LETTER OPEN

Check for updates

AI-based hematological malignancy prediction from peripheral blood smears in a large diagnostic laboratory cohort

Leukemia; https://doi.org/10.1038/s41375-026-02934-1

BACKGROUND

Hematological malignancies represent a wide range of disease entities, most of which arise from dysfunctional proliferation and differentiation of hematopoietic stem and progenitor cells in the bone marrow [1]. Diagnosis requires integration of cytomorphology, molecular genetics, and immunophenotyping from blood or bone marrow. Unlike bone marrow aspiration, assessing cytomorphology in a blood smear is fast, minimally invasive, and provides information on differential cell counts and morphological abnormalities that guide follow-up diagnostic pathways. However, conventional peripheral blood smear analysis involves labor-intensive manual examination of hundreds of cells, which is subject to inter-observer variability. Previous work explored machine-learning for single-cell classification [2, 3], and disease detection [4–9] on curated cohorts. Systematic evaluation across multiple malignancies at their natural clinical distribution remains unexplored.

METHODS

We retrospectively collected digitized white blood cell images from peripheral blood smears from 6610 individuals (6115 patients and 495 healthy stem cell donors) processed at the Munich Leukemia Laboratory (MLL) between 2021 and 2022, encompassing the full spectrum of hematological malignancies. We grouped 168 MLL diagnostic labels into 19 detailed and 8 coarse classes (Fig. 1A). To develop a model for initial diagnosis, we excluded post-chemotherapy patients and ended up with 1634 for model training and 409 for internal testing (Supplementary Fig. 1A). We also tested the model on 1408 patients with unclear diagnoses or borderline classifications.

Using the DinoBloom foundation model [10], we trained a transformer-based AI model, cAltomorph, to predict the coarse classes of hematological malignancies from peripheral blood smear single-cell images (Fig. 1B). cAltomorph consists of three steps: (i) the DinoBloom encoder extracts 768 features from individual white blood cell images. (ii) A transformer [11] aggregator combines the information from 500 cells into a single 512-dimensional representation. (iii) A classifier outputs probabilities for each of the eight classes. Using 5-fold cross-validation, we trained five models and combined their predictions via ensembling testing on a held-out test set (see Supplementary Methods and Supplementary Fig. 1B).

RESULTS

Our model achieves a high sensitivity and precision for acute leukemia (sensitivity/precision=0.74/0.79) and MPN (0.85/0.76), and a moderate performance for MDS (0.71/0.57) (Fig. 1C). Surprisingly, cAltomorph achieves high sensitivity on plasma cell neoplasms (0.77) despite the fact that these are typically not detectable in blood smears. For lymphoma, the sensitivity is expectedly low (0.40/0.66), as well as for MDS/MPN (0.50/0.69) and reactive changes (0.45/0.56). These disease classes are notoriously difficult to diagnose from peripheral blood. The overall accuracy of the model is 0.72. Two-dimensional UMAP projections of patient embeddings reveal clusters corresponding to coarse disease classes, with a clear split between myeloid and lymphoid branches, and a separation of healthy individuals (Fig. 1D). As an auxiliary task, the model predicts hemoglobin values, which correlate significantly with measured values (Pearson correlation coefficient 0.67, $p = 6 \times 10^{-55}$; mean absolute error (MAE) 1.63; Fig. 1E). cAltomorph's output probabilities match true disease probabilities (expected calibration error of 2.4%, Fig. 1F), ensuring reliable confidence in predictions. The model achieves top-2 accuracy of 0.87 (Supplementary Fig. 2).

By summing the probabilities of all malignant class predictions, our model can predict malignancy and thus guide further testing (Fig. 1B). For binary malignant vs. non-malignant prediction (Fig. 1G), our model achieves a sensitivity of 0.94 and specificity of 0.83, with an AUC of 0.97 (Fig. 1H). In our cohort, 13.5% of patients who underwent bone marrow aspiration were subsequently diagnosed with reactive changes, representing potentially avoidable invasive procedures. By adjusting the model's malignancy threshold, we can balance between detecting every leukemia case and avoiding unnecessary bone marrow aspirations. With a 0.5 malignancy threshold, we can reduce the rate of unnecessary bone marrow aspirations from 13.5% to 8.7% (35% relative reduction) while correctly identifying all 46 acute leukemia cases in the test set (Fig. 1I).

