Preprints.org

Article

Blood Cell Revolution: Unveiling 11 Distinct Types with 'Naturalize' Augmentation

Mohamad Abou Ali , Fadi Dornaika [*] , Ignacio Arganda-Carreras

Article

Blood Cell Revolution: Unveiling 11 Distinct Types with 'Naturalize' Augmentation

Mohamad Abou Ali [1] [iD]**, Fadi Dornaika** [1,2,*] [iD]**, and Ignacio Arganda-Carreras** [1,2,3,4] [iD]

[1]　Dept. Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Manuel Lardizabal, 1, 20018 San Sebastian, Spain; mohamad.abouali01@liu.edu.lb (M.A.A.); ignacio.arganda@ehu.eus (I.A.-C.)

[2]　IKERBASQUE, Basque Foundation for Science, Plaza Euskadi, 5, 48009 Bilbao, Spain

[3]　Donostia International Physics Center (DIPC), Manuel Lardizabal, 4, 20018 San Sebastian, Spain

[4]　Biofisika Institute (CSIC, UPV/EHU), Barrio Sarriena s/n, 48940 Leioa, Spain

*　Correspondence: fadi.dornaika@ehu.eus

Abstract: In this paper, we redefine the boundaries of blood cell classification, expanding from 5 to 11 distinct blood cell types using the challenging 11-class PBC dataset. This shift enables a deeper understanding of blood cell diversity, overcoming previous limitations in medical image analysis. Our approach combines state-of-the-art deep learning techniques, including pre-trained ConvNets, ViTb16 models, and custom CNN architectures. We employ transfer learning, fine-tuning, and ensemble strategies, such as CBAM and Averaging ensembles, to achieve unprecedented accuracy and interpretability. Our fully fine-tuned EfficientNetV2 B0 model sets a new standard, with a macro-average precision, recall, and F1-score of 91%, 90%, and 90%, respectively, and an average accuracy of 93%. This breakthrough underscores the transformative potential of 11-class blood cell classification for more precise medical diagnoses. Moreover, our groundbreaking "Naturalize" augmentation technique produces remarkable results. The 2K-PBC dataset generated with "Naturalize" boasts a macro-average precision, recall, and F1-score of 97%, along with an average accuracy of 96% when leveraging the fully fine-tuned EfficientNetV2 B0 model. This innovation not only elevates classification performance but also addresses data scarcity and bias in medical deep learning. Our research marks a paradigm shift in blood cell classification, enabling more nuanced and insightful medical analyses. The "Naturalize" technique's impact extends beyond blood cell classification, emphasizing the vital role of diverse and comprehensive datasets in advancing healthcare applications through deep learning.

Keywords: Convolutional Neural Net (CNN); Vision Transformer (ViT); ImageNet Models; Transfer Learning (TL); Machine Learning (ML); Deep Learning (DP); Blood Cell Classification; Peripheral Blood Cell (PBC); CBAM; Naturalize

---

1. Introduction

Hematopoiesis, the remarkable process responsible for blood cell formation, is orchestrated by a hierarchy of specialized cells. At its core are hematopoietic stem cells, the ultimate architects of this symphony. These versatile cells possess the unique ability to self-renew and differentiate into various lineages. From them, progenitor cells emerge, committed to specific blood cell types like red blood cells, white blood cells, and platelets. Progenitor cells then give rise to precursor cells, which exhibit more specialized traits and characteristics. Finally, precursor cells mature into fully functional blood cells, capable of carrying out their vital roles in oxygen transport, immune defense, and clotting. The presence of precursor cells in the bloodstream is essential to maintain a dynamic reservoir, ready to respond to the body's ever-changing demands for blood cells, ensuring a constant supply to sustain life and preserve health [1].

Illustrated in Figure 1 [1], the marvel of hematopoiesis unveils the genesis of neutrophils, mature white blood cells, commencing from hematopoietic stem cells. An intricate journey unfolds, as they

metamorphose through myeloid progenitors, traverse the precursors (myeloblast, promyelocyte, and myelocyte), evolve into meta-myelocytes, and ultimately reach their zenith as segmented neutrophils, boasting multi-lobed nuclei. This progression orchestrates the assembly of a formidable defense force, primed to combat infections with remarkable efficiency [1].
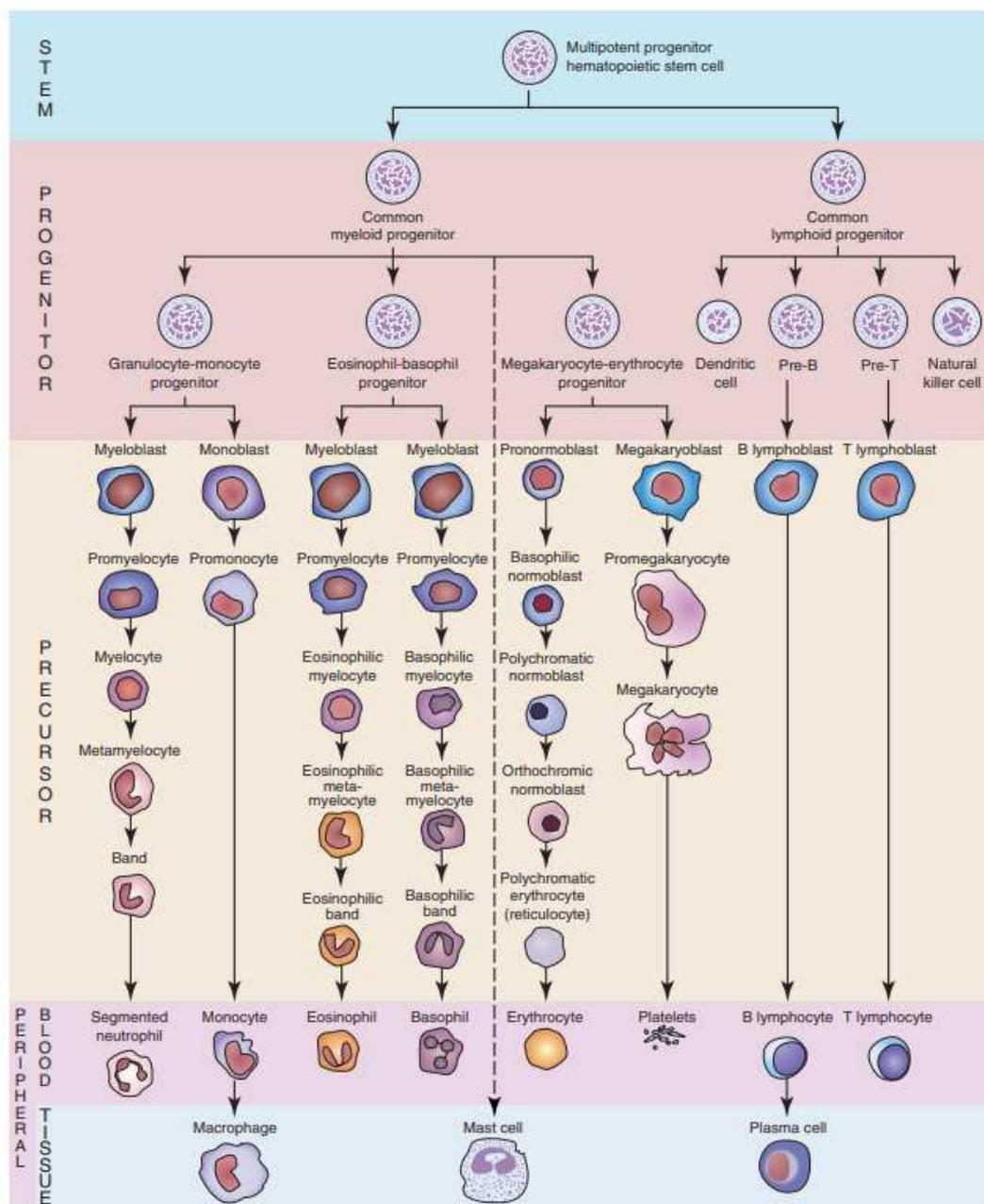
Figure 1. Chart of hematopoiesis.

A peripheral blood smear is a microscopic window into hematopoiesis, offering a quick and insightful view of blood cell types and any abnormalities, crucial for diagnosing and monitoring hematological conditions [2].

In the realm of hematology, the meticulous examination of peripheral blood smears is vital for diagnosing blood disorders and guiding treatment Hematologists shoulder the burden of this labor-intensive manual process, scrutinizing countless blood cells with unwavering attention amidst a high workload and constant distractions. Their dedication is essential for precise diagnostics and patient care; given the criticality of the results they deliver [2].

However, the field of hematology has witnessed a transformative evolution through deep learning, a subset of artificial intelligence. This advancement automates peripheral blood smear analysis, alleviating the challenges faced by hematologists. Deep learning algorithms swiftly process blood cells, identifying subtle abnormalities and providing rapid, consistent results. This automation eases the workload, reduces human error, and empowers healthcare experts to focus on intricate cases [3].

As technology continues to advance, the fusion of deep learning and peripheral blood smear analysis promises to revolutionize blood disorder diagnosis and management, enhancing patient care outcomes [3].

In recent years, the drive to automate peripheral blood smear analysis has gained momentum within hematology. Our study leverages the "Peripheral Blood Cell" (PBC) dataset, encompassing 11 distinct blood cell classes, offering a comprehensive foundation for analysis [4]. Challenges stemming from class imbalance in the PBC dataset are addressed with our innovative "Naturalize" augmentation technique.

