



## Deep learning for bone marrow cell detection and classification on whole-slide images



Ching-Wei Wang<sup>a,b,1,\*</sup>, Sheng-Chuan Huang<sup>c,d,e,1</sup>, Yu-Ching Lee<sup>b</sup>, Yu-Jie Shen<sup>a</sup>, Shwu-Ing Meng<sup>c</sup>, Jeff L. Gaol<sup>a</sup>

<sup>a</sup> Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taipei, 106, Taiwan

<sup>b</sup> Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taipei, 106, Taiwan

<sup>c</sup> Department of Laboratory Medicine, National Taiwan University Hospital, Taipei, 100, Taiwan

<sup>d</sup> Department of Hematology and Oncology, Hualien Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Hualien, Taiwan

<sup>e</sup> Department of Clinical Pathology, Hualien Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Hualien, Taiwan

### ARTICLE INFO

#### Article history:

Received 21 August 2020

Revised 6 October 2021

Accepted 13 October 2021

Available online 16 October 2021

#### 2020 MSC:

68T07

92C50

#### Keywords:

Hematopathology

Whole-slide image

Bone marrow differential cell count

Deep learning

### ABSTRACT

Bone marrow (BM) examination is an essential step in both diagnosing and managing numerous hematologic disorders. BM nucleated differential count (NDC) analysis, as part of BM examination, holds the most fundamental and crucial information. However, there are many challenges to perform automated BM NDC analysis on whole-slide images (WSIs), including large dimensions of data to process, complicated cell types with subtle differences. To the authors best knowledge, this is the first study on fully automatic BM NDC using WSIs with 40x objective magnification, which can replace traditional manual counting relying on light microscopy via oil-immersion 100x objective lens with a total 1000x magnification. In this study, we develop an efficient and fully automatic hierarchical deep learning framework for BM NDC WSI analysis in seconds. The proposed hierarchical framework consists of (1) a deep learning model for rapid localization of BM particles and cellular traits generating regions of interest (ROI) for further analysis, (2) a patch-based deep learning model for cell identification of 16 cell types, including megakaryocytes, mitotic cells, and four stages of erythroblasts which have not been demonstrated in previous studies before, and (3) a fast stitching model for integrating patch-based results and producing final outputs. In evaluation, the proposed method is firstly tested on a dataset with a total of 12,426 annotated cells using cross validation, achieving high recall and accuracy of  $0.905 \pm 0.078$  and  $0.989 \pm 0.006$ , respectively, and taking only 44 seconds to perform BM NDC analysis for a WSI. To further examine the generalizability of our model, we conduct an evaluation on the second independent dataset with a total of 3005 cells, and the results show that the proposed method also obtains high recall and accuracy of 0.842 and 0.988, respectively. In comparison with the existing small-image-based benchmark methods, the proposed method demonstrates superior performance in recall, accuracy and computational time.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

Bone marrow (BM) examination is a crucial step in both diagnosing and treating various hematologic disorders. BM nucleated differential cell count (NDC), as part of the examination, is mandatory to evaluate the hematopoiesis in different cell lineages and to quantify abnormal cells, if present (Lee et al., 2008). In other words, BM NDC is an indispensable key step, which not only generates correct diagnosis but also provides an important clue in prognosis and disease follow-up, especially for hematologic ma-

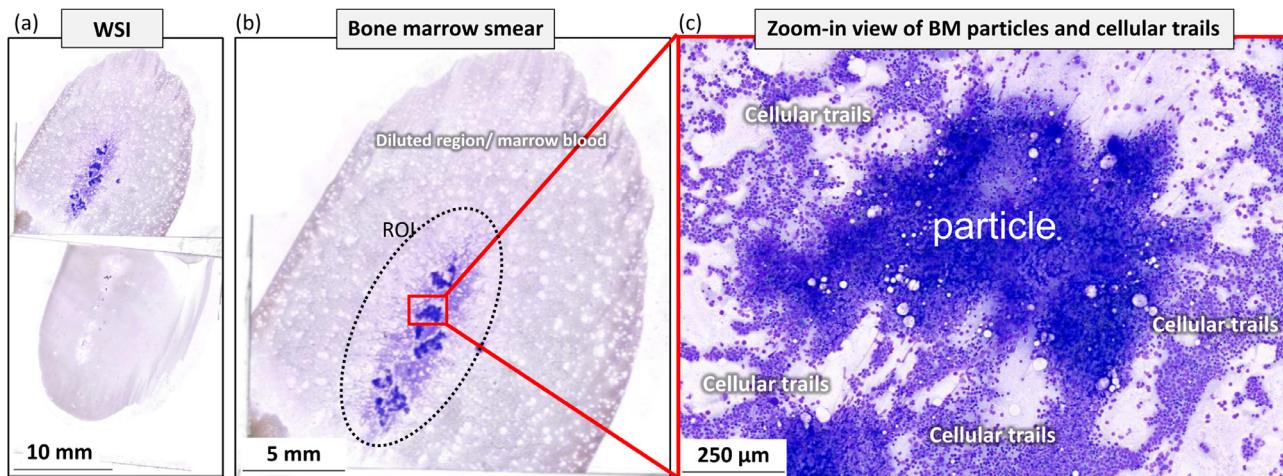
lignancies, such as acute myeloid leukemia (AML) (Döhner et al., 2017), acute lymphoblastic leukemia (ALL) (Swerdlow et al., 2017), myelodysplastic syndrome (MDS) (Greenberg et al., 2012), chronic myeloid leukemia (CML) (Swerdlow et al., 2017), chronic myelomonocytic leukemia (CMML) (Swerdlow et al., 2017), and multiple myeloma (MM) (Kumar et al., 2016). From the French-American-British (FAB) classification (Bennett et al., 1976), proposed in 1976, to World Health Organization (WHO) classification (Swerdlow et al., 2017), updated in 2017, the cytomorphological examination of BM smear remains the gold standard in disease diagnosing and treatment response monitoring.

Fig. 1(a) shows a whole slide image (WSI) from a patient. In routine practice at our hospital, a complete slide consists of a BM smear (upper part) and a peripheral blood smear (lower part),

\* Corresponding author.

E-mail address: [cweiwang@mail.ntust.edu.tw](mailto:cweiwang@mail.ntust.edu.tw) (C.-W. Wang).

<sup>1</sup> These authors are joint first authors.



**Fig. 1.** Illustration of ROI in a WSI. (a) A WSI contains both bone marrow smear in the upper part and peripheral blood smear in the lower part of the same patient. (b) Bone marrow smear with a ROI (denoted by black dotted lines) to mark the region of the BM particles and cellular trails. (c) A zoom-in view of ROI, containing BM particles and cellular trails, in which BM NDC was conducted.

which can offer extra valuable clues in diagnosis. In some circumstances, peripheral blood smear contains more abnormal cells than BM smear, such as certain types of lymphoma, acute megakaryoblastic leukemia, and myelofibrosis (in which BM aspirate often results in so-called "dry tap"). Furthermore, the percentage of abnormal cells in peripheral blood are also included in diagnostic criteria in some diseases. To conduct a complete BM NDC, a trained hematologist, pathologist and/or medical technician firstly locates the BM particles and cellular trails to determine the regions of interest (ROI) under low magnification with 10x objective lens (Fig. 1b and dummyTXdummy-(1c) following the International Council for Standardization in Hematology (ICSH) guidelines (Lee et al., 2008). This initial step is critical because if the counting was performed on diluted regions outside the BM particles and cellular trails, the results can draw to a wrong conclusion and even mislead the treatment direction. After targeting the ROI, examiners then employ a 100x objective lens with oil immersion in order to discriminate diverse cell types. BM NDC analysis is well-known for its labor-intensive, time-consuming, complicated, and poorly reproducible work. The complexity and poor reproducibility emerge from the diversity and delicate intra-lineage difference within the maturation process of BM cells, which is prone to intra- and inter-observer variability (Chandrasevan et al., 2019; Choi et al., 2017; Wilkins et al., 2008; Yu et al., 2019). Finally, according to the ICSH guidelines (Lee et al., 2008), it is required to count at least 1000 nucleated BM cells (500 nucleated cells per slide) for a patient. The overall time spent for a BM NDC on two microscopic slides per patient is at least 30 minutes, even for a well-trained examiner. As a result, an efficient and fully automatic bone marrow whole-slide analysis system is highly demanded to increase the accuracy and speed of the diagnosis, reduce the workload and labor costs and enable reproducibility.

Many efforts have been devoted to developing the automation of blood cell differential counts. Nonetheless, existing modern hematology analyzers fail to show promising results in identifying and classifying blasts, immature granulocytes and basophils (Meintker et al., 2013). While it remains an insurmountable task to build a reliable automated blood cell characterization and quantification system in pathologic samples of peripheral blood, it is undoubtedly that the automated BM NDC poses an even greater challenge due to the operator techniques, quality of smears, staining variation, cells overlapping and complexity, failure to identify nucleated red blood cells and different stages of cell lineages and lipid interference (Fuentes-Arderiu and Dot-Bach, 2009).

Nowadays, pathologists commonly conduct a microscopic investigation on WSIs with the aid of a computer-assisted image analysis system (Bychkov et al., 2018; Campanella et al., 2019; Ianni et al., 2020; Iizuka et al., 2020; Lin et al., 2019; Peikari et al., 2016; Zhang et al., 2019). In digital pathology, glass slides are scanned using a whole-slide scanner generating digitized WSIs for further assessment using the computer-based algorithms. Typically, WSIs are formulated in a multi-resolution pyramid structure from high to low magnification with a size of approximately 250,000 × 500,000 pixels at the highest resolution level. The gigantic size of WSIs makes implementation of an intelligent automatic analysis of BM NDC on WSIs more challenging.

In this study, we present an efficient and fully automatic hierarchical deep learning framework for the BM NDC using WSIs. By exploiting the efficiency of multi-resolution pyramid data structure and Cascade R-CNN, the proposed framework has the capability to identify BM particles and cellular trails and perform BM cell detection-classification inside BM particles to automate BM NDC using WSI analysis in seconds without any examiner-provided annotations. The proposed method has the capability of (1) fast localization of BM particles and cellular trails, (2) BM cell detection-classification and (3) fast stitching model for multi-type patch-based detection. The proposed framework can efficiently differentiate 16 BM cell types, including blast, promyelocyte, myelocyte, metamyelocyte, band neutrophil, segmented neutrophil, eosinophil, monocyte and promonocyte, proerythroblast, basophilic erythroblast, polychromatic erythroblast, orthochromatic erythroblast, lymphocyte, plasma cell, mitotic cell and megakaryocyte. To the authors' best knowledge, this is the first study to perform BM NDC on WSIs with the most abundant cell types, including megakaryocyte, mitotic cell and detailed four stages of erythroblast, which are critical for clinical diagnosis with more information given in Section 5.

The rest of this paper is organized as follows. Section 2 describes related work in BM NDC. Section 3 presents the proposed method. The experimental comparisons against benchmark methods are given in Section 4. Section 5 provides the discussion. Section 6 concludes the paper.

## 2. Related work

### 2.1. Feature extraction and classification

Several studies try to automate BM NDC by using the combinations of image preprocessing, image segmentation, fea-

ture extraction and conventional machine learning methods. [Reta et al. \(2015\)](#) perform image segmentation based on contextual color and textual information using Iterated Conditional Model algorithm to solve a maximum a posterior estimation of Markov Random Field model and the classifications of two acute leukemia families, two lymphoblastic sub-types and three myeloblastic sub-types, using K-nearest neighbor, random forest, simple logistic, SVM and random committee based on geometric, statistical, texture, size ratio and eigenvalues. [Ghosh et al. \(2016\)](#) develop algorithms to perform image preprocessing and segmentation by including Leishman stained region highlighting, background scaling and false regions elimination and conduct classification using different fuzzy and non-fuzzy techniques based on shape, size, color and texture features for six BM cell types. [Karthikeyan and Poornima \(2017\)](#) adopt the assemblage of histogram equalization, median filtering, fuzzy C-means, K-means clustering, Gabor texture extraction method and Support Vector Machine (SVM) providing information whether a patient is affected by leukemia or not. [Mishra et al. \(2017\)](#) develop a framework by integrating histogram equalization, Wiener filtering, marker-based watershed segmentation, gray level co-occurrence matrix, probabilistic principal component analysis and random forest to classify the BM cells into two types. Nonetheless, the approaches mentioned above are impractical for clinical usage due to the limited capability to identify a small number of cell types and to deal with small images instead of WSIs.

