

Article

# Blob Detection and Deep Learning for Leukemic Blood Image Analysis

Cecilia Di Ruberto , Andrea Loddo \*  and Giovanni Puglisi

Department of Mathematics and Computer Science, University of Cagliari, 09124 Cagliari CA, Italy; dirubert@unica.it (C.D.R.); puglisi@unica.it (G.P.)

\* Correspondence: andrea.loddo@unica.it

Received: 9 December 2019; Accepted: 2 February 2020; Published: 10 February 2020



**Abstract:** In microscopy, laboratory tests make use of cell counters or flow cytometers to perform tests on blood cells, like the complete blood count, rapidly. However, a manual blood smear examination is still needed to verify the counter results and to monitor patients under therapy. Moreover, the manual inspection permits the description of the cells' appearance, as well as any abnormalities. Unfortunately, manual analysis is long and tedious, and its result can be subjective and error-prone. Nevertheless, using image processing techniques, it is possible to automate the entire workflow, both reducing the operators' workload and improving the diagnosis results. In this paper, we propose a novel method for recognizing white blood cells from microscopic blood images and classify them as healthy or affected by leukemia. The presented system is tested on public datasets for leukemia detection, the SMC-IDB, the IUMS-IDB, and the ALL-IDB. The results are promising, achieving 100% accuracy for the first two datasets and 99.7% for the ALL-IDB in white cells detection and 94.1% in leukemia classification, outperforming the state-of-the-art.

**Keywords:** leukocyte detection; leukemia classification; blob detection; deep learning

## 1. Introduction

Among medicine's branches, hematology studies the diagnosis and treatment of patients with blood or bone marrow disorders. Hematologists perform a wide range of laboratory tests to produce and interpret results assisting clinicians in the diagnosis and treatment of disease. They are related to the blood and the bone marrow to provide direct clinical care to patients. Indeed, in certain circumstances, predicting how the bone marrow may have contributed to a clinical condition may be more important than identifying the patient's hematologic condition. Hematologists generally receive samples of blood smear and examine them, looking for abnormalities; therefore, if they detect the possible presence of pathologies, they perform a bone marrow biopsy and give a diagnosis in a short time.

Healthy individuals' bone marrow releases only mature and nearly mature cells into the bloodstream. However, diseases, disorders, or deficiencies can induce the bone marrow to release immature or abnormal cells into the bloodstream, affecting the number and type of blood cells produced, their functions, and their lifespan.

Different tests exist to supervise blood cell proportions and numbers. For example, the Complete Blood Count (CBC) indicates the numbers and types of cells in the peripheral circulation. Pathologists compare cell percentages with the reference ranges to determine if they are present in a reasonable proportion, if a cell type is increased or decreased, or if there are any immature cells. Reference ranges for blood tests are the set of values in which 95% of the healthy population falls. Test results performed over a representative selection of the population help to define them.

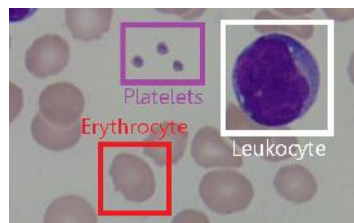
Automatic cell counters or flow cytometry are examples of quick methods to perform CBC automatically. Nevertheless, if they indicate the certain or possible presence of abnormal cells, hematologists realize a blood smear in order to analyze this manually in any case. Although the manual analysis of blood smears depends on the operator's skills, opinion, and tiredness, it is also particularly helpful to categorize and identify pathological conditions through the description of the cells' appearance, as well as any cell abnormalities. The drawbacks are several. Indeed, it is lengthy, repetitive, and error-prone, suffers from the absence of standard procedures, and is subjective because different operators can produce several interpretations of the same scene.

In this work, we investigate and implement image processing techniques to automatize the blood cells counting procedure, particularly the white blood cells, and classify them as affected by leukemia or not.

The rest of the paper is organized as follows. Section 2 illustrates the background regarding peripheral blood analysis. Section 3 presents some related works. The proposed method for cell segmentation, counting, and classification is explained in Section 4, while Section 5 discusses experimental results and gives detail about the datasets used. Finally, Section 6 offers conclusions and future aspects.

## 2. Background

Blood images obtained from healthy patients usually consist of three components, at least: the platelets (or thrombocytes), the Red Blood Cells (RBCs or erythrocytes), and the White Blood Cells (WBCs or leukocytes), as shown in Figure 1.



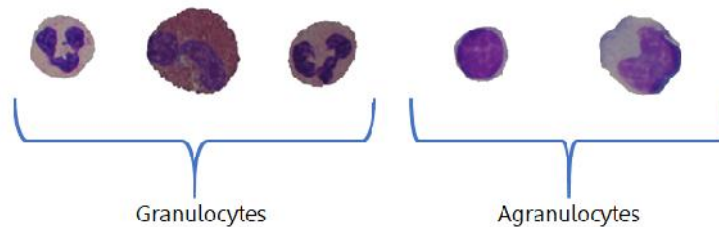
**Figure 1.** Blood smear components. The red rectangle presents an example of an erythrocyte; the white box shows a leukocyte; while the violet one includes some platelets.

The blood platelets are the smallest cells of the blood, averaging about 2 to 4  $\mu\text{m}$  in diameter. They are disc shaped and do not present a nucleus. The tiny granules of platelets contain substances necessary for the clot formation activity because their function is related to homeostasis and the prevention and control of bleeding. For this reason, a discrete number of platelets is necessary. As a consequence, a platelet count is usually evaluated by preparing a blood smear to visualize any anomalies in shape or size directly.

RBCs are round and have the form of biconcave disks with a diameter of 7–8  $\mu\text{m}$ . Peripherally, the presence of hemoglobin gives an area with a central pallor equal to 1–3  $\mu\text{m}$ , approximately 30–45% of the diameter of the cells. RBCs exist in different types, and each one has a particular form. Abnormal variability in their size, unusual variation in shape, and significant differences among erythrocyte types may indicate the presence of a disease [1]. At the same time, their color is representative of the concentration of the hemoglobin in the cell. The cytoplasm of all healthy RBCs does not contain any granule or other structures, e.g., typical of WBCs, while the inclusions are the result of unique conditions, and their identification can be clinically helpful.

