Received: 24 April 2024 | Revised: 14 September 2024 | Accepted: 24 September 2024

DOI: 10.1111/cas.16374

Cancer Science WILEY

ORIGINAL ARTICLE

Diagnosis and typing of leukemia using a single peripheral blood cell through deep learning

Geng Yan[1,2,3] | Gao Mingyang[4] | Shi Wei[3] | Liang Hongping[3] | Qin Liyuan[5] | Liu Ailan[6] | Kong Xiaomei[7] | Zhao Huilan[8] | Zhao Juanjuan[4] | Qiang Yan[1,2,4]

[1]Department of Physiology, Shanxi Medical University, Taiyuan, China

[2]Key Laboratory of Cellular Physiology, Ministry of Education (Shanxi Medical University), Taiyuan, China

[3]Department of Clinical Laboratory, Shanxi Provincial People's Hospital, Taiyuan, China

[4]College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Taiyuan, China

[5]Department of Hematology, Shanxi Provincial People's Hospital, Taiyuan, China

[6]Department of Clinical Laboratory, Second Hospital of Shanxi Medical University, Taiyuan, China

[7]Department of Pulmonary and Critical Care Medicine, First Hospital of Shanxi Medical University, Taiyuan, China

[8]PET/CT Department, Shanxi Coal Center Hospital, Taiyuan, China

Correspondence
Qiang Yan, Department of Physiology, Shanxi Medical University, 56 Xinjiannan Road, Taiyuan, Shanxi 030001, China.
Email: qiangyan@tyut.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 62376183; International (Regional) Cooperation and Exchange Program of the National Natural Science Foundation of China, Grant/Award Number: U21A20469; Scientific Research Special Project for the High-quality Development of the Great Health Industry in Shanxi Province, Grant/Award Number: DJKZXKT2023008

Abstract

Leukemia is highly heterogeneous, meaning that different types of leukemia require different treatments and have different prognoses. Current clinical diagnostic and typing tests are complex and time-consuming. In particular, all of these tests rely on bone marrow aspiration, which is invasive and leads to poor patient compliance, exacerbating treatment delays. Morphological analysis of peripheral blood cells (PBC) is still primarily used to distinguish between benign and malignant hematologic disorders, but it remains a challenge to diagnose and type these diseases solely by direct observation of peripheral blood(PB) smears by human experts. In this study, we apply a segmentation-based enhanced residual network that uses progressive multi-granularity training with jigsaw patches. It is trained on a self-built annotated dataset of 21,208 images from 237 patients, including five types of benign white blood cells(WBCs) and eight types of leukemic cells. The network is not only able to discriminate between benign and malignant cells, but also to typify leukemia using a single peripheral blood cell. The network effectively differentiated acute promyelocytic leukemia (APL) from other types of acute myeloid leukemia (non-APL), achieving a precision rate of 89.34%, a recall rate of 97.37%, and an F1 score of 93.18% for APL. In contrast, for non-APL cases, the model achieved a precision rate of 92.86%, but a recall rate of 74.63% and an F1 score of 82.75%. In addition, the model discriminates acute lymphoblastic leukemia(ALL) with the Ph chromosome from those without. This approach could improve patient compliance and enable faster and more accurate typing of leukemias for early diagnosis and treatment to improve survival.

KEYWORDS
acute myeloid leukemia, deep learning, hematological malignancy, leukemia, peripheral blood cell

1 | INTRODUCTION

Leukemia is a common hematological malignancy with high morbidity and mortality worldwide.[1] As with other cancers, early detection and prompt intervention can reduce tumor burden and prolong patient life.[2] The disease shows a high degree of heterogeneity, requiring different treatments and showing different prognoses.[3–5] Therefore, accurate classification of leukemia, not only the mere diagnosis of the disease itself, is important for effective treatment. However, currently available diagnostic methods, such as morphologic analysis, immunophenotyping by flow cytometry (FCM) analysis, cytogenetics, and molecular genetics, are time-consuming, expensive, and complicated.[6–8] All of these tests rely on bone marrow (BM) aspiration, which is invasive and results in low patient compliance, exacerbating the problem of delayed diagnosis.[9] Moreover, the limited availability of trained professionals and the large and expensive equipment required further limit these methods. As a result, they are not viable options for primary hospitals that serve as the main screening centers for leukemia, or for resource-poor and impoverished countries and regions.[10] Therefore, it is imperative to develop simple and highly effective methods for screening, early diagnosis, and accurate classification of leukemia.

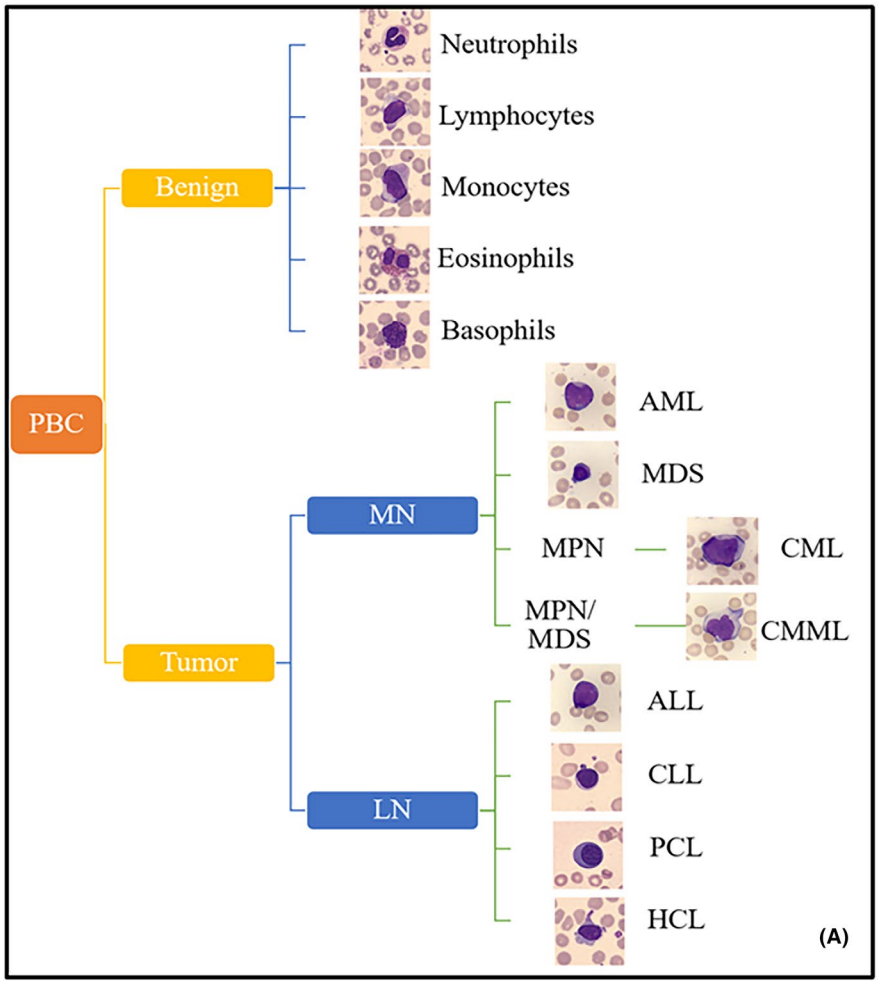
As is well known, the morphological analysis of peripheral blood cells(PBC) is an important cornerstone in the diagnosis of benign and malignant hematologic diseases, certainly including leukemia.[11] The manual inspection method is labor-intensive, susceptible to inaccuracies, and exhibits a high degree of subjectivity. Moreover, training inspection experts to achieve a high level of professional competence is a significant challenge.[12,13] Furthermore, due to physiological constraints, certain valuable information remains unidentified, limiting the ability of PBC morphology to provide comprehensive clinical diagnoses. Although certain cellular morphologic features have been identified to differentiate different types of acute myeloid leukemia (AML),[14] diagnosis by direct visualization of a PB smear by a hematopathologist alone remains challenging. In modern times, numerous computational methods have emerged that have shown remarkable capabilities in medical image interpretation tasks.[15–17] These capabilities can be applied to computer vision for blood cell morphology differentiation. Several studies have focused on the classification of physiological cell types in PB smears by deep learning (DL),[18] limiting their usability for the diagnosis of hematological malignancies. Subsequently, different researchers have used different approaches to accurately distinguish blasts (malignant cells) from healthy cells.[19] In recent years, several studies have focused on automating the cytomorphological classification of PB cells or BM cells, but without diagnosing diseases.[20,21] The majority of previous studies using DL approaches to blood cell classification have focused on relatively small numbers of disease classes.[22,23] Few attempts have been made to diagnose and typify leukemia by blasts that appear on PB smears. In terms of discriminative power and consistency in distinguishing between different entities, DL models have shown impressive and sometimes even superior capabilities compared to humans in certain specific tasks.[13,24] Inspired by imaging genomics, we seek to establish a correlation between the morphology of leukemia cells with specific genetic imprints. Different types of leukemia are characterized by chromosomal aberrations and alterations in transcription factors, epigenetic regulators, and signaling molecules that establish leukemia type-specific transcriptional networks.[25–28] Despite the complexity of these regulatory networks, they ultimately maintain a similar leukemic cell morphology.[29] In summary, our aim is to utilize PBC morphology as a means to efficiently screen neoplastic cells, diagnose and classify leukemia quickly and correctly.

