Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 30, Num. 02 (2023) 89-100

RESEARCH ARTICLE

Machine Learning Algorithms for Peripheral Blood Cell Classification - A Hemovision Project Experience

Algoritmos de Aprendizado de Máquina para Classificação de Células do Sangue Periférico - Uma Experiência do Projeto Hemovision

Mariana Dourado X. S. Santos[1]*, William Laus Bertemes[2], Iaan Mesquita de Souza[1], Mateus Henrique B. Andrades[1], Vinicius Sebba Patto[1]

Abstract: This research explores the use of machine learning algorithms to classify nucleated peripheral blood cells. The ResNet18 convolutional neural network was used to pre-process the images and replace the dense layers; and for the output, the Support Vector Machine (SVM) classifier was chosen. Images from different datasets were used for training and testing the model. Thus, the developed model achieved an accuracy and F1-Score of 99.96%. In face of the obtained results, it was found that machine learning algorithms can be satisfactorily integrated into educational and diagnostic support processes.
Keywords: Support Vector Machine — Convolutional Neural Networks — White Blood Cells — Machine Learning

Resumo: Este trabalho trata do uso de algoritmos de aprendizado de máquina para classificação de células nucleadas do sangue periférico. Foi utilizada a rede neural convolucional ResNet18 para o pré-processamento das imagens e em substituição às camadas densas; e para a saída foi escolhido o classificador Support Vector Machine (SVM). Foram usadas imagens de diferentes datasets para o treino e teste do modelo. Assim, o modelo desenvolvido alcançou uma acurácia e F1-Score de 99.96%. Diante dos resultados encontrados, constatou-se que os algoritmos de aprendizado de máquina podem ser integrados de forma satisfatória aos processos educacionais e de apoio ao diagnóstico.
Palavras-Chave: Máquina de Vetor de Suporte — Redes Neurais Convolucionais — Glóbulos Brancos — Aprendizado de Máquina

[1] *Instituto de Informática, Universidade Federal de Goiás (UFG), Goiânia - Goiás, Brazil*
[2] *Hospital das Clínicas, Universidade Federal de Goiás (UFG), Goiânia - Goiás, Brazil*
***Corresponding author**: marianadximenes@gmail.com
DOI: http://dx.doi.org/10.22456/2175-2745.130669 • **Received:** 06/03/2023 • **Accepted:** 22/06/2023

1. Introduction

The complete blood count (CBC) is the most requested complementary exam in the world; it is an integral part of health screening and an almost indispensable adjunct in the diagnosis and follow-up of chronic diseases in general, infectious diseases, medical emergencies, and the follow-up of chemotherapy and radiotherapy, having a broad relationship with the other clinical pathology exams [1].

With the development of sophisticated automated blood cell analyzers to perform the CBC, the number of samples requiring blood distension for morphological evaluation has greatly decreased and in many clinical settings is around 10% to 15% of routine samples. Nevertheless, blood distention remains a crucial diagnostic aid [2].

In order to obtain the most clinically relevant information from the morphological analysis of a blood distention, the procedure should be performed by a person very experienced in hematological microscopy, be it a clinical pathologist or a medical hematologist. Given that, knowing the unusual changes in the peripheral blood and its morphology allows the analyst to make accurate diagnoses and suggest, in a targeted and cost-effective manner, the performance of further complementary tests [2] [3].

However, often this experience is developed over several years and after experiencing different clinical cases to be able to perform the correct classification of blood cells and their morphological changes. Still, optical microscopy is the gold standard for micromorphological analysis of blood distensions, although it is quite laborious and sometimes subjective, so several objective methods have been proposed to assist in this process and help workflows in the hematology lab, such as the use of Machine Learning (ML) for classification and

pattern recognition in cell images.

The greatest appeal of using these computer models for data interpretation is their ability to learn from examples. Among the most commonly used ML algorithms for laboratory imaging, Artificial Neural Networks (ANN) stand out. As such, ANNs have been widely used for modeling nonlinear, complex, and multidimensional [4] [5] data. It is also worth noting, that ANNs evolved into deep ANNs, which gave rise to the term Deep Learning (DL), a technology that has been widely applied in image processing [6] [7]. Their success is noted in the field of medical imaging [8] [9] [10] for classification [11] object detection [12] and segmentation [13] [14].

Due to this ability, ANNs (including deep RNAs) can be considered a good alternative for pattern recognition and image classification. Several types of research have been conducted to evaluate the effectiveness of ANNs in classifying the differential of peripheral blood leukocytes [15] In addition to ANNs, other ML algorithms such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF) can also be used for blood white cell classification, as in the study conducted by Wibawa [16].

It is important to notice that there are several blood cell databases available, suggesting a great possibility for studies. This scenario allows the use of several ML algorithms in applications to support, for example, diagnostic, predictive, and prescriptive analyses [17]. However, the use of ML algorithms for this purpose is still very distant from the daily life of most clinical pathologists and students in this area; thus, the execution of this work is justified and motivated.

As light microscopy is the gold standard for micromorphological analysis of blood distensions, it is of utmost importance in this work to obtain good hit rates in the classification of blood cells, as this classification is often used in medical diagnostics. Therefore, it is important to perform several tests with different parameters on different ML algorithms found in the state of the art.

Given this, the main objective of this work was to make a study of the most relevant articles on blood cell image classification and replicate them in this work, following the cell types considered in the Hemovision project. Consequently, it was possible to identify the most interesting ML algorithms to be executed in the back-end of the Hemovision application, according to the image resolutions and image banks used in the tests, so far.

This study was developed within the scope of the research project Decision Support Prototypes for Intelligent and Outstanding Findings (SPIN-OFF), whose activities are focused on understanding systems and creating solutions or prototypes for decision support. In this paper, the prototype Hemovision, a subproject of SPIN-OFF, is discussed, in which images of nucleated peripheral blood cells are classified by users and by ML algorithms. Hemovision is composed of a mobile application and ML algorithms on the back-end. One of the goals of Hemovision is to help health students, especially in hematology, to better understand certain blood cells and to promote interest and debate on the subject.

This article is structured as follows: in Section 2, the works related to image classification are presented, in addition to the methodology used in this work, involving the databases used for training and testing the model, in addition to the algorithms implemented and their respective configurations; in Section 3, the results of the experiments developed in this work are presented, in addition to presenting some important considerations about the performance of the experiments and their results. The references used in this article are given in the last section, References.

2. Methods

A review was made on how computer vision and ML techniques were being used to treat and classify images in general and medical images. Thus, works dealing with the classification of blood cell images were prioritized.

After that, this section deals with the databases used, the proposed model, and its performance on different databases.

