SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
VOL. 5, NO. 3, DECEMBER 2022
DOI: 10.35377/saucis.05.03.1196934

Research/Review Article

Automatic Classification of White Blood Cells Using Pre-Trained Deep Models

Oguzhan Katar[1], Ilhan Firat Kilincer[2]
[1]Corresponding Author; Firat University, Department of Software Engineering; okatar@firat.edu.tr; 0000-0002-5628-3543; +90 424 607 81 07
[2]Firat University, Department of Informatics; ifkilincer@firat.edu.tr; 0000-0001-8090-4998

Received 31 October 2022; Revised 1 November 2022; Accepted 22 December 2022; Published online 31 December 2022

Abstract

White blood cells (WBCs), which are a crucial component of the immune system, help our body defend against infections and other diseases. Some diseases may cause our body to produce fewer WBCs than it requires. Therefore, WBCs are of great importance in medical imaging. Artificial intelligence-based computer systems can assist experts in analyzing WBCs. In this study, we proposed an approach for the automatic classification of WBCs into five different classes using a pre-trained model. We trained ResNet-50, VGG-19, and MobileNet-V3-Small pre-trained models with ImageNet weights. For the training, validation, and testing processes of the models, we used a public dataset containing 16,633 images with an uneven class distribution. While the ResNet-50 model achieved an accuracy of 98.79%, the VGG-19 model achieved an accuracy of 98.19%, and the MobileNet-V3-Small model achieved the highest accuracy rate at 98.86%. When examining the predictions of the MobileNet-V3-Small model, we observed that it was not affected by class dominance and was able to correctly classify even the least sampled class images in the dataset. In addition to the high accuracy achieved in the classification of WBCs using the proposed pre-trained deep learning models, we also applied the Grad-CAM method to further understand and interpret the model's predictions.

Keywords: white blood cells, classification, pre-trained models, artificial intelligence, Grad-CAM

1. Introduction

Blood is a vital fluid that helps to nourish the body, maintain acid-base balance, transport hormones, and maintain salt and water balance. Blood consists of three types of cells: erythrocytes, platelets, and leukocytes [1].

Erythrocytes, the most abundant type of blood cell, contain a substance called hemoglobin, which is responsible for transporting oxygen in the body [2]. Oxygen, inhaled into the lungs through respiration and then entering the blood, can be transported to all body tissues with the help of hemoglobin in erythrocytes. Adequate oxygen access to each cell in the body depends on the sufficient number and function of erythrocytes in the blood. Erythrocytes, which are reddish in color and therefore also referred to as red blood cells, obtain their color from the iron mineral in the structure of hemoglobin [3].

Platelets are cell fragments that are formed by the disintegration of cells called megakaryocytes in the bone marrow tissue located in the center of our bones after they mature and enter the blood [4]. Platelets play a vital role in regulating certain chemical reactions that occur in the blood due to the biochemical substances they contain [5]. However, their primary function is in the case of bleeding due to injury to blood vessels; they help to quickly close and repair the wounded area.

Leukocytes, also known as white blood cells (WBCs), are an important part of the immune system and a group of cells that protect the body against infections [6]. When the body encounters foreign organisms, they reproduce rapidly. The primary function of leukocytes is to identify and eliminate antigens such as bacteria, viruses, fungi, and poisonous toxins that have entered the body in various ways. Leukocytes consist of five different types of WBCs, each with its own specific functions:

- Basophils, which are the least common type of leukocyte in the body, fight infections and parasitic infections. By releasing histamine during allergic reactions, basophils enable the body

to produce an antibody called immunoglobulin E. Additionally, by secreting heparin, they increase the fluidity of the blood [7].

- Eosinophils produce enzymes that destroy parasites that cause inflammatory and allergic reactions in the body [8].
- Monocytes are produced in the bone marrow and then enter the bloodstream. These cells are called monocytes when in the bloodstream, but within a few hours, they leave the circulatory system and enter the tissues. The monocyte cells that reach the tissue are called macrophages. They eliminate microorganisms that cause infections and clean up dead cells [9].
- Lymphocyte cells, which are produced in the bone marrow and lymph tissue, secrete chemicals called lymphokines against foreign organisms in the body, stimulating other immune system cells and allowing them to attack the foreign organism [10].
- Neutrophils are the first precursor cells to reach foreign organisms that cause infections in the body. They release and digest chemical enzymes to combat foreign organisms [11].

Leukemia, anemia, cancer, and various other diseases can be diagnosed through the analysis of WBCs [12]. This analysis is often conducted using a peripheral blood smear, which is a common laboratory method. To obtain a sample, a healthcare provider draws blood from a patient's finger or toe using a sterile needle, and the sample is then examined in a laboratory to create a peripheral blood film [13]. This film is manually analyzed by a specialist to identify signs of disease. However, manual analysis can be time-consuming and laborious for experts. As a result, computer-aided systems have been developed to assist with the classification of WBCs. With the advancement of hardware technology, the use of artificial intelligence (AI) in this field has increased. AI-based systems, also known as decision support systems, are designed to minimize errors caused by human factors and are used in various sectors, including healthcare. For example, decision support systems have been successfully used to detect COVID-19 through chest computed tomography images and to detect brain tumors through brain magnetic resonance imaging without human intervention [14].

Many studies have been carried out for the automatic classification of WBCs by AI-based systems. In the study [15], researchers proposed a system that uses the DenseNet-121 model to classify different types of WBCs. A publicly available dataset including eosinophil, lymphocyte, monocyte, and neutrophil classes was used for model training. The dataset contains 12,444 different samples with a resolution of 320×240px. The normalization process was applied to the dataset samples to speed up model training. The number of dataset samples has been increased with data augmentation techniques such as flipping, rotation, brightness, and zooming. The dataset samples were resized to a resolution of 224×224px. After the pre-processing steps, 20,050 WBCs images were obtained, including synthetic images. The model is trained for 10 epochs with the help of the Adam optimizer. Four different training processes were performed and the batch size value was changed to 8, 16, 32, and 64 in each training. The model, which was trained with 8 batch sizes, achieved 98.84% accuracy, 99.33% precision, 98.85% sensitivity, and 99.61% specificity values during the test phase, and achieved more successful results compared to other models.

In the study [16], researchers proposed an approach that can classify WBCs from microscopic blood images. The researchers used a publicly available dataset of images with different values in resolutions ranging from 350×236px to 2592×1944px. AlexNet, ResNet-101, and GoogleNet models were trained to detect five different classes: basophil, eosinophil, lymphocyte, monocyte, and neutrophil. While the dataset samples are resized to 227×227px resolution for training the AlexNet model, this value is 224×224px for the training of the GoogleNet and ResNet-101 models. To compare the success of the pre-trained models in classifying WBCs, 178 test images were given to the relevant models as input. The AlexNet model achieved better results compared to other models with 96.63% accuracy, 97.85% specificity, and 89.18% sensitivity rates.

