*Microscopy and Microanalysis*, 2024, **00**, 1–9
https://doi.org/10.1093/micmic/ozad143

**Original Article**

Microscopy AND Microanalysis

# Digital Microscopy Augmented by Artificial Intelligence to Interpret Bone Marrow Samples for Hematological Diseases

**David Bermejo-Peláez[1], Sandra Rueda Charro[1], María García Roa[2], Roberto Trelles-Martínez[2], Alejandro Bobes-Fernández[2], Marta Hidalgo Soto[3], Roberto García-Vicente[4,5], María Luz Morales[4,5], Alba Rodríguez-García[4,5], Alejandra Ortiz-Ruiz[4,5], Alberto Blanco Sánchez[6], Adriana Mousa Urbina[1], Elisa Álamo[1], Lin Lin[1,7], Elena Dacal[1], Daniel Cuadrado[1], María Postigo[1], Alexander Vladimirov[1], Jaime Garcia-Villena[1], Andrés Santos[7,8], María Jesús Ledesma-Carbayo[7,8], Rosa Ayala[4,5,6], Joaquín Martínez-López[4,5,6,*], María Linares[4,5,9,*], and Miguel Luengo-Oroz[1,*]**

[1]*Spotlab, P.º de Juan XXIII, 36B, Madrid 28040, Spain*
[2]*Department of Hematology, Hospital Universitario Fundación Alcorcón, C. Budapest, 1, Alcorcón 28922, Madrid, Spain*
[3]*Vall Hebron Institute of Oncology (VHIO), Carrer de Natzaret, 115-117, Horta-Guinardó, Barcelona 08035, Spain*
[4]*Department of Translational Hematology, Research Institute Hospital 12 de Octubre (imas12), Av. de Córdoba, s/n, Madrid 28041, Spain*
[5]*Hematological Malignancies Clinical Research Unit H120-CNIO, CIBERONC, C. de Melchor Fernández Almagro, 3, Madrid 28029, Spain*
[6]*Department of Hematology, Hospital Universitario 12 de Octubre, Av. de Córdoba, s/n, Madrid 28041, Spain*
[7]*Biomedical Image Technologies Laboratory, ETSI Telecomunicación, Universidad Politécnica de Madrid, Av. Complutense, 30, Madrid 28040, Spain*
[8]*CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, C. de Melchor Fernández Almagro, 3, Madrid 28029, Spain*
[9]*Department of Biochemistry and Molecular Biology, Pharmacy School, Universidad Complutense de Madrid, Pl. de Ramón y Cajal, s/n, Madrid 28040, Spain*

***Corresponding author:** Miguel Luengo-Oroz, E-mail: miguel@spotlab.org; María Linares, E-mail: mlinares@ucm.es; Joaquín Martínez-López, E-mail: jmarti01@med.ucm.es

## Abstract

Analysis of bone marrow aspirates (BMAs) is an essential step in the diagnosis of hematological disorders. This analysis is usually performed based on a visual examination of samples under a conventional optical microscope, which involves a labor-intensive process, limited by clinical experience and subject to high observer variability. In this work, we present a comprehensive digital microscopy system that enables BMA analysis for cell type counting and differentiation in an efficient and objective manner. This system not only provides an accessible and simple method to digitize, store, and analyze BMA samples remotely but is also supported by an Artificial Intelligence (AI) pipeline that accelerates the differential cell counting process and reduces interobserver variability. It has been designed to integrate AI algorithms with the daily clinical routine and can be used in any regular hospital workflow.

**Key words:** artificial intelligence (AI), bone marrow aspirates, differential cell counting (DCC), digital microscopy

## Introduction

Morphological analysis of cells in samples of bone marrow aspirates (BMAs) is a crucial part of hematological diagnosis that can guide disease management. Visual inspection of samples through light microscopy has been traditionally used for morphological assessment, ensuring accuracy of the diagnosis and thus the subsequent management of the disease affected by a number of factors, including the experience and capabilities of the personnel performing the visual examination of samples. The examination of BMA and peripherical blood smears by Artificial Intelligence (AI) represents a clear opportunity to automate, standardize, and quantify cytomorphological analyses (Fan et al., 2022; The Lancet Haematology, 2022).

In particular, the analysis of BMA samples is a common and essential process for several hematological diseases (acute leukemias, plasma cell disorders, myelodysplastic syndromes, etc.) (Arber et al., 2016; Swerdlow et al., 2016), and despite current technological advances, it still heavily rely on traditional optical microscopes and clinical expertise and the subjectivity of hematologists.

As has been stated in several studies, BMA analysis shows a high level of interobserver variability (Bentley, 1990; Fuentes-Arderiu & Dot-Bach, 2009; Font et al., 2015; Choi et al., 2017). Furthermore, the International Council for Standardization in Hematology guideline for the standardization of bone marrow (BM) specimens and reports suggests extending the number of cells to be counted to more than 500, or even comparing the results with those of another sample and asking another observer to evaluate the sample independently, especially when a disease is suspected. In short, the differential cell count (DCC) in BMA samples has proved to be a time-consuming and error-prone procedure (Hodes et al., 2019)

