

# Using Neural Networks and Hough Transform for Leukocytes Differentiation in Blood Count Images

Yuri Marchetti Tavares[1], Nadia Nedjah[2(✉)], and Luiza de Macedo Mourelle[3]

[1] Naval Weapon Systems Department, Brazilian Navy, Brasilia, Brazil
[2] Department of Electronics Engineering and Telecommunications, Engineering Faculty, State University of Rio de Janeiro, Rio de Janeiro, Brazil
nadia@eng.uerj.br
[3] Department of System Engineering and Computation, Engineering Faculty, State University of Rio de Janeiro, Rio de Janeiro, Brazil

**Abstract.** Automated equipments for blood count have facilitated the clinical routines in medical laboratories. However, when grave anomalies are identified, some samples still need to be analyzed in a traditional way using a microscope. This paper proposes a method to classify the five most common types of leukocytes, present in the human blood using blood smear images. This allows reducing the amount of work, the time spent in the blood test as well as cost of such a test. The results are satisfactory and should contribute positively to the development of an automatic system and/or a smart electronic device, which would come handy for this kind of applications.

## 1 Introduction

The blood count is a medical routine for blood test to evaluate peripherical blood. It is one of the most requested blood test, and it is used for obtaining diagnostics, clinical evaluations and the evolution process of several diseases [1]. Basically, the blood count provides qualitative and quantitative informations about the blood cells.

The Practical Manual of Hematology [2] states that the human blood consists of plasma (55%) and cells (45%). The formed elements, which are made up of the blood cells and platelets, also called cellular portion presents three types of cells in suspension in the plasm: red corpuscles (erythrocytes), white corpuscles (leukocytes) and platelets. The white corpuscles have also other subdivisions: lymphocytes, basophils, monocytes, eosinophils and neutrophils.

Initially, the blood tests were done through microscopes, where the erythrocytes, leukocytes and platelets are observed and counted using a Neubauer's camera and colored glass slide. With electronics development and automation, several equipments were designed to provide these informations in a fast and precise way.

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11620, pp. 643–653, 2019.
https://doi.org/10.1007/978-3-030-24296-1_52

The automate analysis has facilitated the performance of the routine in clinics laboratories. The equipments usually allow analysis of 30 to 120 blood counts per hour. The most simple appliances have the impedance principle as reference, with the change of electric current between two electrodes. When a cell transverse the system, there is a change in this impedance and a electrical impulse is generated, according to the diameter of each cell (erythrocytes, leukocytes or platelets). These equipments allow present clinics laboratories to deal with high volume of samples received daily.

However, some samples analyzed by the automate cell counters still require manual evaluation of the blood smear for anomalies observation. Every time a blood test result has to be confirmed, the technician prepares the blood smear of the patient and analyses it with a microscope. This procedure is precise, but also very slow and demands a laborious process. A specific amount of blood is required to put on a glass slide, to color it and to wait until it dries. Afterwards, the technician must check the glass slide and count cell after cell with the microscope. On the other hand, the informations obtained through images and videos have become an important area of research, especially with the development and refinement of sensors and intelligent equipments capable of capturing, storing, editing and transmitting images.

This paper proposes the development of a method that can automate the process of identification and counting of blood cells, that contributes to the decrease of work, of the time to provide the analysis result and even of the cost for the blood exam in hematology laboratories. The objective is to develop a classifier, based on artificial neural networks, capable of identifying blood cells and differ the leukocytes in exams performed with blood on glass slides. The classifier uses the Hough transform, mathematical morphology and segmentation, also based on neural networks, to automate the process of cell counting.

Section 2 presents some related works. First, in Sect. 3 some blood cell components are introduced. Then, in Sect. 4, we describe the methodology and the structure of the proposed method, including segmentation and characteristics extractions. Subsequently, in Sect. 5, we present some results. Finally, in Sect. 6, we draw some conclusions.

## 2 Related Work

In [3], the authors use the concept of Hough transform showing a methodology for segmenting and counting red cells in images of blood cells. Only the erythrocytes were considered, taking into account their size and shape.

In [4], a system based on neural networks is proposed for classifying blood cells. The aim of the work is the extraction of characteristics only for classifying the three main groups: red cells, white cells and platelets. Aspects of detection and segmentation of cells are not considered, as well as the differentiation of leukocytes.

