Medical & Biological Engineering & Computing (2024) 62:575–589
https://doi.org/10.1007/s11517-023-02955-3

ORIGINAL ARTICLE

CDC-NET: a cell detection and confirmation network of bone marrow aspirate images for the aided diagnosis of AML

Jie Su[1,2] · Yahui Liu[3] · Jing Zhang[1] · Jinjun Han[1] · Jinming Song[4]

Received: 28 February 2023 / Accepted: 20 October 2023 / Published online: 12 November 2023

Abstract

Standardized morphological evaluation in pathology is usually qualitative. Classifying and qualitatively analyzing the nucleated cells in the bone marrow aspirate images based on morphology is crucial for the diagnosis of acute myoid leukemia (AML), acute lymphoblastic leukemia (ALL), and Myelodysplastic syndrome (MDS), etc. However, it is time-consuming and difficult to accurately identify nucleated cells and calculate the percentage of the cells because of the complexity of bone marrow aspirate images. This paper proposed a deep learning analysis model of bone marrow aspirate images, termed Cell Detection and Confirmation Network (CDC-NET), for the aided diagnosis of AML by improving the accuracy of cell detection and recognition. Specifically, we take the nucleated cells in the bone marrow aspirate images as the detection objects to establish the model. Since some cells from different categories have similar morphology, classification error is inevitable. We design a confirmation network in which multiple trained classifiers work as pathologists to confirm the cell category by a voting method. To demonstrate the effectiveness of the proposed approach, experiments on clinical microscopic datasets are conducted. The Recall and Precision of CDC-NET are 78.54% and 91.74% respectively, and the missed rate of our method is lower than those of the other popular methods. The experimental results demonstrated that the proposed model has the potential for the pathological analysis of aspirate smears and the aided diagnosis of AML.

Keywords Digital diagnosis · Convolutional neural network (CNN) · Ensemble learning · AML · Bone marrow aspirate image

Highlights

- A Cell Detection and Confirmation Network is designed to achieve automatic analysis of the bone marrow aspirate images to assist in AML diagnosis.
- It simulates pathologists by introducing a voting mechanism to balance their perspectives and further achieve consistent results, and it possesses excellent scalability.
- The experimental results demonstrated that the proposed model has the potential for the pathological analysis of aspirate smears and the aided diagnosis of AML.

✉ Jie Su
sujie0001@sina.cn

1 School of Information Science and Engineering, University of Jinan, Jinan, China

2 Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, China

3 School of Information Management, Beijing Information Science & Technology University, Beijing, China

4 Department of Hematopathology and Lab Medicines, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA

1 Introduction

The diagnosis of hematologic disorders, such as AML and MDS, is based on bone marrow biopsies to obtain aspirate smears and core biopsies for manual evaluation by experienced hematopathologists to identify increases in abnormal cells, such as blasts, dysplastic cells, or plasma cells [1]. Acute myeloid leukemia is a kind of hematologic malignancy caused by the neoplastic proliferation of immature myeloid hematopoietic stem cells or myeloblasts [2]. Early diagnosis of AML is crucial for the treatment and the life of the patients. Classifying, counting, and qualitatively analyzing the nucleated cells in the bone marrow aspirates by professional hematopathologists according to the World Health Organization (WHO) classification is the key step for the diagnosis of acute myeloid leukemia. The blasts in the peripheral blood can also be regarded as the biomarker to diagnose AML. However, the percentage of blasts in bone marrow is usually more reliable or important for the diagnosis of AML. According to the criteria of the World

Health Organization (WHO) classification, the blast count of less than 3% in the bone marrow is normal, more than 5% suggests high-grade myelodysplastic syndrome, and 20% or more is diagnostic of AML [1]. Some examples of bone marrow aspirate images are shown in Fig. 1. Images (a–c) are from non-AML patients with no increase in blasts, while images (d–f) are from AML patients with increased blasts.

Manual quantification of blasts by experienced hematopathologists is time-consuming and prone to human errors. Different hematopathologists could render very different blast counts, depending on their personal criteria and personal experience. It is therefore better to come up with a more effective and less subjective way to quantify blasts and to improve the diagnostic efficiency. There have been systems, such as CellaVison (https://www.cellavision.com), to automatically quantify the blasts in the peripheral blood. However, they cannot be adapted to aspirate smears because they are much more complex than peripheral blood smears. We compare the peripheral blood smear images with the bone marrow aspirate images, which are shown in Fig. 2. Compared with the peripheral blood smear images, bone marrow aspirate images are denser and have more cell types. There are more overlapping cells in bone marrow aspirate images, which could result in missed cells in detection. Moreover, some cells in the bone marrow look very similar in morphology and could cause cell type ambiguity.

The complexity of bone marrow aspirate images makes it difficult to accurately identify cells and calculate the blast percentage. To achieve an effective diagnosis of acute leukemia, many researchers began to focus on the automatic classification and counting of nucleated cells on the bone marrow aspirate images [3–12]. Traditional methods usually segment the bone marrow aspirate images to obtain individual cells at first, and then extract morphologic features of the individual cells for cell recognition and counting. Therefore, the performance of the traditional automatic classification and counting methods depends on the accuracy of image segmentation. To improve the accuracy of segmentation, Goutam and Sailaja [13] proposed a cell detection method by using k-means feature extraction, Local Directional path (LDP), and support vector machine (SVM). Li et al. [14] introduced a dual-threshold method based on a strategic combination of RGB and HSV color space for white blood cell (WBC) segmentation. Aris et al. [15] described an automated counting of WBCs with an analysis of watershed segmentation for the screening of chronic leukemia images. However, it is difficult to accurately segment individual cells from bone marrow aspirate images, and even more difficult for those adherent cells.

The analysis methods using deep learning look more promising for the diagnosis based on bone marrow aspirate images, especially when all types of cells in the bone marrow aspirate images need to be quantified. Song et al. [16] proposed a synchronized deep autoencoder network for simultaneous detection and classification of the cells in bone marrow aspirate images. Yang et al. [17] reported a