2.1 Related works

Current work in cell image classification is based on four stages: preprocessing [6] [7], segmentation [18], feature extraction [19], and classification [20]. For each of them, a specific algorithm is used to get the best result. When it comes to preprocessing, the data augmentation technique [21] can be applied to increase the number of images and thus improve the accuracy of the ML model.

The article [22] checked the F1-Score of the proposed model, Two-module Weighted Optimized Deformable Convolutional Neural Networks (TWO-DCNN), against the methods: VGG-19, VGG-16, Inception-V3, ResNet-50, SVM, MLP, DT and RF. The best result achieved was 95.7% F1-Score by TWO-DCNN on low-resolution images. In this work, the WBC classes taken into account are four: neutrophils, eosinophils, lymphocytes, and monocytes. Furthermore, it is worth mentioning that the author developed the robustness and generalization of the model by inserting noise in the images and decreasing their resolution, through deformable convolutional layers (DC).

Yet, Wibawa [16] compared a DL method against three traditional ML algorithms: SVM, MLP, and KNN, for the classification of just two classes of white cells: neutrophils and lymphocytes. The best accuracy obtained was 99.5% for the DL method - which used a Convolutional Neural Network (CNN), compared to 85% for the SVM algorithm, 81% for the MLP, and 78.5% for the KNN.

A similar work dealing with a model insensitive to low-resolution images is proposed by Ma et al. [23] the Deep Convolutional Generative Adversarial Network combined with ResNet (DC-GAN-ResNet). The result achieved by this method was 91.68% accuracy, compared to 89.38% by RNN-CNN-ResNet50 with Long Short Term Memory (LSTM); and 87.62% by CNN-ResNet50. However, the classes of

white cells taken into account in the test are four: neutrophils, eosinophils, lymphocytes, and monocytes.

The data augmentation technique is addressed by [24] by flipping the images horizontally and vertically, and generating synthetic data from Generative Adversarial Networks (GAN). After this preprocessing step, the DenseNet-169 network initializes the weights previously trained on the CIFAR-100 [25] dataset and then is compared with DenseNet-121, ResNet-50, ResNet-18, VGG-19 and VGG-16 achieving, respectively, the accuracy of 98.8%, 98.3%, 97.4%, 95.4%, 95.9%, and 95.7%. In this work, the WBC classes taken into account are five: neutrophils, eosinophils, lymphocytes, monocytes, and basophils.

Another work that initializes the network weights with those trained on a similar database is proposed by Khan [26]. This method is known as Transfer Learning and is widely used for its saving training time and for yielding higher accuracy [27]. In addition, feature extraction, by RELIEF [28], is performed to reduce redundancy, classification error, and system resource requirements. The proposed MLANet method using Extreme-Learning Machine (ELM) and RELIEF achieved 99.12% accuracy in classifying four classes of leukocytes: neutrophils, eosinophils, lymphocytes, and monocytes.

Work that also uses the data augmentation technique is covered by Katar et al. [29] by performing flipping, rotation, brightness, and zooming to the images. The trained models ResNet-50, VGG-19, and MobileNet-V3-Small were pre-trained with ImageNet weights [30]. Respectively, the average F1-Score achieved by these models were: 97.61%, 95.9%, and 97.92%. In this work, the leukocyte classes taken into account are five: neutrophils, eosinophils, lymphocytes, monocytes, and basophils.

Wang [31] also uses data augmentation, but with the purpose of balancing the dataset. The mainly used data augmentation methods are translation, rotation, color transformation, and scale transformation. The models VGG-16, ResNet, MobileNet V1, Inception V3, and the proposed one achieved accuracy of 91.6%, 93.2%, 91.8%, 94.1%, and 97.6%, respectively. The classes used in this work are five: neutrophils, eosinophils, lymphocytes, monocytes, and basophils.

The eight classes addressed in this paper are also addressed by Long [32], in which a model is proposed and compared with ResNet-18, VGG-16, AlexNet, and InceptionV3. The main dataset used is [33], then others are formed from that by resizing technique and data augmentation. In general, the proposed model excelled over the other models. However, ResNet-18 obtained the second highest accuracy across all datasets. Thus, in this work, eight classes are addressed: neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (promyelocytes, myelocytes, and metamyelocytes), erythroblasts, and platelets.

A comparison between the AlexNet, ResNet18, and GoogleNet architectures all retrained with the tuning method and then given to classify in the softmax and SVM methods are studied by Cengil [34]. When SVM is used as a classifier in CNN models, it increases the accuracy [35] [36]. The work resulted in Resnet18 giving the highest classification accuracy of 99.83% across the four leukocyte classes, those being neutrophils, eosinophils, lymphocytes, and monocytes.

In summary, regarding the classes used for model training and testing, it is noticeable in the aforementioned works that their authors use five or four cell classes, except Acevedo et al. [15]. However, in this work, different datasets containing eight to four classes are addressed to validate the model. Furthermore, it is important to note that the data augmentation technique for training and testing the model is present in most of these works, being characterized as an essential step of the model. Finally, it is notable in the literature that high accuracy for classifying blood cells can be achieved using Transfer Learning and CNN.

2.2 Datasets

One of the databases used for testing and training the model was provided by a study conducted at the Hospital das Clínicas in Barcelona, which has 17092 images of eight different classes, according to [33]. Among the eight classes there are six of leukocytes (lymphocytes, monocytes, neutrophils, eosinophils, basophils, immature granulocytes), one of erythroblasts (nucleated erythrocytes); and one of platelets (macroplatelets and giant platelets). In addition, the arrangement of cell images by class can be seen in Table 1. For organization purposes we will call this dataset Dataset A.

With this same database, it was balanced using the Data Augmentation technique, as performed by Wang [31]. For purposes of understanding, we will call this dataset Dataset B. In it, the classes with dominance in the dataset smaller than 10%, in Dataset A, underwent Random Flip Vertical and Horizontal, besides Random Rotation with a factor equal to 0.2, resulting in an output rotated by a random amount in the range $[-20\% * 2\pi,\ 20\% * 2\pi]$. Thus, the classes that underwent data augmentation were basophils (previously at 7.13%), monocytes (previously at 8.31%), erythroblasts (previously at 9.07%), and lymphocytes (previously at 7.10%). Their transformation due to this technique can be seen from Figure 1 to 4. In addition, the updated values of the number of images per class in Dataset B can be seen in Table 2.