We next analysed cAltomorph's prediction capabilities with respect to the 19 detailed classes. Our model achieved a 0.70 sensitivity on acute myeloid leukemia (AML) cases, which present with myeloblasts in peripheral blood and are thus easier to identify compared to other classes, assigning 9 out of 37 AML cases to MDS (Fig. 2A). To analyse the misclassified cases, we plot the measured myeloblast ratio vs. the disease probability (Fig. 2B). cAltomorph's AML disease probability correlates with the myeloblast ratio in the blood (Spearman correlation coefficient 0.73, $p = 7 \times 10^{-7}$). All misclassified cases have a myeloblast ratio < 20% and are assigned as MDS.

Myeloproliferative neoplasms (MPN) are clonal disorders characterized by proliferation of one or more myeloid lineages, resulting in mature cell overproduction. Our model correctly classifies 11 out of 12 cases with essential thrombocytosis (ET), polycythemia vera (PV), and chronic myeloid leukemia (CML), and achieved 0.83 sensitivity in other MPN cases. In contrast, MDS is characterized by varying degrees of single or multiple cytopenias [1] and dysplasia. cAltomorph effectively distinguishes these differences from

Received: 29 December 2025 Revised: 18 February 2026 Accepted: 11 March 2026
Published online: 23 March 2026

