Lab Invest 103 (2023) 100055

LABORATORY INVESTIGATION

journal homepage: https://laboratoryinvestigation.org/

Research Article

An AI-Aided Diagnostic Framework for Hematologic Neoplasms Based on Morphologic Features and Medical Expertise

Nan Li[a], Liquan Fan[a], Hang Xu[b,c], Xiwen Zhang[b,c], Zanzhou Bai[b,c], Miaohui Li[a], Shumin Xiong[a], Lu Jiang[a], Jie Yang[b,c], Saijuan Chen[a,*], Yu Qiao[b,c,*], Bing Chen[a,*]

[a] *Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China;* [b] *Institute of Image Processing & Pattern Recognition, Department of Automation, Shanghai Jiao Tong University, Shanghai, China;* [c] *Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China*

ARTICLE INFO

Article history:
Received 29 November 2022
Revised 21 December 2022
Accepted 27 December 2022

Keywords:
artificial intelligence
bone marrow morphology
diagnosis
hematology
multimodal features

ABSTRACT

A morphologic examination is essential for the diagnosis of hematological diseases. However, its conventional manual operation is time-consuming and laborious. Herein, we attempt to establish an artificial intelligence (AI)-aided diagnostic framework integrating medical expertise. This framework acts as a virtual hematological morphologist (VHM) for diagnosing hematological neoplasms. Two datasets were established as follows: An image dataset was used to train the Faster Region-based Convolutional Neural Network to develop an image-based morphologic feature extraction model. A case dataset containing retrospective morphologic diagnostic data was used to train a support vector machine algorithm to develop a feature-based case identification model based on diagnostic criteria. Integrating these 2 models established a whole-process AI-aided diagnostic framework, namely, VHM, and a 2-stage strategy was applied to practice case diagnosis. The recall and precision of VHM in bone marrow cell classification were 94.65% and 93.95%, respectively. The balanced accuracy, sensitivity, and specificity of VHM were 97.16%, 99.09%, and 92%, respectively, in the differential diagnosis of normal and abnormal cases, and 99.23%, 97.96%, and 100%, respectively, in the precise diagnosis of chronic myelogenous leukemia in chronic phase. This work represents the first attempt, to our knowledge, to extract multimodal morphologic features and to integrate a feature-based case diagnosis model for designing a comprehensive AI-aided morphologic diagnostic framework. The performance of our knowledge-based framework was superior to that of the widely used end-to-end AI-based diagnostic framework in terms of testing accuracy (96.88% vs 68.75%) or generalization ability (97.11% vs 68.75%) in differentiating normal and abnormal cases. The remarkable advantage of VHM is that it follows the logic of clinical diagnostic procedures, making it a reliable and interpretable hematological diagnostic tool.

Introduction

The morphologic examination of bone marrow cells (BMCs) is essential for diagnosing hematological diseases.[1,2] Traditional manual microscopy examination is still mainly used in clinical practice. Morphologists make a diagnosis by classifying and counting nucleated cells using a microscope. Such an approach is

These authors contributed equally: Nan Li, Liquan Fan, Hang Xu, and Xiwen Zhang.

* Corresponding author.
E-mail addresses: chenbing_rjh@163.com (B. Chen), qiaoyu@sjtu.edu.cn (Y. Qiao), sjchen@stn.sh.cn (S. Chen).

0023-6837/$ - see front matter © 2022 United States & Canadian Academy of Pathology. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.labinv.2022.100055

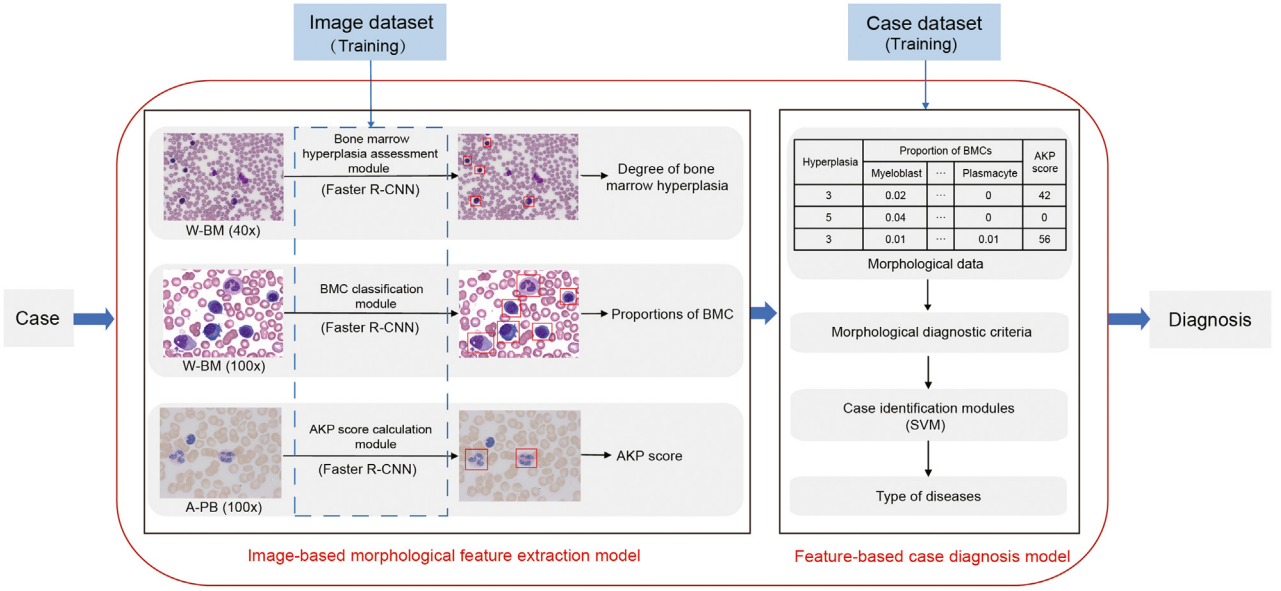


Figure 1.
Virtual hematological morphologist. The whole-process artificial intelligence–aided diagnostic framework containing the following 2main components: (1) an image-based morphologic feature extraction model and (2) a feature-based case diagnosis model. A-PB, AKP-staining peripheral blood; AKP, alkaline phosphatase; BMC, bone marrow cell; Faster-RCNN, Faster Region-based Convolutional Neural Network; SVM, support vector machine; W-BM, Wright staining bone marrow.

time-consuming and laborious. Moreover, the diagnosis depends heavily on the morphologist's experience and is usually subjective and lacks standardized quality control and evaluation. The accurate classification of each nucleated cell may be challenging even for an experienced morphologist. Consequently, an accurate and automated morphologic diagnostic method supported by artificial intelligence (AI) technology is required for clinical practice.

In the past few years, AI technology has developed rapidly in the medical field.[3] Numerous explorations of deep learning algorithms focusing on image recognition, such as the AI-assisted radiologic and pathologic diagnosis of cancers,[4-6] as well as the recognition of peripheral blood smear images in some hematological studies,[7,8] have achieved great success.

However, most hematological diseases are manually diagnosed by bone marrow morphologic features, which mainly rely on the proportion of nucleated cells in bone marrow smears and the positive degree of cytochemical staining. Several studies have shown that the automatic classification of nucleated cells in bone marrow can be realized by using deep learning algorithms,[9-15] even if this approach does not work as well in bone marrow as in peripheral blood because of the higher diversity and density of cells in the former than in the latter.[16-20] Although several studies considered the threshold of BMC proportion for diagnosis,[21-23] only a few have considered the degree of bone marrow hyperplasia or cytochemical staining scores, which are helpful for accurately identifying disease subtypes.

In this study, we aimed to take full advantage of morphologic diagnosis characteristics and morphologists' diagnostic experience to develop a comprehensive and efficient AI-aided hematological diagnostic framework. This knowledge-based framework was designed to use explicit medical expertise by integrating the procedures of multimodal image recognition and case identification to distinguish normal and abnormal cases and precisely diagnose specific diseases in a reliable and interpretable way. The comparison between this framework and a widely used AI-based end-to-end diagnosis framework was also investigated to evaluate its performance.

Materials and Methods

Study Design

Virtual hematological morphologist (VHM), a whole-process AI-aided diagnostic framework that integrated image recognition and case identification, was designed in this study (Fig. 1). The framework had the following 2 main components: (1) the image-based morphologic feature extraction model and (2) feature-based case diagnosis model.