WBCs are the most prominent cells in peripheral circulation and range in size from 10 to 20  $\mu\text{m}$ . They are composed of a darker nucleus and lighter cytoplasm. Therefore, their identification is more intuitive than the other regions. However, WBCs have vast differences in their shapes, dimensions, and edges; consequently, their analysis results are challenging. Figure 2 shows that WBCs differentiate into two main groups: cells containing granules, called granulocytic or myelocytic, and cells without granules, called mononuclear or lymphoid. Thus, we can distinguish the cells according to their shape,

size, the presence of granules in the cytoplasm, and the number of lobes in the nucleus. The lobes are the most significant part of the nucleus, and thin filaments connect them. WBCs mature into five distinct types that include neutrophils, basophils, and eosinophils for the granulocytic cell type and lymphocytes and monocytes for the non-granulocytic one. Neutrophils constitute the majority of WBCs in a healthy adult.



**Figure 2.** The five types of WBCs: basophil, eosinophil, neutrophil, lymphocyte, and monocyte. There exist two main groups of leukocytes: the granulocytes and the agranulocytes. The granulocytes contain granules inside the cytoplasm. They are neutrophils, eosinophils, and basophils and have a multilobed nucleus. The agranulocytes include lymphocytes and monocytes; they do not contain granules and have non-lobular nuclei.

WBCs' life in peripheral circulation is short; therefore, any alteration either in quantity or quality of a particular type can be dramatic for the patient. Several diseases affect the absolute or relative count of WBCs and their appearance on a blood smear. This is the case of parasitic infections that lead to a rise in the number of WBCs, while the most severe cases are certainly leukemias. When this happens, the peripheral smear usually shows a vast number of immature or abnormal cells. In this context, the automated cell counters are not able to distinguish normal from abnormal cells, and even worse, they could fail due to the presence of abnormal cells. Many computer-aided systems for digitized images have been proposed in the last few years to overcome these issues.

### 3. Related Works

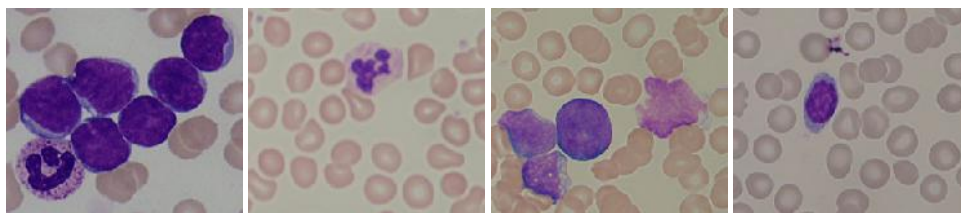
The state-of-the-art presents multiple systems devoted to leukocyte analysis, which face problems like segmentation, detection, or classification, to differentiate among the different types and distinguish if a cell is healthy or not. Here are some examples. Madhloom [2] used arithmetic and thresholding image operations to develop an automated system for WBC nuclei localization and segmentation. Sinha [3] attempted to differentiate the five types of leukocytes using k-means clustering in the HSV color space for WBC segmentation and different classification models for cell differentiation. Generally, images acquired from digital microscopy suffer from the presence of uneven lighting and a brighter central area region, caused by the combination of the lens, the lamplight, and the presence of shaded areas towards the corners. In this case, local approaches for segmentation are more appropriate; Putzu [4] employed a local fuzzy threshold to manage the local illumination variations and the presence of noise or imprecision. Several authors also faced the problem of cell counting as well. Khan [5] proposed a method based on an iterative threshold, determined from the histogram, to binarize the image. Then, the count was performed by extracting the connected components. Overlapping or adjacent cells were ignored. Putzu [6], on the other hand, used the distance transform in combination with the watershed algorithm to separate cell agglomerates and include them in the count. The separation worked only for small or small cell agglomerates. Mahmood and Alomari, instead [7,8], based their RBC and WBC counting methods on the Circular Hough Transform (CHT). Mahmood applied the CHT to binary images obtained from the Lab color space, while Alomari modified the CHT to reduce the number of cell candidates by selecting the one with a higher probability. Alilou [9] proposed a completely different approach: a detection phase using the grey level co-occurrence matrix was applied directly to the original images without any previous segmentation. Consequently, it produced a significant amount of false positives, since it

worked without any restriction on the area of interest. Several methods for leukemia detection exist in literature and typically classify the two most common types of leukemia: Acute Myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL). In particular, Mohapatra [10] used an ensemble of classifiers, and Putzu [6] used an SVM model. In the last few years, neural networks have been increasingly used to face this issue: Vincent [11] and Singh [12] employed them. At the same time, Vogado [13] realized a system for leukemia diagnosis using transfer learning in CNNs and SVM for classification.

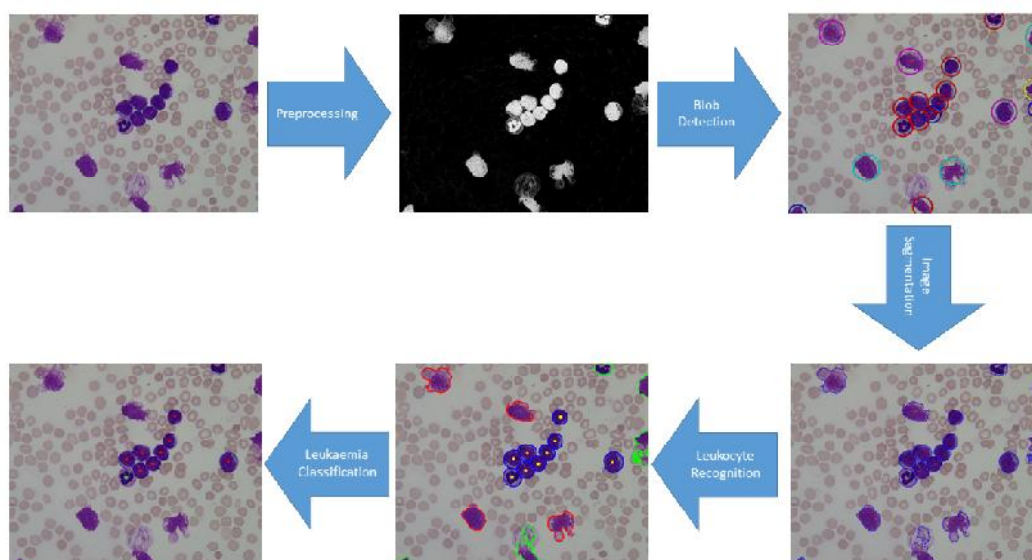
#### 4. Proposed Workflow

The main aim of the proposed solution is the detection and analysis of leukocytes included in blood images. As can be easily seen from visual inspection (see Figure 3), several factors such as the presence of staining artifacts, shape deformations, color variability, and clumps actually make the leukocyte detection task difficult. An algorithm made up of several steps has been developed in order to cope with the aforementioned challenging conditions (see Figure 4):