## 2.2. Deep learning-based algorithm

Deep learning-based approaches have been successfully implemented in many clinical diagnoses ([Campanella et al., 2019; Chlebus et al., 2018; De Fauw et al., 2018; Liu et al., 2020a; Sun et al., 2019](#)). Deep learning has the capability to produce more sensible outputs for a testing dataset compared to the conventional machine learning methods by learning internal data representations from a training dataset using several processing layers ([LeCun et al., 2015](#)). However, the study of fully automatic BM NDC on WSIs using deep learning is poorly explored. Some studies build deep learning approaches for BM NDC using single-cell images, which are manually cropped by examiners. [Choi et al. \(2017\)](#) propose dual-stage CNNs to discriminate the BM cells into ten BM cell types of myeloid and erythroid maturation series. [Kimura et al. \(2019\)](#) develop a novel technique that able to classify the BM cells into 17 BM cell types and 97 morphological features using a CNN and directly generate assisted diagnosis to differentiate into MDS and AA using Extreme Gradient Boosting. [Matek et al. \(2019\)](#) utilize ResNeXt CNN ([Xie et al., 2017](#)) to classify white blood cells in peripheral blood smears. [Sahlot et al. \(2020\)](#) present a CNN with a statistically enhanced Salp Swarm Algorithm to classify the BM cells into two types.

Although there has been progress in conducting BM NDC using multiple-cell images, numerous studies still work on small images instead of WSIs. [Xie et al. \(2018\)](#) implement a structured regression model based on fully residual CNN to merely detect the BM cell without the differentiation process involved. [Chandradevan et al. \(2019\)](#) develop a semi-automatic approach for the BM cell detection and classification of 13 BM cell types using a manually pre-selected regions of interest (ROI). [Hu et al. \(2019\)](#) employ Generative Adversarial Networks (GAN) ([Goodfellow et al., 2014](#)) to learn cell-level visual representation for the nuclei segmentation and cell classification tasks, which deal with at most four BM cell types. [Yu et al. \(2019\)](#) develop deep neural networks, including Path-Aggregation Network, for the BM cell identification and classification of 11 BM cell types. As fully automatic deep learning-based BM NDC on WSIs is poorly explored, two recently published small-image-based approaches

([Chandradevan et al., 2019; Yu et al., 2019](#)) are adopted as benchmark methods. Cascade R-CNN has been widely used for object detection and classification, and it has demonstrated superior performance in object detection and classification ([Aziz et al., 2020; Liu et al., 2020b](#)). In this work, Cascade R-CNN is adopted as the base deep learning approach for both detection BM particles and cellular trails as ROI and classification of BM cells within the ROI.

## 3. Methods

### 3.1. Hierarchical deep learning framework

In this study, we develop an efficient and fully automatic hierarchical patch-based deep learning framework for BM NDC WSI analysis in seconds. The proposed hierarchical patch-based deep learning framework with two Cascade R-CNN deep learning models uses a coarse-to-fine strategy to rapidly locate BM particles and cellular trails and identify BM cells in WSIs by accessing two magnification levels. The overview of the proposed system framework is shown in [Fig. 2](#). Initially, a WSI input is formulated into a multi-resolution pyramid data structure with multiple levels from low to high magnification as shown in [Fig. 2\(a\)](#). At the coarse-level, fast localization of BM particles and cellular trails is performed using the low-level deep learning model ([Fig. 2\(b\)](#)) on the low-magnification level that is closest to the patch unit size, generating a set of ROIs ([Fig. 2\(c\)](#)) as described in [Section 3.1.1](#). At the fine-level, the output of coarse level is then mapped to high-magnification level generating a set of high-magnification ROI which is then formulated to patch-based data structure. BM cell detection-classification is performed on patch-based data structure using the high-level deep learning model ([Fig. 2\(d\)](#)), producing patch-based cell identification results ([Fig. 2\(e,f\)](#)) as described in [Section 3.1.2](#). A fast stitching model ([Fig. 2\(g\)](#)) is then applied to generate final outputs ([Fig. 2\(h\)](#)) as described in [Section 3.1.3](#).

Let  $\mathcal{I}(p, q)$  denote the two dimensional image input space and  $\mathcal{O} = \{o^d(p, q)\}_{d=1}^D = \{(c^d, \mathbf{b}^d)\}_{d=1}^D$  the output set of BM cell, where  $D, c$  denote the number of detected BM cells and the corresponding BM cell type, respectively, and  $\mathbf{b}^d = (b_x^d, b_y^d, b_w^d, b_h^d)$  represents the bounding box of a detected BM cell  $o^d$ . We formulate the two dimensional input data  $\mathcal{I}(p, q)$  into a multi-resolution pyramid data structure with multiple levels  $M$  from high to low magnification  $\{I_m\}_{m=0}^M$ . Our framework  $\Theta$  consists of three models, namely  $\Theta_{trail}$ ,  $\Theta_{cell}$ ,  $\Theta_{stitch}$ , and in this study, the Cascade R-CNN model ([Cai and Vasconcelos, 2019](#)) is adopted as the base deep learning method of the proposed framework as a preliminary test result shows that the Cascade R-CNN outperforms the normal Faster R-CNN in detecting objects that are close to each other.

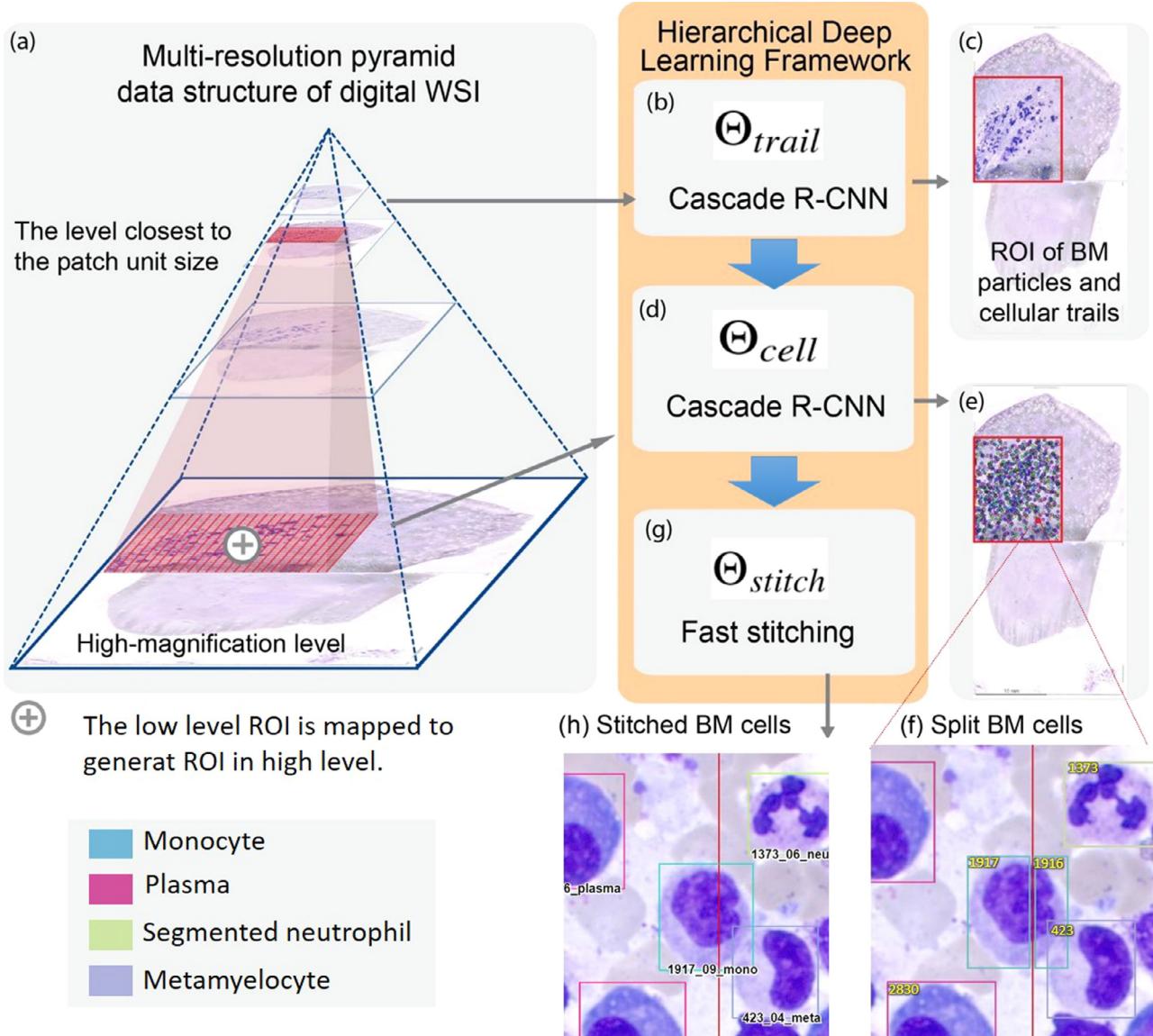
#### 3.1.1. Fast localization of BM particles and cellular trails

The model  $\Theta_{trail}$  aims to rapidly locate BM particles and cellular trails on the low-magnification level  $I_\xi(u, v)$ , generating a set of ROI  $\mathcal{L} = \{k_\xi^n(u, v)\}_{n=1}^N \in I_\xi(u, v)$ , where  $u, v$  denote the horizontal index and vertical index of the low-magnification level, respectively, and  $k, N$  denote the ROI and the number of detected ROI, respectively, formulated as follows:

$$\mathcal{L} = \{k_\xi^n(u, v)\}_{n=1}^N = \Theta_{trail}(I_\xi(u, v)) \quad (1)$$

The model  $\Theta_{trail}$  is developed based on Cascade Region-Based Convolutional Neural Networks (Cascade R-CNN) ([Cai and Vasconcelos, 2019](#)) as described in [Section 3.2](#). The set of ROI  $\mathcal{L}$  is then mapped to the high-magnification level  $I_0(p, q)$  generating a set of ROI  $\mathcal{H} = \{k_0^n(p, q)\}_{n=1}^N \in I_0(p, q)$  calculated as follows:

$$\mathcal{H} = \{k_0^n(p, q)\}_{n=1}^N = \begin{cases} \{k_\xi^n(\lfloor \frac{p}{2^\xi} \rfloor, \lfloor \frac{q}{2^\xi} \rfloor)\}_{n=1}^N, & \xi \in \mathbb{R}_{>0} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$



**Fig. 2.** Illustration of the proposed hierarchical deep learning framework for BM WSI analysis using a coarse-to-fine strategy. (a) Multi-resolution pyramid data structure of WSIs. The proposed hierarchical deep learning framework consists of two cascaded deep learning models: (b) the first layer Cascade R-CNN model conducts fast localization of BM particles and cellular trails to generate (c) ROI, and (d) the second layer Cascade R-CNN model performs BM cell detection-classification inside ROI, generating (e) multi-type patch-based detection results, (f) BM cells could be split around multiple patches where the number in the top left of the bounding box indicates the object detection ID before stitching, (g) a fast stitching model is built to merge patch-based results to produce (h) the final outputs of WSIs.

In this study, the distribution of the total size and the number of detected ROIs per WSI is shown in Fig. 3.

### 3.1.2. BM Cell detection-classification

Initially, the set of ROI  $\mathcal{H}$  is formatted into a patch-based data structure  $\mathcal{H}_{patch} = \{k_{0,\sigma,\tau}^n(i,j)\}_{n=1}^N$ , where  $\sigma, \tau, i, j$  denote the patch column index, patch row index, patch horizontal sub-index and patch vertical sub-index, respectively. The values  $\sigma, \tau, i, j$  are calculated as follows:

$$\begin{aligned} \sigma &= \lfloor p/w_\theta \rfloor, & \tau &= \lfloor q/h_\theta \rfloor, \\ i &= p - \sigma \times w_\theta, & j &= q - \tau \times h_\theta \end{aligned} \quad (3)$$

where  $w_\theta, h_\theta$  denote the patch width and patch height, respectively;  $w_\theta = 640, h_\theta = 640$ . The values  $i, j$  are in range  $\{0, \dots, w_\theta - 1\}$  and  $\{0, \dots, h_\theta - 1\}$ , respectively.

The model  $\Theta_{cell}$  conducts BM cell detection-classification using the patched ROI  $\mathcal{H}_{patch}$  to generate the output set of BM cells in a patch-based data structure  $\mathcal{O}_{patch} = \{\mathcal{O}_{\sigma,\tau}^d(i,j)\}_{d=1}^D$  formulated as

follows:

$$\mathcal{O}_{patch} = \{\mathcal{O}_{\sigma,\tau}^d(i,j)\}_{d=1}^D = \{\Theta_{cell}(k_{0,\sigma,\tau}^n(i,j))\}_{n=1}^N \quad (4)$$

The model  $\Theta_{cell}$  is developed based on Cascade R-CNN as described in Section 3.2. The model for BM cell detection and classification is a patch-based model for detection and classification of multiple cells in a patch.

