LETTER TO THE EDITOR

Accurate automated diagnosis of B-acute lymphoblastic leukemia using deep learning and flow cytometry

Authors

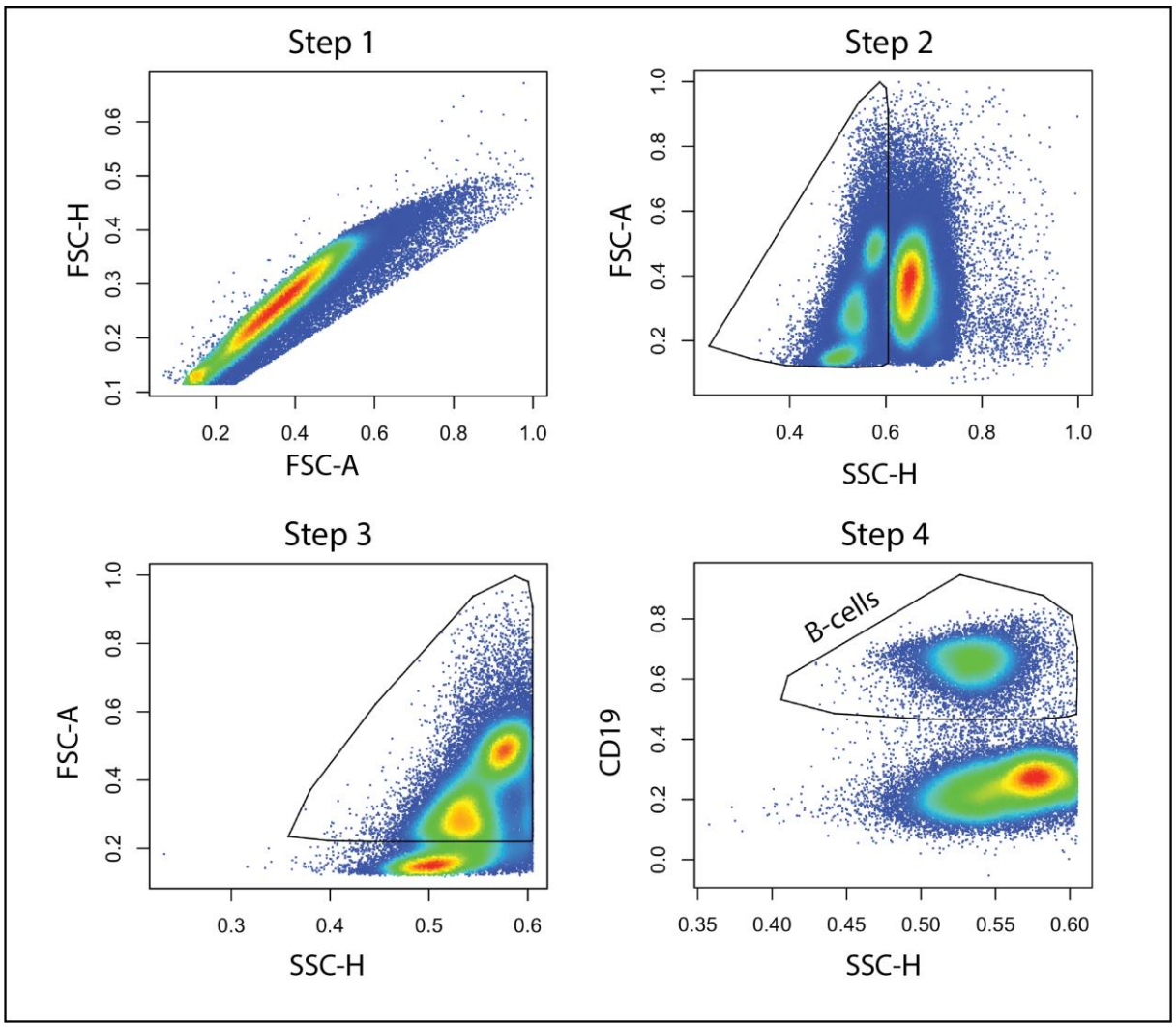
Sulov Chalise,[1] Mikhail Roshal,[1] Sophia Roshal,[2] Jeeyeon Baik,[1] Qi Gao,[1] Anyi Li,[3] Ahmet Dogan,[1] Harini Veeraraghavan[3] and Meng-Lei Zhu[1]