Image recognition was implemented by the image-based morphologic feature extraction model. Multimodal morphologic features were extracted from microscopy images using the following 3 modules: (1) the bone marrow hyperplasia assessment, (2) the BMC classification, and (3) the alkaline phosphatase (AKP) score calculation modules. These modules were developed based on a state-of-the-art deep neural network, Faster Region-based Convolutional Neural Network (Faster R-CNN),[24] which was trained with microscopy images from the image dataset.

Case identification was accomplished by applying the feature-based case diagnosis model, which consisted of 2 case identification modules, using a support vector machine (SVM) algorithm.[25] The case identification modules were trained on the retrospective morphologic data of normal and hematological cases from the case dataset in reference to morphologic diagnostic criteria.

The microscopy images of each case were collected and then analyzed by the image-based morphologic feature extraction model to obtain multimodal morphologic features. With these features as the inputs, the final diagnostic results were obtained through the feature-based case diagnosis model.

To evaluate the performance of our proposed knowledge-based framework in hematological diagnosis, an end-to-end diagnosis framework based on ResNet-101 was established as the baseline for distinguishing normal and abnormal hematological cases.[26] It was trained to automatically learn undefined image features in microscopy images to diagnose diseases directly without considering any explicit medical expertise except the implicit knowledge represented by the labels of the training samples.

Image Collection

Bone marrow and peripheral blood smears were prepared in a wedge-spread manner. Wright and AKP stainings were performed using reagents from Shanghai SunBio Biotechnology Co per the manufacturer's instructions.

In this study, field selection for image collection was conducted manually. Samples with diluted or clotted marrow were excluded. The appropriate regions located on the body-tail junction part of the smear wherein nucleated cells were evenly distributed were selected.

Three types of images were collected, including Wright staining bone marrow (W-BM) images (magnifications, ×40 and ×100) and AKP-staining peripheral blood (A-PB) images (×100). Digital microscopic images of W-BM and A-PB slides were manually photographed using a Zeiss Axio Scope A1 microscope with an AxioCamHRc camera and the AxioVision software version 4.8.2.2. Both the microscope and camera were manufactured by Carl Zeiss MicroImaging GmbH. After adjusting the white balance in the blank area, the slide was focused under the field of view, and the shooting parameters were set and fixed (with an exposure time of 176 milliseconds, intensity of 90%, and color temperature of −0.3). After image acquisition was completed, postprocessing parameters were adjusted to a brightness of −0.45, contrast of 1.32, and gray scale of 0.7 to obtain the final images.

W-BM images (×40) were collected with a high-power lens (×40) for the evaluation of marrow hyperplasia, which was determined as the number of nucleated cells in a high-power field (HPF) in cytomorphology. For the Zeiss Axio Scope A1 microscope used in this study, the area of a real HPF was 0.2595 mm^2. The real size of a single digital image collected with the AxioCamHRc camera was 0.03412 mm^2, which was approximately one-seventh of the area of HPF. Therefore, the number of nucleated cells in 7 adjacent corresponding images was used to evaluate marrow hyperplasia.

W-BM (×100) and A-PB images (×100) were consecutively collected with an oil immersion lens (×100) by manually moving the objective table in a "N" shape for the evaluation of the proportions of nucleated cells and AKP score.

Datasets

Two datasets were established. The microscopic images collected from W-BM and A-PB smears were used to build an image dataset consisting of 3 subsets. The W-BM (×100) subset contained the images of 18 types of BMCs from 6 cell lineages (myeloid, erythroid, lymphoid, monocytic, plasmacytic, and megakaryocytic) with normal morphology. These BMCs included myeloblasts, promyelocytes, neutrophilic myelocytes, neutrophilic metamyelocytes, neutrophilic stab granulocytes, neutrophilic segmented granulocytes, eosinophils, basophils, normoblasts, basophilic normoblasts, polychromatic normoblasts, orthochromatic normoblasts, lymphocytes, monocytes, plasmacytes, granular megakaryocytes, platelet-producing megakaryocytes, and naked nuclei. The W-BM (×40) subset contained images reflecting marrow hyperplasia. The A-PB (×100) subset contained images of mature neutrophils with different AKP-staining grades ranging from "−" to "++++" (Supplementary Method 1).

A case dataset was built using patients' retrospective morphologic diagnostic data between January 2010 and October 2020 from Ruijin Hospital, affiliated with the Shanghai Jiao Tong University School of Medicine. In addition to normal cases, the abnormal cases represented 8 subtypes of hematological diseases, including acute promyelocytic leukemia with *PML::RARA* fusion, acute myeloid leukemia (AML) defined by differentiation (including AML without maturation, AML with maturation, acute myelomonocytic leukemia, and acute monocytic leukemia),[1] acute lymphoblastic leukemia, chronic myelogenous leukemia in chronic phase (CML-CP), chronic lymphocytic leukemia, *BCR::ABL1*-negative myeloproliferative neoplasms,[1] multiple myeloma, and cases with hypoplastic myelogram (mainly aplastic anemia). For each case, the collected diagnostic parameters included the proportions of 15 types of nucleated cells (myeloblasts, promyelocytes, neutrophilic myelocytes, neutrophilic metamyelocytes, neutrophilic stab granulocytes, neutrophilic segmented granulocytes, eosinophils, basophils, normoblasts, basophilic normoblasts, polychromatic normoblasts, orthochromatic normoblasts, lymphocytes, monocytes, and plasmacytes), marrow hyperplasia, and AKP score. Hyperplasia and nucleated cell proportions had no missing values in case dataset. A missing AKP score was denoted as "−1" in evaluating the case identification model and VHM framework.

Cell Annotation

In the image dataset, nucleated cells in W-BM (×40) images, 18 types of BMCs in W-BM (×100) images, and mature neutrophils with different staining grades in A-PB (×100) images were manually annotated using customized labeling software (Supplementary Fig. S1). Four experienced morphologists (L.F., S.X., and B.C. were included) from Ruijin Hospital reviewed and corrected the cell annotations.

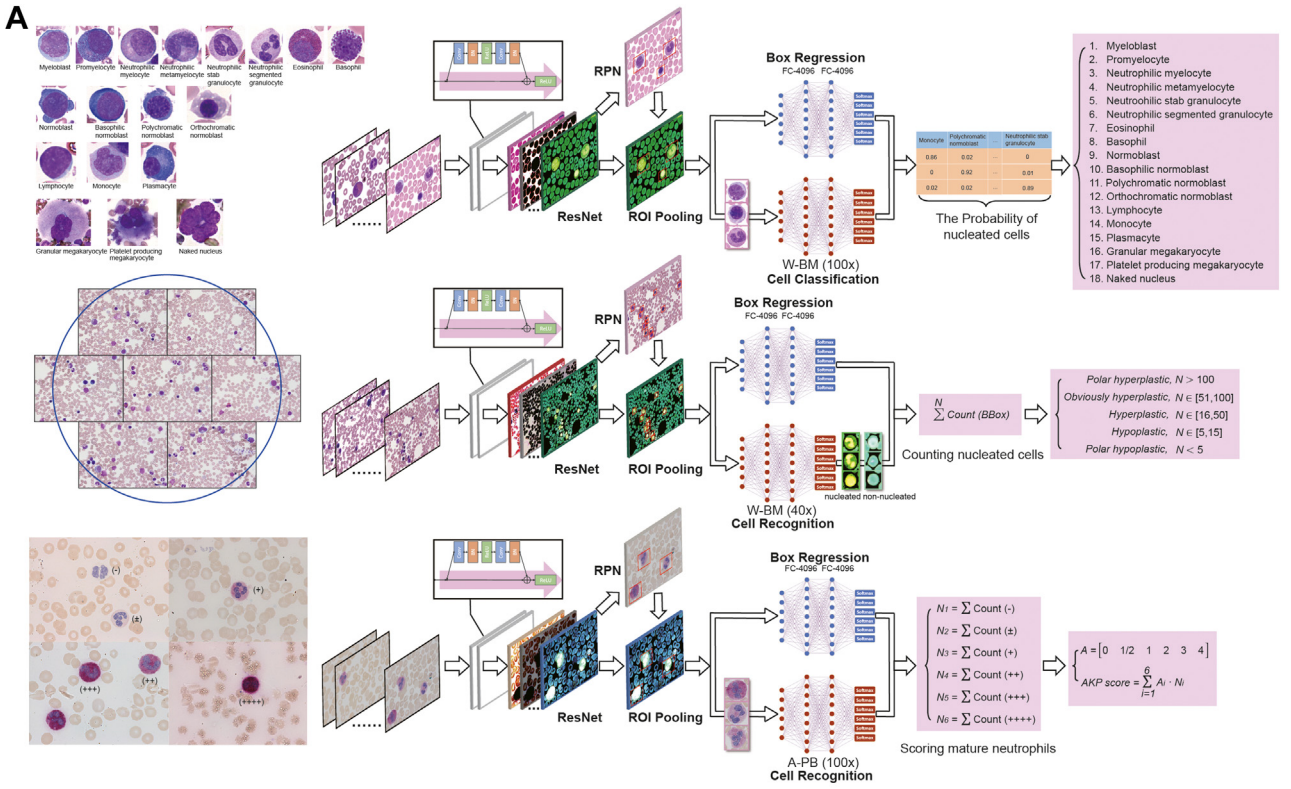
Artificial Intelligence Algorithms

Image Recognition

Faster R-CNN algorithm (Supplementary Fig. S2) was introduced into the 3 modules to perform bone marrow hyperplasia assessment, BMC classification, and AKP score calculation (Fig. 2A).[24]