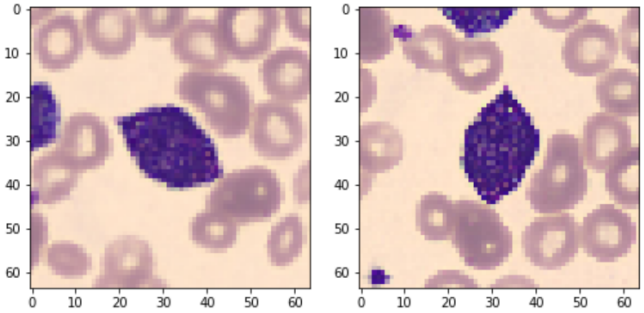

Figure 1. (a) Original image of a basophil and (b) Same image after data augmentation.

Table 1. Types and numbers of cells in each group in Dataset A [33].

CELL TYPE	TOTAL OF IMAGES BY TYPE	%
neutrophils	3329	19.48
eosinophils	3117	18.24
basophils	1218	7.13
lymphocytes	1214	7.10
monocytes	1420	8.31
immature granulocytes (metamyelocytes, myelocites, and promyelocytes)	2895	16.94
erythroblasts	1551	9.07
platelets (thrombocytes)	2348	13.74
Total	17092	100

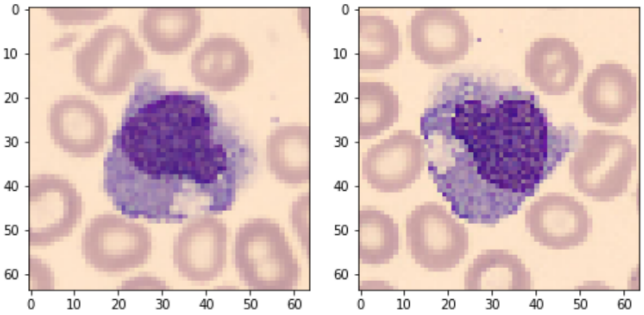

Figure 2. (a) Original image of a monocyte and (b) Same image after data augmentation.

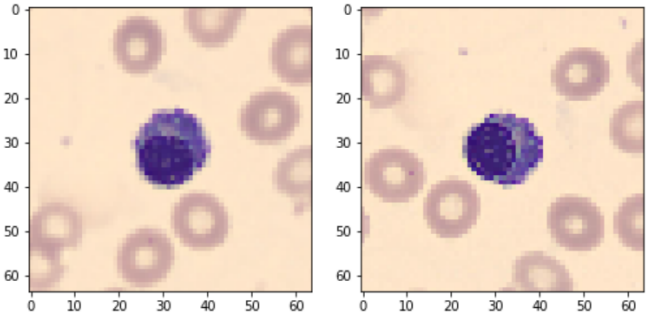

Figure 3. (a) Original image of an erythroblast and (b) Same image after data augmentation.

As Dataset C, let's call the Blood Cell Images dataset available in Kaggle [37]. In this one, there are in total 3120 images of eosinophils, 3113 lymphocytes, 3098 monocytes and 3123 neutrophils. For the test about 20% of each class was separated by the authors, and 80% for the training. The classes and their respective prevalence in dataset C can be seen in Table 3.

Besides this, there is also the Raabindata [38] dataset, which was taken as Dataset D in this work. Dataset D consists of a total of 301 images of basophils, 1066 images of

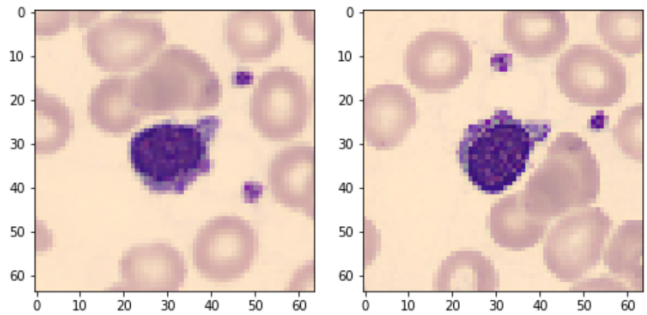

Figure 4. (a) Original image of a lymphocyte and (b) Same image after date augmentation.

eosinophils, 3461 images on lymphocytes, 795 images of monocytes, and 8884 images of neutrophil. For each one, 30% were separated – by the authors – for testing the model, 70% for training. Thus, we have the classes and their respective prevalence in dataset D presented in Table 4.

From Dataset A, we create Daset E, in which all the images have been transformed by the color correction function, introduced by Reinhard et al.[39]. In this function, first the average of the pixels in the image is subtracted. Then, the pixels are scaled by factors determined by their standard deviations. After this transformation, the resulting pixels have standard deviations that conform to those of the source image. Next, instead of adding the averages that were subtracted previously, we add the averages calculated for the image. In the Figure 5 it is possible to see the difference before and after the correction of colors.

Finally, Dataset F is composed by half of the images from each Dataset A, C and D, totaling 22021 cells. The arrangement of each class can be seen in Table 5. For training, 80% of each class was used, while 20% for testing, totaling 17617 cell images for training and 4404 for testing.

In summary, for visualization purposes, a neutrophil image from each dataset has been grouped together in Figure 7. It

Table 2. Types and numbers of cells in each group in Dataset B.

CELL TYPE	TOTAL OF IMAGES BY TYPE	%
neutrophils	3329	14.80
eosinophils	3117	13.86
basophils	2436	10.83
lymphocytes	2428	10.79
monocytes	2840	12.62
immature granulocytes (metamyelocytes, myelocites, and promyelocytes)	2895	12.87
erythroblasts	3102	13.79
platelets (thrombocytes)	2348	10.44
Total	22495	100

Table 3. Types and numbers of cells in each group in Dataset C, and the split done for training and testing.

CELL TYPE	TOTAL OF IMAGES FOR TRAINING (%)	TOTAL OF IMAGES FOR TESTING (%)	TOTAL OF IMAGES BY TYPE
eosinophils	2497 (80.04%)	623 (19.96%)	3120
lymphocytes	2483 (80.09%)	620 (19.91%)	3113
monocytes	2478 (80%)	620 (20%)	3098
neutrophils	2499 (80.02%)	624 (19.98%)	3123
Total	9957 (80.02%)	2487 (19.98%)	12444

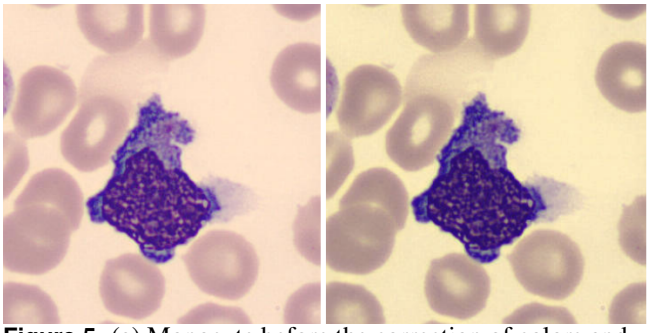

Figure 5. (a) Monocyte before the correction of colors and (b) after the correction of colors.