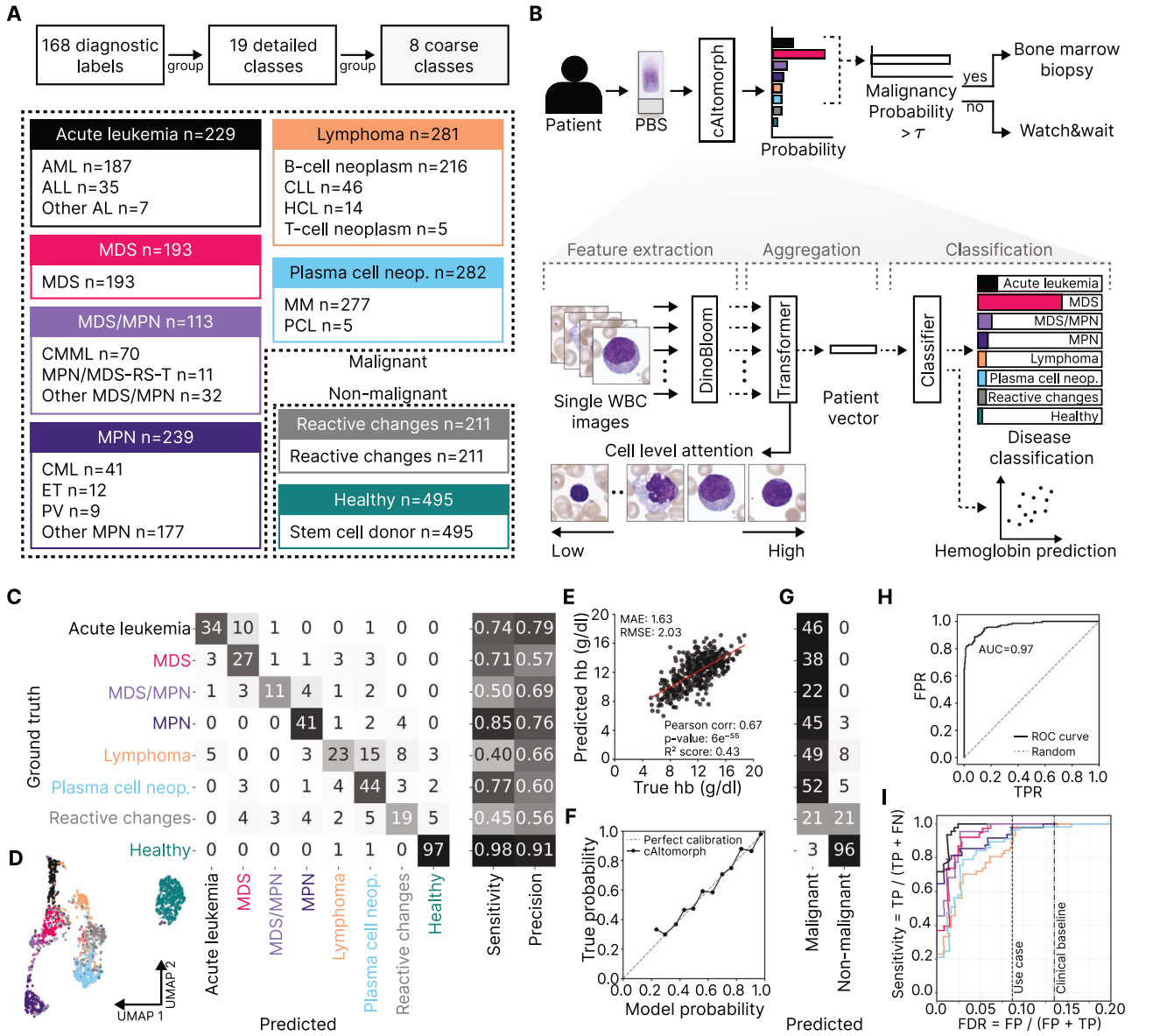
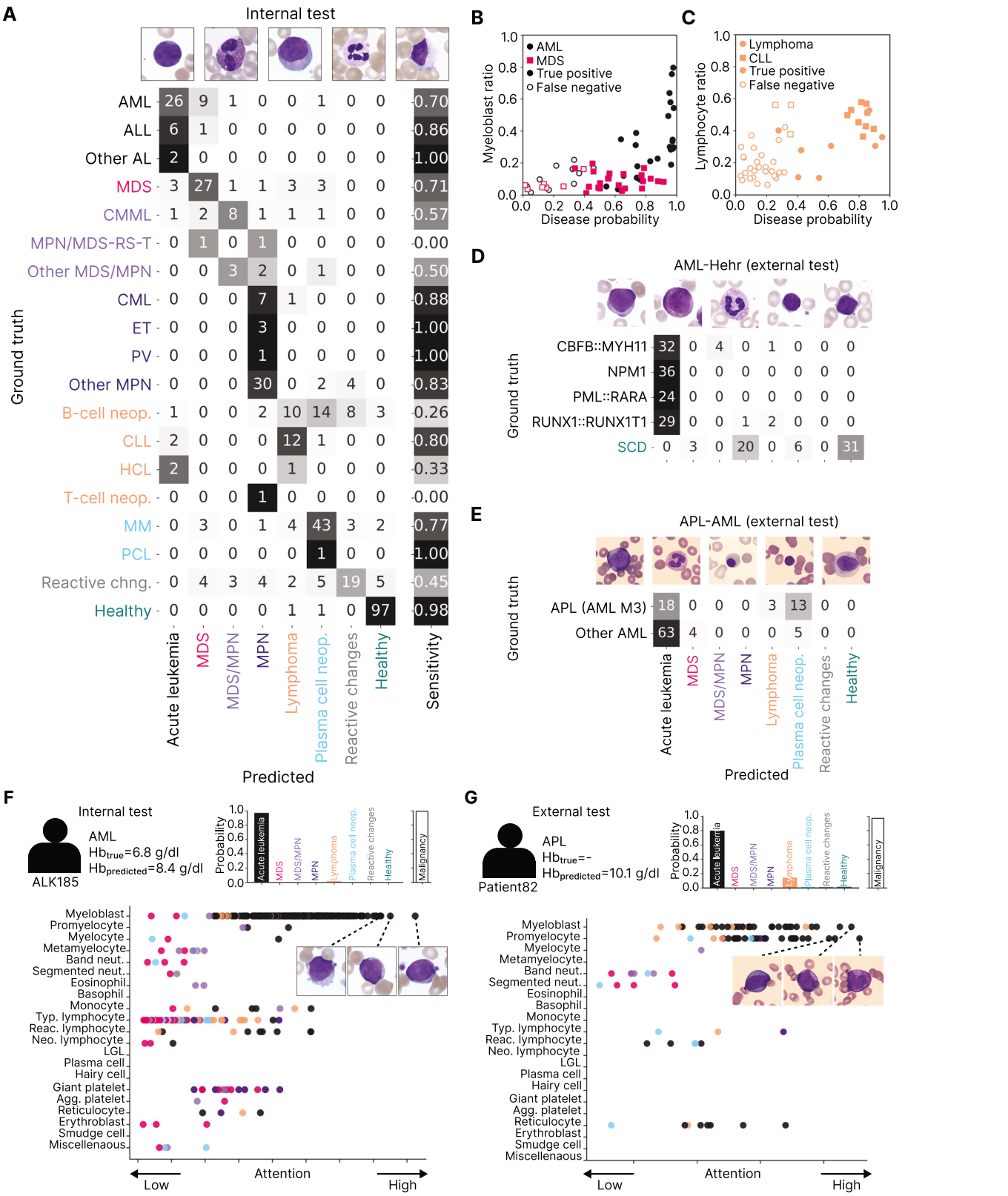