In [5], a method for segmentation, identification and classification of white cells is proposed, based on the equalization of histogram, thresholding and edge detection. However, the proposed method classifies only three of five subclasses of leukocytes: neutrophils, lymphocytes and basophils.

The present work extends the concept of Hough transform to leukocytes too. One segmentation, based on neural networks, is proposed in order to separate the images in four distinct classes: erythrocytes, cytoplasm, nucleus and background. As part of the segmentation, mathematical morphology and edge detection are used to separate the leukocytes from the rest of the image. These leukocytes are differentiated in five classes (neutrophils, lymphocytes, monocytes, eosinophils and basophils) through the use of another neural network.

## 3 Blood Cells

The blood consists of a fluid, in which there are cells in suspension, molecules and ions dissolved in water. An important characteristic of the blood is its steady chemical composition and physical properties, assuring physical conditions for the cell function. The cell part consists of 45% of the whole volume.

To perform the exams with microscope, the blood is spread on a glass slide and, then, colored. Coloring is an important step, since it facilitates the microscopic observation of the cells, differentiating them according to their characteristics. Typical dyes used are: methyl blue, which gives a blue color to DNA, RNA, nucleoproteins, basophil granules and, weakly, eosinophil granules, and violet eusinate, which gives an orange color to hemoglobin [1].

Erythrocytes, or red globules, are small circular cells with 7.5 mm of diameter. They do not have a nucleus, but only plasmatic membrane and cytoplasm. They are the most numerous cells of the blood, with 4.5 to 6.5 millions of cells per $mm^3$. The normal erythrocytes, cells number 1 in Fig. 1, have the shape of a biconcave disk.

Platelets are small cells in the shape of a disk, with a diameter of 2 to 3 µm, corresponding to cells number 2 in Fig. 1. They are formed of fragments of cytoplasm of the megakaryocyte, original cell, and are responsible for hemostasis and blood clotting.

Leukocytes, or white blood cells, form the most heterogeneous group of blood cells, both from the morphological and physiological point of view. Although they perform the role of organism defense, each subtype holds very specific and distinct functions among themselves, which together, define the immune system [1]. The leukocytes are present in the blood in a very lower number than the erythrocytes, with around 4000 to 10000 leukocytes per $mm^3$ of blood.

Neutrophils are the leukocytes found in higher quantity in the human blood and are known for presenting a neutral tonality in blood coloring. They have small cytoplasmatic granules, weakly colored in purple-reddish. The mature neutrophil, cell number 3 in Fig. 1, have 12 to 15 µm of diameter. The core is divided in 2 to 5 distinct lobes. The precursor cell of the segmented neutrophil is a band neutrophil. They are so called because their core does not present segmentation.

Eosinophils are slightly bigger than neutrophils, measuring from 12 to 17 µm. The cores have usually two lobes, connected through on filament. The cytoplasm is covered by spherical granules, that are colored in orange-reddish, corresponding to cell number 5 in Fig. 1.

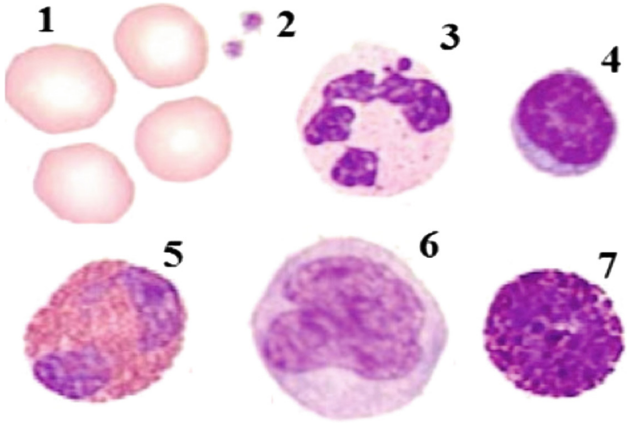



**Fig. 1.** Different kind of blood Cells (Color figure online)

Basophils have almost the same size as neutrophils, measuring from 10 to 14 μm of diameter. Their central core have two or three lobes, that are usually obscured by cytoplamatic granules colored in black-blueish, as shown by cell number 7 in Fig. 1.

Monocytes are the biggest leukocytes, measuring from 12 to 20 μm of diameter. They are cells of irregular form, with a big and unic core, centrally located. The cytoplasm have a blue-grayish color, as shown by cell number 6 in Fig. 1.