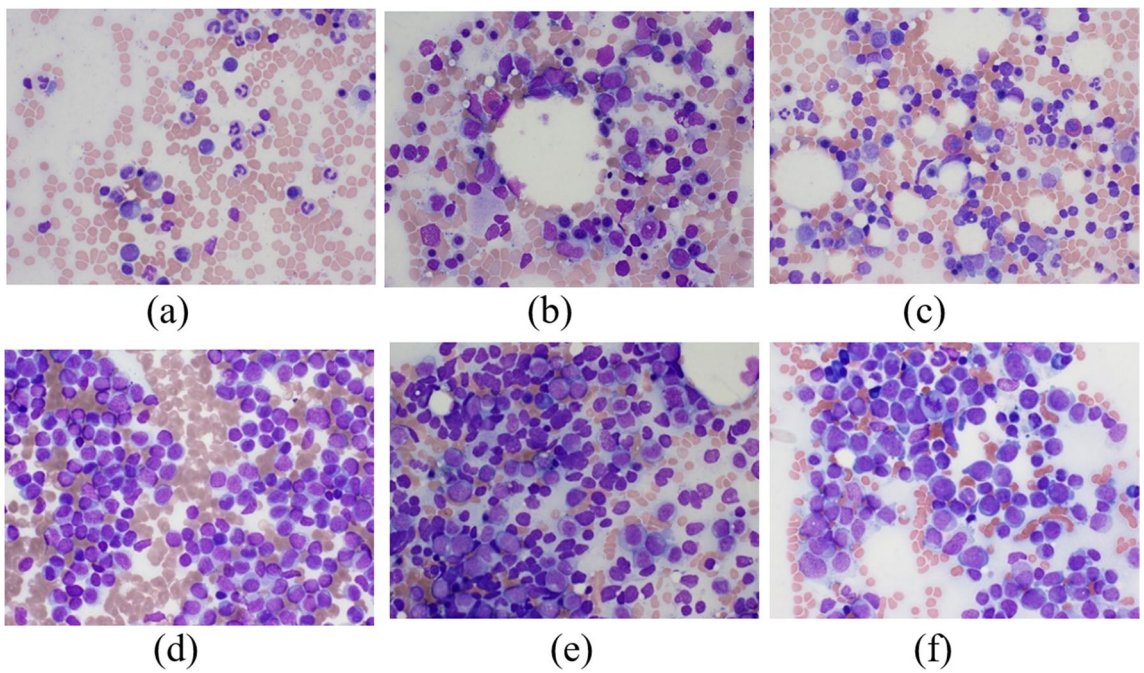
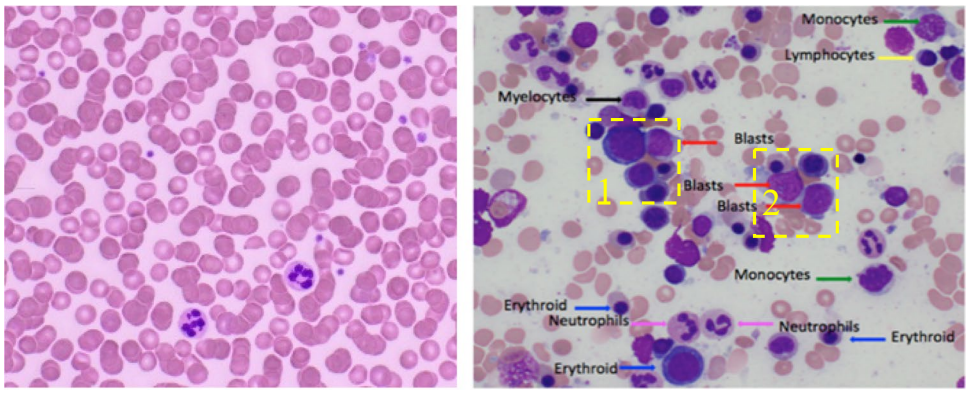


Fig. 1 Examples of bone marrow aspirate smear images. **a**–**c** Bone marrow aspirate images from non-AML patients with no increase in blasts. **d**–**f** Bone marrow aspirate images from AML patients with increased blasts



(a) Peripheral blood smears image (b) Bone marrow aspirate image

Fig. 2 Comparison of peripheral blood smears images and bone marrow aspirate images. Compared with the peripheral blood images, there are more cell types in the bone marrow aspirate images, and the cell morphology is more complex. Cell adhesion, cell overlap, and ambiguous morphology make it difficult to classify the cells in the bone marrow aspirate images. For example, cells in region 1 in the image (**b**) are morphologically intermediate between erythroid and blast cells, which makes definitive classification difficult. Cell adhesion (see the cells in region 2 in the image (**b**)) makes it difficult to isolate the cells accurately. **a** Peripheral blood smears image, **b** Bone marrow aspirate image

new deep neural network employing both complementary and correlated relationships between medical images and clinical information to improve the accuracy of computer-aided diagnosis. Haoyi et al. [18] suggested an end-to-end leukocyte localization and segmentation method, in which a deep convolutional neural network trained on pixel-level prior information was used to locate the region of interest (ROI) of white blood cells and to obtain white blood cell segmentation.

Although great progress has been made in this research field, the classification accuracy needs to be further improved. Therefore, this paper provides a model, i.e., Cell Detection and Confirmation Network (CDC-NET). Specifically, we take the nucleated cells in bone marrow aspirate images as the detecting objects to design and train a detection model. Besides, we designed a Cell Confirmation Network (CC-NET) to improve the classification accuracy by further confirming the types of the detected cells, especially of the cells with ambiguous types. It simulates the diagnosis workflow of hematopathologists and continuously improves the skills of the pathologists and the technicians, i.e., improves the performance of CC-NET. Moreover, we increase the difficulty of the dataset and design a practical evaluation method, to improve the robustness of the proposed model. The major contributions of this paper are listed below:

- For the application of AML diagnosis, we simulate the classification and counting process of the nucleated cells in bone marrow aspirate images under a microscope by pathologists, and design a deep learning model to achieve automatic analysis of the bone marrow aspirate images to assist AML diagnosis. Different from the existing methods, this model simulates pathologists by introducing a voting mechanism to balance their perspectives and further achieve consistent results.
- Since classification error is inevitable even for pathologists, we propose a more practical voting mechanism, i.e., CC-NET, to improve the classification accuracy by further confirming the types of the detected cells. It treats multiple trained classifiers as pathologists or technicians to analyze and judge the cells with ambiguous types. Moreover, pathologists or technicians can intervene in decision-making by interacting with the machine.
- Comprehensive experiments on clinical datasets with more complex images demonstrated the effectiveness of the proposed approach. We also introduce a comprehensive experimental setting to evaluate the performance. The experimental results demonstrated that our approach outperforms the existing methods. Importantly, from the perspective of the pathologist, the missed rate and classification accuracy are more acceptable.

2 Related work

2.1 Classifier

The CNN-based classification network has become one of the most common models in the classification system. AlexNet proposed by Krizhevsky et al. [19] successfully applied some methods such as ReLU, Dropout, and LRN in CNN for the first time and proved the strong feature extraction capability of CNN. Kaiming et al. [20] presented a

residual learning framework to train networks that are substantially deeper than the previously reported, and the effectiveness of this network was demonstrated at the 2015 ImageNet competition. Szegedy et al. [21] proposed Inception V1, which used dense components to approximate the optimal local sparse junction and was demonstrated to be an efficient method to increase network size. However, the training of Deep Neural Networks is complicated by the fact that the distribution of the inputs in each layer changes during training [22]. To solve this problem, Szegedy et al. [23] used standardization as part of the model and used the method for batch normalization (BN) to implement standardization for each small batch of training. Then, they improved Inception V2 and Inception V3 models on the basis of Inception V1. In the task of cell classification using AlexNet, ResNet-50, and Inception V3, the paper [24] focused on the automatic classification of leukocytes. The experimental results revealed that ResNet-50 exhibits the highest classification accuracy, particularly in handling noisy and blurry cell images. Another paper [25] investigated the automatic classification of malaria-infected images, showing excellent performance from all three models in the classification task of malaria-infected images, with Inception V3 demonstrating the best performance. Moreover, the paper [26] conducted cell classification of breast fine needle aspiration cytology images, and comparative experiments indicated that ResNet-50 and Inception V3 models outperformed others in terms of classification accuracy, especially for complex cell morphology and structure. Additionally, the AlexNet model continues to exhibit good performance in specific tasks.

2.2 Detection methods

The CNN-based detection method of nucleated cells in bone marrow aspirate images usually consists of two tasks, i.e., localization and classification. Traditional detection methods usually take part of the image as a candidate region by using different sizes of sliding windows, and then extract the visual features related to the candidate region and identify the objects by classifiers. Ren et al. [27] proposed Faster R-CNN and introduced a Region Proposal Network (RPN) that shared full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. Liu et al. [28] introduced a method named SSD to detect objects in images using a single deep neural network. Redmon et al. [29] described a new approach that detected objects extremely fast, in which object detection was treated as a regression problem to spatially separate bounding boxes and associated class probabilities. However, this method has weak generalization capability and large positioning errors on small groups that are close to each other. Upon this, a series of improved models have been proposed [30, 31]. Bochkovskiy et al. [32] reported a model that combines many cell features, shortened the information path between the bottom and top features, and made full use of feature fusion to obtain more semantic information.

