



Research Article

HematoNet: Expert level classification of bone marrow cytology morphology in hematological malignancy with deep learningSatvik Tripathi^{a,*}, Alisha Isabelle Augustin^b, Rithvik Sukumaran^c, Suhani Dheer^d, Edward Kim^e^a College of Computing and Informatics, College of Arts and Sciences, Drexel University, Philadelphia, USA^b College of Engineering, Drexel University, Philadelphia, USA^c College of Computing and Informatics, Drexel University, Philadelphia, USA^d College of Arts and Sciences, Drexel University, Philadelphia, USA^e College of Computing and Informatics, Drexel University, Philadelphia, PA 19104, USA

ARTICLE INFO

Keywords:

Deep learning

Bone marrow

Cytology

ABSTRACT

There have been few efforts made to automate the cytomorphological categorization of bone marrow cells. For bone marrow cell categorization, deep-learning algorithms have been limited to a small number of samples or disease classifications. In this paper, we proposed a pipeline to classify the bone marrow cells despite these limitations. Data augmentation was used throughout the data to resolve any class imbalances. Then, random transformations such as rotating between 0° to 90°, zooming in/out, flipping horizontally and/or vertically, and translating were performed. The model used in the pipeline was a CoAtNet and that was compared with two baseline models, EfficientNetV2 and ResNext50. We then analyzed the CoAtNet model using SmoothGrad and Grad-CAM, two recently developed algorithms that have been shown to meet the fundamental requirements for explainability methods. After evaluating all three models' performance for each of the distinct morphological classes, the proposed CoAtNet model was able to outperform the EfficientNetV2 and ResNext50 models due to its attention network property that increased the learning curve for the algorithm which was represented using a precision-recall curve.

1. Introduction

The human-based examination and characterization of bone marrow (BM) cells is one of the most important yet time expensive procedures in cancerous and non-cancerous hematological conditions [1–4].

The cytomorphologic examination is still a critical initial step in the diagnosis of many intra- and extramedullary illnesses even though numerous advanced procedures like cytogenetics, immunophenotyping, and molecular genetics are now accessible [5,6]. The function of BM cytology, which was created in the 19th century, is still very significant because of its relatively rapid results and extensive availability [7,8]. Microscopic inspection and single-cell morphology categorization are still the primary responsibility of human clinicians due to the difficulty in automating this process. In certain circumstances, such as those involving ambiguous BM smears, the process of manually evaluating the specimens may be arduous and time-consuming [9,10]. It has been observed that examiner classifications are prone to significant inter-and intra variability, which means that the number of high-quality cytolog-

ical exams is constrained since subject matter experts are hard to come by [11–15].

It is also challenging to integrate this procedure with other diagnostic methods that provide more quantitative data since the analysis of individual cell morphologies is qualitative by nature. Few efforts have been made to automate the cytomorphological categorization of BM cells. Hand-crafted single-cell characteristics extracted from digital pictures are often used to categorize cells. Furthermore, most prior research on automated cytomorphologic classification focused on the classification of physiological cell types or peripheral blood smears, restricting their applicability to the classification of leukocytes in the BM for the diagnosis of hematological malignancies [5,16–22]. For BM cell categorization, deep-learning algorithms have been limited to a small number of samples or disease classifications and/or have not made the related data accessible [23–29].

Convolutional neural networks have helped to significantly increase the accuracy of computer vision classification tasks in the last few years [30–35]. These methods have also been used to identify multiple types of

* Corresponding author.

E-mail addresses: st3263@drexel.edu (S. Tripathi), aia43@drexel.edu (A.I. Augustin), rs3673@drexel.edu (R. Sukumaran), sd3589@drexel.edu (S. Dheer), ek826@drexel.edu (E. Kim).

cancer, predict progression of tumor, and classify various types of skin diseases [36–41]. Because of this, the effective use of CNNs for image classification depends on the availability of a substantial quantity of image data and high-quality annotation, which may be difficult to achieve because of the cost associated in collecting labels by medical specialists [42–45]. Human examiners are required to supply the ground truth labels for network training and assessment in instances like cytological inspection of BM, when there is no underlying technological gold standard.

2. Related work

2.1. Deep learning approaches

Deep Learning (DL) is an area of machine learning that uses artificial neural networks to model the brain's structure and function. Deep learning powers numerous artificial intelligence (AI) apps and companies that automate analytical and physical processes. For cancer detection, Oncologists have been using Deep Learning methods to improve patient diagnosis, prognosis, and therapy selection by combining genomic, transcriptomic, and histopathological data. The purpose of DL is to develop decision-making tools to aid cancer researchers in their studies and health professionals in the clinical care of cancer patients [46]. In another case, Tayebi et al. worked on building an end-to-end deep learning-based model for automated bone marrow cytology [47]. A computerized full slide picture of the patient's blood was used to quickly and automatically determine areas appropriate for cytology, which was followed by the identification and classification of all of the blood cells inside those areas. Cell type diversity in bone marrow is quantified by using the Histogram of Cell Types (HCT), a new visual depiction that serves as a cytological "patient fingerprint." A great degree of precision was achieved in the method's area detection [48].

2.2. Hematological malignancy

In a similar research conducted on leukemia cancer, the study compared two leukemia detection methods. The first method was a genomic sequencing method and the second was multi-class classification model. Both employed a Convolutional neural network (CNN) as network design and also used three-way cross-validation to separate their datasets. The findings indicated that the genomic model performed better, with 98 percent accuracy in predicting values, whereas the Multi-class classification model has an accuracy of 81 percent. On the other hand, another research looked at the clinical usefulness of an array-based genome-wide screen in leukemia, as well as the technological obstacles and an interpretive procedure [13].

2.3. Bone marrow morphology

Another study by focused on developing an accurate bone marrow cell identification technique for quantitative analysis. The YOLOv5 network [49], trained by minimizing a new loss function, was used in this study to offer a bone marrow cell identification technique. The suggested new loss function was based on a classification algorithm for detecting bone marrow cells. As per the results, the proposed loss function was beneficial in improving the algorithm's efficiency, and the proposed bone marrow cell identification algorithm outperformed other cell detection techniques [50]. Another research work evaluated how useful flow cytometry, karyotype, and a fluorescence *in situ* hybridization (FISH) panel are in detecting myelodysplastic syndrome in children (MDS). The study concluded that flow cytometry and MDS FISH may be used in conjunction with morphological examination and karyotype to discover anomalies in specific cases [51].

3. Methods

3.1. Dataset

The dataset, acquired by Matek et al., contains 171,375 images from a cohort of 945 patients diagnosed with various hematological diseases at MLL Munich Leukemia Laboratory [52]. The minimum patient age was 18.1 years, and the maximum was 92.2 years. The average patient age, based off of the median, was 69.3 years. The mean age was 65.6 years.