In this study, we introduce a segmentation-based enhanced residual network that utilizes progressive multigranularity training of jigsaw patches. This network demonstrates the ability to differentiate between benign and malignant cells and type leukemia. Moreover, the network distinguishes acute promyelocytic leukemia (APL) from other types of acute myeloid leukemia (non-APL), as well as acute lymphoblastic leukemia (ALL) containing the Ph chromosome (Ph+ALL) from those without (Ph−ALL). These findings could improve patient compliance and enable faster typing of leukemias for early diagnosis and treatment.

2 | MATERIALS AND METHODS

2.1 | Sample selection and data preparation

The study was approved by the Ethics Committee of Shanxi Medical University, including the explicit waiver of informed consent from individual patients due to the de-identification of all samples in accordance with the Declaration of Helsinki. From 2020 to 2022, we enrolled 161 patients diagnosed with hematological malignancies and 118 controls from the Second Hospital of Shanxi Medical University and the Shanxi Provincial People's Hospital. The gold standard in this study is the diagnosis by hematopathologists according to the latest guidelines. All diagnoses were confirmed by independent hematopathologists based on clinical information, BM cell morphology, BM biopsy results, FCM, and genetic data. The control cohort included 68 individuals undergoing routine physical examinations and 50 patients with conditions unrelated to hematological malignancies. The leukemia cohort included several subtypes[30]: AML, myelodysplastic syndromes (MDS), chronic myelomonocytic leukemia (CMML), chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), plasma cell leukemia (PCL), and hairy cell leukemia (HCL) (Figure 1A). Slides were processed on an SP-10 automated slicer (Sysmex) and stained with May–Grünwald Giemsa. These slides were then imaged using the DI-60 automated digital cell image analyzer (Sysmex). To ensure the collection of diagnostically relevant cells, two to three slides were prepared per tumor sample, while a single slide was sufficient for each control sample. The imaging process focused on nonerythrocytic cells, including platelets, platelet aggregates, staining artifacts, and similar elements. Approximately 120 images (250×250 pixels) were captured per slide.



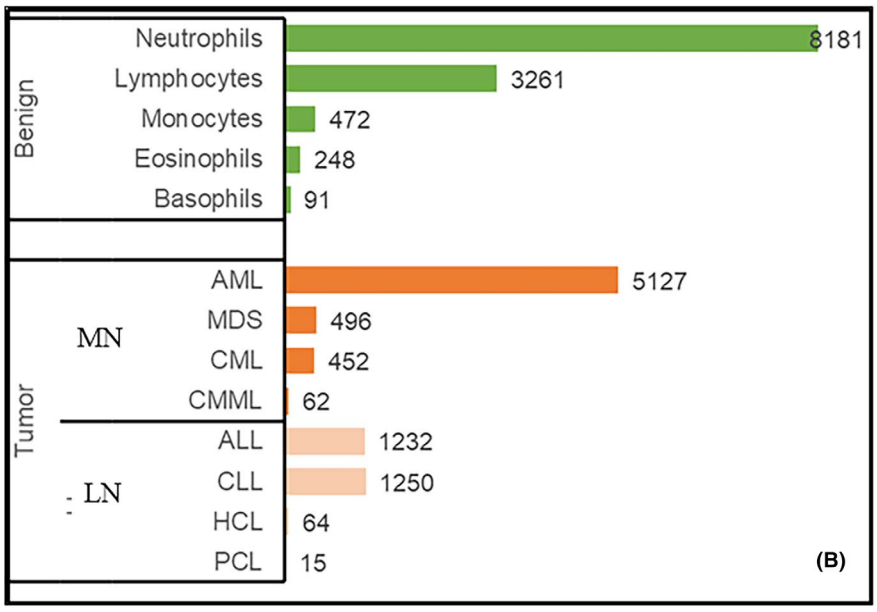


FIGURE 1 Overview of the (A) classification of blood cancers and (B) number of benign cells and blasts in the peripheral blood cell (PBC). Depending on which types of aberrant leukocytes can become cancerous, blood cancers can be categorized as myeloid or myelogenous neoplasms (MNs) and lymphoblastic or lymphocytic neoplasms (LNs). According to the 2016 WHO classification, MNs are divided into four main groups: acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), MDS/myeloproliferative neoplasms (MPN), and MPN. Chronic myelomonocytic leukemia (CMML) is a common typical subtype of MDS/MPN; chronic myeloid leukemia (CML) is a typical subtype of MPN. The classification of LNs is quite complex and diverse, not entirely limited to the division into acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), plasma cell leukemia (PCL), and hairy cell leukemia (HCL).