1. Preprocessing
2. Blob detection
3. Image segmentation
4. Leukocyte recognition
5. Leukemia classification



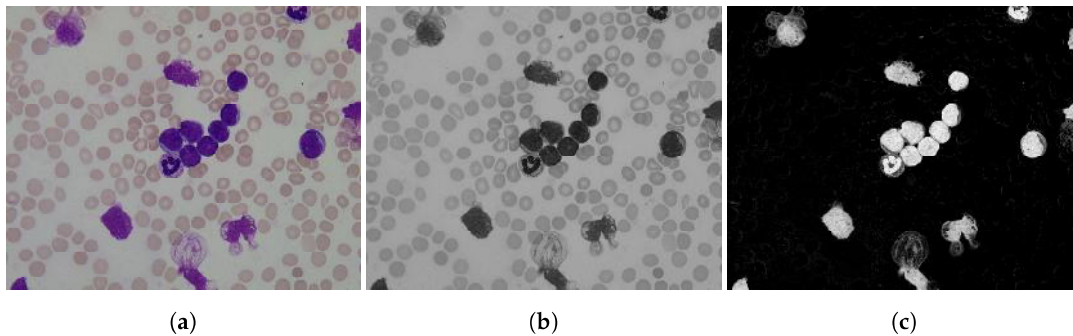
**Figure 3.** Blood images containing leukocytes, extracted from ALL-IDB1. First of all, they present different coloration and illumination conditions; therefore, their analysis should consider the consequent color variability. Moreover, further issues are represented. From left to right: leukocytes' clump formation, light blur, and low contrast between the background and the cytoplasm of WBCs, the presence of deformed WBCs and staining artifacts (light violet regions, similar to WBC's shape), poor contrast, and the presence of an artifact.



**Figure 4.** The overall scheme of the proposed solution.

#### 4.1. Preprocessing

A simple preprocessing algorithm was employed to visualize better the leukocytes included in a blood image. Specifically (see Figure 5), leukocytes have a peculiar color. Instead of merely performing a grayscale conversion, the original image  $I_{RGB}$  was converted from the RGB to the HSV (Hue, Saturation, and Value) color space. Then, thresholding in the hue space was applied, and a Boolean mask  $M_H$  was computed considering pixel values belonging to a specific range experimentally found. Finally, a new image, called  $I_{PRE}$ , was obtained, multiplying the saturation channel ( $I_S$ ) by the Boolean hue mask  $M_H$ . As can be seen from Figure 5, the preprocessed image  $I_{PRE}$  could be considered a better input with respect to the grayscale ones.



**Figure 5.** (a) Original RGB image, (b) grayscale image, and (c) preprocessed image exploiting both the hue and saturation channels.

#### 4.2. Blob Detection

Starting from the complement of the image  $I_{PRE}$ , a multiscale blob detection step based on the scale-normalized LoG (Laplacian of Gaussian) was employed [14] (see Figure 6):

$$L = \sigma^2(G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma)) \tag{1}$$

where:

$$G(x, y, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

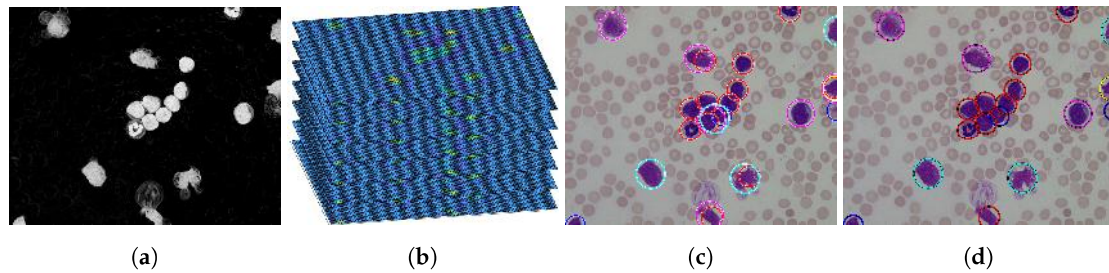
Moreover, considering a specific scenario in terms of image acquisition conditions (e.g., imaging device properties, zoom factor, and so forth), the typical sizes of leukocytes could be easily estimated. Then, all the parameters of the scale-normalized LoG (i.e.,  $\sigma$  values) could be set to work properly in the specified range of scales. As first, the input image was convoluted with a series of scale-normalized LoG with increasing  $\sigma$  values ( $\sigma_1, \sigma_2, \dots, \sigma_N$ ) obtaining  $N$  output images  $I_{\sigma_1}, I_{\sigma_2}, \dots, I_{\sigma_N}$ . The range of  $\sigma$  applied in the convolutions was strictly related to the radii of the blobs that had to be detected (from  $r_{min}$  to  $r_{max}$  in our tests). It is worth noting that the Laplacian achieved a maximum response to a binary circle of radius  $r$  when  $\sigma = r\sqrt{2}$ .

Once the three-dimensional  $(x, y, \sigma)$  scale-space was generated, local maxima were detected and considered as candidate blobs. To retain only the strongest local maxima, filtering based on an adaptive threshold, defined as follows, was then applied:

$$Th_{blob} = t \times V_{max} \tag{3}$$

where  $V_{max}$  is the maximum value computed in the scale-space and  $t \in [0, 1]$ .

Finally, a non-maxima suppression step was employed to refine the final results, as shown in Figure 6.



**Figure 6.** (a) Preprocessed image, (b) scale-space generated considering increasing  $\sigma$  values, (c) all the detected blobs at different scales obtained from the strongest local maxima, and (d) detected blobs after non-maxima suppression.