### 3.1.3. Fast stitching model for integrating multi-type patch-based identification results

As WSIs are usually with extremely high resolution, it is not feasible to put the images directly into a deep learning model due to the enormous computational burden. To preserve image details, many studies (Coudray et al., 2018; Zhang et al., 2019; Campanella et al., 2019) organize the high-resolution images into computationally memory-affordable patches, which are used as the model inputs. The results of patches are integrated with the stitching model. The output set  $\mathcal{O}_{patch}$  in the patch-based data structure

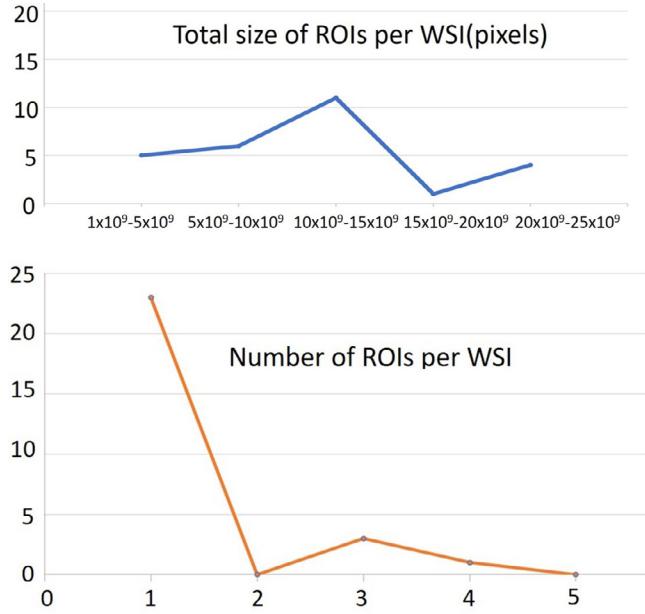


Fig. 3. Distribution of the total size and the number of detected ROIs per WSI.

is transformed back to the two dimensional data structure of the original image input  $\mathcal{I}(p, q)$ , generating  $\mathcal{O} = \{o^d(p, q)\}_{d=1}^D$ . The values  $p, q$  can be calculated as follows:

$$p = i + \sigma \times w_\theta, \quad q = j + \tau \times h_\theta \quad (5)$$

Due to the patch-based data structure, an individual detected BM cell could be split into multiple detection results around different patches, as shown in Fig. 4. Thus, we develop a fast stitching model  $\Theta_{stitch}$  to integrate split detected BM cells, formulated as follows:

$$\mathcal{O} = \{o'^d(p, q)\}_{d=1}^{D' < D} = \Theta_{stitch}(\{o_{\sigma, \tau}^d(i, j)\}_{d=1}^D) \quad (6)$$

where  $o', D'$  denote the detected BM cell and the number of detected BM cells after the stitching, respectively.

Let  $\mu$  denote the distance between the two points,  $i, j$ . The value  $\mu$  is calculated as follows:

$$\mu(i, j) = \sqrt{dx^2 + dy^2} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (7)$$

where  $dx, dy$  denote the difference between the  $x$ -coordinates of the two points and the difference between the  $y$ -coordinates, respectively.

For each detected BM cell  $o_{\sigma, \tau}^d(i, j)|_{\xi^d > \lambda}$  where  $\xi^d$  is the detection probability and  $\lambda = 0.1$  in this study, we utilize the information of BM cell type  $c^d$  and calculate the four corner coordinates of the associated bounding box  $\{e_\phi^d\}_{\phi=\{1,2,3,4\}}$ . A four neighborhood system is designed for the fast stitching model  $\Theta_{stitch}$ . A set of the average distances  $\zeta^d$  is computed from the detected BM cell  $o_{\sigma, \tau}^d(i, j)$  to the detected BM cells  $\{o_{\alpha, \beta}^t(i, j)\}_{t \neq d} \wedge |\alpha - \sigma| + |\beta - \tau| = 1$  in the four neighboring patches. The value  $\zeta^d$  is calculated using  $\zeta^d = (C_1^d + C_2^d)/2$ , where  $C_1^d, C_2^d$  denote the smallest and second smallest distances between the corner points of the detected BM cell  $o^d$  and the neighboring BM cell  $o^t$ . The smallest distance  $C_1^d$  is calculated as follows:

$$C_1^d = \min_{\phi, \psi} \mu(e_\phi^d, e_\psi^t) \quad (8)$$

To calculate the second smallest distance  $C_2^d$ , we exclude the corner points  $\phi', \psi'$ , which have been used to calculate the smallest distance  $C_1^d$ , formulated as follows:

$$\phi', \psi' = \operatorname{argmin}_{\phi, \psi} \mu(e_\phi^d, e_\psi^t) \quad (9)$$

Next, the second smallest distance  $C_2^d$  can be calculated as follows:

$$C_2^d = \min_{\phi, \psi} \mu(e_\phi^d, e_\psi^t) |_{(\phi, \psi) - (\phi', \psi')} \quad (10)$$

The detected BM cell  $o^d$  is stitched with the neighboring detected BM cells  $\{o^t\}$  generating a single BM cell  $o'^d$  under the following condition:

$$o'^d = (c'^d, b'^d) = o^d \bigcup \{o^t\} |_{\zeta^d < \epsilon} \quad (11)$$

where  $\epsilon$  denotes the distance threshold. In this study, the distance threshold  $\epsilon$  is set as 25.

### 3.2. Modified Cascade R-CNN

The Cascade R-CNN model is a multi-stage object detection deep learning architecture (Cai and Vasconcelos, 2019), which consists of one region proposal network (RPN) and three Faster R-CNN (Ren et al., 2017) detectors. RPN proposes a set of low quality candidate bounding boxes from an image to determine the occurrence of an object, and therefore the subsequent detectors are developed to be more selective for lower quality candidates. In this study, four modifications are made for optimization as follows.

- The performance of an AI model is affected because the data set is not fully labeled, and many unlabeled cells are defined as background tissues, which severely confuse AI models during training. To deal with this issue, we have replaced the original loss function with an improved soft-sampling loss function. The soft-sampling technology has been demonstrated to be useful in OpenImagesV3 challenge (Krasin et al., 2017), which is a real-world dataset with missing annotations (Wu et al., 2019). In the original Cascade R-CNN, the  $L_{loc}$  loss is characterized by a tuple of true bounding-box regression targets, and the  $R_{loc}$  is the bounding box risk. The bounding box-regressor  $\Psi_\vartheta(o^d, b^d)$  aims to regress a candidate bounding box  $b^d$  of an object proposal image  $a^d$  into a referenced bounding box  $g^d$  and is learned from a training set  $(b^d, g^d)$  by minimizing the risk formulated as follows:

$$L_{loc}(\mathbf{r}, \mathbf{s}) = \sum_d \text{smooth}_{L1}(r^d - s^d) \quad (12)$$

$$R_{loc}(\Psi_\vartheta) = \sum_d L_{loc}(\Psi_\vartheta(a^d, b^d), g^d) \quad (13)$$

where  $\mathbf{r}$  are bounding-box regression offsets, and  $\mathbf{s}$  are ground-truth bounding-box regression targets.

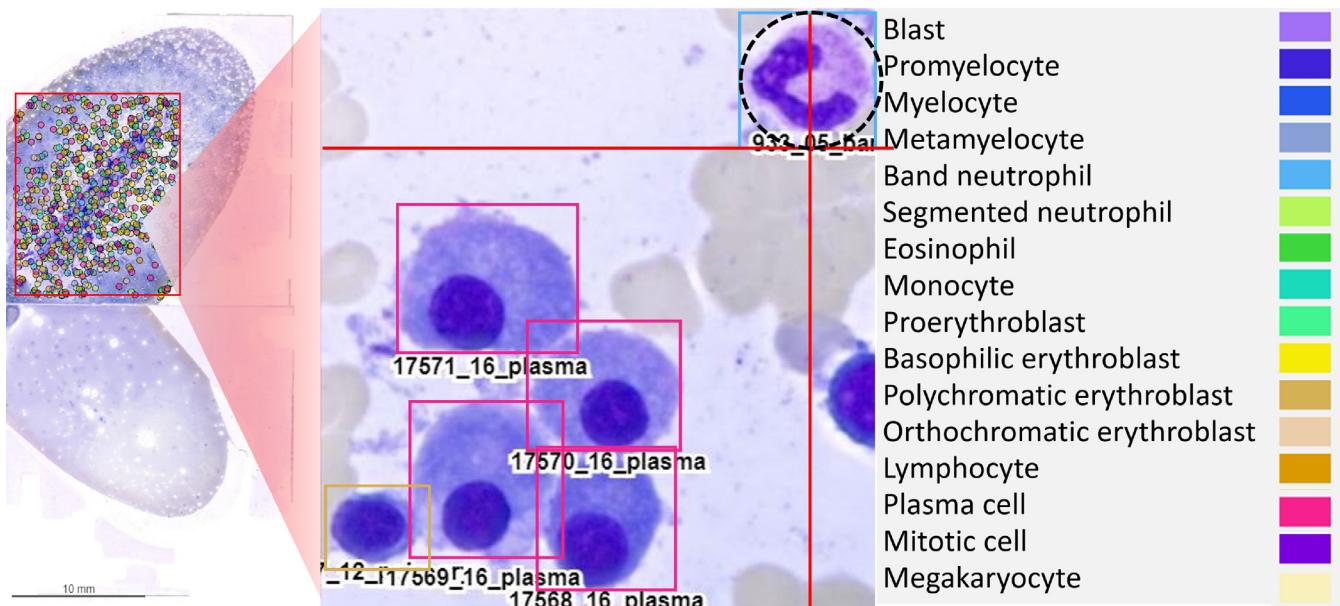
The improved soft-sampling loss  $L_{loc}$  is reformulated as follows by adding the weight assigned in consideration of IoU value and initiated with a Gompertz function as shown below.

$$L_{loc}(\mathbf{r}, \mathbf{s}) = \sum_d w_\varpi^d \times \text{smooth}_{L1}(r^d - s^d) \quad (14)$$

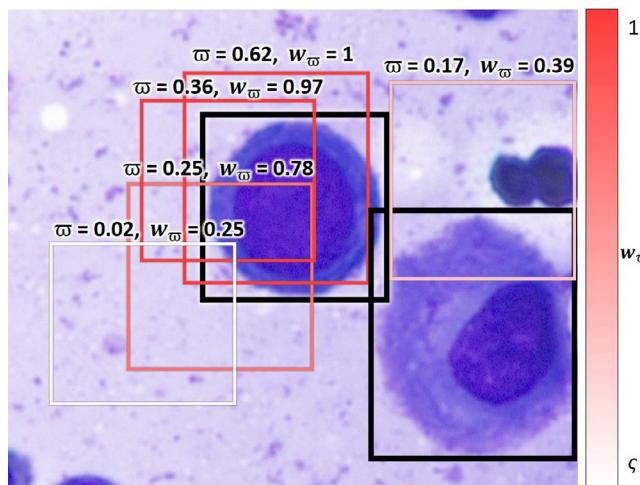
$$w_\varpi = \varsigma + (1 - \varsigma)e^{-se^{-v\varpi}} \quad (15)$$

where  $\varpi$  is the intersection over union (IoU) between the reference standards  $g^d$  and the candidate bounding boxes  $b^d$ , and  $\varsigma=0.25$ ,  $s=50$ ,  $v=20$  (as in Wu et al. (2019)).

Soft Sampling re-weights the gradients of proposal samples, i.e. candidate bounding boxes, as a function of overlap with positive instances and guarantees that the uncertain background regions are given a smaller weight compared to the hard-negatives. The IoU based soft sampling approach reduces the gradients as an IoU function with positive examples to ameliorate the effect of false negatives, which ensures that all sampled boxes contribute in training, but more confident samples are given more focus as shown in Fig. 5. The effectiveness of IoU based soft sampling has been demonstrated



**Fig. 4.** Sample BM cell stitching results by the proposed method for merging of cells (denoted by black dotted circle) split among patches (denoted by red lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** Soft-Sampling samples. The proposals are weighted by overlap based soft-sampling. Black bounding boxes are the reference standard by medical experts.

on multiple datasets, including PASCAL VOC, OpenImagesV3 (Wu et al., 2019) and OpenImagesV4 (Gao et al., 2018), and the method using soft sampling ranks top 2nd among more than 450 submissions worldwide in ECCV 2018 OpenImagesV4 object detection challenge (Gao et al., 2018).