**Received:** July 7, 2023. **Revised:** November 15, 2023. **Accepted:** December 22, 2023

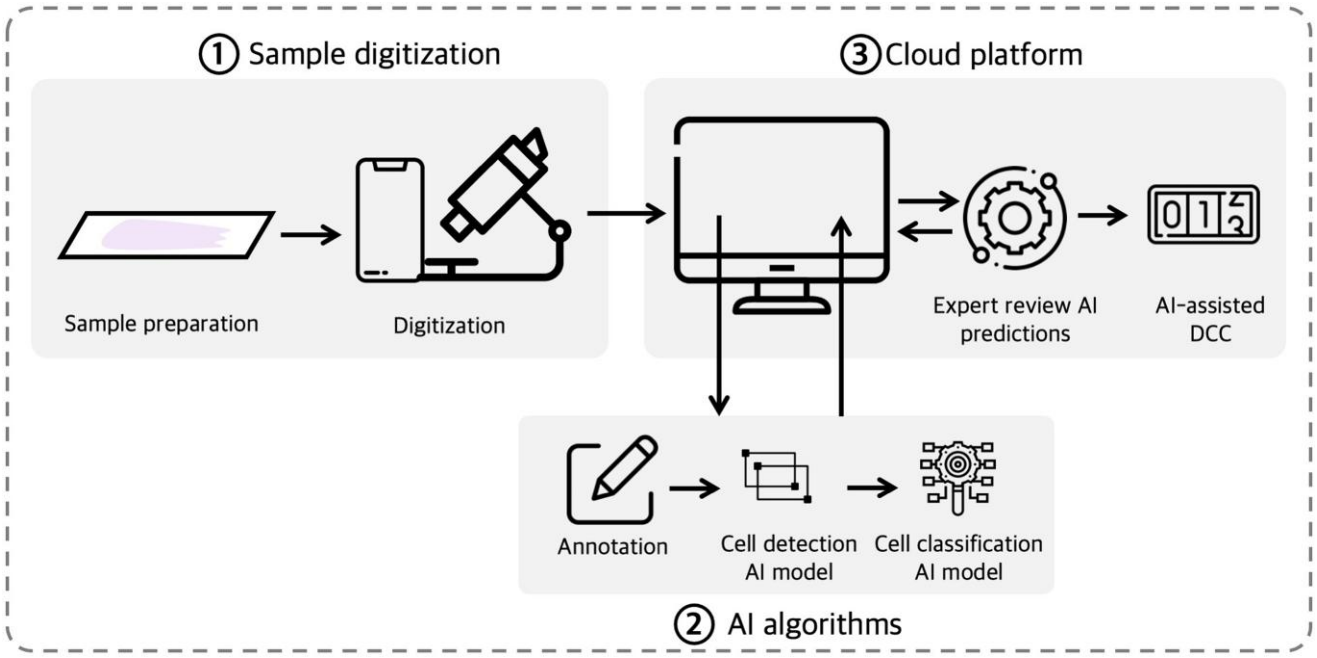


**Fig. 1.** A workflow of the system, composed of three main components: (1) digitization, (2) AI algorithms, and (3) a cloud visualization platform.

that requires a highly specialized professional in cytology (hematologists or hematopathologists) to ensure reliable results (WHO, 2017).

Recent studies have proposed AI-based tools to assist professionals in detecting and classifying nucleated cells in BMA samples. Traditional machine learning techniques based on handcrafted features from cell images have been proposed; however, these methods may be limited by the labor-intensive feature extraction process, the subjectivity of feature selection (Theera-Umpon & Dhompongsa, 2007; Liu et al., 2019), and low robustness. To address these limitations, more recent approaches have turned to convolutional neural networks (CNNs) to perform a morphological examination of BMA samples. The following have been proposed: the use of dual-stage CNNs to classify nucleated cells in BMA samples (Choi et al., 2017), Faster R-CNNs (Liu et al., 2022) together with VGG16 CNN (Chandradevan et al., 2020) and Xception CNN (Eckardt et al., 2022) for cell classification, ResNet50 networks (Wang et al., 2022*a*, 2022*b*), ResNetXt-50 to classify BMA cells in an extensive expert-annotated dataset (Matek et al., 2021), Mask R-CNN for cell detection and segmentation (Ouyang et al., 2021), DenseNet to automatically select appropriate regions for BMA cytology together with YOLO network to detect and classify cellular and noncellular object in the sample (Tayebi et al., 2022), semisupervised deep learning models to overcome the need for large labeled datasets for training (Nakamura et al., 2022), CNNs specially designed to tackle the class imbalance problem (Guo et al., 2022), CoAtNet for BMA morphology classification merging CNNs and transformers models (Tripathi et al., 2022), and a YOLOX-s model for cell detection together with a specific CNN that utilizes multi-level features for cell classification (Wang et al., 2023).

However, despite all this effort, most prior studies rely on the need for complex and expensive commercially available scanners to digitize BMA whole-slide images, also failing in considering how these tools are proposed to be implemented in an actual clinical practice that allows interaction between professional and suggested results by AI.

In this work, we aimed to develop and evaluate an integrated microscopy digital system to cover the entire process, from BMA sample digitization without the need for complex digitization devices to DCC facilitated by human–AI interaction. The proposed system is based on a simple three-dimensional (3D) printed device that couples a smartphone to a conventional optical microscope, allowing standardized and easy acquisition of microscopy images. It also includes a web-based telemedicine platform that integrates the analysis from AI algorithms. By taking advantage of this holistic digital system, it could be possible to implement the solution in any hematology department in the world without incorporating specific and complex medical electronic devices into the clinical workflow.

## Materials and Methods

### System

The workflow of the AI-assisted digital DCC system is presented in Figure 1. The pipeline is as follows: first, the user digitizes a BMA sample using a 3D-printed microscopy arm that allows attaching a smartphone to a conventional optical microscope. The user uses a specific mobile app and digitizes at least 20 different fields (which always will contain at least 500 cells for differential counting) of the sample using a 100× objective. It should be noted that this system can also be easily used for the digitization of a larger number of fields and therefore the analysis of a larger number of cells, which can lead to a much more accurate DCC. In order to accelerate the process that lasts a few minutes and make image acquisition easier, the user uses a foot pedal that triggers the image capture, so that both users' hands are free to operate the microscope. The acquired images are automatically uploaded through the mobile app to a web telemedicine platform, where they are subsequently processed by an AI algorithm that automatically detects and differentiates all nucleated cells in the sample into six possible cell lineages. The user can review AI predictions and confirm or edit them in case they disagree with the AI model. Last, cell series proportions are calculated by generating the DCC.

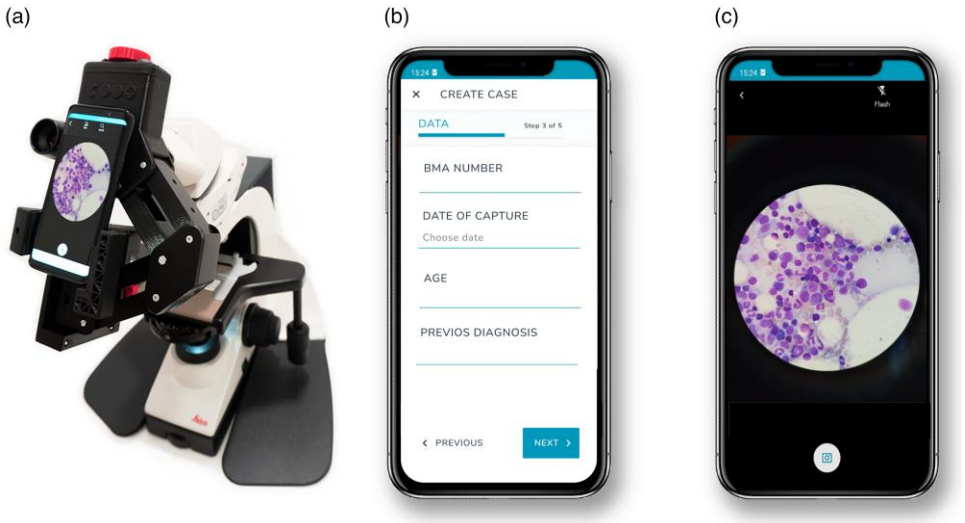


**Fig. 2.** A sample digitization system. (**a**) A three-dimensional–printed microscopy adapter arm. (**b**, **c**) Screenshots of the mobile app for standardized BMA image acquisition, which allows collecting metadata associated to each BMA sample.