Images of bone marrow smears were stained using May-Grünwald-Giemsa/Pappenheim staining. A brightfield microscope with 40x magnification and oil immersion were used to acquire the images. The original images were 2452x2056 in size. After individual cells were annotated into 21 different classes by morphologists, 250x250 square regions were extracted from the original images, each region containing one annotated cell.

Fig. 1 indicates the classes of each image, as well as an example image of that class and the number of images from the original dataset in that class. The distribution of classes in the dataset can be seen in Fig. 2.

3.2. Preprocessing

In this study we performed data augmentation in order to rectify class imbalances. Data augmentation is a technique used to generate new data from a set of existing data. In the case of images, new data can be created by applying a variety of transformations to an image. Some such transformations are rotations, translations, zooming in or out, noise insertion, cropping, and flipping horizontally or vertically.

Data augmentation is especially effective when used to address uneven distribution of class data. In our case, some of the classes had as few as 8 images, while others had as many as 29,000. Augmenting the data helped address this issue, increasing the number of images we had for classes that were under-represented. This reduces bias towards overrepresented classes.

We augmented the under-represented classes to roughly 20,000 images per class. The random transformations we performed included rotating between 0° and 90°, zooming in/out, flipping horizontally and/or vertically, and translating. Examples of these augmented images are shown in Fig. 3.

We also removed images that were irrelevant to our study. Images classified as 'Artefacts', 'Other', and 'Unidentifiable' were discarded before the augmentation process.

3.3. Models implemented

For comparison, we implemented EfficientNetV2 [53] and ResNeXT50 models [54,55], while our main model for the pipeline is CoAtNet [56]. We chose the CoAtNet because of its Convolution and Attention based model architecture.

3.3.1. ResNeXT50

In the ImageNet Large Scale Visual Recognition Challenge 2016 competition, we employed the ResNeXT-50 architecture built by Xie et al, a successful image classification network that came in second place. 36 Peripheral blood smears have previously been classified using a network-like topology,¹⁴ making it an obvious candidate for BM cell classification. The limited amount of hyperparameters in the ResNet architecture is one of its advantages. The base architecture with 32 cardinalities and a 4d bottleneck width built ResNeXT-50. For ease of reference, this network is referred to as ResNeXT-50 (32x4d). It should be noted that the template's input/output width is set to 256-d and that the feature map's subsampled widths are always doubled.

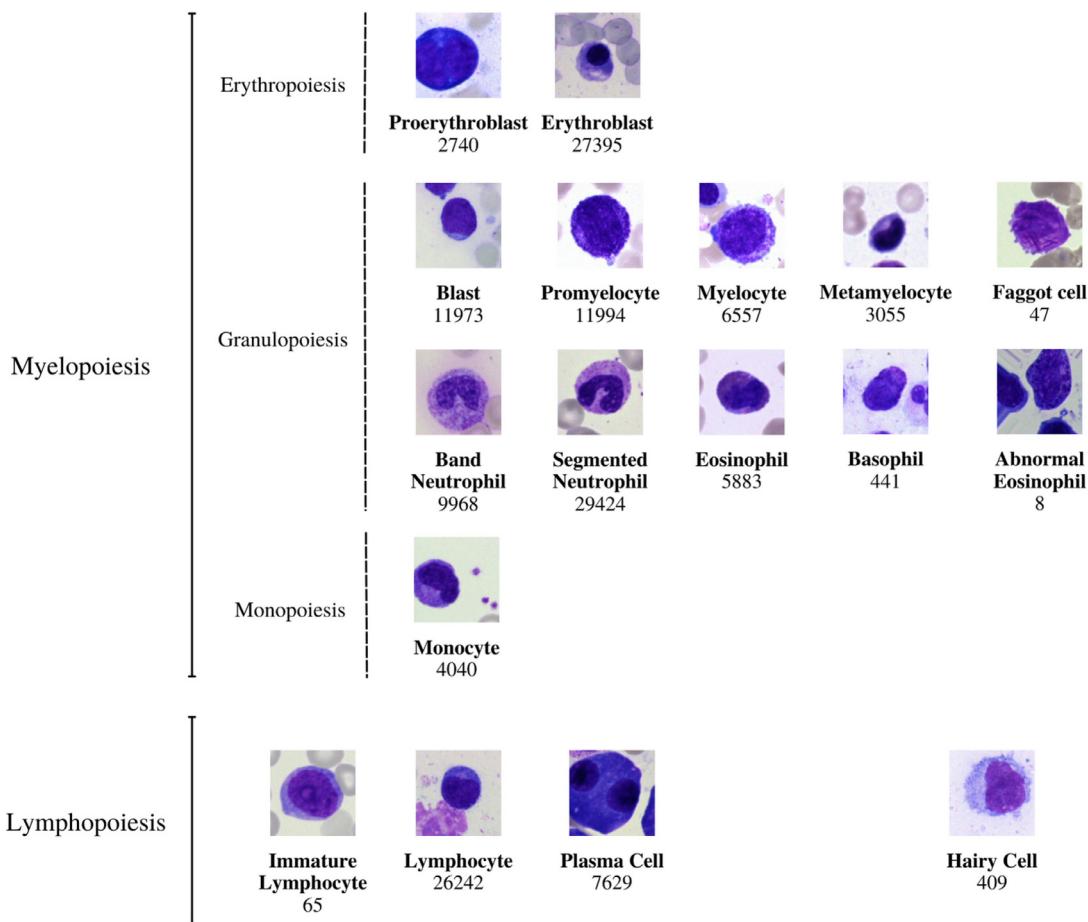


Fig. 1. The 21 morphological classifications of BM cells employed in this investigation have a similar structure. The classes are arranged into hematopoietic lineages in the following order: In accordance with standard practice, the main physiological classes of myelopoiesis and lymphopoiesis, as well as typical pathological classes and, are all included in the classification. As described in further detail in the main text, all cells were stained with the May-Grunwald-Giemsa/Pappenheim stain and photographed at a magnification of 340.

Neurons in artificial neural networks perform inner product which can be thought of as a form of aggregating transformation:

$$\sum_{i=1}^D w_i x_i \quad (1)$$

where $x = [x_1, x_2, \dots, x_D]$ is a D-channel input vector to the neuron and w_i is a filter's weight for the i th channel. This is the elementary transformation that's done by the convolutional and fully-connected layers.