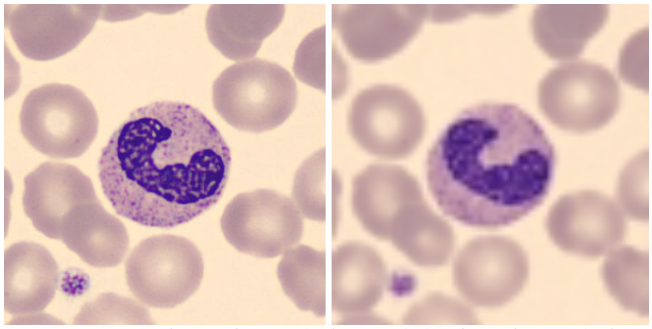

Figure 6. (a) 360 x 363 resolution and (b) 64 x 64 resolution.

is important to note that reduction of the resolution of the images was employed in all datasets used in this work. Thus, the image is transformed from the original size to 64x64, in order to make the model more robust, because in this way, one can simulate the camera scenario of a cell phone with a low resolution camera. The visual difference can be seen in Figure 6.

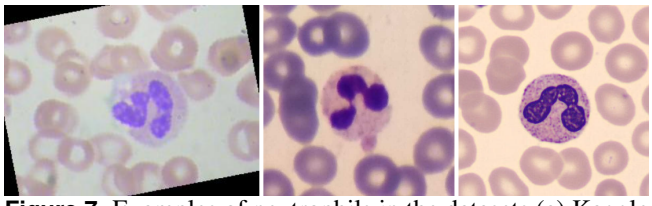

Figure 7. Examples of neutrophils in the datasets (a) Kaggle Blood Cells Images [37], (b) Raabin-WBC [38], and (c) Acevedo et al. [33].

Table 4. Cell types and numbers in each group in Dataset D, and the split done for training and testing.

CELL TYPE	TOTAL OF IMAGES FOR TRAINING (%)	TOTAL OF IMAGES FOR TESTING (%)	TOTAL OF IMAGES BY TYPE
basophils	212 (70.54%)	89 (29.56%)	301
eosinophils	744 (69.80%)	322 (30.20%)	1066
lymphocytes	2427 (70.13%)	1034 (29.87%)	3461
monocytes	561 (70.57%)	234 (29.43%)	795
neutrophils	6231 (71.14%)	2653 (29.86%)	8884
Total	10175 (71.14%)	4332 (29.86%)	14507

Table 5. Types and numbers of cells in each group in Dataset F.

CELL TYPE	TOTAL OF IMAGES BY TYPE	%
neutrophils	7665	34.8
eosinophils	3651	16.57
basophils	759	3.4
lymphocytes	3894	17.68
monocytes	2.656	12.06
immature granulocytes (metamyelocytes, myelocites, and promyelocytes)	1.447	6.5%
erythroblasts	775	3.5
platelets (thrombocytes)	1.174	5.3
Total	22021	100

2.3 Proposed Model

In this work, CNN ResNet18 was used for image pre-processing, due to its high accuracy and reasonable validation time, as shown in Figure 8, and the Support Vector Machine (SVM) classifier was chosen to replace the dense layers - referring to the neural network - due to the fact that it has presented good results in several studies [22] [35] [36]. In addition to SVM requiring less computational resources compared to dense layers, the result achieved by both is very similar in classifying the images. Tests were made using only ResNet18 to preprocess and classify the images. The F1-Score achieved by this model on Dataset A, for example, averaged 97.89%. Nevertheless, on Dataset C an F1-Score of on average 83.2% was achieved. Remember that the F1-Scores achieved in Dataset A and C by the proposed model are: 97% and 82.85%.

The Figure 9 represents the architecture of the model. However, because SVM is a binary classifier, the modification made to involve multiclassing is based on the "One-Vs-Rest" method, in which each class is compared individually with all the others - simulating a single class - until all have been compared. The other configurations used in the SVM were chosen based on the Grid Search method, which consists of testing different combinations of parameters when training the model and returning the one that resulted in the best accuracy. Three experiments were performed and the regularization parameters tested were 0.1, 1, 10, and 100; additionally, the Kernel Radial Basis Function (RBF), sigmoid, polynomial, and linear functions were tested.

All experiments were carried out in the Google Collaboratory environment, which provides a free Jupyter notebook.

3. Results and Discussion

In order to better understand which features of the cell the model is taking into account, we use a technique called Class Activation Map (CAM). Thus, CAM allows highlighting discriminative regions used by CNN to identify a class in the image. The Figure 10(a) and (b) are examples of CAM in a basophil and a neutrophil, respectively. In these figures it is noticeable that the nucleus of the cell is the main feature taken into account for the classification of the cell.

Regarding the methods for evaluating the model, the metrics used in this paper were accuracy, precision, sensitivity, and F1-Score.