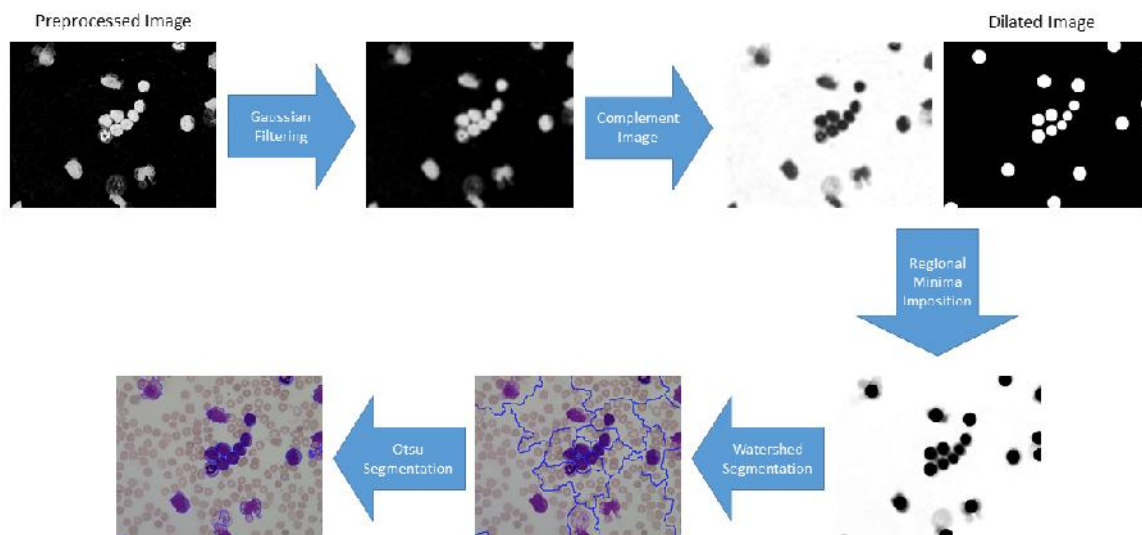
#### 4.3. Image Segmentation

Although the blob detection step provided useful information about leukocytes candidate locations, to improve the accuracy of the overall solution, a segmentation step based on the watershed algorithm [15] was employed (see Figure 7). Gaussian filtering was firstly applied to reduce the amount of noise in the preprocesses image  $I_{PRE}$ , and the complement image  $I_C$  was then computed. Later, the binary image generated from the locations of blob centroids was dilated with a disk-shaped structuring element. To avoid intersections, the radius of the dilated region related to a generic centroid  $c_k$  was set as follows:

$$r_k = 0.97 \times \min(r_{min}, dist_{min}/2) \quad (4)$$

where  $dist_{min}$  is the minimum Euclidean distance from  $c_k$  to the other centroids.

Both complement image  $I_C$  and dilated image  $I_D$  were then exploited to impose regional minima. Later, the watershed segmentation algorithm partitioned the entire images into  $K$  regions. Finally, a further segmentation step based on the Otsu algorithm [16] removed pixels belonging to the image background.



**Figure 7.** The overall scheme of the image segmentation step.

#### 4.4. Leukocyte Recognition

The previous steps of the developed pipeline could properly segment blob-like objects with hue values belonging to a specific range. Staining artifacts were also included in the set of segmented objects, due to their visual appearance both in terms of shape and color. A machine learning solution was then designed to perform a binary classification task and retain only leukocytes. As first, all the

regions obtained in the segmentation step were superimposed on an artificial background to remove the influence of neighboring elements. Each element detected in the original image  $I_{RGB}$  was then used to generate a new image containing only the object to be classified  $R_i$ . Two additional images were generated from  $R_i$  by applying a rotation followed by an averaging  $R_{i_{RA}}$  or a superimposition  $R_{i_{RS}}$  step to improve the overall accuracy of the method. Feature extraction was then performed with a pre-trained Convolutional Neural Network (CNN). In this work, an AlexNet [17] pre-trained with ImageNet [18] was used. This CNN contained five convolutional layers followed by three fully connected layers. Max-pooling layers were also employed. Feature vectors obtained separately from the aforementioned images  $R_i, R_{i_{RA}}, R_{i_{RS}}$  were then concatenated into a single vector  $v_f$ . Finally, the binary classification (leukocytes vs. staining artifacts) was performed by a linear support vector machine (see Figure 8).

