classification by leveraging common parameters shared across diverse FC protocols.

We employed FC data from 215 samples (110 AML, 105 non-neoplastic) collected in five institutions using different panel configurations as model training set, and another 196 similarly collected samples (90 AML and 106 non-neoplastic) for independent validation set. The framework employs GMM-SVM classification based on 16 common parameters (FSC-A, FSC-H, SSC-A, CD7, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117, and HLA-DR) that are consistently present across various panel designs. The framework demonstrated robust performance with 98.15 % accuracy, 99.82 % area under curve (AUC), 97.30 % sensitivity, and 99.05 % specificity. Independent validation on 196 additional samples further confirmed the framework's effectiveness, maintaining high performance with 93.88 % accuracy and 98.71 % AUC.

This research establishes the viability of standardized FC analysis across diverse panel configurations and instruments through machine learning implementation. The framework's robust performance suggests promising applications for harmonized multi-center FC analysis, potentially resolving current standardization challenges in flow cytometry interpretation.

1. Introduction

Multiparameter Flow Cytometry (MFC) is vital for hematological disease diagnostics and monitoring, with panels typically measuring 15 or more parameters. The difficulties associated with high-dimensional data are amplified by the increasing use of fluorochrome/antibody combinations and custom testing protocols, particularly in complex leukemia and lymphoma tests. However, many laboratories still depend on manual gating, a process that requires human experts to apply a sequential gating procedure to large sets of bidimensional plots to identify and label cell populations of interest. The expert-dependent manual interpretation process is time-consuming, subjective, and

* Corresponding author. No. 57, Lane 155, Section 3 of Keelung Rd, Taipei, Taiwan, Department of Hematological Oncology, National Taiwan University Cancer Center, Taipei, Taiwan, School of Medicine, National Taiwan University College of Medicine, Taipei 106, Taiwan.
E-mail addresses: bskomd@ntu.edu.tw, kevinkomd@ntuh.gov.tw (B.-S. Ko).

https://doi.org/10.1016/j.compbiomed.2025.110394
Received 10 March 2025; Received in revised form 23 April 2025; Accepted 5 May 2025
Available online 21 May 2025

prone to inter-interpreter variability. Furthermore, a shortage of trained professionals in flow cytometry intensifies workloads, extends patient wait times, and raises safety concerns. A survey of the clinical cytometry workforce revealed that the vacancy rate for clinical flow cytometry technologists tripled from 2014 to 2022, while the retirement rate surged significantly from 2020 to 2022 [1,2]. Increasing clinical flow cytometry test demand coupled with a lack of trained professionals further complicates access to cytometry technologies and their application across various fields.

Differences in assay design, reagent choices, and interpretation practices often lead to inconsistent results and complicate standardization. In the past decades, there were many attempts to develop guidance and consensus for flow cytometry panels in various applications from cell sorting [3], PNH [4], multiple myeloma (MM) measurable residual disease (MRD) [5], acute myeloid leukemia (AML) MRD [6,7], and B-ALL MRD [8]. However, it remains challenging to achieve consensus among laboratories regarding panel design for particular clinical applications. Various institutions may prioritize different markers based on regional preferences, reagent availability, or historical methods, resulting in diverse diagnostic practices for similar medical conditions. This inconsistency complicates the comparison of patient results between labs and may impede collaborative clinical and translational research.

Artificial intelligence (AI), particularly machine learning (ML), holds significant potential for assisting physicians in managing hematolymphoid diseases by simplifying the interpretation of complex flow cytometry data. ML approaches have been applied to various aspects of flow cytometry analysis, aiding in tasks such as diagnosis, risk stratification, and predicting treatment responses. Numerous computational solutions have been developed for specific applications, including quality control algorithms to filter low-quality events (e.g., flowAI, flowClean, PeaCoQC [9] and normalization algorithms to mitigate batch effects (e.g., CytoNorm [10]). Approaches like clustering for automated cell population identification utilize density-based methods such as FlowSOM [11], while automated gating algorithms and dimensionality reduction techniques, such as principal component analysis (PCA), t-SNE, and PhenoGraph [12], enable improved visualization of flow cytometry data.

More recently, deep learning methods have been implemented for classification tasks, with examples like FlowCat and EnsembleCNN [13]. While many of these tools focus on identifying cell populations [14], machine learning-driven approaches for disease classification using MFC data are also emerging. Such examples include UMAP-RF [13], which classifies diseases, and CNN-based methods for subtype classification of B-cell non-Hodgkin lymphomas (B-NHL) [15]. These developments demonstrate the growing potential of ML in advancing clinical applications of flow cytometry.

Furthermore, many of these tools could only support data analysis measured with the same panel, therefore, it is hard to apply the one optimized pipeline developed to analyze one dataset readily applicable to analyze data acquired using another flow cytometry panel. Additionally, several of these tools lack accessible interfaces and require users to write computer languages such as R or Python, making them less practical for clinical environments. This shortcoming emphasizes the demand for better options for comprehensive immunophenotype assessment. All of these phenomena underscore the urgent need for user-friendly bioinformatic solutions that can efficiently analyze complex flow cytometry data without requiring advanced programming skills.

Addressing these pressing issues requires a multifaceted approach that includes the development of intuitive, automated analysis software, improved training programs, standardized protocols to minimize variability, and initiatives aimed at enhancing job satisfaction and retention for laboratory staff.

We have previously shown that supervised machine learning approaches can effectively identify leukemia at sample level through multiple retrospective studies at single centers: 1) AML MRD assessment using 5333 flow cytometry data from National Taiwan University Hospital (NTUH), achieving accuracies between 84.6 % and 92.4 % and AUCs reaching between 92.1 % and 95.0 % [16]; 2) Acute leukemia subtype classification with 592 flow cytometry data from UPMC achieving 94.1 % accuracy and 99.6 % AUC [17]; and 3) AML MRD assessment with 209 flow cytometry data from Mayo Clinic showing 88 % accuracy and 91.3 % AUC [18]. 4) AML MRD assessment with 1040 MFC data from Roswell Park Comprehensive Cancer Center (RPCCC) [19], 5) Automatic hematologic malignancy classification on Munich Leukemia Laboratory (MLL) data [20]. These findings suggest that AI-driven classification of flow cytometry data is not only fast and highly accurate but could also enhance the efficiency of specimen triage.

Our feature selection analysis from these above-mentioned studies revealed that we could develop classifiers that perform comparably using only a subset of parameters from the entire panel. Given the overlap of common parameters in various panels for diagnosing hematological diseases, we hypothesize that our methods could be utilized across different panels and instruments, employing shared parameters to develop classifiers effectively.

In this study, we employed datasets collected from five sites with different panel and instrument usage scenarios, and utilized our previously developed machine learning framework to develop cross-panel sample classification for AML and non-neoplastic conditions.