3 Methods

An overview of CDC-NET is illustrated in Fig. 3. To improve the interpretability, CDC-Net is designed as a cascade model, which consists of the cell detection network

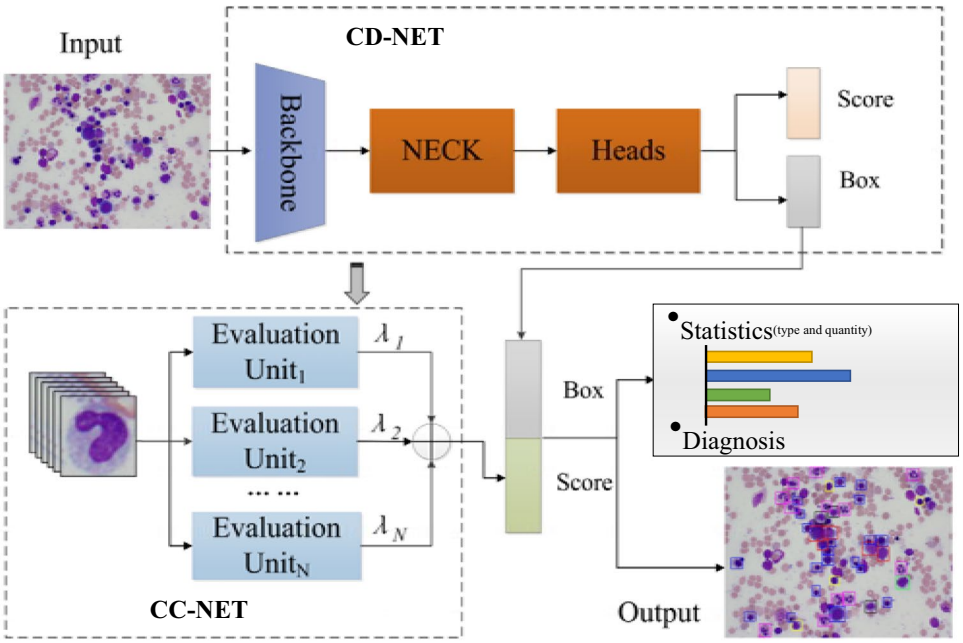


Fig. 3 An overview of the CDC-NET. The CD-NET aims to locate and recognize the nucleated cells in bone marrow aspirate images, while the CC-NET is designed to confirm the types of the detected ambiguous cells. ⊕ denotes the combination of the evaluated units

(CD-NET) and the cell confirmation network (CC-NET). The cell detection network is realized by a pre-trained detection model. The cell confirmation network is designed to realize an assessment and voting mechanism, in which multiple trained classifiers serve as the pathologists to score the cells and to vote for the cell types. Reference information for pathological analysis and diagnosis is generated in the form of text and images.

3.1 Cell detection network

The common detection model consists of three parts, i.e., the backbone pre-trained on the ImageNet, the head used to predict classes bounding boxes of objects, and the neck lying between the backbone and the head. The CD-NET is designed on the base of YOlO and there are some improvements. First, we analyzed the distribution of the size of boxes containing various intact cells, and designed anchors according to it. The nine anchor box sizes generated by clustering are (47, 36), (54, 49), (56, 62), (64, 56), (71, 64), (79, 79), (85, 60), (100, 87), (119, 100). Moreover, for the purpose of trying its best to detect all nucleated cells in the bone marrow aspirate images, the CD-NET pays more attention to the localization task rather than classification. Because the cytoplasmic boundary of the nucleated cells is often unclear, we incorporate the uIoU-based evaluation metrics [33] into the detection model. United Intersection over Union (uIoU) keeps the main properties of IoU [34], and it is defined to correct the ambiguous labels (caused by the unclear cytoplasmic boundary) by keeping the predicted bounding boxes containing relationships in the positive sample as much as possible.

3.2 Cell confirmation network

The CC-NET is designed to confirm the types of nucleated cells, as all the pathologists will do. Since the pathological data with high-quality manual annotation is expensive and limited, it is difficult to train an accurate classification model. The ensemble learning method is employed to design the CC-NET, which dynamically fuses multiple classifiers, to reduce the variance and the bias to improve the accuracy of prediction. The architecture of CC-NET is shown in Fig. 4.

Multiple evaluation units are treated as pathologists or technicians to simulate the clinical diagnosis process. More specifically, the original training set is divided for cross-validation, and each classifier is trained and assessed on different training schemes. After the trained evaluation units pass the examination (human examination and metric assessing), the weight of each evaluation unit is measured according to the examination results to form a complete assessment network. The design principles are as follows:

- Each classifier has independent decision-making ability. The training set is selected from the original dataset but with replacement, and there are differences among the training sets.
- Each evaluation unit approximates the optimal global solution. Different classifiers are obtained by using different objective functions and different optimization methods.
- The weights of the evaluation units are calculated dynamically. The weight of each evaluation unit is adjusted automatically during the training process until the optimal weight is achieved.

3.2.1 Assessing mechanism

In the CC-NET, we design an assessment mechanism for individual classifiers to assess the weakly supervised classifiers to obtain a more comprehensive model by integrating the mature classifiers. In order to facilitate the downstream tasks, the classifier is trained on the dataset with 8 types of cells. When training each classifier, the training sets are randomly sampled to reduce the relationship among the training

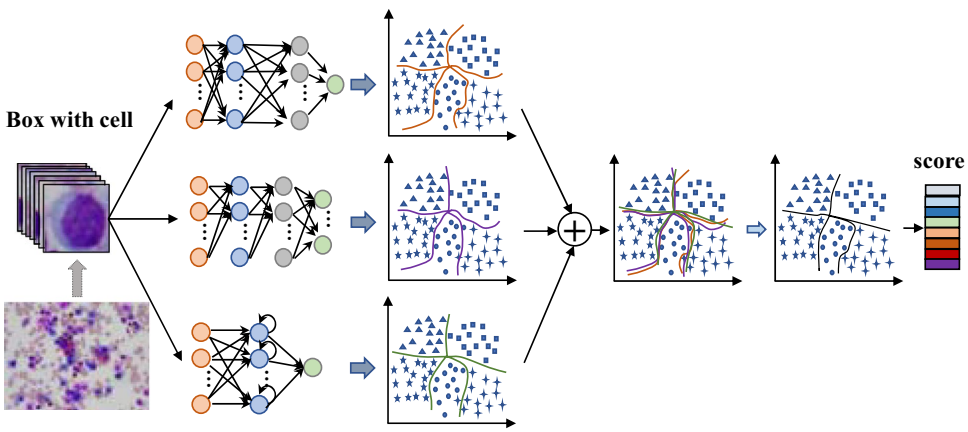


Fig. 4 The architecture of CC-Net. Multiple trained classifiers are employed to confirm the types of the detected cells. ⊕ denotes the combination of the evaluated units

sets as much as possible. Two crucial aspects of CC-NET are shown as follows.

Training the classifier For the training data set $D_{train} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$,$x_i$ is an image with a single cell, and y_i is the corresponding category of the cell image x_i. For each classifier, we randomly selected 90% samples from D_{train} as the training set and the remaining 10% samples as the validation set. The trained classifier is put into the classifier set $D_C = \{c_1, c_2, \cdots, c_k\}$.
Reliability calculation *RL* is the reliability, which is used to evaluate the dynamic performance of the trained classifier. The test set is $D_{test} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ and it does not overlap with the D_{train}. For the trained classifier $c_j \in D_c$, the Precision and Recall are calculated and the Precision-Recall curve [35] is plotted. The reliability of the predicted sample RL_j is calculated according to the Precision Recall Curve. The classifier that fails to pass the assessment is deleted from the classifier set D_C. The specific calculation formula is as

$$RL_j = \frac{1}{11} \times \sum_{i}^{11} \int_0^1 p_{smooth}(r_i) dr_i, \tag{1}$$

where r_i is the Recall value of point i *and* $i \in \{0, 0.1, 0.2, \cdots, 0.9, 1\}$, $j \in \{0, 1, 2, \cdots, k\}$, P_{smooth} is the lower area of the Precision-Recall curve between 0 and 1.