Aggregated transformations are presented as:

$$\mathcal{F}(x) = \sum_{i=1}^C T_i(x) \quad (2)$$

where $T_i(x)$ can be an arbitrary function. Analogous to a simple neuron, T_i should project x into an embedding and then transform it. The cardinality of the network, C , is the size of the set of transformations that will be aggregated can be an arbitrary number. While the dimension of width is related to the number of simple transformations, the dimension of cardinality controls the number of more complex transformations.

$$y = x + \sum_{i=1}^C T_i(x) \quad (3)$$

3.3.2. EfficientNetV2

EfficientNetV2 is a new family of convolutional networks with quicker training speeds and higher parameter economy. These models are developed using a mix of neural architecture search and scaling,

which are both designed to optimize training time and parameter efficiency. The EfficientNetV2 backbone differs significantly from the original EfficientNet in many key ways: MBConv and the new fused-MBConv are utilized extensively in the early layers of EfficientNetV2[53]. It also likes a lower MBConv expansion ratio since smaller expansion ratios often have less memory access overhead. In order to compensate for the decreased receptive field caused by the smaller 3x3 kernel size, EfficientNetV2 adds extra layers. Because of its huge parameters and memory access cost, it is possible that EfficientNetV2 totally eliminates stride-1 in the original EfficientNet.

EfficientNetV1 used MBConv layers with depth-wise convolutions in its design. However, current accelerators typically cannot fully use the reduced parameters and FLOPs of depthwise convolutions. As a result, reducing FLOPs doesn't always translate into faster training. EfficientNetV2 uses MBConv and Fused-MBConv instead of depthwise convolution. The MBConv and Fused-MBConv blocks are shown in the figure. In both blocks, an 1x1 convolution is applied to the SE module. The MBConv block employs an 1x1 conv and a depthwise convolution with a 3x3 filter in the early stages. A 3x3 filter is applied to a single convolution layer created by fusing 1x1 and 3x3 convolution together in the Fused-MBConv block.

3.3.3. CoAtNet

CoAtNet is a hybrid model that merges Convolutional Neural Networks and Transformer models. CoAtNet was developed because it combines the capabilities of both ConvNet and Transformer into a single network. CoAtNe delivers cutting-edge performance with varying data

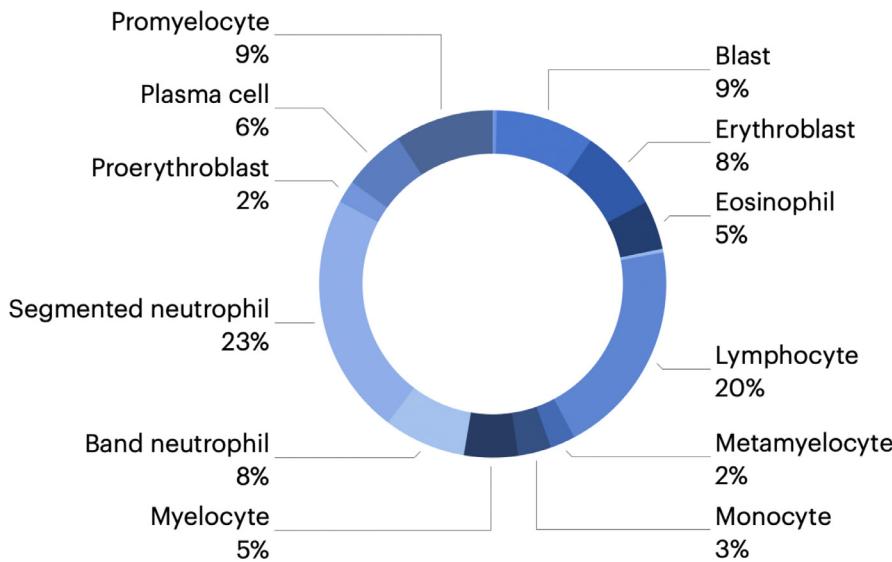


Fig. 2. The pie chart shows the data distribution, we divided up the 171 374 nonoverlapping photos into the several classifications that were utilized.

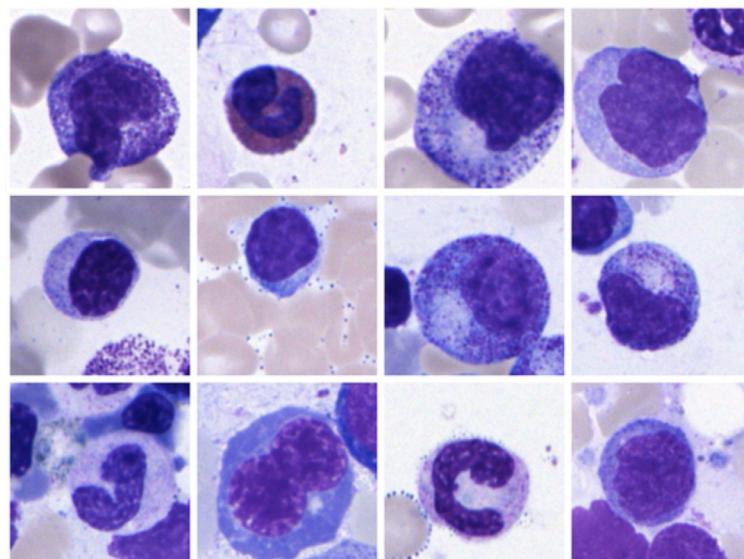


Fig. 3. The augmented data from each classes are shown in the figure. The random transformations we performed included rotating between 0° to 90°, zooming in/out, flipping horizontally and/or vertically, and translating.

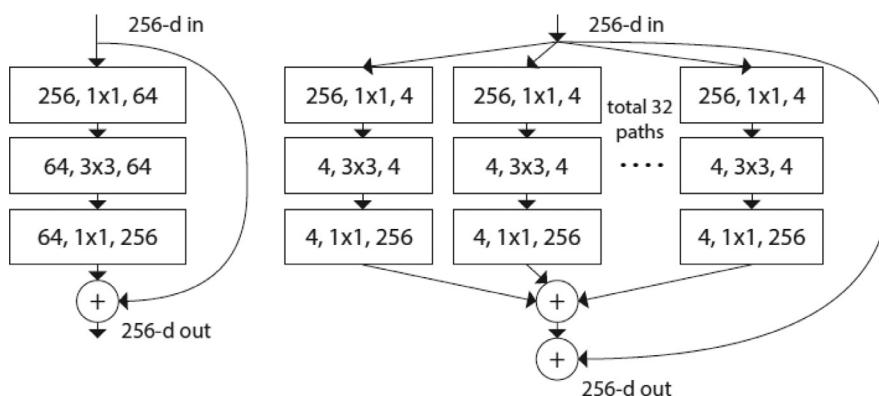


Fig. 4. ResNet [57] is shown on the left. A ResNeXt block with cardinality = 32 and nearly the same complexity as the previous block is shown on the right. A layer is represented as (number of in channels, filter size, number of out channels).

volumes while using the same resource. Inductive biases in CoAtNet allow it to generalize like ConvNets. In addition, CoAtNet benefits from better transformer scalability and quicker convergence, increasing its efficiency.