In one study [17], researchers designed a deep convolutional neural network (CNN) model to classify microscopic images of WBCs. They proposed a new data augmentation method based on feature concentration to enhance the dataset and address the small number of samples. The training, validation,

and testing processes for the CNN model, which was designed to automatically classify the neutrophil, lymphocyte, monocyte, eosinophil, and basophil classes, were carried out using a special dataset provided by Sichuan Meisheng Biotech Company. This dataset consists of 8600 leukocyte images with a resolution of 1024×768px collected from various individuals. These images were divided into 217×217px pieces, resulting in a total of 11,658 sub-images. 80% of the dataset samples were reserved for training, while the remaining 20% were used for validation. The proposed model achieved an average test accuracy of 97.6% in classifying the five different WBCs.

In another study [18], researchers proposed an approach for classifying WBCs in microscopic images. Samples from a publicly available dataset containing a total of 352 images were augmented using various image augmentation techniques, resulting in 12,444 images. The dataset included samples belonging to the eosinophil, lymphocyte, monocyte, and neutrophil classes. A seven-layer convolutional neural network with an input size of 120×160px was created to automatically classify these samples. To this end, all of the dataset samples were resized to 120×160px. The proposed model was subjected to two different training processes to examine its binary and multiclass classification performance. In binary classification, a mononuclear class was created using eosinophil and neutrophil samples, and a polynuclear class was created using lymphocyte and monocyte samples. The model achieved an accuracy of 96.30% in binary classification and 87.93% accuracy in multiclass classification.

In another study [19], the researchers proposed a system that can simultaneously detect and classify WBCs in an image. This system is based on the F-RCNN and YOLOv4 architectures. The models were trained on samples from the Blood Cell Count Dataset (BCCD), which includes samples of four different WBCs: neutrophils, eosinophils, monocytes, and lymphocytes. The F-RCNN model achieved an accuracy of 96.25% and the YOLOv4 model achieved 95.75% accuracy during the testing phase.

In yet another study [20], the researchers proposed a U-Net-based approach for WBCs segmentation. In the U-Net encoder network, ResNet-50 blocks were integrated instead of the default layers, and squeeze-and-excitation blocks were added to the decoder network. The training and testing stages of the model were conducted using samples from the BCISC and LISC datasets. Using various data augmentation techniques, the number of samples for each dataset was increased to 10,000. The dataset samples were divided into 80% for training, 10% for validation, and 10% for testing. The ResNet-50-based U-Net model was trained for 200 epochs with a batch size of 8 and Adam optimization. It was reported that the model achieved a Dice score of 98.13% and a mean Intersection over Union (mIoU) rate of 96.36% during the testing phase using the BISC dataset samples.

The primary objective of this study is to use deep learning to automatically detect WBCs from microscopic blood images, thereby assisting specialists in the early diagnosis of diseases related to WBCs counts. The main contributions of this study are as follows:

- Demonstrating the effectiveness of existing deep learning models on a new dataset.
- Achieving high performance on a non-uniformly distributed dataset without using data augmentation for WBCs classification.
- Visualizing, using Gradient-weighted Class Activation Mapping (Grad-CAM), which pixel areas the deep learning models focus on during the decision-making phase, thereby providing an explainable structure for pre-trained models.
- Reducing human errors and subjectivity by using deep learning structures to perform these tasks, which are currently carried out by experts visually.

The remainder of this paper is organized as follows: Section 2 presents the proposed method for this study, including the dataset used, pre-trained deep learning models, classification performance measures, and the Grad-CAM algorithm. Section 3 presents the parameters and environments used in the training phase, the numerical values of the model during the training phase, the test phase predictions, and performance values. The discussion and conclusion sections of the study are presented in Section 4 and Section 5, respectively.

2. Material and Methods

An approach has been proposed for the deep learning-based automated classification of WBCs from microscopic blood images. The block representation of the proposed method is given in Figure 1.

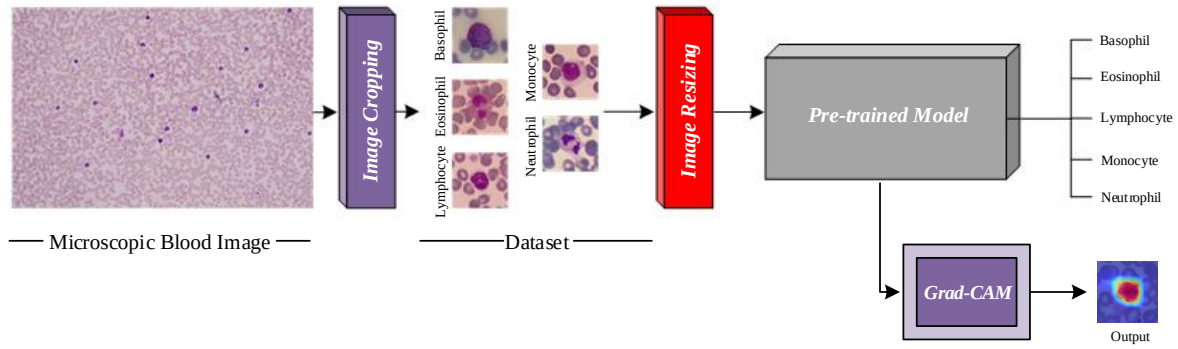


Figure 1 A Block Representation of The Proposed Method

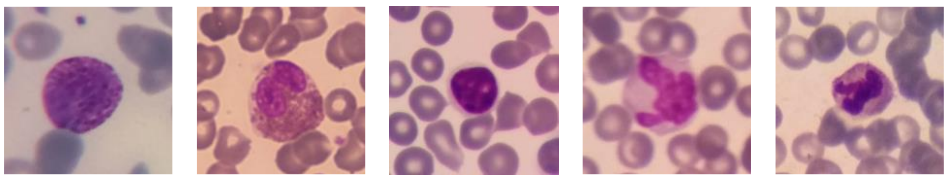
In the proposed method, the image given as input to the deep learning model is classified as a basophil, eosinophil, lymphocyte, monocyte, or neutrophil at the model's output. The choice of dataset and model is critical for achieving high success rates in this classification process. The quality of the dataset directly affects the performance of the deep learning model, and therefore it is important that it is created or verified by experts. This can be a resource and time-intensive process. However, several researchers have created and publicly shared WBCs datasets, as listed in Table 1.

Table 1 Publicly Available WBCs Datasets

Dataset	Basophil	Eosinophil	Lymphocyte	Monocyte	Neutrophil	Total
LISC [21]	54	42	59	55	56	266
BCCD [22]	3	86	33	19	208	349
MISP [23]	0	42	36	33	38	149
ALL-IDB [24]	1	2	60	3	18	84
Zheng et al. [25]	1	22	53	48	176	300
Raabin-WBC [26]	301	1066	3609	795	10,862	16,633

2.1 Dataset

In this study, the Raabin-WBC dataset [26] was used for the training, validation, and testing of the models. The Raabin-WBC dataset was created using 72 peripheral blood films collected from Shariati Hospital, which were examined using Olympus Cx18 and Zeiss microscopes. A total of 16,633 WBCs images with a resolution of 575×575px were obtained, and these images were labeled by two experts: 301 were labeled as basophils (Bas), 1066 as eosinophils (Eos), 3609 as lymphocytes (Lym), 795 as monocytes (Mon), and 10,862 as neutrophils (Neu). Samples of each class in the dataset are shown in Figure 2.



