Poster #: 1112

XXI IWCLL International Workshop on CLL

Artificial Intelligence–Driven Clustering of Peripheral Blood Cell Morphology Reveals Disease Heterogeneity and Prognostic Relevance in CLL

Yan Wang[1,2],Ziyuan Zhou[1,2],Luomengjia Dai[1,2],Yiwen Hua[1,2],Yujie Wu[1,2],Rong Wang[1,2],Lei Fan[1,2],Jianyong Li[1,2],Juejin Wang[2],Huayuan Zhu[1,2]

1: Department of Hematology, the First Affiliated Hospital with Nanjing Medical University, Jiangsu Province Hospital

2: Nanjing Medical University

OBJECTIVES

- we developed an AI-driven system for objective, high-throughput evaluation of peripheral blood cell morphology in CLL.

CONCLUSIONS

- Our large-scale, AI-driven morphologic analysis reveals not only striking inter-patient heterogeneity in CLL blood cell phenotype but also robust associations between cell morphology, molecular features, and clinical outcomes. This approach offers a scalable avenue for routine, objective, and high-throughput risk monitoring in CLL and provides a template for future multi-center validation and real-world clinical integration.

江苏省人民医院 JIANGSU PROVINCE HOSPITAL 南京医科大学第一附属医院 THE FIRST AFFILIATED HOSPITAL WITH NANJING MEDICAL UNIVERSITY 江苏省妇幼保健院 JIANGSU WOMEN AND CHILDREN HEALTH HOSPITAL

南京医科大学 NANJING MEDICAL UNIVERSITY

INTRODUCTION

- Real-time monitoring is essential for the management of chronic lymphocytic leukemia (CLL) patients. Peripheral blood analysis offers advantages of affordability, convenience, and suitability for longitudinal disease tracking. However, current blood film evaluation methods lack automation and are subject to observer bias, affecting reproducibility and clinical translation. Moreover, CLL exhibits complex morphological and biological heterogeneity, with subtle cellular changes potentially heralding disease progression or transformation. To address these unmet needs, we developed an AI-driven system for objective, high-throughput evaluation of peripheral blood cell morphology in CLL.

METHODS

- Peripheral blood cell morphology were collected from blood films using whole-slide imaging at high resolution. Regions of interest (ROI) were automatically selected using a ResNet-18 based deep neural network classifier, and lymphocyte segmentation and feature extraction were performed via a VGG16 convolutional neural network. Based on this AI-driven system, we calculated the distribution of lymphocyte cell areas and analyzed different patterns of each patient and cell. Both unsupervised clustering and manual pattern recognition were performed based on cell size distribution curves and distinct morphological cell types. Then we systematically correlated morphological clusters and patient cluster types with key molecular markers, laboratory parameters, survival data. A multiple instance learning (MIL) model was built using aggregated cell features to predict clinical and molecular risk profiling.

RESULTS

- Building on our previous work[1], we conducted a large-scale retrospective study of 260 CLL patients at our center, extracting morphological features from over 1.5 million peripheral blood cells using a high-throughput, AI-driven imaging pipeline. Cell size areas analysis revealed patient-specific lymphocyte distribution patterns and peak types, reflecting underlying morphological heterogeneity. Clustering based on cell size distribution patterns successfully stratified patients into groups (e.g., unimodal-narrow, unimodal-broad, and multimodal). (Figure1)

Figure1. Association Between Different Kurtosis Levels and Prognosis

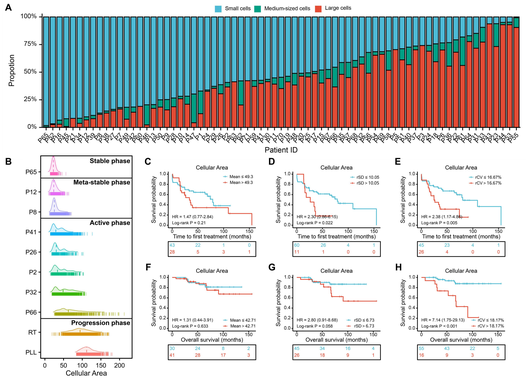
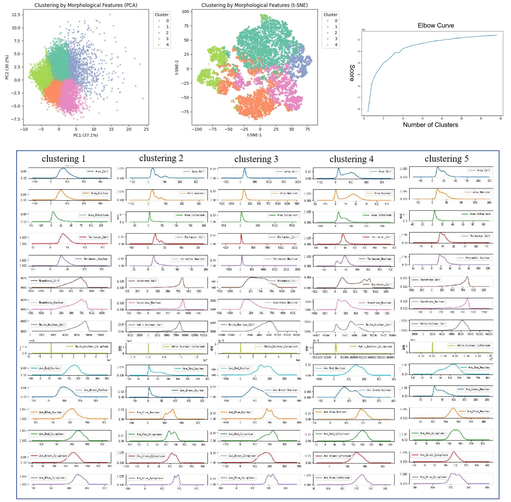