Given the diagnostic significance of blasts in hematological malignancies, only blasts from PB smears were selected for the malignant cohort. In addition, promyelocytes indicative of APL were included in the malignant dataset. In the control group, only five types of normal leukocytes were included; any suspected or ambiguous malignant cells were omitted from the control dataset. To increase data diversity, neutrophils with toxic granules and atypical lymphocytes were included. Images containing multiple cells, platelets and platelet aggregates, blue cells, senescent and degenerated cells, ruptured cells, staining artifacts, nucleated erythrocytes, and

other irrelevant cellular components were excluded. After clinical diagnosis and morphological evaluation, the cell images were consolidated into the final dataset. The refined dataset contained 21,208 individual cell images, including 8955 images of malignant cells from 119 patients and 12,253 images of five types of normal leukocytes from 118 controls. Detailed characteristics of the leukemia and control cohorts are shown in Table 1 and Figure 1B. Techniques such as image rotation, cropping, and color jittering are used for data augmentation when dealing with class imbalance due to natural morbidity.

2.2 | Deep learning model and training

The classification of leukemia, due to the inherently subtle intra-class object variation, is a fine-grained visual classification. Multiple categories make this task considerably much more challenging than traditional classification tasks. Based on the idea that the segmentation task can facilitate the classification task, we performed cell segmentation using the Segment Anything Model (SAM).[31] We used the ViT-L model and employed SamAutomaticMaskGenerator to automatically generate segmentation masks for objects in images. Subsequently, we manually inspected these masks and confirmed the selection of correctly segmented intact cells. Next, we used an improved ResNeXt framework developed by Du et al.[32] The refined residual network by progressive multigranularity (PMG) training of jigsaw patches emphasizes identifying the most discriminative granularities within local regions and leveraging information across diverse granularities to enhance classification accuracy.

Compared to the usual multiclass classification, there are evident differences in the classification of leukemia. As shown in Figure 1A, the diagnosis and typing of leukemia was multilevel in the taxonomy tree and its topological structure. The features of the root nodes are better distinguished than those of the distal nodes. It is more difficult to identify cells than to determine whether they are benign or malignant. Furthermore, when benign cells are predicted, they are no longer classified as leukemia types, while predicted malignant cells are no longer classified as normal cells. In order to diagnose and type leukemia, a multistage framework is proposed. Based on the hierarchy in diagnosis and leukemia type, the inference process includes five PMG models: Stage 1, Stage 2B (benign), Stage 2T (tumor), Stage 3M (myeloid), and Stage 3L (lymphoid). These models classify benign or malignant, leukemia type and subtype levels, respectively. In our training process, the output of the last stage PMG is used for the next stage PMG loss calculation and parameter update, which constrains and guides the classification of the next network (Figure 2).

TABLE 1 Statistics of the dataset.

	Images (patients)	Class	Patients	Images
Benign	12,253 (118)	Neutrophils	–	8181
		Lymphocytes	–	3261
		Monocytes	–	472
		Eosinophils	–	248
		Basophils	–	91
Tumor	8955 (99)	AML	50	5127
		MDS	22	496
		CML	5	452
		CMML	3	62
		ALL	18	1232
		CLL	17	1250
		PCL	3	64
		HCL	1	15

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; CMML, chronic myelomonocytic leukemia; HCL, hairy cell leukemia; MDS, myelodysplastic syndromes; PCL, plasma cell leukemia.

The network was trained on a server with two NVIDIA RTX A5000 graphics processing units and 64 GB of memory, and the PMG model took 48 h to run. In our comparative analysis, we utilized the ViT-B/16 and ResNeXt50 architectures, both of which have demonstrated exceptional performance in visual processing tasks. The models were initialized with the Xavier initialization method during the early training stages, facilitating expedited convergence. We used the Stochastic Gradient Descent (SGD) optimizer, with a learning rate set of 1e-4, a weight decay parameter of 5e-5, and a total of 150 training iterations.

We found that, in real life, cell images from new subjects are never available a priori for training; a realistic classifier would require training on data from some subjects and testing on data from completely unseen subjects. This limits the performance of the classifier on data from future subjects, as subject-specific features also contribute to class discrimination. Therefore, our data is divided into a training set and a test set at the subject level. In other words, all cell images belonging to a subject are split into the training set or the test set. This stratified train–test split was done in a random manner. For the training of individual networks reported in this article, we used 80% of the available images for each class, while 20% were used to evaluate the trained network at the subject level, not at the cell level.

3 | RESULTS

3.1 | Classification performance of the PMG

Compared to ResNeXt and the one-stage PMG, our trained multistage PMG has notably improved its top-1 accuracy to 99.53% (Table 2). For most of the morphological categories in our scheme, the classification precision, recall, and F1 score hover around 80% (Table 3). As expected from data-driven learning algorithms, classification performance often improves with an increase in the available training sample images. Specifically, for AML, we achieved a classification precision of 81.42% ± 5.12%, a recall of 96.94% ± 4.71%, and an F1 score of 88.5% ± 2.59%. However, for HCL, all performance metrics remained at approximately 50%. Given the low incidence