[1]Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; [2]Carnegie Mellon University, Pittsburgh, PA and [3]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Correspondence:
H. VEERARAGHAVAN - veerarah@mskcc.org
M. ZHU- zhum1@mskcc.org

https://doi.org/10.3324/haematol.2025.288277

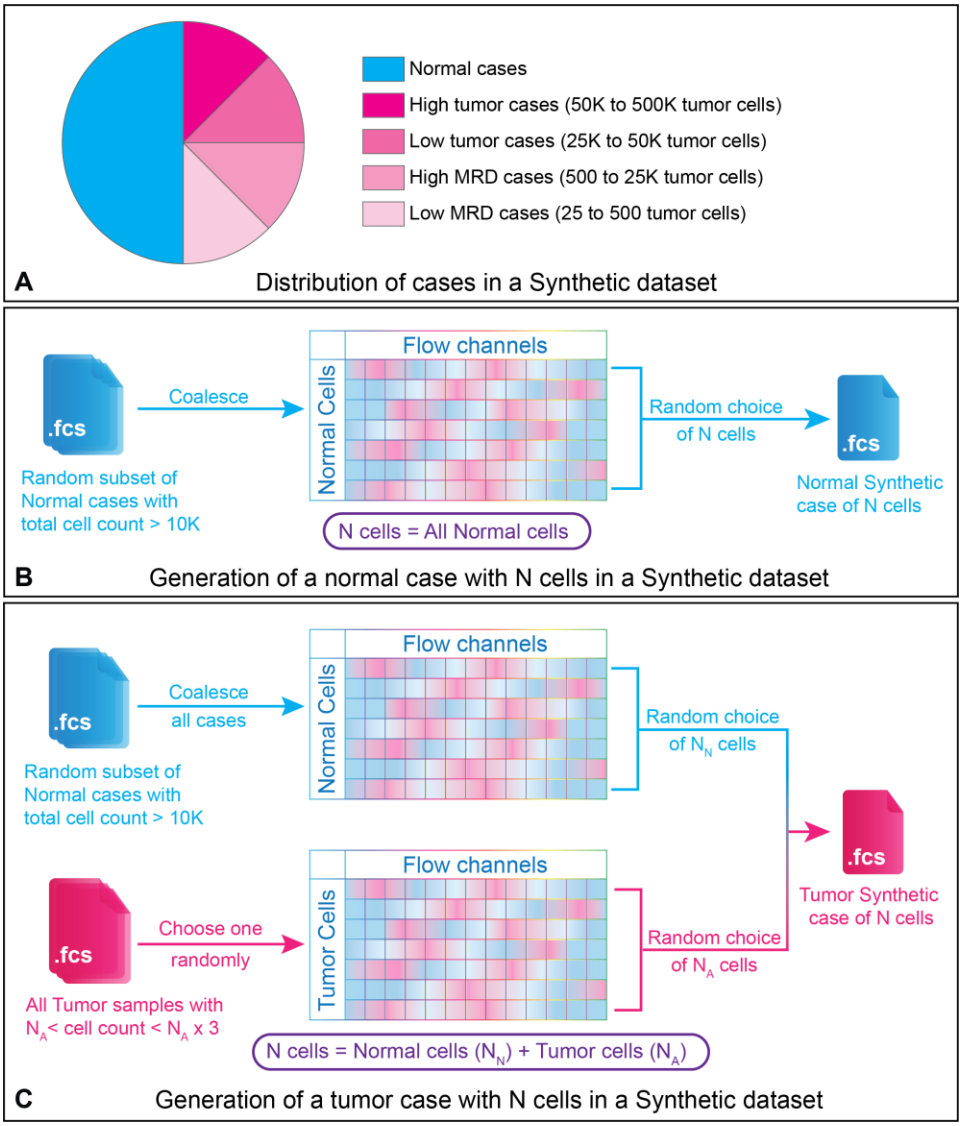
Received: May 30, 2025.
Accepted: December 24, 2025.
Early view: January 8, 2026.



Supplementary Figure 1

Automated B-cell extraction via flowDenstiy

A four-step automated gating strategy for CD19+ B-cell isolation from flow cytometry data using flowDensity. Step 1: Quality control filtering removes margin events and doublets. Margin events are excluded using empirically defined FSC and SSC minimum/maximum thresholds, while doublets are identified and removed based on FSC-A/FSC-H ratio exceeding median plus four standard deviations. Steps 2-3: Sequential gating isolates viable mononuclear cell populations using side scatter parameters (Step 2: SSC-H <0.605; Step 3: SSC-A >0.22). Step 4: CD19+ B-cells are identified from the mononuclear population using density-based clustering with a CD19 expression threshold >0.3. All threshold values were optimized by comparing automated gating results to manual expert annotations across a validation set of 15 randomly selected patient cases, ensuring concordance between automated and manual B-cell populations. This standardized approach enables consistent, reproducible B-cell extraction across large-scale flow cytometry datasets without manual intervention.