(a) Basophil *(b) Eosinophil* *(c) Lymphocyte* *(d) Monocyte* *(e) Neutrophil*

Figure 2 Dataset Samples [26]

Upon examination of the class-based distribution of samples in the Raabin-WBC dataset, it was observed that the Neu class is dominant. Data augmentation methods, which involve creating synthetic images, can be used to balance the distribution of classes. However, in this study, no data augmentation was performed in order to test the performance of pre-trained models under challenging conditions.

2.2 Pre-trained Models

Transfer learning is a machine learning technique that involves using the weights of a previously trained model as initial weights in the training phase of CNN. This allows the model, which was previously trained on a task, to be reused for different tasks. Transfer learning is highly effective for achieving good performance with a small amount of data. It is now a widely used method, especially for tasks related to image or natural language processing, as it allows researchers to use pre-trained models that have already learned how to classify images and have learned general features such as edges and shapes. Examples of pre-trained models that are often used as the basis for transfer learning include ResNet [27], VGG [28], and MobileNet [29], which were trained using the ImageNet [30] database. Pre-trained models can be grouped into three categories based on the number of parameters they contain: low (less than 15 M), medium (between 15 M - 70 M), and high (more than 70 M). Information on the pre-trained models is provided in Table 2 [31].

Table 2 Pre-trained Models Used in This Study [31]

Model	Default Input Size	Parameters (Million)	Category
ConvNeXtXLarge	224×224	350.1	High
ConvNeXtLarge	224×224	197.7	High
VGG-19	224×224	143.7	High
VGG-16	224×224	138.4	High
EfficientNetV2L	480×480	119	High
NASNetLarge	331×331	88.9	High
ConvNeXtBase	224×224	88.5	High
EfficientNetB7	600×600	66.7	Medium
ResNet152	224×224	60.4	Medium
ResNet152V2	224×224	60.4	Medium
InceptionResNetV2	299×299	55.9	Medium
EfficientNetV2M	480×480	54.4	Medium
ConvNeXtSmall	224×224	50.2	Medium
ResNet101	224×224	44.7	Medium
ResNet101V2	224×224	44.7	Medium
EfficientNetB6	528×528	43.3	Medium
EfficientNetB5	456×456	30.6	Medium
ConvNeXtTiny	224×224	28.6	Medium
ResNet50	224×224	25.6	Medium
ResNet50V2	224×224	25.6	Medium
InceptionV3	299×299	23.9	Medium
Xception	299×299	22.9	Medium
EfficientNetV2S	384×384	21.6	Medium
DenseNet201	224×224	20.2	Medium
EfficientNetB4	380×380	19.5	Medium
EfficientNetV2B3	300×300	14.5	Low
DenseNet169	224×224	14.3	Low
EfficientNetB3	300×300	12.3	Low
EfficientNetV2B2	260×260	10.2	Low
EfficientNetB2	260×260	9.2	Low
EfficientNetV2B1	240×240	8.2	Low
DenseNet121	224×224	8.1	Low
EfficientNetB1	240×240	7.9	Low
EfficientNetV2B0	224×224	7.2	Low
Mobilenet_v3_large	224×224	5.4	Low
NASNetMobile	224×224	5.3	Low
EfficientNetB0	224×224	5.3	Low
MobileNet	224×224	4.3	Low
MobileNetV2	224×224	3.5	Low
Mobilenet_v3_small	224×224	2.9	Low

To directly assess the effect of the number of parameters on model performance, three pre-trained models were randomly selected from the categories specifically created for this study: ResNet-50, VGG-19, and MobileNet-V3-Small.

2.3 Performance Metrics

Various metrics can be calculated using True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) to evaluate the performance of models. In this study, four metrics were used to evaluate the models for each class. These metrics and their corresponding equations are as follows:

- Accuracy is a performance metric that measures the percentage of correct predictions made by a classification model. It is the most widely used performance metric, but it may not fully reflect the performance of a model and can sometimes be misleading. For example, in a dataset where some classes are more represented than others, accuracy may not be a sufficient metric.
- Precision measures the percentage of predictions made by a model that are correct. The main difference between precision and accuracy is that precision only considers correct predictions, while accuracy considers all predictions. Therefore, precision is often a more precise metric and is given greater consideration when evaluating the performance of classification models.
- Sensitivity is a performance metric that measures the success of a classification model. It shows the percentage of data that the model predicted correctly. The main difference with other metrics is that sensitivity only evaluates correct predictions. For example, a model may have low sensitivity even though it has high accuracy. In this case, most of the data that the model predicts correctly are misclassified data, indicating that the model is not performing well.
- The F-1 score is a combination of sensitivity and precision ratios, used to evaluate the performance of a classification model, especially for multi-label data. The advantage of the F-1 score is that it does not rely solely on accuracy values, allowing it to show whether the model has balanced performance for all classes.

$$Precision\ (P) = \frac{TP}{TP + FP} \tag{1}$$

$$Sensitivity\ (S) = \frac{TP}{TP + FN} \tag{2}$$

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

$$F1\ Score\ (F1) = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{4}$$

2.4 Grad-CAM Algorithm

Grad-CAM is a technique that visualizes the regions of an image that are most important for a CNN to make a prediction. It allows us to understand which parts of an image a CNN is using to make a decision, and can be used to generate heatmaps that highlight these regions [32]. Grad-CAM works by using the gradients of the output of the CNN with respect to the input image to produce a weighted sum of the feature maps in the final convolutional of the network. The resulting heatmap is then overlaid on the input image to show which regions had the greatest influence on CNN's prediction. The architecture of the Grad-CAM algorithm is depicted in Figure 3.

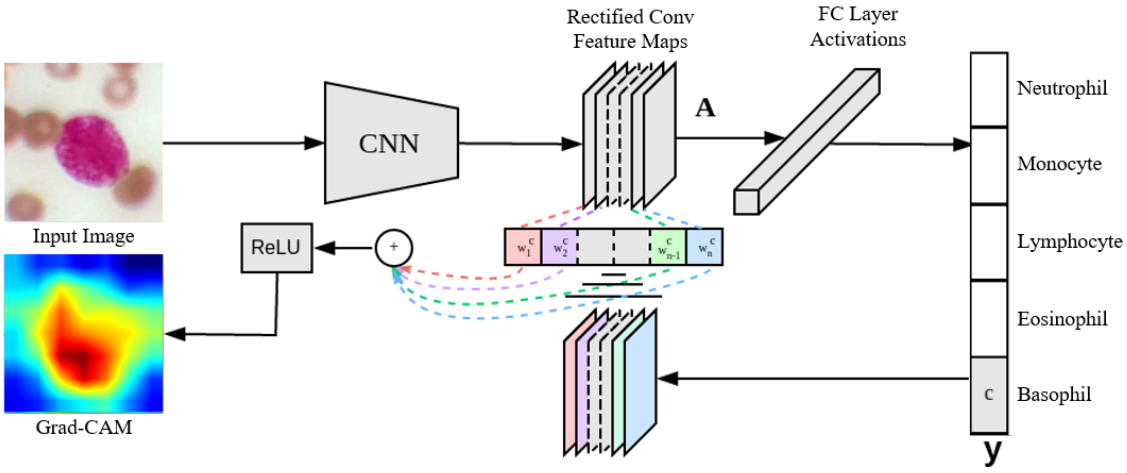


Figure 3 The architecture of the Grad-CAM [32]

The process for creating a Grad-CAM visualization for a pre-trained CNN is as follows:

1. Feed the input image through the CNN to generate a prediction.
2. Compute the gradients of the output of the CNN with respect to the feature maps in the final convolutional layer.
3. Take the weighted sum of the feature maps, using the gradients as weights.
4. Resize the resulting heatmap to the size of the input image.
5. Overlay the heatmap on the input image to highlight the regions that were most important for CNN's prediction.