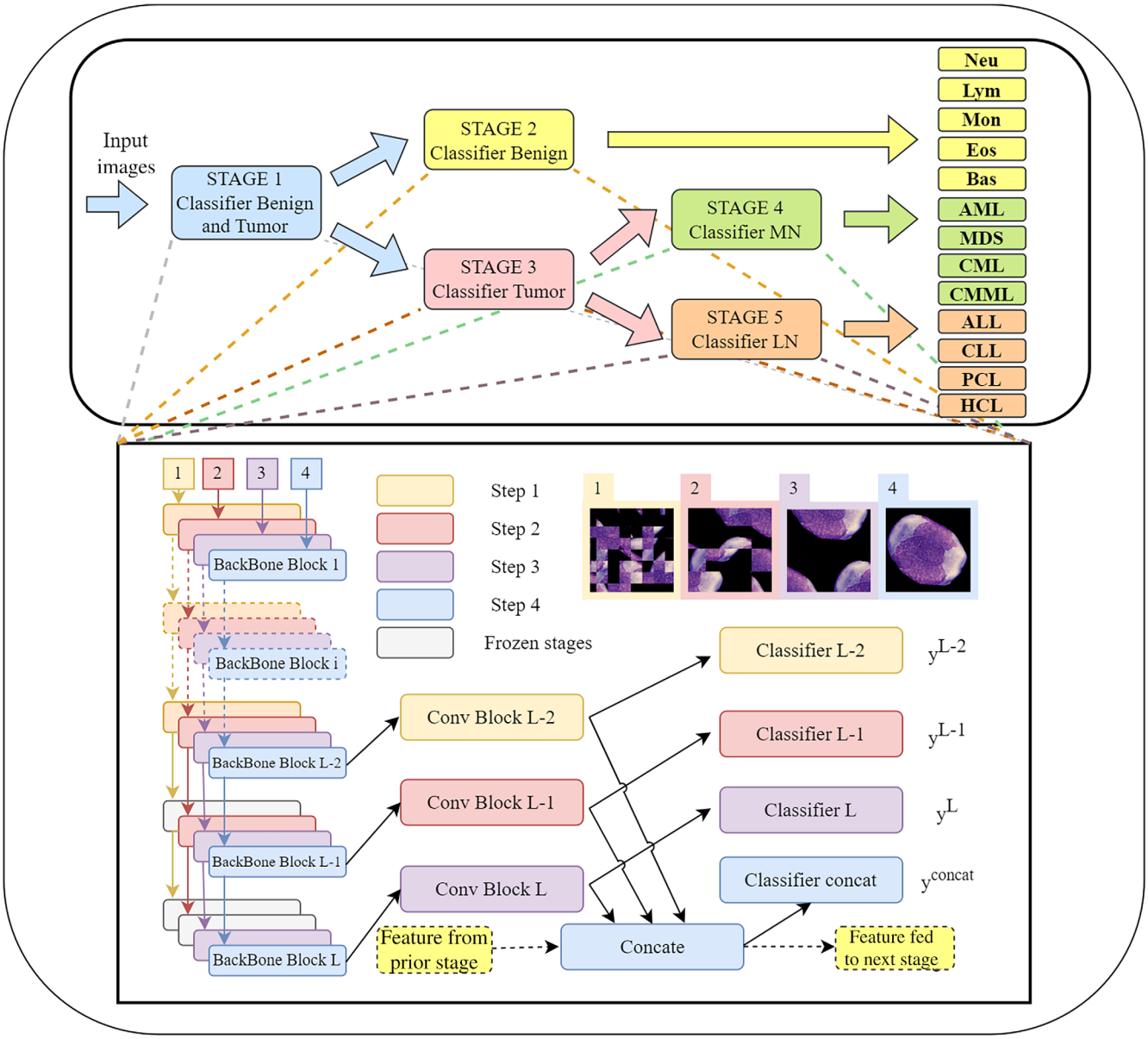


FIGURE 2 Multistage progressive multigranularity (PMG) training framework. Our model consists of five PMGs. The progressive training procedure consists of four steps at each iteration (here S = 3 for explanation). The conv block represents the combination of two convolutional layers with a max pooling layer, and the classifier represents two fully connected layers with a softmax layer at the end. At each iteration, the training data is augmented by the jigsaw generator and sequentially input into the network by S + 1 steps. ALL, acute lymphoblastic leukemia; AML; acute myeloid leukemia; Bas, basophil; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; CMML, chronic myelomonocytic leukemia; Eos, eosinophil; HCL, hairy cell leukemia; LN, lymphocytic neoplasm; Lym, lymphocyte; MDS, myelodysplastic syndromes; MN, myeloid or myelogenous neoplasm; Mon, monocyte; Neu, neutrophil; PCL, plasma cell leukemia.

rates of HCL and PCL, we experimented with data augmentation techniques, but unfortunately observed minimal improvements.

The model demonstrates a remarkable ability to recognize the distinct morphological features of the five benign cell types. However, disease classification is more challenging than cell classification. While the benign and malignant classifications achieved 96% precision, recall, and F1 score, these performance indicators decrease as the classifications become more specific. We face challenges due to unbalanced category data resulting from natural incidence rates. Despite attempts to balance the data by oversampling or undersampling and the use of data augmentation techniques, the improvements were minor.

Beyond the issue of data imbalance, the inherent difficulty in distinguishing between closely related myeloid neoplasms such as AML, MDS, CML, and CMML is consistent with previous reports.[21] The confusion matrix highlights the misclassifications among these categories (Figure 3). In particular, the CML classification struggles with a precision of 79.75% ± 0.97%, a recall of 25.93% ± 2.98%, and an F1 score of 39.13% ± 1.48%. We infer that this is in part due to the unique pathologic characteristics of CML. In CML, the presence of

TABLE 2 Comparison results of deep learning models.

	Top-1 accuracy	Mean precision	Mean recall	Mean F1 score
ViT	76.25 ± 5.03	64.97 ± 7.91	54.03 ± 7.61	53.69 ± 6.04
ResNeXt	72.45 ± 6.85	58.19 ± 5.09	58.3 ± 4.57	55.15 ± 4.06
ResNeXt + PMG	82.43 ± 5.46	67.5 ± 5.32	65.59 ± 6.66	65.91 ± 3.64
ResNeXt + PMG + multistage	99.53 ± 3.05	89.26 ± 2.31	89.89 ± 3.67	90.28 ± 1.23

Abbreviation: PMG, progressive multigranularity.

TABLE 3 Class-wise precision and recall of the multistage progressive multigranularity (PMG) classifier.

Class	Images (patients)	Precision	Recall	F1 score
Neutrophils	8181	99.69 ± 6.86	99.37 ± 5.87	99.53 ± 6.29
Lymphocytes	3261	97.17 ± 0.55	97.17 ± 0.42	97.17 ± 0.36
Monocytes	472	94.15 ± 2.55	89.58 ± 3.24	91.98 ± 6.86
Eosinophils	248	90.7 ± 1.30	95.12 ± 1.32	92.86 ± 6.86
Basophils	91	79.17 ± 4.81	100 ± 6.24	88.37 ± 7.69
AML	5127 (50)	81.42 ± 5.12	96.94 ± 4.71	88.5 ± 2.59
MDS	496 (22)	78.57 ± 5.27	67.35 ± 1.20	72.53 ± 2.37
CML	452 (5)	79.75 ± 0.97	25.93 ± 2.98	39.13 ± 1.48
CMML	62 (3)	42.42 ± 7.50	57.14 ± 3.03	48.7 ± 5.47
ALL	1232 (18)	93.9 ± 6.96	81.63 ± 6.13	87.33 ± 1.55
CLL	1250 (17)	82.31 ± 0.39	94.16 ± 0.95	87.84 ± 0.33
PCL	64 (3)	75 ± 8.09	11.54 ± 1.09	20 ± 2.74
HCL	15 (1)	42.86 ± 7.06	50 ± 5.43	46.15 ± 3.76

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; CMML, chronic myelomonocytic leukemia; HCL, hairy cell leukemia; MDS, myelodysplastic syndromes; PCL, plasma cell leukemia.