Lynfocytes, corresponding to cell number 4 in Fig. 1, are the second most abundant of leukocytes in the circulation. The are small mononucleus cells, without granules in the cytoplasm, slightly bigger than erythrocytes, measuring from 10 to 16 μm of diameter. Their core is big and homogeneous, showing an intense blueish color.

## 4 Methodology

For the execution of the proposed method, the procedure is divided into five steps: image acquisition, preprocessing, characteristics extraction, Differential Artificial Neural Network (DANN) training and DANN utilization. The evaluation of the proposed method is also planned.

For the acquisition step, images available from Hemosurf [6] were used as data base. They consist of blood cells obtained from a Zeiss Axioskop microscope and Sony DXC-3000A 3CCD video camera, freely available by Bern University. For this work, images with 1000x zoom were used from the set "Normal White Blood Cells", the set "Examine and Compare Normal Leukocytes", the normal leukocytes set "Blood Film 4" and the basophils and eosinophils set "Chronic Myelocytic Leukemia".

The preprocessing step includes the identification and segmentation of the leukocytes from the acquired images of the previous step. After this step, smaller

images were obtained, having isolatedly each white globule. These images were separated by class, so that 70% of the segmented images were used for the DANN and 30% were used for the generalization evaluation.

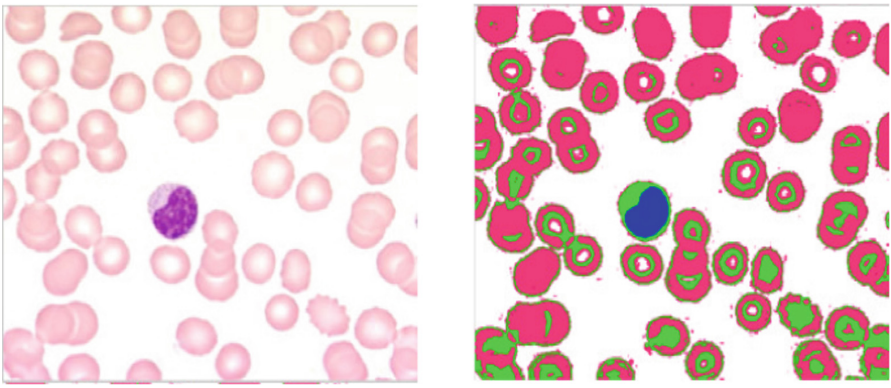
The next step is the attributes extraction of each isolated white globule. These attributes are inserted as inputs of the DANN.

With the training data base having the characteristics of each cell, the DANN can be trained. Once trained, this net can be used for classifying the five classes of interest, implementing the differentiation of the leukocytes. For the evaluation of the model and of the DANN, a test of the generalization is implemented, with those 30% of the data separated in the preprocessing step.

### 4.1 Preprocessing

Image segmentation is the main part of the preprocessing step, where leukocytes are identified and separated. For this function, another neural network is used, called Segmentation Artificial Neural Network (SANN). This neural network is of type Multilayer Perceptron, with a hidden layer, a hyperbolic tangent function as transfer function of the hidden layer and a linear function as output transfer function. Cross validation i used with anticipated stop and the error is estimated using the quadratic medium error.

Pixels of the four areas, that we wished to separate, were used as input data of the RANN: red blood cells, leukocytes core and cytoplasm, and background. RGB values of each pixel were considered as inputs. In order to form the training set, small samples of each segmentation class in two images of the data base were selected. Figure 2 shows the segmentation result obtained for one of the images. The pink elements refer to the red blood cells, the green elements are the cytoplasm, the blue elements are the core and the white areas are the background. This is the best achieved result. Three neurons in the hidden layer were used. In order to deal with the fact that part of some red blood cells were wrongly



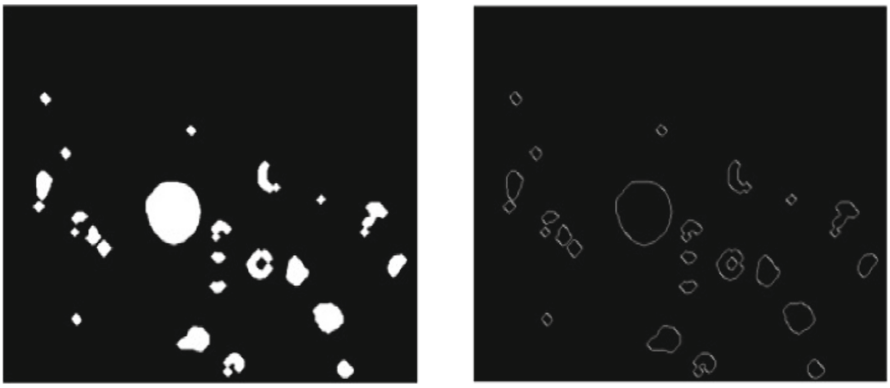
(a) Original image (b) ANN Segmented image