Faster R-CNN is a deep neural network. It aims to detect the accurate locations of various BMCs or peripheral blood cells and classify these cells. In Faster R-CNN, a convolutional neural network named ResNet-101 was first used to extract the feature maps of the input image. The region proposal network module generated several bounding boxes for the locations of suspected cells. The regions-of-interest pooling module mapped the bounding boxes to the feature map to obtain the local features. Fully connected layers were used to assemble these local features into global feature vectors through the weight matrix. Finally, the feature vectors were used to implement the regression of the bounding boxes and the classification of the cells.[24]

For VHM application, the input of the BMC classification, bone marrow hyperplasia assessment, and AKP score calculation modules were W-BM (×100), W-BM (×40), and A-PB images (×100), respectively. Cells in these images were detected, classified, and counted to extract morphologic features, including BMC proportions, bone marrow hyperplasia degree, and AKP score.



		Case 1	…	Case N
Hyperplasia		3	…	5
BMC proportion	Myeloblast	0	…	0.63
	Promyelocyte	0.02	…	0
	…	…	…	…
	…	…	…	…
	Normoblast	0	…	0
	…	…	…	…
	…	…	…	…
	Lymphocyte	0.18	…	0.33
	Monocyte	0.04	…	0.03
	Plasmacyte	0	…	0
AKP score		52	…	92.5

17-dimentional diagnostic features

		Case 1	…	Case N
Hyperplasia		3	…	5
BMC proportion	Myeloblast	0	…	0.63
	Neutrophils in intermediate stage	…	…	…
	Eosinophils	0.03	…	0
	Basophils	0	…	0
	G/E	5	…	199
AKP score		52	…	92.5

7-dimentional diagnostic features

Figure 2.

Application of artificial intelligence algorithms in virtual hematological morphologist. (A) Image recognition. Top left: Eighteen types of nucleated cells in Wright staining bone marrow (W-BM) (magnification, ×100) images were used to train the bone marrow cell (BMC) classification module. Top right: Architecture of the BMC classification module. Nucleated cells were detected and classified by the Faster Region-based Convolutional Neural Network (Faster R-CNN). Middle left: Seven adjacent W-BM (×40) images representing a high-power field. These images were used to train the bone marrow hyperplasia assessment module. Middle right: Architecture of the bone marrow hyperplasia assessment module. Nucleated cells recognized by the Faster R-CNN were counted to assess the degree of bone marrow hyperplasia. Bottom left: Mature neutrophils with different alkaline phosphatase (AKP) staining degrees in AKP-staining peripheral blood (A-PB) (×100) images were used to train the AKP score calculation module. The symbols "-" to "++++" represent the AKP staining degrees from negative to strong positive. Bottom right: Architecture of the AKP score calculation module. The AKP score was calculated by counting and scoring the mature neutrophils recognized by the Faster R-CNN. (B) Case identification. Top: Seventeen-dimensional diagnostic features extracted from morphologic data were used as the input of the support vector machine (SVM)-based abnormal case identification module to distinguish normal and abnormal cases. Bottom: Seven-dimensional diagnostic features extracted from the morphologic data were used as the inputs of the SVM-based chronic myelogenous leukemia in chronic phase (CML-CP) case identification module to identify CML-CP cases precisely. G/E, the ratio of granulocytes to erythroblasts; ROI, region of interest; RPN, region proposal network.

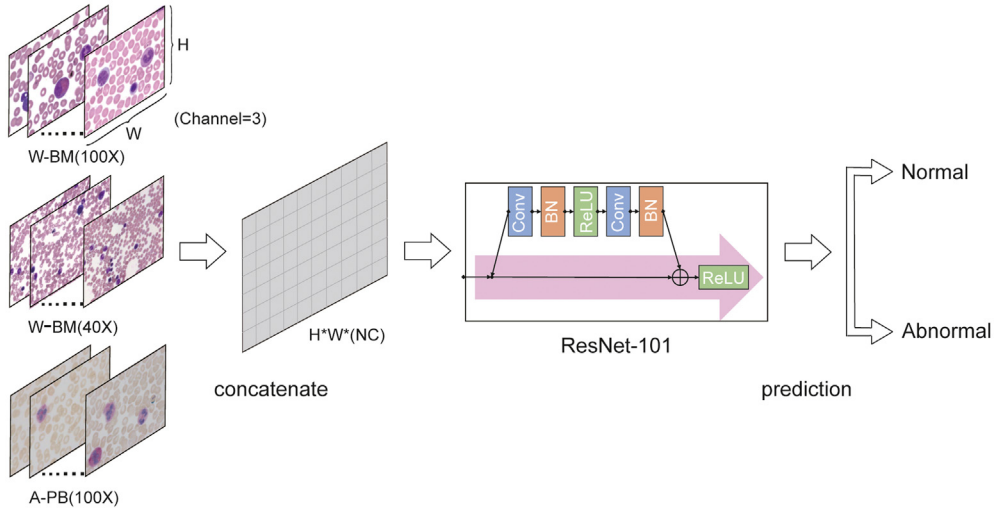


Figure 3.
Application of the artificial intelligence algorithm in the end-to-end diagnosis framework. A group of images from each case, including Wright staining bone marrow (W-BM) (magnification, ×100), W-BM (×40), and alkaline phosphatase staining peripheral blood (A-PB) (×100) images, were concatenated and sent into the classical ResNet-101 deep neural network for feature extraction and case classification. BN, batch normalization; Conv, convolution; H, height; NC, number of channels; ReLU, rectified linear unit; W, width.

Case Identification

SVM algorithm (Supplementary Fig. S3) was introduced into the feature-based case diagnosis model for case identification in VHM (Fig. 2B).[25] For the case identification model training, the inputs were morphologic features obtained from the case dataset. For VHM application, the inputs were morphologic features acquired from the image-based morphologic feature extraction model. SVM is 1 of the most influential supervised learning approaches for 2-class classification. It aims to use a hyperplane to separate 2 groups of points with the largest margin. SVM can deal well with classifying linearly inseparable points in the original data space. In SVM, the points are transformed into a designed high-dimensional feature space for classification in which these points become linearly separable.

End-To-End Framework

ResNet-101 algorithm (Supplementary Fig. S4) was introduced to develop an end-to-end diagnostic framework (Fig. 3).[26] ResNet is a deep convolutional neural network commonly used for image classification. It has special residual modules to increase the depth of the network and avoid the vanishing gradient or the exploding gradient problems. ResNet-101 is a widely used ResNet that has 101 convolutional layers. The end-to-end diagnostic framework and VHM use the same deep neural network structure for classification. In the end-to-end diagnostic framework, the W-BM (×100), W-BM (×40), and A-PB images (×100) of each case were concatenated to generate the input of ResNet-101. The output of ResNet-101 was the diagnostic result.