Supplementary Figure 2

Synthetic Dataset Generation Strategy for Sample-Level Model Training

Construction of balanced synthetic flow cytometry datasets for training, validation, and testing of the sample-level module. (A) Distribution of cases in synthetic datasets: Each dataset maintains a 50:50 balance between normal and tumor-positive cases, with tumor cases further stratified into four burden categories: low MRD (25-500 tumor cells), high MRD (500-25,000 tumor cells), low tumor burden (25,000-50,000 tumor cells), and high tumor burden (50,000-500,000 tumor cells). A minimum of 25 tumor cells was required for all positive cases. (B) Normal case generation workflow: Normal cases are synthesized by: (1) randomly selecting multiple normal patient samples with combined cell count >10,000, (2) coalescing these samples, and (3) randomly sampling N cells (ranging from 10,000-50,000) to create individual synthetic normal cases. This approach captures the natural variability in cell counts and phenotypes observed in clinical samples. (C) Tumor case generation workflow: Synthetic tumor cases combine normal and tumor cell populations by: (1) generating a normal B-cell background (NN cells) using the process from panel B, (2) selecting a tumor sample containing NA < cell count < 3× NA tumor cells (where NA is the desired tumor cell count), and (3) randomly sampling NA tumor cells to combine with the normal population, yielding cases with N = NN + NA total cells. This strategy generated 15,000 training, 6,000 validation, and 9,000 testing synthetic samples from annotated patient data, ensuring robust model development across the full spectrum of tumor burdens encountered clinically, from MRD to overt disease.

Module	Input	Output	Key Implementation & Hyperparameters
Cell-level Module	(1 × 12) Each cell with 12 cell markers	(2 × 1) probability for normal and tumor class	Random weight initialization; Loss function: cross-entropy + Orthogonal Projection Loss (gamma=0.5, alpha=1); Trained for 100 epochs, batch size =800, Stochastic gradient descent optimizer with momentum=0.9, weight decay=1e-3; learning rate =0.01 decaying 10% at epochs (5,10,20,30,50,70); class imbalance handled via imbalanced sampler; hyperparams selected by AUROC metric
Sample-level Module	(3 × 7,500 × 14) Matrix of top 7,500 cells after reordering and added cell-level probability for normal and tumor + FFT real & imaginary of that matrix	(2 × 1) probability for normal vs tumor class	Random weight initialization; Loss function: cross-entropy + Orthogonal Projection Loss (gamma=0.5, alpha=1); Trained for 100 epochs, batch size =10, Stochastic gradient descent optimizer with momentum=0.9, weight decay=1e-3; learning rate =0.01 decaying 10% at epochs (5,10,20,30,50,70); FFT added for noise robustness; hyperparams selected by AUROC metric
Quantification Module	Same as sample-level module with the value at [0,0,0] replaced with the pre-truncation cell count for that sample	(1 × 1) value between 0 and 1 multiplied by input size to get absolute tumor count	Same as sample-level module but with Huber loss (delta=0.007); trained only on synthetic B-ALL positive samples; hyperparams selected by R^2 metric

Supplementary Table 1

FlowARC Module Architecture and Training Parameters

Summary of neural network architectures and hyperparameters for the three FlowARC modules.Cell-level Module:Processes individual cells (1×12 feature vector) using 1D ResNet-18 (15) to output binary tumor/normal probabilities (2×1). Trained with combined cross-entropy and Orthogonal Projection Loss (γ=0.5, α=1) to handle class imbalance via imbalanced sampling. Sample-level Module:Analyzes top 7,500 reordered cells as a 3×7,500×14 tensor (cells × markers × probability scores + FFT transformations) using 2D ResNet-101 to classify entire samples as normal or tumor. FFT components enhance noise robustness.Quantification Module:Employs the same architecture as the sample-level module but replaces probability scores with pre-truncation cell counts and uses Huber loss (δ=0.007) for regression to estimate absolute tumor cell numbers. All modules utilize stochastic gradient descent with momentum (0.9), weight decay (10^{-3}), and learning rate scheduling (initial 0.01, 10% decay at epochs 5, 10, 20, 30, 50, 70). Hyperparameters were optimized using AUROC for classification modules and R^2 for the quantification module on validation datasets. The quantification module was trained exclusively on synthetic B-ALL positive samples to ensure accurate tumor burden estimation.