2. Materials and methods

2.1. Flow cytometry datasets

Retrospective clinical FC datasets from five different institutions were collected for this study: National Taiwan University Cancer Center (NTUCC) in Taiwan, Roswell Park Comprehensive Cancer Center (RPCCC) in United States, Taichung Veteran General Hospital (VGH) in Taiwan, University of Pittsburgh (UPMC) in United States, and National Taiwan University Hospital (NTUH) in Taiwan. Bone marrow (BM) specimens were collected and analyzed with the standard diagnostic protocols at each clinical FC laboratory for the identification of acute myeloid leukemia (AML) or non-neoplastic hematological conditions including cytopenia(s), and the diagnosis of AML was made according to WHO 2016 classification [21] in all centers. Prior to the study's initiation, all biospecimens and their corresponding FCS files were de-identified. A total of 110 AML and 105 non-neoplastic FC samples were collected for training and cross-validating our approach. Detailed descriptions including dataset sources, types of instruments and panel, and case number are listed in Table 1a. With the exception of 9 non-neoplastic cases from NTUCC, which were monitored for AML residual disease (MRD) and interpreted as negative, all other samples were collected at first diagnosis. These cases were measured with five types of panels, including Euroflow AML/MDS, ClearLLab 10C, and three different laboratory developed test (LDT) panels, and acquired from three different instrument models, including BD FACSCantoII, BD FACSLyric, and Beckman Coulter NaviosEX. Each panel used a unique combination of markers and channels for AML and non-neoplastic diagnosis, as outlined in Supplementary Table S1. Of these, sixteen parameters were shared across all panels: FSC-A, FSC-H, SSC-A, CD7, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117, and HLA-DR.

In addition to the training datasets, we collected 90 AML and 106 non-neoplastic FC samples, as described in Table 1b, for use as independent validation datasets to assess the performance of the model's classification. These samples were not included in any model training. Non-neoplastic samples from NTUCC, RPCCC, and NTUH were MRD negative cases. AML cases from RPCCC and NTUH were MRD cases with residual disease percentage greater than 80 % and 50 %, respectively. Except for the samples from RPCCC, all validation samples were measured using the same instrument models and panels as the corresponding training datasets. However, the cases from RPCCC were

Table 1
Training and validation set across different institutions.

1a) Training and Validation Datasets

Dataset	NTUCC	RPCCC	VGH	UPMC	NTUH	All Sites
Flow Panel	Euroflow AML/MDS	ClearLLab 10C	LDT	LDT	LDT	5 panels
Instrument	BD FACSLyric	Navios EX	BD FACSCantoII	BD FACSCantoII	BD FACSCantoII	3 models
Number of tubes	7	4	4	4	13	
File format	FCS3.1	FCS3.0	FCS3.0	FCS3.0	FCS2.0	
Average Event Per Tube	600,000	70,000	260,000	28,000	100,000	
Sample Number (AML, non-neoplastic)	17, 19	27, 21	17, 18	24, 25	25, 22	110, 105

1b) Independent Validation Datasets

Dataset	NTUCC	RPCCC	VGH	UPMC	NTUH	All sites
Flow Panel	Euroflow AML/MDS	AML MRD LDT	LDT	LDT	LDT	5 panels
Instrument	BD FACSLyric	BD FACSCantoII	BD FACSCantoII	BD FACSCantoII	BD FACSCantoII	2 models
Number of tubes	7	5	4	4	13	
File format	FCS3.1	FCS3.0	FCS3.0	FCS3.0	FCS2.0	
Average Event Per Tube	600,000	700,000	260,000	28,000	100,000	
Sample Number (AML, non-neoplastic)	16, 16	18, 17	6, 31	25, 25	25, 17	90, 106

Abbreviation: NTUCC: National Taiwan University Cancer Center, RPCCC: Roswell Park Comprehensive Cancer Center, VGH: Taichung Veterans General Hospital, UPMC: University of Pittsburgh Medical Center, NTUH: National Taiwan University Hospital, AML: Acute Myeloid Leukemia, MDS: Myeloid Dysplastic Syndrome, LDT: laboratory developed test.

measured using panels specifically designed for AML MRD monitoring, with several different combinations of tubes employed. Detailed descriptions of the RPCCC panels are provided in Supplementary Table S2.

This study was approved by the Institutional Review Boards (IRBs) in accordance with the Declaration of Helsinki. The IRB approval codes for each institution are as follows: RPCCC-STUDY00001445, TCVGH-SE20116A, NTUCC-202204104RS, UPMC-STUDY21080145, and NTUH-201906018RINB, while informed consents are waived in all approvals.

2.2. Machine learning based sample classification framework

We developed a comprehensive machine learning based framework for cross-panel flow cytometry analysis. The workflow consists of four primary components: parameter alignment, data preprocessing, sample-level representation encoding, and classification, as illustrated in Fig. 1. This section details each component and its implementation.

2.2.1. Parameter alignment and data preprocessing

The initial preprocessing stage is crucial for ensuring compatibility across different panels and data quality. Raw FC data undergoes sequential processing steps:

1. Parameter alignment to extract common parameters presented across all panels
2. Application of compensation matrices to correct fluorescence spillover between channels
3. Random down-sampling per sample to ensure computational efficiency
4. Max-min normalization of fluorescent channel values to standardize measurements

Each preprocessed FC data $X \in R^{T\times D}$ is utilized for the following sample-level representation encoding and classification, where T is the total cell number across tubes of one FC sample and D is the number of common parameters measured across all panels.

2.2.2. Sample-level encoding with Fisher vector

The sample encoding process employs a two-stage approach to transform variable-length flow cytometry data into fixed-length

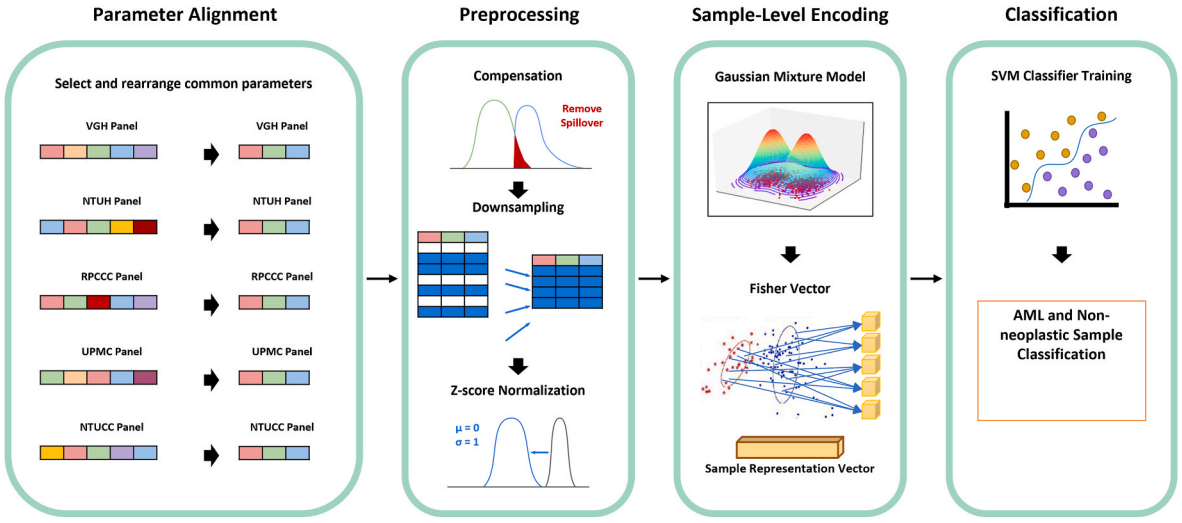


Fig. 1. Workflow of the presented panel-agnostic machine learning (ML) approach for AML versus non-nonneoplastic sample classification. This workflow is based on our previous sample classification method for a single panel. Common markers/parameters across different panels are selected and rearranged first before following cross-panel analysis.

representations suitable for machine learning analysis.