Statistical Analysis

Recall and precision were used to evaluate the performance of the image recognition modules by assessing whether the predictions were consistent with ground truth. Balanced accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUROC) were used to evaluate the performance of the case identification modules and VHM. In the test on image recognition modules, the ground truth was the labels of the cells that were checked by 4 experienced morphologists. The ground truth in the test on case identification modules was the clinical morphological diagnosis. In the evaluation of the performance of VHM, the ground truth was determined following World Health Organization diagnostic criteria, which included diagnoses based on morphology, immunology, cytogenetics, and molecular biology.[1] For example, the diagnosis of CML-CP was supported by the detection of the chromosomal translocation (9;22) (q34;q11) and the *BCR::ABL1* fusion gene. Recall, precision, sensitivity, specificity, accuracy, and balanced accuracy are related to true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) rates with the following formulas:

$$1.\ \text{Recall} = \frac{TP}{TP + FN}$$

$$2.\ \text{Precision} = \frac{TP}{TP + FP}$$

$$3.\ \text{Sensitivity} = \frac{TP}{TP + FN}$$

$$4.\ \text{Specificity} = \frac{TN}{TN + FP}$$

$$5.\ \text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}$$

$$6.\ \text{Balanced accuracy} = \frac{TPR + TNR}{2} = \frac{\frac{TP}{TP+FP} + \frac{TN}{TN+FN}}{2}$$

Generalization Ability

The generalization ability of an AI system represents its ability to perform well on previously unobserved inputs. Notably, the generalization ability is of key importance to the performance of a deep neural network. It indicates the learning ability of a deep network. In this study, the generalization ability of the diagnostic framework was defined by the ratio of testing accuracy to training accuracy. Testing accuracy refers to the classification accuracy in the situation where the training set is used for training and the testing set is used for test,

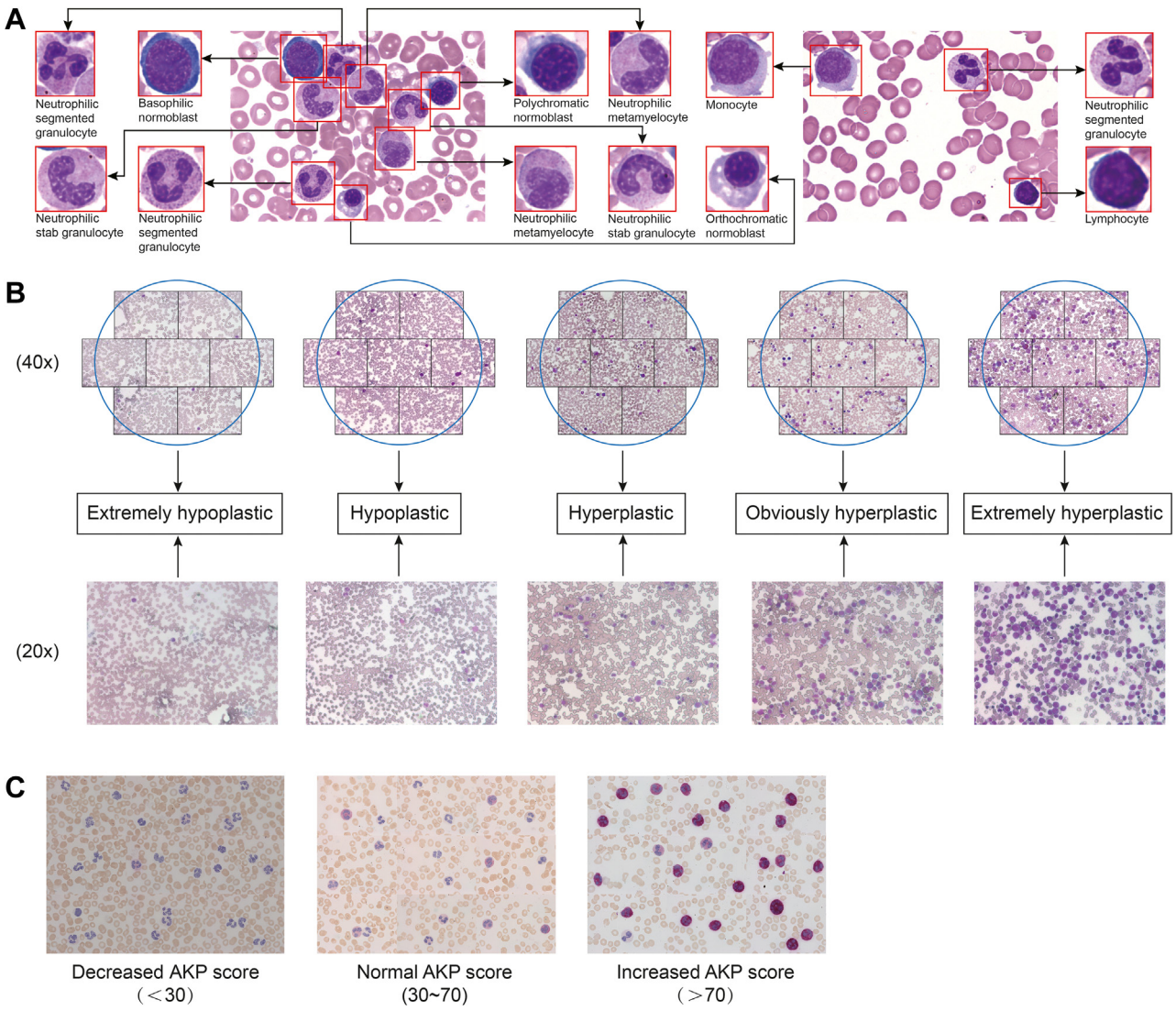


Figure 4.
Test results of the image-based morphologic feature extraction model. (A) Examples of bone marrow cells detected and classified by the bone marrow cell classification module. (B) Examples of different degrees of bone marrow hyperplasia with Wright staining bone marrow images (magnification, ×40) judged by the bone marrow hyperplasia assessment module, as well as their corresponding low-power (×20) images. (C) Examples of cases judged on the basis of different levels of alkaline phosphatase (AKP) scores by the AKP score calculation module.

while training accuracy refers to the classification accuracy in the situation where the training set is used for training and the the same training set is used for test.

$$\text{Generalization ability} = \frac{\text{Testing accuracy}}{\text{Training accuracy}}$$

In general, the testing accuracy is poorer than the training accuracy. Therefore, the generalization ability usually varies within the interval (0,1).

Results

Performance of the Image-Based Morphological Feature Extraction Model

The image-based morphologic feature extraction model consisted of the following 3 parallel image recognition modules: (1) the BMC classification, (2) the bone marrow hyperplasia assessment, and (3) the AKP score calculation modules.

Bone Marrow Cell Classification Module

The W-BM (×100) subset in the image dataset was composed of 12,466 W-BM (×100) images, which contained 50,084 nucleated cells from 18 main cell types of 6 cell lineages in normal myelograms. A BMC classification module was established on the basis of the Faster R-CNN network to detect and classify nucleated cells (Fig. 4A) and was evaluated through 4 random tests. For each random test, images in the subset were randomly divided into the training and testing sets at the ratio of 4:1 (Supplementary Table S1). The average result of the 4 random tests is shown in Table 1. The overall recall and precision of the module for the 18 types of nucleated cells were 94.65% and 93.95%, respectively. This module had a recall above 95% for 11 cell types (myeloblasts, promyelocytes, eosinophils, basophils, polychromatic normoblasts, orthochromatic normoblasts, lymphocytes, monocytes, plasmacytes, granular megakaryocytes, and naked nuclei) and

Table 1
Average results of 4 random tests of bone marrow cell classification

Bone marrow cell types	Recall (%)					Precision (%)				
	RT_1	RT_2	RT_3	RT_4	RT_AVG	RT_1	RT_2	RT_3	RT_4	RT_AVG
Myeloblast	97.92	96.94	96.92	98.60	97.60	93.08	93.42	91.86	92.98	92.84
Promyelocyte	94.12	95.86	96.9	96.75	95.91	90.78	94.56	88.34	94.04	91.93
Neutrophilic myelocyte	91.86	90.78	90.98	91.81	91.36	98.22	98.99	98.55	98.07	98.46
Neutrophilic metamyelocyte	96.99	94.71	94.01	93.73	94.86	83.33	83.42	84.32	83.56	83.66
Neutrophilic stab granulocyte	89.28	90.65	91.65	89.43	90.25	92.15	89.47	86.99	90.24	89.72
Neutrophilic segmented granulocyte	92.54	91.14	89.46	92.01	91.29	92.97	94.31	95.76	92.62	93.91
Eosinophil	99.71	99.72	98.86	99.73	99.5	99.14	98.62	99.14	99.73	99.16
Basophil	99.35	99.3	98.8	100	99.36	98.09	100	97.62	98.72	98.61
Normoblast + Basophilic normoblast	92.64	94.9	94.19	93.9	93.91	95.41	97.89	97.48	99.82	97.65
Polychromatic normoblast	96.79	97.47	97.76	98.48	97.63	97.61	97.95	96.64	97.08	97.32
Orthochromatic normoblast	95.96	97.65	95.43	95.67	96.18	94.25	92.04	92.61	93.21	93.03
Lymphocyte	99.48	99.03	99.28	99.08	99.22	96.11	96.78	95.73	97.08	96.43
Monocyte	94.85	95.63	93.3	97.72	95.37	93.88	94.42	94.95	94.38	94.4
Plasmacyte	100	98.18	98.03	98.39	98.65	99.39	98.78	98.68	98.92	98.94
Granular megakaryocyte	94.52	95.71	96.64	95.8	95.67	83.64	91.23	91.14	91.95	89.49
Platelet-producing megakaryocyte	73.87	84.21	86.79	88.18	83.26	87.23	78.43	90.2	91.51	86.84
Naked nucleus	97.87	95.35	97.92	97.5	97.16	95.83	91.11	92.16	90.7	92.45
Overall	94.53	94.7	94.47	94.91	94.65	93.89	94.08	93.6	94.22	93.95

AVG, average; RT, random test.

above 90% for most of the other cell types. The detailed results of the 4 random tests are shown in the confusion matrix in Supplementary Table S2.