Fig. 1 Our diagnostic laboratory dataset contains over 3.2 million single-cell images from blood smears of 6115 patients and 495 healthy stem cell donors with 19 detailed and 8 coarse hematological disease classes. A 168 diagnostic labels were hierarchically grouped into 19 detailed classes, which were then further consolidated into broader 8 coarse classes. **B** For routine diagnosis, both peripheral blood smears (PBS) and bone marrow samples were obtained from patients. Cytomorphology, DNA sequencing, immunophenotyping and cytogenetics from bone marrow were used to determine the ground truth label. Model architecture: In the feature extracting stage, the DinoBloom foundation model provides a latent space encoding of each single-cell image. During aggregation, a transformer combines these encodings into a single latent space vector. For classification, we employ a multi-layer perceptron to predict probabilities for the 8 coarse classes and hemoglobin values. We compute the malignancy probability by summing the predicted probabilities of the malignant classes (acute leukemia, MDS, MDS/MPN, MPN, lymphoma, plasma cell neoplasm). If it exceeds the malignancy threshold (0.5 in this study), cAltomorph recommends a bone marrow biopsy. Cell-level attention from the transformer aggregator allow identifying diagnostically relevant cells. **C** Our model distinguishes acute leukemias, MPN and healthy patients with a sensitivity and precision of over 0.74. cAltomorph shows surprisingly high performance on MDS and plasma cell neoplasms. **D** UMAP latent space embeddings show well-distinguishable patient clusters corresponding to different disease classes. Colors correspond to the classes shown in the confusion matrix. **E** Predicted and measured hemoglobin correlate with a Pearson coefficient of 0.67. **F** Model-predicted probabilities align well with true disease distributions, indicating good calibration. **G** Confusion matrix for malignancy vs. non-malignancy predictions at a malignancy threshold of 0.5. **H** cAltomorph achieves 0.97 area under the curve for malignant vs. non-malignant separation. **I** By tuning cAltomorph's malignancy threshold, we can reduce the rate of unnecessary bone marrow aspirations (FDR) from 13.5% (Clinical baseline) to 8.7% while maintaining 100% sensitivity for acute leukemia.

peripheral blood cell images, misclassifying only 1 out of 38 MDS cases as MPN (Fig. 2A). However, there is an expected confusion with MDS/MPN cases, rare clonal bone marrow disorders with overlapping features of both MDS and MPN: 50.0% are predicted correctly; 13.6% are predicted as MDS, and 18.2% are predicted as MPN (Fig. 2A).

For lymphoid malignancies, 12 out of 15 CLL cases are correctly classified (0.80 sensitivity), but 26% of lymphoma cases are misclassified as plasma cell neoplasm, predominantly B-cell neoplasms (Fig. 2A). By plotting lymphocyte ratio in blood vs. lymphoma model probability, we find that the model detects lymphoma cases more accurately when lymphocyte counts are



elevated (Fig. 2C). Notably, most of these cases belong to the CLL subtype.

We further evaluate our model on an extended internal test set including patients with unclear/rare diagnoses, double diagnoses, and MGUS cases previously excluded (Supplementary Fig. 4A). Our model is able to classify borderline cases into one of the suspected classes. For instance, out of 11 'MDS-AML borderline' cases, 5 are assigned to acute leukemia and 4 are assigned to MDS; out of 4 'MPN in blast crisis' cases, 3 are assigned to acute leukemia, and 1 to MPN. Notably, cAltomorph classifies 77 out of 199 MGUS cases

Fig. 2 cAltomorph predicts hematological malignancies from peripheral blood images in internal and external datasets while highlighting clinically relevant cells. A cAltomorph achieves high sensitivities in subtypes with visible aberrations or high cellular counts, such as AML, ALL, CLL, ET, and CML. cAltomorph also achieves high sensitivity on multiple myeloma cases, although it is not typically diagnosed from morphology. As expected, performance was lower in disease classes not diagnosed from peripheral blood, such as lymphoma (excluding CLL). The model almost perfectly identified healthy donors, but, as expected, it struggles with patients exhibiting reactive changes. **B** cAltomorph acute leukemia prediction probability correlates with myeloblast ratio in the blood. Acute leukemia patients with a low myeloblast ratio are misclassified as MDS. **C** cAltomorph lymphoma prediction probability correlates with lymphocyte ratio in the blood. The cases with a high lymphocyte ratio belong to the CLL subtype. **D** Model performance on the AML-Hehr dataset. Ground-truth labels include four genetic subtypes of AML and stem cell donors (SCD). cAltomorph identifies all acute leukemia cases with high sensitivity and precision. **E** Confusion matrix for predictions on the APL-AML dataset. **F** Two exemplary patients are presented: one from the internal test set and one from the external APL-AML dataset. For each patient, we provide metadata, predicted hemoglobin, eight-class hematological disease probabilities, malignancy probability, and attention distribution over cells. Individual cells are passed through cAltomorph to better understand their relation to diseases. Cell dots are colored based on model predictions. An AML patient from the internal test set was confidently predicted as acute leukemia by cAltomorph. Myeloblasts receive the highest attention. **G** An APL patient from the APL-AML dataset. Promyelocytes and myeloblasts receive the highest attention, as expected.