In the training process, we apply data augmentation to boost up positive samples with four variation factors, including rotation, contrast, saturation and brightness. To obtain gradient stability of the training dataset and prevent overflow from data augmentation, a mean normalization mechanism is applied (Pal and Sudeep, 2016) as follows:

$$\mathbf{Z}' = \mathbf{Z} - \mathbf{W} \quad (16)$$

where  $\mathbf{Z}'$  is the normalized data,  $\mathbf{Z}$  is the original data and  $\mathbf{W} = (104, 117, 124)$  is trained from the ImageNet dataset (Deng et al., 2009).

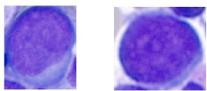
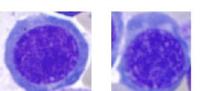
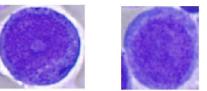
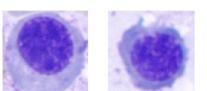
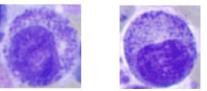
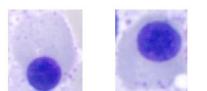
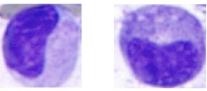
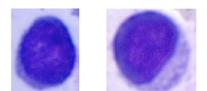
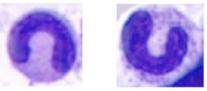
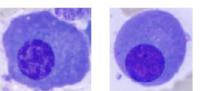
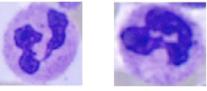
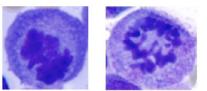
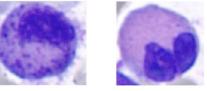
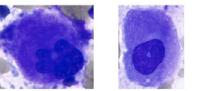
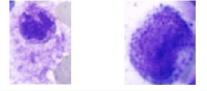
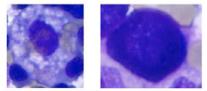
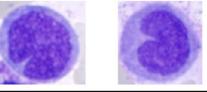
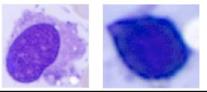
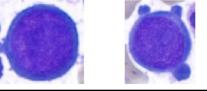
- To decrease memory usage and accommodate limited memory size of cheap GPU cards like 1080 Ti, we modify the input size from  $800 \times 1320$  to  $640 \times 1120$ .
- The original design of Cascade R-CNN utilizes a fixed step size, namely 120000, for decreasing the learning rate 10% every 120,000 iterations. However, when the lower the learning rate becomes, the less the training samples contributes to the model learned. For a large data set such as gigapixel WSIs, some samples may not be learned properly after the learning rate drops. To resolve this issue, we modify the step size to a flexible data-oriented value that is 10 times of the number of patches of the whole training set.

In this study, we use three detectors  $\eta = 3$  with the IoU threshold  $\varphi = \{0.5, 0.6, 0.7\}$  and  $\lambda = 1$ . The network is trained using cross loss entropy function, and a pre-trained ResNet-101 (He et al., 2016) model is used for initialization. The entire network is trained with the learning rate of  $5e - 4$  and the weight decay of  $1e - 4$ .

## 4. Experiments and results

### 4.1. Datasets

Two datasets of BM aspirate smears were built in this study and collected from the Department of Laboratory Medicine, National Taiwan University Hospital, with ethical approval (201911074RINB). The first dataset contains 12,426 annotated cells as shown in Fig. 6, which are gathered from 27 de-identified patients' smear slides from January 2017 to November 2019, including 12 smears for non-neoplastic/normal marrow, eight for myeloid diseases (six for AML, one for MDS with excess blasts and one for CMML), three for ALL and four for MM. The first dataset is used for 10-fold cross validation in evaluation as shown in Table 1. The second separate dataset contains 3005 annotated cells as shown in Table 2, gathered from 10 de-identified patients' smear slides from April 2021 to May 2021, including three normal smears and seven diseased smears (two MM, two AML, one CMML, one CML and one MDS). The second dataset is used as a separate testing set to examine the model generalizability.

Cell Types	Number of cells annotated	Examples	Cell Types	Number of cells annotated	Examples
Blast	2197		Basophilic erythroblast	379	
Promyelocyte	872		Polychromatic erythroblast	1195	
Myelocyte	825		Orthochromatic erythroblast	804	
Metamyelocyte	688		Lymphocyte	427	
Band neutrophil	542		Plasma cell	1262	
Segmented neutrophil	1201		Mitotic cell	75	
Eosinophil	267		Megakaryocyte	300	
<sup>a</sup> Basophil	16		Other cells <sup>a,b</sup>	26	
Monocyte and promonocyte	365		Smudge cell <sup>a</sup>	896	
Proerythroblast	89		Total	12426	

<sup>a</sup> These cell types are not included in the training set.

<sup>b</sup> Other cells include histiocytes/macrophages, osteoblasts, osteoclasts, and non-hematopoietic cells such as metastatic carcinoma cells.

**Fig. 6.** List of total annotated cells in the first dataset according to cell types with reference images.

The clinical diagnosis is made by qualified hematologists in accordance with the clinical information, laboratory data, immunophenotyping, cytogenetic analysis and molecular study.

The slides are stained with Liu's stain, a modified Romanowsky stain, and digitized at 0.25 μm/pixel (×40 objective) using the Panoramic 250 Flash III WSI scanner (3DHISTECH, Budapest, Hungary). The averaged pixel resolution and slide dimensions are 253,070 × 482,247 pixels and 21.83 × 41.59 mm<sup>2</sup>, respectively. The WSIs are annotated by two medical experts, including a hematologist with three years of experience (examiner 1) and a senior medical technician with 20 years of expertise (examiner 2) based on the cytomorphological criteria established by the College of American Pathologist (Glassy, 1998). All annotations from the hematologist and the medical technician are used for quantitative evaluation. Both annotators annotated cells of partial slides and produced approximately 500 annotations per slide for the first dataset and 300 annotations per slide for the second dataset, respectively. The annotators differentiate BM cells into 19 cell types, including blast, promyelocyte, myelocyte, metamyelocyte, band neutrophil, segmented neutrophil, eosinophil, monocyte and promono-

cyte, proerythroblast, basophilic erythroblast, polychromatic erythroblast, orthochromatic erythroblast, plasma cell, mitotic cell, megakaryocyte, basophil, smudge cell, lymphocyte and others (histiocyte/macrophage, osteoblast, osteoclast and non-hematopoietic cell). In this study, we use 16 cell types to be trained and evaluated.

The remaining three cell types, basophil, smudge cell and others are excluded from training model because of following reasons. First, basophils naturally exist in bone marrow with very few numbers, accounting only 0 to 0.5 percent of total BM cells (Bain et al., 2016). Therefore, the number of annotations of basophils is too low to train an AI model. Second, smudge cell is excluded in this study because examiners require complete cells (i.e. with both intact cytoplasm and nucleus) in order to obtain definite cell classification. However, during sample collection and preparation, cells might be broken or distorted, and thus those cells will not be selected for classification to avoid misdiagnosis due to inadequate information. From time to time, there are only limited numbers of sample cells in some specimens, examiners are reluctantly forced to conduct cell classification based on

**Table 1**

Quantitative comparison with the benchmark methods by BM cell type using cross validation on the 1st dataset.

Method	Chandrasevan et al. (2019)		Yu et al. (2019)		The proposed method			
	Data type		Manual ROI-based		Patch wise 10-fold CV		Slide wise LOOCV	
	small images		small images		WSIs		WSIs	
Cell type	Recall	Accuracy	Recall	Accuracy	Recall	Accuracy	Recall	Accuracy
Blast	0.850	0.974	0.815	0.956	<b>0.974±0.022</b>	<b>0.978±0.009</b>	<b>0.882±0.222</b>	0.958±0.052
Promyelocyte	0.870	0.963	0.854	<b>0.984</b>	<b>0.877±0.049</b>	0.976±0.008	0.857±0.212	0.968±0.051
Myelocyte	<b>0.820</b>	0.972	0.708	0.969	0.757±0.112	<b>0.982±0.008</b>	0.733±0.308	<b>0.979±0.027</b>
Metamyelocyte	0.840	0.971	0.847	0.968	<b>0.876±0.075</b>	<b>0.986±0.006</b>	0.733±0.338	<b>0.985±0.018</b>
Band neutrophil	<b>0.900</b>	0.988	0.876	0.974	0.852±0.147	<b>0.991±0.006</b>	0.786±0.267	0.985±0.025
Segmented neutrophil					<b>0.946±0.057</b>	<b>0.991±0.005</b>	<b>0.929±0.174</b>	0.985±0.019
Eosinophil	0.940	0.993	<b>0.987</b>	0.990	0.923±0.159	<b>0.999±0.003</b>	0.828±0.339	<b>0.997±0.006</b>
Monocyte	0.690	0.973	<b>0.825</b>	0.982	0.651±0.284	<b>0.990±0.008</b>	0.611±0.345	<b>0.986±0.015</b>
Erythroblast	0.860	0.979	0.971	<b>0.984</b>	<b>0.980±0.020</b>	0.979±0.020	<b>0.971±0.040</b>	<b>0.991±0.016</b>
Pro.ery.	-	-	-	-	<b>0.900±0.108</b>	<b>0.996±0.004</b>	<b>0.678±0.410</b>	<b>0.997±0.006</b>
Baso.ery.					<b>0.892±0.013</b>	<b>0.983±0.005</b>	<b>0.815±0.304</b>	<b>0.995±0.011</b>
Poly.ery.					<b>0.996±0.023</b>	<b>0.973±0.012</b>	<b>0.970±0.078</b>	<b>0.997±0.007</b>
Ortho.ery.					<b>0.992±0.088</b>	<b>0.984±0.010</b>	<b>0.990±0.033</b>	<b>0.999±0.002</b>
Lymphocyte	0.680	0.970	0.750	0.927	<b>0.945±0.093</b>	<b>0.996±0.002</b>	<b>0.919±0.245</b>	<b>0.994±0.009</b>
Plasma cell	0.890	0.990	-	-	<b>0.987±0.015</b>	<b>0.997±0.002</b>	<b>0.987±0.040</b>	0.987±0.029
Mitotic cell	-	-	-	-	<b>1±0</b>	<b>1±0.001</b>	<b>0.950±0.194</b>	<b>0.999±0.004</b>
Megakaryocyte	-	-	-	-	<b>0.997±0.009</b>	<b>1±0.001</b>	<b>0.983±0.064</b>	<b>0.998±0.008</b>
Basophil	0.890	0.988	0.818	0.987	-	-	-	-
Invalid/Unknown	0.920	0.928	0.666	0.877	-	-	-	-
Average	0.846	0.974	0.829	0.964	<b>0.905±0.078</b>	<b>0.989±0.006</b>	<b>0.878±0.093</b>	<b>0.988±0.009</b>

**Table 2**

Quantitative evaluation of the proposed method on the 2nd dataset.

Cell type	Cell number	The proposed			
		$M_{Patch-10foldCV}$		$M_{Slide-LOOCV}$	
		Recall	Accuracy	Recall	Accuracy
Blast	126	0.342	0.965	0.529	<b>0.978</b>
Promyelocyte	147	0.700	<b>0.983</b>	<b>0.893</b>	<b>0.987</b>
Myelocyte	310	0.613	0.954	0.546	<b>0.975</b>
Metamyelocyte	265	0.819	0.963	0.815	<b>0.980</b>
Band neutrophil	298	<b>0.933</b>	<b>0.991</b>	0.865	<b>0.990</b>
Segmented neutrophil	341	<b>0.986</b>	0.987	<b>0.998</b>	0.983
Eosinophil	189	<b>0.994</b>	0.985	<b>0.994</b>	<b>0.993</b>
Monocyte	89	0.149	0.975	0.210	<b>0.985</b>
Erythroblast	946	<b>0.978</b>	0.975	<b>0.992</b>	<b>0.985</b>
Pro.ery.	109	<b>0.928</b>	<b>0.996</b>	<b>0.944</b>	<b>0.998</b>
Baso.ery.	165	<b>0.979</b>	<b>0.990</b>	<b>0.988</b>	<b>0.993</b>
Poly.ery.	278	<b>0.957</b>	<b>0.991</b>	<b>0.986</b>	<b>0.996</b>
Ortho.ery.	394	<b>0.997</b>	<b>0.998</b>	<b>0.997</b>	<b>0.997</b>
Lymphocyte	54	<b>0.929</b>	0.950	<b>1</b>	<b>0.972</b>
Plasma cell	95	<b>0.936</b>	<b>0.988</b>	<b>0.948</b>	<b>0.992</b>
Mitotic cell	14	<b>0.875</b>	<b>0.999</b>	<b>0.611</b>	<b>0.997</b>
Megakaryocyte	131	<b>1</b>	<b>0.998</b>	<b>1</b>	<b>0.996</b>
Total	3005	-	-	-	-
Average	-	0.750	<b>0.983</b>	0.842	<b>0.988</b>

the information from smudge cells. Some cells could be successfully identified through specific characters (for example, neutrophil can be identified easily through segmented nucleus even without intact cytoplasm); while the majority of smudge cells could not be identified successfully. Thirdly, other cells include histiocytes/macrophages, osteoblasts, osteoclasts, and non-hematopoietic cells such as metastatic carcinoma cells are also excluded from AI training model because these cells originally are not included in BM NDC (Lee et al., 2008). Furthermore, these cells also exist in BM with low numbers that will not be enough to train an AI model. Based on the above reasons, these three cell types are excluded in our training model.