Figure 1. Relationship between cellular-size distribution heterogeneity and clinical prognosis.
(A) Proportion of small, medium-sized, and large cells across patients.
(B) Distribution of cellular area in representative cases at different disease phases.
(C–H) Kaplan–Meier survival curves for time to first treatment and overall survival, stratified by mean cellular area, relative standard deviation (rSD), and relative coefficient of variation (rCV).

Abbreviations: rSD, relative standard deviation; rCV, relative coefficient of variation.

- Notably, patients with broad/multimodal peaks pattern showed significantly shorter overall survival (OS) (P=0.04) and were enriched for high-risk molecular features, including TP53 mutation and complex karyotype, indicating worse prognosis linked to underlying morphological diversity. For each patient, the proportion of different cell categories was quantified, and unsupervised mass clustering at single-cell level defined distinct morphological cell types, resulting in five robust patient clusters. (Figure2)

Figure2. Clustering of Morphological Features into Five Distinct Categories and Inter-Cluster Differences



TFigure 2.
Clustering analysis of cellular morphological features identifies five distinct groups.
Top row: Visualization of clusters by Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE), showing clear separation of the five clusters. Elbow curve indicates the optimal number of clusters.
Bottom panels: Density distributions of morphological parameters within each cluster, highlighting differences in cellular and nuclear area, perimeter, roundness, and color channel intensities. These profiles demonstrate the distinct morphological characteristics that define each cluster.

- This patient-level composition of morphology-based cell clusters was strongly correlated with key molecular risk factors and could inform individualized patient monitoring. (Figure3)

Figure 3. Survival Analysis of CLL Patients Based on Cellular Morphological Clustering

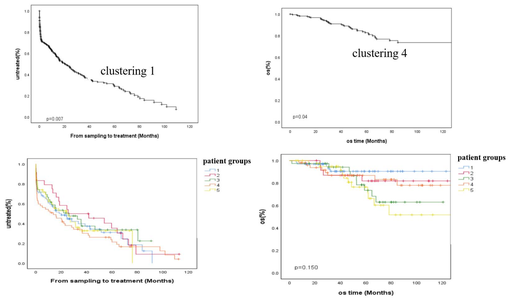


Figure 3.
Upper panels: The relationship between cellular morphological clusters and patient outcomes. Kaplan–Meier analyses show that patients with a higher proportion of cell type 1 are associated with shorter time to first treatment, while those with more cell type 4 are linked to overall survival (OS) in CLL.
Lower panels: Further clustering of patients based on the proportions of the five cellular categories reveals new patient groups. Among these, there is no significant difference in time to first treatment (TFS), but overall survival (OS) differs markedly between groups.

- Furthermore, we developed a MIL model built on aggregated cell features to clinical and molecular features[2,3]. This MIL model achieved an AUC of 0.83 for the prediction of TP53 mutation and 0.68 for complex karyotype, highlighting the translational potential of cell morphology as a noninvasive biomarker for integrated risk prediction in CLL.(Figure4)

Figure 4. Performance of MIL Model for Prediction of Key Molecular Abnormalities in CLL

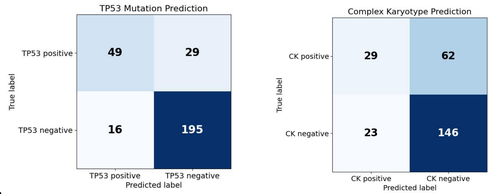


Figure 4.
Confusion matrices showing the predictive performance of the MIL model for Complex Karyotype (CK) (left) and TP53 mutation (right) in 260 CLL patients.

REFERENCES

1. Wang, Y., et al., Enhancing morphological analysis of peripheral blood cells in chronic lymphocytic leukemia with an artificial intelligence-based tool. Leuk Res, 2023. 130: p. 107310.
2. Lu, M.Y., et al., Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng, 2021. 5(6): p. 555-570.
3. Couture HD, M.J.P.C., Multiple instance learning for heterogeneous images: Training a cnn for histopathology, in International Conference on Medical Image Computing and Computer-Assisted Intervention. 2018. p. 254-262.

ACKNOWLEDGMENTS

We would like to thank Bionovation for generously providing the scanners and technical support, which was essential for the data acquisition in this study.