Grad-CAM is relatively simple to implement and can be used with any CNN, regardless of its architecture. It is also an efficient method, as it only requires a single forward and backward pass through the network to generate the visualization. However, there are some limitations to Grad-CAM. For example, it can only provide visualizations for a single class at a time, and it is sensitive to the specific layer chosen for visualization. Additionally, the visualizations produced by Grad-CAM may not always align perfectly with human intuition, as they are based on the internal representation of the CNN rather than the visual features that a human might use to classify the image.

3. Experimental Results

The results of models trained to classify WBCs from microscopic blood images are presented in this section. In addition, an analysis of the experimental findings with performance metrics is shown in the following sections.

3.1 Experimental Setups

The default input size of the ResNet-50, VGG-19, and MobileNet-V3-Small models used in this study is 224×224px, so all of the dataset samples were resized to this value. Before training the model, 70% of the resized dataset samples were randomly divided for use in the training, 20% for validation, and 10% for testing. The visual representation of these processes is shown in Figure 4.

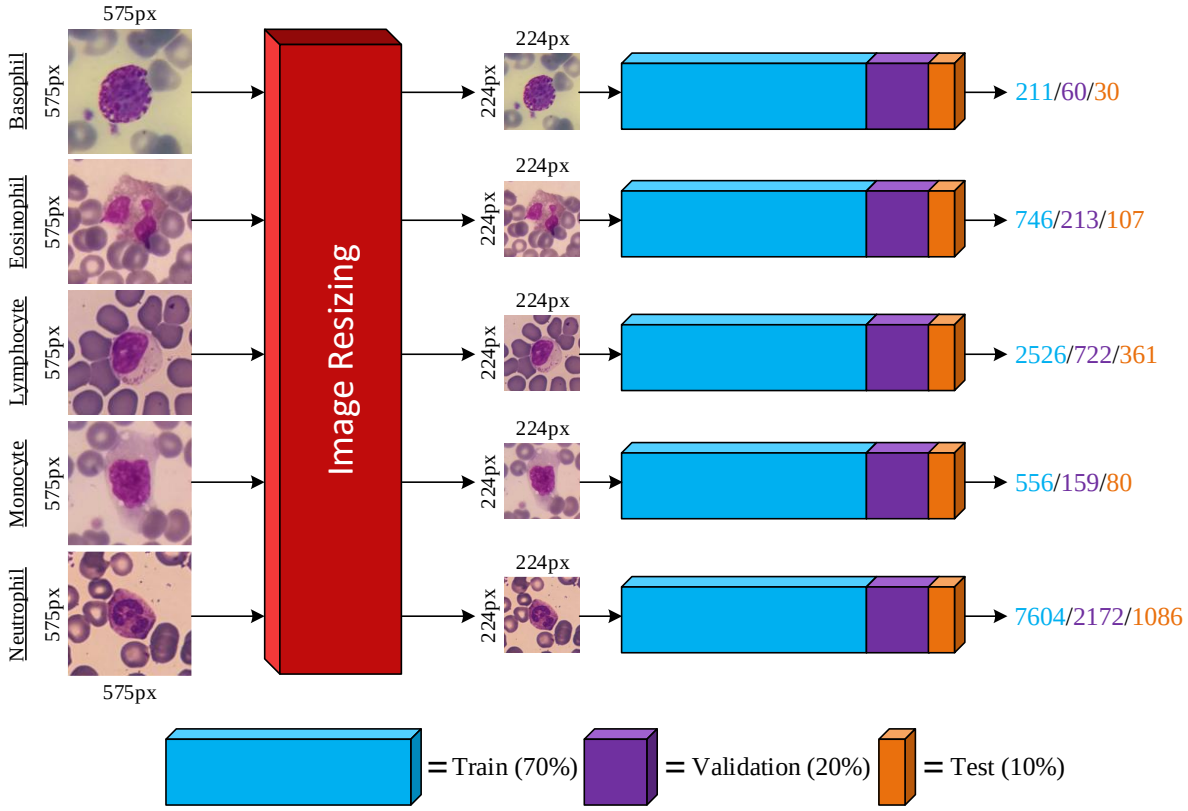
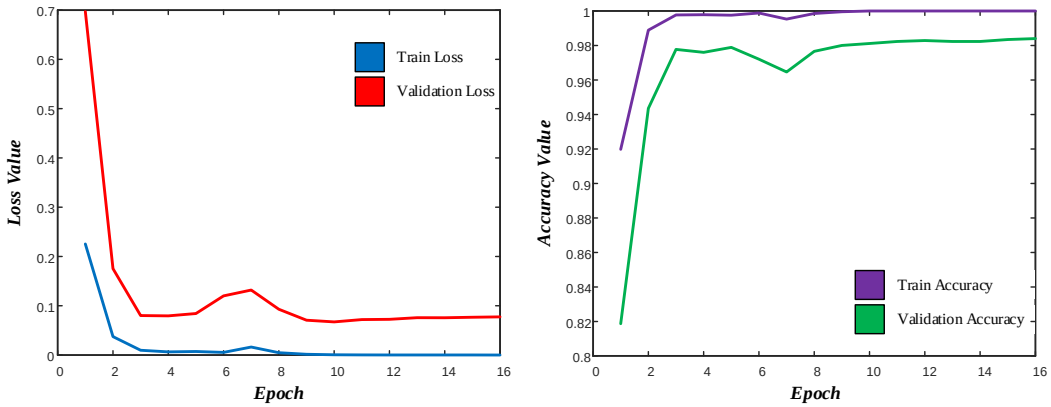