Our approach incorporates state-of-the-art deep learning models, including ImageNet ConvNets [5] and Vision Transformer (ViT) [6], through techniques like transfer learning [7–9], fine-tuning, and ensembling [10]. Evaluation encompasses quantitative assessments using confusion matrices, classification reports [11], and visual evaluations using tools like Score-CAM [12].

These advancements, combined with the pioneering application of "Naturalize," signify substantial progress in automating peripheral blood smear analysis. This paper delves deep into the methodologies and outcomes of these cutting-edge approaches, illuminating their potential to reshape the field of hematology.

After this introduction, the rest of the paper will continue as follows: Section 2 highlights the relevant literature related to the detection and classification of blood cells using pre-trained CNNs and ViTs, and Section 3 describes the methodology used in this study. In addition, Section 4 presents the experimental results obtained using pre-trained ImageNet models and Google ViT for the blood cell classification. In Section 5, an in-depth analysis of results is made. Finally, the conclusions of this work are presented in Section 6.

2. Related Works

Within the realm of machine learning, the conventional approach involved manual extraction of image features followed by classification. The advent of deep learning revolutionized this process by automating image classification, particularly impactful in blood cell analysis. Previous studies primarily concentrated on five white blood cell (WBC) types: neutrophils, eosinophils, basophils, lymphocytes, and monocytes. Distinguishing our research is the comprehensive classification of 11 distinct blood cell types, inclusive of segmented and banded neutrophils, introducing Meta-myelocyte, Myelocyte, Pro-myelocyte, erythroblasts, and platelets to the array. This expansion broadens the classification spectrum, redefining the scope of blood cell analysis.

Jung et al. [13] introduced W-Net, a CNN-based model for white blood cell (WBC) classification, achieving 97% accuracy with 10-fold cross-validation. Their work focused on classifying five WBC types: neutrophils, eosinophils, basophils, lymphocytes, and monocytes. Sahlol et al. [14] combined VGG-16 and a feature reduction algorithm (SESSA) to achieve 95.67% accuracy for WBC leukemia image classification. They categorized WBCs into five classes.

Almezhghwi et al. [15] used ImageNet pre-trained architectures (VGG, ResNet, DenseNet) and data augmentation, with DenseNet-169 achieving 98.8% accuracy. Their study involved the classification of five WBC types. Tavakoli et al. [16] developed a segmentation algorithm and used SVM for WBC classification, achieving high accuracy in multiple datasets. They focused on five WBC types.

Chen et al. [17] proposed a hybrid deep model combining ResNet and DenseNet with a spatial and channel attention module (SCAM), outperforming previous methods. They worked with datasets

containing five WBC types. Katar et al. [18] applied transfer learning with ImageNet models, with MobileNet-V3-Small achieving 98.86% accuracy. They classified five WBC types.

Nahzat et al. [19] designed a CNN-based model using the Kaggle BCCD dataset, achieving competitive results. They also focused on five WBC types. Heni et al. [20] introduced the EK-Means method for WBC image segmentation, achieving a validation accuracy of 96.24% with VGG-19. They categorized five WBC types. Ziquan Zhu [21] and [22] presented DLBCNet models that used ResNet-50 for feature extraction and achieved impressive accuracy. They worked with datasets containing five WBC types.

However, what sets our research apart is the groundbreaking "Naturalize" augmentation technique. It directly addresses two major deep learning challenges: data insufficiency and class imbalance. In deep learning, the availability of diverse and balanced data is critical for model performance. "Naturalize" overcomes these hurdles by generating high-quality blood cell samples, effectively augmenting our dataset. This innovation significantly enhances classification performance, pushing the boundaries of what deep learning can achieve in blood cell analysis. Our approach not only improves accuracy but also ensures robustness and reliability in dealing with diverse and imbalanced blood cell classes.

By expanding the classification scope to 11 blood cell types and introducing the game-changing "Naturalize" method, our research pioneers a new era in automated blood cell analysis. It effectively tackles long-standing challenges in deep learning, transforming the field and promising substantial improvements in diagnostic precision, ultimately enhancing patient care and medical outcomes.

3. Materials and Methods

This section offers an detailed exposition of our methodology, designed to elevate blood cell image classification within the challenging Peripheral Blood Cell (PBC) dataset to an exceptional level of precision. Our classification endeavor leveraged the power of pre-trained Deep Learning (DL) architectures, prominently featuring the "ImageNet ConvNets" and "Keras Vision Transformers (ViT)." We harnessed the ingenious "Naturalize" augmentation technique, strategically addressing data insufficiencies in certain classes, such as pro-myelocyte, while concurrently expanding the available image data in other classes. The sheer breadth and sophistication of our methodology are vividly depicted in Figure 2, a testament to our unwavering commitment to advancing the frontiers of image classification.

3.1. PBC Dataset

3.1.1. Original 8-Class PBC Dataset

The original PBC dataset [4], sourced from an online repository, encompasses 17,092 images categorized into eight distinct classes of blood cells. These eight classes include neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes, erythroblasts, and platelets (thrombocytes).

This means the original PBC contains 8 folders. The "Immature Cells" folder contains three classes labeled as "Pro-myelocyte", "Myelocyte" and "Meta-myelocyte". Also, the "Neutrophil" folder contains two labeled classes "Banded Neutrophil" and "Segmented Neutrophil". Thus, the total number of classes in the original PBC dataset are 11 classes.

Table 1 provides an overview of the distribution of the eight blood cell classes within the original PBC dataset.

The images in the PBC dataset adhere to a standard size of 360 × 363 pixels [4], which closely matches the input size of ImageNet models and the Google ViT, minimizing the need for significant resizing.
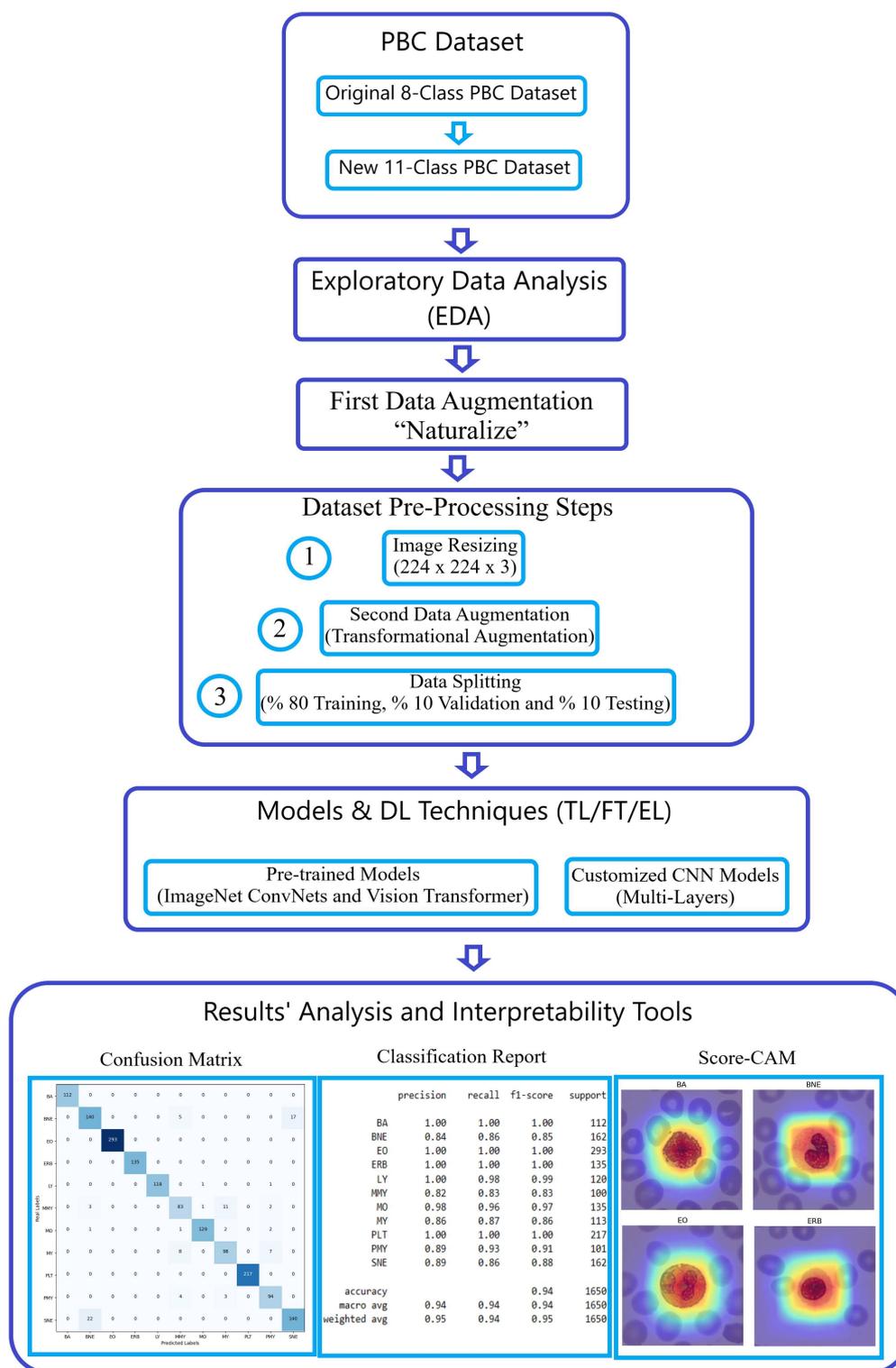
Figure 2. Methodology workflow using the PBC dataset.

Table 1. Summary OF the PBC dataset.