#### 4.2. Evaluation method

The evaluation was conducted on non-overlap patches. Cross validation (CV) is a common technique for evaluating the model

generalizability on unseen data (Tabe-Bordbar et al., 2018). In this study, we evaluate the proposed method with both patch wise 10-fold CV and slide wise leave one out CV (LOOCV). For the patch wise 10-fold CV, as the number of cells for each cell type is unbalanced, evaluation is performed considering the number of cells in each cell class with non-overlap image patch used as a data unit; image patches with the cell type that has the least number of cells are first distributed evenly to each subset, and then the remaining patches with the cell type that has the second least number of cells are distributed evenly to each subset. The data allocation process continues until there is no remaining patch.

For quantitative evaluation, we utilize the recall and accuracy to compare and measure the performance of the benchmark approaches and the proposed method. The metrics are calculated as follows:

$$R = \frac{TP}{TP + FN} \quad (17)$$

**Table 3**  
Evaluation of the proposed method on the diseased data.

Cell type	$M_{Patch-10foldCV}$				$M_{Slide-LOOCV}$			
	2nd Dataset		1st Dataset		2nd Dataset		1st Dataset	
	Recall	Accuracy	Recall	Accuracy	Recall	Accuracy	Recall	Accuracy
Blast	0.309	0.958	<b>0.996</b>	<b>0.999</b>	0.491	0.971	<b>0.969</b>	<b>0.988</b>
Promyelocyte	0.714	<b>0.986</b>	<b>0.980</b>	<b>0.996</b>	<b>0.932</b>	<b>0.993</b>	<b>0.923</b>	<b>0.989</b>
Myelocyte	0.598	0.965	<b>0.972</b>	<b>0.997</b>	0.555	<b>0.987</b>	0.809	<b>0.990</b>
Metamyelocyte	0.772	0.970	<b>0.938</b>	1	0.738	<b>0.984</b>	<b>0.952</b>	<b>0.998</b>
Band neutrophil	<b>0.915</b>	<b>0.990</b>	1	1	0.854	<b>0.990</b>	0.819	<b>0.998</b>
Segmented neutrophil	<b>0.983</b>	0.982	<b>0.955</b>	<b>0.998</b>	<b>0.997</b>	0.978	<b>0.939</b>	<b>0.996</b>
Eosinophil	1	0.98	0.958	1	1	<b>0.991</b>	0.866	<b>0.999</b>
Monocyte	0.144	0.979	<b>0.929</b>	<b>0.998</b>	0.233	0.989	0.641	<b>0.996</b>
Erythroblast	<b>0.978</b>	0.975	<b>0.995</b>	<b>0.998</b>	<b>0.989</b>	0.982	<b>0.991</b>	<b>0.997</b>
Pro.ery.	<b>0.939</b>	<b>0.996</b>	1	1	<b>0.917</b>	<b>0.998</b>	1	1
Baso.ery.	<b>0.994</b>	<b>0.988</b>	1	1	<b>0.982</b>	<b>0.992</b>	<b>0.984</b>	1
Poly.ery.	<b>0.960</b>	<b>0.989</b>	<b>0.975</b>	<b>0.998</b>	<b>0.980</b>	<b>0.970</b>	<b>0.998</b>	<b>0.998</b>
Ortho.ery.	<b>0.996</b>	<b>0.997</b>	1	1	<b>0.996</b>	<b>0.996</b>	1	<b>0.999</b>
Lymphocyte	<b>0.900</b>	0.948	<b>0.900</b>	1	1	<b>0.972</b>	<b>0.875</b>	<b>0.998</b>
Plasma cell	<b>0.910</b>	0.983	<b>0.990</b>	<b>0.994</b>	<b>0.927</b>	<b>0.989</b>	<b>0.986</b>	<b>0.992</b>
Mitotic cell	<b>0.750</b>	<b>0.998</b>	1	1	<b>0.417</b>	<b>0.996</b>	1	1
Megakaryocyte	1	<b>0.997</b>	<b>0.964</b>	<b>0.999</b>	1	<b>0.996</b>	<b>0.979</b>	<b>0.999</b>
Average	0.758	<b>0.985</b>	<b>0.975</b>	<b>0.999</b>	0.824	<b>0.988</b>	<b>0.924</b>	<b>0.996</b>

**Table 4**

10-fold CV results in detection of BM particles and cellular trails.

10-fold CV	Recall	Precision	Accuracy	IOU
Object-based	0.93±0.11	1±0	0.93±0.11	0.75±0.19
Slide-based	1±0	1±0	1±0	0.74±0.14

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

where  $R$ ,  $A$  denote recall and accuracy, respectively; true (T) and false (F) represent BM cell identified correctly or otherwise, respectively; positive (P) and negative (N) denote the different categories.

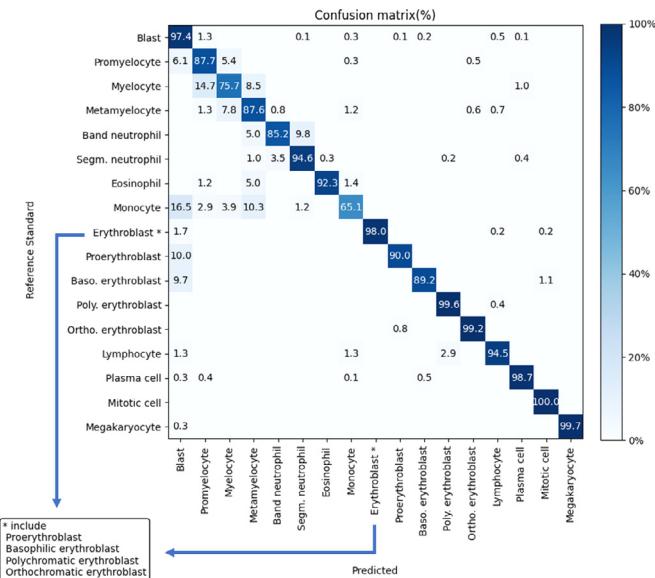
#### 4.3. Quantitative evaluation

The evaluations were conducted in four parts, including comparison and quantitative evaluation on the 1st dataset using patch wise 10-fold CV and slide wise LOOCV (see Table 1), evaluation on the separate 2nd dataset (see Table 2), evaluation on the diseased data of both datasets (see Table 3) and evaluation on detection of BM particle and cellular trails (see Table 4).

##### 4.3.1. Comparison and evaluation on the 1st dataset

Firstly, an evaluation was conducted for identification of individual BM cell types and overall performance by the proposed method using patch wise 10-fold CV and slide wise LOOCV on the 1st dataset and the benchmark approaches (Chandradevan et al., 2019; Yu et al., 2019), as shown in Table 1, in which there are 17 cell types in evaluation, containing individual 16 cell types and a combined erythroblast type (including proerythroblast, basophilic erythroblast, polychromatic erythroblast, orthochromatic erythroblast). The reported numbers of Chandradevan et al. (2019), Yu et al. (2019) are referred in Table 1.

Overall, the proposed method demonstrates promising performance with a recall of  $0.905 \pm 0.078$  and an accuracy of  $0.989 \pm 0.006$  for the patch wise 10-fold CV and a recall of  $0.878 \pm 0.093$  and an accuracy of  $0.988 \pm 0.009$  for the slide wise LOOCV. In comparison, the proposed method outperforms the benchmark methods (Chandradevan et al., 2019), which obtains a recall of 0.846, and an accuracy of 0.974, and Yu et al. (2019), which acquires a recall of 0.829 and an accuracy of 0.964.



**Fig. 7.** Confusion matrix for 17 cell types, which contains individual 16 cell types and a combined erythroblast type (in percentage %) by comparing the output of the proposed method and the reference standard of validation dataset using the patch wise 10-fold CV on the first dataset.

Furthermore, our framework surpasses the benchmark methods (Chandradevan et al., 2019; Yu et al., 2019) for 13 out of the 17 BM cell types w.r.t. recall and for 15 out of the 17 cell types w.r.t. accuracy in the patch wise 10-fold CV and for 11 out of the 17 BM cell types w.r.t. recall and for 12 out of the 17 cell types w.r.t. accuracy in the slide wise LOOCV as highlighted in bold in Table 1.

Furthermore, based on the confusion matrix in the patch wise 10-fold CV provided in Fig. 7, the proposed method shows excellent performance for 11 out of 17 BM cell types with more than 0.9 recall and for 7 out of 17 BM cell types with more than 0.95 recall. Results of BM NDC analysis by the proposed method on a normal and a diseased marrow sample are displayed in Fig. 9.

##### 4.3.2. Evaluation on the separate 2nd dataset

Secondly, an evaluation was conducted on a separate dataset to test the model generalizability of the proposed method; the two best models, i.e.  $M_{Patch-10foldCV}$  and  $M_{Slide-LOOCV}$ , generated from

the first dataset using the patch wise 10-fold CV and the slide wise LOOCV are tested on the 2nd dataset. Table 2 shows that even for a separate testing set, the proposed method still outperforms two benchmark approaches (Chandrasevan et al., 2019; Yu et al., 2019) for 12 out of the 17 BM cell types w.r.t. recall and for 8 out of the 17 BM cell types w.r.t. accuracy using  $M_{Patch-10foldCV}$  and for 12 out of the 17 BM cell types w.r.t. recall and for 16 out of the 17 BM cell types w.r.t. accuracy using  $M_{Slide-LOOCV}$  as highlighted in bold. Furthermore, the proposed method shows excellent performance for 11 out of 17 BM cell types with more than 0.9 recall and for 16 out of 17 BM cell types with more than 0.95 accuracy using  $M_{Patch-10foldCV}$  and for 10 out of 17 BM cell types with more than 0.9 recall and for 17 out of 17 BM cell types with more than 0.95 accuracy using  $M_{Slide-LOOCV}$ . The results show that the model trained from slide wise LOOCV obtains better performance on the new dataset. However, the results also show that current model is limited in identification of the four BM cell types with recall less than 0.8, including blast, promyelocyte, myelocyte, and monocyte, and further improvement with more training data could be conducted in future work.

#### 4.3.3. Evaluation on diseased data

Thirdly, a detailed analysis of the results with respect to diseased cells is provided in Table 3, showing that for the diseased cells, the proposed method performs better overall and outperforms two benchmark approaches (Chandrasevan et al., 2019; Yu et al., 2019) for more than 12 out of the 17 BM cell types in recall and for more than 13 out of the 17 BM cell types in accuracy as highlighted in bold for both datasets.