Bone Marrow Hyperplasia Assessment Module

The W-BM (×40) subset contained 139 images collected from myelograms with different degrees of bone marrow hyperplasia. With the support of the Faster R-CNN network, a bone marrow hyperplasia assessment module was trained to judge the degree of bone marrow hyperplasia. The degree of marrow hyperplasia was categorized into 5 grades ranging from extremely hypoplastic, hypoplastic, hyperplastic, obviously hyperplastic, to extremely hyperplastic in accordance with the nucleated cell counts in an HPF, which was approximately equivalent to the area of 7 adjacent W-BM images (×40). Detailed information on the grading criteria for the degree of bone marrow hyperplasia is shown in Supplementary Table S3. Images in this subset were used as the training set, and 12 cases formed a testing set to evaluate the performance of this module. Three regions with well-distributed cells were randomly selected for each case in the testing set. A triplicated test was performed, and the average nucleated cell counts in these 3 regions were used to assess the grade of marrow hyperplasia. The consistency rate between the AI and manual work in the case-level assessment of bone marrow hyperplasia was 100% (Supplementary Table S4). The results representing different degrees of bone marrow hyperplasia and the corresponding low-power (×20) images are presented in Figure 4B.

Alkaline Phosphatase Score Calculation Module

The A-PB (×100) subset consisted of 1,409 images containing 1,894 mature neutrophils with different degrees of AKP staining. An AKP score calculation module based on the Faster R-CNN network was established to recognize AKP-stained cells automatically. As per AKP-staining degrees, mature neutrophils in peripheral blood smears were usually marked as "−" to "++++," which indicated scores from "0" to "4," respectively (Supplementary Method 1). The AKP score was defined as the sum of the scores of 100 mature neutrophils, with a normal range of 30 to 70 points. Decreased and increased AKP scores help differentiate the subtypes of hematological diseases. For example, in CML, the AKP score is significantly decreased, and is usually less than 10 points.[3] The 1,409 images in the A-PB (×100) subset were collected from 14 cases (3 cases with decreased AKP scores, 5 with normal AKP scores, and 6 with increased AKP scores). These images were randomly divided into training and testing sets in a 4:1 ratio to evaluate the module's performance at the cell classification level. The overall recall rate was 93.47% (Supplementary Table S5). The consecutively collected A-PB images (×100) of 12 cases were tested, and the AKP score was calculated to evaluate the performance at the case level. Results for the different AKP score levels are presented in Figure 4C. The judgment of normal, increased (>70 points), or decreased (<30 points) levels based on the AKP scores provided by the AI module were consistent with those obtained through manual operation (Supplementary Table S6).

Performance of the Feature-Based Case Diagnosis Model

The case dataset consisted of the retrospective morphologic diagnostic data of 593 normal and 2,073 abnormal cases, including 235 cases with acute promyelocytic leukemia with *PML::RARA* fusion, 273 cases with AML defined by differentiation (17 cases with AML without maturation, 75 cases with AML with maturation, 83 cases with acute myelomonocytic leukemia, and 98 cases with acute monocytic leukemia), 214 cases with acute lymphoblastic leukemia, 370 cases with CML-CP, 184 cases with chronic lymphocytic leukemia, 184 cases with *BCR::ABL1*-negative myeloproliferative neoplasms, 223 cases with multiple myeloma, and 390 cases with hypoplastic myelograms (mainly aplastic anemia). Two case identification modules were developed based on SVM. The first was designed to distinguish normal and abnormal cases, and the second was used for identifying CML-CP cases.

Case Identification Module for Distinguishing Normal and Abnormal Cases

The proportions of 15 cell types (the 3 subtypes of megakaryocytes were not included in the proportion calculation), the degrees of bone marrow hyperplasia, and AKP scores were selected

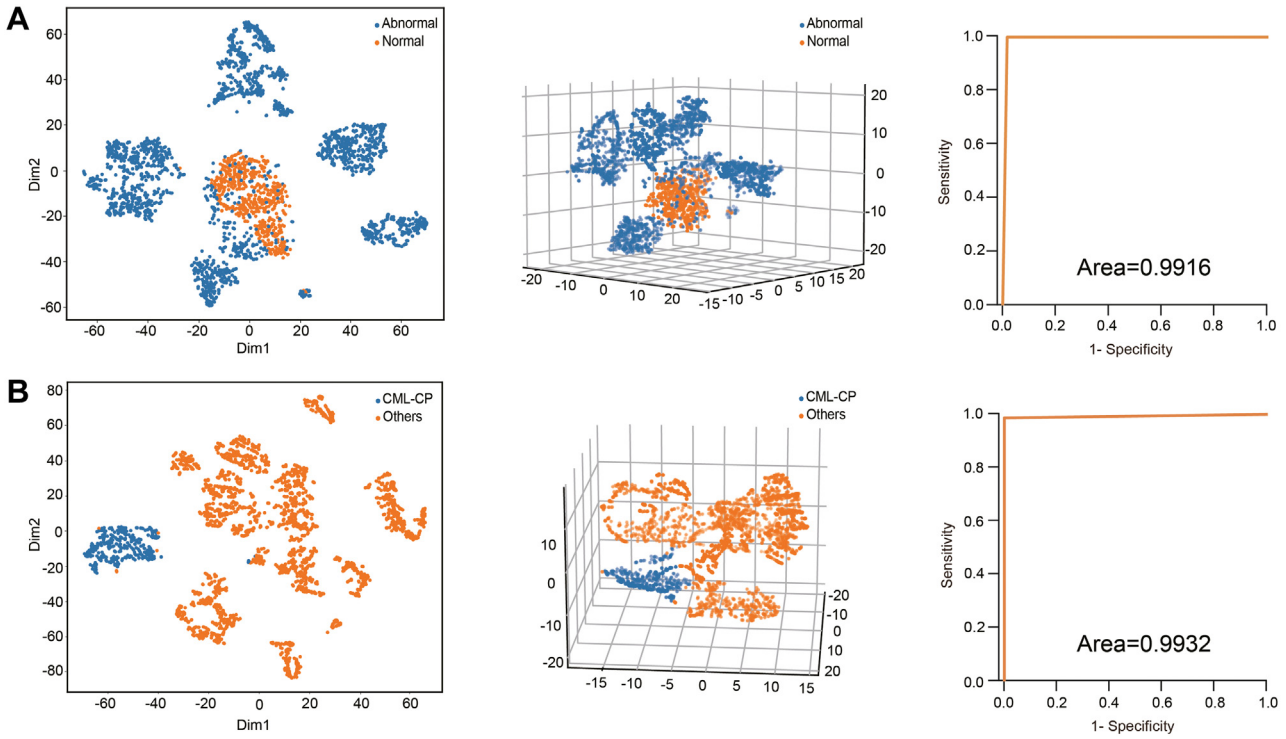


Figure 5.
Visualization of case distribution and test results of the feature-based case diagnosis model. (A) Left and middle columns: Two- and 3-dimensional visualization of the distribution of normal and abnormal cases in the 17-dimensional morphologic feature space. Right column: Receiver operating characteristic curve of the case identification module for distinguishing normal and abnormal cases. (B) Left and middle columns: Two- and 3-dimensional visualization of the distribution of chronic myelogenous leukemia in chronic phase (CML-CP) and other cases in the 7-dimensional morphologic feature space. Right column: Receiver operating characteristic curve of the case identification module for identifying CML-CP cases.

as morphologic diagnostic criteria to distinguish normal and abnormal cases. A 17-dimensional morphologic feature space was constructed accordingly, and the *t*-distributed stochastic neighbor embedding method was used to visualize the distribution of all 2,666 cases. The results showed that the normal case points clustered together and were separated from the abnormal points in 2- and 3-dimensional mapping spaces (Fig. 5A, left and middle column). In addition, 593 normal and 2,073 abnormal cases in the case dataset were randomly divided into the training and testing sets in a 4:1 ratio (Table 2). The balanced accuracy, sensitivity, specificity, and AUROC of the case identification module for identifying abnormal cases was 99.76%, 100%, 98.32%, and 0.9916, respectively (Table 3, Fig. 5A, right column).