Generalization and model capacity are the two main elements from which hybridizing convolution and attention in machine learning is

examined. The research demonstrates that convolutional layers have greater generalization while attention has higher model capacity. We can get greater generalization and capacity by merging convolutional and attention layers.

This hybrid model is more focused on image classification and is based on two key features: (a) Depthwise convolution and self-attention

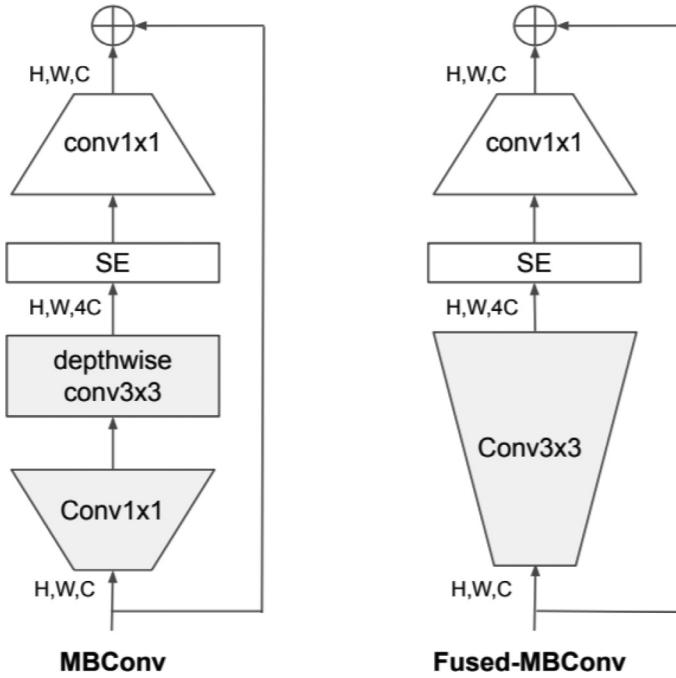


Fig. 5. The MBConv and Fused-MBConv blocks are shown in the diagram above.

may readily be combined using simple relative attention. (b) By stacking convolution layers and attention layers in a logical way, generalization, capacity, and efficiency are greatly increased.

The model architecture, shown in Fig. 6, consist of convolution and self-attention operations. The convolutional layer reduces the dimensionality of the input. The MBconv blocks are a type of image residual block that has an inverted structure. The first MBconv block expands the input by 4x before performing a depthwise convolution to capture the spatial interaction and the second block will compress it before adding a residual. Depthwise convolution performs convolution for each channel separately can be expressed as:

$$y_i = \sum_{j \in \mathcal{L}(i)} w_{i-j} \odot x_j \quad (4)$$

where $x_i, y_i \in \mathbb{R}^D$ are the input and output at position i respectively, and $\mathcal{L}(i)$ denotes a local neighborhood of i , e.g., a 3×3 grid centered at i in image processing. For the self-attention blocks and feed-forward network (FNN) module uses the same expansion-compression structure, similar to the MBConv blocks. In comparison, self-attention allows the receptive field to be the entire spatial locations and computes the weights based on the re-normalized pairwise similarity between the pair (x_i, x_j) :²

$$y_i = \sum_{j \in \mathcal{G}} \underbrace{\frac{\exp(x_i^\top x_j)}{\sum_{k \in \mathcal{G}} \exp(x_i^\top x_k)}}_{A_{i,j}} x_j \quad (5)$$

where \mathcal{G} indicates the global spatial space. The last three stages of a CoAtNet can be either a Convolution or a Transformer block, resulting in multiple combinations for the model architecture. Table 1 displays the family of CoAtNet models that have different sizes, number of blocks and hidden channels in the model architecture.

Input-Adaptive Weighting makes self-attention more prone to capture the relationships between different elements in the input and Global Receptive Field is the larger receptive field that's used in self attention. An optimal model architecture involves Input-Adaptive Weighting and Global Receptive Field characteristics of self-attention and the Translation Equivariance that is featured in CNNs as a way to improve generalization for a limited size dataset. The overall idea is to sum a global

static convolution kernel with the adaptive attention matrix, after or before the softmax initialization:

$$y_i^{\text{post}} = \sum_{j \in \mathcal{G}} \left(\frac{\exp(x_i^\top x_j)}{\sum_{k \in \mathcal{G}} \exp(x_i^\top x_k)} + w_{i-j} \right) x_j \quad (6)$$

$$y_i^{\text{pre}} = \sum_{j \in \mathcal{G}} \frac{\exp(x_i^\top x_j + w_{i-j})}{\sum_{k \in \mathcal{G}} \exp(x_i^\top x_k + w_{i-k})} x_j \quad (7)$$

3.4. Evaluation metrics

Most morphological classifications were accurately predicted by our trained model. Because neural networks are data-driven learning algorithms, their classification performance improves as the number of training sample data increases. Precision and recall, which are commonly used measurements of accuracy, precision, and recall, were utilized to evaluate our training method.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

A class's "true positives (TP)" and "true negatives (TN)" are determined by counting the number of predictions that agree with the ground truth in each category. "False positives (FP)" and "false negatives (FN)", on the other hand, represent the amount of predictions that are categorized or not classified into a specific class despite the fact that the ground truth contradicts their classification.

The accuracy is calculated by dividing the number of properly identified samples by the total number of samples in the assessment data set. This measure is one of the most often utilized in ML applications in medicine, but it is also renowned for being deceptive in the situation of various class proportions, since just assigning all samples to the predominant class is a simple method to get high accuracy.

Recall determines the ratio of properly categorized positive samples to all samples allocated to the positive class, also known as the sensitivity or the True Positive Rate (TPR).

The precision reflects the fraction of the recovered samples that are relevant and is determined as the ratio of properly categorized samples to all samples allocated to that class. When the costs of False Positives are substantial, precision is a suitable metric to use.

4. Results and discussion

The use of neural networks has been demonstrated to be effective in a variety of image classification challenges [58]. Using BM smears from a large patient cohort, we offer a comprehensive annotated high-quality data set of microscopic images collected from BM smears that may be utilized as a reference for developing machine learning algorithms for morphological categorization of diagnostically significant leukocytes. To the best of our knowledge, this picture database is the most comprehensive one currently accessible in the literature in terms of the number of patients, diagnoses, and single-cell images that are included within it. We utilized the data set to train and evaluate a novel convolutional and attention network-based model for cytology morphological classification, which was then compared with existing state-of-the-art CNN models.