as plasma cell neoplasms (39%), which is a precursor condition of multiple myeloma. Predicted hemoglobin values correlate well with measured values (Pearson correlation coefficient 0.66, Supplementary Fig. 4B).

We evaluate cAltomorph's generalizability on two external datasets: AML-Hehr [4] includes four genetic AML subtypes and healthy stem cell donors, while APL-AML [7] comprises only AML patients, grouped into acute promyelocytic leukemia (APL) and other AML subtypes. In AML-Hehr, our model misclassifies only 8 out of 129 AML patients, achieving a sensitivity of 0.94 for acute leukemias and an AUROC of 0.99 in distinguishing malignant from non-malignant samples (Fig. 2D, Supplementary Fig. 3A). For APL-AML, having a distinct staining background, cAltomorph achieves a 0.76 sensitivity for detecting acute leukemias and a 1.0 sensitivity for malignancy detection (Fig. 2E). Notably, external cases aligned closely with internal test samples in a low-dimensional UMAP embedding, forming coherent clusters consistent with disease types (Supplementary Fig. 3B).

To illustrate clinical utility, we visualize cell-level attentions for representative cases in an interactive dashboard at https://github.com/marrlab/cAltomorph. The model successfully assigns high attention to myeloblasts when classifying an acute leukemia case, while assigning minimal attention to typical lymphocytes (Fig. 2F). In an APL patient from the external APL-AML dataset, cAltomorph correctly highlights promyelocytes and myeloblasts (Fig. 2G). In a patient with myeloproliferative neoplasm (MPN), the model highlights giant platelets that can hint at the diagnosis of MPN (Supplementary Fig. 5A). In a patient with *CBFB::MYH11* fusion from the external AML-Hehr dataset, the model identifies myeloblasts and monocytic cells, consistent with the monocytic differentiation typical of this AML subtype (Supplementary Fig. 5B).

CONCLUSION

We developed cAltomorph, a transformer-based model for multi-disease classification of hematological malignancies from peripheral blood smears. In a diagnostic laboratory cohort, the model achieved high accuracy for acute leukemias, myeloproliferative neoplasms, and healthy controls, with robust performance on external validation cohorts. Based on test-set performance, the model may help reduce unnecessary bone marrow aspirations (by up to 35% in our test set) while preserving sensitivity for acute leukemia, although prospective validation is needed. Notably, image-based diagnostics outperformed models based on differential cell counts and demographic variables (age and sex; Supplementary Table 3).

Important limitations are the retrospective, single-center study design, class imbalance affecting the detection of rare diseases, and a healthy control group limited to stem cell donors. Prospective multi-center validation is required to assess real-world clinical utility. Nevertheless, cAltomorph demonstrates the feasibility of AI-based peripheral blood screening as a complementary tool for initial diagnostic triage in hematological malignancies, potentially reducing the burden of invasive procedures while maintaining diagnostic sensitivity.

Muhammed Furkan Dasdelen[1,2,9], Ivan Kukuljan[1,9], Peter Lienemann[1,3], Fatih Ozlugedik[1], Ario Sadafi[1,4], Matthias Hehr[1,5], Karsten Spiekermann[3,6], Christian Pohlkamp[7] and Carsten Marr[1,3,4,6,8]✉

[1]*Institute of AI for Health, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany.* [2]*International School of Medicine, Istanbul Medipol University, Istanbul, Turkey.* [3]*Department of Medicine III, Ludwig-Maximilian-University Hospital, Munich, Germany.* [4]*Munich Center for Machine Learning (MCML), Munich, Germany.* [5]*Dr. von Haunersches Kinderspital, Ludwig-Maximilians-University Munich, Munich, Germany.* [6]*DKTK, German Cancer Consortium, Munich, Germany.* [7]*Munich Leukemia Laboratory, Munich, Germany.* [8]*Department of Physics, University of Munich, Munich, Germany.* [9]*These authors contributed equally: Muhammed Furkan Dasdelen, Ivan Kukuljan.*
✉*email: carsten.marr@helmholtz-munich.de*