**Fig. 2.** Blood cells image before and after segmentation (Color figure online)

marked as cytoplasm, mathematical morphology is inserted into the preprocessing step.

Mathematical morphology consisted, first, of opening and closing operations with a disk type morphological element of $3 \times 3$. The aim is thus to fully fill the leukocyte cytoplasm. Afterwards, erosion and dilation operations were implemented to eliminate the red blood cells marked as cytoplasm. The result obtained for Fig. 2 is shown in Fig. 3.

At this processing step, the only thing missing is the identification of the leukocyte. For this function, the Hough transform is used. This transform was proposed by Hough [7] for detecting subatomic particles path, popularized by Rosenfeld [8] and largely used for detecting geometric forms, such as lines, circles and ellipses [9]. This method converts the problem of curve detection in an efficient search for peaks in the parameters space.



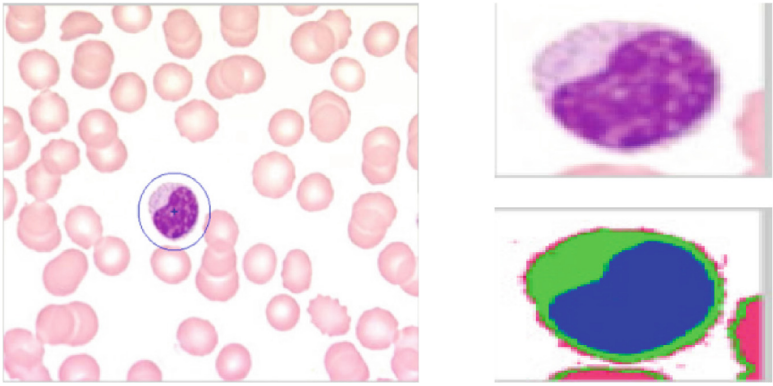
(a) Using morphology (b) Using contours

**Fig. 3.** Blood cells image of Fig. 2 digitalized using morphology *vs.* contours

The Hough transform fits perfectly in the detection of the circle center, which represents the leukocyte in the digitalized image. The search parameters were adjusted such that only circles of diameter very close to the leukocytes diameter were found. We also observed that the Hough transform is more efficient when only the objects perimeter were considered. Therefore, before implementing the search, the image is submitted to the Sobel filter for board detection. The resulting image can be seen in Fig. 3. Finally, the leukocyte is extracted independently from the rest of the image, as shown in Fig. 4.

### 4.2 Characteristics Extraction

The chosen characteristics for the DANN input were selected such that they were able to separate the classes in the best and clearer way possible. The size of the cells, the size of the core and the rate of the red color in the cytoplasm,



(a) Original image (b) Leukocyte

**Fig. 4.** Leukocyte found in the blood cells image of Fig. 2

as indicated by rate $R$ were considered. For each of the chosen cells, these characteristics were calculated using Eqs. (1), (2) and (3):

$$T_T = P_N + P_C \tag{1}$$

$$T_N = \frac{P_N}{T_T} \tag{2}$$

$$R = \frac{\sum R_i}{P_C}, \tag{3}$$

where $T_T$ is the total size of the cell, $P_N$ is the number of pixels corresponding to the core, $P_C$ is the number of pixels corresponding to the cytoplasm, $T_N$ is the size of the cell core, $R$ is the red rate of the cytoplasm and $R_i$ is the value of pixel $i$ red component of the cytoplasm. We observe that the segmentation precision of SANN is of fundamental importance for the calculation of these parameters.

Sometimes, besides the leukocytes, some of the red blood cells are also selected. In order to eliminate these undesirable data, a threshold is defined, when calculating the characteristics: if the size of the core is less than 25% of the cell total size, the datum is eliminated. This way, we can guarantee that only the leukocytes that have core are considered in the construction of the data base.