Case Identification Module for Identifying Chronic Myelogenous Leukemia in Chronic Phase Cases

On the basis of the staging criteria and morphologic manifestations defined by the World Health Organization for CML, 7 indicators were selected as the morphologic diagnostic criteria for the identification of CML-CP cases, including the degree of bone marrow hyperplasia, AKP score, the proportions of myeloblasts, neutrophils in the intermediate stage (sum of neutrophilic myelocytes, neutrophilic metamyelocytes, and neutrophilic staff cells), eosinophils, basophils, and the ratio of granulocytes to erythroblasts.[1,27,28] The *t*-distributed stochastic neighbor embedding method was used to visualize the distribution of all 2,666 cases with these 7 morphologic features. The results revealed that the CML-CP case points clustered well and were remarkably separated from the others in 2- and 3-dimensional mapping spaces (Fig. 5B, left and middle columns). In addition, 593 normal cases, 370 cases with CML-CP, and 1,703 other abnormal cases were randomly divided into training and testing sets at a 4:1 ratio (Table 2). The balanced accuracy, sensitivity, specificity, and AUROC of the case identification module in identifying patients with CML-CP was 99.89%, 98.65%, 100%, and 0.9932, respectively (Table 3, Fig. 5B, right column).

Performance of Virtual Hematological Morphologist and Its Comparison With That of the End-To-End Framework

VHM is a whole-process AI-aided diagnostic framework that was established by integrating an image-based morphologic feature extraction model and a feature-based case diagnosis model. A 2-stage strategy was used in the case diagnosis model (Fig. 6A). The differential diagnosis of normal and abnormal cases was performed in the first stage. The predicted abnormal cases in the first stage were further precisely classified into CML-CP and others in the second stage. A total of 50 normal, 50 CML-CP, and 60 other abnormal cases (Table 2) were used as an integrated testing set to test the performance of VHM. For each case, 7 adjacent W-BM (×40), W-BM (×100) containing at least 500 nucleated cells, and A-PB (×100) images containing at least 100 mature neutrophils were collected as the integrated inputs of VHM. The balanced accuracy, sensitivity, specificity, and AUROC was 97.16%, 99.09%, 92.00%, and 0.9555, respectively, for the differential diagnosis of normal and abnormal cases in the first stage (Table 4, Fig. 6B), and 99.23%, 97.96%, 100%, and 0.9898, respectively, for the precise diagnosis of CML-CP cases in the second stage (Table 4, Fig. 6C).

Table 2
Case distribution in training and testing sets

Types of diseases	Case identification modules		VHM	End-to-end framework	
	Training set, cases (%)	Testing set, cases (%)	Integrated testing set, cases	Training set, cases (%)/images (%)	Testing set, cases (%)/images (%)
Normal	474 (79.93)	119 (20.07)	50	40 (80)/12130 (79.17)	10 (20)/3192 (20.83)
AML defined by differentiation[a]	219 (80.22)[b]	54 (19.78)[c]	10[d]	8 (80)/1403 (78.64)[e]	2 (20)/381 (21.36)[f]
APL with *PML::RARA* fusion	188 (80)	47 (20)	10	8 (80)/1088 (80.89)	2 (20)/257 (19.11)
CML-CP	296 (80)	74 (20)	50	40 (80)/7130 (80.09)	10 (20)/1773 (19.91)
ALL	171 (79.91)	43 (20.09)	10	8 (80)/1260 (79.10)	2 (20)/333 (20.90)
CLL	147 (79.89)	37 (20.11)	10	8 (80)/1581 (82.30)	2 (20)/340 (17.70)
BCR::ABL1-negative MPN	148 (80.43)	36 (19.57)	10	8 (80)/2252 (77.68)	2 (20)/647 (22.32)
MM	178 (79.82)	45 (20.18)	10	8 (80)/2434 (81.19)	2 (20)/564 (18.81)
Cases with hypoplastic myelogram	312 (80)	78 (20)	0	0/0	0/0
Total	2133 (80.01)	533 (19.99)	160	128 (80)/29251 (79.62)	32 (20)/7487 (20.38)

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; APL, acute promyelocytic leukemia; CLL, chronic lymphocytic leukemia; CML-CP, chronic myelogenous leukemia in chronic phase; MM, multiple myeloma; MPN, myeloproliferative neoplasms; VHM, virtual hematological morphologist.

[a] The AML family of "AML defined by differentiation" cover the AML subtypes of AML without maturation (AML-M_1), AML with maturation (AML-M_2), acute myelomonocytic leukemia (AML-M_4), and acute monocytic leukemia (AML-M_5).

[b] Fourteen cases with AML-M_1, 60 cases with AML-M_2, 67 cases with AML-M_4, and 78 cases with AML-M_5.

[c] Three cases with AML-M_1, 15 cases with AML-M_2, 16 cases with AML-M_4, and 20 cases with AML-M_5.

[d] Five cases with AML-M_2, 4 cases with AML-M_4, and 1 case with AML-M_5.

[e] Four cases with AML-M_2, 3 cases with AML-M_4, and 1 case with AML-M_5.

[f] One case with AML-M_2 and 1 case with AML-M_4.

The end-to-end framework is commonly used for image-based diagnosis.[29-32] Therefore, this study also developed this framework to compare its performance with VHM in distinguishing normal and abnormal cases. A total of 50 normal and 110 abnormal cases were randomly divided into the training and testing sets in a 4:1 ratio (Table 2). For each case, the images used in the end-to-end framework were the same as the following input of VHM: (1) W-BM ($\times$40, 7 images per case), (2) W-BM (100 x, 134 $\pm$ 76 images per case on average), and (3) A-PB ($\times$100, 88 $\pm$ 32 images per case on average) images. The training and testing accuracies and the generalization abilities, represented by the ratio of testing accuracy to training accuracy, are shown in Table 5. The testing accuracy and generalization ability of the end-to-end framework was 68.75% and 68.75%, respectively. In contrast, the testing accuracy and generalization ability of VHM reached 96.88% and 97.11%, respectively.

Discussion

Recently, the development of AI technology for morphologic examination has attracted increasing attention. Most research on AI-based bone marrow cytomorphology has focused on automated BMC classification.[21-23] These approaches do not take into account other diagnostic features that are frequently used by medical experts. Herein, VHM, a promising AI-aided diagnostic framework that made full use of practical clinical diagnostic criteria, was developed for the morphologic diagnosis of hematologic diseases.

Besides the image dataset, which was used for cell recognition, a case dataset was established in this study. The morphologic diagnostic indices extracted from the case dataset are helpful for introducing medical knowledge into the diagnostic model. Based on these 2 datasets, the image-based morphologic feature extraction model and the feature-based case diagnosis model in our AI-aided diagnosis framework could be trained reliably.

The image-based morphologic feature extraction model consisted of 3 parallel image recognition modules. Previous research on automatic morphologic diagnostic methods mainly focused on BMC classification without considering other morphologic features.[22-24] However, the degree of bone marrow hyperplasia and cytochemical staining scores are important for diagnosing hematological malignancies. The degree of bone marrow hyperplasia varies in hematological diseases. In proliferative hematological malignancies, such as leukemia, the cellularity of the bone marrow is extremely hyperplastic. Hypocellularity indicates the possible existence of diseases with bone marrow failure, such as aplastic anemia and myelofibrosis. Cytochemical staining scores are helpful in differentiating the subtypes of hematological diseases of which the AKP score is an important one. Decreased AKP

Table 3
The performance of case identification modules

Case identification module	Prediction, n (%)		Balanced Accuracy (%)	Sensitivity (%)	Specificity (%)	AUROC
	True	False				
Case identification for abnormal cases			99.76	100	98.32	0.9916
Abnormal	414 (100)	0 (0)				
Normal	117 (98.32)	2 (1.68)				
Case identification for CML-CP			99.89	98.65	100	0.9932
CML-CP	73 (98.65)	1 (1.35)				
Others	459 (100)	0 (0)				

AUROC, area under the receiver operating characteristic curve; CML-CP, chronic myelogenous leukemia in chronic phase.