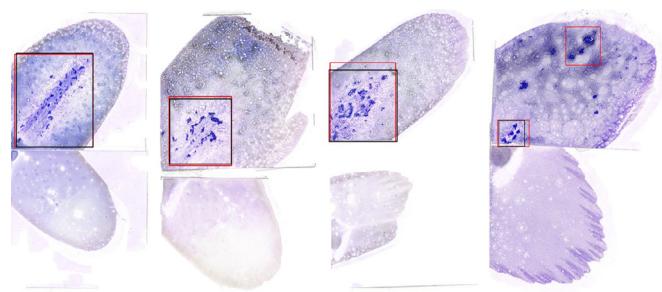
#### 4.3.4. Evaluation on BM particle and cellular trail detection

Fourthly, an evaluation was conducted for detection of BM particle and cellular trails by the proposed method. Table 4 shows the 10-fold cross validation results of the proposed method in localization of BM particles and cellular trails. Analysis results based on two kinds of measurement units, i.e. annotation (object-based) and the total annotated area (slide-based) are provided. For the slide based-detection, when the IoU between the AI prediction and the reference standard on a WSI greater than 0.5 is counted as successful detection; for the object based-detection, when the IoU between the AI prediction and the reference standard on a single ROI greater than 0.5 is counted as successful detection. In the case of object-based detection, the proposed method achieves the recall, precision, accuracy, and IOU of  $0.93 \pm 0.1135$ ,  $1 \pm 0$ ,  $0.93 \pm 0.1135$ , and  $0.7518 \pm 0.1852$  while in the case of slide-based detection, the proposed method achieves the recall, precision, accuracy, and IOU of  $1 \pm 0$ ,  $1 \pm 0$ ,  $1 \pm 0$ , and  $0.7428 \pm 0.1357$ , respectively. Qualitative results of our framework for detection of BM particles and cellular trails are displayed in Fig. 8.

#### 4.4. Computational time

Due to the enormous size of WSIs, the computational time for WSI analysis is critical for practical clinical usage. Therefore, we analyzed the data transfer time and AI inference time using different hardware settings. For data transfer time, data extraction of a WSI from a remote NAS takes 42 seconds while the data extraction from a local hard drive (HD) takes 6 seconds. For AI inference time in WSI BM NDC analysis, the proposed method takes 44 seconds using four NVIDIA TITAN RTX GPU cards and 94 seconds using two cheap NVIDIA Geforce GTX 1080 Ti GPU cards, respectively, showing that the proposed approach is an efficient and affordable solution for practical clinical usage (see Table 5).

In comparison with the benchmark approach (Chandrasevan et al., 2019), we perform a regression analysis to estimate our framework performance for 500 BM cells using



**Fig. 8.** Testing results of 10-fold CV in detection of BM particles and cellular trails by the proposed method where the black boxes are the reference standard by medical experts, and the red boxes are the detected BM particles and cellular trails by the proposed method, demonstrating that the proposed method is able to locate the areas of the BM particles and cellular trails. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 5**

Time for a WSI analysis by the proposed method.

Data Transfer		
Data Location Time (seconds)	Remote NAS	Local HD
42		6
AI Inference		
CPU	Intel Xeon Gold 6148	Intel Core i9-7900x
Memory	512 GB	128 GB
GPU	NVIDIA TITAN RTX $\times 4$	NVIDIA Geforce GTX 1080 Ti $\times 2$
Time (seconds)	44	94

**Table 6**

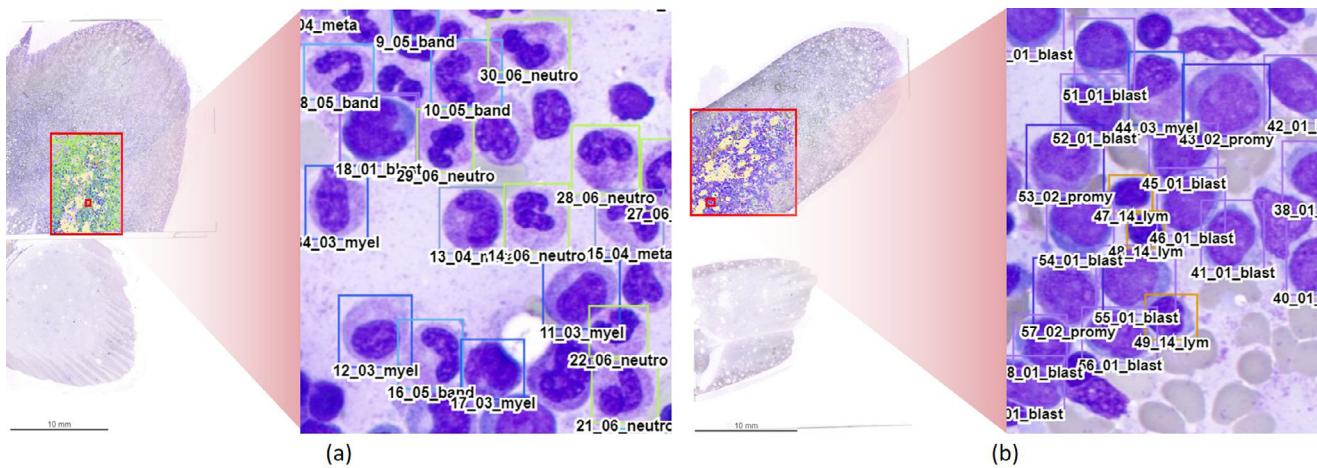
Intra- and inter-observer agreement using Cohen's kappa metric.

Intra-observer agreement analysis	kappa
Intra-examiner 1	0.745 (good)
Intra-examiner 2	0.897 (good)
Inter-observer agreement analysis	kappa
Inter-examiner 1 vs. 2	0.794 (good)
Inter-AI vs. examiner 1	0.918 (very good)
Inter-AI vs. examiner 2	0.931 (very good)

least squares. For analysis of 500 BM cells, the proposed method takes only 10.16 seconds using four NVIDIA TITAN RTX GPU cards and 22.51 seconds using two cheap NVIDIA Geforce GTX 1080 Ti GPU cards while the benchmark method (Chandrasevan et al., 2019) costs 162 seconds. The proposed method based on two different hardware sets is 16 times faster and 7.2 times faster than the benchmark method, respectively.

#### 4.5. Intra- and inter-observer variability

Cohen's kappa statistic is used to analyze the annotation agreement of intra- and inter-observer. Conventionally, a kappa value of <0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80 and 0.81–1.00 are interpreted as a poor, fair, moderate, good and excellent agreement, respectively (Buckens et al., 2013; Gianelli et al., 2014). In this study, the intra-observer analysis is performed based on two sets of annotations produced at an interval of one week. For the intra-observer variability, we perform the kappa analysis on 602 randomly selected BM cells from three WSIs, and for the inter-observer variability between two examiners, we perform the kappa analysis on 620 cells. The inter-observer analysis between AI and individual examiners is performed on 2154 randomly selected BM cells from 27 WSIs. Table 6 presents the kappa results. The intra-observer reliabilities of examiner 1 and 2 are interpreted as good with kappa values of 0.745 and 0.897, respectively, and the inter-observer relia-



**Fig. 9.** Sample BM cell detection-classification results by our proposed method for (a) normal and (b) diseased marrow.

bility between the two examiners is also good with a kappa value of 0.794. For the intra-observer reliabilities of AI and examiners are both very good with kappa values of 0.918 and 0.931, respectively, showing that the proposed AI model is reliable and consistent to examiners' decisions. Fig. 10 shows the result of inter-observer analysis between AI and examiner 1 on 2154 cells with the detailed number of cells for each type.

## 5. Discussion

In this study, we present an efficient and fully automatic hierarchical deep learning framework for BM NDC WSI analysis in seconds. This is an important issue considering BM NDC is the gold standard in diagnosing and managing various hematologic disorders. To the authors' best knowledge, this is the first study to perform BM NDC on WSIs with the most abundant cell types, including megakaryocyte, mitotic cell and detailed four stages of erythroblast, which were not reported before. Existing modern hematology analyzers fail to generate reliable results in differentiating BM cells (Meintker et al., 2013). Moreover, manual BM NDC is poorly reproducible, labor-intensive, time-consuming and complicated in clinical practice. As a heavy workload, examiners are required to analyze the BM smear cellular trails behind the particles and for at least 1,000 BM cells for two slides per patient (Lee et al., 2008). Furthermore, the differentiation task in BM NDC is difficult due to the small inter-class difference within a maturation stage and prone to intra- and inter-observer variations. In the related BM studies, the kappa of the inter-observer analysis ranges from 0.46 to 0.76 with a mean of 0.63 in Gianelli et al. (2014) and from 0.143 to 0.483 with a mean of 0.352 in Birkhead et al. (1987), respectively, which indicate that BM NDC is difficult, even for well-trained medical experts. Based on the related BM studies and our experiments, there exist inevitable variations on intra- and inter-observer indicated by the kappa values.

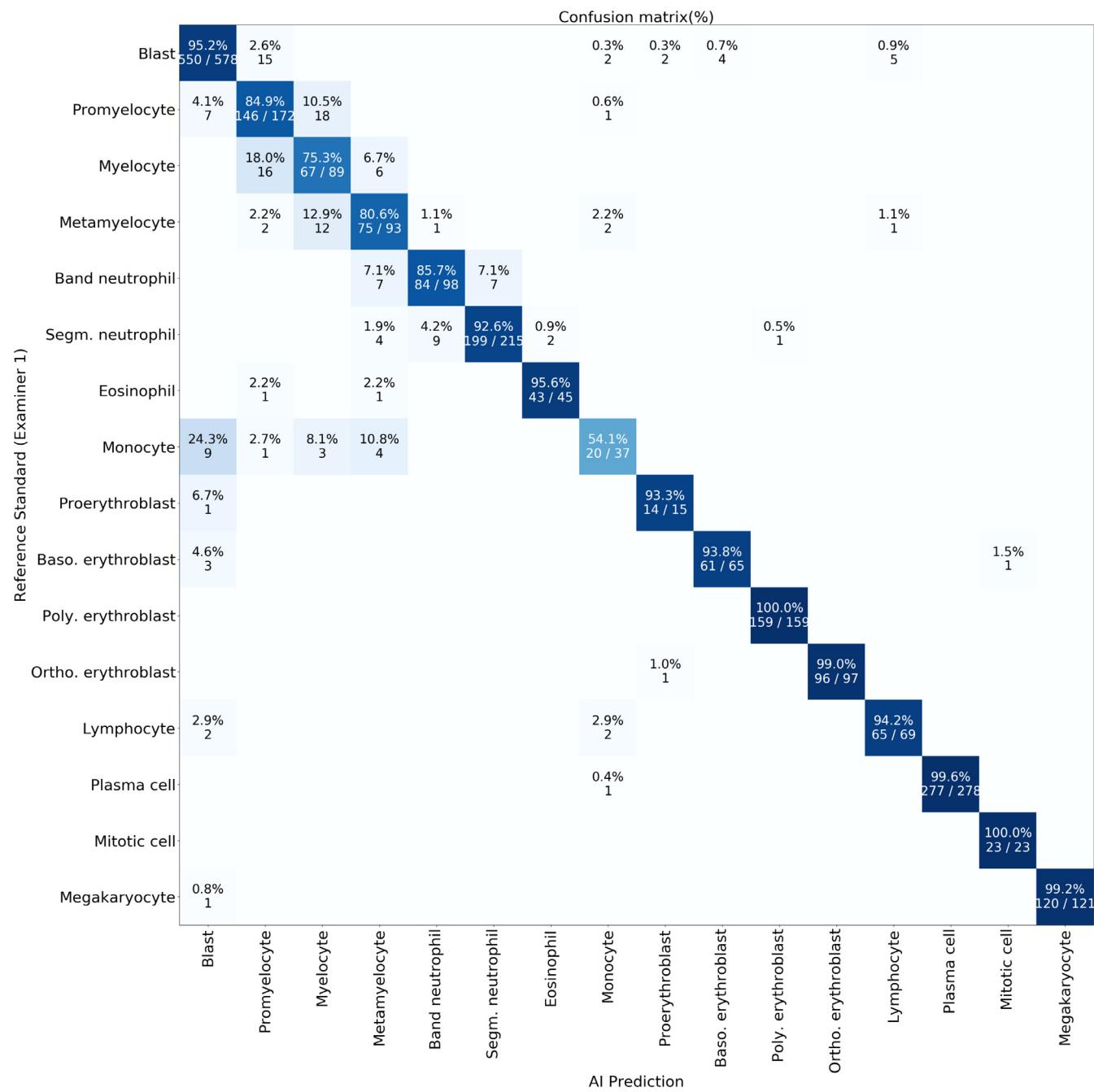
Nevertheless, due to large dimensions of data to process, the automatic analysis using WSIs remains a challenging problem. Several studies try to tackle these difficulties using the combinations of image preprocessing, image segmentation, feature extraction and conventional machine learning methods (Ghosh et al., 2016; Karthikeyan and Poornima, 2017; Mishra et al., 2017; Reta et al., 2015). However, the aforementioned works are impractical for clinical usage due to the limited capability to identify a small number of cell types and to deal with small images instead of WSIs. With growing interests in deep learning applications, there have been several studies that utilized deep learning models to auto-

mate BM NDC. Some classification-based deep learning methods are proposed to deal with only pre-selected single-cell BM cell images, which are manually cropped by examiners (Choi et al., 2017; Kimura et al., 2019; Sahlool et al., 2020). To reduce the workload of manual intervention by examiners, several methods are proposed to deal with pre-selected small images with multiple-cells, manually chosen by medical experts (Chandradevan et al., 2019; Hu et al., 2019; Xie et al., 2018; Yu et al., 2019). For practical clinical use, we develop a fully automatic hierarchical deep learning framework, which is capable of automatically detect the regions of BM cellular trails of WSIs, without any examiner-provided annotation. More importantly, the proposed method is able to process WSI analysis fast and efficiently in seconds while achieving state-of-the-art performance in BM NDC analysis without manual intervention.