Figure 5 shows the leukocytes distribution, used in the DANN training, in the characteristics space. The red circles represent the neutrophils, the blue circles are the lymphocytes and the gray circles are the monocytes. The black crosses represent the basophils and the magenta crosses are the eosinophils.

### 4.3 Differentiation Artificial Neural Network

In order to classify the five classes of leukocytes, an artificial neural network of type Multilayer Perceptron is used, considered an universal approximator [10]. The used topology consists of an input layer, with each input related to

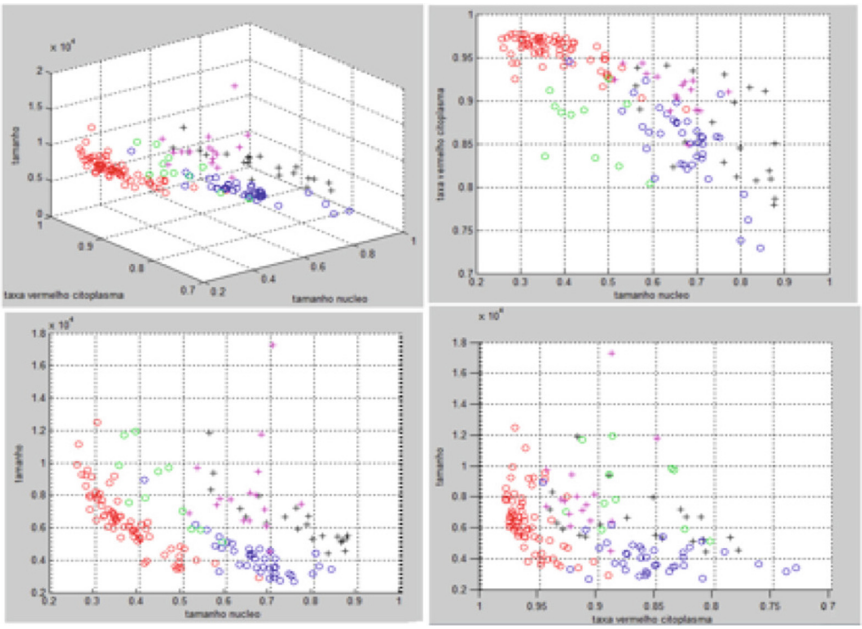



**Fig. 5.** Representation of the training leukocytes in the feature space (Color figure online)

each characteristic defined in the previous section, and one output, with linear function. Only one hidden layer is used, with hyperbolic tangent as activation function.

During the supervised training, with crossed validation and anticipated stop, the network parameters were adjusted so that its output assumed the pre-established value for each class, according to Table 1. Limited values were established at the output, so that the classificator output showed only discrete values between 1 and 5.

## 5 Results

In order to implement all the proposed procedure, tha MATLAB 8.3.0532 (R2014a) software is used, running on an ASUS computer, with Intel Core i7-2630 QM 2GHz, 6GB RAM and 64 bits Windows 7 Home Premium operating system. Images processing and artificial neural networks were implemented using the MATLAB toolbox. A independent script is used for the Hough transform. The DANN is then trained with data from the preprocessing step, using the medium squared error as stop criteria, with crossed validation (30%) and anticipated stop.

The best network is identified empirically and has eight neurons in the hidden layer. The training is repeated for 15 times, resulting in a medium squared error

**Table 1.** Expected values of the DANN output

| Blood cell | Class | DANN output |
|---|---|---|
| Neutrophil | 1 | [0.5, 1.5 ] |
| Lymphocytes | 2 | [1.5, 2.5 ] |
| Monocyte | 3 | [2.5, 3.5 ] |
| Eosinophil | 4 | [3.5, 4.5 ] |
| Basophil | 5 | [4.5, 5.5 ] |

of 16.18%. After the training, the network is used for classifying the training data and the confusion matrix showed in Table 2 shows the collected results. We can observe that the network has learned well how to classify the neutrophils with a hit rate of 100% and that the hit rate of the other classes is higher than 54%.

**Table 2.** Confusion matrix obtained with the training data

| Found class | Real class | | | | |
|---|---|---|---|---|---|
| | Neutrophil | Lymphocytes | Monocyte | Eosinophil | Bashophil |
| Neutrophil | 66 | 2 | 0 | 0 | 0 |
| Lymphocytes | 0 | 29 | 4 | 0 | 0 |
| Monocyte | 0 | 6 | 6 | 2 | 0 |
| Eosinophil | 0 | 1 | 1 | 12 | 0 |
| Basophil | 0 | 0 | 0 | 2 | 12 |

In order to evaluate the generalization of DANN and validate the method, the test set is used. The overall error is 25.35%. To evaluate the error rate, the false positive and the false negative rates for each class, the confusion matrix of Tables 3 and 4 are exploited.