## *Digitization*

To digitize BMA samples, the system includes a 3D-printed device that allows coupling and aligning a smartphone's camera with a conventional optical microscope's eyepiece lens. In this way, this device converts any optical microscope into a digital one and can be used to obtain digital images of BMA samples without the need for additional and expensive devices for capturing BMA images. The smartphone uses a mobile app developed to acquire patient metadata, which was previously customized specifically for fast, standardized, and easy digitization of BMA microscopy images (Fig. 2). The system complies with privacy regulations and follows all requirements from General Data Protection Regulation. In addition, high-security industry standards have been implemented, and all data collected through the mobile app can be encrypted and anonymized.

Users employed this digitization system to acquire photographs of 20 microscope fields for each BMA sample, ensuring that more than 500 cells were digitized. All acquired images were uploaded from the mobile app to a web-based telemedicine platform that allowed a remote visualization and analysis of BMA images. Patient data such as age and previous diagnosis were also collected through the mobile app and transferred to the telemedicine platform.

BMA samples used for training the AI algorithm were digitized using a BQ Aquaris X2 smartphone model, while samples used for the evaluation of the entire AI-assisted digital DCC system were digitized using two different smartphone models (BQ Aquaris X2 and Xiaomi Redmi Note 8T), to increase the variability in terms of mobile models and reduce the possible bias of this variable. All images were acquired at an image resolution of 12 megapixels. The smartphones were attached with the 3D-printed device to the ocular of a light microscope (Leica DM 2000 LED) and using a 100× objective (1,000× total magnification).

## *AI Algorithm*

The AI algorithm was developed for the automatic identification and differentiation of nucleated cells in BMA images into seven different classes or lineages: myeloid, erythroid, monocytic, lymphoid, blasts (myeloblasts, monoblasts, and megakaryoblasts), plasma cells, and artifacts (cells without optimal cytologic characteristics). Examples of each cell class are shown in Figure 3. The algorithm consisted of a two-stage cell detection and lineage classification deep learning–based model. First, it detects all nucleated cells without regard to the cell lineage. The detected cells are subsequently introduced into the classification model, which can distinguish between different cell lineage classes. More detailed information about each algorithm for cell detection and differentiation can be found in the Development and Training of AI Algorithms subsection.

Once an entire image representing a microscope field of view is processed by the AI algorithm, cell series proportions are calculated automatically for all the six lineage classes under study. The proportion of the "artifact" class was not reported, as it has no clinical relevance.

## *AI-assisted Web Telemedicine Platform*

All acquired images with the mobile app are transferred to the cloud telemedicine platform, where images can be visualized in an easy-to-use dashboard, allowing scrolling and zooming of the images. This platform also allows image labeling, which experts can use to digitally analyze BMA samples. In addition, these manually generated labels can be used for training the algorithms.

On the other hand, the developed AI algorithm is embedded into the platform, so that all uploaded images are processed by the model, and AI annotations are visualized. The platform includes a review terminal, which is designed for experts to review results produced by the AI and confirm or edit AI suggestions in case they disagree with them by either reclassifying misdifferentiated cells in the image or labeling new undetected cells. A comparison between AI predictions and reviewed classification can be performed to assess AI model performance and usability of the system. This process is designed to speed up the diagnosis process as the user must only review and modify those cells with which they disagree, rather than manually differentiating and counting the required 500 cells per sample.

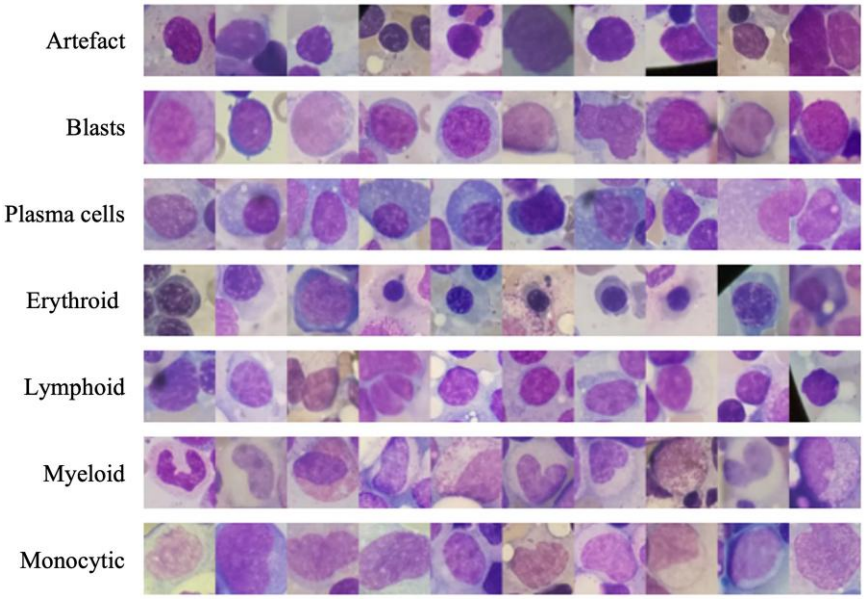


**Fig. 3.** Examples of seven different morphological classes of BMA cells stained with May–Grünwald–Giemsa and digitized at 100× magnification.

## Study Design and Sample Collection

Two different datasets were collected for this study. The first one was collected for the development of an AI algorithm for the automatic identification and classification of hematopoietic stem cells in BMA samples. This dataset consisted of 101 BMA randomly selected samples that were retrospectively extracted between 2019 and 2021 at the Haematology Service of the University Hospital 12 de Octubre (Madrid, Spain). The age of included patients ranged from 1 to 87 (mean 57.5 years old).

Second, we validated the entire AI-based digital system for assisting hematologists to perform a DCC. For this purpose, a second dataset was collected, which was composed of 16 BMA samples that were extracted between December 2021 and February 2022 at the same center, with an age range from 2 to 82 years old (mean of 55.5 years). The BMA samples were selected considering that different hematological diseases should be represented to validate the algorithm. Patient characteristics of both study cohorts are presented in Table 1.

All BMA samples were prepared according to standard protocols and using May–Grünwald–Giemsa staining. We assessed sample preparation quality to discard those BMA samples without proper quality staining, those with insufficient lumps, or those with a certain level of dysplasia.