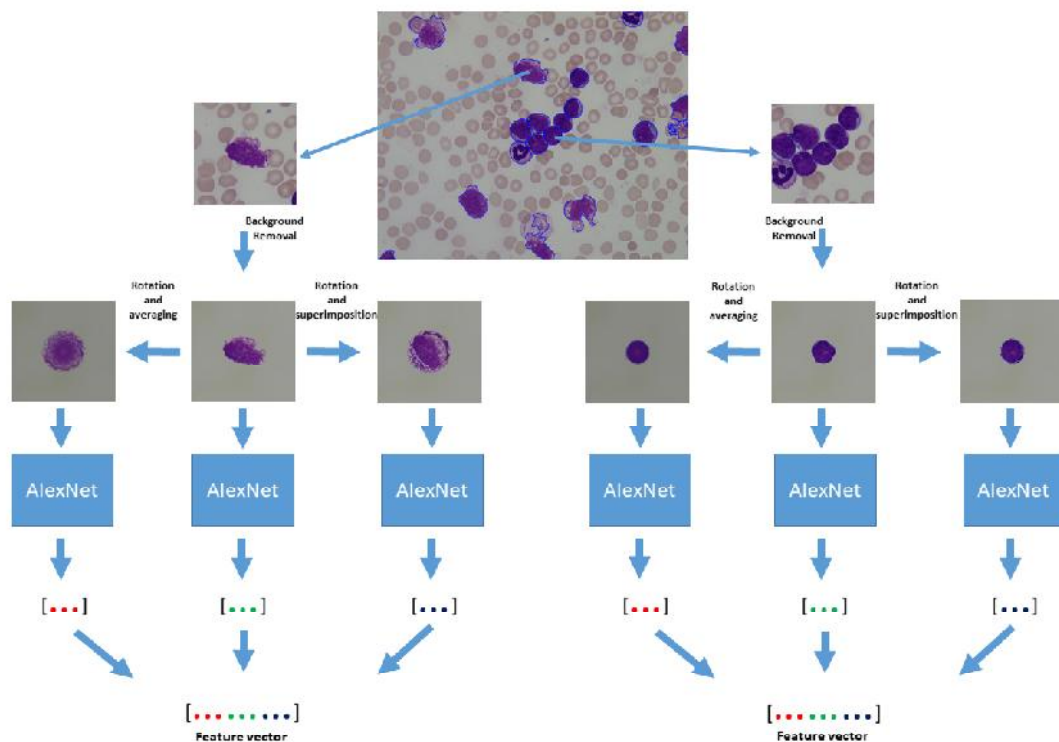


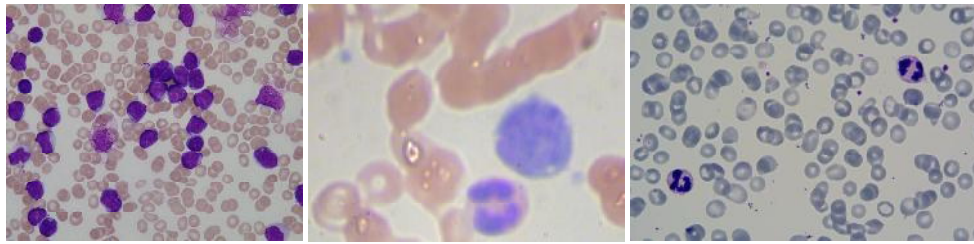
Figure 8. Feature extraction step applied to several segmented elements of the input image.

#### 4.5. Leukemia Classification

The final step of the proposed solution was devoted to recognizing leukocytes affected by leukemia. Specifically, the segmented regions previously identified as leukocytes ( $RL_i$ ) were considered at three different scales ( $s_1, s_2,$  and  $s_3$ ) and superimposed on an artificial background generating three novel images:  $RL_{s_1}, RL_{s_2}, RL_{s_3}$ . Later, feature extraction was performed with a pretrained convolutional neural network (AlexNet [17]). Finally, feature vectors obtained by  $RL_{s_1}, RL_{s_2}, RL_{s_3}$  were independently classified by three different linear SVMs and their results combined together applying a simple voting strategy.

### 5. Experimental Results

All the modules of the proposed system for leukocyte detection and classification were implemented in MATLAB. Three different and public datasets were used to test the proposed system. Each one is represented by a sample image in Figure 9.



**Figure 9.** Sample images extracted from the described datasets. From left to right: ALL-IDB1, SMC-IDB, and IUMS-IDB sample images.

### 5.1. Dataset Description: ALL-IDB

ALL-IDB [19] is a public image dataset of peripheral blood samples from healthy individuals and leukemia patients. The experts collected these samples at the M. Tettamanti Research Centre for childhood leukemia and hematological diseases, Monza, Italy. The ALL-IDB database is composed of two versions: ALL-IDB1 and ALL-IDB2, and its images are in JPG format with 24 bit color depth.

ALL-IDB1 is composed of 108 original RGB images captured with an optical laboratory microscope coupled with an Olympus Optical C2500L camera or a Canon PowerShot G5 camera. The first 33 images have a  $1712 \times 1368$  resolution, while the remaining images have a  $2592 \times 1944$  resolution. The images were taken with different magnifications of the microscope, ranging from 300 to 500, which brought the color and brightness differences. ALL-IDB1 presents full images containing either cells and agglomerates; therefore, it can be used to test the segmentation capability of algorithms, as well as image pre-processing methods or classification systems.

ALL-IDB2 is a collection of cropped areas of interest of healthy and blast cells extracted from ALL-IDB1. It contains 260 images, and 50% of these represent lymphoblasts.

We created an entirely new ground truth because ALL-IDB was designed to test the classification and not the segmentation performances. Moreover, to the best of our knowledge, no ground truths for segmentation evaluations are publicly available, not even from third parties.

### 5.2. Dataset Description: SMC-IDB

SMC-IDB was proposed in [20] and presented at IEEE's 2012 SMC conference; it is available at SMC-IDB ([www.mathworks.com/matlabcentral/fileexchange/36634](http://www.mathworks.com/matlabcentral/fileexchange/36634)). It is a public image dataset of peripheral blood samples, acquired from Giemsa-stained slides, as in ALL-IDB. Nevertheless, the images are different, since they were acquired with a different combination of microscopes and cameras. This dataset provides a total of 367 peripheral blood images of size  $640 \times 480$ .

### 5.3. Dataset Description: IUMS-IDB

The Iran University of Medical Science [21] dataset, available at IUMS-IDB (<http://misp.mui.ac.ir/fa/download>), presents 195 microscopic images of size  $3872 \times 2592$ , taken from the peripheral blood of eight healthy subjects. They visually differ from the ALL-IDB and SMC-IDB images.

### 5.4. Results and Comparison with State-of-the-Art

The performances of the proposed method were evaluated with the accuracy, precision, recall, and F-measure metrics, which according to the literature, are used most often for counting purposes. The results were obtained using leave-one-out cross-validation. It is essential to notice that the detection results were promising because the method produced only one False Negative (FN) and correctly identified 270 cells (out of 271) arranged in the first 33 images of ALL-IDB. In detail, 416 objects were detected. Then, the recognition step eliminated the cells touching the borders; therefore, the count decreased from 416 to 345. It identified correctly 270 cells (TP), misidentified one WBC (FN), and recognized correctly 74 artifacts (TN). We compared the performances of the proposed method with those obtained by other authors who used ALL-IDB to test their approaches for WBC counting.



It must be noted that several authors attempted to count WBCs in ALL-IDB; they are indicated in Table 1 and compared to our results. In particular, Mahmood [7] performed a color space conversion and then used the CHT with no restrictions on the area of interest. This method was only tested on ten images. On the contrary, Di Ruberto [22], Bhavnani [23] and Loddo [24] applied the CHT after a segmentation phase to restrict the search to a specific area of interest. They completed the counting on the first 33 images and 14 random images, respectively. In detail, Bhavnani [23] segmented the images using a fixed threshold over the G channel of the RGB color space. In contrast, Loddo [24] segmented the images using a machine learning approach that combined nearest neighbor and support vector machine. Putzu [6] combined the distance transform with the watershed algorithm to separate the agglomerates produced by the segmentation step. They both worked on the first 33 images of ALL-IDB1. Finally, Alilou [9] used a sliding windows strategy to classify each candidate bounding box using hand-crafted features; he tested his method on an unknown number of images.