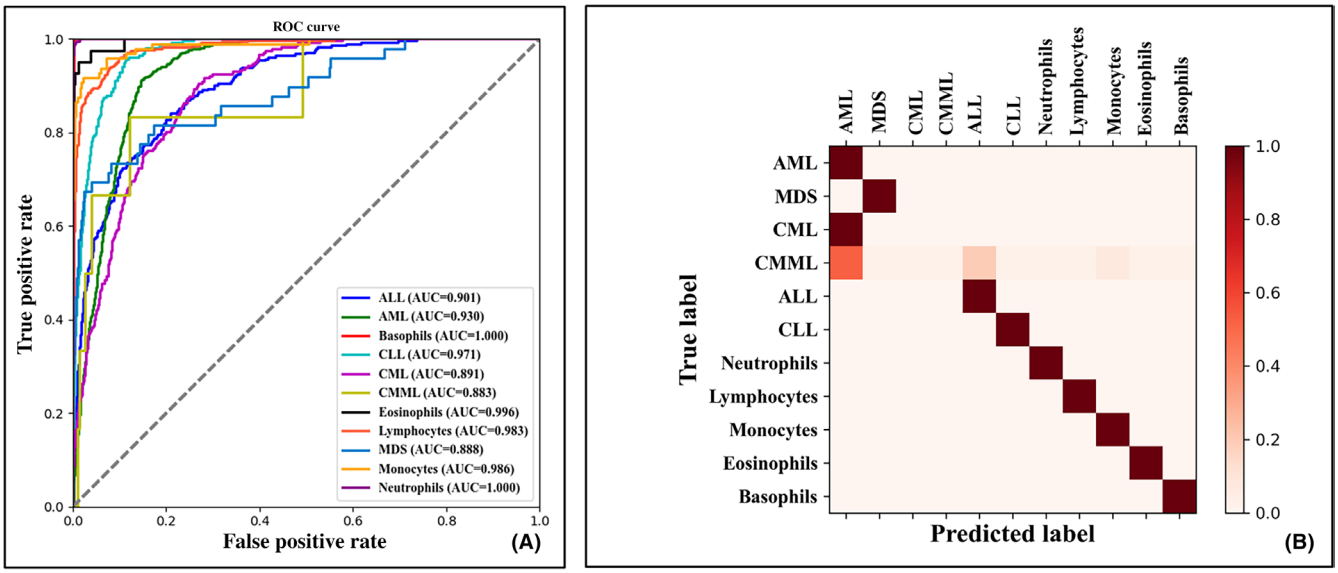


FIGURE 3 Accurate multistage progressive multigranularity (PMG) prediction for most classes. (A) Receiver operating characteristic (ROC) for multiclassification of leukemia on peripheral blood cell images. (B) Confusion matrix of predictions obtained by the PMG classifier on the test database annotated with gold standard labels provided by human experts. The number of single cell images included in each category is indicated in the logarithmic plot on the right. ALL, acute lymphoblastic leukemia; AML; acute myeloid leukemia; AUC, area under the ROC curve; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; CMML, chronic myelomonocytic leukemia; MDS, myelodysplastic syndromes.

TABLE 4 Results of each stage in the multistage progressive multigranularity (PMG) classifier.

	Top-1 Acc	Mean precision	Mean recall	Mean F1 score
Stage1	96.07 ± 0.43	96.07 ± 0.72	95.79 ± 0.41	95.92 ± 0.20
Stage 2B	97.72 ± 0.70	92.25 ± 1.18	96.25 ± 2.46	93.98 ± 0.93
Stage 2T	93.03 ± 0.33	96.67 ± 1.05	89.24 ± 1.77	92.26 ± 0.46
Stage 3M	80.58 ± 2.16	59.93 ± 3.67	47.55 ± 2.23	50.04 ± 1.12
Stage 3L	87.59 ± 1.05	88.11 ± 0.91	87.89 ± 1.63	87.59 ± 1.64

Note: Stage 1, benign or malignant classification; Stage 2B (benign), classification of benign cells; Stage 2T (tumor), classification of malignant cells; Stage 3M (myeloid), classification of myeloid leukemia subtypes; Stage 3L (lymphoid), classification of lymphoid leukemia types.

blasts is not essential for diagnosis. Instead, CML diagnosis relies primarily on an abnormally high WBC count and the presence of numerous basophils and eosinophils. This finding suggests a potential bias in our previous dataset and guides future optimizations. In the 2T stage, we attempted broader leukemia classifications. Distinguishing between acute and chronic forms, or myeloid and lymphocytic leukemias, showed limited effectiveness for the former (F1 score just over 50%) but strong performance for the latter over 90% accuracy (Table 4). This unexpected finding underscores a strong relationship between cell morphology and specific genetic markers, validating our classification approach.

Except for generally demonstrating the correlation between cell morphology and specific genetic imprints, we found two special cases to further establish this mapping. We trained the networks separately on two datasets: one to distinguish APL from non-APL, and another to identify Ph+ALL from Ph−ALL. As expected, the classification precision reached 89.34% and 92.86% (Tables 5 and 6, Figure 4A,B). Moreover, while pathologists may struggle to identify morphological differences associated with the genomic truth, we describe an algorithm using DL to learn the morphological features predictive of genotype in ALL (Figure 5A,B).

3.2 | Prediction performance and external validation

To test the generalizability of our model, it is necessary to evaluate the network's predictions on an external dataset not used during training. We collected 795 new cell images from three cases of AML (one case of AML did not show any fusion genes that are routinely detected in clinical tests), one case of CML, and one case of CLL. The precision rates were 92.44%, 83.75%, and 92.85% respectively.

We also evaluated our model on two annotated public datasets in The Cancer Imaging Archive (TCIA): AML-Cytomorphology_LMU by Matek et al.[33] and the CNMC 2019 dataset: ALL Challenge dataset of ISBI 2019 (C-NMC 2019) by Gupta et al.[34] We selected all cell images of myeloblast (MYO) from AML-Cytomorphology_LMU and 2397 ALL single cell images from the fold 0 ALL group of the CNMC training data. The precision was 96.96% and 95.45%, respectively. Table 7 reveals a good performance of the classifier on the external dataset, suggesting that the generalization is robust. It should be noted that the heterogeneity of image pixels and format, staining, or imaging system affects the diagnosis and type. There is also considerable variation in imaging and annotation strategies. On the classic ALL-IDB2,[35] the performance of ALL identification is 46.51%. Consider that the images of ALL-IDB2 are captured by a laboratory optical microscope combined with a Canon PowerShot G5 camera, and these images are segmented by an algorithm into a 257 × 257 pixel TIF format. Obviously, there are evident differences in staining techniques and imaging methods from the self-built dataset, and these differences are likely to cause domain shift problems, which in turn adversely affect the performance of our model.

3.3 | Visualization and explainability

To gain a better understanding of the classification decisions made by deep networks and to improve their explainability, we undertook an analysis on the PMG model using Grad-CAM.[36] Our objective was to identify what the model had learned and which regions of the input images were important for the classification decisions. To achieve this, we utilized the outputs of the DL model for the last three layers of the network and used these outputs on a per-cell basis within the validation cohort to construct a feature space to visualize these cells. We selected two typical parts of our models: APL versus non-APL and Ph+ALL versus Ph−ALL. Figures 4C and 5C show the masked image overlay of cell image and heatmap generated in Python; image regions containing relevant features are colored red. Acute promyelocytic leukemia focuses on the cytoplasm, which is densely packed or agglomerated with large granules, causing a blurring of the boundaries between the nucleus and cytoplasm. In some cells the cytoplasm is filled with fine dust-like granules. Non-APL focuses on the nucleus. Considering Ph+ALL as another illustrative case, we observed in the heatmaps (Figure 5C) that nuclear chromatin is coarser, more condensed and less homogeneous in lymphocytoblasts from individuals with Ph+ALL, reflecting the morphological features of aged nuclei. Conversely, nuclear chromatin distribution appeared more refined and homogeneous in Ph−ALL lymphocytes. By applying cell-level classifications associated with Ph+ baseline values, we demonstrate that DL model is able to differentiate Ph+ALL in the independent patient cohort, and we extracted learned information from the trained network to identify previously undescribed morphological features of Ph+ALL. The current model is based on retrospective data, which requires further investigation

TABLE 5 Class-wise precision and recall of acute promyelocytic leukemia (APL) versus non-APL classifier.