Accuracy is defined as the ratio of correct predictions to

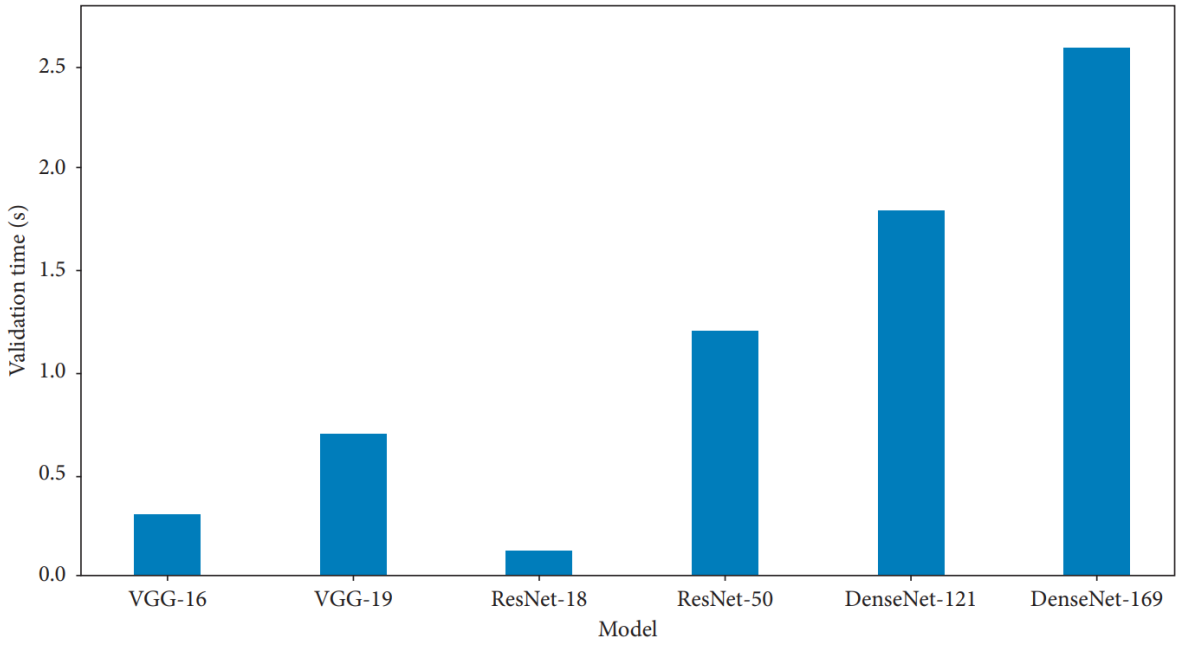

Figure 8. Time for models to perform inference on validation data [24].

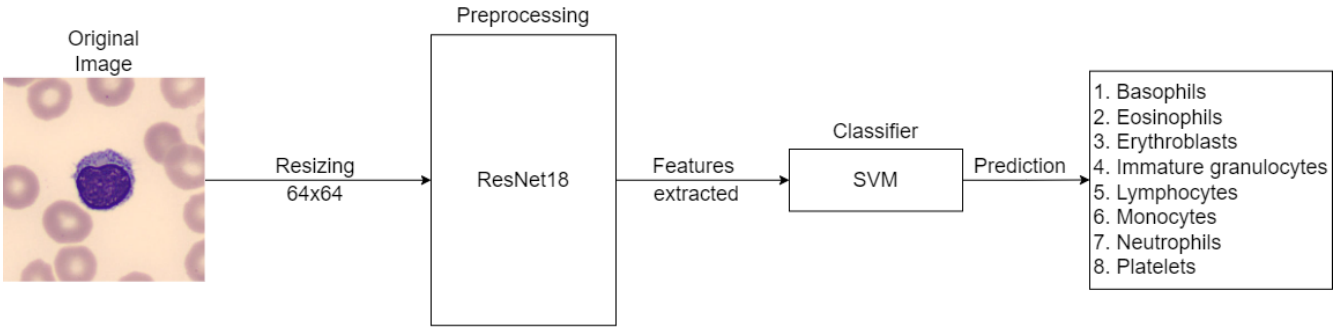

Figure 9. Architecture of the model of this work.

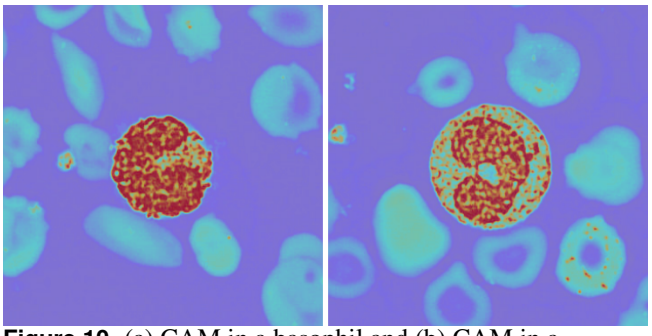

Figure 10. (a) CAM in a basophil and (b) CAM in a neutrophil.

total predictions made. Accuracy is calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision is defined as the ratio of correctly classified positive samples (True Positive) to the total number of classified positive samples (the sum of True and False Positives). Precision is calculated as follows:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Sensitivity, also known as Recall, can be described as the metric used for evaluating a model's ability to predict the true positives of each available category. The recall is calculated as follows:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score (or F-score) is defined as the harmonic mean of the precision and sensitivity. It combines the precision and sensitivity metrics into a single metric. The F1-score has been designed to work well on imbalanced data. F1-score is calculated as follows:

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2*TP}{2*TP+FP+FN} \quad (4)$$

Table 6. Comparison of the datasets used in this article and their respective accuracy achieved by the model.

Datasets	Classes	Blood Cell Types	Images Number	Data Augmentation	Split	F1-Score
Dataset A	8	neutrophils, eosinophils, lymphocytes, monocytes, basophils, immature granulocytes (metamyelocytes, myelocites, and promyelocytes), erythroblasts, platelets (thrombocytes)	17092	No	Ours (70/30)	97%
Dataset B	8	neutrophils, eosinophils, lymphocytes, monocytes, basophils, immature granulocytes (metamyelocytes, myelocites, and promyelocytes), erythroblasts, platelets (thrombocytes)	22495	Yes	Ours (70/30)	97.47%
Dataset C (Authors' split)	4	neutrophils, eosinophils, lymphocytes, monocytes	12444	No	Authors	82.85%
Dataset C	4	neutrophils, eosinophils, lymphocytes, monocytes	12444	No	Ours (70/30)	99.96%
Dataset D (Authors' split)	5	neutrophils, eosinophils, lymphocytes, monocytes, basophils	14507	No	Authors	99.91%
Dataset D	5	neutrophils, eosinophils, lymphocytes, monocytes, basophils	14507	No	Ours (70/30)	96.31%

During the Grid Search tests for the SVM classifier, the best regularization parameter compared between the values 0.1, 1, 10, and 100 was equal to 10; the best Kernel function found was the Radial Basis Function (RBF), compared to the sigmoid, polynomial and linear functions.

The proposed model reached, in Dataset A, an accuracy of 97.2%, and the F1-Score of 97% considering all classes; moreover, the normalized confusion matrix can be visualized in Figure 11, in which the lines represent the real class, and the columns the predicted class. It is notable that the eosinophils were correctly classified in 99.89% of the tests, neutrophils in 98.30%, and platelets in 100%. Immature granulocytes were correctly classified in 94.25% of the tests and were responsible for the misclassification of basophils in 81% and monocytes in 95%.

Nevertheless, the highest average accuracy achieved by the model in Dataset B, after three different tests with the same model, was 97.43%. Moreover, its F1-Score was 97.47%. The normalized confusion matrix can be seen in Figure 12.

Since balancing the dataset with data augmentation was not helpful enough compared to the computational costs, it was decided not to perform this technique on the following datasets.