3.2.2 Voting mechanism

A weighted voting mechanism based on a deep neural network is employed to improve the accuracy and diversity of the model, which takes advantage of multiple evaluation units. Different from the previous multi-classifier methods, the weight of each classifier in CC-NET is computed dynamically. The initial weights of the classifiers are obtained by the experience values. The calculation formula of voting prediction is defined as

$$H(x) = \sum_{i=1}^{N} \lambda_i h_i(x), \tag{2}$$

where $H(x)$ is the final score of the single cell image x, $h_i(x)$ is the predicted score of the *i-th* evaluation unit for the single cell image x, λ_i is the weight of the evaluation unit h_i and each weight satisfies the relation

$$\sum_{i=1}^{N} \lambda_i = 1. \tag{3}$$

We keep updating the model parameters along the opposite direction of the first derivative of the loss function $Loss(\theta)$ until $Loss(\theta)$ reaches the minimum point, where parameter $\theta \in R$. By this method, the model converges quickly. The loss function is defined as

$$Loss(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(H_\theta(x_i) - y_i\right)^2, \tag{4}$$

where m is the number of samples. When we set the updating step size to α, the weight updating function formula is defined as

$$\lambda_j = \lambda_j - \alpha \times \frac{1}{m} \sum_{i=1}^{m} \left(H_\theta(x_i) - y_i\right) \times x_{ij}, \tag{5}$$

The pseudo-code of weight calculation in the voting mechanism is shown in Algorithm 1.

3.3 Evaluation metrics

In clinical application, the task of the proposed aided analysis and diagnostic model is to identify and count all types of nucleated cells in the bone marrow aspirate images and calculate the percentage of blasts quantitatively. To better evaluate the model from the perspective of clinical application, we introduce several more realistic evaluation indexes.

3.3.1 Evaluation metrics of CD-NET

Missed Rate (MR) is used to measure the performance of CD-NET. The MR is defined as:

$$MR = 1 - \frac{TP_B}{TP_B + FN_B}, \tag{6}$$

where TP_B is the True Positive and is defined as the number of the detected cells among the labeled cells. FN_B is the False Negative and is defined as the number of cells not detected by the cell detector among the labeled cells.

3.3.2 Evaluation metrics of CC-NET

Since there are cells with ambiguous types, pathologists will need to achieve consensus on the type of cell by voting. We simulate this process by the CC-NET to confirm the type of the detected cells. Precision and Recall are used as the measurements of the performance of the CC-NET, which are defined respectively by

$$Recall = \frac{TP_C}{TP_C + FN_C}, \tag{7}$$

and

$$Precision = \frac{TP_C}{TP_C + FP_C}, \tag{8}$$

Algorithm 1 Multi-classifier voting mechanism

Training Process

Training set: $D_{train} = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$;

Number of classifiers: C;

Input: Train epochs: T;

Classifier prediction function: $h(x)$;

The initial weights: λ_0.

For *t=1,2, ..., T*:

For *j=1,2, ..., m:*

Calculate the score of the cell image x by $H(x) \leftarrow \sum_{i=1}^{C} \lambda_i h_i(x) + b_t$.

Update the weight of the evaluation unit by $\lambda_t \leftarrow \lambda_t - \alpha \times \frac{1}{m} \sum_{i=1}^{m} (H_\theta(x_i) - y_i) \times x_{ij}$.

Update the parameter b_t by $b_t \leftarrow b_t - \alpha \times \frac{1}{m} \sum_{i=1}^{m} (H_\theta(x_i) - y_i) \times x_{ij}$.

End

End

Output: Weight λ_t

where TP_C is the True Positive and is defined as the number of blast cells being correctly classified. FP_C is the False Positive and is defined as the number of cells that are misclassified as blast cells. FN_C is the False Negative and is defined as the number of blasts that are misclassified.

F1-Measure ($F1$) is introduced to comprehensively represent Precision and Recall, which is defined as

$$F_1 = \frac{2 \times precision \times \mathrm{Re}call}{precision + \mathrm{Re}call}. \tag{9}$$

4 Experiment

4.1 The experimental dataset

Some image examples from the dataset are shown in Fig. 5.

In order to facilitate the research of automated diagnosis of hematological disorders, we have constructed a high-quality Bone Marrow Aspirate Smear Image Dataset (BMASID), which contains 230 bone marrow aspirate images, all with corresponding images manually labeled by

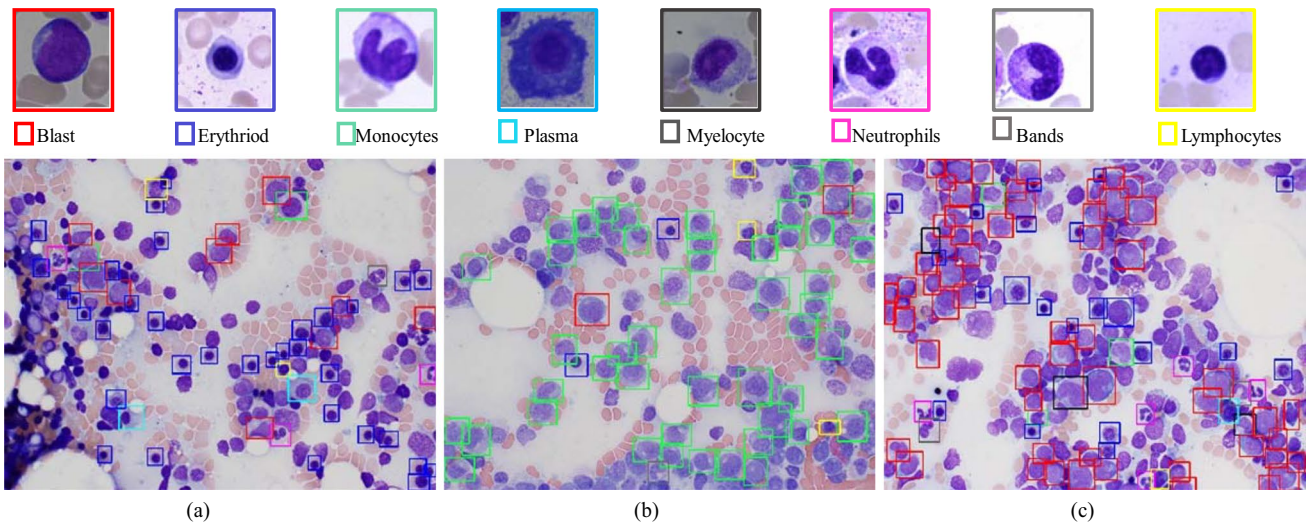


Fig. 5 Examples from the clinical micro-image dataset. The nucleated cells in the bone marrow aspirate images are classified to 8 types according to the aided analysis and diagnosis tasks. **a**, **b** and **c** show the samples of images full of the dense cells and adherent cells, along with garbage and necrotic cells. **b** shows a sample of poorly stained image

hematopathologists [33]. All images were microscopically imaged using Olympus-DP21, UPLanFLN-40X/0.75∞/0.17/FN26.5 equipment. The exposure time is 0.05 ms, the object distance is 17 mm, the focal length is 35 mm, and the resolution is 1200*1600 pixels. There are unavoidable problems in the bone marrow aspirate images, i.e., cell damage and uneven staining. Therefore, it is very difficult to accurately identify all kinds of bone marrow nucleated cells. In each bone marrow smear image, we labeled as many as possible of the eight types of nucleated cells (Fig. 5) that are used for the diagnosis of hematological malignancies (blasts, erythroid cells, monocytes, plasma cells, myelocytes, neutrophils, lymphocytes, and bands) according to the standard and experiences of pathologists.

- We increased the difficulty of the training dataset in terms of staining differences and image complexity. The number of images in the training set also increased from 230 to 270.
- We increased the difficulty of the test set. Images in the test set have different staining qualities and are taken with different equipment. There are also a variety of noise, overlapping cells, adherent cells, cells with blurred borders, garbage cells, and cells with ambiguous types in the images. The garbage cells, such as damaged cells, red blood cells, dead cells, etc., were not evaluated or labeled (see Fig. 5).