**Table 1.** WBC counting performances compared with the state-of-the-art. Di Ruberto’s work [25] was tested on the entire ALL-IDB. The column for this work indicates the results obtained on the first 33 images.

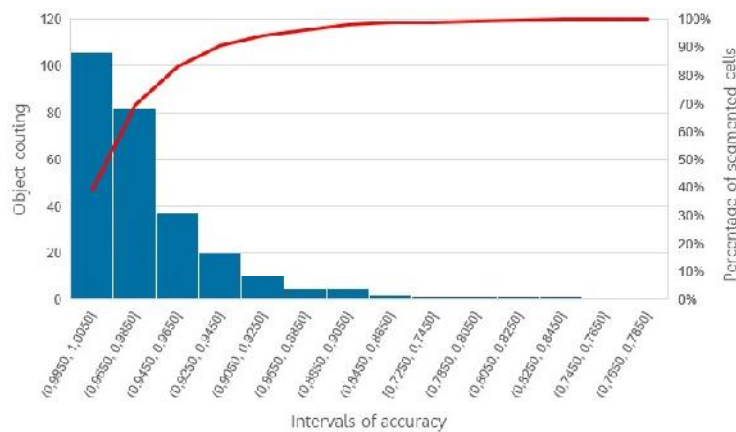
	Mahmood [7]	Alilou [9]	Putzu [6]	Di Ruberto [22]	Bhavnani [23]	Loddo [24]	Di Ruberto [25]	Our Method
N° images	10	Unknown	33	33	14	33	33	33
Accuracy	81%	-	-	99.2%	94.2%	-	98.0%	99.7%
Precision	-	94%	-	100%	-	89.0%	98.7%	100%
Recall	-	88%	92%	99.2%	-	98.0%	99.3%	99.6%
F-measure	-	-	-	99.6%	-	93.0%	98.9%	99.8%

To further assess the performance of the proposed segmentation algorithm, we computed the pixel-wise accuracy by comparing our results with the ground truth images. Some interesting results are presented in Figures 10 and 11. In particular, the first one was subdivided into intervals spaced 0.25% and showed that more than 90% of the cells were segmented with a pixel-wise accuracy higher than 95%. Moreover, it must be noted that 256 objects (out of 271) were segmented with an accuracy higher than 90%. While Figure 10 analyses in detail the accuracy of the objects, Figure 11 focuses on the segmented images. It shows the pixel-wise accuracy for each one and a red line that indicates the average accuracy of 97.02%. However, it is essential to specify why some images had a lower accuracy than the average. For example, in Image Numbers 9 and 30, which are depicted in Figure 12, the segmentation of the leukocytes inside the red rectangle was harder due to the presence of a light WBC’s cytoplasm. Due to this, the method was not accurate in recognizing all the pixels of the cell, even though the overall accuracy remained higher than 90% for both cases. The proposed approach also improved the health status classification results; in particular, it obtained an accuracy of 94.1%, improving Putzu’s [6] results, which obtained 93.2% for the same task in the same environment setting.

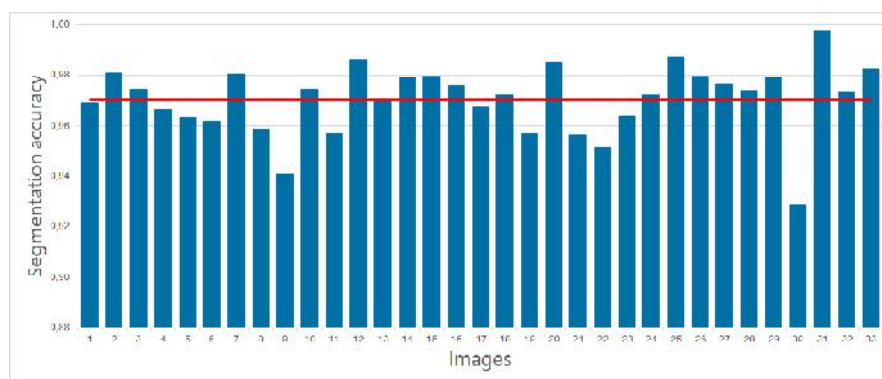
We also performed two further detection experimentations on IUMS-IDB and SMC-IDB depicted in Section 5, datasets composed of images with very different characteristics from ALL-IDB. The final result for WBC detection was 100% in both cases and matched the performances achieved by Di Ruberto [22]. Visual results are reported in Figure 13.

It is crucial to notice that our approach was dependent only on three key parameters, which were the interval of the hue space, the blob dimension, and the value of  $t$ , defined in Equation (3). We searched for blue objects; therefore, we tuned the range of hue accordingly, acting as a filter on the saturation channel image, and the same for every experiment. More precisely, in the [0–1] interval of the hue channel, where blue can be found around 2/3 of it, we selected the [0.65–0.9] range. By choosing a larger interval, the detection results could include artifacts (see Figure 14), to be removed in the next steps of the system, increasing the computational workload. The blob dimension, strictly related to WBC’s size, was dependent on the acquisition condition of the images contained in the datasets, and one of them was sufficient to obtain the corresponding value. The last one was the  $t$  parameter. It was chosen to remove the unnecessary local maxima, to reduce or avoid

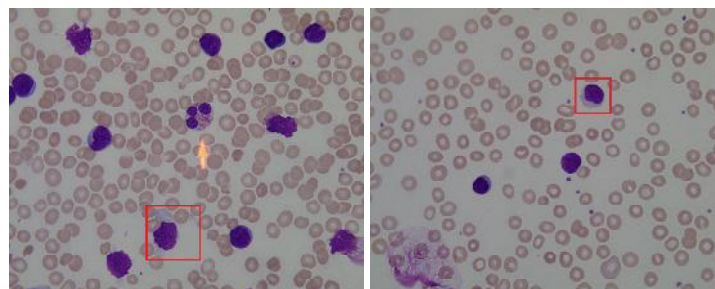
any false positive. Figure 15 describes the detection performances by varying the value of  $t$ , defined in Equation (3), on the IUMS-IDB and the SMC-IDB datasets. This showed that the detection results were close (or equal) to 100% if the value of  $t$  was chosen in the range [0.4–0.6], avoiding or reducing the detection of false positives. In general, the detection performances were high even in a wider range of  $t$ , but were the best in the interval mentioned above. In conclusion, we can state that our method could be generalized to datasets with different illumination conditions and resolutions, taking into account the characteristics of the parameters, easily obtainable from a sample image of each dataset.