The accurate distinction of various morphological classes, as previously stated, may be challenging, particularly when they are strongly tied to the leukocyte differentiation lineage; however, this is not always the case. Consequently, certain predictions of the network may be regarded as reasonable, even if they deviate from the ground truth by the human annotator, because of the inherent uncertainty associated with

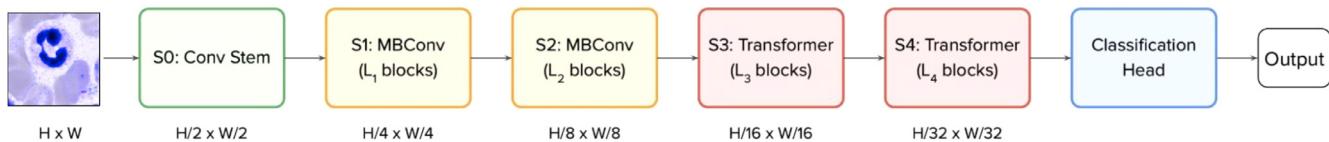


Fig. 6. The CoAtNet architecture is shown in the figure. An input picture of size $H \times W$, we use convolutions at the first stem stage (S0) to decrease the size to $H/2 \times W/2$, which is the final size. With each step, the size of the object continues to shrink. The number of layers is denoted by the letter L_n . Then, the first two stages (S1 and S2) mostly use MBConv building pieces, which are composed of depth wise convolution operations. The last two stages (S3 and S4) are mostly comprised of Transformer blocks with a high degree of relative self-attention. In contrast to the preceding Transformer blocks in ViT, we employ pooling between stages in this block, which is comparable to the Funnel Transformer block. We then use a classification head to provide predictions.

Table 1

L specifies the number of blocks, and D denotes the number of channels in the hidden dimension. We always utilize the kernel size 3 for all Conv and MBConv blocks, no matter what. Following [22], we increase the size of each attention head in all Transformer blocks to 32. The expansion rate for the inverted bottleneck is always 4, while the expansion (shrink) rate for the SE is always 0.25. The inverted bottleneck is also known as the inverted bottleneck.

Stage	Size	CoAtNet-0		CoAtNet-1		CoAtNet-2		CoAtNet-3		CoAtNet-4	
S0-Conv	1/2	L=2	D=64	L=2	D=64	L=2	D=128	L=2	D=192	L=2	D=192
S1-MbConv	1/4	L=2	D=96	L=2	D=96	L=2	D=128	L=2	D=192	L=2	D=192
S2-MBConv	1/8	L=3	D=192	L=6	D=192	L=6	D=256	L=6	D=348	L=12	D=348
S3-TFMRel	1/16	L=5	D=398	L=14	D=398	L=14	D=512	L=14	D=768	L=28	D=768
S4-TFMRel	1/32	L=2	D=768	L=2	D=768	L=2	D=1024	L=2	D=1536	L=2	D=1536

Table 2

Comparison of Multi-Class Classification Metrics Results amongst Various Models Used. The C, R, and E represents CoAtNet [56], ResNextT50 [54,55], and EfficientNetV2 [53] respectively.

Class Name	Accuracy			Precision			Recall		
	C	R	E	C	R	E	C	R	E
Band neutrophils	0.96	0.89	0.85	0.97	0.84	0.78	0.96	0.87	0.75
Segmented neutrophils	0.97	0.91	0.89	0.95	0.89	0.86	0.97	0.85	0.88
Lymphocytes	0.91	0.82	0.81	0.94	0.83	0.84	0.93	0.78	0.81
Monocytes	0.77	0.59	0.47	0.81	0.62	0.56	0.79	0.63	0.57
Eosinophils	0.85	0.68	0.64	0.91	0.65	0.62	0.88	0.69	0.6
Basophils	0.64	0.27	0.14	0.74	0.11	0.15	0.7	0.17	0.15
Metamyelocytes	0.88	0.21	0.29	0.91	0.15	0.21	0.89	0.12	0.24
Myelocytes	0.85	0.75	0.72	0.88	0.65	0.69	0.87	0.59	0.66
Promyelocytes	0.97	0.89	0.86	0.98	0.79	0.8	0.98	0.73	0.78
Blasts	0.96	0.91	0.92	0.94	0.71	0.84	0.98	0.88	0.8
Plasma cells	0.94	0.89	0.87	0.93	0.74	0.85	0.95	0.75	0.82
Proerythroblasts	0.84	0.68	0.57	0.89	0.53	0.58	0.85	0.5	0.55
Erythroblasts	0.98	0.87	0.88	0.99	0.78	0.86	0.98	0.76	0.83
Hairy cells	0.93	0.51	0.45	0.92	0.46	0.4	0.88	0.41	0.42
Abnormal eosinophils	0.43	0.18	0.12	0.42	0.11	0.09	0.39	0.13	0.14
Immature lymphocytes	0.63	0.22	0.16	0.65	0.13	0.15	0.66	0.12	0.16
Faggot cells	0.77	0.23	0.19	0.83	0.16	0.1	0.87	0.18	0.11

the morphological categorization. In the case of segmented neutrophils and band neutrophils, which are subsequent morphological classes in the ongoing process of myelopoiesis, misunderstanding between the two might be deemed tolerated in certain circumstances [59].

As shown in Table 2, the accuracy, precision, and recall values achieved by the CoAtNet (C), ResNext2 (R), and EfficientNetV2 (E) for each of the distinct morphological classes. The CoAtNet model outperformed both of the other models because of its attention network property that increased the learning curve for the algorithm which is represented using a precision-recall curve in Fig. 7.

For classes in which there are just a few training samples available, such as faggot cells or diseased eosinophils, the classifier performs less well, as would be anticipated for a data-driven strategy [60,61]. There would be a greater need for training data if the image-classification job was focused on the detection of these particular cell types. It is also possible that training a binary classifier rather than a complete multiclass classifier will result in improved prediction performance. The false positive and false negative cases depicts the similarity between each of these cells and this re-enforcing the importance as well as the complexity of the task of classification of cell morphology.

Because neural networks are created in a data-driven manner based on the training set, the classification judgments made by neural networks do not lend themselves to straightforward human interpretation. However, in order to obtain insight into the classifications made by these algorithms, a number of explainability approaches have recently been created to aid in their investigation [62]. As part of our effort to determine which regions of the input images are important for the network's classification decisions, we analyzed the CoAtNet model using SmoothGrad[63] and Grad-CAM[64], two recently developed algorithms that have been shown to meet the fundamental requirements for explainability methods (ie, sensitivity to data and model parameter randomization) [65].