**Table 3.** Confusion matrix obtained during generalization

| Found class | Real class | | | | |
|---|---|---|---|---|---|
| | Neutrophil | Lymphocytes | Monocyte | Eosinophil | Bashophil |
| Neutrophil | 30 | 1 | 1 | 0 | 0 |
| Lymphocytes | 2 | 11 | 1 | 1 | 0 |
| Monocyte | 0 | 4 | 4 | 1 | 1 |
| Eosinophil | 0 | 1 | 0 | 4 | 3 |
| Basophil | 0 | 0 | 0 | 2 | 4 |

**Table 4.** Hit rate, false positive rate and false negative rate for each considered class

| | Neutrophil | Linfocyte | Monocyte | Eosinophil | Bashophil |
|---|---|---|---|---|---|
| Hit | 93,75% | 64,7% | 66,7% | 50% | 50% |
| False positive | 6,25% | 26,7% | 60% | 50% | 33,33% |
| False negative | 6,25% | 35,29% | 33,3% | 50% | 50% |

We can observe that the network displays the same pattern shown during the training, proving the generalization. Again, the neutrophils hit rate is very high. These results are due to the fact that the neutrophils are the kind of leukocytes that are most abundant in the blood, presenting greater number of data.

It is noteworthy to point out that the result is also influenced by the low quantity of data available. Compared to the red blood cells and platelets, the amount of leukocytes present per $mm^3$ is very low. This fact hinders the generation of a rich and consistent data base. In order to train the network correctly, it is necessary to have statistical information about the samples. For instance, in the case of the monocytes, the network exploited 11 samples and the generalization test is implemented with 8 samples only. It is noteworthy to emphasize that the difference in contrast and brightness during the images acquisition can influence the results, if there is not enough data.

## 6 Conclusion

The proposed method was considered promising, showing a satisfactory result and generalization capacity. We can also say that it is applicable to real world systems, in which the resolution conditions, size and brightness of the microscope images can be controlled, providing even also the increase in the results performance.

For future work, we foresee the inclusion of characteristics related to the shape of the blood cell to provide a differentiation of stick neutrophils and segmented neutrophils, and also the identification of abnormal cells. In this last case, it is of great importance when elaborating the evaluation manual of blood smears.

**Acknowledgement.** We are grateful to the State of Rio de Janeiro funding agency FAPERJ (http://www.faperj.br) for supporting this research.

## References

1. Monteiro, F.G.: Comparação dos Resultados de Hemogramas do Contador Eletrónico ABX Pentra 60 com a Microscopia, M.Sc. Dissertation, Federal University of Rio Grande do Sul, Brazil (2005)
2. Vivas, W.L.P.: Manual Prático de Hematologia. http://docente.ifsc.edu.br/rosane.aquino/MaterialDidatico/AnalisesClinicas/hemato/Manual%20de%20Hematologia.pdf. Accessed June 2019

3. Maitra, M., Gupta, R.K., Mukherjee, M.: Detection and couting of red blood cells in blood cell images using hough transform. Int. J. Comput. Appl. **53**(16) (2007)
4. Khashman, A.: Blood cell identification using a simple neural network. Int. J. Neural Syst. **18**(5), 453–458 (2008)
5. Hiremath, P.S., Bannigidad, P., Geeta, S.: Automated identification and classification of white blood cells (Leukocytes) in digital microscopic images. IJCA Special Issue Recent Trends Image Process. Pattern Recogn. (2010)
6. Hemosurf (2014). http://hemosurf.ehb.be/Data/Data_E/wbc.htm
7. Hough, P.V.C.: A Method and Means for Recognizing Complex Patterns, U.S. Patent 3,069,654, December (1962)
8. Rosenfeld, A.: Picture processing by computer. ACM Comput. Surv. **1**, 147–176 (1969)
9. Mukhopadhyay, P., Chaudhuri, B.B.: A survey of hough transform. Pattern Recogn. **48**, 87–116 (2015)
10. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Networks **2**, 359–366 (1989)