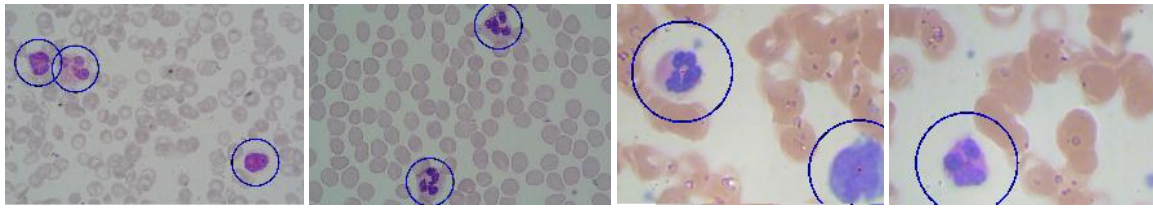
**Figure 10.** Pixel-wise accuracy of cell segmentation with the proposed method. Objects accuracy is subdivided into intervals. The Pareto chart shows that our method offers high pixel-wise accuracy for the majority of the objects and that more than 90% of the total cells are segmented with a pixel-wise accuracy of more than 95%.



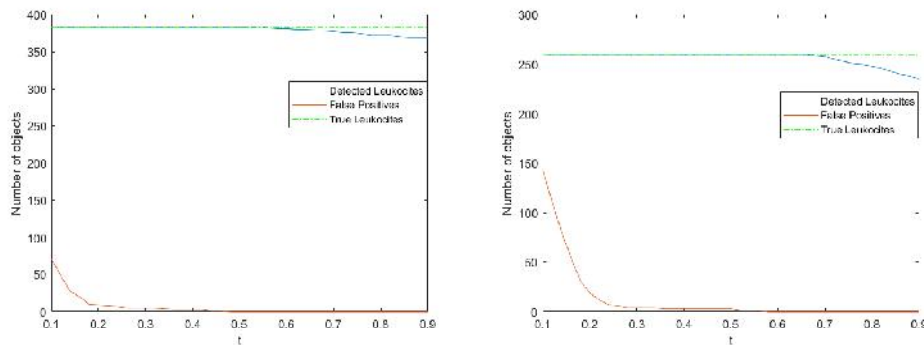
**Figure 11.** Pixel-wise segmentation accuracy for every tested image. The red line indicates an average accuracy of 97.02% on the first 33 images.



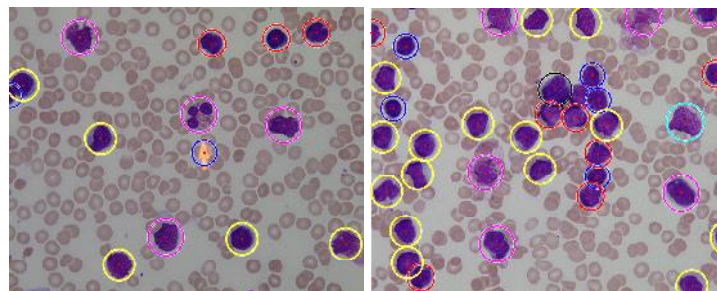
**Figure 12.** Example of images with a pixel-wise segmentation accuracy lower than the average. The reason is related to the presence of a light cytoplasm, which makes segmentation harder than the nucleus region. From left to right: Image Numbers 9 and 30.



**Figure 13.** WBC detection results on two images extracted from IUMS-IDB (left) and two from SMC-IDB (right).



**Figure 14.** WBCs detection performances on SMC-IDB and IUMS-IDB, respectively. They are close (or equal) to 100% if  $t$  value lies in the range [0.4–0.6].



**Figure 15.** Example of the application with H channel values in the range [0–1]. On the left, the algorithm selects the yellow arrow as a blob, while on the right, we show a complex situation in which a WBC inside the cells agglomerate is not selected as a valid blob.

## 6. Conclusions

This paper proposed a novel and efficient system for leukocyte counting and classification, providing an automated procedure to recognize the ALL, trying to overcome the typical human operator's issues, subjective opinion, or tiredness due to lengthy and repetitive tasks. Most of the works available in literature presented only methods for segmentation (e.g., [2,4,5,7–9]) or classification tasks separately. Only a few proposed a complete system to overcome both detection and classification problems (e.g., [10,11,13]), even though the authors tested their proposed system on local datasets ([10,11]) or on hybrid datasets [13]. Consequently, it is not possible to reproduce the same experiments of the works mentioned above, in order to make a useful performance comparison. Another novel aspect of our system was related to the detection of single cells from agglomerates without introducing any computational workload. This is a very difficult problem, often not considered or addressed by specific steps [6]. We firstly evaluated the performances of the proposed approach on the first 33 images of ALL-IDB, a public dataset for leukemia detection. The test was performed on the first 33 images because they were acquired with the same camera and presented the same illumination conditions. In particular, the results showed that it correctly detected 99.7% of the WBCs present in the first 33 images of ALL-IDB. Moreover, the system correctly identified 94.1% of the lymphoblasts, in the classification step. In both cases, the system outperformed the state-of-the-art methods based

on computer vision techniques. WBC detection performances were also evaluated on IUMS-IDB and SMC-IDB. The accuracy results on both datasets reached 100%, matching the performances achieved by Putzu [22]. Further steps will be devoted to the development of the identification phase to increase the overall accuracy of the system. It is essential to study and analyze the use of new features that may be discriminant for this analysis. Moreover, the analysis of all images included in MP-IDB [26], both for cell and parasite detection and for quantifying the parasitemia level, is another target for this research. As a future direction, a supervised approach to classify the detected cells can be introduced to provide the number of cell subtypes and the number of abnormal cells. In conclusion, our method proposed in this study is a very promising technique to identify ALL.