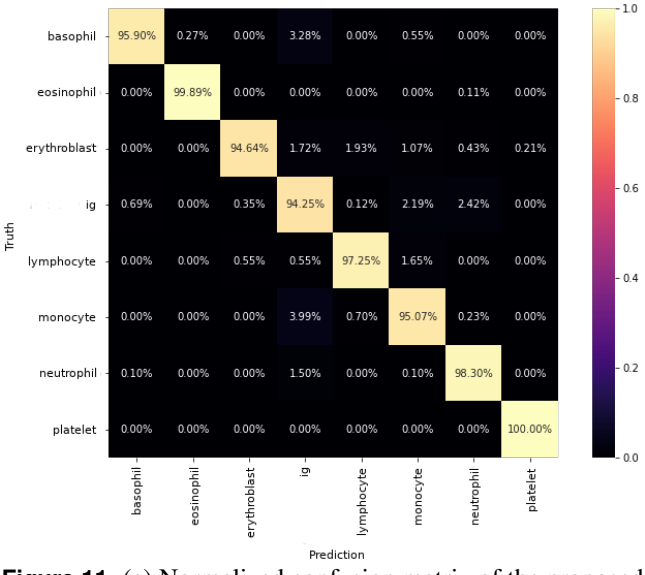

Figure 11. (a) Normalized confusion matrix of the proposed model in Dataset A, where the rows represent the actual class and the columns the predicted class.

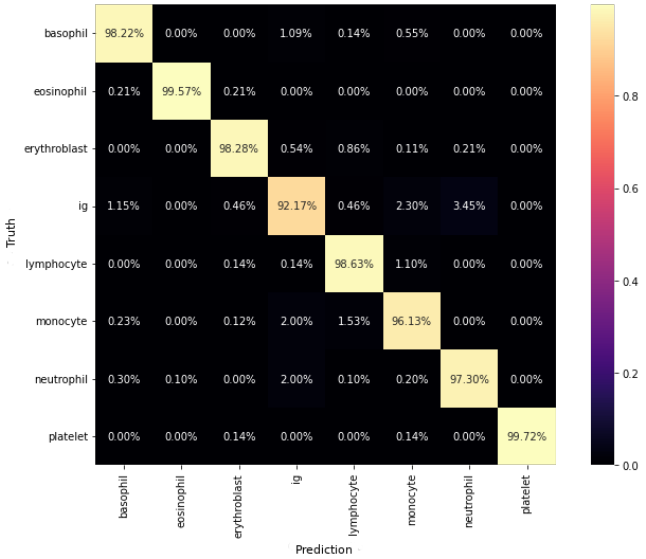

Figure 12. (a) Normalized confusion matrix of the proposed model in Dataset B, where the rows represent the actual class and the columns the predicted class.

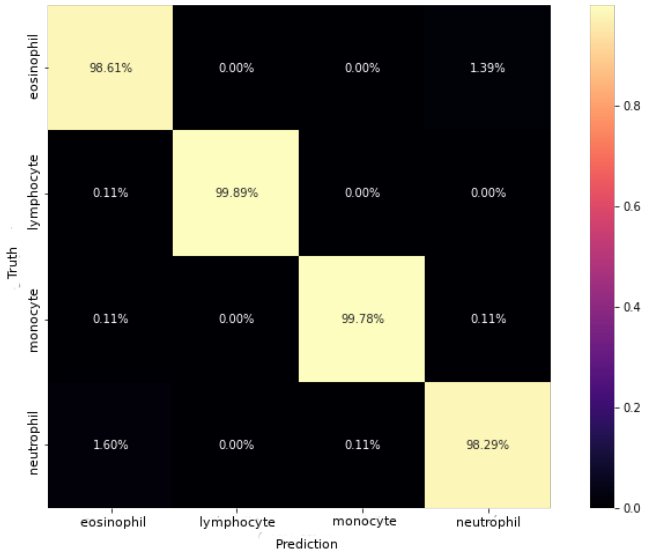

Figure 14. (a) Normalized confusion matrix of the proposed model in Dataset C, with our split, in which the rows represent the actual class and the columns the predicted class.

On dataset C, using the authors' training/test partition, the model achieved an accuracy of 82.37% and an F1-Score of 82.85%. Furthermore, the values of accuracy and F1-Score in training were 99.96% both. Thus, characterizing it as an overfitting of the model. However, when ignoring the authors' split and performing a split of its own – of 30% for testing and 70% for training – the model achieved accuracy and an F1-Score of 99.91%. Suspicion of model overfitting was ceased by realizing a training accuracy and F1-Score at 99.14%. The normalized confusion matrix of the two aforementioned cases can be checked in Figures 13 and 14, respectively.

The accuracy and F1-Score achieved in Dataset D with the authors' training/test split were 96.12% and 93.58%. While using our own split mentioned earlier, the accuracy achieved was 97.88% and the F1-Score was 96.31%. Its normalized confusion matrix can be seen in Figures 15 and 16.

In Dataset E, the accuracy and F1-Score achieved were very close to the obtained in dataset A. 97.51% and 97.18%. Thus, one can see that the color correction presented does not make much difference to the model.

Finally, in Dataset F, the achieved accuracy was 90.89% and F1-Score was 89.14%.

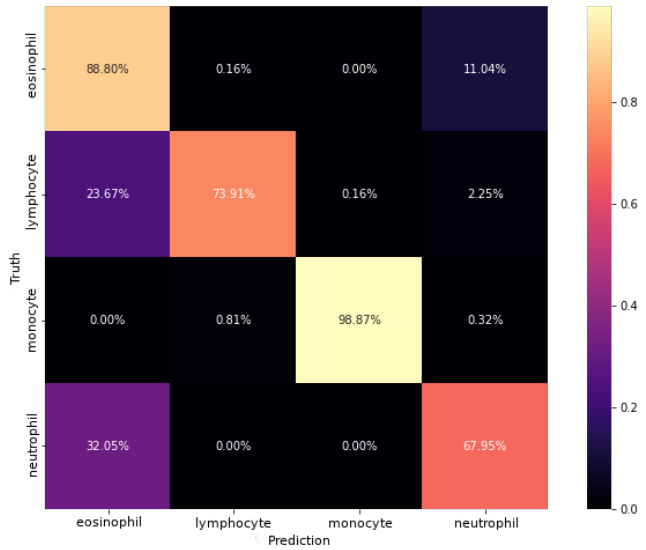

Figure 13. (a) Normalized confusion matrix of the proposed model in Dataset C, with the authors' split, in which the rows represent the actual class and the columns the predicted class.