As seen in Fig. 8, the model has trained to focus on the important input of a single-cell patch (ie, the primary leukocyte depicted in it) while disregarding background characteristics such as erythrocytes, cell debris, and pieces of other cells visible in the patch. Furthermore, specific defining structures that are known to be relevant to human examiners when classifying cells appear to play a role in the network's attention pattern, such as the cytoplasm of eosinophils and the cell membrane of hairy cells, which appear to play a role in the network's attention

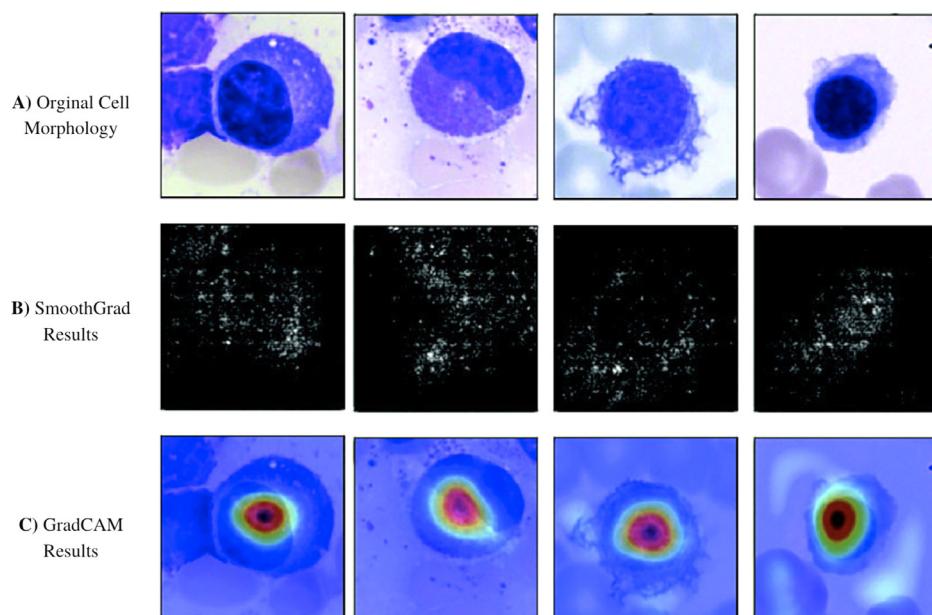
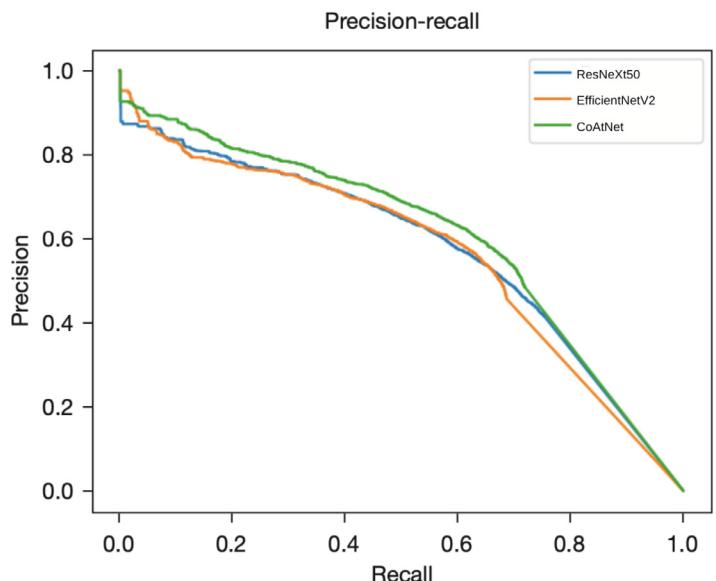


Fig. 7. The figure shows precision-recall curve for each of the models we trained. The trade off between precision and recall may be seen in the precision-recall curve for various threshold values. A low false negative and false positive rate is associated with a higher area under the curve.

Fig. 8. Original photos classified properly by the network are presented in the top row. All cells were stained with the May-Gruenwald-Giemsa/Pappenheim stain and photographed at 340 magnification, as described in the primary text. The center row displays analysis using the SmoothGrad algorithm. The brighter a pixel looks, the more it contributes to the categorization judgment made by the network. Results of a second network analysis approach, the Grad-CAM algorithm, are exhibited in the bottom row as a heat map superimposed on the original picture. The areas of the image that contain essential information are highlighted in red. Both analytical approaches imply that the network has trained to concentrate on the leukocyte while disregarding background structure. Note the network's focus on traits known to be significant for certain cell types, such as the cytoplasmic organization of eosinophils or the nuclear architecture of plasma cells.

pattern. Although these analyses, when used as post classification explanations, do not in and of themselves guarantee the correctness of a particular classification decision, they can help to increase confidence that the network has learned to focus on relevant features of the single-cell images and that predictions are based on reasonable characteristics.

As a whole, the results are positive and promising, with good accuracy and recall values achieved for the vast majority of diagnostically relevant classes studied. Our results are consistent with previous findings in other areas of medical imaging, where attention-based image classification tasks have outperformed approaches that rely on the extraction of image features to attain higher standards [66–69]. The most important component of the successful use of CNNs is a training data set that is sufficiently big and of good quality [70].

5. Conclusion

As part of the current investigation, we mostly used a single-center strategy, with all BM smears included for training having been stained, captured, and processed in the same laboratory. The network presented

in this paper performs well in such an environment, which is quite promising. Our proposed convolution and attention network model (CoAtNet) outperformed the current state-of-the-art CNN models in classifying cells and even took a huge leap in accurately classifying classes with low sample sizes. This shows how the attention network could be used in similar datasets in the future as well.

Future research may help to lessen the importance of label noise (eg, by using semi- or unsupervised methods as having been applied to processes such as erythrocyte assessment[46] or cell cycle reconstruction). Increased growth of the morphological database, preferably as part of multi-centric research and using a variety of scanner hardware, will almost certainly improve the performance and resilience of the network, particularly for classes with a smaller size in the current data set.

However, because of the large number of cases and diagnoses included in our study, we anticipate that the data set will fairly accurately represent the morphological variation seen in most cell groups. The purpose of this research is to evaluate the morphology of the adult BM. It would be fascinating to expand the study to include samples from newborns and young children, particularly for lymphoid cells. The per-

formance of our network in a real-world diagnostic environment will need more investigation. The range of diagnostic modalities employed in hematology suggests that the integration of supplementary data (for example, from flow cytometry or molecular genetics) would improve the quality of predictions made by neural networks.