Stage 1: Gaussian Mixture Model Modeling

The Gaussian Mixture Model (GMM) is trained through an expectation-maximization algorithm with the cell distribution of each FC sample to obtain a set of parameters λ:

$$\lambda = \omega_k, \mu_k, \sigma_k;\ k = 1...K$$

where ω_k, μ_k, and σ_k represent respectively the mixing weights, mean vectors, and covariance matrices of k-th Gaussian component. K denotes the total number of Gaussian components. For a single cell data $x_t \in X$, the GMM probability density function is defined as:

$$P(x_t \mid \lambda) = \sum_{k=1}^{K} \omega_k P_k(x_t \mid \omega_k, \mu_k)$$

Stage 2: Fisher Vector Computation

The Fisher Vector encoding quantifies the sample distribution's deviation from the modeled GMM through first and second-order statistics. The first-order and second-order statistics are computed as:

$$g_{\mu_k}^{X} = \frac{1}{T\sqrt{\omega_k}} \sum_{t=1}^{T} \gamma_t(k) \left(\frac{x_t - \mu_k}{\sigma_k} \right)$$

$$g_{\sigma_k}^{X} = \frac{1}{T\sqrt{2\omega_k}} \sum_{t=1}^{T} \gamma_t(k) \left(\left(\frac{x_t - \mu_k}{\sigma_k} \right)^2 - 1 \right)$$

where $\gamma_t(k)$ represents the posterior probability for x_t:

$$\gamma_t(k) = P(i \mid x_t, \lambda) = \frac{\omega_i P_i(\lambda)}{\sum_{j=1}^{N} \omega_j P_j(\lambda)}$$

This process generates a high-dimensional feature vector F with dimensionality $2KD$, where K is the component number of modeled GMM and D is the number of FC parameters.

2.2.3. Support Vector Machine classification

The computed Fisher Vectors serve as the input to a linear Support Vector Machine (SVM) classifier. The SVM optimization problem is formulated as:

$$\min_{w,b,\xi} \frac{1}{2} ||w||^2 + C \sum_i \xi_i$$

subject to : $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$

2.2.4. Algorithm implementation

The procedures of training and predicting algorithms for the presented classification method are described as below:

Algorithm 1. Flow Cytometry Classification Training

Input: 1. Preprocessed FC data $\boldsymbol{X}$, where $\boldsymbol{X} \in \boldsymbol{R}^{T \times D}$
2. Clinical diagnosis $\boldsymbol{Y}$ for corresponding $\boldsymbol{X}$, where $\boldsymbol{Y} \in [\boldsymbol{AML}, \boldsymbol{non\text{–}neoplastic}]$
Output: 1. Parameters for trained GMM
2. Parameters for trained SVM
Procedure:
1 Train GMM with training FC data $\boldsymbol{X}$
2 Compute feature vector $\boldsymbol{F}$ with trained GMM and Fisher Vector encoding for each FC data $\boldsymbol{X}$
3 Train SVM with feature vectors $\boldsymbol{F}$ and corresponding clinical diagnosis $\boldsymbol{Y}$

Algorithm 2. Flow Cytometry Classification Predicting

Input: 1. Preprocessed FC data $\boldsymbol{X}'$, where $\boldsymbol{X}' \in \boldsymbol{R}^{T \times D}$ and $\boldsymbol{X}' \notin \boldsymbol{training\ data\ X}$
2. Parameters for trained GMM
3. Parameters for trained SVM
Output: 1. Prediction label $\boldsymbol{Y}'$
2. Prediction probability $\boldsymbol{P}'$
Procedure:
1 Compute feature vector $\boldsymbol{F}'$ with trained GMM and Fisher Vector encoding for each FC data $\boldsymbol{X}'$
2 Output predicted diagnosis label $\boldsymbol{Y}'$ and prediction probability $\boldsymbol{P}'$ for each FC case with feature vector $\boldsymbol{F}'$ and trained SVM

2.3. Sample classifier performance evaluation

The workflow predicts diagnostic labels with probabilities for each sample. In this study, the applicability and accuracy of this algorithm is evaluated across five diagnostic laboratories using different reagent panels. The model was trained and validated with 3-fold cross validation. The classification performance was evaluated by area under the receiver operating characteristic curve (AUC), accuracy (defined as the concordance of model prediction to the ground truth), sensitivity (true positive rate, TPR), specificity (true negative rate, TNR), false negative rate (FNR), and false positive rate (FPR). Notably, the threshold, a cutoff value applied to the model's predicted probabilities, on the ROC curve can be adjusted to balance the TPR and TNR, allowing optimization of the model's sensitivity or specificity depending on clinical priorities.

2.4. Classification result visualization

The prediction result would be presented in two sample-level visualizations. The first one is a 3D visualization. This method includes principal component analysis (PCA) -compressed features and prediction probabilities as visualization coordinates. The second method is Ranked prediction probabilities visualization. The predicted probability for each FC case is ranked to observe the probability threshold of wrongly predicted cases.

2.5. Explainability analysis

The importance of each FC parameter to the sample classification was assessed with three methods, including Single Parameter Selected experiment, Single Parameter Masked experiment, and Forward Sequential Feature Selection. The first method trains and validates the model with one of the original parameters alone and compares all obtained performance to see which parameter achieves better by itself. The second approach trains the model with the original parameter set, but validates it with one of the parameters being masked with zero value. The last method selects a relevant parameter subset from the original set by starting with an empty parameter set and iteratively adding a parameter that improves model performance the most at each parameter count.

2.6. Software and programming resources

All experiments were conducted using programs coded with Python 3.8 on a Linux server (Ubuntu 22.04). This platform includes 48 Intel(R) Xeon(R) CPUs and 500 GB of RAM for data analysis.

3. Results

3.1. Cross-site AML vs non-neoplastic sample classification

We evaluated the feasibility and performance of cross-panel AML and non-neoplastic sample classification using various dataset and parameter combinations. Table 2 summarizes the results of eleven

Table 2
Model performance with different datasets and parameter sets.

2a) Performance of models trained with site-dependent parameters

Model	A	B	C	D	E
Dataset	NTUCC	RPCCC	VGH	UPMC	NTUH
Panel	Euroflow AML/MDS	ClearLLab 10C	LDT	LDT	LDT
Site-Dependent Parameter (n)	**23**	**31**	**22**	**30**	**21**
Instrument	BD FACSLyric	Navios EX	BD FACSCantoII	BD FACSCantoII	BD FACSCantoII
Sample Number (AML, non-neoplastic)	17, 19	27, 21	17, 18	24, 25	25, 22
AUC	100.00 %	100.00 %	100.00 %	99.48 %	100.00 %
Accuracy	97.22 %	97.92 %	94.19 %	97.92 %	100.00 %
Sensitivity	94.44 %	96.30 %	87.78 %	100.00 %	100.00 %
Specificity	100.00 %	100.00 %	100.00 %	95.83 %	100.00 %
FNR	5.56 %	3.70 %	12.22 %	0.00 %	0.00 %
FPR	0.00 %	0.00 %	0.00 %	4.17 %	0.00 %