**Author Contributions:** The work described in this article is the collaborative development of all the authors. All authors have read and agreed to the published version of the manuscript.

**Funding:** The Regione Autonoma della Sardegna partially supported the research in this paper with the research project “Algorithms and Models for Imaging Science [AMIS]” (finanziato con risorse FSC 2014–2020—Patto per lo Sviluppo della Regione Sardegna).

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

WBC	White Blood Cell
RBC	Red Blood Cell
ALL-IDB	Acute Lymphoblastic Leukemia-Image Data Base
SMC-IDB	System, Man and Cybernetics-Image Data Base
IUMS-IDB	Iran University of Medical Science-Image Data Base
MP-IDB	Malaria Parasite Image Data Base
TP	True Positive
TN	True Negative
FN	False Negative

## References

1. Erhabor, O.; Adias, T.C. *Hematology Made Easy*; AuthorHouse: Bloomington, IN, USA, 2013.
2. Madhloom, H.T.; Kareem, S.A.; Ariffin, H.; Zaidan, A.A.; Alanazi, H.O.; Zaidan, B.B. An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold. *J. Appl. Sci.* **2010**, *10*, 959–966. [[CrossRef](#)]
3. Sinha, N.; Ramakrishnan, A.G. Automation of differential blood count. *TENCON Conf. Converg. Technol. Asia-Pac. Reg.* **2003**, *2*, 547–551.
4. Di Ruberto, C.; Putzu, L. Accurate Blood Cells Segmentation through Intuitionistic Fuzzy Set Threshold. In Proceedings of the International Conference SITIS on Signal-Image Technology and Internet-Based Systems, Marrakech, Morocco, 23–27 November 2014; pp. 57–64.
5. Khan, S.; Khan, A.; Khattak, F.S.; Naseem, A. An accurate and cost effective approach to blood cell count. *Int. J. Comput. Appl.* **2012**, *50*, 975–8887. [[CrossRef](#)]
6. Putzu, L.; Caocci, G.; Di Ruberto, C. Leucocyte classification for leukemia detection using image processing techniques. *AIM* **2014**, *62*, 179–191. [[PubMed](#)]
7. Mahmood, N.H.; Lim, P.C.; Mazalan, S.M.; Razak, M.A.A. Blood cells extraction using color based segmentation technique. *Int. J. Life Sci. Biotechnol. Pharma Res.* **2013**, *2*, 233–240.
8. Alomari, Y.M.; Sheikh Abdullah, S.N.H.; Zaharatul Azma, R.; Omar, K. Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm. *Comput. Math. Methods Med.* **2014**, *2014*, 979302. [[CrossRef](#)] [[PubMed](#)]
9. Alilou, M.; Kovalev, V. Automatic object detection and segmentation of the histocytology images using reshaping agents. *Image Anal. Stereol.* **2013**, *32*, 89–99. [[CrossRef](#)]

10. Mohapatra, S.; Patra, D.; Satpathy, S. An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images. *Neural Comput. Appl.* **2014**, *24*, 1887–1904. [[CrossRef](#)]
11. Vincent, I.; Kwon, K.; Lee, S.; Moon, K. Acute lymphoid leukemia classification using two-step neural network classifier. In Proceedings of the 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), Mokpo, Korea, 28–30 January 2015; pp. 1–4. [[CrossRef](#)]
12. Singh, G.; Bathla, G.; Kaur, S. Design of new architecture to detect leukemia cancer from medical images. *Int. J. Appl. Eng. Res.* **2016**, *11*, 7087–7094.
13. Vogado, L.H.; Veras, R.M.; Araujo, F.H.; Silva, R.R.; Aires, K.R. Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. *Eng. Appl. Artif. Intell.* **2018**, *72*, 415–422. [[CrossRef](#)]
14. Lindeberg, T. Feature detection with automatic scale selection. *Int. J. Comput. Vis.* **1998**, *30*, 77–116.
15. Meyer, F. Topographic distance and watershed lines. *Signal Process.* **1994**, *38*, 113–125. [[CrossRef](#)]
16. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**. [[CrossRef](#)]
18. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.
19. Labati, R.D.; Piuri, V.; Scotti, F. All-IDB: The acute lymphoblastic leukemia image database for image processing. In Proceedings of the 2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, 11–14 September 2011; pp. 2045–2048.
20. Mohamed, M.; Far, B.; Guaily, A. An efficient technique for white blood cells nuclei automatic segmentation. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea, 14–17 October 2012; pp. 220–225.
21. Sarrafzadeh, O.; Rabbani, H.; Talebi, A.; Banaem, H.U. Selection of the best features for leukocytes classification in blood smear microscopic images. *Proc. SPIE* **2014**, *9041*, 8.
22. Di Ruberto, C.; Loddo, A.; Putzu, L. A leucocytes count system from blood smear images. *Mach. Vis. Appl.* **2016**, 1–10. [[CrossRef](#)]
23. Bhavnani, L.A.; Jaliya, U.K.; Joshi, M.J. Segmentation and Counting of WBCs and RBCs from Microscopic Blood Sample Images. *Int. J. Image Graph. Signal Process.* **2016**, *8*, 32–40. [[CrossRef](#)]
24. Loddo, A.; Putzu, L.; Di Ruberto, C.; Fenu, G. A Computer-Aided System for Differential Count from Peripheral Blood Cell Images. In Proceedings of the International Conference on Signal-Image Technology Internet-Based Systems (SITIS), Naples, Italy, 28 November–1 December 2016; pp. 112–118.
25. Di Ruberto, C.; Loddo, A.; Putzu, L. Detection of red and white blood cells from microscopic blood images using a region proposal approach. *Comput. Biol. Med.* **2020**, *16*, 103530. [[CrossRef](#)] [[PubMed](#)]
26. Loddo, A.; Di Ruberto, C.; Kocher, M.; Prod'Hom, G. MP-IDB: The Malaria Parasite Image Database for Image Processing and Analysis. In *Processing and Analysis of Biomedical Information, SaMBa 2018*; Springer: Cham, Switzerland, 2019; Volume 11379, pp. 57–65.