Figure 4 Image Resizing and Splitting Method

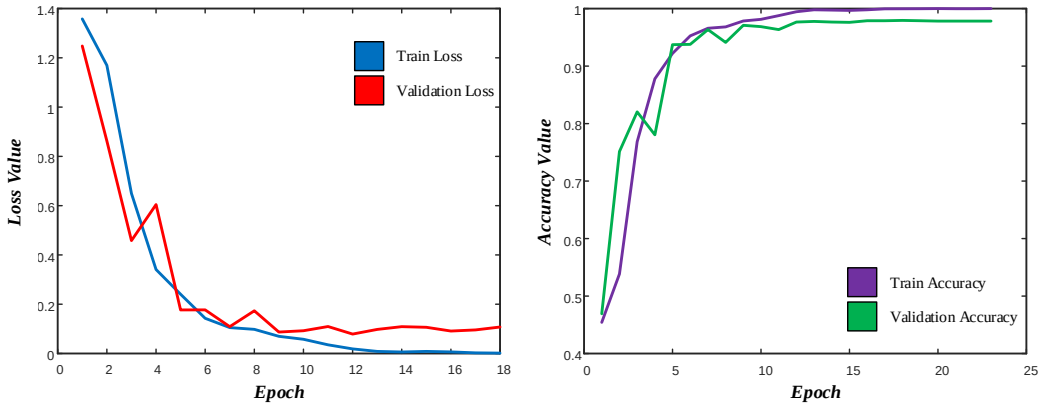
The pre-trained ResNet-50, VGG-19, and MobileNet-V3-Small models were included in the training using the Keras library. Since the models will only make predictions for five different classes, the Dense layers were revised and the softmax activation function was used. The models were compiled with an Adam optimizer and a learning rate of 0.0001. ImageNet weights were used instead of random initial weights for the training of the models. The models were trained with a constant batch size of 64, and training was carried out for a maximum of 50 epochs using the early stopping function. If the monitored validation accuracy value does not improve for five consecutive epochs, the early stopping function terminates the training phase and the weights of the epoch with the highest validation accuracy value are recorded in the '.h5' format. All of these processes were performed in the Google Colab environment.

3.2 Results

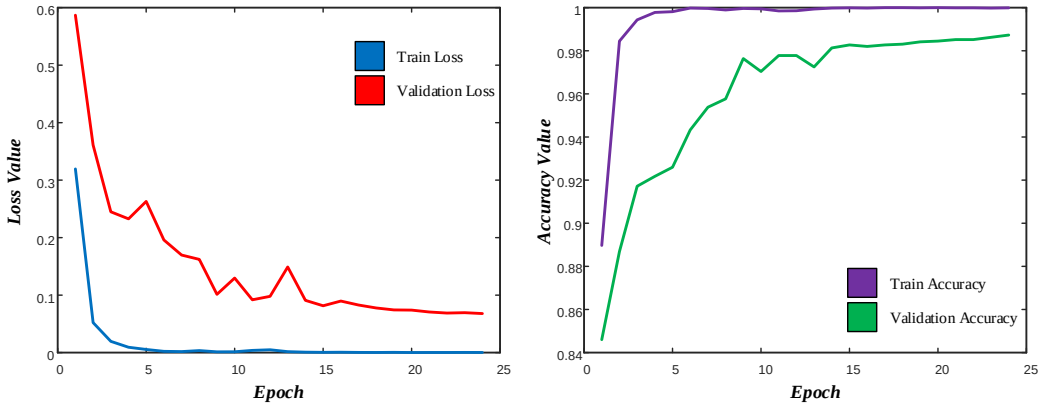
Three different deep-learning models were trained with the same parameters. The time required to complete the model training processes is directly proportional to the number of parameters and layers they have. The training stages of the models were carried out using the early stopping function, and the weights of the epoch that achieved the highest validation accuracy were recorded. ResNet-50 reached the highest validation accuracy after 16 epochs, VGG-19 reached the highest validation accuracy after 23 epochs, and MobileNet-V3-Small reached the highest validation accuracy after 24 epochs. The loss and accuracy graphs for the models during the training and validation phases are shown in Figure 5.



(a) ResNet-50



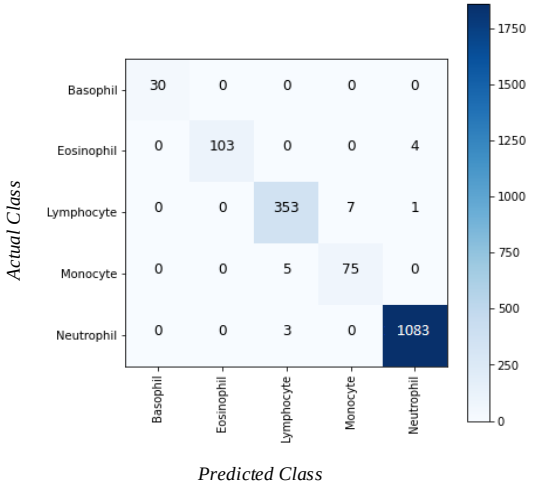
(b) VGG-19



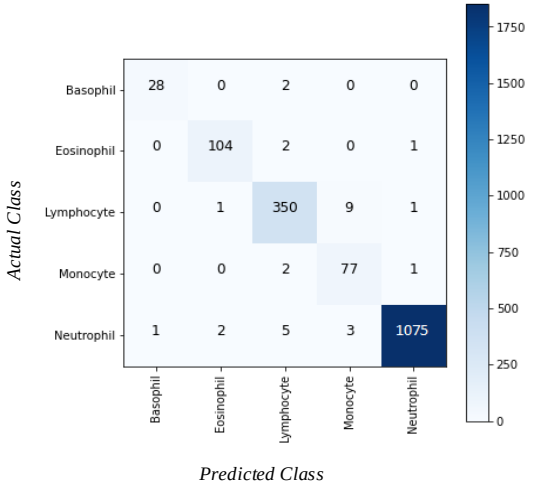
(c) MobileNet-V3-Small

Figure 5 Loss and Accuracy Graphs

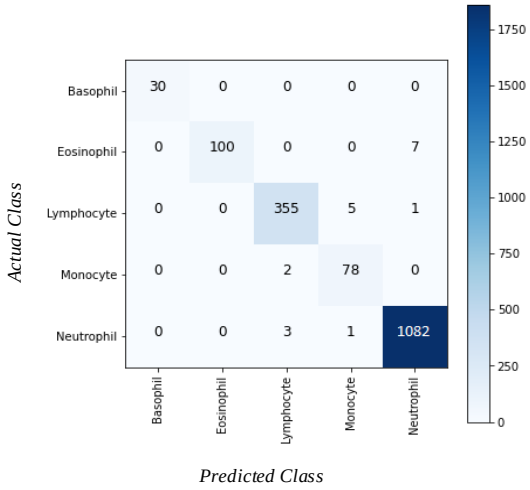
When the validation accuracy values were examined for the three models that completed training, it was observed that a rate of more than 95% could be achieved in less than 25 epochs. This is due in part to the fact that the models were trained using ImageNet weights instead of starting with random weights. Even though the models were trained with a dataset that is not evenly balanced, the lack of overfitting indicates the success of the pre-trained models. Performance metrics were used to compare the classification performance of the three different models trained to classify five different WBCs from microscopic blood images. For this, images that were not included in the training and validation phases but were reserved solely for use in the testing phase were given as input to each model. The confusion matrices generated by the predictions of the models for these inputs are shown in Figure 6.