Number	Cell Type	Total of Images by Type	Percent
1	Neutrophils	3,329	19.48
2	Eosinophils	3,117	18.24
3	Basophils	1,218	7.13
4	Lymphocytes	1,214	7.10
5	Monocytes	1,420	8.31
6	Immature Granulocytes (IG)	2,895	16.94
7	Erythroblasts	1,551	9.07
8	Platelets (Thrombocytes)	2,348	13.74
	Total	17,092	100

3.1.2. New 11-Class PBC dataset

To expand the original 8-Class PBC dataset into 11 distinct blood cell types or classes, a separation of labeled classes in two PBC folders "Immature Granulocytes" and "Neutrophil" was applied, as illustrated in Figure 3. These eleven classes include banded neutrophils, basophils, eosinophils, erythroblasts, lymphocytes, meta-myelocytes, monocytes, myelocytes, platelets, pro-myelocytes and segmented neutrophils.
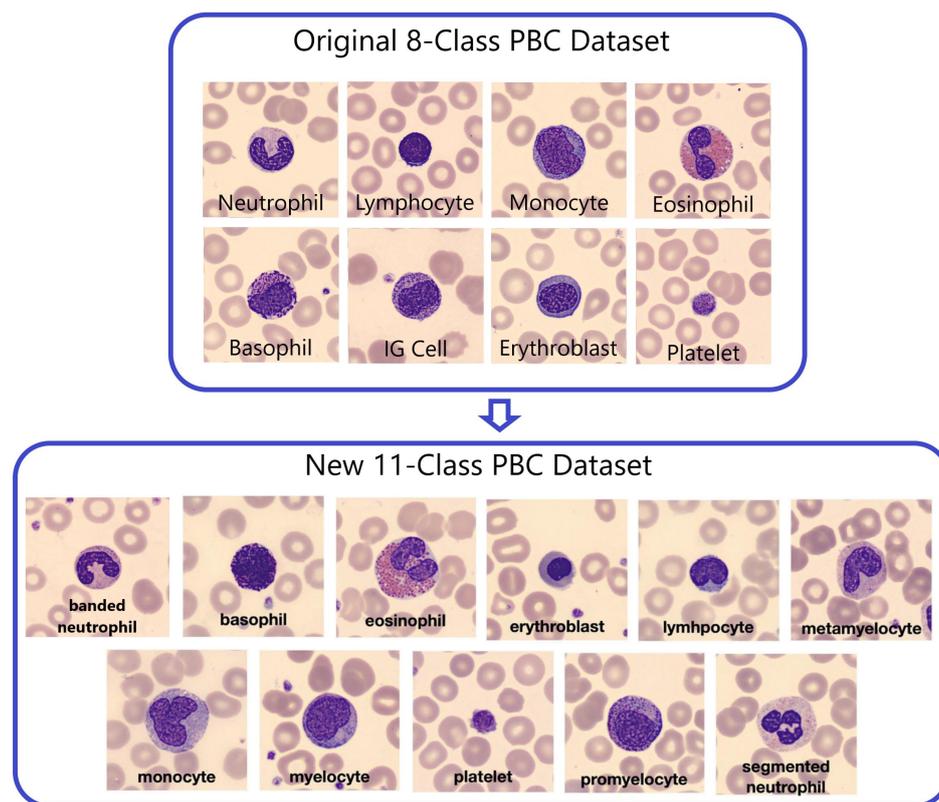
Figure 3. Transition from the 8-class PBC into the 11-class PBC.

The new Peripheral Blood Cell (PBC) dataset has been grouped into 11 specific blood cell categories, as outlined in Table 2 [4]. To achieve this classification, certain entries from the original PBC dataset were eliminated because they didn't align with the new 11 categories. For instance, 151 images initially categorized as immature granulocytes lacked further subclassification (pro-myelocyte, myelocyte, or meta-myelocyte), resulting in a reduction in the total number of images in the revised 11-class PBC dataset to 16,891. Notably, the analysis of Table2 reveals a significant underrepresentation of the Pro-myelocyte (PMY) category among these classes.

Table 2. Summary of The New 11-Class PBC Dataset

#	Cell Class	Symbol	Total Images By Class	%
1	Basophil	BA	1,218	7
2	Banded Neutrophil	BNE	1,633	10
3	Eosinophil	EO	3,117	18
4	Erythroblast	ERB	1,551	9
5	Lymphocyte	LY	1,214	7
6	Meta-myelocyte	MMY	1,015	6
7	Monocyte	MO	1,420	8
8	Myelocyte	MY	1,137	7
9	Platelet	PLT	2,348	14
10	Pro-myelocyte	PMY	592	4
11	Segmented Neutrophil	SNE	1,646	10
	Total		16,891	100

3.2. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to gain insights into the nature of the dataset. This involved training and testing pre-trained ImageNet models with the new 11-class PBC dataset, analyzing the confusion matrix, and generating a classification report.

EDA uncovered issues: multi-class images caused confusion and were removed; fragments from other classes also confused models (Figure 4). Scarce pro-myelocyte images affected performance, addressed with "Naturalize" augmentation to match original image quality.
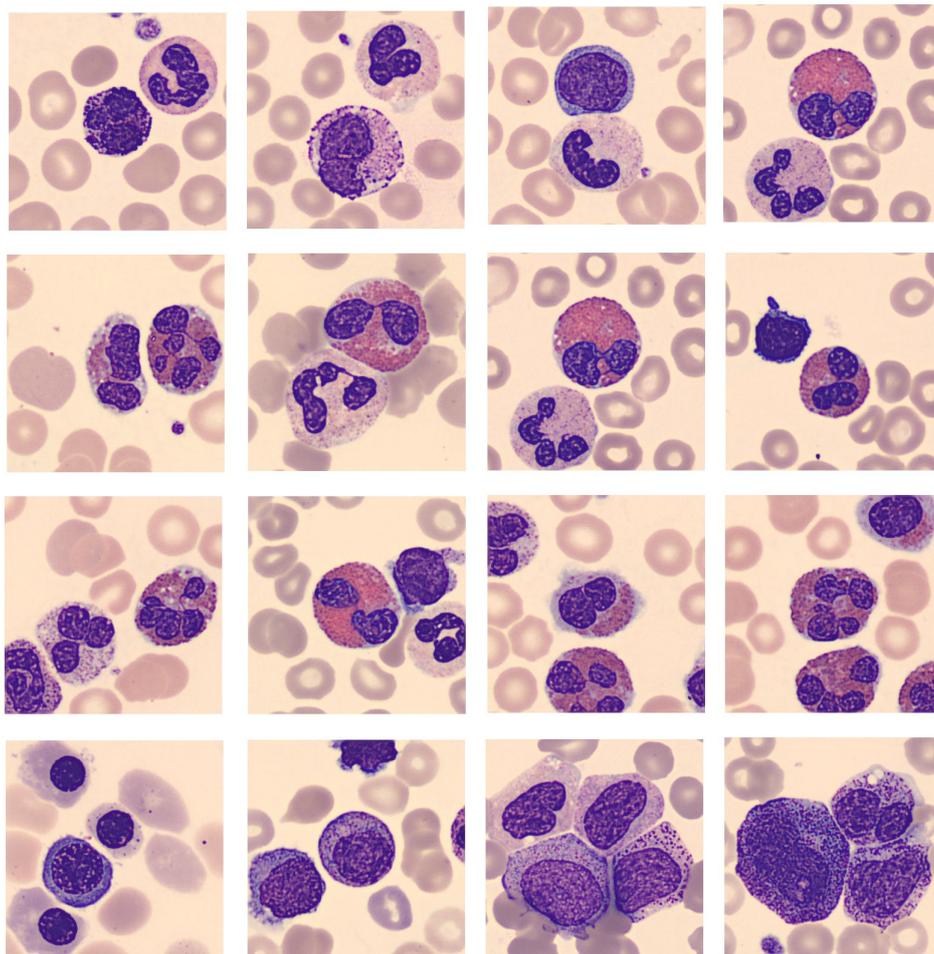
Figure 4. PBC images containing more than one class.

3.3. First Data Augmentation "Naturalize"

The pseudocode shown in Algorithm 1 demonstrates the principle behind the "Naturalize" augmentation technique and how it works.

Algorithm 1 Naturalize Algorithm

```
 1: # Imports and Paths
 2: import os, random, image_processing, SAM_model
 3: Define file paths and import essential libraries
 4:
 5: # Load SAM_model and PBC Dataset
 6: mount Google_drive
 7: load peripheral blood smear images from PBC dataset
 8: SAM = load_model(SAM_model)
 9:
10: # Segment PBC Dataset Using SAM
11: segment all PBC images using SAM into segmented "RBCs, WBCs, PLTs"
12: save segmented "RBCs, WBCs, PLTs" into new datasets on Google_drive
13:
14: # Collision Function
15: function CHECK_COLLISION(positions, x, y, width, height)
16:     Function to check for image collisions
17: end function
18:
19: # Composite Image Creation
20: for i in range(num_images) do
21:     Load background image and initialize positions list
22:     Add 1 WBC image
23:     Add 8 RBC images with collision avoidance
24:     Save the composite image on Google_drive
25: end for
```

The "Naturalize" augmentation technique consists of two primary steps:

- **Step 1 - Segmentation**:

    A portion of the initial PBC dataset images underwent segmentation and were separated into six distinct datasets employing the "Segment Anything Model (SAM)" developed by Meta AI [23]. These datasets encompass segmented images of Red Blood Cells (RBC), Banded Neutrophils (BNE), Meta-myelocytes (MMY), Myelocytes (MY), Pro-myelocytes (PMY), and Segmented Neutrophils (SNE).