The ethical approval for the study was obtained from the Hospital 12 de Octubre Ethics Committee for Research with medicinal products (Num. Ref. 20/430). This study was conducted in accordance with the Declaration of Helsinki; all BMA samples used in this study were anonymized according to local guidelines, and informed consent was obtained from all subjects.

**Table 1.** Patient Characteristics of the Study Cohort.

| | Development of the AI Algorithm | System Evaluation |
|---|---|---|
| Age | | |
| Mean ± SD (years) | 57.56 ± 15.71 | 55.5 ± 19.89 |
| Gender | | |
| Female | 57 (60.64%) | 9 (56.25%) |
| Male | 37 (39.36%) | 7 (43.75%) |
| Diagnosis | | |
| Acute lymphoblastic leukemia | 6 (6.38%) | 2 (12.5%) |
| Acute myeloid leukemia | 20 (21.28%) | 4 (25%) |
| Chronic lymphocytic leukemia | 2 (2.13%) | 1 (6.25%) |
| Chronic myeloid leukemia | 5 (5.32%) | 1 (6.25%) |
| Diffuse large B-cell lymphoma | 4 (4.26%) | 1 (6.25%) |
| Follicular lymphoma | 5 (5.32%) | — |
| Multiple myeloma and other plasma cells dyscrasias | 33 (35.11%) | 5 (31.25%) |
| Other lymphomas | 7 (7.45%) | 1 (6.25%) |
| Reactive bone marrow | 5 (5.32%) | — |
| Others | 7 (7.45%) | 1 (6.25%) |

## Development and Training of AI Algorithms

The cell labeling process for training the AI algorithms was performed using the web-based telemedicine platform, where expert hematologists manually annotated individual cells using a point-based annotation tool by placing a label point corresponding to one of the seven possible classes (six different cell lineages and artifact class) in the center of each cell. These annotated images were used for training and validating the classification algorithm. Additionally for training the cell detection algorithm, cells were annotated without regard to the cell class by manually drawing a bounding box around all cells present in a given image (microscope field). Figure 4 shows an example of an annotation procedure for both cases (bounding boxes for training the cell detection algorithm and point-based labels for training the cell classification algorithm), and both of these training activities were performed in the web telemedicine platform.

For training the classification algorithm, a total of 61,344 cells were identified and classified with one of the classes under study by a panel of 3 different hematologists (randomly

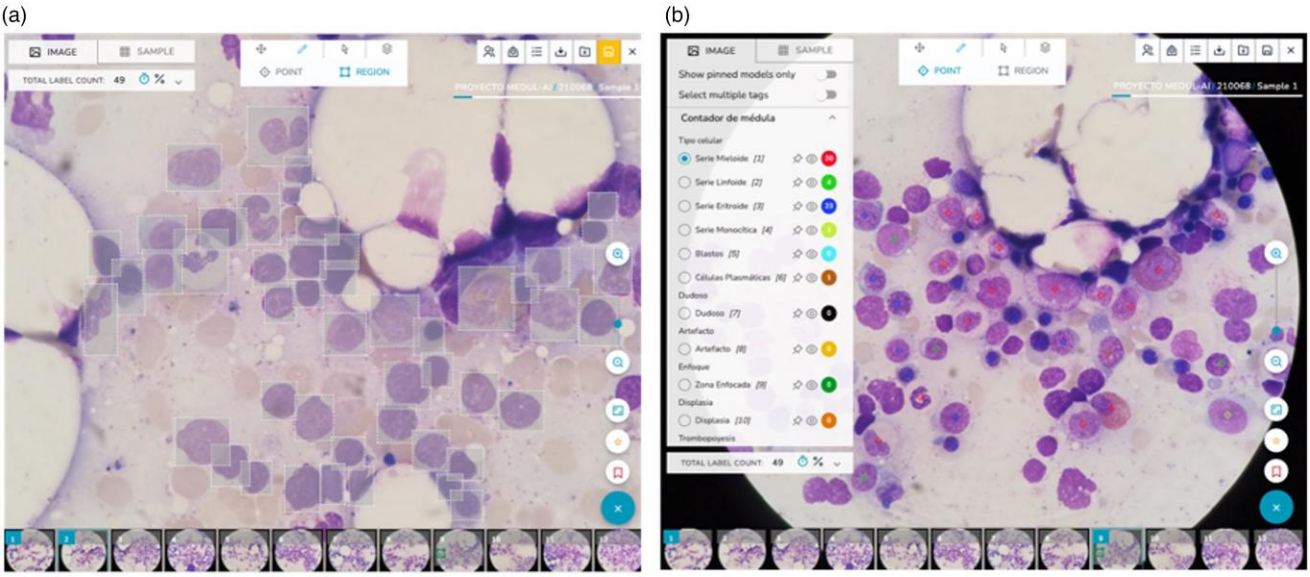


**Fig. 4.** Screenshots of a web telemedicine platform for image labeling. (**a**) Manually placed bounding boxes around all cells in a given image for training the detection algorithm. (**b**) Point-based annotations for training the classification algorithm.

selected from the 6 who participated in the study) according to 7 different classes: myeloid, erythroid, monocytic, lymphoid, blasts, plasma cells, and artifact (cells without optimal cytologic characteristics). From these labels and in order to train the algorithm only with those cells with full agreement among observers, we discarded 4,171 cells, and thus, we only used 57,173 labels for training the algorithm, with the following distribution: myeloid ($n = 13{,}497$), erythroid ($n = 7{,}164$), monocytic ($n = 1{,}055$), lymphoid ($n = 2{,}793$), blasts ($n = 2{,}444$), plasma cells ($n = 1{,}105$), and artifact ($n = 29{,}115$). From these 57,173 total labels, 80% were used for training the classification algorithm, while the remaining 20% was used for a preliminary assessment of AI performance.

On the other hand, a total of 13,494 cells annotated in the form of bounding boxes were used for training ($n = 4{,}358$) and validating ($n = 9{,}136$) the detection algorithm.

The proposed algorithm for cell detection was based on the single-shot detection (SSD) (Liu et al., 2016) architecture together with the MobileNet V2 (Sandler et al., 2018) backbone network. The SSD constitutes a powerful object detection technique that employs a feed-forward CNN to predict object classes and their corresponding bounding boxes simultaneously, eliminating the need for region proposal networks found in other two-stage object detection approaches such as Faster R-CNN. This results in a faster and more efficient detection pipeline while maintaining high accuracy levels. On the other hand, MobileNet V2, serving as the backbone network, is based on depthwise separable convolutions, which factorize standard convolutions into separate operations handling spatial and channel-wise correlations. Further enhanced by the introduction of intermediate expansion layers and linear bottleneck layers, MobileNet V2 provides improved accuracy and efficiency compared with standard convolutional networks. Both networks have been designed to be lightweight and computationally efficient, and their combination is well-suited for embedding into real-time clinical tools.