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.

References

- [1] Wintrobe MM. Clinical hematology. *Acad Med* 1962;37(1):78.
- [2] Theml H, Diem H, Haferlach T. Color atlas of hematology: practical microscopic and clinical diagnosis. Thieme; 2004.
- [3] Hoffman R, Benz Jr EJ, Silberstein LE, Heslop H, Anastasi J, Weitz J. Hematology: basic principles and practice. Elsevier Health Sciences; 2013.
- [4] Löffler H, Rastetter J. Atlas of clinical hematology. Springer Science & Business Media; 2012.
- [5] Kratz A, Lee S-h, Zini G, Riedl JA, Hur M, Machin S, et al. Digital morphology analyzers in hematology: ICSHreview and recommendations. *Int. JLabHematol* 2019;41(4):437–47.
- [6] Salakij C, Salakij J, Apibal S, Narkkong N-A, Chanhome L, Rochanapat N. Hematology, morphology, cytochemical staining, and ultrastructural characteristics of blood cells in king cobras (ophiophagus hannah). *Vet Clin Pathol* 2002;31(3):116–26.
- [7] Thomas X. First contributors in the history of leukemia. *World J Haematol* 2013;2(62).
- [8] Tkachuk D, Hirschmann J. Approach to the microscopic evaluation of blood and bone marrow. *Wintrobe Atlas Clin Haematol* Lippincott Williams Wilkins 2007.
- [9] Briggs C, Longair I, Slavik M, Thwaite K, Mills R, Thavaraja V, et al. Can automated blood film analysis replace the manual differential? An evaluation of the cellavision dm96 automated image analysis system. *Int J Lab Hematol* 2009;31(1):48–60.
- [10] Angulo J, Flandrin G. Automated detection of working area of peripheral blood smears using mathematical morphology. *AnalCellPathol* 2003;25(1):37–49.
- [11] Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nat Mach Intell* 2019;1(11):538–44.
- [12] Nabity MB, Harr KE, Camus MS, Flatland B, Vap LM. ASVCP guidelines: allowable total error hematology. *Vet Clin Pathol* 2018;47(1):9–21.
- [13] Simons A, Sikkema-Raddatz B, de Leeuw N, Konrad NC, Hastings RJ, Schoumans J. Genome-wide arrays in routine diagnostics of hematological malignancies. *HumMut* 2012;33(6):941–8.
- [14] Fuentes-Arderiu X, Dot-Bach D. Measurement uncertainty in manual differential leukocyte counting. *Clin Chem Lab Med* 2009;47(1):112–15.
- [15] Font P, Loscertales J, Soto C, Ricard P, Novas CM, Martín-Clavero E, et al. Interobserver variance in myelodysplastic syndromes with less than 5% bone marrow blasts: unilineage vs. multilineage dysplasia and reproducibility of the threshold of 2% blasts. *AnnHematol* 2015;94(4):565–73.
- [16] Krappe S, Benz M, Wittenberg T, Haferlach T, Münenmayer C. Automated classification of bone marrow cells in microscopic images for diagnosis of leukemia: a comparison of two classification schemes with respect to the segmentation quality. *Medical imaging 2015: computer-aided diagnosis, 9414*. International Society for Optics and Photonics; 2015. 94143I.
- [17] Reta C, Altamirano L, Gonzalez JA, Diaz-Hernandez R, Peregrina H, Olmos I, et al. Segmentation and classification of bone marrow cells images using contextual information for medical diagnosis of acute Leukemias. *PLoS One* 2015;10(6):e0130805.
- [18] Chandrasevan R, Aljudi AA, Drumheller BR, Kunanthaseelan N, Amgad M, Gutman DA, et al. Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells. *Lab Investig* 2020;100(1):98–109.
- [19] Song T-H, Sanchez V, Daly HE, Rajpoot NM. Simultaneous cell detection and classification in bone marrow histology images. *IEEE JBiomedHealth Inform* 2018;23(4):1469–76.
- [20] Krappe S, Wittenberg T, Haferlach T, Münenmayer C. Automated morphological analysis of bone marrow cells in microscopic images for diagnosis of leukemia: nucleus-plasma separation and cell classification using a hierarchical tree model of hematopoiesis. *Medical imaging 2016: computer-aided diagnosis, 9785*. International Society for Optics and Photonics; 2016. 97853C.
- [21] Scotti F. Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images. In: CIMSA. 2005 IEEE international conference on computational intelligence for measurement systems and applications, 2005.. IEEE; 2005. p. 96–101.
- [22] Kimura K, Tabe Y, Ai T, Takehara I, Fukuda H, Takahashi H, et al. A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA. *Sci Rep* 2019;9(1):1–9.
- [23] Mori J, Kaji S, Kawai H, Kida S, Tsubokura M, Fukatsu M, et al. Assessment of dysplasia in bone marrow smear with convolutional neural network. *Sci Rep* 2020;10(1):1–8.
- [24] Wu Y-Y, Huang T-C, Ye R-H, Fang W-H, Lai S-W, Chang P-Y, et al. A hematologist-level deep learning algorithm (bmsnet) for assessing the morphologies of single nuclear balls in bone marrow smears: algorithm development. *JMIR MedInform* 2020;8(4):e15963.
- [25] Anilkumar K, Manoj V, Sagi T. A survey on image segmentation of blood and bone marrow smear images with emphasis to automated detection of leukemia. *Biocybern Biomed Eng* 2020;40(4):1406–20.
- [26] Rehman A, Abbas N, Saba T, Rahman Slu, Mahmood Z, Kolivand H. Classification of acute lymphoblastic leukemia using deep learning. *Microsc Res Tech* 2018;81(11):1310–17.
- [27] Jin H, Fu X, Cao X, Sun M, Wang X, Zhong Y, et al. Developing and preliminary validating an automatic cell classification system for bone marrow smears: a pilot study. *J Med Syst* 2020;44(10):1–10.
- [28] Yu T-C, Chou W-C, Yeh C-Y, Yang C-K, Huang S-C, Tien FM, Yao C-Y, Cheng C-L, Chuang M-K, Tien H-F, et al. Automatic bone marrow cell identification and classification by deep neural network. *Blood* 2019;134:2084.
- [29] Choi JW, Ku Y, Yoo BW, Kim J-A, Lee DS, Chai YJ, et al. White blood cell differential count of maturation stages in bone marrow smear using dual-stage convolutional neural networks. *PLoS One* 2017;12(12):e0189259.
- [30] Suzuki K. Overview of deep learning in medical imaging. *RadiolPhysTechnol* 2017;10(3):257–73.
- [31] Fu G-S, Levin-Schwartz Y., Lin Q.-H., Zhang D.. Machine learning for medical imaging. 2019.
- [32] Zhang D, Song Y, Liu D, Jia H, Liu S, Xia Y, et al. Panoptic segmentation with an end-to-end cell R-CNNfor pathology image analysis. In: International conference on medical image computing and computer-assisted intervention. Springer; 2018. p. 237–44.
- [33] Tripathi S. Artificial intelligence: a brief review. In: Analyzing future applications of AI, Sensors, and robotics in society; 2021. p. 1–16.
- [34] Tripathi S, Musiolik TH. Fairness and ethics in artificial intelligence-based medical imaging. In: Ethical implications of reshaping healthcare with emerging technologies. IGI Global; 2022. p. 71–85.
- [35] Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput* 2017;29(9):2352–449.
- [36] Tripathi S, Augustin AI, Moyer EJ, Zavalny A, Dheer S, Sukumaran R, Schwartz D, Gorski B, Dako F, Kim E. Radgennets: deep learning-based radiogenomics model for gene mutation prediction in lung cancer. *bioRxiv* 2022. doi:10.1101/2022.04.13.488208.
- [37] Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA: Cancer JClin* 2019;69(2):127–57.
- [38] Le E, Wang Y, Huang Y, Hickman S, Gilbert F. Artificial intelligence in breast imaging. *ClinRadiol* 2019;74(5):357–66.
- [39] Lee E-J, Kim Y-H, Kim N, Kang D-W. Deep into the brain: artificial intelligence in stroke imaging. *JStroke* 2017;19(3):277.
- [40] Huang J, Shlobin NA, Lam SK, DeCuyperre M. Artificial intelligence applications in pediatric brain tumor imaging: a systematic review. *World Neurosurg* 2022;157:99–105.
- [41] Esteva A, Topol E. Can skin cancer diagnosis be transformed by AI? *Lancet* 2019;394(10211):1795.
- [42] Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans-MedImaging* 2016;35(5):1153–9.
- [43] Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *AnnuRevBiomedEng* 2017;19:221–48.
- [44] Schaeckermann M, Beaton G, Habib M, Lim A, Larson K, Law E. Capturing expert arguments from medical adjudication discussions in a machine-readable format. In: Companion proceedings of the 2019 world wide web conference; 2019. p. 1131–7.
- [45] Group S-I, Community FR, et al. Artificial intelligence and medical imaging 2018: French radiology community white paper. *Diagn Interv Imaging* 2018;99(11):727–42.
- [46] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* 2021;13(1):1–17.
- [47] Tayebi RM, Mu Y, Dehkharghanian T, Ross C, Sur M, Foley R, et al. Automated bone marrow cytology using deep learning to generate a histogram of cell types. *Commun Med* 2022;2(1):1–14.
- [48] Tayebi R.M., Mu Y., Dehkharghanian T., Ross C., Sur M., Foley R., Tizhoosh H.R., Campbell C.J.. Histogram of cell types: deep learning for automated bone marrow cytology. *arXiv preprint arXiv:210702293*; 2021.
- [49] Zhu X, Lyu S, Wang X, Zhao Q. Tph-yolov5: improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 2778–88.
- [50] Huang D, Cheng J, Fan R, Su Z, Ma Q, Li J. Bone marrow cell recognition: training deep object detection with a new loss function. In: 2021 IEEE international conference on imaging systems and techniques (IST). IEEE; 2021. p. 1–6.
- [51] Chisholm KM, Xu M, Davis B, Ogi A, Pacheco MC, Geddis AE, et al. Evaluation of the utility of bone marrow morphology and ancillary studies in pediatric patients under surveillance for myelodysplastic syndrome. *Am J Clin Pathol* 2018;149(6):499–513.
- [52] Matek C, Krappe S, Münenmayer C, Haferlach T, Marr C. Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. *Blood J Am SocHematol* 2021;138(20):1917–27.
- [53] Tan M, Le Q. Efficientnetv2: smaller models and faster training. In: International conference on machine learning. PMLR; 2021. p. 10096–106.
- [54] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1492–500.