(a) ResNet-50



(b) VGG-19



(c) MobileNet-V3-Small

Figure 6 Confusion Matrices for Each Model

To evaluate the Grad-CAM outputs of the deep learning model, it is necessary to first assess the performance of the model during the training phase. This helps to understand the accuracy of the model's predictions and assess the reliability of the model. When the predictions are analyzed, it is apparent that the models have learned to classify WBCs. In Table 3, the performance metric values achieved by the relevant models during the testing phase are provided.

Table 3 The Results of The Pre-trained Models

Model	**Basophil**			**Eosinophil**			**Lymphocyte**			**Monocyte**			**Neutrophil**		
	P (%)	**S (%)**	**F1 (%)**	**P (%)**	**S (%)**	**F1 (%)**	**P (%)**	**S (%)**	**F1 (%)**	**P (%)**	**S (%)**	**F1 (%)**	**P (%)**	**S (%)**	**F1 (%)**
ResNet-50	100	100	100	100	96.26	98.09	97.78	97.78	97.78	91.46	93.75	92.59	99.54	99.72	99.62
VGG-19	96.55	93.33	94.91	97.19	97.19	97.19	96.95	96.95	96.95	86.51	96.25	91.12	99.72	98.98	99.34
MobileNet-V3-Small	100	100	100	100	93.45	96.61	98.61	98.33	98.46	92.85	97.50	95.11	99.26	99.63	99.44

Following the training phase, the model should be evaluated using the test data. The resulting outputs should be carefully analyzed to interpret how the Grad-CAM outputs describe the images and identify the features that the model considers important. Figure 7 presents the Grad-CAM outputs for a selection of randomly chosen images from the test set.

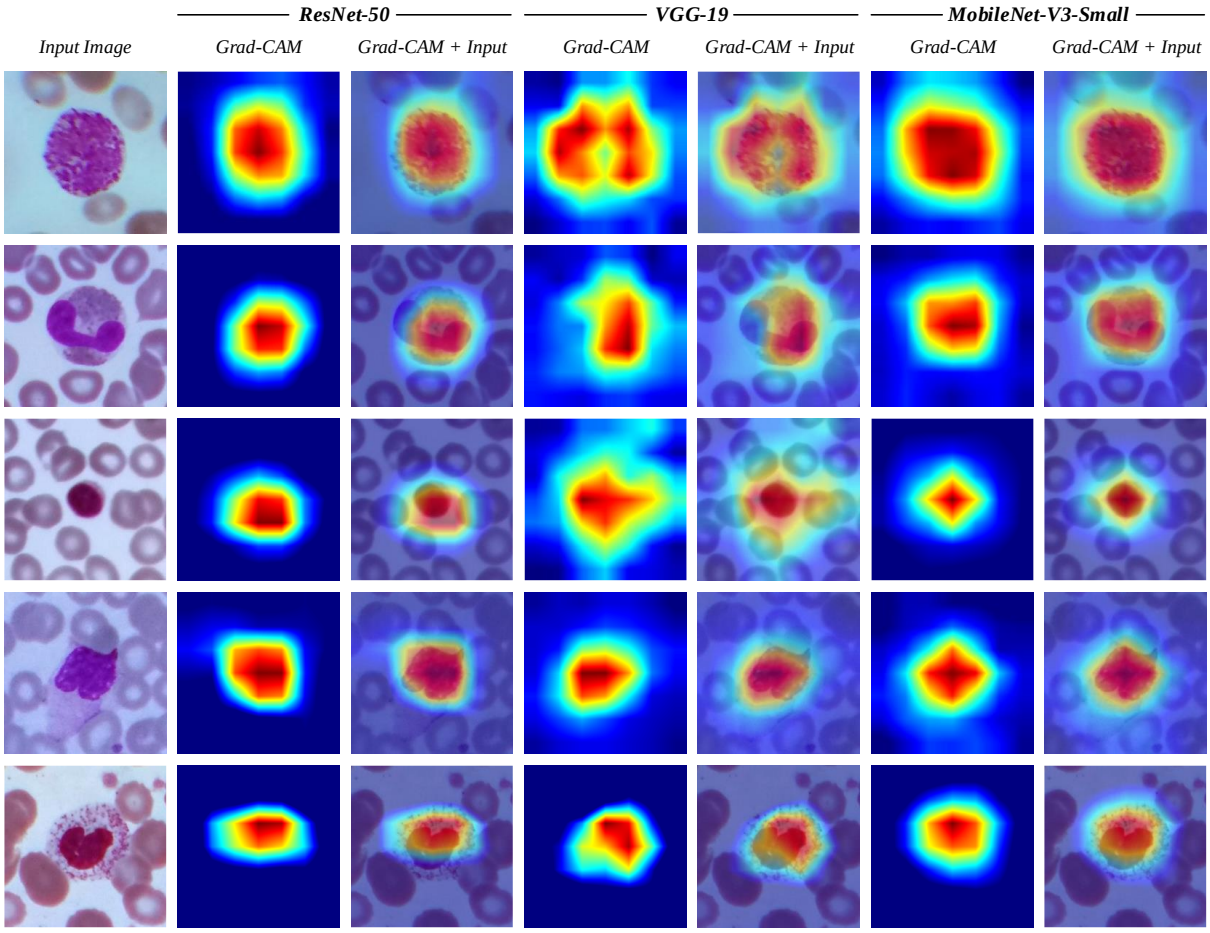


Figure 7 The Grad-CAM outputs

Upon examination of the Grad-CAM outputs, it was observed that all three models effectively identified the relevant features in the images with a high degree of accuracy. As the model performance is consistent with expectations, it is not necessary to further adjust the hyperparameters or layer configurations of the models.

4. Discussion

The detection of WBCs using microscopic blood images is a topic of active research. Table 4 presents a selection of studies on this subject that have been curated by hand. Yildirim and Cinar [33] employed AlexNet, ResNet-50, DenseNet-201, and GoogleNet architectures on a dataset comprising 9,663 images. For each model, three different training stages were conducted using original data, data filtered with a Gaussian filter, and data filtered with a median filter. The highest accuracy rate of 83.44% was achieved by the DenseNet-201 model trained with Gaussian-filtered data. Ekiz et al. [34] classified 12,442 WBCs images using both a CNN model and a Con-SVM model, with the Con-SVM model found to be more accurate, achieving an accuracy rate of 85.96%, compared to the CNN model's accuracy rate of 83.91%. Sharma et al. [15] implemented a deep learning model based on the DenseNet121 architecture for the classification of various types of WBCs. The model was optimized with normalization and data augmentation and achieved an accuracy of 98.84%. Girdhar et al. [35] proposed a method that demonstrated the ability to accurately classify WBCs types in a shorter number of epochs/time compared to other approaches. The performance of the proposed method was evaluated using the Kaggle dataset, resulting in an overall accuracy of 98.55%. Nahzat et al. [36] aimed to develop a CNN-based model for the classification of WBCs. They used images of WBCs from the Kaggle dataset to train and evaluate their proposed model, testing it with various optimizers to determine the best performance. They also compared the performance of their model with four pre-trained CNN models