4.2 Experimental results

4.2.1 Experiments and performance of CD-NET

We conducted experiments on the extended BMASID to demonstrate the effectiveness of the proposed approach. To train the CD-NET, the batch size was set to 8 and the initial learning rate was 1e-4. The weight obtained on the training set by 1000 iterations was used as the pre-training model to initiate the model, and the model was fine-tuned by setting the maximum number of iterations to 800. We used the validation set to evaluate the performance of each network. Since the Missed Rate is essential for the effectiveness of the statistical results, we compared the CD-NET with three previously reported popular detection methods (i.e., Faster RCNN, SSD, YOLOV4). They are trained on the same dataset. The Missed Rate is shown in Fig. 6.

In particular, we employed uIoU and IoU as the measurement index in the methods respectively. The results in Fig. 6 show that our approach (CDN + uIoU) achieves more satisfactory results in Missed Rate when compared with the previously reported detection methods. The methods with uIoU are more stable than those with IoU. With the increase of the threshold value, the advantage of uIoU became more obvious. As can be seen, it keeps the ideal results when the threshold value of uIoU is less than 0.6.

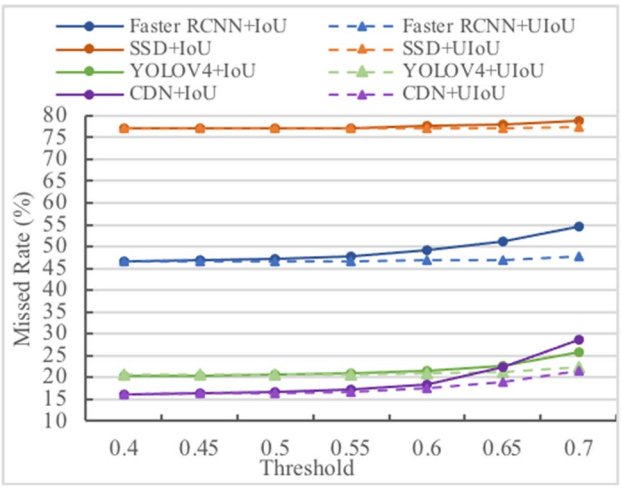


Fig. 6 The comparison of the Missed Rates of the cell detection methods. The threshold of the IoU and uIoU is set to 0.4 to 0.7

4.2.2 Experiments and performance of CC-NET

We designed the CC-NET with better scalability and update capability. Evaluation units of the CC-NET can be dynamically improved and adjusted with the requirements. This paper trained and evaluated the classic classifiers, i.e. AlexNet [19], ResNet-50 [20], and Inception V3 [23], to find suitable components for bone marrow cell type evaluation networks. In the experiments, AlexNet, ResNet-50, and Inception V3 were used as the baseline of each evaluation unit of the CC-NET, respectively. The batch size was set to 16 and the initial learning rate was 1e-3 for each evaluation unit. The evaluation unit was fine-tuned by setting the maximum number of iterations to 800. The weight of the voting mechanism was the best weight selected from 10 K training with a step size of 5e-5. We conducted ten times of tests and used average precision, recall, and F1 to report the performance of the evaluation units. Notably, the main purpose of this experiment is to compare the performances of different combinations of evaluation units, instead of improving the classification performance by specifying feature selection. Because this paper takes the assisted diagnosis of AML as a case, the eported metrics are only related to blast cells. The performance of the cell evaluation network is shown in Table 1.

As shown in Table 1, the classifier with the Inception V3 as the baseline achieves the best Recall and F1, i.e., Recall is 87.734% and the F1 is 86.935%, when compared with the other two classifiers. The Recall is the essential measurement index in this experiment because we focus more on the maximal number of correctly recognized cells, which is

Table 1 The performance comparison of different fusions of evaluation units

Different fusions of evaluation units			Recall (%)	Precision (%)	F1(%)
AlexNet [19]	Inception V3 [23]	ResNet-50 [20]			
√			74.747 ± 0.28	88.889 ± 0.21	81.207 ± 0.25
	√		87.374 ± 0.44	86.500 ± 0.32	86.935 ± 0.38
		√	83.333 ± 0.39	89.918 ± 0.28	86.500 ± 0.34
√	√		87.626 ± 0.37	87.406 ± 0.31	87.516 ± 0.34
√		√	83.838 ± 0.41	91.209 ± 0.42	87.368 ± 0.42
	√	√	91.162 ± 0.50	89.136 ± 0.53	90.137 ± 0.52
√	√	√	88.889 ± 0.48	90.488 ± 0.51	89.681 ± 0.49

proven to minimize the possibility of AML misdiagnosis. When fusing two classifiers (i.e., ResNet-50 + Inception V3), the Recall and F1 are 91.162% and 90.137%, respectively. The Recall and F1 are improved obviously.

4.2.3 Experiments and performance of CDC-NET

To further verify the effectiveness of each evaluation unit, we fused the ResNet-50 and Inception V3 to form CC-NET-I, and fused Inception V3, AlexNet, and ResNet-50 to form CC-NET-II. CC-NET-I and CC-NET-II were cascaded with Faster RCNN, YOLOV4, and CD-NET respectively. The results of ten times tests are shown in Table 2.

From Table 2, we can see that the precision value obtained by YOLOV4 with CC-NET-I is 93.515% and the Recall value is 69.192%. YOLOV4 has demonstrated noticeable improvements in the accuracy of object localization and classification compared to its previous versions. However, YOLOV4 has limitations when it comes to detecting densely packed small objects. This is due to the utilization of low-resolution feature maps in the object detection process of YOLOV4, which may result in missed detections or false positives for particularly small and densely populated targets. The Recall and F1 are improved when we employ CD-NET. When we cascade CD-NET and CC-NET-II, the comprehensive performance is further improved, i.e., Recall is 78.535%, Precision is 91.74% and the F1 is 84.625%. It indicates that the model performs well while considering both recall and precision, which is of great significance for imbalanced datasets or tasks that require attention to both precision and recall. When cascading with CD-NET, the recall of the CC-NET-I (i.e., the recall of the fusion of ResNet-50 and Inception V3) is reduced by 0.252%, which is caused by the cell detecting error (i.e., the detecting error caused by some missed cells). According to the requirements of nucleated cell detection in bone marrow aspirate images, CC-NET needs to improve the Precision with a high Recall. Therefore, CC-NET is the result of the fusion of three evaluation units, i.e., AlexNet, ResNet-50, and Inception V3.

Due to the fact that the Precision-Recall curve focuses more on TP and can effectively evaluate the classifiers trained on imbalanced data, we plotted the Precision-Recall curves to compare the models in Table 2 with better performance, as shown in Fig. 7. The AUC of the model (YOLOV4 with CC-NET-II) is about 0.9084, while the AUC of the model (CD-NET with CC-NET-II) is about 0.9281.