    This selective approach, using only these six classes in SAM, was adopted based on the findings of our earlier Exploratory Data Analysis (EDA) and classification report, which demonstrated an overall enhancement in classification performance, contrasting with the use of all 11 classes. Notably, for each cell class, SAM generates masks that partition the images into three distinct components: background, segmented RBCs, and segmented WBCs.

    To offer further clarification, Figure 5 utilizes Pro-myelocytes (PMY) as a demonstration of the SAM model's application. This figure showcases the segmentation process for PMY images, illustrating the division into background, segmented RBCs, and segmented PMYs.
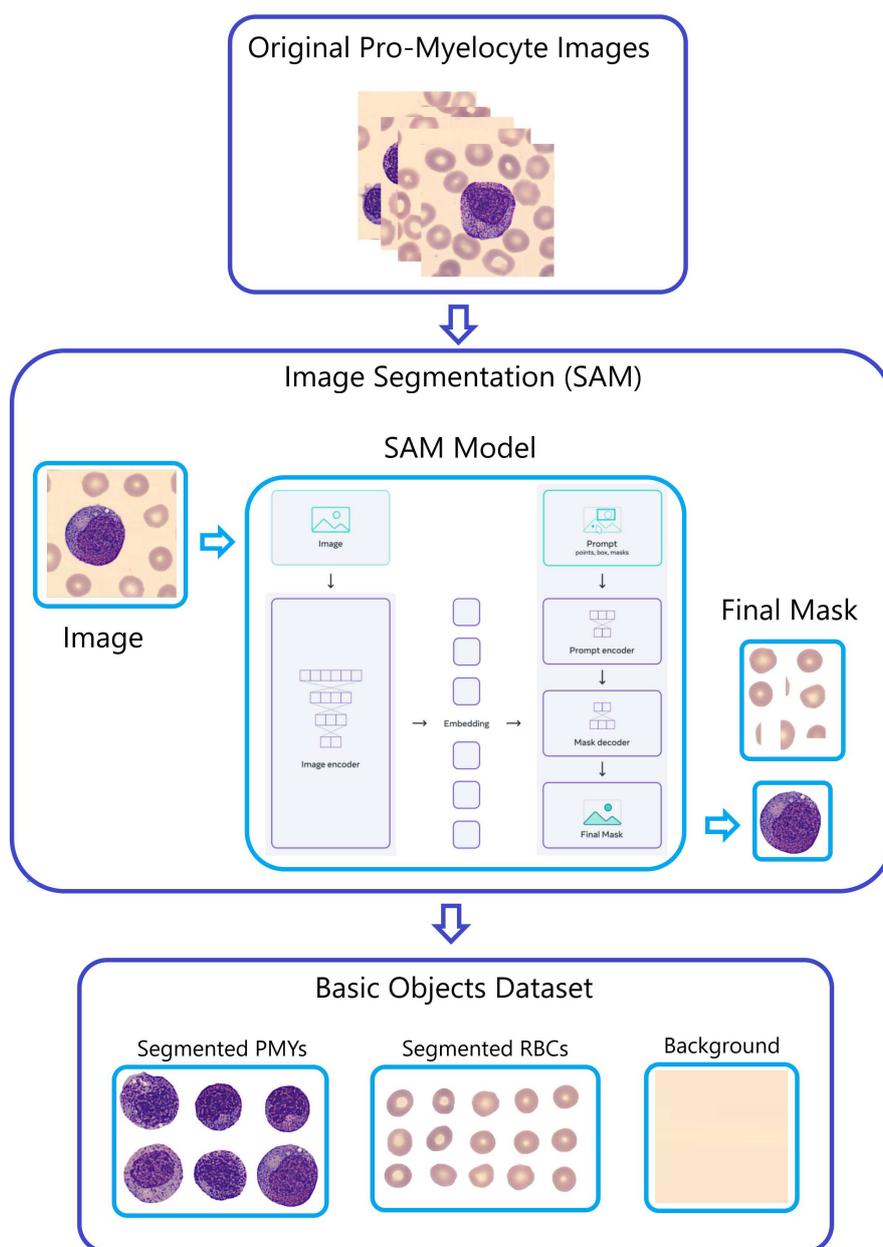
Figure 5. The "Naturalize" first step - Segmentation.

- **Step 2 - Generating Composite Images** (Figure 6):

  The SAM-segmented RBCs were combined with randomly selected WBCs from the "BNE, MMY, MY, PMY, and SNE" datasets to create composite images. The primary objective in this selection and fusion of segmented RBCs and WBCs is to generate entirely distinct images each time the Naturalize code is run.

  Converting the segmented cells into the RGBA format is aimed at resetting the alpha transparency channel within these segments, preventing their black backgrounds from appearing in the newly generated images using the Naturalize method. However, the size of segmented RBCs and WBCs remains unchanged to mirror the real dimensions of blood cells in peripheral blood smears.

  Determining the size (width and height) of the segmented RBCs and WBCs functions as input for the Collision-Free mechanism, ensuring that the segmented RBC images avoid overlapping

or excessive stacking on either themselves or the segmented WBC images when added to the background image. This function's output results in well-organized peripheral blood cell images without cells accumulating on top of each other.

This procedural demonstration is depicted in Figure 5 and Figure 6, highlighting the creation of composite PMY images as an illustrative example.
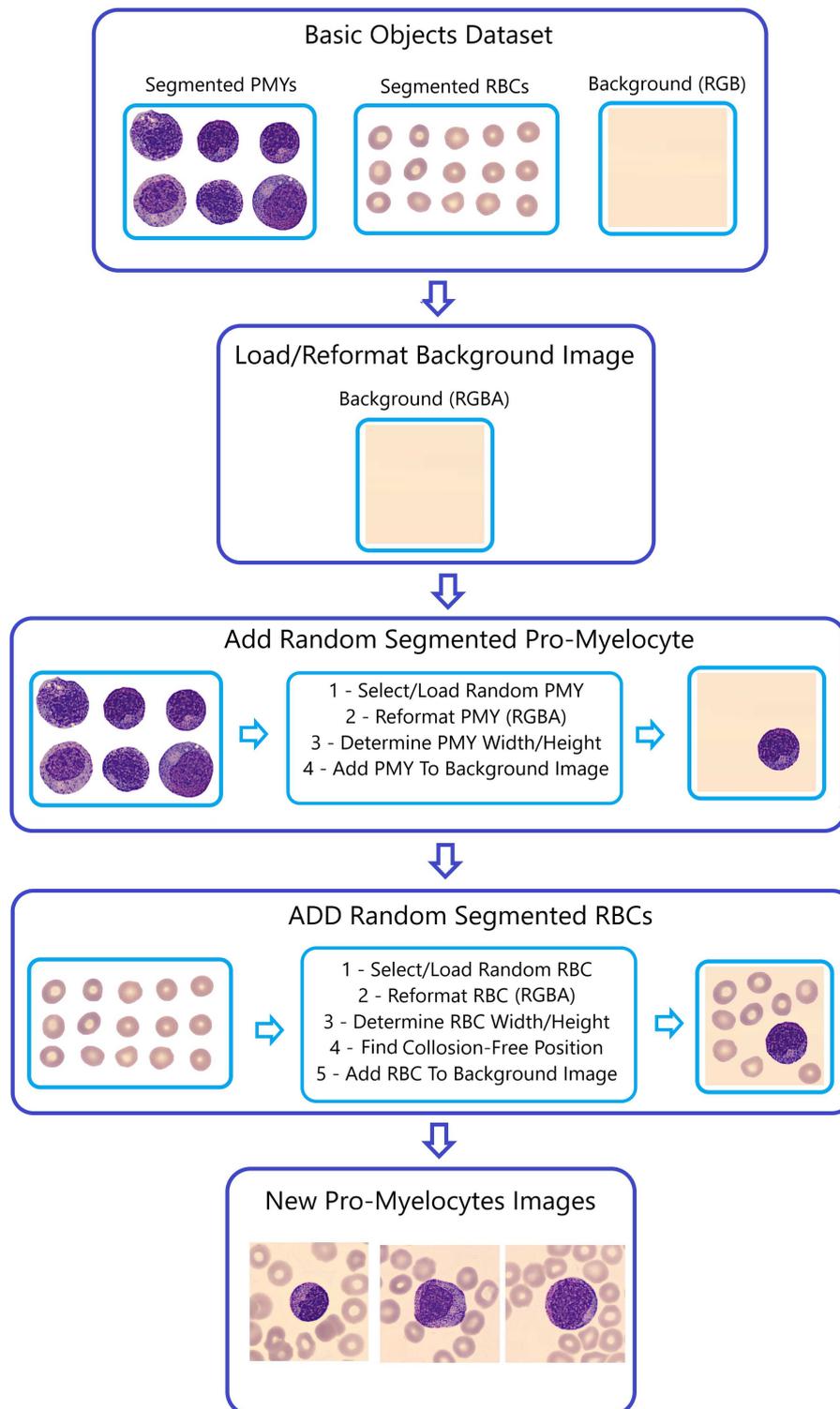

Figure 6. The "Naturalize" second step - Composite Image Generation.

Informed by our comprehensive exploratory data analysis (EDA), we have carefully curated specific images across multiple classes, underscoring our dedication to maintaining data quality. For instance, upon the exclusion of multi-class images from the pro-myelocyte category, the count within the original 11-class PBC dataset reduced from 592 to 419 images, a process similarly carried out across other classes. Following this pruning process, the overall count of images in the original PBC dataset consolidates to 16,115 images.

During the initial application of the "Naturalize" technique, the count of pro-myelocytes surged from 419 images to 1,011 images. Table 3 vividly represents the remarkable evolution of the original 419-PMY PBC dataset across 11 distinct blood cell classes, notably bolstered by the addition of new created 592 PMY images. Consequently, the cumulative count of images within the augmented 419-PMY PBC dataset now elevates to 16,345 images.