- [55] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *IntJComputVis* 2015;115(3):211–52.
- [56] Dai Z, Liu H, Le QV, Tan M. Coatnet: marrying convolution and attention for all data sizes. *Adv Neural Inf Process Syst* 2021;34:3965–77.
- [57] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
- [58] Doan M, Sebastian JA, Caicedo JC, Siegert S, Roch A, Turner TR, et al. Objective assessment of stored blood quality by deep learning. *Proc Natl Acad Sci* 2020;117(35):21381–90.
- [59] Krappe S, Wittenberg T, Haferlach T, Münzenmayer C. Automated morphological analysis of bone marrow cells in microscopic images for diagnosis of leukemia: nucleus-plasma separation and cell classification using a hierarchical tree model of hematopoiesis. *Medical imaging 2016: computer-aided diagnosis*, 9785. International Society for Optics and Photonics; 2016. 97853C.
- [60] Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput* 2017;29(9):2352–449.
- [61] Schouten JP, Matek C, Jacobs LF, Buck MC, Bošnacki D, Marr C. Tens of images can suffice to train neural networks for malignant leukocyte detection. *Sci Rep* 2021;11(1):1–8.
- [62] Samek W, Müller K-R. Towards explainable artificial intelligence. In: Explainable AI: interpreting, explaining and visualizing deep learning. Springer; 2019. p. 5–22.
- [63] Smilkov D, Thorat N, Kim B, Viégas F., Wattenberg M.. Smoothgrad: removing noise by adding noise. arXiv preprint arXiv:170603825; 2017.
- [64] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 618–26.
- [65] Adebayo J, Gilmer J, Muelly M, Goodfellow I, Hardt M, Kim B. Sanity checks for saliency maps. *Adv Neural Inf Process Syst* 2018;31.
- [66] Jing B, Xie P, Xing E.. On the automatic generation of medical imaging reports. arXiv preprint arXiv:171108195; 2017.
- [67] Khanh TLB, Dao D-P, Ho N-H, Yang H-J, Baek E-T, Lee G, Kim S-H, Yoo SB. Enhancing u-net with spatial-channel attention gate for abnormal tissue segmentation in medical imaging. *Appl Sci* 2020;10(17):5729.
- [68] Li X, Hu X, Yu L, Zhu L, Fu C-W, Heng P-A. Canet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE Trans Med Imaging* 2019;39(5):1483–93.
- [69] He X, Deng Y, Fang L, Peng Q. Multi-modal retinal image classification with modality-specific attention network. *IEEE Trans Med Imaging* 2021;40(6):1591–602.
- [70] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.