2b) Performance of models trained with common parameters across panels

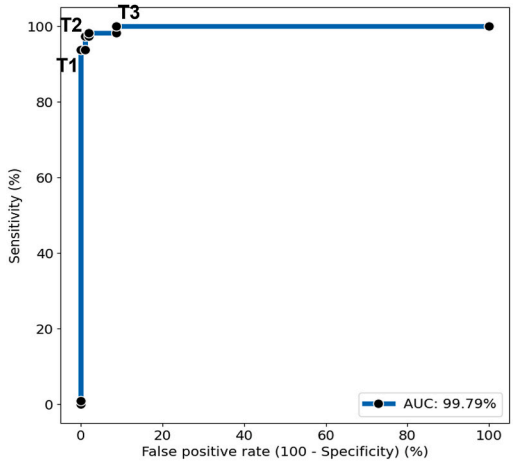
Model	F	G	H	I	J	K
Dataset	NTUCC	RPCCC	VGH	UPMC	NTUH	All Sites
Panel	Euroflow AML/MDS	ClearLLab 10C	LDT	LDT	LDT	5 panels
Parameters (n)	**16 common parameters[a] across all panels**					
Instrument	BD FACSLyric	Navios EX	BD FACSCantoII	BD FACSCantoII	BD FACSCantoII	3 models
Sample Number (AML, non-neoplastic)	17, 19	27, 21	17, 18	24, 25	25, 22	110, 105
AUC	100.00 %	99.47 %	100.00 %	98.96 %	100.00 %	99.82 %
Accuracy	94.44 %	93.75 %	94.19 %	95.83 %	100.00 %	98.15 %
Sensitivity	88.89 %	96.30 %	87.78 %	95.83 %	100.00 %	97.30 %
Specificity	100.00 %	90.48 %	100.00 %	95.83 %	100.00 %	99.05 %
FNR	11.11 %	3.70 %	12.22 %	4.17 %	0.00 %	2.70 %
FPR	0.00 %	9.52 %	0.00 %	4.17 %	0.00 %	0.95 %

(a) Models A to E were trained with individual dataset and parameters. (b) Model F to J were trained with individual dataset, but only using common parameters across different panels. Model K was trained with all five datasets and common parameters.
Abbreviation: AML: Acute Myeloid Leukemia, NTUCC: National Taiwan University Cancer Center, RPCCC: Roswell Park Comprehensive Cancer Center, VGH: Taichung Veterans General Hospital, UPMC: University of Pittsburgh Medical Center, NTUH: National Taiwan University Hospital, AUC: area under the receiver operating characteristic curve, FNR: false negative rate, FPR: false positive rate.
** All metric values are the averaged performance across three-fold cross validation.
[a] Common parameters: FSC-A, FSC-H, SSC-A, CD7, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117, and HLA-DR.

classification models, trained and tested under different configurations. Models A to E were trained on individual datasets using site-specific parameters, while Models F to J were trained using individual datasets but constrained to the 16 parameters commonly measured across all five flow cytometry panels. Finally, Model K was trained on the combined datasets from all five sites using the same 16 common parameters: FSC-A, FSC-H, SSC-A, CD7, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117, and HLA-DR.

Models A to E, which utilized an expanded feature set specific to each site, achieved AUCs between 99.48 % and 100 %. Models F to J, trained on individual datasets with the 16 common parameters, achieved comparable classification performance, with AUC values ranging from 98.96 % to 100 % (Table 2). Performance across models trained on the same dataset was largely consistent, with only marginal performance differences. For example, Model C and Model H (both trained on the VGH dataset) achieved identical results: AUC of 100 %, accuracy of 94.19 %, sensitivity of 87.78 %, and specificity of 100 %. Models trained on NTUH datasets (Models E and J) demonstrated equivalent AUC of 100 % and accuracy of 100 %.

Model K, the panel-agnostic classifier trained on all datasets



Threshold (Prediction Probability)	T1 (73.67%)	T2 (45.69%)	T3 (16.42%)
Accuracy	96.74%	98.14%	95.81%
Sensitivity	93.64%	98.18%	100.00%
Specificity	100.00%	98.10%	91.43%
FNR	6.36%	1.82%	0.00%
FPR	0.00%	1.90%	8.57%

Abbreviation: AUC: area under the receiver operating characteristic curve, FNR: false negative rate, FPR: false positive rate
* All metric values are the overall performance of the three-fold cross validation, which considers predictions from multiple folds as one same fold for evaluation.

Fig. 2. Receiver operating characteristic (ROC) curve of Model K and its performance with different thresholds. The blue line represents the ROC curve of Model K, along with the black points as different classification thresholds. By adjusting the threshold, the workflow is able to regulate its prediction capability on AML and non-neoplastic samples.

combined and using only the 16 common parameters, demonstrated outstanding performance, achieving an AUC of 99.82 % and accuracy of 98.15 %. This result highlights the utility of the reduced, standardized parameter set for cross-panel applicability without compromising diagnostic performance.

The receiver operating characteristic (ROC) curve of Model K (Fig. 2) shows flexibility in performance depending on the selected threshold. At threshold T2, Model K achieved an accuracy of 98.14 %, a sensitivity of 98.18 %, and specificity of 98.10 %. When the threshold was adjusted to T1 to prioritize specificity, the model achieved perfect specificity (100 %) but slightly lower sensitivity (93.64 %) and accuracy (96.74 %). Conversely, at threshold T3, the model maximized sensitivity (100 %) while slightly reducing accuracy (95.81 %) and specificity (91.43 %). These results demonstrate that the model can be adapted to different operational needs by optimizing the classification threshold. ROC curves and corresponding confusion matrices for all models (A through K) are presented in Supplementary Fig. S1.

3.2. Independent validation

Independent validation datasets were utilized to evaluate the generalizability of Model K, with performance metrics summarized in Table 3. Model K achieved an overall AUC of 98.71 % and an accuracy of 93.88 %, effectively distinguishing AML-positive cases from non-neoplastic cases across the independent datasets. Of particular note, the RPCCC dataset, which was measured using the AML-MRD panel, contained panel compositions that were not included in the training dataset and incorporated several composition variations across different tubes (Supplementary Table S2). Despite this variability, Model K demonstrated robust performance on the RPCCC dataset, achieving an AUC of 100 %, accuracy of 97.14 %, sensitivity of 100 %, and specificity of 94.12 % (Table 3). These findings underscore the model's flexibility and resilience in accurately classifying AML and non-neoplastic cases across diverse datasets and panel configurations.

3.3. Visualization of classification results

To further interpret and compare classification results, we deployed 3D visualizations (Fig. 3a) and ranked prediction probability distribution plots (Fig. 3b and c) to assess Model K's performance based on its predictions and ground truth labels. In the 3D visualization, each point represents a single flow cytometry (FC) case, with the center color indicating the ground truth diagnosis and the edge color representing the predicted label. Out of the 215 cases analyzed, Model K produced four discordant predictions, as detailed in Supplementary Table S3.

The 3D visualization (Fig. 3a), based on principal component analysis (PCA)-compressed features, demonstrated clear separation between AML and non-neoplastic cases, highlighting Model K's ability to extract meaningful, panel-agnostic sample-level features. The ranked prediction probability plots (Fig. 3b) revealed that misclassifications occurred when AML prediction probabilities dipped below 45.69 % or when non-neoplastic probabilities dropped below 28.47 %. Similarly, Fig. 3c, which emphasizes independent validation datasets represented in deeper colors, showed misclassifications when AML probabilities fell below 46.62 % or non-neoplastic probabilities fell below 41.68 %. Discordant cases identified within the independent validation datasets are listed in Supplementary Table S4.

These visualizations effectively highlight Model K's performance and its ability to generalize across diverse datasets, while clearly identifying the thresholds at which misclassifications occur. This approach further emphasizes Model K's consistency, and interpretability in accurately classifying AML and non-neoplastic cases, even across varying panel configurations.

3.4. Assessing contribution of features to model performance

We implemented multiple feature selection approaches to evaluate the relative contribution of each parameter to Model K's classification performance and to assess parameter interactions and their effect on model performance. Fig. 4 illustrates the individual contributions of the 16 common parameters to Model K's performance. In single-parameter training experiments, AUC values ranged from 62.9 % (CD13) to 96.9 % (CD117), while accuracy values ranged between 57.7 % (CD13) and 91.2 % (SSC-A). The most influential parameters, each achieving an AUC above 82 %, were identified as CD117, SSC-A, CD34, HLA-DR, and CD16. Conversely, CD13, CD56, and CD45 demonstrated lower individual impact, with AUC values lower than 65 %.