Class	Images (patients)	Precision	Recall	F1 score
APL	216 (7)	89.34 ± 0.32	97.37 ± 1.34	93.18 ± 1.07
Non-APL	4911 (43)	92.86 ± 0.98	74.63 ± 1.41	82.75 ± 0.78

Abbreviation: Non-APL, other types of acute myeloid leukemia.

TABLE 6 Class-wise precision and recall of acute lymphoblastic leukemia(ALL) containing Ph chromosome (Ph+ALL) versus ALL without Ph chromosome (Ph−ALL) classifier.

Class	Images (patients)	Precision	Recall	F1 score
Ph+ALL	247 (7)	92.86 ± 0.66	83.06 ± 0.42	87.68 ± 0.66
Ph−ALL	985 (9)	87.25 ± 1.23	94.78 ± 2.02	90.86 ± 1.73

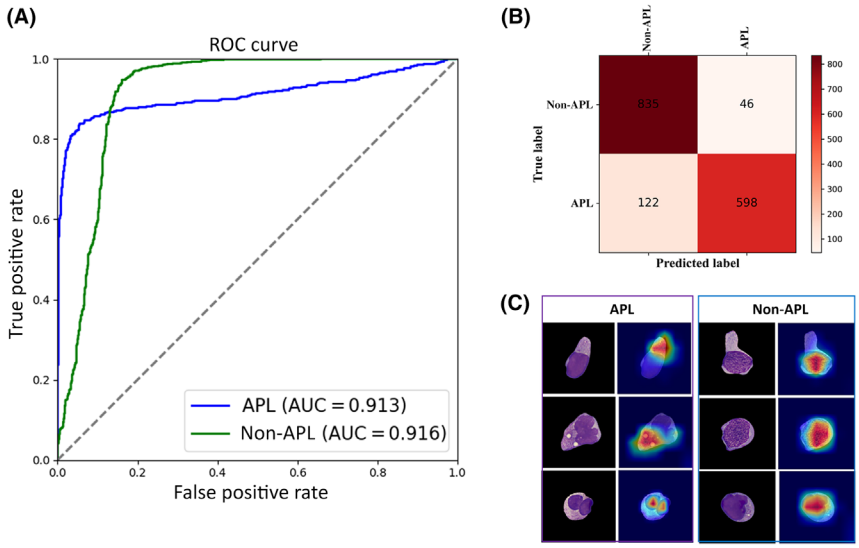


FIGURE 4 Performance measures of the acute promyelocytic leukemia (APL) and non-APL classification model. (A) Receiver operating characteristic (ROC) for binary classification of APL on PBC images. (B) Confusion matrix of the predictions obtained by the progressive multigranularity (PMG) classifier on the test database, the number of single cell images included in each category is indicated on the right. (C) Heatmap of the Grad-CAM algorithm generated in Python. The scale indicates the importance of certain image areas for correct class prediction. The more red the area, the more it contributes to the classification decision made by the network. AUC, area under the ROC curve.

in a prospective setting to confirm its diagnostic utility. In contrast, the ResNeXt model shows inconsistent foci, such as scattered dots in the cytoplasm or nucleus, which is inconsistent with human intuitive interpretation (Figures S1 and S2).

4 | DISCUSSION

We established a comprehensive and reliable database of PB cells containing approximately 20,000 images. Shortly after diagnosis, patients have to start chemotherapy, which inevitably leads to changes in cell morphology that affect the correct diagnosis. Patients also go through a period of myelosuppression approximately 1 week after drug treatment, during which there are not many WBCs to collect. Furthermore, the diagnostic workflow does not preserve PB smears. The period from initial diagnosis to being left untreated is too short to capture. In our investigation, we obtained PB smears at designated time intervals from individuals suspected of hematological malignancies, prior to any administration of chemotherapy. Subsequently, the subjects were subjected to monitoring, and only those with confirmed diagnoses were incorporated into the study. Additionally, data were annotated by two hematology experts with more than 10 years o experience, in accordance with the final clinical diagnosis. We also paid more attention to the details of data processing. The data were divided into training and testing sets at the subject level to prevent similarities between individual cells. In addition, any apparently normal cells that could contain unidentifiable abnormalities from cancer patients were eliminated to ensure an unbiased judgment. It is widely recognized that the quality of the data determines the maximum potential of the model, emphasizing the need for high quality data information.

What is different from previous studies is that we diagnosed diseases based on cell morphology, rather than just classifying the cells. Cell images contain much more information than we can discern

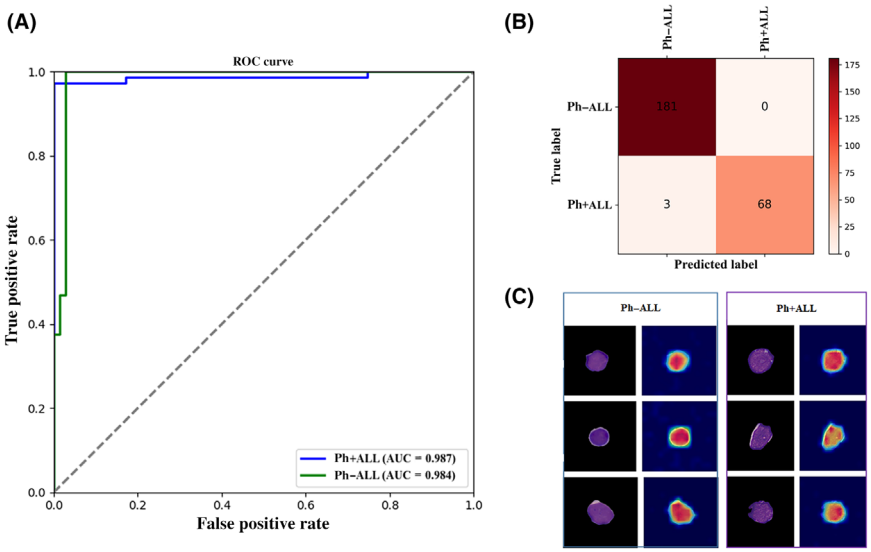


FIGURE 5 Performance measures of the acute lymphoblastic leukemia (ALL) containing the Ph chromosome (Ph+ALL) and ALL without Ph chromosome (Ph−ALL) classification model. (A) Receiver operating characteristic (ROC) for binary classification of Ph+ALL and Ph−ALL on peripheral blood cell images. (B) Confusion matrix of the predictions obtained by the progressive multigranularity (PMG) classifier on the test database; the number of single cell images included in each category is indicated on the right. (C) Heatmap of the Grad-CAM algorithm generated in Python. The heatmap shows a cell-specific class evaluation by the deep learning model generated in Python, showing areas that the deep learning model focused on. Red image areas are more important for class predictions. AUC, area under the ROC curve.