Table 2 The performance comparison of different fusions of evaluation units when cascading with different detection methods

Method	Different combinations of evaluation units			Recall (%)	Precision (%)	F1 (%)
	NO-CC-NET	CC-NET-I	CC-NET-II			
Faster RCNN	√			51.768 ± 0.28	81.673 ± 0.31	63.370 ± 0.30
		√		53.283 ± 0.36	87.190 ± 0.34	66.144 ± 0.38
			√	52.020 ± 0.32	89.177 ± 0.31	65.709 ± 0.34
YOLOv4	√			70.960 ± 0.43	89.206 ± 0.42	79.044 ± 0.43
		√		69.192 ± 0.38	93.515 ± 0.36	79.535 ± 0.38
			√	69.949 ± 0.40	92.333 ± 0.39	79.597 ± 0.40
CD-NET	√			77.020 ± 0.31	87.896 ± 0.26	82.099 ± 0.29
		√		76.768 ± 0.39	91.291 ± 0.29	83.402 ± 0.35
			√	78.535 ± 0.42	91.740 ± 0.31	84.625 ± 0.38

CC-NET-I is the fusion of ResNet-50 and Inception V3. CC-NET- II is the fusion of AlexNet, Inception V3 and ResNet-50. No-CC-NET means there is no evaluation network

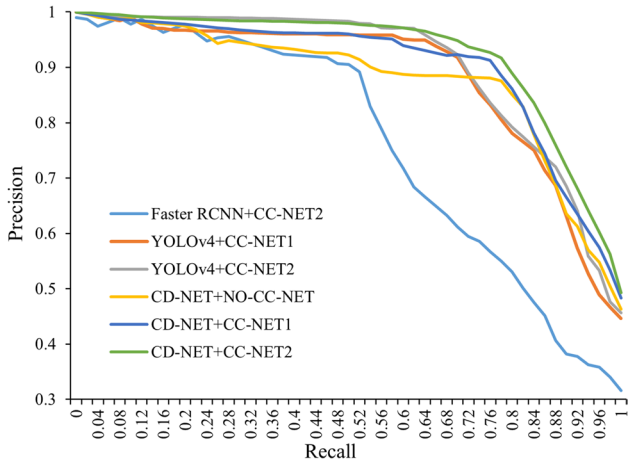


Fig. 7 The Precision-Recall curves of the models in Table 2 with better performance

Table 3 Comparison of the diagnostic results. IoU threshold is 0.4

Method	Clinical-I(100) / Clinical-II(50)			
	Faster-RCNN	YOLOV4	CD-NET	Manual film-reading
Non-CC-NET	25/13	25/12	26/13	26/13
CC-NET	26/13	26/13	26/13	

4.2.4 Experimental results in the aided diagnosis of AML

To demonstrate the effectiveness of CC-NET in clinical applications, we extracted clinical data during different periods. Clinical-I is the bone marrow aspirate image data from 100 patients during the first period and clinical-II is the bone marrow aspirate image data from 50 patients during the second period. We verified the CDC-NET on Clinical-I and clinical-II, by comparing the diagnostic results of different methods. Confirmed by the pathologist on bone marrow aspirate images, Clinical-I consists of 26 patients with AML and Clinical-II consists of 13 patients with AML. The experimental results are shown in Table 3.

From Table 3, we can see that when employing CC-NET, Faster-RCNN, YOLOV4, and CD-NET correctly identified all AML patients on the base of the bone marrow aspirate images. However, there is an identification error if we just use the Faster RCNN or YOLOV4 as the diagnosis model because some misclassified nucleated cells result in the error of differential counting.

4.2.5 Visualizing the detection results

In order to display and compare the experimental results of different models for the same group of bone marrow aspirate images, boxes in different colors were used to label the correctly detected nucleated cells, the missed nucleated cells, the misclassified nucleated cells, and the additionally detected nucleated cells. The experimental results are shown in Fig. 8.

Incorrectly located or missed cells are boxed in orange and the correctly located but misclassified cells are boxed in red. As can be seen, CD-NET achieves more than 90% locating effectiveness. The missed rate is lower than that of YOLOV4, as can be easily seen in the areas with dense cells. When employing CC-NET, some misclassified nucleated cells have been corrected, and the accuracy of cell identification is improved. In addition, it is interesting to notice that CDC-NET can identify the nucleated cells that the pathologists overlooked but are meaningful for the diagnosis (see the cells boxed in blue in Fig. 8).

5 Discussion

To promote the development of deep learning-based auxiliary diagnosis methods for AML, we proposed a model named CDC-NET. Since an effective deep learning model significantly depends on large and high-quality annotated datasets and constructing a high-quality standard dataset of medical images is not easy, it is difficult to improve the performance of the auxiliary diagnosis model on a limited dataset. The proposed CD-NET aims to train a cell detection model with a low missed rate. As seen from Fig. 6, CD-NET achieves more satisfactory performance than previously reported detection methods in terms of missed rate. Compared with the IoU measure index, uIoU seems to be more effective in our task in areas full of cells with ambiguous boundaries in bone marrow aspirate images. We proposed CC-NET to produce a strong supervision model. It fuses multiple evaluation units. Table 1 shows the effectiveness of CC-NET. When cascading with CD-NET, the Recall of CDC-NET is higher than that of CD-NET (see Table 3.). The experiments on 50 clinical data (bone marrow aspirate images from 50 patients) demonstrate that our approach achieves better performance.

Our previous research introduced a Cell Recognition Network (CRNet) utilizing YOLOV3 and uIoU for cell recognition. However, when confronted with more complex data, specifically with the addition of 40 complex bone marrow images to the dataset, the recognition error increased (as depicted in CASEII-C in Fig. 8). To address this challenge, we propose a voting strategy-based method to mitigate recognition errors, building upon the CRNet research (as illustrated in CASEII-D in Fig. 8). This approach aligns with the pathologists' methodology in the pathological diagnosis of Acute Myeloid Leukemia (AML), as it assists in identifying problematic types of nucleated cells.