Parameter masking experiments revealed the critical importance of specific markers in the classification process. During these experiments, individual parameters were systematically masked during model validation, and the impact on performance was assessed using AUC and accuracy metrics. HLA-DR, CD117, CD64, and CD56 emerged as particularly crucial parameters, with their masking resulting in substantial performance decreases: AUC reductions of 43.09 %, 24.89 %, 4.50 %, and 3.15 % (Fig. 5a), respectively, and accuracy decreases of 48.38 %, 45.14 %, 26.01 %, and 26.02 % (Fig. 5b). The marked reduction in accuracy following HLA-DR and CD117 masking underscores their significance in distinguishing between AML and non-neoplastic cases. Notably, while CD64 and CD56 showed modest and poor performance in single-parameter analysis, their substantial impact in masking experiments suggests their value lies primarily in synergistic interactions with other parameters rather than as standalone markers.

Forward sequential feature selection was employed to evaluate the interaction and relative importance of different parameter combinations. A total of 136 parameter combinations were assessed, with detailed results presented in Supplementary Table S5.

Table 3
Performance model K on independent validation datasets.

Independent Validation Dataset	NTUCC	RPCCC	VGH	UPMC	NTUH	All Sites
Panel	Euroflow AML/MDS	AML MRD LDT	LDT	LDT	LDT	5 panels
Parameters (n)	**16 common parameters[a] across all panels from training datasets**					
Instrument	BD FACSLyric	BD FACSCantoII	BD FACSCantoII	BD FACSCantoII	BD FACSCantoII	2 models
Sample Number (AML, non-neoplastic)	16, 16	18, 17	6, 31	25, 25	25, 17	90, 106
AUC	100.00 %	100.00 %	100.00 %	96.96 %	98.12 %	98.71 %
Accuracy	96.88 %	97.14 %	97.30 %	90.00 %	90.48 %	93.88 %
Sensitivity	93.75 %	100.00 %	83.33 %	96.00 %	100.00 %	96.67 %
Specificity	100.00 %	94.12 %	100.00 %	84.00 %	76.47 %	91.51 %
FNR	6.25 %	0.00 %	16.67 %	4.00 %	0.00 %	3.33 %
FPR	0.00 %	5.88 %	0.00 %	16.00 %	23.53 %	8.49 %

Abbreviation: NTUCC: National Taiwan University Cancer Center, RPCCC: Roswell Park Comprehensive Cancer Center, VGH: Taichung Veterans General Hospital, UPMC: University of Pittsburgh Medical Center, NTUH: National Taiwan University Hospital, LDT: laboratory developed test, AUC: area under the receiver operating characteristic curve, FNR: false negative rate, FPR: false positive rate.

[a] Common parameters: FSC-A, FSC-H, SSC-A, CD7, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117, and HLA-DR.

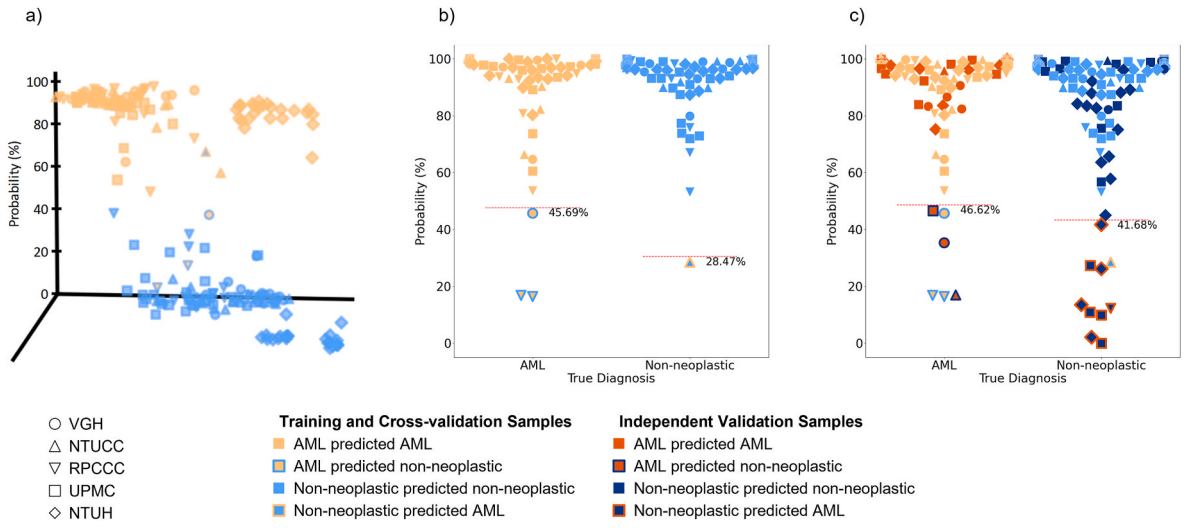


Fig. 3. Prediction visualization of Model K. (a) 3D visualization with principal component analysis (PCA) -compressed features and prediction probabilities. (b) Ranked prediction probabilities. (c) Ranked prediction probabilities of both original datasets and independent validation datasets. Each dot represents a FC case. Wrongly predicted cases are noted with opposite color on the edge.

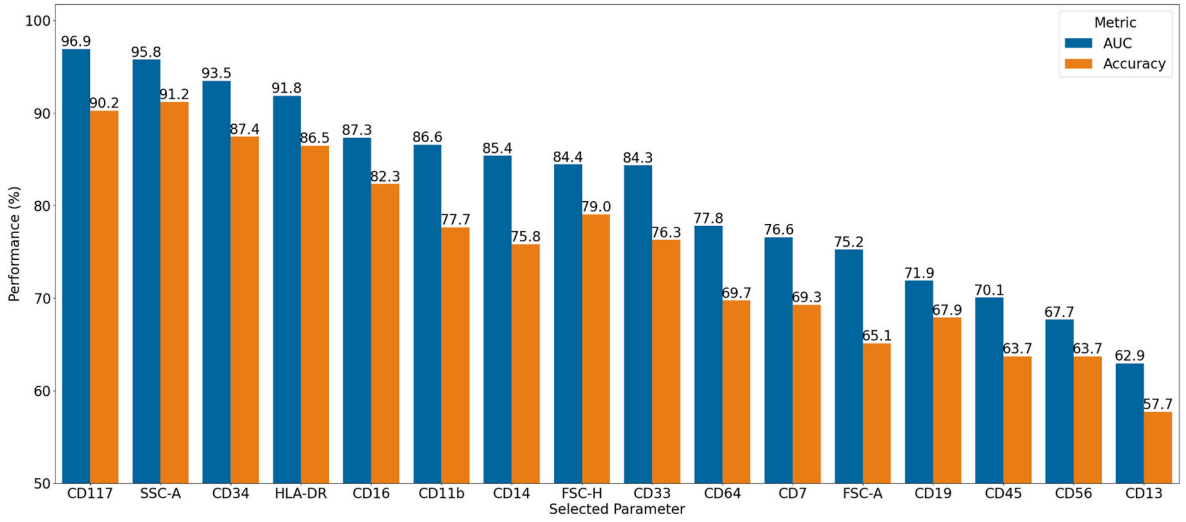


Fig. 4. Results of Single Parameter Selected experiment. This experiment assesses the importance of each parameter by training and cross-validating the presented workflow with only one marker.