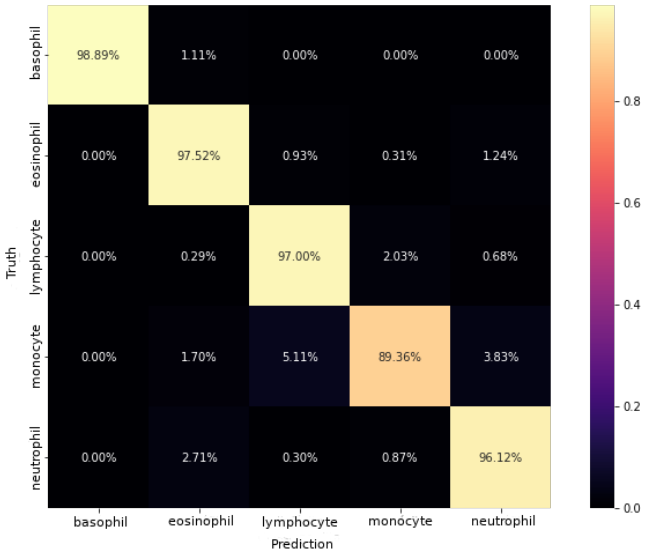

Figure 15. (a) Normalized confusion matrix of the proposed model in Dataset D, with the authors' split, in which the rows represent the actual class and the columns the predicted class.

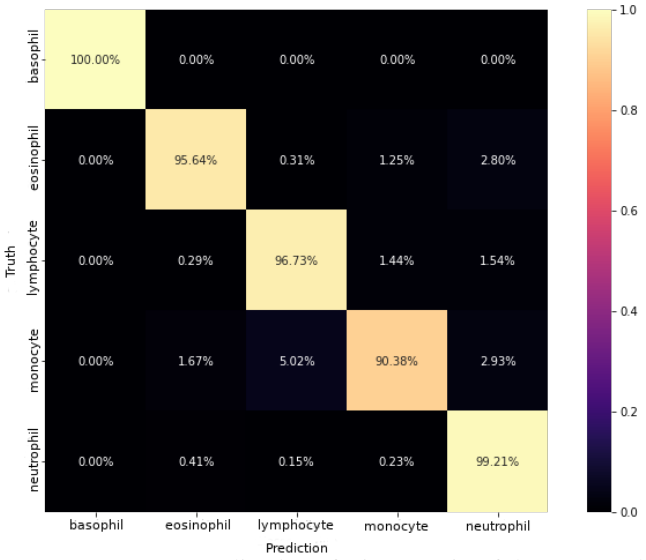

Figure 16. (a) Normalized confusion matrix of the proposed model in Dataset D, with our split, in which the rows represent the actual class and the columns the predicted class.

From the technical point of view, it is important to note that the limit of RAM memory in Colab makes it difficult to perform experiments with high-resolution images and makes them slow; this takes a long time to complete them and increases the possibility of problems such as memory overflow and losing all processing. Regarding the results obtained, the positive highlights were the excellent hit rates in the classification of platelets, eosinophils, and neutrophils. On the other hand, the negative highlights, possibly caused by the low number of images, were the monocytes, basophils, and erythroblasts. Although with a large number of images in the immature granulocyte class, there were negative impacts on the monocyte and basophil classifications. It is noteworthy that the immature granulocyte class is composed of three distinct cell types with different morphological characteristics, which may have impaired the classifications of monocytes and basophils.

Another positive aspect worth noting was that the few misclassifications occurred between morphologically similar classes that can commonly raise doubts in a trained professional.

In general, the model performed well, considering the reduced resolution and the imbalance between the number of images in each class of the database. Moreover, the hit rates were higher than those presented in many works cited in the state of the art and that used a smaller number of classes in their algorithms.

In summary, the objective of this work is the automatic classification of nucleated cells of peripheral blood, through the use of machine learning algorithms. The purpose of the study of classification models will be the implementation of the HemoVision application, designed so that students in the health area, especially in hematology, can use it to solve their doubts quickly, making the elucidation process efficient and autonomous. In addition, because it is a social network, interest and debate on the subject will be encouraged through posts and comments from the community.

In view of the results obtained in this work, machine learning algorithms can be a good tool to assist in the classification tasks of peripheral blood microscopy images and be employed as a support to professionals and students, both for educational purposes and for diagnostic support.

Author contributions

In this research, the conceptualization was mainly done by William L. Bertemes, who had the initial idea to develop a social network for health students and professionals interested in blood cell images classification, and Vinicius S. Patto, who improved this idea, proposing the use of machine learning to classify images and help users to improve their ability to classify blood cell images.

Considering the methodology, Mariana D. X. S. Santos and Mateus H. B. Andrades were mainly involved in the designing study, including selecting the research design, choice of algorithms, the sample size, and the data collection methods, Vinicius had small contributions in this stage, working as an advisor for Mariana and Mateus.

William, Mariana and Mateus conducted the data collection, gathering images and analyzing their quality and diversity.

Data analysis was mainly done by Mariana and Mateus, who were responsible for instance to analyzing the Accuracy, F1 Score and other metrics, considering the confusion matrix.

Iaan M. Souza contributed to the development of HemoVision mobile app, elicitating, analyzing and validating requirements. He developed a functional prototype architecting the front-end and back-end of the application to integrate it with the machine learning algorithms presented in this research. William was considered the product owner.

Mariana was the main author of this article; she had the collaboration of William, Vinicius and Mateus.

References

[1] FAILACE, R.; FERNANDES, F. *Hemograma: manual de interpretação*. 6. ed. Porto Alegre: Artmed, 2015.

[2] BAIN, B. J. Diagnosis from the blood smear. *New England Journal of Medicine*, Oxford, v. 353, p. 498–507, aug. 2005.

[3] CHABOT-RICHARDS, D. S.; FOUCAR, K. Does morphology matter in 2017? an approach to morphologic clues in non-neoplastic blood and bone marrow disorders. *International Journal of Laboratory Hematology*, Oxford, v. 39, n. S1, p. 23–30, may 2017.

[4] JUN, G. et al. Convolutional neural networks for computer-aided detection or diagnosis in medical image

analysis: an overview. *Mathematical Biosciences and Engineering*, Springfield, v. 16, n. 6, p. 6536–6561, jul. 2019.

[5] CHEN, S.; BILLINGS, S. A.; GRANT, P. M. Non-linear system identification using neural networks. *International Journal of Control*, Amsterdam, v. 51, n. 6, p. 1191–1214, mar. 1990.

[6] HE, J. et al. The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, New York, v. 25, p. 30–36, jan 2019.

[7] JORDAN, M.; MITCHELL, T. Machine learning: trends, perspectives, and prospects. *Science*, New York, v. 349, n. 6245, p. 255–260, jul. 2015.

[8] LIANG, G. et al. Combining convolutional neural network with recursive neural network for blood cell image classification. *IEEE*, New York, v. 6, p. 36188–36197, jul. 2018.