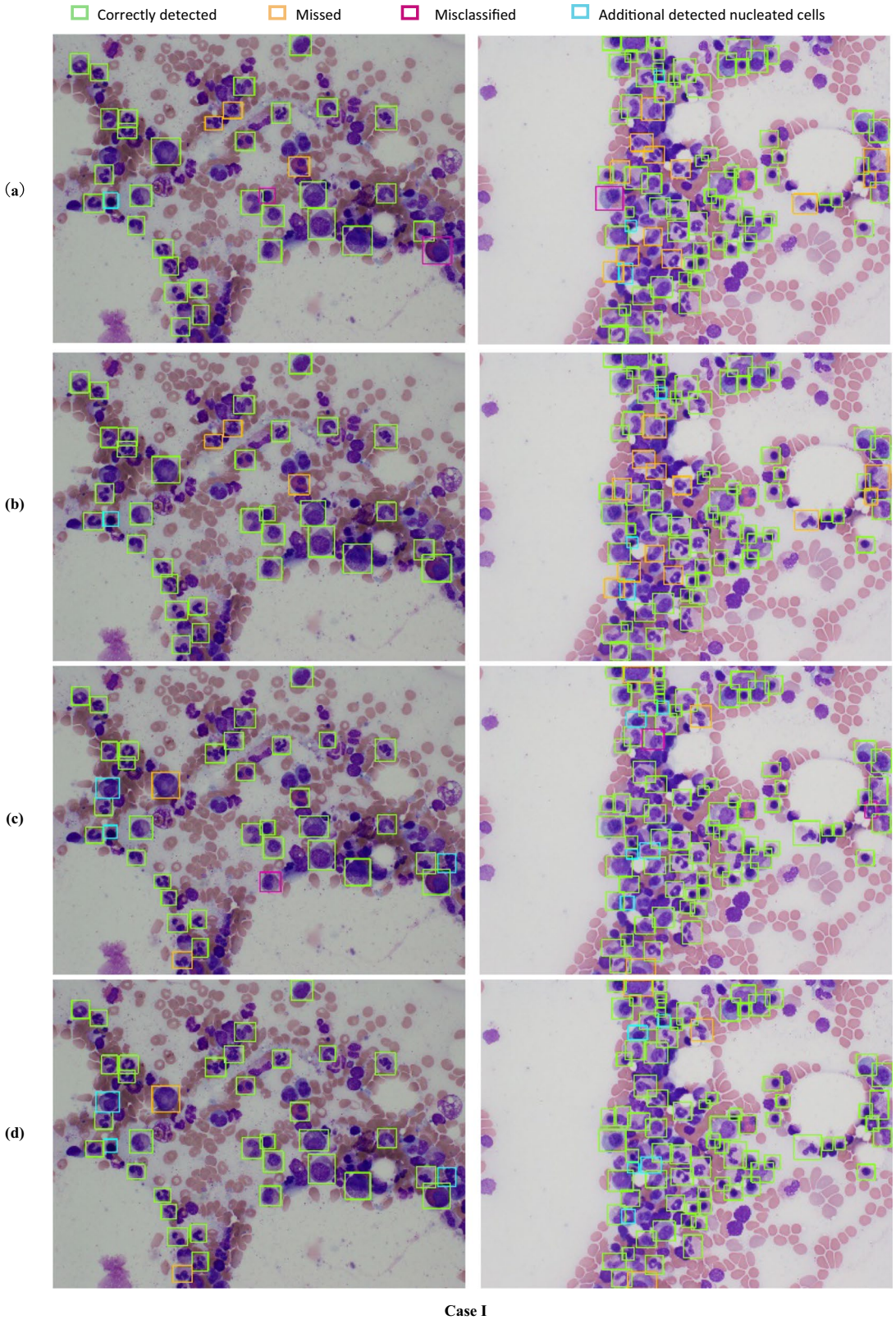


Fig. 8 Comparison of the methods. (**a**) the detection results by YOLOV4, (**b**) the detection results by YOLOV4 + CC-NET. (**c**) the detection results by CD-NET. (**d**) the detection results by CDC-NET

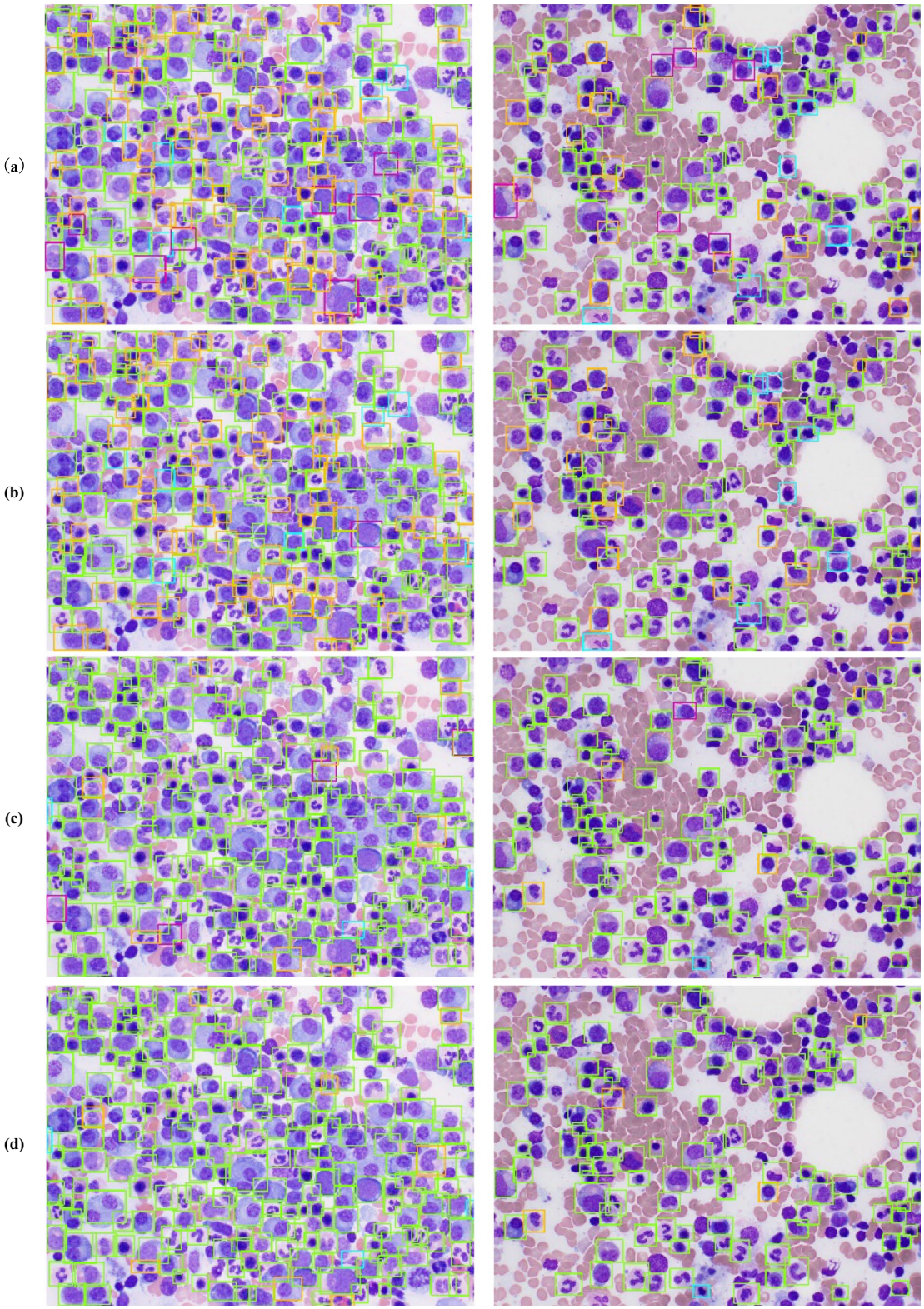


Fig. 8 (continued)

In bone marrow aspirate images, there are inherent challenges such as uneven staining, cell damage, overlapping or adherent cells, cells with unclear borders, and cells with ambiguous types. These factors contribute to the difficulty in accurately identifying various types of bone marrow nucleated cells. To facilitate the expandability of the model, this approach incorporates an average voting mechanism and focuses on experiments using widely used classifiers like AlexNet, ResNet-50, and Inception V3. However, it is crucial to conduct additional experiments in the future to explore further possibilities. Furthermore, to enhance the accuracy of digital auxiliary diagnosis, it is necessary to improve the quality of staining. This improvement in staining quality would contribute to more precise results in the digital diagnostic process.

In the future, our research will extend to studying the auxiliary diagnosis of other types of blood diseases, such as myelodysplastic syndromes (MDS) and acute lymphoblastic leukemia (ALL). This expansion will enable us to explore the potential application of our method in diagnosing a wider range of blood disorders.

6 Conclusion

This paper provides an analytical model based on the deep learning method for the auxiliary diagnosis of AML. We focus on crucial technology i.e., nucleated cell detection and cell differential counting. Different from the existing methods, we design a cell detection network (CD-NET) and an ensemble learning-based confirmation network (CC-NET) to improve the accuracy of the detection and classification of nucleated cells in the bone marrow aspirate images. In particular, CC-NET fuses multiple classifiers with independent decision-making capability and is trained dynamically, which aims to improve robustness and scalability. Experiments were conducted on a clinic dataset, which consists of 230 bone marrow aspirate images. When we took uIoU as the metrics, the cell detection network (CD-Net) reduced the missing rate of cells significantly. When we cascaded the CC-NET with CD-Net, the diagnosis accuracy on the bone marrow aspirate images was notably improved. Experimental results demonstrated that our approach outperforms other deep learning methods and achieved 91.74% precision, which demonstrates that our approach has the potential to be an important component of the auxiliary diagnosis system for AML.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (No. 52001039), National Natural Science Foundation of China under Grand (No.52171310), Shandong Natural Science Foundation in China (No. ZR2019LZH005), Research fund from Science and Technology on Underwater Vehicle Technology Laboratory (No.2021JCJQ-SYSJJ-LB06903). University Innovation Team Project of Jinan (No.2019GXRC015). Science and technology improvement project for small and medium-sized enterprises in Shandong Province (No. 2021TSGC1012).

Data availability The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Conflicts of interest The authors declare that they have no conflicts of interest.