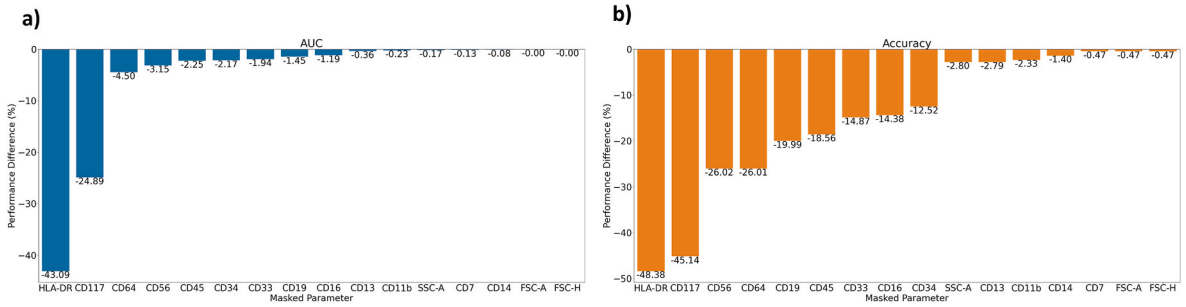


Fig. 5. Performance difference of Single Parameter Masked experiment and Model K. This experiment assesses the parameter importance by cross-validating Model K (trained with all dataset and common parameters) with one parameter masked with zero value. The more performance dropped, the more important the parameter was.

Table 4 summarizes the best-performing parameter combination for each parameter count. Using CD117 alone, the model achieved an AUC of 96.92 %, consistent with its strong performance in single-parameter selection. As additional parameters were incorporated, the AUC remained relatively stable with minor fluctuations. A seven-parameter subset (FSC-A, SSC-A, CD11b, CD13, CD33, CD56, CD117) achieved

Table 4
Results of forward sequential feature selection.

Count	Parameter Subset	Added	AUC	ACC
1	CD117	CD117	96.92 %	90.23 %
2	SSC-A, CD117	SSC-A	99.01 %	94.88 %
3	SSC-A, CD11b, CD117	CD11b	99.97 %	98.6 %
4	SSC-A, CD11b, CD13, CD117	CD13	100.0 %	99.53 %
5	FSC-A, SSC-A, CD11b, CD13, CD117	FSC-A	100.0 %	99.53 %
6	FSC-A, SSC-A, CD11b, CD13, CD33, CD117	CD33	100.0 %	98.60 %
7	**FSC-A, SSC-A, CD11b, CD13, CD33, CD56, CD117**	**CD56**	**100.0 %**	**99.54 %**
8	FSC-A, SSC-A, CD11b, CD13, CD33, CD34, CD56, CD117	CD34	100.0 %	98.14 %
9	FSC-A, FSC-H, SSC-A, CD11b, CD13, CD33, CD34, CD56, CD117	FSC-H	100.0 %	98.60 %
10	FSC-A, FSC-H, SSC-A, CD11b, CD13, CD16, CD33, CD34, CD56, CD117	CD16	100.0 %	98.60 %
11	FSC-A, FSC-H, SSC-A, CD11b, CD13, CD16, CD33, CD34, CD45, CD56, CD117	CD45	100.0 %	99.07 %
12	FSC-A, FSC-H, SSC-A, CD11b, CD13, CD14, CD16, CD33, CD34, CD45, CD56, CD117	CD14	100.0 %	98.60 %
13	FSC-A, FSC-H, SSC-A, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD117	CD19	100.0 %	98.61 %
14	FSC-A, FSC-H, SSC-A, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117	CD64	99.97 %	98.61 %
15	FSC-A, FSC-H, SSC-A, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117, HLA-DR	HLA-DR	99.97 %	98.15 %
16	FSC-A, FSC-H, SSC-A, CD7, CD11b, CD13, CD14, CD16, CD19, CD33, CD34, CD45, CD56, CD64, CD117, HLA-DR	CD7	99.82 %	98.15 %

Abbreviation: AUC: area under the receiver operating characteristic curve, ACC: Accuracy.

optimal performance, with an AUC of 100 % and accuracy and 99.54 %, and slightly surpassed the original parameter set.

4. Discussion

4.1. Panel-agnostic framework and parameter reduction

Achieving consensus on flow cytometry panel compositions for similar clinical applications remains a challenge due to variability in laboratory practices, instrumentation, and evolving technological advancements. Recognizing this variability, we propose a flexible, panel-agnostic framework that accommodates differences in marker panels and instrument models while providing standardized classification solutions to reduce the analytical burden on laboratories.

Previous studies from our group have demonstrated that machine learning (ML)-based classification of acute myeloid leukemia (AML) samples can maintain high performance even with reduced parameter sets [16,17,19]. For instance, our UPMC single-center study showed that a model trained on just four parameters (FSC-A, FSC-H, SSC-H, and CD117) achieved 91.9 % accuracy, compared to 94.2 % accuracy when utilizing all 37 parameters [17]. In the current study, our panel-agnostic approach using 16 common parameters demonstrated equivalent performance across both training datasets (Table 2) and independent validation datasets (Table 3). These findings underscore the potential for further parameter reduction while maintaining robust classification performance across diverse laboratory settings.

4.2. Model performance and feature selection

The receiver operating characteristic (ROC) curve analysis (Fig. 2) demonstrated the flexibility of classification thresholds, which can be adjusted to optimize either specificity (minimizing false positives) or sensitivity (minimizing false negatives). When combined with the sample visualization approach (Fig. 3), the model's predictions and associated probabilities provide valuable decision support for objective, sample-level classification and efficient prioritization of cases requiring manual review.

Feature selection analysis of Model K (Fig. 4) identified CD117, HLA-DR, CD34, SSC-A, and CD16 as particularly informative parameters, each capable of achieving >85 % AUC in AML versus non-neoplastic classification independently. Parameter masking experiments (Fig. 5) revealed that individual masking of HLA-DR, CD117, CD64, and CD56 resulted in the most substantial performance decreases, with AUC reductions and accuracy drops exceeding 20 %. Forward feature selection analysis (Table 4) further confirmed the importance of CD117, SSC-A, and CD56 among the 16 common parameters. These findings align with the biological significance of these markers: CD117 (c-lit), a tyrosine kinase receptor expressed on early myeloid progenitor, serves as a crucial role for confirming the myeloid origin of blast cells, determining their maturation stage, and even harboring therapeutic implications with tyrosine kinase inhibitors [22,23]. CD34 is a marker of hematopoietic stem and early progenitor cells, and its expression indicates a high proportion of immature blasts, which is a common feature of many AML cases [24]. CD16 expression is typically associated with mature neutrophils and natural killer cells. In the context of AML, its expression helps determine the degree of differentiation, especially in cases where blasts show granulocytic maturation [24]. HLA-DR, a major histocompatibility complex class II antigen, is instrumental in distinguishing AML subtypes. Although most AML cases express HLA-DR, its absence is a hallmark of acute promyelocytic leukemia (APL), a subtype with unique clinical intervention [25]. CD56 is predominantly expressed in NK cells [22]. Moreover, previous studies have demonstrated that aberrant CD56 expression in AML correlates with poor prognosis [26,27]. Taking together, CD34 confirms the immature nature of the blasts, HLA-DR helps in distinguishing AML subtypes (notably the HLA-DR-negative APL), CD117 supports the myeloid lineage identification and can hint at underlying genetic mutations, CD16 indicates the degree of differentiation in granulocytic lineage cells, and CD56 identifies aberrant expression patterns of AML blasts. It is therefore not surprising that these markers are critical for our cross-institute model performance.