TABLE 7 Class-wise precision and recall of multistage progressive multigranularity (PMG) classifier on external dataset.

External dataset	Class	Images	Precision	Recall	F1 score
New self-built	AML	519	92.44 ± 5.27	94.89 ± 1.20	84.51 ± 3.36
	CML	94	83.75 ± 2.67	40.53 ± 1.68	50.13 ± 1.48
	CLL	182	92.85 ± 3.40	94.14 ± 1.03	89.7 ± 2.65
AML-Cytomorphology_LMU	AML	3268	96.96 ± 5.93	98.17 ± 4.23	97.56 ± 5.59
CNMC 2019	ALL	2397	95.45 ± 1.53	98.52 ± 1.35	90.17 ± 1.39
ALL-IDB2	ALL	130	46.51% ± 0.94%	52.17% ± 1.32%	48.98% ± 1.58%

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia.

with the naked eye. The heatmap visualization of the classification without segmentation reveals that numerous regions crucial for classification are located outside the cells. So, we adopted the idea of segmentation followed by classification, which brings the focus of discrimination back to the cell. This likely explains why the segmentation task facilitates the classification task.

The classification of leukemia cells is much more challenging than traditional classification tasks due to the heterogeneity of leukemia. In this study, we used PMG, which proposes a new, progressive training strategy for the ResNeXt skeleton, with the results demonstrating the effectiveness.[32] The leukemia cell classes are hierarchical, with the upper layer providing a coarse-grained classification and the lower layer offering a finer-grained classification. This hierarchical structure allows for the rapid identification of broad categories through coarse-grained classification, followed by gradual refinement in classification granularity, ultimately leading to precise classification and identification. We not only separated tumor cells from normal and benign reactive infected cells, but also identified leukemia types based on immature WBCs.

It is worth mentioning that our classification also accurately identified the AML cases lacking the leukemia fusion gene. Currently, a wide range of genetic markers are under investigation for their relevance to leukemia in the clinical setting, where their identification is essential for the diagnosis, classification, treatment, and prognostic assessment of the disease. Among these markers, fusion genes—common genetic abnormalities in leukemia—have been officially recognized in the new guidelines as a criterion for leukemia diagnosis.[30] In clinical practice, fusion genes serve as molecular markers in AML. Notably, when fusion genes are confirmed positive, the diagnostic criteria for AML with recurrent genetic abnormalities no longer require the presence of at least 20% blast cells. This adjustment not only simplifies the diagnostic approach but also improves diagnostic precision, particularly when blast cells are below 20%, or in differentiating AML from MDS characterized by elevated blast cells, as well as distinguishing trilineage dysplastic AML from MDS. In such scenarios, the detection of fusion genes is especially critical. Nonetheless, a negative result from this analysis does not eliminate the possibility

that some patients may possess other uncommon variant fusion genes. Our method identified a case of AML with all 13 fusion genes tested negative solely through cytomorphology. This highlighted the importance of our research in demonstrating the feasibility of morphology-based classification in the absence of typical molecular features. Therefore, we used the model as a biomarker for diagnosis and also successfully distinguished APL from non-APL, thereby validating the feasibility of genetic imprint identification in visual representations.[13] This further supports the claim. We also pioneered the prediction of Ph+ALL, a high-risk leukemia subtype with high relapse rates and poor prognosis. The observations indicate that Ph+ALL cells manifest rough, aggregated, heterogeneous, and senescent nuclei, which is attributed to the function of the Ph chromosome. The fusion gene *BCR-ABL*, generated by the Ph chromosome, suppresses the apoptotic pathway, resulting in impaired differentiation capacity, enhanced proliferation, and an elevated nuclear-to-cytoplasmic ratio. Similarly, the report shows specific cupped cells have been reported in AML with nucleophosmin 1 (NPM1) mutations.[29,37] These discoveries consistently underscore a profound correlation between these morphological traits and the underlying molecular genetic abnormalities, establishing them as reliable biomarkers for prediction and diagnosis.

There are several works that need to be improved in the future. Although the top 1 accuracy of the automated diagnostic system exceeded 95%, our system still serves as an aided tool. In conclusion, a correct diagnosis of hematological malignancies is strictly dependent on clinical information, BM examination, flow cytometric data, and genetic tests. Further efforts are required to collect a diverse range of tumor cells to enhance the identification of different types of hematological malignancies. Dataset imbalance seriously affects classifier performance. Most classification algorithms aim to maximize overall accuracy, which often results in a bias towards the majority class in imbalanced datasets. Although this may enhance overall accuracy, it leads to poor classification performance for the minority class. Despite our efforts to augment the minority class through image rotation and flipping, the marginal improvement in accuracy remains limited. This limited sample size restricts the classifier's ability to learn effective features, weakening its recognition of the minority class. During training, the decision boundary tends to favor the majority class, increasing the likelihood of misclassifying minority samples. While adjusting weights and loss functions could address the imbalance, we chose not to do so, as it would distort the true incidence rates and hinder accurate recognition of the majority class during testing. Nevertheless, our work remains valuable. We utilized blast analysis on PB samples to screen for a wide range of leukemias and hematological malignancies, offering a rapid, convenient, and efficient diagnostic approach. In certain instances, leukemia cells have not broken through the BM to be released into the PB, and the information from PB cells is limited. In our future work, we plan to integrate morphological and genetic information from BM to classify disease subtypes and risk levels for stratified and targeted therapy.

AUTHOR CONTRIBUTIONS

Geng Yan: Conceptualization; project administration; supervision; writing – original draft; writing – review and editing. **Gao Mingyang:** Methodology; software; validation; visualization; writing – original draft; writing – review and editing. **Shi Wei:** Data curation; resources; writing – original draft; writing – review and editing. **Liang Hongping:** Data curation; formal analysis; resources; validation; writing – original draft; writing – review and editing. **Qin Liyuan:** Data curation; formal analysis; resources; validation; writing – original draft; writing – review and editing. **Liu Ailan:** Data curation; resources; writing – original draft; writing – review and editing. **Kong Xiaomei:** Formal analysis; investigation; validation; writing – original draft; writing – review and editing. **Zhao Huilan:** Formal analysis; investigation; validation; writing – original draft; writing – review and editing. **Zhao Juanjuan:** Methodology; software; validation; visualization; writing – original draft; writing – review and editing. **Qiang Yan:** Conceptualization; funding acquisition; project administration; supervision; writing – original draft; writing – review and editing.

FUNDING INFORMATION

This work was supported by the National Natural Science Foundation of China (No. 62376183) and the Funds for International Cooperation of the National Natural Science Foundation of China (No. U21A20469).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.Not a member of the all authors and corresponding author is a JAC member.