References

1. Bruneau J, Molina TJ (2020) WHO Classification of tumors of hematopoietic and lymphoid tissues. In: Molina TJ (eds) Hematopathology. Encyclopedia of Pathology. Springer, Cham, pp 501–505. https://doi.org/10.1007/978-3-319-95309-0_3817
2. Bonnet D, Dick JE (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 3(7):730–737
3. Porcu S, Loddo A, Putzu L, Di Ruberto C (2018) White blood cells counting via vector field convolution nuclei segmentation. In: Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (4):227–234
4. Das DK, Maiti AK, Chakraborty C (2018) Automated identification of normoblast cell from human peripheral blood smear images. J Microsc 269(3):310–320
5. Ghribi O, Maalej A, Sellami L, Slima MB, Maalej MA, Mahfoudh KB, Dammak M, Mhiri C, Hamida AB (2019) Advanced methodology for multiple sclerosis lesion exploring: Towards a computer aided diagnosis system. Biomed Signal Process Control 49:274–288
6. Matek C, Schwarz S, Spiekermann K, Marr C (2019) Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nat Mach Intell 1(11):538–544
7. El Alaoui Y, Elomri A, Qaraqe M, Padmanabhan R, YasinTaha R, El Omri H, El Omri A, Aboumarzouk O (2022) A review of artificial intelligence applications in hematology management: current practices and future prospects. J Med Internet Res 24(7):e36490
8. Raina R, Gondhi NK, Chaahat R, Singh D, Kaur M, Lee HN (2023) A systematic review on acute leukemia detection using deep learning techniques. Arch Comput Methods Eng 30(1):251–270
9. Saleem S, Amin J, Sharif M, Mallah GA, Kadry S, Gandomi AH (2022) Leukemia segmentation and classification: A comprehensive survey. Comput Biol Med 150:106028
10. Akram N, Adnan S, Asif M, Imran SMA, Yasir MN, Naqvi RA, Hussain D (2022) Exploiting the multiscale information fusion capabilities for aiding the leukemia diagnosis through white blood cells segmentation. IEEE Access 10:48747–48760
11. Mustaqim T, Fatichah C, Suciati N (2023) Deep learning for the detection of acute lymphoblastic leukemia subtypes on microscopic images: A systematic literature review. IEEE Access (11):16108–16127. https://doi.org/10.1109/ACCESS.2023.3245128
12. Li N, Fan L, Xu H, Zhang X, Bai Z, Li M, Xiong S, Jiang L, Yang J, Chen S, Qiao Y (2023) An AI-Aided diagnostic framework for hematologic neoplasms based on morphologic features and medical expertise. Lab Invest 103(4):100055
13. Goutam D, Sailaja S (2015) Classification of acute myelogenous leukemia in blood microscopic images using supervised classifier.

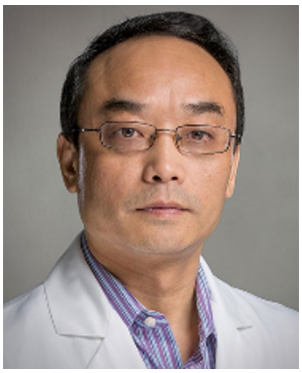
In: 2015 IEEE International Conference on Engineering and Technology (ICETECH). IEEE, pp 1–5
14 Li Y, Zhu R, Mi L, Cao Y, Yao D (2016) Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method. Comput Math Methods Med 2016:9514707
15. Aris TA, Nasir AA, Mustafa WA (2018) Analysis of distance transforms for watershed segmentation on chronic leukaemia images. J Telecommun Electron Comput Eng (JTEC) 10(1–16):51–56
16. Song TH, Sanchez V, Daly HE, Rajpoot NM (2018) Simultaneous cell detection and classification in bone marrow histology images. IEEE J Biomed Health Inform 23(4):1469–1476
17. Yang S, Liu X, Zheng Z, Wang W, Ma X (2021) Fusing medical image features and clinical features with deep learning for computer-aided diagnosis. arXiv preprint arXiv:2103.05855, 2021
18. Fan H, Zhang F, Xi L, Li Z, Liu G, Xu Y (2019) LeukocyteMask: An automated localization and segmentation method for leukocyte in blood smear images using deep neural networks. J Biophotonics 12(7):e201800488
19. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
20. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
21. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
22. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, pp 448–456. pmlr
23. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
24. Yu W, Chang J, Yang C, Zhang L, Shen H, Xia Y, Sha J (2017) Automatic classification of leukocytes using deep neural network. In: 2017 IEEE 12th international conference on ASIC (ASICON). IEEE, pp 1041–1044
25. Suriya M, Chandran V, Sumithra MG (2022) Enhanced deep convolutional neural network for malarial parasite classification. Int J Comput Appl 44(12):1113–1122
26. Zerouaoui H, Idri A, Nakach FZ, Hadri RE (2021) Breast fine needle cytological classification using deep hybrid architectures. In: Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part II 21. Springer International Publishing, pp 186–202
27. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(06):1137–1149
28. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, pp 21–37
29. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
30. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7263–7271
31. Redmon J, Farhadi A (2018) YOLOv3: An incremental improvement. https://doi.org/10.48550/arXiv.1804.02767
32. Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: Optimal speed and accuracy of object detection. https://doi.org/10.48550/arXiv.2004.10934
33. Su J, Han J, Song J (2021) A benchmark bone marrow aspirate smear dataset and a multi-scale cell detection model for the diagnosis of hematological disorders. Comput Med Imaging Graph 90:101912
34. Rahman, MA, Wang, Y (2016) Optimizing intersection-over-union in deep neural networks for image segmentation. In: Bebis G, et al. Advances in Visual Computing. ISVC 2016. Lecture Notes in Computer Science, vol 10072. Springer, Cham. https://doi.org/10.1007/978-3-319-50835-1_22
35. Raghavan V, Bollmann P, Jung GS (1989) A critical investigation of recall and precision as measures of retrieval system performance. ACM Trans Inf Syst (TOIS) 7(3):205–229

Jie Su received the Ph. D. degree in Computer Science and Technology from Harbin Engineering University in 2010. From 2014 to 2015, she was working with the University of South Florida as a visiting scholar. Currently, she is an Associate Professor at the School of Information Science and Engineering, University of Jinan, China. She is also a fellow of the Shandong Provincial Key Laboratory of Network-Based Intelligent Computing. Her research interests include computer vision, assistive robotics, and medical image analysis.

Yahui Liu received the M. S. degree in Computer Applied Technology from the Harbin University of Science and Technology, Harbin, China, in 2004, and the Ph. D. degree in Mechatronic Engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011. She is currently an Associate Professor with the School of Information Management. She has been involved in research on target detection, image segmentation, and time series forecasting.

Jing Zhang received the B.Eng., M.Eng., and Ph.D. degrees in computer science and technology from Harbin Engineering University in 1987, 1998, and 2005, respectively. From 2005 to 2007, she was doing her postdoctoral research in the area of image processing with Harbin Institute of Technology, Harbin. From 2003 to 2004, she was working with the Melbourne University in Australia as a visiting scholar. From 2015 to 2016, she was doing research at University of York, in UK as an academic visitor followed the guidance of Prof. Edwin Hancock. Currently she is a professor in University of Jinan. She has published more than 100 papers in journals, edited books, and refereed conferences. Her research interests are image processing and virtual reality.

Jinjun Han received the M.S. degree from the University of Jinan in 2022. She was focusing on the research of pathological image analysis algorithms and intelligent diagnosis system development.



Jinming Song is board certified in anatomic pathology, surgical pathology, and hematopathology. His specialty is the diagnosis of acute and chronic leukemia, lymphoma, and other benign or malignant hematopoietic disorders. His expertise includes evaluation and interpretation of peripheral blood smears, bone marrow biopsies, lymphoid and non-lymphoid tissues, flow cytometry, molecular pathology especially NextGen sequencing, and cytogenetics. Dr. Song's research interest is personized medicine, the use of gene mutations and cytogenetic abnormalities for the diagnosis, and the prediction of prognosis and treatment responses in patients with hematopoietic disorders.