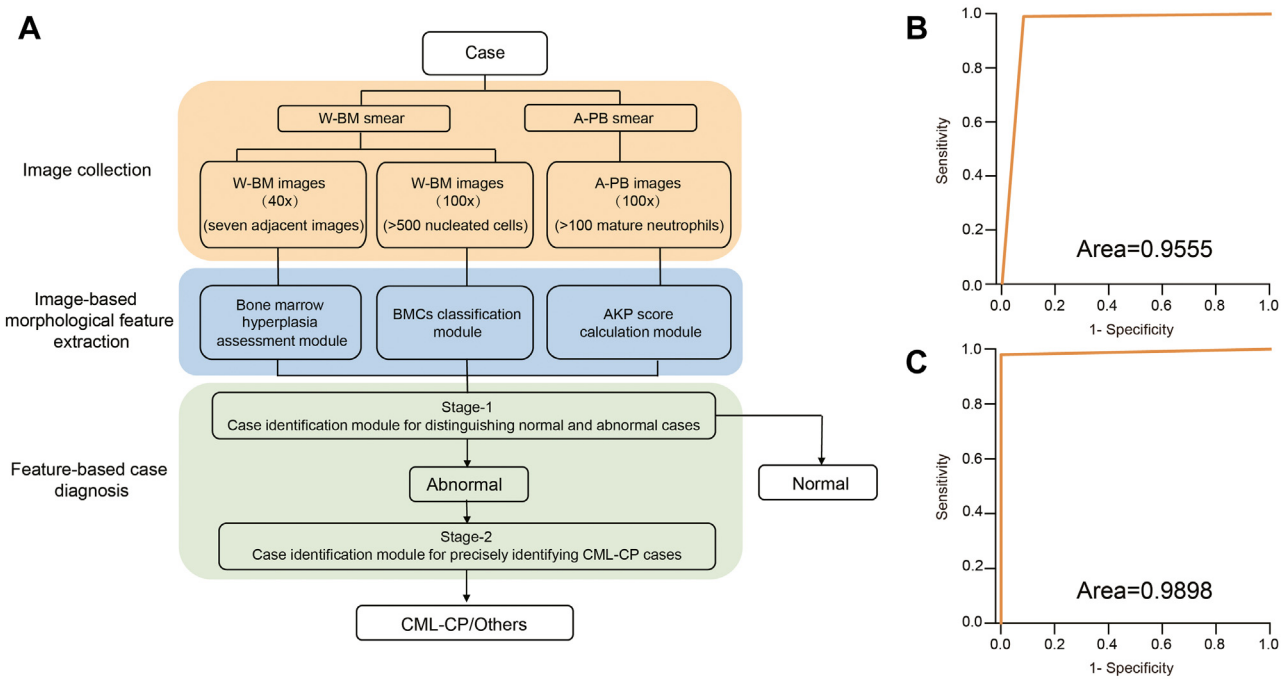


Figure 6.
Workflow and performance of virtual hematological morphologist. (A) The diagnostic strategy of virtual hematological morphologist. (B) Receiver operating characteristic curve for the differential diagnosis of normal and abnormal cases. (C) Receiver operating characteristic curve for the precise diagnosis of chronic myelogenous leukemia in chronic phase (CML-CP) from predicted abnormal cases. A-PB, AKP-staining peripheral blood; AKP, alkaline phosphatase; BMCs, bone marrow cells; W-BM, Wright staining bone marrow.

scores are frequently observed in CML-CP cases, whereas increased AKP scores are common in myeloproliferative diseases and infection.[33] In our study, 2 image recognition modules were designed to extract the degree of bone marrow hyperplasia, AKP scores, and the proportion of BMCs. The entire model performed well. Specifically, the recall and precision of the BMC classification module reached 94.65% and 93.95%, respectively. Moreover, we found a 100% consistency rate in the results of the bone marrow hyperplasia assessment and AKP score calculation modules with those of manual work. The integration of these 3 modules enriched the morphologic features and considerably improved diagnostic accuracy. Indeed, without the support of the degree of bone marrow hyperplasia and AKP score, the final results of balanced accuracy, sensitivity, and specificity in the differential diagnosis of normal and abnormal cases was only 84.64%, 96.36%, and 54%, respectively (Supplementary Table 7), which had improved to 97.17%, 99.09%, and 92%, respectively.

In the feature-based case diagnosis model, multiple diagnostic morphologic indices were selected to construct multidimensional feature spaces. The selection of various morphologic features corresponded to clinical diagnostic criteria, and therefore, benefited from the medical knowledge and experience of morphologic experts. Furthermore, in contrast to the manual diagnostic method based on the thresholding a single criterion, our case diagnosis model based on the SVM algorithm fully used the correlations and interactions among multiple morphologic features. In this work, normal and abnormal, CML-CP, and other cases were well separated in the high-dimensional morphologic feature space. These 2 classification tasks could well be accomplished through the SVM approach even though both classification problems were not linearly separable problems. In addition, SVM presented a superior performance in this model compared with the 2 other algorithms widely used, random forest and multilayer perceptron. In the 2 case identification modules, the balanced accuracies of SVM were 99.76% and 99.89%, 99.64% and 99.89% for the random forest, and 98.24% and 99.10% for multilayer perceptron (Table 6). The application of SVM may be helpful for optimizing or discovering new underlying diagnostic criteria for hematological diseases in the future.

Table 4
The performance of the 2 stages of virtual hematological morphologist

Two stages of VHM	Prediction, n (%)		Balanced accuracy	Sensitivity	Specificity	AUROC
	True	False				
First stage of VHM			97.16	99.09	92	0.9555
Abnormal	109 (99.09)	1 (0.91)				
Normal	46 (92)	4 (8)				
Second stage of VHM			99.23	97.96	100	0.9898
CML-CP	48 (97.96)	1 (2.04)				
Others	64 (100)	0 (0)				

AUROC, area under the receiver operating characteristic curve; CML-CP, chronic myelogenous leukemia in chronic phase; VHM, virtual hematological morphologist.

Table 5
Comparison of end-to-end strategy and knowledge-based strategy (virtual hematological morphologist) in distinguishing normal and abnormal cases

Artificial intelligence frameworks	Training accuracy, (%)	Testing accuracy, (%)	Generalization ability, (%)[a]
End-to-end framework	100	68.75	68.75
Knowledge-based framework (VHM)	99.76	96.88	97.11

VHM, virtual hematological morphologist.
[a] Generalization ability = (testing accuracy)/(training accuracy).

VHM, which was established through the combination of 2 effective AI models, achieved the desired performance with a balanced accuracy of 97.16% in the identification of abnormal cases in the first stage and 99.23% in the precise diagnosis of CML-CP in the second stage. Notably, in addition to fully considering multimodal morphologic features and medical knowledge, the appropriate deep learning strategy was crucial. The end-to-end framework based on the deep neural network is the most widely used in various image-based AI-aided applications and is well implemented in image-based diagnosis for several diseases.[29-32] Therefore, we established an end-to-end diagnostic framework as the baseline for a fair comparison with VHM in distinguishing normal and abnormal hematologic cases. Our results regarding framework comparison showed that the knowledge-based VHM greatly outperformed the end-to-end framework. Although the training accuracy of VHM and the end-to-end framework reached 99.76% and 100%, respectively, the end-to-end framework's testing accuracy and generalization ability were considerably less desirable than those of VHM (68.75% vs 96.88%, 68.75% vs 97.11%, respectively) (Table 5).

The core idea of the end-to-end framework is to learn unknown image features automatically by mapping an input vector to an output vector without the guidance of explicit knowledge. However, the current end-to-end framework may not be able to accomplish a complex AI task that requires time to think and reflect to perform or cannot be viewed as mapping one vector to another.[34] The task of hematological diagnosis based on morphologic features requires the extraction and quantitative analysis of multimodal features, such as the density and proportion of nucleated cells in a certain area of bone marrow smears and positive degree of cytochemical staining. Therefore, the end-to-end framework may be unsuitable in this case. In this scenario, the poor testing accuracy (68.75%) of the end-to-end framework indicated that its desired training accuracy (100%) signified the drawback of overfitting rather than its advantage.[35] This drawback was expressed in the framework's poor generalization ability (68.75%), which is impractical in clinical diagnosis applications. In contrast, VHM fully used explicit medical expertise, such as multimodal morphologic features and diagnostic criteria, in the diagnosis procedure. The desired testing accuracy (96.88%) and generalization ability (97.11%) of VHM indicated that the knowledge-based framework outperformed the end-to-end framework in hematological diagnosis.