DATA AVAILABILITY

We release the test set of our comprehensive peripheral blood cytomorphology image dataset. This data encompasses 409 individuals, categorized as follows: 46 acute leukemia, 38 MDS, 22 MDS/MPN, 48 MPN, 57 lymphoma, 57 plasma cell neoplasm, 42 reactive changes and 99 healthy. The dataset covers the entire spectrum of hematological malignancies, reflecting their relative frequencies within the population. Altogether, it consists of 201,560 blood cell images. The test dataset is available at https://doi.org/10.82296/hmgu-nefeli.9bv4e-3ag16. External test sets are available under the corresponding links: AML-Hehr: https://www.cancerimagingarchive.net/collection/aml-cytomorphology_mll_helmholtz/. APL-AML: https://www.kaggle.com/datasets/eugeneshenderov/acute-promyelocytic-leukemia-apl.

CODE AVAILABILITY

Our code and model weights are available at https://github.com/marrlab/cAltomorph.

REFERENCES

1. Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. Leukemia. 2022;36:1703–19.
2. Matek C, Krappe S, Münzenmayer C, Haferlach T, Marr C. Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. Blood. 2021;138:1917–27.
3. Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nat Mach Intell. 2019;1:538–44.

4. Hehr M, Sadafi A, Matek C, Lienemann P, Pohlkamp C, Haferlach T, et al. Explainable AI identifies diagnostic cells of genetic AML subtypes. PLoS Digit Health. 2023;2:e0000187.
5. Manescu P, Narayanan P, Bendkowski C, Elmi M, Claveau R, Pawar V, et al. Detection of acute promyelocytic leukemia in peripheral blood and bone marrow with annotation-free deep learning. Sci Rep. 2023;13:2562.
6. Eckardt J-N, Middeke JM, Riechert S, Schmittmann T, Sulaiman AS, Kramer M, et al. Deep learning detects acute myeloid leukemia and predicts NPM1 mutation status from bone marrow smears. Leukemia. 2022;36:111–8.
7. Sidhom J-W, Siddarthan IJ, Lai B-S, Luo A, Hambley BC, Bynum J, et al. Deep learning for diagnosis of acute promyelocytic leukemia via recognition of genomically imprinted morphologic features. NPJ Precis Oncol. 2021;5:38.
8. de Almeida JG, Gudgin E, Besser M, Dunn WG, Cooper J, Haferlach T, et al. Computational analysis of peripheral blood smears detects disease-associated cytomorphologies. Nat Commun. 2023;14:1–14.
9. Wei B-H, Tsai XC-H, Sun K-J, Lo M-Y, Hung S-Y, Chou W-C, et al. Annotation-free deep learning for predicting gene mutations from whole slide images of acute myeloid leukemia. NPJ Precis Oncol. 2025;9:35.
10. Koch V, Wagner SJ, Kazeminia S, Sancar E, Hehr M, Schnabel J, et al. DinoBloom: a foundation model for generalizable cell embeddings in hematology. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature; 2024. pp 520–30.
11. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations (ICLR). 2021. https://arxiv.org/abs/2010.11929

ACKNOWLEDGEMENTS

We acknowledge support from the High-Tech Agenda Bayern.

AUTHOR CONTRIBUTIONS

CP, CM and IK conceptualized the study idea. CP, PL and IK curated the data. MFD, IK and FO trained the models, conducted formal analysis, interpreted and visualized the results. IK, MFD, MH and CM. drafted the original manuscript. CM and CP supervised the study. KS and MH provided clinical expertise. KS, AS, MH, PL, and CP critically revised the manuscript. All authors had full access to the data, made a review of the paper, and approved the final submission.

FUNDING

CM received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement 866411 & 101113551). Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS STATEMENT

All experiments were conducted in accordance with the Declaration of Helsinki. The retrospective analysis of images used in this study received approval from the Ethics Committee of the Medical Faculty, Ludwig-Maximilians-Universität (LMU) Munich, Germany (Approval No. 25-0744).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41375-026-02934-1.

Correspondence and requests for materials should be addressed to Carsten Marr.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.