Table 3. Summary of the augmented 419-PMY PBC dataset

#	Cell Class	Symbol	Total Images By Class	%
1	Basophil	BA	1,123	6.8
2	Banded Neutrophil	BNE	1,621	9.8
3	Eosinophil	EO	2,937	17.8
4	Erythroblast	ERB	1,351	8.2
5	Lymphocyte	LY	1,202	7.3
6	Meta-myelocyte	MMY	1,007	6.1
7	Monocyte	MO	1,350	8.2
8	Myelocyte	MY	1,134	6.9
9	Platelet	PLT	2,178	13.2
10	Pro-myelocyte	PMY	1,011	6.1
11	Segmented Neutrophil	SNE	1,620	9.8
	Total		16,345	100

The dataset underwent a significant expansion through two additional applications of the "Naturalize" augmentation technique, resulting in the creation of the 1K-PBC and 2K-PBC datasets. Each of these datasets was enriched by adding 1,000 and 2,000 images per sub-dataset from the following five blood classes: "BNE, MMY, MY, PMY, and SNE." Consequently, the 1K-PBC dataset now comprises: 1,123 BA, 2,621 BNE, 2,937 EO, 1,351 ERB, 1,202 LY, 2,007 MMY, 1,350 MO, 2,134 MY, 2,178 PLT, 2,011 PMY, and 2,620 SNE. Additionally, the 2K-PBC dataset includes: 1,123 BA, 3,621 BNE, 2,937 EO, 1,351 ERB, 1,202 LY, 3,007 MMY, 1,350 MO, 3,134 MY, 2,178 PLT, 3,011 PMY, and 3,620 SNE.

The decision behind the sole inclusion of the following five blood cell classes "BNE, MMY, MY, PMY, and SNE" in the second and third "Naturalize" application is based on the results of the classification report and aims to improve classification performance.

This forward-thinking enhancement approach propelled the augmented 419-PMY dataset to an impressive 21,534 images, rebranded as the "1K-PBC" dataset, and further expanded it to 26,534 images, now recognized as the "2K-PBC" dataset. This heralds a transformative era of robust and enriched data resources.

The prior datasets, namely 419-PMY, 1K-PBC, and 2K-PBC, generated using the Naturalize augmentation technique, all exhibited imbalanced class distributions. In response to this imbalance issue, we've created a new dataset called Balanced PBC, where all classes have been rebalanced with 2,000 images each. As a result, certain classes (BA, BNE, ERB, LY, MMY, MO, MY, PMY, and SNE) have been augmented to include 2,000 images, while the number of images in the remaining classes (EO and PLT) has been reduced to 2,000 each.

3.4. Comparison between Naturalize and conventional augmentation techniques

Conventional image augmentation involves standard transformations such as rotation, flipping, and color adjustments to expand datasets. These techniques focus on general image manipulation to increase variety. On the other hand, the "Naturalize" augmentation method is more complex and

specific. It employs a selective segmentation process using the "Segment Anything Model" to isolate particular cell classes from the dataset. This segmentation approach is based on findings from prior data analysis, emphasizing the enhanced performance of a subset of classes.

Moreover, the number of segmented Red Blood Cells (RBCs), White Blood Cells (WBCs), and Platelets (PLTs) generated from the original Peripheral Blood Cell (PBC) dataset using the SAM model is significantly large. The random addition of these segmented objects results in an incredibly vast number of unique and realistic replicas of the original PBC dataset.

In addition, "Naturalize" offers control over the quality of the created images by enabling the regulation of the number of added segmented RBCs with the selected segmented WBCs. For instance, one can selectively choose to include 4 or 8 segmented RBCs, providing a level of precision and control not commonly found in traditional augmentation techniques.

Furthermore, the versatility of the "Naturalize" technique extends beyond medical imaging. By segmenting all objects in the original images and reintroducing them into the background images, "Naturalize" can be applied to various applications, both within the medical field and beyond. This adaptability showcases the potential for widespread use, not limited to medical image augmentation.

This method maintains the realism of cell sizes, preserving the authentic dimensions of blood cells. It also ensures that the combined cells do not overlap excessively through a "Collision-Free" mechanism, a feature not typically found in traditional augmentation methods. Overall, "Naturalize" focuses on medical image authenticity and diversity, tailoring the augmentation process to suit specific requirements rather than applying generic transformations.

3.5. Augmented 419-PMY PBC Dataset preprocessing

The preprocessing of the augmented 419-PMY PBC dataset involved three primary steps:

- **Step 1 - Image Resizing:** The images were resized to match the standard "224x224" image input size required by pre-trained ImageNet ConvNets and ViT models.
- **Step 2 - Second Data Augmentation (Transformational):** The training dataset underwent a second augmentation step, which included geometric transformations such as horizontal flips, vertical flips, and 90-degree rotations. This augmentation aimed to mitigate overfitting of models.
- **Step 3 - Data Splitting:** The final 11-class PBC dataset was split into three subsets: an 80% training set, a 10% validation set, and a 10% testing set.

3.6. Models & DL techniques (TL/FT/EL)

Three types of model architectures were utilized in this study: pre-trained ImageNet ConvNets, pre-trained Vision Transformers (ViT), and customized multi-layer CNN models. Additionally, three DL techniques were employed to train pre-trained models: Transfer Learning (TL), Fine-Tuning (FT), and Ensemble Learning (EL).

3.6.1. Pre-trained ImageNet ConvNets

Pre-trained ImageNet models are an explicit example of ConvNets, which are trained on a large dataset.

Pre-trained ImageNet models were used as the foundation for the study. Notable models employed in this research included ConvNexTBase [25], DenseNet-121 [26], DenseNet-169 [26], DenseNet-201 [26], EfficientNetV2 B0 [27], and VGG-19 [28]. Figure 7 provides an example of the architecture of the VGG-19 model when applied to classify neutrophils.
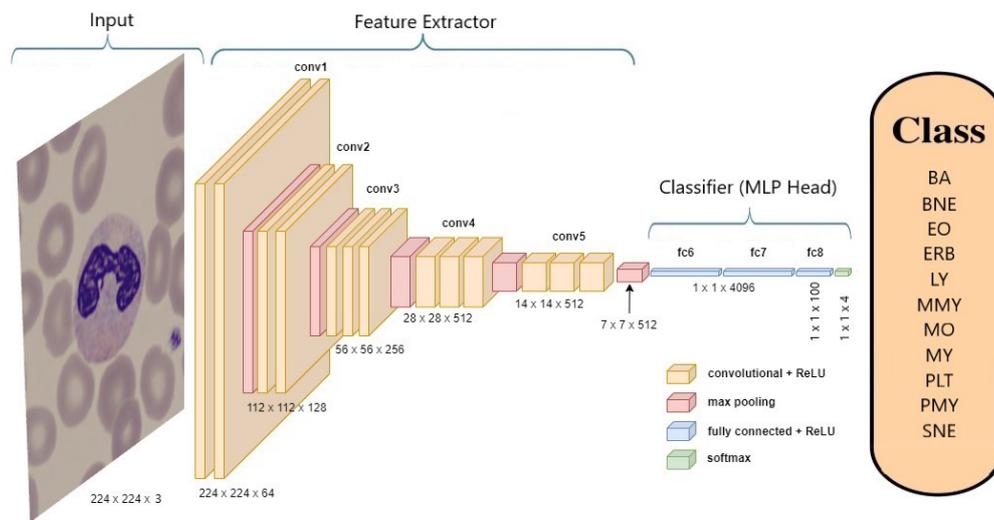
Figure 7. Architecture of VGG-19 model classifying a neutrophil.

3.6.2. Pre-trained Vision Transformer (ViT)

The Vision Transformer (ViT) [6] architecture, based on the transformer architecture commonly used in Natural Language Processing (NLP), was applied. It involved splitting input images into small patches and processing each patch through a transformer encoder. Unlike convolutional layers, ViT used self-attention mechanisms to extract features from input images, enabling the model to consider the entire image at once. The study utilized the "ViTb16" architecture with 12 encoder blocks, and Figure 8 illustrates its application to classify neutrophils [6].

Figure 8. Architecture of ViT classifying a neutrophil.

