4. SPECIAL CONDITIONS

AUTOMATED PANEL-AGNOSTIC FLOW CYTOMETRIC MINIMAL RESIDUAL DISEASE DETECTION IN MULTIPLE MYELOMA USING QUALITY CONTROL-AWARE MACHINE LEARNING

E. Akar, M. Baysal, S. Cakmak, S. Akpinar, B. Turgut

Hematology Department, Tekirdag Namik Kemal University Hospital, Turkey

https://doi.org/10.3324/haematol.2026.s2.14062

Introduction. Flow cytometric minimal residual disease (MRD) assessment in multiple myeloma is highly operator-dependent, relying on expert-driven manual gating that varies with acquisition depth, panel composition, and fluorochrome configuration. The aim of this study was to evaluate automated MRD detection using a quality control (QC)–aware, panel-agnostic machine learning (ML) framework applied directly to raw flow cytometry data.

Methods. Bone marrow samples were analyzed by multiparametric flow cytometry. Raw, ungated, list-mode event data were exported as CSV files following matrix-based fluorescence compensation. All panels contained core plasma cell markers (CD38, CD138, CD19, CD56, CD200, CD27, CD81, CD45). MRD status was reassessed by two experts and finalized by consensus at a limit of quantification (LOQ) of 10^{-4}. Samples were stratified by acquisition depth into Tier-A (≥500,000 events) for model development and Tier-B (<500,000 events) reserved for independent stress testing. Feature generation followed a fully scripted pipeline applied independently per sample, approximating manual plasma cell and aberrant plasma cell (aPC) gating using distribution-based rules. Acquisition stability, singlet integrity, event depth, distributional features, phenotype correlations, and gate concordance measures were quantified and included as model features. ML analysis followed a staged framework. Stage-1 targeted binary classification of LOQ-positive MRD (≥50 aPC) versus MRD-negative (<20 aPC), excluding positive-not-quantifiable (PNQ; 20–49 aPC) samples from training. Subsequent stages evaluated negative reliability and PNQ behavior. L2-regularized logistic regression and XGBoost models were evaluated using nested cross-validation, with probability thresholds derived in Tier-A and applied unchanged to Tier-B.

Results. A total of 125 samples were analyzed: 79 (63.2%) MRD-positive, 36 (28.8%) MRD-negative, and 10 (8.0%) PNQ. Ninety-three samples (74.4%) were Tier-A and 32 (25.6%) Tier-B. Across all configurations, Stage-1 XGBoost showed the best performance. Tier-A out-of-fold evaluation yielded a PR-AUC of 0.85 and a ROC-AUC of 0.80, with sensitivity 0.93 at a probability threshold of 0.5 (Brier score 0.17). Discrimination was preserved in Tier-B stress testing (PR-AUC 0.95; ROC-AUC 0.85). Dual probability thresholds enabled clinically interpretable rule-out (sensitivity 0.96 at probability 0.36) and rule-in (specificity 0.96 at probability 0.89) strategies. Predicted probability distributions showed substantial overlap between PNQ (median 0.83; IQR 0.61–0.86) and LOQ-positive cases (median 0.88; IQR 0.80–0.90), whereas MRD-negative samples clustered at lower probabilities (median 0.27; IQR 0.23–0.36). Logistic regression features achieved lower but competitive performance (Tier-A PR-AUC 0.76). Multiclass analyses (Stages 3–4) with either model showed poor PNQ discrimination with zero recall.

Conclusions. This study demonstrates that MRD assessment can be automated using a panel-agnostic ML framework applied directly to raw flow cytometry data. Incorporation of acquisition-level QC metrics and biologically anchored, in-sample feature generation reduces operator dependence and supports probabilistic MRD interpretation. Despite limitations related to retrospective design, single-center data, and sample size, this framework provides a reproducible foundation for scalable, operator-independent MRD assessment and warrants external validation.

Abstracts printing tool powered by pagepress.org