[9] BO, B. et al. Optogenetic excitation of ipsilesional sensorimotor neurons is protective in acute ischemic stroke: a laser speckle imaging study. *IEEE Trans.Biomed. Eng.*, New York, v. 66, n. 5, p. 1372–1379, may 2018.

[10] LIU, Q. et al. Monitoring acute stroke in mouse model using laser speckle imaging-guided visible-light optical coherence tomography. *IEEE Trans. Biomed. Eng.*, New York, v. 65, n. 10, p. 2136–2142, may 2017.

[11] IRVIN, J. et al. Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Arxiv preprint arXiv:1901.07031*, [S. l.], p. 1–9, jan 2019.

[12] ISENSEE, F. et al. No new-net. In: *4th International MICCAI Brain lesion Workshop*. Cham, Switzerland: Springer, 2018. p. 234–244.

[13] ÇIÇEK, Ö. et al. 3d u-net: Learning dense volumetric segmentation from sparse annotation. *Arxiv preprint arXiv: 1606.06650*, [S. l.], p. 1–8, jun. 2016.

[14] DUAN, Y. et al. Leukocyte classification based on spatial and spectral features of microscopic hypers-pectral images. *Optics LaserTechnology 112*, Amsterdam, v. 112, p. 530–538, apr. 2019.

[15] ACEVEDO, A. et al. Recognition of peripheral blood cell images using convolutional neural networks. *Computer Methods and Programs in Biomedicine*, Amsterdam, v. 180, p. 1–16, oct. 2019.

[16] WIBAWA; S., M. A comparison study between deep learning and conventional machine learning on white blood cells classification. In: *International Conference on Orange Technologies (ICOT)*. New York: IEEE, 2018. p. 1–6.

[17] WEBER, E. U.; COSKUNOGLU, O. Descriptive and prescriptive models of decision- making: implications for the development of decision aids. *IEEE transactions on Systems, Man, and Cybernetics*, New York, v. 20, n. 2, p. 310–317, mar. 1990.

[18] PUTTAMADEGOWDA, J.; PRASANNAKUMAR, S. White blood cell segmentation using fuzzy c means and snake. In: *Computation System and Information Technology for Sustainable Solutions (CSITSS)*. New York: IEEE, 2016. p. 47–52.

[19] GAUTAM, A.; BHADAURIA, H. White blood nucleus extraction using k-mean clustering and mathematical morphing. In: *Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit*. New York: IEEE, 2014. p. 549–554.

[20] ALREZA, Z. K. K.; KARIMIAN, A. Design a new algorithm to count white blood cells for classification leukemic blood image using machine vision system. In: *6th International Conference on Computer and Knowledge Engineering (ICCKE)*. New York: IEEE, 2016. p. 251–256.

[21] PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. *Arxiv preprint arXiv:1712.04621*, [S. l.], p. 1–8, dec. 2017.

[22] YAO, X. Classification of white blood cells using weighted optimized deformable convolutional neural networks. *Artificial Cells, Nanomedicine, and Biotechnology*, London, v. 49, n. 1, p. 147–155, feb. 2021.

[23] MA, L.; SHUAI, R.; RAN, X. E. A. Combining dc-gan with resnet for blood cell image classification. *Medical & Biological Engineering & Computing*, New York, v. 58, p. 1251–1264, mar. 2020.

[24] ALMEZHGHWI, K.; SERTE, S. Improved classification of white blood cells with the generative adversarial network and deep convolutional neural network. *Computational Intelligence and Neuroscience*, New York, v. 2020, p. 1–12, jul. 2020.

[25] KRIZHEVSKY, A. The cifar dataset. *Universidade de Toronto*, 2019. Disponível em: ⟨https://www.cs.toronto.edu/~kriz/cifar.html⟩. Acesso em:22 junho 2023.

[26] KHAN, A. et al. White blood cell type identification using multi-layer convolutional features with an extreme-learning machine. *Biomedical Signal Processing and Control*, Oxford, v. 69, p. 102932, aug. 2021.

[27] RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, Cambridge, v. 29, n. 9, p. 2352–2449, sep. 2017.

[28] KONONENKO, I. Estimating attributes: Analysis and extensions of relief. In: *Machine Learning: ECML-94*. Cham, Switzerland: Springer, 1994. p. 171–182.

[29] KATAR, O.; KILINCER, I. F. Automatic classification of white blood cells using pre-trained deep models. *Sakarya University Journal of Computer and Information Sciences*, Turkey, v. 5, n. 3, p. 462–476, dec. 2022.

[30] DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer*

Vision and Pattern Recognition. New York: IEEE, 2009. p. 248–255.

[31] WANG, Y.; CAO, Y. Human peripheral blood leukocyte classification method based on convolutional neural network and data augmentation. *Medical Physics*, Woodbury, v. 47, n. 1, p. 142–151, jan. 2019.

[32] LONG, F. et al. Bloodcaps: A capsule network based model for the multiclassification of human peripheral blood cells. *Computer Methods and Programs in Biomedicine*, Amsterdam, v. 202, p. 105972, apr. 2021.

[33] ACEVEDO, A. et al. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Computer Methods and Programs in Biomedicine*, Amsterdam, v. 30, p. 1–5, jun. 2019.

[34] CENGIL, E.; ÇINAR, A.; YILDIRIM, M. A hybrid approach for efficient multi-classification of white blood cells based on transfer learning techniques and traditional machine learning methods. *Concurrency and Computation: Practice and Experience*, Chichester, v. 34, n. 6, p. e6756, mar. 2021.

[35] MANOHAR, N. et al. Convolutional neural network with svm for classification of animal images. In: SRIDHAR, V.; PADMA, M.; RAO, K. R. (Ed.). *Emerging Research in Electronics, Computer Science and Technology*. Singapore: Springer Singapore, 2019. p. 527–537.

[36] AGARAP, A. F. An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *Arxiv preprint arXiv: 1712.03541*, [S. l.], p. 1–4, dec. 2017.

[37] MOONEY., P. Blood cell images. *Kaggle*, 2020. Disponível em: ⟨https://www.kaggle.com/paultimothymooney/blood-cells⟩. Acesso em: 22 junho 2023.

[38] KOUZEHKANAN, Z. M. et al. Raabin-wbc: a large free access dataset of white blood cells from normal peripheral blood. *bioRxiv*, Cold Spring Harbor Laboratory, p. 1–24, 2021. (No prelo). Disponível em: ⟨https://www.biorxiv.org/content/early/2021/05/29/2021.05.02.442287⟩.

[39] REINHARD, E. et al. Color transfer between images. *IEEE Computer Graphics and Applications*, New York, v. 21, n. 5, p. 34–41, jul. 2001.