3.6.3. Customized multi-layer CNN

Customized multi-layer CNN models were developed and tested using the final 11-class PBC dataset. Figure 9 provides a schematic representation of a typical customized CNN model architecture.

```
_____
Layer (type)                 Output Shape              Param #
===================================================================
conv2d (Conv2D)              (None, 224, 224, 16)       448

max_pooling2d (MaxPooling2D  (None, 112, 112, 16)       0
)

conv2d_1 (Conv2D)            (None, 112, 112, 32)       4640

max_pooling2d_1 (MaxPooling  (None, 56, 56, 32)         0
2D)

conv2d_2 (Conv2D)            (None, 56, 56, 64)         18496

max_pooling2d_2 (MaxPooling  (None, 28, 28, 64)         0
2D)

conv2d_3 (Conv2D)            (None, 28, 28, 128)        73856

max_pooling2d_3 (MaxPooling  (None, 14, 14, 128)        0
2D)

flatten (Flatten)            (None, 25088)              0

dense (Dense)                (None, 128)                3211392

dense_1 (Dense)              (None, 11)                 1419

===================================================================
Total params: 3,310,251
Trainable params: 3,310,251
Non-trainable params: 0
_____
```
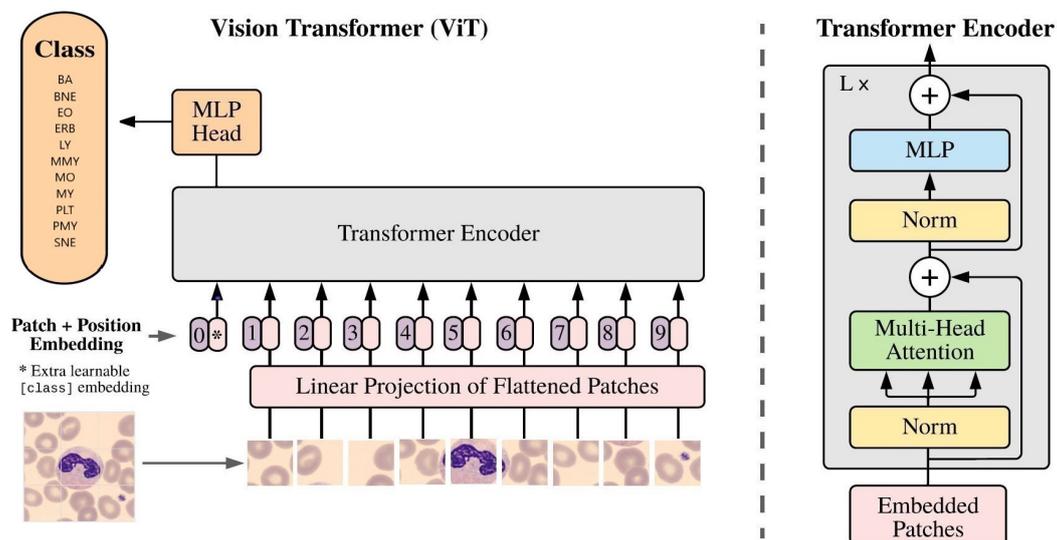
Figure 9. Customized multi-layer CNN model's architecture.

3.6.4. DL Techniques (TL / FT / EL)

Pre-trained ImageNet model has Convolutional Base (feature extractor) and Classifier (MLP head). Transfer learning swaps MLP head for a new one, retraining on a custom dataset. Convolutional base stays non-trainable. Fine-tuning retrains Convolutional Base and MLP head, adapting both for a new learning task.

Ensemble learning (EL) unites results from high-performing transfer-learned or fine-tuned models, utilizing either averaging or the Convolutional Block Attention Module (CBAM)[24]. The objective is to attain enhanced overall model performance. In their study[17], the authors applied Ensemble-CBAM with their top-scoring pre-trained models. To replicate their findings, this work also utilizes the Ensemble-CBAM model.

CBAM is a tool that turbocharges the effectiveness of convolutional neural networks (CNNs) by focusing on two critical aspects: the inter-dependencies among channels (channel attention) and the spatial locations within feature maps (spatial attention). When added to a CNN, CBAM directs the model's attention toward relevant data while filtering out irrelevant noise, leading to better

performance and more resilient representations. The CBAM module is performed on the merged features of the two nets (average or concatenation).

Equation 1 ([30]) outlines how the channel attention block weights are computed:

$$\alpha_c = \sigma\left(\text{MLP}\left(\text{AvgPool}(x_c)\right) + \text{MLP}\left(\text{MaxPool}(x_c)\right)\right) \tag{1}$$

Here, α_c stands for the attention weights for channel c within the CNN's output feature maps. x_c represents the feature maps for channel c, which are the input to the CAM module. σ denotes the sigmoid function. MLP denotes a multi-layer perceptron. The MaxPool and AvgPool operations denote the max pooling and average pooling operations over the spatial dimensions of the feature maps, respectively.

Equation 2 ([30]) demonstrates the calculation for the spatial attention block weights:

$$\beta_{i,j} = \sigma\left(\text{MLP}\left([x_{i,j}; \text{AvgPool}(x); \text{MaxPool}(x)]\right)\right) \tag{2}$$

In this equation, $\beta_{i,j}$ represents the attention weight assigned to the spatial location i, j in the input feature maps x. The global average pooling and max pooling operations, denoted by $\text{AvgPool}(x)$ and $\text{MaxPool}(x)$ respectively, are used to extract global information from the feature maps and reduce their spatial dimensionality. The concatenated feature vector $[x_{i,j}; \text{AvgPool}(x); \text{MaxPool}(x)]$ for the spatial location i, j is passed through an MLP layer, which applies a non-linear transformation to the feature vector. The output of the MLP layer is passed through a sigmoid activation function, denoted by σ, to obtain the final attention weight $\beta_{i,j}$.

In the Ensemble-CBAM model, the top-performing models, namely "EfficientNetV2 B0 and DenseNet-169," are combined using CBAM and an MLP head.

Three deep learning techniques (TL/FT/EL) are employed in this study to create an optimal DL PBS tool.

3.7. Results' analysis and interpretability tools

In addition to the accuracy metrics (accuracy and loss), three results' analysis and interpretability tools are used. These are confusion matrix, classification reports, and Score-CAM.

3.7.1. Confusion matrix

A confusion matrix [31], or error matrix, visualizes algorithm performance, especially in supervised learning. Rows represent actual classes; columns represent predicted classes. Figure 10 displays one in multi-class classification, showing "TN and TP" for correctly classified negative and positive cases, and "FN and FP" for misclassified cases.
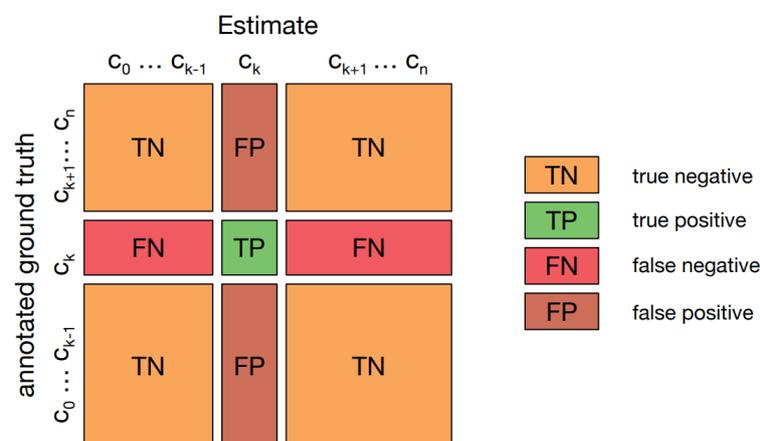

Figure 10. Confusion matrix for multi-class classification.

3.7.2. Classification report

A classification report [31] evaluates prediction quality using precision, recall, and F1-score per class, along with macro and weighted average accuracies. Accuracy, calculated as a percentage of correct predictions, is determined by equation 3 [31]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Precision measures the quality of a positive prediction made by the model and the equation 4 [31] demonstrates its computational process:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Recall measures how many of the true positives (TPs) were recalled (found) and calculated using the equation 5 [31]:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F1-Score is the harmonic mean of precision and recall and can be calculated using the equation 6 [31]:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{6}$$

3.7.3. Score-CAM

A Score-CAM [12] is a score-weight visual explanation method based on a class activation mapping (CAM) for CNN models. It helps to understand the internal mechanism of CNN models.

4. Results & Discussions

4.1. Results

This section provides a comprehensive overview of experiments on 11-class blood cell classification using diverse models, including pre-trained ImageNet models ("ConvNextBase," "DenseNets," "EfficientNetB0," "VGG-19," "ViTb16"), customized CNNs, and ensembles. Experiments were conducted on the challenging PBC dataset, addressing class imbalance by introducing the "Naturalize" augmentation technique, creating "augmented 419-PMY," "1K-PBC," "2K-PBC," and "Balanced PBC" datasets. Model performance was evaluated numerically using confusion matrices, classification reports, and visually with Score-CAM.

4.1.1. Augmented 419-PMY PBC Dataset Results

Initially, all pre-trained models were fitted using transfer learning, but it was observed that fine-tuning led to better results. Table 4 presents the accuracy scores of the training, validation, and testing subsets of the augmented 419-PMY PBC dataset for the fine-tuned models. Notably, the EfficientNetV2 B0 model achieved the highest training and validation accuracies, while the DenseNet-169 model recorded the highest testing accuracy.

Table 5 provides the macro-average precision, recall, and F1-score of the testing subset of the augmented 419-PMY PBC dataset for the fine-tuned models, with the DenseNet-169 model achieving the best results.

Given the superior performance of the EfficientNetV2 B0 model in the training and validation subsets, it was selected for subsequent trials.

Table 4. Summary of Models' Training, Validation, and Testing Accuracies.

Model	Accuracy		
	Training	Validation	Testing
ConvNexTBase	94.57 %	92.42 %	90.85 %
Customized-CNN	90.15 %	88.67 %	87.27 %
DenseNet-121	93.68 %	94.67 %	93.52 %
DenseNet-169	94.81 %	94.67 %	94.48 %
DenseNet-201	93.94 %	95.09 %	93.94 %
EfficientNetV2 B0	96.26 %	96.06 %	93.82 %
Ensemble-Average	95.41 %	94.24 %	94.00 %
Ensemble-CBAM	96.11 %	95.12 %	94.08 %
ViTb16	89.46 %	89.70 %	89.12 %
VGG-19	94.68 %	93.12 %	91.75 %

Table 5. Summary of Models' Macro-average Precisions, Recalls, and F1-Scores.