The detection network detected all cells appearing in a given image (microscope field) without regard to the cell lineage. The detected cells were subsequently processed by a classification algorithm for assigning them to one of the seven possible classes.

For cell differentiation, small image patches of size $200 \times 200$ pixels (approximately equivalent to $16 \times 16\ \mu m^2$) were extracted around the center of each detected cell's bounding box. These patches were then classified using the Xception convolutional network architecture (Chollet, 2017), which is built upon the core idea of depthwise separable convolutions similar to MobileNet V2, but with structural differences that enable improved feature learning and representation. To strengthen the classification algorithm's robustness and improve accuracy, data augmentation techniques were applied. These included rotation, horizontal, and vertical flips to account for variations in cell orientation and perspective. Moreover, brightness, contrast, and color modifications were introduced to mimic potential discrepancies arising from different microscope and smartphone setups used for sample digitization, as well as variations in sample staining procedures. No additional preprocessing step was performed in images to train the algorithms.

All algorithms were implemented in Python (version 3.8.10) using the TensorFlow framework (version 2.11), and training was conducted in a workstation equipped with an NVIDIA Tesla T4 16GB GPU.

## Evaluation of the System

For comparative purposes and to evaluate the entire AI-assisted digital DCC system, all 16 BMA samples were analyzed in the telemedicine platform by 4 different expert hematologists. For each sample preparation, one hematologist performed the digital analysis of the sample in a blinded fashion without the assistance of the AI (i.e., labeling and performing the DCC from scratch), while the other three hematologists analyzed the same sample assisted by the AI algorithm. All hematologists were asked to analyze and count at least 500 nucleated cells on each aspirate smear, following international recommendations on the diagnosis of hematology disorders (WHO, 2017; Acosta et al., 2022).

**Table 2.** Performance of the AI Algorithm for Detecting and Classifying Cells in BMA Samples.

| | 20% Validation Set | | Independent Validation Set | | Total Validation | | |
|---|---|---|---|---|---|---|---|
| | Accuracy | *n* | Accuracy | *n* | Accuracy (95% CI) | | *n* |
| Lymphoid | 86.56 | 424 | 86.82 | 258 | 86.66 | (84.10–89.21) | 682 |
| Erythroid | 96.18 | 1,284 | 92.43 | 1,030 | 94.51 | (93.58–95.44) | 2,314 |
| Myeloid | 98.17 | 2,565 | 98.25 | 2,286 | 98.21 | (97.83–98.58) | 4,851 |
| Monocytic | 56.10 | 205 | 67.63 | 139 | 60.76 | (55.60–65.92) | 344 |
| Blasts | 87.42 | 469 | 82.60 | 638 | 84.64 | (82.52–86.77) | 1,107 |
| Plasma cells | 79.70 | 197 | 58.00 | 50 | 75.30 | (69.93–80.68) | 247 |
| **Average** | 93.35 | 5,144 | 92.52 | 4,401 | **92.97** | **(92.46–93.48)** | **9545** |

*n*, number of cells used for evaluating the performance.
Average values across the different cell lineages are highlighted in bold.

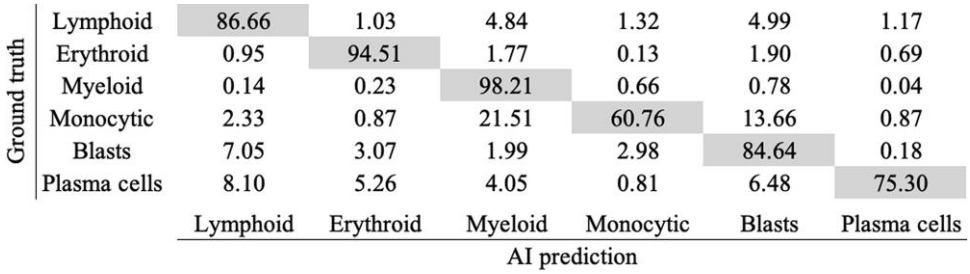


**Fig. 5.** Performance evaluation of the AI algorithm. A confusion matrix comparing ground truth and AI predictions.

Hematologists who were assisted by the AI were presented with the predictions made by the algorithm and were asked to review each of the classified cells and relabel those incorrectly classified or undetected. The platform only displays the predictions of those cells that have been detected by the first detection algorithm and that were subsequently classified by the classification algorithm as a class other than an artifact. The user can always add a label to a cell that has not been labeled by the AI. Those hematologists who did not have the assistance of AI manually labeled from scratch at least 500 cells from each preparation.

The selection of the hematologists who were assisted by the AI and who performed the blinded analysis was done on a rotating basis.

The relative percentages of each cell series for all preparations were calculated and compared for both methods (digital DCC and AI-assisted DCC).

The time needed to complete the analysis of a BMA sample when AI assistance was used and when the analysis was performed in a blinded fashion was measured. Analysis times were compared to assess whether the AI-assisted system could reduce DCC diagnosis time. Additionally, as each of the BMA samples was analyzed by four experienced hematologists, we also quantified the interobserver variability by measuring the agreement in the cell classification among different experts.

Last, we also assessed the performance of the AI algorithm on these 16 BMA-independent samples by comparing the AI predictions against a consensus labeling defined as the majority among the 4 experts for each cell.

## Results

From the 61,344 cells manually identified and classified along the 101 BMA samples collected for algorithm development, a total of 5,144 cells were separated for preliminary performance evaluation (20% of the available images). Cells annotated as artifact were not included in this analysis.

On the other hand, each of the 16 BMA samples for evaluating the entire system was analyzed by 4 different experienced hematologists. Each expert identified and classified at least 500 cells for each sample. Experts placed 7,840, 7,922, 7,908, and 8,375 labels, respectively, among the 16 BMA samples. From these labels, only 4,401 corresponded to the same cell, so that each of these cells was classified by the 4 experts, in addition to the AI algorithm.

### Performance of the AI Algorithm

The detection algorithm achieved a high overall accuracy for detecting BM nucleated cells in microscopy images, with a sensitivity of 91.6% (95% CI 90.6–92.7%) and a precision of 91.3% (95% CI 90.2–92.3%).