Table 6
The balanced accuracy of different algorithms in case identification

Case identification	RF, (%)	MLP, (%)	SVM, (%)
Normal and abnormal	99.64	98.24	99.76
CML-CP and others	99.89	99.10	99.89

CML-CP, chronic myelogenous leukemia in chronic phase; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine.

The interpretational ability of the AI diagnostic framework was also important. The learning process of the end-to-end diagnostic framework was similar to a black box wherein extracted intermediate features stayed unknown, which led to poor interpretability.[35] In contrast, VHM integrated different AI models with other purposes under the guidance of the medical diagnostic procedure. Features extracted from images covered various practical clinical diagnostic indices, including marrow hyperplasia, nucleated cell proportions, and cytochemical staining scores, which made VHM a medically interpretable diagnostic tool.

Various studies have proven the powerful potential of image-based AI diagnosis.[3-6] However, a critical problem regarding the diagnosis of various diseases centers around which AI-aided framework is suitable; whether it is an end-to-end framework, knowledge-based framework, or another new framework that has yet to be developed. On the basis of the results of this study, we concluded that for the diagnosis of diseases that require complex feature processing rather than simple mapping, the integration of explicit medical expertise into a comprehensive AI-aided workflow in a way that is similar to the clinical diagnostic procedure is remarkably valuable. VHM is one such good example.

Finally, further improvements to the AI-aided hematological diagnostic framework based on morphologic features should be considered in the future. One of its limitations is that the number of samples was small, so the image and case datasets need to be continuously enlarged through data accumulation. Furthermore, although we successfully identified CML-CP cases based on the well-trained modules of BMC recognition, marrow hyperplasia, and AKP score assessment in this study, various image and case identification modules should be trained to extract additional morphologic features, such as the features of different types of leukemia cells (abnormal promyelocytes, monoblasts, and lymphoblasts) and other types of cytochemical staining (myeloperoxidase, periodic acid-Schiff, naphthol AS-D chloroacetate esterase, and naphthol AS-D acetate esterase),[36] to identify additional subtypes of leukemia precisely. Other efficient AI techniques may also be selected to replace Faster R-CNN and SVM to improve performance, and new approaches need to be developed for the precise diagnosis of other hematological diseases. Moreover, external and multicenter validation is required.

Acknowledgments

The authors sincerely appreciate the guidance and support from Prof Zhu Chen during the whole project.

Author Contributions

B.C., Y.Q., and S.C. designed and supervised the study. N.L., M.L., and L.F. collected the image and case data. H.X., X.Z., and Z.B. developed and coded the algorithms. N.L., H.X., and X.Z. analyzed the data. L.F., S.X., and B.C. reviewed the annotated image data. N.L., H.X., L.J., Y.Q., and B.C. wrote the paper. All authors read and approved the final manuscript.

Data Availability

For original data, please contact chenbing_rjh@163.com.

Funding

This study was supported by the National Natural Science Foundation of China grant (81670137), Shanghai Municipal Education Commission-Gaofeng Clinical Medicine Grant Support grant (20152501), and State Key Laboratory of Medical Genomics Support grant (201802).

Declaration of Competing Interest

None reported.

Ethics Approval and Consent to Participate

The study was approved by the review boards of the State Key Laboratory of Medical Genomics and conducted in accordance with the Declaration of Helsinki. All the images and cases were anonymized and deidentified during the development of artificial intelligence algorithms.

Supplementary Material

The online version contains supplementary material available at https://doi.org/10.1016/j.labinv.2022.100055

References

1. Khoury JD, Solary E, Abla O, et al. The 5th edition of the World Health Organization Classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms. *Leukemia*. 2022;36(7):1703–1719.
2. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. *Leukemia*. 2022;36(7):1720–1748.
3. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31–38.
4. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500–510.
5. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019;20(5):e253–e261.
6. Xie X, Fu CC, Lv L, et al. Deep convolutional neural network-based classification of cancer cells on cytological pleural effusion images. *Mod Pathol*. 2022;35(5):609–614.
7. Bibi N, Sikandar M, Ud Din I, Almogren A, Ali S. IoMT-based automated detection and classification of leukemia using deep learning. *J Healthc Eng*. 2020;2020:6648574.
8. Karar ME, Alotaibi B, Alotaibi M. Intelligent medical IoT-enabled automated microscopic image diagnosis of acute blood cancers. *Sensors (Basel)*. 2022;22(6):2348.
9. Jin H, Fu X, Cao X, et al. Developing and preliminary validating an automatic cell classification system for bone marrow smears: A pilot study. *J Med Syst*. 2020;44(10):184.
10. Fu X, Fu M, Li Q, et al. Morphogo: an automatic bone marrow cell classification system on digital images analyzed by artificial intelligence. *Acta Cytol*. 2020;64(6):588–596.
11. Wu YY, Huang TC, Ye RH, et al. A hematologist-level deep learning algorithm (BMSNet) for assessing the morphologies of single nuclear balls in bone marrow smears: algorithm development. *JMIR Med Inform*. 2020;8(4):e15963.
12. Choi JW, Ku Y, Yoo BW, et al. White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. *PLoS One*. 2017;12(12):e0189259.
13. Su J, Han J, Song J. A benchmark bone marrow aspirate smear dataset and a multi-scale cell detection model for the diagnosis of hematological disorders. *Comput Med Imaging Graph*. 2021;90:101912.
14. Chandradevan R, Aljudi AA, Drumheller BR, et al. Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells. *Lab Invest*. 2020;100(1):98–109.
15. 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.
16. Al-Qudah R, Suen CY. Improving blood cells classification in peripheral blood smears using enhanced incremental training. *Comput Biol Med*. 2021;131:104265.
17. Reena MR, Ameer PM. Localization and recognition of leukocytes in peripheral blood: A deep learning approach. *Comput Biol Med*. 2020;126:104034.
18. Hegde RB, Prasad K, Hebbar H, Singh BMK. Feature extraction using traditional image processing and convolutional neural network methods to classify white blood cells: A study. *Australas Phys Eng Sci Med*. 2019;42(2):627–638.
19. Chari PS, Prasad S. Pilot study on the performance of a new system for image based analysis of peripheral blood smears on normal samples. *Indian J Hematol Blood Transfus*. 2018;34(1):125–131.
20. Shahin AI, Guo Y, Amin KM, Sharawi AA. White blood cells identification system based on convolutional deep neural learning networks. *Comput Methods Programs Biomed*. 2019;168:69–80.
21. Ouyang N, Wang W, Ma L, et al. Diagnosing acute promyelocytic leukemia by using convolutional neural network. *Clin Chim Acta*. 2021;512:1–6.
22. Boldú L, Merino A, Acevedo A, Molina A, Rodellar J. A deep learning model (ALNet) for the diagnosis of acute leukaemia lineage using peripheral blood cell images. *Comput Methods Programs Biomed*. 2021;202:105999.
23. Zhou M, Wu K, Yu L, et al. Development and evaluation of a leukemia diagnosis system using deep learning in real clinical scenarios. *Front Pediatr*. 2021;9:693676.
24. Ren S, He K, Girshick R, Sun J, Faster R-CNN. towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(6):1137–1149.
25. Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24(12):1565–1567.
26. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition*. Las Vegas, NV: Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. Accessed December 10, 2015.
27. Jabbour E, Kantarjian H. Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. *Am J Hematol*. 2020;95(6):691–709.
28. Shen T, Zhao YQ. *Criteria for Diagnosis and Efficacy of Hematological Diseases*. Science Press; 2018, 4th ed.
29. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118.
30. Lin D, Xiong J, Liu C, et al. Application of Comprehensive Artificial intelligence Retinal Expert (CARE) system: A national real-world evidence study. *Lancet Digit Health*. 2021;3(8):e486–e495.
31. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25(1):65–69.
32. Chuang WY, Chen CC, Yu WH, et al. Identification of nodal micrometastasis in colorectal cancer using deep learning on annotation-free whole-slide images. *Mod Pathol*. 2021;34(10):1901–1911.
33. Hayhoe FG. Alkaline phosphatase in haematology. *Cell Biochem Funct*. 1983;1(2):74–76.
34. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016.
35. Shouval R, Fein JA, Savani B, Mohty M, Nagler A. Machine learning and artificial intelligence in haematology. *Br J Haematol*. 2021;192(2):239–250.
36. Riley RS, Williams D, Ross M, et al. Bone marrow aspirate and biopsy: a pathologist's perspective. II. interpretation of the bone marrow aspirate and biopsy. *J Clin Lab Anal*. 2009;23(5):259–307.