Model	Macro Average		
	Precision	Recall	F1-Score
ConvNexTBase	89%	89%	89%
Customized-CNN	87%	87%	87%
DenseNet-121	92%	93%	92%
DenseNet-169	**94%**	**94%**	**94%**
DenseNet-201	93%	93%	93%
EfficientNetV2 B0	93%	93%	93%
Ensemble-Average	93%	93%	93%
Ensemble-CBAM	93%	93%	93%
ViTb16	86%	87%	86%
VGG-19	89%	90%	90%

4.1.2. EfficientNetV2 B0

Table 6 presents the classification report of the fine-tuned EfficientNetV2 B0 model using the original PBC dataset.

Table 6. EfficientNetV2 B0 - Classification report for the original PBC.

Class	Precision	Recall	F1-score	Support
BA	1.00	1.00	1.00	112
BNE	0.88	0.77	0.82	162
EO	1.00	1.00	1.00	293
ERB	1.00	1.00	1.00	135
LY	0.99	1.00	1.00	120
MMY	0.77	0.88	0.82	100
MO	1.00	0.96	0.98	135
MY	0.74	0.80	0.77	113
PLT	1.00	1.00	1.00	217
PMY	0.80	0.59	0.68	59
SNE	0.83	0.91	0.87	162
Accuracy			0.93	1608
Macro Avg	0.91	0.90	0.90	1608
Weighted Avg	0.93	0.93	0.93	1608

Table 7 displays the classification report of the fine-tuned EfficientNetV2 B0 model using the augmented 419-PMY PBC dataset.

Table 7. EfficientNetV2 B0 - Classification report for the augmented 419-PMY PBC

Class	Precision	Recall	F1-score	Support
BA	1.00	1.00	1.00	112
BNE	0.88	0.77	0.82	162
EO	1.00	1.00	1.00	293
ERB	1.00	1.00	1.00	135
LY	0.99	1.00	1.00	120
MMY	0.75	0.88	0.81	100
MO	0.99	0.96	0.98	135
MY	0.87	0.80	0.83	113
PLT	1.00	1.00	1.00	217
PMY	0.91	0.89	0.90	101
SNE	0.83	0.91	0.87	162
Accuracy			0.94	1650
Macro Avg	0.93	0.93	0.93	1650
Weighted Avg	0.94	0.94	0.94	1650

Tables 8, 9, and 10 illustrate the classification reports of the fine-tuned EfficientNetV2 B0 model using the 1K-PBC dataset, 2K-PBC dataset, and Balanced PBC dataset, respectively.

Table 8. EfficientNetV2 B0 - Classification report for the 1K-PBC.

Class	Precision	Recall	F1-score	Support
BA	0.99	1.00	1.00	112
BNE	0.90	0.91	0.90	262
EO	1.00	1.00	1.00	293
ERB	1.00	0.99	1.00	135
LY	0.98	1.00	0.99	120
MMY	0.87	0.94	0.90	200
MO	1.00	0.98	0.99	135
MY	0.92	0.88	0.90	213
PLT	1.00	1.00	1.00	217
PMY	0.98	0.94	0.96	201
SNE	0.91	0.90	0.91	262
Accuracy			0.95	2150
Macro Avg	0.96	0.96	0.96	2150
Weighted Avg	0.95	0.95	0.95	2150

Table 9. EfficientNetV2 B0 - Classification report for the 2K-PBC

Class	Precision	Recall	F1-score	Support
BA	0.98	1.00	0.99	112
BNE	0.92	0.96	0.94	362
EO	1.00	1.00	1.00	293
ERB	0.99	0.99	0.99	135
LY	0.98	1.00	0.99	120
MMY	0.93	0.96	0.94	300
MO	0.98	0.98	0.98	135
MY	0.94	0.92	0.93	313
PLT	1.00	1.00	1.00	217
PMY	0.97	0.95	0.96	301
SNE	0.96	0.92	0.94	362
Accuracy			0.96	2650
Macro Avg	0.97	0.97	0.97	2650
Weighted Avg	0.96	0.96	0.96	2650

Table 10. EfficientNetV2 B0 - Classification report for the Balanced PBC.

Class	Precision	Recall	F1-score	Support
BA	1.00	1.00	1.00	200
BNE	0.90	0.86	0.88	200
EO	1.00	1.00	1.00	200
ERB	1.00	0.99	1.00	200
LY	1.00	1.00	1.00	200
MMY	0.92	0.92	0.92	200
MO	0.99	0.98	0.98	200
MY	0.87	0.93	0.90	200
PLT	1.00	1.00	1.00	200
PMY	0.95	0.90	0.93	200
SNE	0.88	0.91	0.89	200
Accuracy			0.95	2200
Macro Avg	0.95	0.95	0.95	2200
Weighted Avg	0.95	0.95	0.95	2200

5. Discussion

This research represents a significant step forward in the application of deep learning to automate the reading, analysis, and classification of blood cells. Table 11 summarizes the macro-average precisions, recalls, F1-scores, and accuracy of the fine-tuned EfficientNetV2 B0 model across the testing subsets of the imbalanced "augmented 419-PMY," "1K-PBC," "2K-PBC" datasets, and the "Balanced PBC" dataset.

Table 11. EfficientNetV2 B0 - Macro-average precisions, recalls, and F1-Scores.

PBC Datasets	Macro Average			
	Precision	Recall	F1-Score	Accuracy
Imbalanced Datasets				
Original PBC	0.91	0.90	0.90	0.93
Augmented 419-PMY PBC	0.93	0.93	0.93	0.94
1K-PBC	0.96	0.96	0.96	0.95
2K-PBC	0.97	0.97	0.97	0.96
Balanced Dataset				
Balanced PBC	0.95	0.95	0.95	0.95

Table 11 demonstrates that while the imbalanced nature of the 11 blood classes led to some minor overfitting, the application of the "Naturalize" augmentation technique substantially improved macro-average precision, recall, and F1-score, increasing from 90% with the original PBC dataset to 97% with the 2K-PBC dataset. These findings are further supported by the use of Score-CAM, as shown in Figure 11, which visualizes the fine-tuned pre-trained DenseNet-169 model with the augmented 419-PMY PBC dataset.

Conventional image augmentation methods, which involve actions like rotation and color adjustments, primarily aim to enhance general diversity in datasets. In stark contrast, the "Naturalize" technique represents a highly sophisticated approach that leverages precise segmentation to extract specific cell classes such as Red Blood Cells (RBCs), White Blood Cells (WBCs), and Platelets (PLTs) from the dataset. Through the combination of these segmented elements, "Naturalize" creates an extensive and authentic expansion of the original dataset. What sets this technique apart is its unparalleled level of control, allowing for precise selection and regulation of specific cell quantities, all while maintaining realism and preventing overlap thanks to a 'Collision-Free' mechanism.
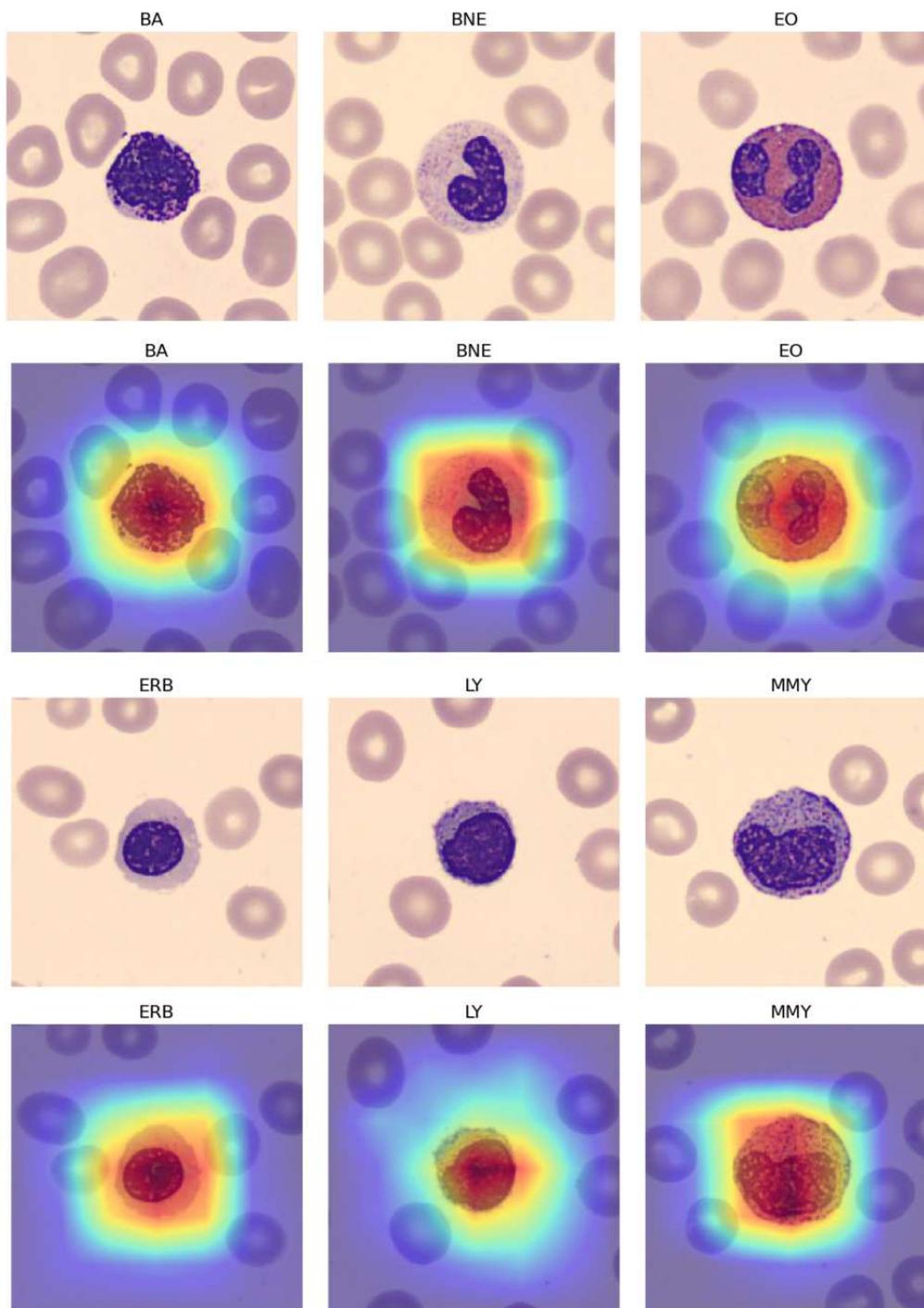
Figure 11. Score-CAM for fine-tuned DenseNet-169.