(MobileNetV2, DenseNet121, InceptionV3, and ResNet50) and found that the proposed model, despite having the lowest number of trainable parameters and training time, outperformed the others with an accuracy of 99.5%. Karakuş and Özbay [37] used CNN models and combined them with three different machine learning classifiers. They applied contrast-limited adaptive histogram equalization (CLAHE) and Gaussian filters to images from the Kaggle dataset, which were then reclassified using the three CNN networks. The results showed that the classification performance was higher when the images were preprocessed with these filters compared to the original data. Jung et al. [38] proposed W-Net, a CNN-based method for the classification of WBCs. To evaluate W-Net, they used a large-scale dataset of 6,562 real images of the five WBCs types, obtained from The Catholic University of Korea. The results showed that W-Net achieved an average accuracy of 97%. Wang et al. [39] proposed a deep CNN called WBC-AMNet for automatically classifying WBCs subtypes based on a focused attention mechanism. This method uses feature fusion strategies, combining Squeeze-and-Excitation and Gather-Excite modules, to obtain more localized attention from the CNN. The WBC-AMNet achieved an overall accuracy of 98.39. They also used Grad-CAM to visualize the attention heatmaps of different feature maps. Roy and Ameer [40] applied a semantic segmentation technique using a deep learning network to accurately segment WBCs from microscopic blood images. The proposed model employed the DeepLabv3+ architecture with a ResNet-50 network as the feature extractor. The model was evaluated on three different public datasets containing five categories of WBCs, using 10-fold cross-validation to assess its effectiveness. The average segmentation accuracy achieved by the proposed model was 96.1% IoU. Wu et al. [20] proposed a WBC image segmentation network based on U-Net that combines residual networks. The encoder structure of the network uses ResNet50 residual blocks as the main unit. The proposed model achieved 96.36% mIoU.

Table 4 Comparison of Our Work With Some State-of-the-art Studies

Study	**Number of Class**	**Number of Images**	**Model**	**Explainability**	**Task**	**Performance**
Yildirim and Cinar [33]	4 (Eos, Lym, Mon, Neu)	9,663	DenseNet-201	Black-box	Classification	Acc=83.44%
Ekiz et al. [34]	4 (Eos, Lym, Mon, Neu)	12,442	Con-SVM	Black-box	Classification	Acc=85.96%
Sharma et al. [15]	4 (Eos, Lym, Mon, Neu)	12,444	DenseNet-121	Black-box	Classification	Acc=98.84%
Girdhar et al. [35]	4 (Eos, Lym, Mon, Neu)	12,444	CNN	Black-box	Classification	Acc=98.55%
Nahzat et al. [36]	4 (Eos, Lym, Mon, Neu)	12,444	Hybrid CNN	Black-box	Classification	Acc=99.50%
Karakuş and Özbay [37]	4 (Eos, Lym, Mon, Neu)	12,444	CNN	Black-box	Classification	Acc=97.10%
Jung et al. [38]	5 (Bas, Eos, Lym, Mon, Neu)	6,562	W-Net	Black-box	Classification	Acc=97.00%
Wang et al. [39]	4 (Eos, Lym, Mon, Neu)	16,873	WBC-AMNet	Grad-CAM	Classification	Acc=98.39%
Roy and Ameer [40]	5 (Bas, Eos, Lym, Mon, Neu)	642	DeepLabv3+	Black-box	Segmentation	mIoU=96.10%
Wu et al. [20]	5 (Bas, Eos, Lym, Mon, Neu)	516	U-Net	Black-box	Segmentation	mIoU=96.36%
The proposed our study	5 (Bas, Eos, Lym, Mon, Neu)	16,633	MobileNet-V3-Small	Grad-CAM	Classification	Acc=98.86%

In this study, we employed a pre-trained MobileNet-V3-Small model for automated WBCs classification. Our results demonstrated a high accuracy of 98.86%, which is higher than the accuracy reported in most other studies. This suggests that the model in our study was able to accurately classify the images into the appropriate categories. Our study included a larger number of classes (5) compared to many other studies (which often have only 4 classes). This increased complexity made the task more challenging and required a more sophisticated model. Our dataset was also relatively large, with 16,663 images, which may have contributed to the robustness and generalizability of our model. We also employed Grad-CAM as an explainability method to provide insights into the model's decision-making process and identify any potential biases or weaknesses.

It is worth noting that some other studies have focused on image segmentation, a task distinct from classification. Image segmentation involves predicting a pixel-level mask for each class in the image, while classification simply involves predicting a single class label for the entire image. In this study, we employed the MobileNet-V3-Small model architecture, which may not be optimal for all tasks and datasets. Alternative model architectures may yield better performance in certain cases. Some other studies have utilized models with more layers and a greater number of parameters (e.g. DenseNet-201, DenseNet-121), which may improve performance but also require more computational resources and may be more prone to overfitting.

The limitations of this study are as follows:

- The dataset consists of only 16,633 images, which may not be sufficient to fully capture the variability and complexity of the WBCs being analyzed.
- Our study only evaluated the performance of three pre-trained models (ResNet-50, VGG-19, and MobileNet-V3-Small) on the WBCs classification task.
- The durability of models against changes due to variations in lighting, background, or other factors that may affect the appearance of WBCs in images has not been validated.
- As k-fold cross-validation was not employed, the model was only evaluated on a single split of the data.

In future research, it would be beneficial to augment the dataset with a larger number of images that have a more balanced distribution of classes. This would likely lead to more robust and accurate classifications. It would also be useful to evaluate the model on a range of different datasets to assess its generalizability and performance on diverse types of images. While the models in this study demonstrated high accuracy rates, there is always a potential for further improvement. Additional research could be conducted to optimize the models and enhance their performance. While the models in this study demonstrated high accuracy in classifying WBCs, it would be valuable to assess their performance in real-world settings. This might involve testing the models on images from actual medical cases or incorporating the models into existing medical imaging systems for use by healthcare professionals.

5. Conclusion

In recent years, advances in hardware technology have enabled the use of machine learning techniques in the field of healthcare, specifically in the automatic classification of WBCs using microscopic blood images. Accurate identification of WBCs is crucial for medical diagnosis and research. This study proposes a deep learning-based approach for the automatic classification of WBCs using microscopic blood images and investigates its effectiveness through experiments on a dataset of 16,633 different WBCs images. Several popular pre-trained models, including MobileNet-V3-Small, were employed for the deep learning models. The MobileNet-V3-Small model achieved the highest accuracy rate of 98.86%. To understand how the model was making its predictions, we employed a visualization technique called Grad-CAM to identify the pixel areas that the model was focusing on. The findings of this study suggest that deep learning may be a useful tool for the automated identification of WBCs in medical diagnosis and research. However, further research is needed to fully evaluate the robustness

and generalizability of these results, as well as to explore the potential for using deep learning in other aspects of medical diagnosis and treatment.

References

[1] C. J. Walsh and C. A. Luer, "Elasmobranch hematology: identification of cell types and practical applications," *The Elasmobranch Husbandry Manual: Captive Care of Sharks, Rays and their Relatives*, pp. 307-323, 2004.