ACKNOWLEDGEMENTS

Thank Professor Ma Ruiqing and Dr Li Aoyu for revising the manuscript of the article.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. Data sharing requests should be sent to Qiang Yan (qiangyan@tyut.edu.cn).

ETHICS STATEMENT

Approval of the research protocol by an institutional review board: The study received approval from the Ethics Committee of Shanxi Medical University.

Informed consent: The need for informed consent from individual patients was waived due to the de-identification of all samples in accordance with the Declaration of Helsinki.

The study is registry with the Ethics Committee of Shanxi Medical University and the registration no. of the study: No. 2022321.

Animal studies: N/A.

ORCID

Geng Yan https://orcid.org/0009-0006-5626-2925

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249.
2. Cabannes-Hamy A, Brissot E, Leguay T. High tumor burden before blinatumomab has a negative impact on the outcome of adult patients with B-cell precursor acute lymphoblastic leukemia. A real-world study by the GRAALL. *Haematologica*. 2022;107(9):2072-2080.
3. Brown PA, Shah B, Advani A, et al. Version 2.2021, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw*. 2021;19(9):1079-1109.
4. Kayser S, Schlenk RF, Platzbecker U. Management of patients with acute promyelocytic leukemia. *Leukemia*. 2018;32(6):1277-1294.
5. DiNardo CD, Wei AH. How I treat acute myeloid leukemia in the era of new drugs. *Blood*. 2020;135(2):85-96.
6. Short NJ, Rytting ME, Cortes JE. Acute myeloid leukaemia. *Lancet*. 2018;392(10147):593-606.
7. Leonard JP, Martin P, Roboz GJ. Practical implications of the 2016 revision of the World Health Organization classification of lymphoid and myeloid neoplasms and acute leukemia. *J Clin Oncol*. 2017;35(23):2708-2715.
8. D€ohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447.
9. Briggs C, Longair I, Slavik M, et al. Can automated blood film analysis replace themanual differential? An evaluation of the CellaVision DM96 automated image analysis system. *Int J Lab Hematol*. 2009;31(1):48-60.
10. Eckardt JN, Schmittmann T, Riechert S, et al. Deep learning identifies acute Promyelocytic leukemia in bone marrow smears. *BMC Cancer*. 2022;22(1):201.
11. Theml H, Diem H, Haferlach T. *Hematology: Practical Microscopic and Clinical Diagnosis*. 2nd revised ed. Thieme; 2004.
12. Tomczak A, Ilic S, Marquardt G, et al. Multi-task multi-domain learning for digital staining and classification of leukocytes. *IEEE Trans Med Imaging*. 2021;40(10):2897-2910.
13. Sidhom JW, Siddarthan IJ, Lai BS, et al. Deep learning for diagnosis of acute promyelocytic leukemia via recognition of genomically imprinted morphologic features. *NPJ Precis Oncol*. 2021;5(1):38.
14. Gao Z, Mao A, Wu K, et al. Childhood leukemia classification via information bottleneck enhanced hierarchical multi-instance learning. *IEEE Trans Med Imaging*. 2023;42(8):2348-2359.
15. Morris A. Technology watch: AI can diagnose diabetic retinopathy. *Nat Rev Endocrinol*. 2018;14(2):65.
16. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks [published correction appears in nature. 2017;546(7660):686]. *Nature*. 2017;542(7639):115-118.
17. McKinney SM, Sieniek M, Godbole V, et al. Addendum: international evaluation of an AI system for breast cancer screening. *Nature*. 2020;586(7829):E19.
18. Shahin AI, Guo Y, Amin KM, Sharawi AA. White blood cells identification system based on convolutional deep neural learning networks. *Comput Methods Prog Biomed*. 2019;168:69-80.
19. Qin F, Gao N, Peng Y, Wu Z, Shen S, Grudtsin A. Fine-grained leukocyte classification with deep residual learning for microscopic images. *Comput Methods Prog Biomed*. 2018;162:243-252.
20. Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acutemyeloid leukaemia with convolutional neural networks. *NatMach Intell*. 2019;1(11):538-544.
21. 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(20):1917-1927.
22. Kimura K, Tabe Y, Ai T, et al. A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA. *Sci Rep*. 2019;9(1):13385.
23. Sahasrabudhe M, Sujobert P, Zacharaki EI, et al. Deep multi-instance learning using multi-modal data for diagnosis of lymphocytosis. *IEEE J Biomed Health Inform*. 2021;25(6):2125-2136.
24. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1(6):e271-e297.
25. Assi SA, Imperato MR, Coleman DJL, et al. Subtype-specific regulatory network rewiring in acute myeloid leukemia. *Nat Genet*. 2019;51(1):151-162.
26. Churpek JE, Bresnick EH. Transcription factor mutations as a cause of familial myeloid neoplasms. *J Clin Invest*. 2019;129(2):476-488.
27. Robbe P, Ridout KE, Vavoulis DV, et al. Whole-genome sequencing of chronic lymphocytic leukemia identifies subgroups with distinct biological and clinical features. *Nat Genet*. 2022;54(11):1675-1689.
28. Adamo A, Chin P, Keane P, et al. Identification and interrogation of the gene regulatory network of CEBPA-double mutant acute myeloid leukemia. *Leukemia*. 2022;37:102-112.
29. Eckardt JN, Middeke JM, Riechert S, et al. Deep learning detects acute myeloid leukemia and predicts NPM1 mutation status from bone marrow smears. *Leukemia*. 2022;36(1):111-118.
30. Baram DV, Asaulenko ZP, Spiridonov IN, Krivolapov YA. WHO classification of tumors of hematopoietic and lymphoid tissues (5th edition). 2022.
31. Kirillov A, Mintun E, Ravil N, et al. Segment anything. In Meta AI Research, FAIR 2023.
32. Du R, Chang D, Bhunia AK, et al. Fine-grained visual classication via progressive multi-granularity training of jigsaw patches. *European Conference on Computer Vision*. Springer International Publishing; 2020.
33. Matek C, Schwarz S, Marr C, Spiekermann K. A single-cell morphological dataset of leukocytes from aml patients and non-malignant controls. 2019.
34. Gupta R, Gehlot S, Gupta A. C-NMC: B-lineage acute lymphoblastic leukaemia: a blood cancer dataset. *Med Eng Phys*. 2022;103:103793.
35. Labati RD, Piuri V, Scotti F. All-IDB: the acute lymphoblastic leukemia image database for image processing. In the 18th IEEE International Conference on Image Processing 2011.
36. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017. In IEEE International Conference on Computer Vision (ICCV); pp. 616–626.
37. Claire B, Thomas B. t (6;9) (p23;q34.1) acute myeloid leukaemia with cup-like blasts. *Br J Haematol*. 2018;182:9.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yan G, Mingyang G, Wei S, et al. Diagnosis and typing of leukemia using a single peripheral blood cell through deep learning. *Cancer Sci*. 2025;116:533-543. doi:10.1111/cas.16374