By effectively addressing the challenges posed by data insufficiency and class imbalance through the "Naturalize" augmentation method, this study lays a solid foundation for the broader utilization of deep learning algorithms and models in blood cell classification. The implementation of "Naturalize" holds tremendous promise for the development of comprehensive and balanced peripheral blood smear datasets.

Furthermore, it's important to note that 'Naturalize' is not confined to the realm of medical imaging alone; its adaptability extends to various fields, underscoring its potential for widespread

application. This method stands out for its unwavering commitment to authenticity and tailored diversity, ensuring the preservation of genuine cell sizes and the fulfillment of specific augmentation requirements.

6. Conclusions

In this study, we have presented a comprehensive approach to automated blood cell classification, addressing the challenges posed by an imbalanced dataset and limited sample sizes within specific classes. Our methodology leverages state-of-the-art deep learning models, including pre-trained ConvNets and ViTb16, along with a customized multi-layer CNN model. Through the application of transfer learning, fine-tuning, and ensemble learning techniques, we have achieved remarkable results on the 11-class PBC dataset.

Our most significant achievement lies in the fully fine-tuned EfficientNetV2 B0 model, which demonstrated exceptional performance on the original PBC dataset, with a macro-average precision, recall, and F1-score of 91%, 90%, and 90%, respectively, and an impressive average accuracy of 93%. This success underscores the effectiveness of our approach in addressing the inherent challenges of blood cell classification.

Furthermore, we introduced the "Naturalize" augmentation technique, a novel and innovative approach to generating synthetic blood cell samples with the same quality as the original dataset. The resulting 2K-PBC dataset, augmented using "Naturalize," achieved outstanding results, with a macro-average precision, recall, and F1-score of 97%, and an average accuracy of 96% when employing the fully fine-tuned EfficientNetV2 B0 model.

Our research not only provides a robust solution for blood cell classification but also contributes to the broader field of medical image analysis by addressing the common issues of insufficient and imbalanced data. The "Naturalize" augmentation technique opens up new possibilities for generating high-quality synthetic data, which can enhance the performance and generalization of deep learning models in various medical image analysis tasks.

In conclusion, this study represents a significant step forward in the automated classification of blood cells, offering a powerful tool for medical professionals and researchers. The combination of advanced deep learning models and innovative data augmentation techniques has the potential to revolutionize the field, leading to more accurate and reliable diagnostic tools for blood-related disorders and diseases. We anticipate that our work will inspire further research and advancements in this critical area of healthcare and medical imaging.

Author Contributions: Conceptualization, M.A.A. and F.D.; methodology, M.A.A. and F.D.; software, M.A.A.; validation, M.A.A.; formal analysis, M.A.A., F.D. and I.A.-C.; investigation, M.A.A.; resources, M.A.A. and F.D.; data curation, M.A.A.; writing—original draft preparation, M.A.A. and F.D.; writing—review and editing, M.A.A., F.D. and I.A.-C.; supervision, F.D. and I.A.-C.; project administration, F.D. and I.A.-C.; funding acquisition, F.D. and I.A.-C. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1.   S. B. McKenzie and L. Williams, Clinical laboratory hematology, 3rd ed. Upper Saddle River, NJ: Pearson, 2014.
2.   Rodak, B. F., & Carr, J. H. (2012). Clinical Hematology Atlas (4th ed.). Saunders.

3.   R. Al-qudah and C. Y. Suen, "Peripheral blood smear analysis using deep learning: Current challenges and future directions," in *Computational Intelligence and Image Processing in Medical Applications*, WORLD SCIENTIFIC, 2022, pp. 105–118. https://doi.org/10.1142/9789811257452_0007

4.   A. Acevedo, A. Merino, S. Alférez, Á. Molina, L. Boldú, and J. Rodellar, "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems," *Data Brief*, vol. 30, no. 105474, p. 105474, 2020. https://doi.org/10.1016/j.dib.2020.105474

5.   O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," 2014. https://doi.org/10.1007/s11263-015-0816-y

6.   A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020. https://doi.org/10.48550/arXiv.2010.11929.

7.   J. Puigcerver et al., "Scalable transfer learning with expert models," *Computing Research Repository (CoRR)*, 2020. https://doi.org/10.1109/CVPR42600.2020.00395

8.   H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Med. Imaging*, vol. 22, no. 1, p. 69, 2022. https://doi.org/10.1186/s12880-022-00793-7

9.   K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, 2016. https://doi.org/10.1186/s40537-016-0043-6

10. A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 2, pp. 757–774, 2023. https://doi.org/10.1016/j.jksuci.2023.01.014

11. I. Düntsch and G. Gediga, "Confusion matrices and rough set data analysis," 2019. https://doi.org/10.1088/1742-6596/1229/1/012055

12. H. Wang et al., "Score-CAM: Score-weighted visual explanations for convolutional neural networks," 2019. https://doi.org/10.48550/arXiv.1910.01279

13. C. Jung, M. Abuhamad, J. Alikhanov, A. Mohaisen, K. Han, and D. Nyang, "W-Net: A CNN-based architecture for white blood cells image classification," *arXiv* [eess.IV], 2019. https://doi.org/10.48550/arXiv.1910.01091

14. [14] A. T. Sahlol, P. Kollmannsberger, and A. A. Ewees, "Efficient classification of white Blood Cell Leukemia with improved swarm optimization of deep features," *Sci. Rep.*, vol. 10, no. 1, p. 2536, 2020. https://doi.org/10.1038/s41598-020-59215-9

15. K. Almezhghwi and S. Serte, "Improved classification of white blood cells with the generative adversarial network and deep convolutional neural network," *Comput. Intell. Neurosci.*, vol. 2020, p. 6490479, 2020. https://doi.org/10.1155/2020/6490479

16. S. Tavakoli, A. Ghaffari, Z. M. Kouzehkanan, and R. Hosseini, "New segmentation and feature extraction algorithm for classification of white blood cells in peripheral smear images," *Sci. Rep.*, vol. 11, no. 1, p. 19428, 2021. https://doi.org/10.1038/s41598-021-98599-0

17. H. Chen et al., "Accurate classification of white blood cells by coupling pre-trained ResNet and DenseNet with SCAM mechanism," *BMC Bioinformatics*, vol. 23, no. 1, p. 282, 2022. https://doi.org/10.1186/s12859-022-04824-6

18. O. Katar and İ. F. Kilinçer, "Automatic classification of white blood cells using pre-trained deep models," *Sakarya University Journal of Computer and Information Sciences*, vol. 5, no. 3, pp. 462–476, 2022. https://doi.org/10.35377/saucis...1196934

19. S. Nahzat, "White Blood Cell Classification Using Convolutional Neural Network," *Technology and Engineering Research*, vol. 3, no. 1, pp. 32–41, 2022. https://doi.org/10.53525/jster.1018213

20. A. Heni, "Blood Cells Classification Using Deep Learning with customized data augmentation and EK-means segmentation," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 3, 2023. https://www.jatit.org/volumes/Vol101No3/11Vol101No3.pdf

21. Z. Zhu, S.-H. Wang, and Y.-D. Zhang, "ReRNet: A deep learning network for classifying blood cells," *Technol. Cancer Res. Treat.*, vol. 22, 2023. https://doi.org/10.1177/15330338231165856

22. Z. Zhu et al., "DLBCNet: A deep learning network for classifying blood cells," *Big Data Cogn. Comput.*, vol. 7, no. 2, p. 75, 2023. https://doi.org/10.3390/bdcc7020075

23. A. Kirillov et al., "Segment Anything," 2023. https://doi.org/10.48550/arXiv.2304.02643

24. Zhang, A., Lipton, Z.C., Li, M., Smola, A.J.: "Dive into deep learning." *Computing Research Repository (CoRR)*, (2021). https://doi.org/10.1109/CVPR.2017.243

25. Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," 2022. https://doi.org/10.48550/arXiv.2201.03545

26. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: "Densely connected convolutional networks". In: *2017 IEEE Conference on Com- puter Vision and Pattern Recognition (CVPR)*, pp. 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243

27. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional Neural Networks," 2019. https://doi.org/10.48550/arXiv.1905.11946

28. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. https://doi.org/10.48550/arXiv.1409.1556

29. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Computer Vision – ECCV 2018, Cham: Springer International Publishing, 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

30. A. Ganguly, A. U. Ruby, and G. C. Chandran, "Evaluating CNN architectures using attention mechanisms: Convolutional Block Attention Module, Squeeze, and Excitation for image classification on CIFAR10 dataset," *Research Square*, 2023. https://doi.org/10.21203/rs.3.rs-3245213/v1

31. H. Dalianis, "Evaluation Metrics and Evaluation," in Clinical Text Mining, Cham: *Springer International Publishing*, 2018, pp. 45–53. https://doi.org/10.1007/978-3-319-78503-5_6