Notably, quantum yield of the fluorochromes is not likely to impact the model performance of our algorithm. With the information on the fluorochromes used for the top-performing single features, i.e. CD117, CD34, HLA-DR and CD16, we can observe that the markers were conjugated to fluorochromes with varying brightness levels (dim to bright) across centers (Supplement Table S6). The finding suggests that the brightness or quantum yield of the fluorophore alone may not significantly account for classification performance. Instead, our data processing framework, coupled with the inherent characteristics of the Gaussian Mixture Model, appears to effectively mitigate fluorescence intensity variability. This enables the encoding of biologically relevant features from the data, resulting in robust sample-level classification regardless of fluorochrome intensity.

Model performance variability was still observed to some extent in our study and can be attributed to certain intrinsic differences between institutions. For example, the average event number per tube was significantly lower in the UPMC dataset, which may have affected sample representation (Tables 1a and 1b). Additionally, the FCS file format used in the NTUH dataset was FCS2.0, whereas other datasets employed FCS3.0 or FCS3.1. These differences, along with other site-specific factors such as panel variability, sample handling, and instrument calibration, may partially explain why the cross-panel model exhibited differing performance across subsets. One practical approach to mitigate these site-specific differences is to implement institution-specific probability thresholds when applying the model. As illustrated in Fig. 2, adjusting the classification threshold allows for an optimized balance between sensitivity and specificity based on local validation

results. Through this customized calibration approach, institutions can effectively manage their predominant challenges by selectively reducing false positives or false negatives according to their specific needs. Moving forward, we can enhance cross-site performance consistency through optimized harmonization strategies and model refinement using expanded, multi-institutional datasets.

4.3. Analysis of misclassified cases

To investigate potential sources of discordance, we performed a detailed manual review of cases misclassified by Model K. In the cross-validation analysis, four discordant cases were identified: three AML cases incorrectly classified as non-neoplastic (with probabilities of 16.42 %, 16.82 %, and 45.69 %, respectively) and one non-neoplastic case misclassified as AML (with a probability of 28.47 %, Supplementary Table S3). Analysis of the independent validation datasets revealed additional misclassifications, including three AML cases incorrectly identified as non-neoplastic (with probabilities of 16.95 %, 35.31 %, and 46.62 %, respectively) and nine non-neoplastic cases from RPCCC, UPMC, and NTUH that were misclassified as AML (with probabilities ranging from 0 % to 41.68 %, Supplementary Table S4). Notably, all misclassifications occurred with relatively low prediction probabilities, suggesting that implementing probability thresholds could effectively flag these cases for manual review.

Selected scatter plots from the manual analysis are provided in Supplementary Fig. S2. A breakdown of misclassified cases is detailed below.

4.4. AML cases misclassified as non-neoplastic

A total of six AML cases were misclassified as non-neoplastic: three in the cross-validation set (Supplementary Table S3) and three in the independent validation set (Supplementary Table S4). Within the RPCCC dataset, two AML cases were misclassified (Supplementary Fig. S2a–1 & 2); both displayed a higher proportion of monocytes that were CD64-positive but only partially CD14-positive. Similarly, two AML cases from VGHTC were also misclassified (Supplementary Fig. S2a–3 & 4). One case revealed a prominent abnormal population within the CD45dim region, which was positive for CD13, CD34, CD117, and CD56, but showed dim expression of HLA-DR (Supplementary Fig. S2a–3). The other case suffered from poor specimen quality, with over 50 % of events negative for CD45, while the abnormal population was positive for CD13, CD34, CD117, CD33, CD7, and HLA-DR (Supplementary Fig. S2a–4). Additionally, one NTUHCC case was misclassified, exhibiting a $CD34^{-}$, HLA-DR- phenotype while being positive for CD117, CD11b, CD13, CD33, and CD38 in the CD45dim gate (Supplementary Fig. S2a–5). Furthermore, one AML case from UPMC was misclassified as non-neoplastic. However, a manual review concluded that there was no flow evidence of leukemia (Supplementary Fig. S2a–6). This case illustrates how the AI prediction model can assist in identifying instances that may be misinterpreted during routine evaluations.

Excluding the one case requiring diagnosis revision to non-neoplastic disease, the blast percentage of the remaining 5 misclassified cases showed no statistically significant difference compared to AML cases in the cross- and independent validation cohorts (median ± standard deviation: 38.8 ± 18.3 % vs. 43.1 ± 14.8 %), suggesting blast percentage was not a critical predictive feature for the model. Notably, after reviewing the immunophenotypes of our mis-classified cases and independent test ones, we found that 60 % (3/5) of the misclassified AML cases lacked CD34 expression, which was higher than the overall frequency 23.3 % (21/90) in our independent validation dataset. Given that CD34 emerged as a heavily weighted feature for AML classification, its absence likely contributed significantly to these misclassification events, highlighting a model limitation in identifying CD34-negative cases. Similarly, HLA-DR-negative cases were overrepresented among misclassifications, with 40 % (2/5) lacking HLA-DR expression compared to 14.4 % (13/90) in the validation cohort. CD117-negative expression was observed in one misclassified case (20 %) versus 16.6 % (15/90) in the validation dataset. In general, reduced expression of $CD34^{-}$ and HLA-DR in leukemia cells seemed to impact substantially on model performance, likely due to their relative underrepresentation in our AML dataset. To address this limitation in the future, future iterations of the model could benefit from a more balanced cohort enriched with CD34-negative and/or HLA-DR-negative AML cases, thereby improving classification accuracy for these clinically important subtypes.

4.5. Non-neoplastic cases misclassified as AML

A total of ten non-neoplastic cases were misclassified as AML: one in the cross-validation set (Supplementary Table S3) and nine in the independent validation set (Supplementary Table S4).

One misclassified case from NTUHCC was examined through manual review and revealed sparse myeloid progenitors, but marker expression was consistent with normal epitope density, providing no evidence of leukemia (Supplementary Fig. S2b–1). Another misclassified case from RPCCC was determined to be B-cell Acute Lymphoblastic Leukemia (B-ALL) upon review. The abnormal population in this case exhibited features of CD45 dim, CD13 dim, CD33 high, CD34 positive, CD38 bright, CD10 bright, and CD19 positive, consistent with a B-ALL immunophenotype rather than AML (Supplementary Fig. S2b–2).

Four UPMC non-neoplastic cases were misclassified as AML. After manual review, two cases showed no evidence of acute leukemia (Supplementary Fig. S2b–3 & 6), one case was consistent with myelodysplastic syndrome (MDS), with an abnormal population showing distinct immunophenotypes (Supplementary Fig. S2b–4), and one case exhibited a paucity of granulocytes and a B-cell population that was insufficient to definitively support a diagnosis of leukemia (Supplementary Fig. S2b–5).

Lastly, four NTUH non-neoplastic cases were misclassified as AML. Upon manual review, it was found that two cases were consistent with AML (Supplementary Fig. S2b–7 & 8), one case displayed no evidence of acute leukemia (Supplementary Fig. S2b–9), and one case showed an atypical myeloid population that could not definitively support a diagnosis of leukemia (Supplementary Fig. S2b–10).