Two data sets were used to evaluate the performance of the classification model. First, the performance was evaluated on the validation set (20% of the training set), which included the above-mentioned 5,144 cells. The algorithm was also evaluated on an independent test set obtained from the 16 BMA samples used for evaluating the entire system comprising 4,401 cells classified by 4 different experts, where the ground truth was established using the majority voting rule (consensus) among the 4 experts. The total validation set consisted of 9,545 cells. Table 2 shows the average performance of the model in the total validation set as well as detailed performance of each cell lineage class for both validation sets independently. The confusion matrix for the total validation set (9,545 cells) is shown in Figure 5. The overall accuracy was 92.97%, although it is worth noting the decrease in performance in those classes with fewer training images, such as monocytic and plasma cells. However, this fact is easily remedied by increasing the number of training images for these classes.

**Table 3.** Cohen's Kappa Score Representing Interobserver Agreement and Agreement Between Observers and AI Predictions.

| | Obs 1 | Obs 2 | Obs 3 | Obs 4 | AI |
|---|---|---|---|---|---|
| Obs 1 | — | 0.920 | 0.892 | 0.892 | 0.865 |
| Obs 2 | | — | 0.897 | 0.895 | 0.856 |
| Obs 3 | | | — | 0.876 | 0.889 |
| Obs 4 | | | | — | 0.873 |
| Average | 0.895 | | | | 0.871 |

**Table 4.** Interobserver Agreement When BMA Samples are Analyzed in a Blinded Fashion (Left) and When Hematologists Are Assisted by AI (Right).

| Observers | Blinded Kappa (95% CI) | AI-Assisted Kappa (95% CI) | *p*-Value |
|---|---|---|---|
| 1 versus 2 | 0.91 (0.90–0.93) | 0.93 (0.91–0.94) | <0.0001**** |
| 1 versus 3 | 0.88 (0.87–0.90) | 0.89 (0.88–0.91) | <0.0001**** |
| 1 versus 4 | 0.88 (0.87–0.90) | 0.91 (0.89–0.92) | <0.0001**** |
| 2 versus 3 | 0.88 (0.86–0.90) | 0.91 (0.90–0.92) | <0.0001**** |
| 2 versus 4 | 0.86 (0.84–0.88) | 0.92 (0.90–0.93) | <0.0001**** |
| 3 versus 4 | 0.86 (0.83–0.87) | 0.89 (0.88–0.91) | <0.0001**** |
| Mean | 0.88 (0.85–0.92) | 0.93 (0.89–1) | <0.0001**** |

****indicates statistical significance.

## Evaluation of the System

### Interobserver Variability

The BMA analysis presents a high level of interobserver variability when differentiating hematopoietic stem cell lineages. Cohen's kappa score was used to evaluate interobserver agreement in the classification of cell lineages. The agreement was performed on 4,401 BM cells from the 16 cases used for evaluation annotated by 4 experts. Table 3 presents the kappa results for all pairs of observers as well as for the agreement between each observer and the AI predictions. The mean interobserver agreement was 0.895. A similar agreement was found when comparing experts and AI predictions, with a mean kappa score of 0.871.

On the other hand, to assess whether the assistance of the AI can reduce the interobserver variability among hematologists when analyzing BMA samples, we have computed interobserver agreement on those samples that were analyzed in a blinded fashion, as well as on samples where hematologists were assisted by the AI. Table 4 shows the comparison between both interobserver agreement values (blinded and AI-assisted analysis), and, as it can be seen, the interobserver agreement is considerably increased when AI is used, from 0.88 to 0.93. The difference in interobserver agreement between both methods is statistically significant ($p$-value < 0.0001).

### Analysis Time

To assess whether AI assistance can reduce analysis time, we have measured the time needed to complete a BMA analysis when users were assisted by AI and when they had to perform the analysis from scratch (i.e., without the assistance of AI).

Table 5 summarizes the average time needed to analyze a single cell for each of the 4 involved hematologists, as well as the time needed to complete an entire BMA analysis, which usually comprises the identification of at least 500 cells. As derived from the table, it is shown that when users used the AI assistance, the analysis time was reduced by a factor of 18.75%.

### Usability Assessment

As part of the objectives of this work, we wanted to evaluate the usability of the proposed AI-assisted digital DCC system. For this purpose, we designed an anonymized usability survey that was completed by the four hematologists who used the proposed system. The participants showed a high level of satisfaction in using the system (average score of 4.59 out of 5) and considered it a useful tool in their daily work. Additionally, the results derived from the usability survey show that the automatic hematopoietic stem cell lineage percentages predicted by the AI algorithm were sufficiently reliable, even though some cells were misclassified and had to be modified in the review process.

## Discussion

The use of AI in the field of hematology has the potential to improve the way hematological disorders are diagnosed and treated. The ability of AI-based algorithms to analyze large amounts of data quickly and accurately can greatly improve diagnostic accuracy and efficiency, which can ultimately lead to better patient outcomes. In this context, the digitization of hematology is crucial to take full advantage of AI's capabilities, by providing a large amount of data that can be used for developing and training AI-based algorithms and easy access to the data for analysis and follow-up.

In this work, we presented a digital microscopy system that allows counting cell types from BMAs that integrate AI algorithms with the clinical practice in a regular hospital workflow. The proposed system not only provides an accessible and simple way to digitize, store, and label BMAs remotely but also is supported by an AI pipeline that speeds up the time required for the analysis and reduces the interobserver variability.

Digitizing BMAs has some direct implications as it is possible to easily examine the samples remotely, to share them for second opinions, or to analyze samples from the same patient over time. Furthermore, it allows precise quantitative analysis and research of cell types and morphologies, as well as testing new clinical hypotheses. Some clinical interpretations (clinical remissions and response to treatment) rely on an expert analysis of a limited number of fields or cells, and with the proposed

**Table 5.** Comparison of the Time Needed to Complete an Analysis of a Single Cell (Left) and a Whole BMA Sample (Right) When the User Used the AI Assistance and When the Analysis Was Performed in a Blinded Fashion.

| Time/Cell (s) | | | Time/BMA Sample (500 Cells) (mm:ss) | | | |
|---|---|---|---|---|---|---|
| | Blinded | AI-Assisted | | Blinded | AI-Assisted | Time Reduction (%) |
| Obs 1 | 3.0 | 2.7 | Obs 1 | 25:15 | 22:47 | 10.00 |
| Obs 2 | 4.0 | 2.8 | Obs 2 | 33:18 | 23:18 | 30.00 |
| Obs 3 | 2.8 | 2.5 | Obs 3 | 23:24 | 20:40 | 10.71 |
| Obs 4 | 2.9 | 2.5 | Obs 4 | 23:49 | 20:40 | 13.79 |
| Mean | 3.2 | 2.6 | Mean | 26:26 | 21:52 | 18.75 |

mm, minutes; s/ss, seconds.

system working in an optimized manner, it will be possible to analyze a higher number of cells without increasing the time of analysis, hence being able to measure with more precision.