Our framework has the capability to identify BM particles and cellular trails and perform BM cell detection-classification inside BM particles to automate BM NDC using WSI analysis in seconds. The proposed method is capable of differentiating 16 BM cell types, including blast, promyelocyte, myelocyte, metamyelocyte, band neutrophil, segmented neutrophil, eosinophil, monocyte and promonocyte, proerythroblast, basophilic erythroblast, polychromatic erythroblast, orthochromatic erythroblast, lymphocyte, plasma cell, mitotic cell and megakaryocyte. With the success of BM NDC on WSIs, our framework possess the potential to be utilized in pathology and cytology digital slides in different specimens, especially in finding suspicious cells to support the diagnosis and follow-up of malignant diseases.

Our study is the first to explore megakaryocyte, mitotic cell and detailed stages of erythroblasts due to their clinical significance. Although not included in the BM NDC (Lee et al., 2008), the megakaryocyte cell is a major BM component and required in evaluating the quality of BM aspirate and the condition of hematopoiesis, especially for patients receiving chemotherapy or differential diagnosis of thrombocytopenia. The mitotic cell is also vital to provide the value of growth rate, which is an important feature in malignant diseases. Finally, the detailed stages of erythroblast could aid in diagnosing dyserythropoiesis (such as MDS) and pure erythroid leukemia (a rare leukemia type), which relies on the proportion of erythroid cells and specific proerythroblasts in all BM nucleated cells (Swerdlow et al., 2017).

In evaluation, the proposed method achieves a recall more than 0.9 for 11 out of 17 BM cell types. The proposed method also surpasses the benchmark methods (Chandradevan et al., 2019; Yu et al., 2019) with a recall of  $0.905 \pm 0.078$  and an accuracy of



**Fig. 10.** Inter-observer analysis between AI and examiner1 with detailed cell number for each cell type.

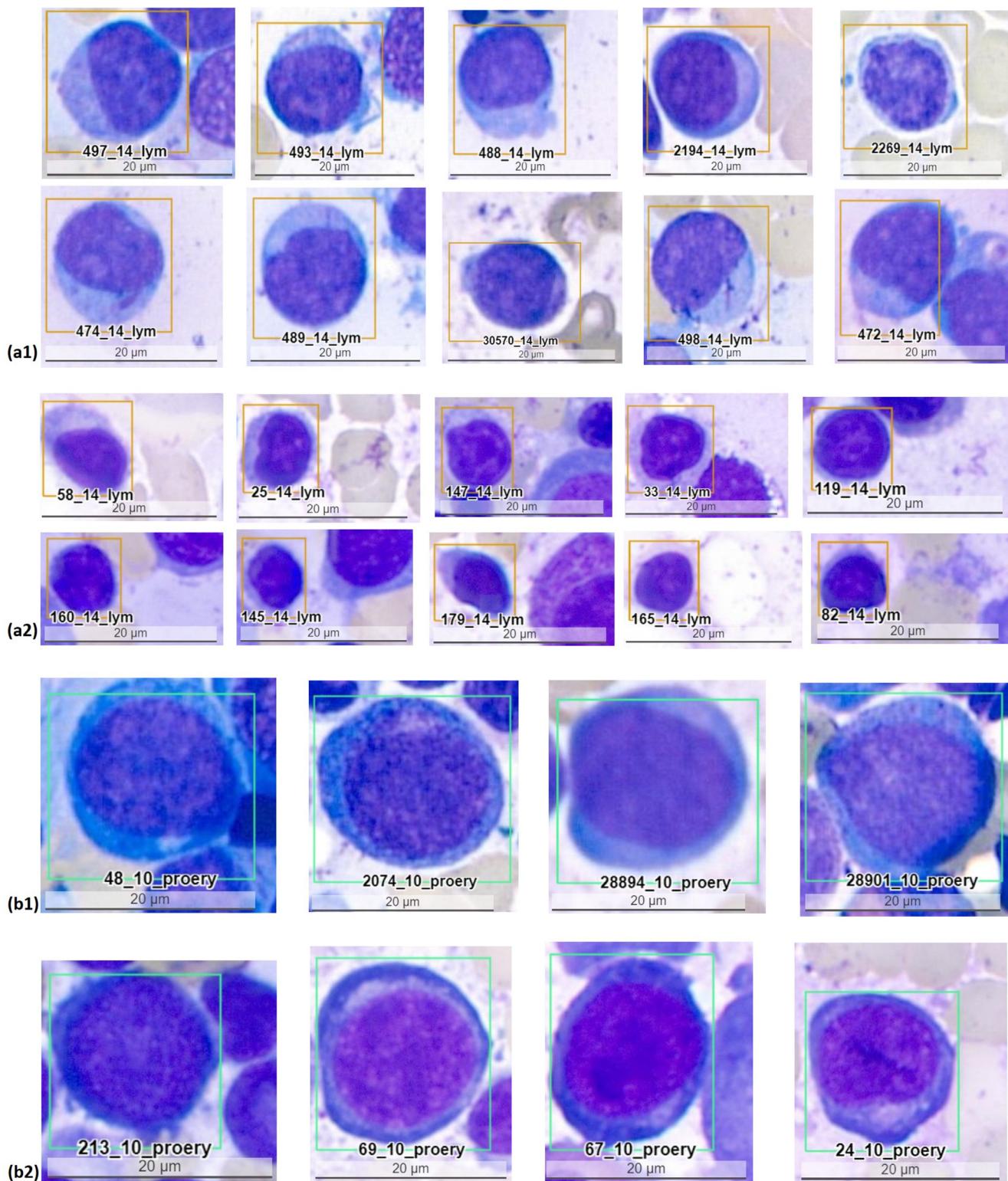
$0.989 \pm 0.006$ . In this study, basophil is not included as the number of annotations is relatively low.

We postulate that both the rarity of basophil in BM particles and the dissolved granules of basophil make the recognition task difficult.

By increasing the number of annotations of basophil and incorporating peripheral blood smear, we shall overcome the limitation.

To examine the generalizability of our model, we performed an evaluation on a second independent dataset. However, the recall of four types of BM cells, blast, promyelocyte, myelocyte and monocyte, dropped below 0.8, but accuracy remained excellent (**all above 0.95**). Promyelocyte, myelocyte and metamyelocyte, which are all belong to myeloid precursors, represent successional maturation stages that share overlapping cytomorphological features. Therefore, some cells show ambiguous phenotypes, which are dif-

ficult to be classified affirmatively even for a well-trained medical expert. By adding more cells for training, the uncertainty shall be overcome. Blast, which is in fact a broad and heterogeneous category, includes precursors from different lineages, including myeloblast, lymphoblast, monoblast/promonocyte and occasionally bizarre lymphoma cells. Blasts are a group of immature cells defined by high nucleus to cytoplasm (N:C) ratio, finely reticular chromatin with distinct nucleoli and basophil cytoplas, and the size of a blast cell varies dramatically, ranging from 10 to 25  $\mu\text{m}$  depending on the cell origin (Glassy, 1998). Hence, blasts could be misidentified as other cell types in the following circumstances: (1) Cells are not well spread, resulting in the pseudo-high-N:C phenotype. By adjusting the AI model to select well-spreading regions of smear slide, this mistake shall be eliminated. (2) Some blasts, which are smaller in size and contain more abun-



**Fig. 11.** Misidentified cases of blasts by the proposed model in the 2nd dataset and samples of misidentified types with similar looks in the 1st dataset. (a1) Blasts misidentified as lymphocytes, (a2) Lymphocytes in the 1st dataset, (b1) Blasts misidentified as proerythroblasts, (b2) Proerythroblasts in the 1st dataset.

dant cytoplasm, could be misidentified as lymphocytes as shown in Fig. 11. By recognizing their finely chromatin, the presence of nucleoli and larger cellular size of blasts (compared to lymphocyte, 6.2–7.4 μm (Kuse et al., 1985)), this misidentification shall be avoided. (3) Proerythroblast, the “blast” form of erythroid lineage, is not included to blast due to its distinct maturation pathway.

Due to its immaturity, proerythroblast also shares common features with other blast, including large cellular size, high N:C ratio and basophilic cytoplasm. The key distinct phenotype of proerythroblast is that its nucleus is in the center of cell and the nucleoli is dark blue, compared with other blasts, which contain light gray nucleoli as shown in Fig. 11. Furthermore, in the first dataset,

there are only 89 proerythroblasts, which is definitely inadequate for AI training. By increasing annotations of proerythroblast, our AI model shall be able to discriminate these two cell types for future work. Monocyte, which has indented nucleus, abundant gray blue cytoplasm, sometimes with granules or vacuoles in cytoplasm, and is also variable in size (12–20  $\mu\text{m}$ ) (Glassy, 1998), truly possess a great challenge in the recognition task. Looking into the detail of misidentification, the major confusion comes from lymphocyte. By emphasizing the different size between monocyte (6.2–7.4  $\mu\text{m}$ ) and lymphocyte (12–20  $\mu\text{m}$ ), this problem shall be solved.

For future work, we would like to explore possible data normalization strategies as it may be helpful to perform data normalization not only w.r.t. to mean but also w.r.t. to variance. Another improvement could be made for the future work is to modify the proposed method with the processing of overlapping patches as the current framework does not integrate mismatching classification.

Currently, the major bottleneck in digital pathology is the lengthy slide scanning process. Without assistance of artificial intelligence, a slide scanner blindly scans the whole area in each slide and costs approximately 50 minutes for a slide of 21.83  $\times$  41.59 mm<sup>2</sup>, producing a WSI with the pixel resolution of 253,070  $\times$  482,247 pixels and the file size of 8 GB. This bottleneck could be resolved by integrating the proposed AI models to automatically capture and analyze data in regions of interests and greatly reducing the total scanning and data analysis time together. The time for scanning and analysis of the low resolution 512  $\times$  512 image for the first layer model is less than 1 second in total. By integrating the proposed hierarchical AI models, the first layer model would filter out most of unnecessary regions for further scanning or analysis by rapidly locating the cellular trails and particles, which usually account for 5% ~ 18% region in a WSI and contain thousands of cells far more than 500 cells required according to ICSH guidelines (Lee et al., 2008), and thus reduce scanning time for high resolution tiles down to 5% ~ 18%. In other words, the scanning time would be reduced to 2.5 to 9 mins. The second layer model would run in parallel during scanning. Overall, the total scanning and data analysis time together could be reduced to 2.5 to 9 mins per slide.

## 6. Conclusion

Fully automatic BM WSI analysis can bring the revolutionized change in labor-intensive and time-consuming manual BM NDC by dramatically diminishing the required time from at least 30 minutes to 5 minutes per patient and providing correct counting results without inter-observer bias. In this way, it can markedly shorten the time required from initial presentation to diagnosis and therefore enable the physician to grasp the optimal treatment timing. In addition, it also shows great value in clinical utility, for example, distinguishing AML from MDS and CMMI by differentiating BM cells component; aiding in the risk stratification of MDS via blast percentage; determining treatment response in patients with AML after receiving chemotherapy by whether blast percentage is above five or not; clarifying the etiology of thrombocytopenia through evaluation the number and morphology of megakaryocytes.

In this paper, we present an efficient and fully automatic bone marrow WSI analysis under 40x objective magnification with the advantage of fast localization of BM particles and cellular trails, high-performance BM cell detection-classification and fast stitching model for integrating multi-type patch-based detection. In evaluation, the proposed method demonstrates superior performance compared to the benchmark methods (Chandrasevan et al., 2019; Yu et al., 2019), which only handle small manually selected ROI images and require human intervention by examiners. Comprehensive experiments demonstrate that the proposed method has the dis-

criminative representative ability in BM NDC analysis on WSIs and achieves state-of-the-art performance.

In the future, by integrating results of other studies, such as BM biopsy, immunophenotyping, cytogenetics and gene mutation, diagnosis and treatment modality can be made more quickly and more precisely.