[2] A. Glenn and C. E. Armstrong, "Physiology of red and white blood cells," *Anaesthesia & Intensive Care Medicine*, vol. 20, no. 3, pp. 180-174, 2019.

[3] R. Van Zwieten, A. J. Verhoeven and D. Roos, "Inborn defects in the antioxidant systems of human red blood cells," *Free Radical Biology and Medicine*, vol. 67, pp. 377-386, 2014.

[4] I. Andia and N. Maffulli, "Platelet-rich plasma for managing pain and inflammation in osteoarthritis," *Nature Reviews Rheumatology*, vol. 9, no. 12, pp. 721-730, 2013.

[5] B. Olas and B. Wachowicz, "Role of reactive nitrogen species in blood platelet functions," *Platelets*, vol. 18, no. 8, pp. 555-565, 2007.

[6] M. Habibzadeh, M. Jannesari, Z. Rezaei, H. Baharvand and M. Totonchi, "Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception," *Tenth international conference on machine vision*, vol. 10696, pp. 274-281, 2018.

[7] A. L. Gillen and J. Conrad, "Our Impressive Immune System: More Than a Defense", *Faculty Publications and Presentations*, 135, 2014.

[8] H. Kita and B. S. Bochner, "Biology of eosinophils", *Middleton's allergy principles and practice*, vol. 8 pp. 265-279, 2013.

[9] F. Ginhoux and S. Jung, "Monocytes and macrophages: developmental pathways and tissue homeostasis", *Nature Reviews Immunology*, vol. 14, no. 6, pp. 392-404, 2014.

[10] Y. Ueda, M. Kondo and G. Kelsoe, "Inflammation and the reciprocal production of granulocytes and lymphocytes in bone marrow", *The Journal of experimental medicine*, vol. 201, no. 11, pp. 1771-1780, 2005.

[11] E. Bronze-da-Rocha and A. Santos-Silva, "Neutrophil elastase inhibitors and chronic kidney disease", *International journal of biological sciences*, vol. 14, no.10, pp. 1343-1360, 2018.

[12] A. Shahzad, M. Raza, J. H. Shah, M. Sharif and R. S. Nayak, "Categorizing white blood cells by utilizing deep features of proposed 4B-AdditionNet-based CNN network with ant colony optimization," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3143-3159, 2022.

[13] L. B. Maedel and K. Doig, "Examination of the peripheral blood film and correlation with the complete blood count," *Hematology: clinical principles and applications*, pp. 192-209, 2013.

[14] O. Katar and E. Duman, "Deep Learning Based Covid-19 Detection With A Novel CT Images Dataset: EFSCH-19," *Avrupa Bilim ve Teknoloji Dergisi*, vol. 29, pp. 150-155, 2021.

[15] S. Sharma, S. Gupta, D. Gupta, S. Juneja, P. Gupta, G. Dhiman and S. Kautish, "Deep learning model for the automatic classification of white blood cells," *Computational Intelligence and Neuroscience*, 2022.

[16] M. J. Macawile, V. V. Quiñones, A. Ballado, J. D. Cruz and M. V. Caya, "White blood cell classification and counting using convolutional neural network," *3rd International conference on control and robotics engineering (ICCRE)*, pp. 259-263, 2018.

[17] Y. Wang and Y. Cao, "Human peripheral blood leukocyte classification method based on convolutional neural network and data augmentation," *Medical physics*, vol. 47, no. 1, pp. 142-151, 2020.

[18] M. Sharma, A. Bhave and R. R. Janghel, "White blood cell classification using convolutional neural network," *Soft Computing and Signal Processing*, pp. 135-143, 2019.

[19] J. Yao et al., "High-efficiency classification of white blood cells based on object detection", *Journal of Healthcare Engineering*, 2021.

[20] J. Wu et al., "WBC Image Segmentation Based on Residual Networks and Attentional Mechanisms," *Computational Intelligence and Neuroscience*, 2022.

[21] S. H. Rezatofighi and H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," *Computerized Medical Imaging and Graphics*, vol. 35, no. 4, pp. 333-

343, 2011.

[22] M. Mohamed, B. Far and A. Guaily, "An efficient technique for white blood cells nuclei automatic segmentation," *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 220-225, 2012.

[23] O. Sarrafzadeh, H. Rabbani, A. Talebi and H. U. Banaem, "Selection of the best features for leukocytes classification in blood smear microscopic images," *Medical Imaging 2014: Digital Pathology*, vol. 9041, pp. 159-166, 2014.

[24] R. D. Labati, V. Piuri and F. Scotti, "All-IDB: The acute lymphoblastic leukemia image database for image processing," *18th IEEE international conference on image processing*, pp. 2045-2048, 2011.

[25] X. Zheng, Y. Wang, G. Wang and J. Liu, "Fast and robust segmentation of white blood cell images by self-supervised learning," *Micron*, vol. 107, pp. 55-71, 2018.

[26] Z. M. Kouzehkanan et al., "A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm," *Scientific reports*, vol. 12, no. 1, pp. 1-14, 2022.

[27] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409*, 1556, 2014.

[29] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications", *arXiv preprint arXiv:1704.04861*, 2017.

[30] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," *IEEE conference on computer vision and pattern recognition*, pp. 248-255, 2009.

[31] [Online]. Available: https://keras.io/api/applications/ [Accessed in December 2022].

[32] R. R. Selvaraju et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization," *IEEE international conference on computer vision*, pp. 618-626, 2017.

[33] M. Yildirim and A. Çinar, "Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease," *Rev. d'Intelligence Artif.*, vol. 33, no. 5, pp. 335-340, 2019.

[34] A. Ekiz, K. Kaplan and H. M. Ertunç, "Classification of white blood cells using CNN and Con-SVM," *29th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, 2021.

[35] A. Girdhar, H. Kapur and V. Kumar, "Classification of White blood cell using Convolution Neural Network," *Biomedical Signal Processing and Control*, vol. 71, 2022.

[36] S. Nahzat, F. Bozkurt and M. Yağanoğlu, "White Blood Cell Classification Using Convolutional Neural Network. Journal Of Science," *Technology And Engineering Research*, vol. 3, no. 1, pp. 32-41, 2022.

[37] M. Ö. Karakuş and E. Özbay, "Lökosit Tespiti İçin Beyaz Kan Hücrelerinin Esa Kullanılarak Sınıflandırılması," *Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 9, no. 17, pp. 333-344, 2022.

[38] 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 preprint arXiv:1910.01091*, 2019.

[39] Z. Wang, J. Xiao, J. Li, H. Li and L. Wang, "WBC-AMNet: Automatic classification of WBC images using deep feature fusion network based on focalized attention mechanism," *Plos one*, vol. 17, no. 1, 2022.

[40] R. M. Roy and P. M. Ameer, "Segmentation of leukocyte by semantic segmentation model: A deep learning approach," *Biomedical Signal Processing and Control*, vol. 65, 102385, 2021.