The study presented in this work has required the creation of an extensive database for training an algorithm able to classify different cell types. It is worth mentioning that all the samples used come from one single hospital, and though the bias depending on sample preparation is limited, further multicentric validation is suggested to understand the bias introduced by the variability of the preparation. It is also worth pointing out that the digitization process enabled by mobile phones has standardized acquisition parameters available in any android smartphone and results in no difference between smartphones. It is essential to acknowledge the potential impact of smartphone camera quality on the performance of AI models, particularly in capturing fine details of cellular structures and textures. A lower quality camera may compromise the ability to discern fine features, potentially affecting predictive accuracy. In light of this consideration, we emphasize that our study utilized smartphones with a resolution of 12 megapixels, which, while considered medium-low by contemporary standards, represents a common and widely available imaging technology. Despite the prevalence of higher resolution cameras, our results demonstrate the robustness of our AI model even under these conditions, showcasing its adaptability to standard and more affordable smartphone cameras commonly found in the market.

Digitalization of microscopy samples based on a smartphone without the requirement of scanners and complex equipment lowers barriers to scale all the benefits of digitalization of BMAs including accuracy, the possibility to share and do remote analysis, and reproducibility.

Efforts to digitize BMA samples and develop AI-based analysis systems have been made (Choi et al., 2017; Liu et al., 2022; Chandradevan et al., 2020; Eckardt et al., 2022; Wang et al., 2022*a*; Wang et al., 2022*b*; Matek et al., 2021; Ouyang et al., 2021; Tayebi et al., 2022; Nakamura et al., 2022; Guo et al., 2022; Tripathi et al., 2022; Wang et al., 2023); however, they faced challenges, often relying on complex and expensive scanners. In this context, our approach employs a cost-effective 3D-printed adapter coupled with smartphones to digitize BMA samples. This innovative setup converts conventional microscopes into smart digital platforms, ensuring accessibility across various settings and lowering barriers to scale all the benefits of digitalization of BMAs, including accuracy, the possibility to share and do remote analysis, and reproducibility. Unlike many previous works, this study also addresses the practical integration of AI into clinical practice. We emphasize the critical human–AI interaction, recognizing the need for seamless collaboration between healthcare professionals and the AI system for effective implementation in real-world scenarios.

By digitizing data and considering real-life systems that go beyond AI algorithms that work in isolation, we show a system with all the required components: in this case not only a two-stage algorithm but also a visualization and interaction tool to integrate human and AI support. The current challenge in the field of AI in the medical field revolves around the integration of AI systems into clinical workflows. To address this challenge, technologists should design systems that optimize clinical outcomes, emphasizing the inherently multidisciplinary nature of these projects.

It is worth noting that the majority of the AI models previously proposed in the literature used to analyze BMA samples are designed to classify not only cell lineages but also cell types at different stages of maturation. Future work should include improving the accuracy of the AI pipeline to include new cell types as well as the ability to distinguish and identify cases with dysplasia, adding other biomarkers, and the unsupervised discovery of relevant patterns and cell morphologies correlated with other types of analysis including genetic data. From a technical perspective, we could foresee an algorithm with higher levels of precision, which would be contingent upon the sufficiency of the training data, allowing near-complete automation of the counting process. From cell counting to disease prediction, digitalization can be key to unlocking a big potential for hematology. With the arrival of large language models in AI (Brown et al., 2020; Kung et al., 2023), which can interpret the latest scientific research together with multimodal patient data (Acosta et al., 2022), we envision a future where data from cytomorphological samples, genetic analyses, and other datasets are combined, and an AI assistant supports hematologists to take the best decision at each moment. Such a system will also require further research on how to create systems that integrate human knowledge with insights obtained by AI, resulting in human–AI feedback loops.

## Availability of Data and Materials

For data-sharing requests, please contact Miguel Luengo-Oroz (miguel@spotlab.ai).

## Author Contributions Statement

D.B.-P., S.R.C., A.M.U., E.A., E.D., M.P., A.S., M.J.L.-C., R.A.D., J.M.-L., M.L., and M.L.-O. conceived and designed the study. M.G.R., R.T.-M., A.B.-F., M.H.S., R.G.-V., M.L.M., A.R.-G., A.O.-R., and A.B.S. contributed, digitized, and analyzed the samples. D.B.-P., L.L., D.C., A.V., and J.G.-V. designed and developed the software. D.B.-P., S.R.C., and M.L.-O. analyzed and interpreted the results. D.B.-P. and S.R.C. wrote the main manuscript text. All authors interpreted the results and revised and edited the manuscript.

## Financial Support

This project has been partially funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 881062; the European Union - NextgenerationEU under the "Plan de Recuperación, Transformación y Resiliencia" of the Spanish Government, the Instituto de Salud Carlos III (ISCIII) under grants PMPTA22/00169, PMPTA22/00088, PMPTA22/00041, PMPTA22/00023, PMPTA22/00101; and the Centro para el Desarrollo Tecnológico y la Innovación (CDTI) under grant EXP 00156466 / IDI-20230066. D.B.-P. was supported by grant PTQ2020-011340/AEI/10.13039/501100011033 funded by the Spanish State Investigation Agency. R.G.-V. holds a Formación de Profesorado Universitario (FPU19/04933) grant from the Ministry of Science, Innovation and Universities of Spain Government. A.R.-G. holds the FEHH 2021 research grant from the Spanish Society of Haematology. L.L. was supported by a predoctoral grant IND2019/TIC-17167 (Comunidad de Madrid, Spain).

## Conflict of Interest

D.B.-P., S.R.C., A.M.U., E.A., L.L., E.D., DC, M.P., A.V., J.G.-V., M.J.L.-C., A.S., M.L., and M.L.-O. hold shares or

phantom shares of Spotlab. The rest of the authors declare that they have no competing interest.

## References


Acosta JN, Falcone GJ, Rajpurkar P & Topol EJ (2022). Multimodal biomedical AI. *Nat Med* **28**(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M & Vardiman JW (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**(20), 2391–2405. https://doi.org/10.1182/blood-2016-03-643544

Bentley SA (1990). Automated differential white cell counts: A critical appraisal. *Baillieres Clin Haematol* **3**(4), 851–869. https://doi.org/10.1016/S0950-3536(05)80138-6

Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C & Amodei D (2020). Language models are few-shot learners. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*. December 2020, Article No: 159, pp. 1877–1901.