## Program Availability

The program and a WSI data of this study with a video demo could be downloaded from GitHub repository (<https://github.com/cwwang1979/BoneMarrow>), enabling the readers to run a WSI BM NDC analysis on a real WSI slide.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Ching-Wei Wang:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Sheng-Chuan Huang:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing. **Yu-Ching Lee:** Investigation, Visualization. **Yu-Jie Shen:** Methodology, Software, Investigation, Visualization. **Shwu-Ing Meng:** Data curation. **Jeff L. Gaol:** Writing – original draft, Investigation.

## Acknowledgments

This research is supported by the Ministry of Science and Technology of Taiwan under grants (MOST 109-2221-E-011-018-MY3) and Tri-Service General Hospital-National Taiwan University of Science and Technology, Taipei, Taiwan (TSGH-NTUST-111-05).

## References

- Aziz, L., bin Haji Salam, S., Ayub, S., 2020. Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: a comprehensive review. *IEEE Access*.
- Bain, B.J., Bates, I., Lafsan, M.A., 2016. *Dacie and Lewis Practical Haematology E-Book: Expert Consult: Online and Print*. Elsevier Health Sciences.
- Bennett, J.M., Catovsky, D., Daniel, M.-T., Flandrin, G., Galton, D.A., Gralnick, H.R., Sultan, C., 1976. Proposals for the classification of the acute leukaemias french-american-british (fab) co-operative group. *Br. J. Haematol.* 33 (4), 451–458.
- Birkhead, B., Salt, D., Jackson, R., 1987. Reliability and consistency of bone marrow assessments of complete remission in acute myelogenous leukaemia. *Hematol. Oncol.* 5 (2), 139–145.
- Buckens, C.F., de Jong, P.A., Mol, C., Bakker, E., Stallman, H.P., Mali, W.P., van der Graaf, Y., Verkooijen, H.M., 2013. Intra and interobserver reliability and agreement of semiquantitative vertebral fracture assessment on chest computed tomography. *PLoS ONE* 8 (8), e71204.
- Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P.E., Verrill, C., Wallander, M., Lundin, M., Haglund, C., Lundin, J., 2018. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* 8 (1), 1–11.
- Cai, Z., Vasconcelos, N., 2019. Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* doi:[10.1109/TPAMI.2019.2956516](https://doi.org/10.1109/TPAMI.2019.2956516).
- Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Silva, V.W.K., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* 25 (8), 1301–1309.
- Chandrasevan, R., Aljudi, A., Drumheller, B., Kunananthaseelan, N., Amgad, M., Gutman, D., Cooper, L., Jaye, D., 2019. Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells. *Lab. Invest.* 100. doi:[10.1038/s41374-019-0325-7](https://doi.org/10.1038/s41374-019-0325-7).
- Chlebus, G., Schenk, A., Moltz, J.H., van Ginneken, B., Hahn, H.K., Meine, H., 2018. Automatic liver tumor segmentation in ct with fully convolutional neural networks and object-based postprocessing. *Sci. Rep.* 8 (1), 1–7.

- Choi, J.W., Ku, Y., Yoo, B.W., Kim, J.-A., Lee, D.S., Chai, Y.J., Kong, H.-J., Kim, H.C., 2017. White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. *PLoS ONE* 12 (12).
- Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyo, D., Moreira, A.L., Razavian, N., Tsirigos, A., 2018. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24 (10), 1559–1567.
- De Fauw, J., Ledsam, J.R., Romero-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentini, D., et al., 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* 24 (9), 1342–1350.
- Deng, J., Dong, W., Socher, R., Li, L., Kai Li, Li Fei-Fei, 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. doi:10.1109/CVPR.2009.5206848.
- Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F.R., Büchner, T., Dombovy, H., Ebert, B.L., Fenaux, P., Larson, R.A., et al., 2017. Diagnosis and management of aml in adults: 2017 eln recommendations from an international expert panel. *Blood* 129 (4), 424–447.
- Fuentes-Arderiu, X., Dot-Bach, D., 2009. Measurement uncertainty in manual differential leukocyte counting. *Clinical Chemistry and Laboratory Medicine (CCLM)* 47 (1), 112–115.
- Gao, Y., Bu, X., Hu, Y., Shen, H., Bai, T., Li, X., Wen, S., 2018. Solution for large-scale hierarchical object detection datasets with incomplete annotation and data imbalance arXiv preprint arXiv:1810.06208.
- Ghosh, P., Bhattacharjee, D., Nasipuri, M., 2016. Blood smear analyzer for white blood cell counting: a hybrid microscopic image analyzing technique. *Appl. Soft. Comput.* 46, 629–638.
- Gianelli, U., Bossi, A., Cortinovis, I., Sabattini, E., Tripodo, C., Boveri, E., Moro, A., Valli, R., Ponzoni, M., Florena, A.M., et al., 2014. Reproducibility of the who histological criteria for the diagnosis of philadelphia chromosome-negative myeloproliferative neoplasms. *Mod. Pathol.* 27 (6), 814–822.
- Glassy, E.F., 1998. Color Atlas of Hematology: An Illustrated Field Guide Based on Proficiency Testing. College of American Pathologists.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., pp. 2672–2680.
- Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Solé, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F., Kantarjian, H., Kuendgen, A., Levis, A., Malcovati, L., Cazzola, M., Cermak, J., Fonatsch, C., Le Beau, M.M., Slovak, M.L., Krieger, O., Luebbert, M., Maciejewski, J., Magalhaes, S.M.M., Miyazaki, Y., Pfeilstocker, M., Sekeres, M., Sperr, W.R., Stauder, R., Tauro, S., Valent, P., Vallespi, T., van de Loosdrecht, A.A., Germing, U., Haase, D., 2012. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* 120 (12), 2454–2465. doi:10.1182/blood-2012-03-420489.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:10.1109/CVPR.2016.90.
- Hu, B., Tang, Y., Chang, E.I., Fan, Y., Lai, M., Xu, Y., 2019. Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. *IEEE J. Biomed. Health Inform.* 23 (3), 1316–1328. doi:10.1109/JBHI.2018.2852639.
- Ianni, J.D., Soans, R.E., Sankarapandian, S., Chamarthi, R.V., Ayyagari, D., Olsen, T.G., Bonham, M.J., Stavish, C.C., Motaparthi, K., Cockerell, C.J., et al., 2020. Tailored for real-world: a whole slide image classification system validated on uncurated multi-site data emulating the prospective pathology workload. *Sci. Rep.* 10 (1), 1–12.
- Iizuka, O., Kanavati, F., Kato, K., Rambeau, M., Arihiro, K., Tsuneki, M., 2020. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci. Rep.* 10 (1), 1–11.
- Karthikeyan, T., Poornima, N., 2017. Microscopic image segmentation using fuzzy c means for leukemia diagnosis. *Leukemia* 4 (1), 3136–3142.
- Kimura, K., Tabe, Y., Ai, T., Takehara, I., Fukuda, H., Takahashi, H., Naito, T., Komatsu, N., Uchihashi, K., Ohsaka, A., 2019. A novel automated image analysis system using deep convolutional neural networks can assist to differentiate mds and aa. *Sci. Rep.* 9 (1), 1–9.
- Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Veit, A., Belongie, S., Gomes, V., Gupta, A., Sun, C., Chechik, G., Cai, D., Feng, Z., Narayanan, D., Murphy, K., 2017. Openimages: A public dataset for large-scale multi-label and multi-class image classification. Dataset available from <https://github.com/openimages>.
- Kumar, S., Paiva, B., Anderson, K.C., Durie, B., Landgren, O., Moreau, P., Munshi, N., Lonial, S., Bladé, J., Mateos, M.-V., et al., 2016. International myeloma working group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncol.* 17 (8), e328–e346.
- Kuse, R., Schuster, S., Schübbe, H., Dix, S., Hausmann, K., 1985. Blood lymphocyte volumes and diameters in patients with chronic lymphocytic leukemia and normal controls. *Blut* 50 (4), 243–248.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444.
- Lee, S.-H., Erber, W., Porwit, A., Tomonaga, M., Peterson, L., Hematology, I.C.S.I., 2008. ICSH guidelines for the standardization of bone marrow specimens and reports. *Int. J. Lab. Hematol.* 30 (5), 349–364.
- Lin, H., Chen, H., Graham, S., Dou, Q., Rajpoot, N., Heng, P., 2019. Fast scannet: fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection. *IEEE Trans. Med. Imaging* 38 (8), 1948–1958. doi:10.1109/TMI.2019.2891305.
- Liu, J., Liu, H., Tang, Z., Gui, W., Ma, T., Gong, S., Gao, Q., Xie, Y., Niyyotida, J.P., 2020. Iouc-3dscnn: segmentation of brain tumors via iou constraint 3d symmetric full convolution network with multimodal auto-context. *Sci. Rep.* 10 (1), 1–15.
- Liu, Y., Yang, F., Hu, P., 2020. Small-object detection in uav-captured images via multi-branch parallel feature pyramid networks. *IEEE Access* 8, 145740–145750.
- 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.
- Meintker, L., Ringwald, J., Rauh, M., Krause, S.W., 2013. Comparison of automated differential blood cell counts from abbott sapphire, siemens advia 120, beckman coulter dxh 800, and sysmex xe-2100 in normal and pathologic samples. *Am. J. Clin. Pathol.* 139 (5), 641–650.
- Mishra, S., Majhi, B., Sa, P.K., Sharma, L., 2017. Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection. *Biomed Signal Process Control* 33, 272–280.
- Pal, K.K., Sudeep, K., 2016. Preprocessing for image classification by convolutional neural networks. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE, pp. 1778–1781.
- Peikari, M., Gangeh, M.J., Zubovits, J., Clarke, G., Martel, A.L., 2016. Triaging diagnostically relevant regions from pathology whole slides of breast cancer: a texture based approach. *IEEE Trans. Med. Imaging* 35 (1), 307–315. doi:10.1109/TMI.2015.2470529.
- 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 (6), 1137–1149. doi:10.1109/TPAMI.2016.2577031.
- Reta, C., Altamirano, L., Gonzalez, J.A., Diaz-Hernandez, R., Peregrina, H., Olmos, I., Alonso, J.E., Lobato, R., 2015. Segmentation and classification of bone marrow cells images using contextual information for medical diagnosis of acute leukemias. *PLoS ONE* 10 (6).
- Sahil, A.T., Kollmannsberger, P., Ewees, A.A., 2020. Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. *Sci. Rep.* 10 (1), 1–11.
- Sun, Y., Zhu, S., Ma, K., Liu, W., Yue, Y., Hu, G., Lu, H., Chen, W., 2019. Identification of 12 cancer types through genome deep learning. *Sci. Rep.* 9 (1), 1–9.
- Swerdlow, S., Campo, E., Harris, N.L., Jaffe, E., Pilger, S., Stein, H., Thiele, J., Arber, D., Hasserjan, R., Le Beau, M., et al., 2017. Who classification of tumours of haematopoietic and lymphoid tissues (revised 4th edition). IARC: Lyon 421.
- Tabe-Bordbar, S., Emad, A., Zhao, S.D., Sinha, S., 2018. A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models. *Sci. Rep.* 8 (1), 1–11.
- Wilkins, B.S., Erber, W.N., Bareford, D., Buck, G., Wheatley, K., East, C.L., Paul, B., Harrison, C.N., Green, A.R., Campbell, P.J., 2008. Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes. *Blood* 111 (1), 60–70. doi:10.1182/blood-2007-05-091850.
- Wu, Z., Bodla, N., Singh, B., Najibi, M., Chellappa, R., Davis, L.S., 2019. Soft sampling for robust object detection. *BMVC* 225.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995. doi:10.1109/CVPR.2017.634.
- Xie, Y., Xing, F., Shi, X., Kong, X., Su, H., Yang, L., 2018. Efficient and robust cell detection: a structured regression approach. *Med. Image. Anal.* 44, 245–254.
- Yu, T.-C., Chou, W.-C., Yeh, C.-Y., Yang, C.-K., Huang, S.-C., Tien, F.M., Yao, C.-Y., Cheng, C.-L., Chuang, M.-K., Tien, H.-F., Zhang, Q.-Y., Hsu, W.-H., Chou, S.-C., 2019. Automatic bone marrow cell identification and classification by deep neural network. *Blood* 134. doi:10.1182/blood-2019-125322. 2084–2084
- Zhang, Z., Chen, P., McGough, M., Xing, F., Wang, C., Bui, M., Xie, Y., Sapkota, M., Cui, L., Dhillon, J., et al., 2019. Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. *Nat. Mach. Intell.* 1 (5), 236–245.