4.6. Clinical implication and future direction

In conclusion, our findings demonstrate that machine learning-based approaches can effectively identify complex patterns and abnormalities in flow cytometry data while maintaining robust performance across diverse laboratory settings. The panel-agnostic framework we have developed offers several significant advantages: it enables the implementation of simplified flow cytometry panels, ensures consistent analysis quality, and accommodates variations in panel composition and instrumentation across different laboratories. This flexibility allows operators to benchmark their results against expertly curated datasets, providing valuable quality assurance support. Moreover, the framework's ability to maintain high diagnostic accuracy with reduced parameter sets makes it particularly valuable in resource-limited environments, where it can significantly decrease the analytical burden on laboratory personnel without compromising diagnostic standards. Future work will focus on extending this framework to support different clinical applications, such as multiple hematological malignancies classification and disease monitoring, further enhancing its utility across various clinical and research settings.

CRediT authorship contribution statement

Yu-Fen Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration,

Investigation, Data curation, Conceptualization. **En-Ping Chu:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Fong-Ci Lin:** Writing – review & editing, Writing – original draft, Visualization. **Huan-Yu Chen:** Writing – review & editing, Writing – original draft, Software, Methodology. **Tsung-Chih Chen:** Project administration, Investigation, Data curation. **Joseph Hanson:** Project administration, Data curation. **Joseph D. Tario:** Project administration, Investigation, Data curation. **Kai Fu:** Investigation, Data curation. **Sara A. Monaghan:** Investigation, Data curation. **Paul K. Wallace:** Writing – review & editing, Visualization, Validation, Supervision, Investigation. **Chi-Chun Lee:** Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Bor-Sheng Ko:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Ethics statement

This study was approved by the Institutional Review Boards (IRBs) in accordance with the Declaration of Helsinki. The IRB approval codes for each institution are as follows: RPCCC-STUDY00001445, TCVGH-SE20116A, NTUCC-202204104RS, UPMC-STUDY21080145, and NTUH-201906018RINB, while informed consents are waived in all approvals.

Declaration of generative AI in scientific writing

During the preparation of this work the author(s) used GPT4o in order to correct grammar, remove redundancies and improve sentence structure. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Funding

This study received funding from AHEAD Medicine Corporation, San Jose, CA, United States

Declaration of competing interest

The study is conducted with the collaboration of industry (AHEAD Intelligence Ltd, Taipei, Taiwan and AHEAD Medicine Corporation, San Jose, CA, U.S.) and hospitals.

Paul K. Wallace is a consultant to AHEAD Medicine Corporation, San Jose, CA, U.S.

Acknowledgment

We thank Dr. Chieh-Lin Teng from Taichung Veteran General Hospital for his support of this multi-disciplinary collaboration; and we also thank Dr. Chi-Cheng Li (National Taiwan University Hospital, Taipei, Taiwan), Dr. Su Hwei Lee (National Taiwan University Hospital, Taipei, Taiwan) and Dr. Shao-Min Han (National Taiwan University Cancer Center, Taipei, Taiwan) for providing their expert opinion on flow cytometry data interpretation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.110394.

References

[1] E. Garcia, et al., The American society for clinical pathology 2022 vacancy survey of medical laboratories in the United States, Am. J. Clin. Pathol. 161 (3) (2024) 289–304.
[2] E. Garcia, et al., The American Society for Clinical Pathology's 2014 vacancy survey of medical laboratories in the United States, Am. J. Clin. Pathol. 144 (3) (2015) 432–443.
[3] A. Cossarizza, et al., Guidelines for the use of flow cytometry and cell sorting in immunological studies (second edition), Eur. J. Immunol. 49 (10) (2019) 1457–1973.
[4] T. Oldaker, et al., ICCS/ESCCA consensus guidelines to detect GPI-deficient cells in paroxysmal nocturnal hemoglobinuria (PNH) and related disorders part 4 - assay validation and quality assurance, Cytometry B Clin Cytom 94 (1) (2018) 67–81.
[5] K.T. Soh, et al., Evaluation of multiple myeloma measurable residual disease by high sensitivity flow cytometry: an international harmonized approach for data analysis, Cytometry B Clin Cytom 102 (2) (2022) 88–106.
[6] M. Heuser, et al., Update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD Working Party, Blood 138 (26) (2021) 2753–2767, 2021.
[7] G.J. Schuurhuis, et al., Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party, Blood 131 (12) (2018) 1275–1291.
[8] M.J. Borowitz, et al., Measurable residual disease detection in B-acute lymphoblastic leukemia: the children's oncology group (COG) method, Curr Protoc 2 (3) (2022) e383.
[9] A. Emmaneel, et al., PeacoQC: peak-based selection of high quality cytometry data, Cytometry A 101 (4) (2022) 325–338.
[10] S. Van Gassen, et al., CytoNorm: a normalization algorithm for cytometry data, Cytometry A 97 (3) (2020) 268–278.
[11] S. Van Gassen, et al., FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data, Cytometry A 87 (7) (2015) 636–645.
[12] J.A. DiGiuseppe, et al., PhenoGraph and viSNE facilitate the identification of abnormal T-cell populations in routine clinical flow cytometric data, Cytometry B Clin Cytom 94 (5) (2018) 588–601.
[13] W. Dinalankara, et al., Comparison of three machine learning algorithms for classification of B-cell neoplasms using clinical flow cytometry data, Cytometry B Clin Cytom 106 (4) (2024) 282–293.
[14] N. Aghaeepour, et al., Critical assessment of automated flow cytometry data analysis techniques, Nat. Methods 10 (3) (2013) 228–238.
[15] M. Zhao, et al., Hematologist-level classification of mature B-cell neoplasm using deep learning on multiparameter flow cytometry data, Cytometry A 97 (10) (2020) 1073–1080.
[16] B.S. Ko, et al., Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome, EBioMedicine 37 (2018) 91–100.
[17] S.A. Monaghan, et al., A machine learning approach to the classification of acute leukemias and distinction from nonneoplastic cytopenias using flow cytometry data, Am. J. Clin. Pathol. 157 (4) (2022) 546–553.
[18] B.S. Ko, et al., Artificial intelligence-assisted classification tool for minimal residual disease (MRD) assessment by flow cytometry in patients with acute myeloid leukemia (AML), International Clinical Cytometry Society. Baltimore, MD (2021).
[19] Y.F. Wang, et al., Using artificial intelligence to interpret clinical flow cytometry datasets for automated disease diagnosis and/or monitoring, Methods Mol. Biol. 2779 (2024) 353–367.
[20] J.L. Li, et al., A chunking-for-pooling strategy for cytometric representation learning for automatic hematologic malignancy classification, IEEE J Biomed Health Inform 26 (9) (2022) 4773–4784.
[21] D.A. Arber, et al., The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia, Blood 127 (20) (2016) 2391–2405.
[22] H.T. Maecker, J.P. McCoy, R. Nussenblatt, Standardizing immunophenotyping for the human immunology Project, Nat. Rev. Immunol. 12 (3) (2012) 191–200.
[23] S.G.C. Mestrum, et al., The potential of proliferative and apoptotic parameters in clinical flow cytometry of myeloid malignancies, Blood Adv. 5 (7) (2021) 2040–2052.
[24] W. Li, Flow cytometry in the diagnosis of leukemias, in: W. Li (Ed.), Leukemia, 2022. Brisbane (AU).
[25] V.T. Tran, et al., The diagnostic power of CD117, CD13, CD56, CD64, and MPO in rapid screening acute promyelocytic leukemia, BMC Res. Notes 13 (1) (2020) 394.
[26] N. Iriyama, et al., CD56 expression is an independent prognostic factor for relapse in acute myeloid leukemia with t(8;21), Leuk. Res. 37 (9) (2013) 1021–1026.
[27] M. Breccia, et al., Aberrant phenotypic expression of CD15 and CD56 identifies poor prognostic acute promyelocytic leukemia patients, Leuk. Res. 38 (2) (2014) 194–197.