Chandradevan R, Aljudi AA, Drumheller BR, Kunananthaseelan N, Amgad M, Gutman DA, Cooper LAD & Jaye DL (2020). Machine-based detection and classification for bone marrow aspirate differential counts: Initial development focusing on nonneoplastic cells. *Lab Invest* **100**(1), 98–109. https://doi.org/10.1038/s41374-019-0325-7

Choi JW, Ku Y, Yoo BW, Kim J-A, Lee DS, Chai YJ, Kong H-J & Kim HC (2017). White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. *PLoS One* **12**(12), e0189259. https://doi.org/10.1371/journal.pone.0189259

Chollet F (2017). Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

Eckardt J-N, Schmittmann T, Riechert S, Kramer M, Sulaiman AS, Sockel K, Kroschinsky F, Schetelig J, Wagenführ L, Schuler U, Platzbecker U, Thiede C, Stölzel F, Röllig C, Bornhäuser M, Wendt K & Middeke JM (2022). Deep learning identifies acute promyelocytic leukemia in bone marrow smears. *BMC Cancer* **22**(1), 201. https://doi.org/10.1186/s12885-022-09307-8

Fan BE, Wang SSY, Natalie Aw MY, Chia MF, Yi Chen DT, Ramanathan K, Wong MS, Ponnudurai K & Winkler S (2022). Artificial intelligence in peripheral blood films: An evolving landscape. *Lancet Haematol* **9**(3), e174. https://doi.org/10.1016/S2352-3026(22)00029-1

Font P, Loscertales J, Soto C, Ricard P, Novas CM-, Martín-Clavero E, López-Rubio M, Garcia-Alonso L, Callejas M, Bermejo A, Benavente C, Ballesteros M, Cedena T, Calbacho M, Urbina R, Villarrubia J, Gil S, Bellón JM, Diez-Martin JL & Villegas A (2015). Interobserver variance in myelodysplastic syndromes with less than 5% bone marrow blasts: Unilineage vs. multilineage dysplasia and reproducibility of the threshold of 2% blasts. *Ann Hematol* **94**(4), 565–573. https://doi.org/10.1007/s00277-014-2252-4

Fuentes-Arderiu X & Dot-Bach D (2009). Measurement uncertainty in manual differential leukocyte counting. *Clin Chem Lab Med* **47**, 112–115. https://doi.org/10.1515/CCLM.2009.014

Guo L, Huang P, Huang D, Li Z, She C, Guo Q, Zhang Q, Li J, Ma Q & Li J (2022). A classification method to classify bone marrow cells with class imbalance problem. *Biomed Signal Process Control* **72**, 103296. https://doi.org/10.1016/j.bspc.2021.103296

Hodes A, Calvo KR, Dulau A, Maric I, Sun J & Braylan R (2019). The challenging task of enumerating blasts in the bone marrow. *Semin Hematol* **56**(1), 58–64. https://doi.org/10.1053/j.seminhematol.2018.07.001

Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-Candido G, Maningo J & Tseng V (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit Health* **2**(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC, Leibe B, Matas J & Sebe N (2016). SSD: Single shot multibox detector. In *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*, Leibe B, Matas J, Sebe N & Welling M. (Eds.), **9905**, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Liu H, Cao H & Song E (2019). Bone marrow cells detection: A technique for the microscopic image analysis. *J Med Syst* **43**(4), 82. https://doi.org/10.1007/s10916-019-1185-9

Liu J, Yuan R, Li Y, Zhou L, Zhang Z, Yang J & Xiao L (2022). A deep learning method and device for bone marrow imaging cell detection. *Ann Transl Med* **10**(4), 208. https://doi.org/10.21037/atm-22-486

Matek C, Krappe S, Münzenmayer C, Haferlach T & Marr C (2021). Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. *Blood* **138**(20), 1917–1927. https://doi.org/10.1182/blood.2020010568

Nakamura I, Ida H, Yabuta M, Kashiwa W, Tsukamoto M, Sato S, Ota S, Kobayashi N, Masauzi H, Okada K, Kaga S, Miwa K, Kanai H & Masauzi N (2022). Evaluation of two semi-supervised learning methods and their combination for automatic classification of bone marrow cells. *Sci Rep* **12**(1), 16736. https://doi.org/10.1038/s41598-022-20651-4

Ouyang N, Wang W, Ma L, Wang Y, Chen Q, Yang S, Xie J, Su S, Cheng Y, Cheng Q, Zheng L & Yuan Y (2021). Diagnosing acute promyelocytic leukemia by using convolutional neural network. *Clin Chim Acta* **512**, 1–6. https://doi.org/10.1016/j.cca.2020.10.039

Sandler M, Howard A, Zhu M, Zhmoginov A, & Chen L-C (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520.

Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Swerdlow SH, Campo E, Harris NL & WHO (2017). *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, vol. 2, 4th edition. WHO Guidelines.

Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD & Jaffe ES (2016). The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**(20), 2375–2390. https://doi.org/10.1182/blood-2016-01-643569

Tayebi RM, Mu Y, Dehkharghanian T, Ross C, Sur M, Foley R, Tizhoosh HR & Campbell CJV (2022). Automated bone marrow cytology using deep learning to generate a histogram of cell types. *Commun Med* **2**(1), 45. https://doi.org/10.1038/s43856-022-00107-6

Theera-Umpon N & Dhompongsa S (2007). Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification. *IEEE Trans Inf Technol Biomed* **11**(3), 353–359. https://doi.org/10.1109/TITB.2007.892694

The Lancet Haematology (2022). Artificial intelligence—Refinement and possibilities. *Lancet Haematol* **9**(1), e1. https://doi.org/10.1016/S2352-3026(21)00379-3

Tripathi S, Augustin AI, Sukumaran R, Dheer S & Kim E (2022). HematoNet: Expert level classification of bone marrow cytology morphology in hematological malignancy with deep learning. *Artif Intell Life Sci* **2**, 100043. https://doi.org/10.1016/j.ailsci.2022.100043

Wang M, Dong C, Gao Y, Li J, Han M & Wang L (2022*b*). A deep learning model for the automatic recognition of aplastic anemia, myelodysplastic syndromes, and acute myeloid leukemia based on bone marrow smear. *Front Oncol* **12**, 844978. https://doi.org/10.3389/fonc.2022.844978

Wang W, Luo M, Guo P, Wei Y, Tan Y & Shi H (2023). Artificial intelligence-assisted diagnosis of hematologic diseases based on bone marrow smears using deep neural networks. *Comput Methods Programs Biomed* **231**, 107343. https://doi.org/10.1016/j.cmpb.2023.107343

Wang C, Wei X-L, Li C-X, Wang Y-Z, Wu Y, Niu Y-X, Zhang C & Yu Y (2022*a*). Efficient and highly accurate diagnosis of malignant hematological diseases based on whole-slide images using deep learning. *Front Oncol* **12**, 879308. https://doi.org/10.3389/fonc.2022.879308