

Methodology of High Accuracy, Sensitivity and Specificity in the Counts of Erythrocytes and Leukocytes in Blood Smear Images



Ana Carolina Borges Monteiro , Yuzo Iano , Reinaldo Padilha França 
and Rangel Arthur 

Abstract The evaluation of human blood is an important diagnostic method for the detection of diseases. The analysis of the erythrocytes contained in the blood contributes to the detection of anemia and leukemia, whereas the leukocyte analysis allows the diagnosis of inflammation and/or infections. The blood is analyzed through of the complete blood count test (CBC), which is dependent on automated and/or manual methodologies. The dependence of medical areas on new technologies leads the present study to the goal of developing an image segmentation algorithm that meets the criteria of efficiency and reliability for detection and counting of blood cells. The algorithm was developed through the Matlab software, being the image processing methodology based on the union of the Watershed transform and Morphological Operations, originating the WT-MO methodology. For the simulations, 30 blood smear images containing erythrocytes and leukocytes were used in a non-pathological state. The results showed that the WT-MO methodology presents high sensitivity (99%), specificity (96%) and accuracy (98,3%) when compared with the manual methodology. Therefore, the WT-MO methodology is an accurate, reliable and low-cost technique and can be applied as a third more accessible methodology to perform of the complete blood count test (CBC) in populations of underdeveloped and developing countries.

Keywords Blood cells · Watershed transform · Morphological operations · Blood smear · Image processing · Algorithm · ROC curve

A. C. B. Monteiro (✉) · Y. Iano · R. P. França · R. Arthur
School of Electrical and Computer Engineering (FEEC), University of Campinas – UNICAMP,
Av. Albert Einstein – 400, Barão Geraldo, Campinas, SP, Brazil
e-mail: monteiro@decom.fee.unicamp.com

Y. Iano
e-mail: yuzo@decom.fee.unicamp.com

R. P. França
e-mail: padilha@decom.fee.unicamp.com

R. Arthur
e-mail: rangel@ft.unicamp.br

© Springer Nature Switzerland AG 2019
Y. Iano et al. (eds.), *Proceedings of the 4th Brazilian Technology Symposium (BTSym'18)*, Smart Innovation, Systems and Technologies 140,
https://doi.org/10.1007/978-3-030-16053-1_8

1 Introduction

During the embryonic stage, the yolk sac is responsible for the formation of the first blood cells, later during fetal development, this function is transferred to the liver and spleen. During adulthood, the interior of the bone marrow is composed of a hematopoietic tissue, being present in the long bones and in the axial skeleton. Bone marrow presents a suitable microenvironment for the development of hematopoietic cells and proliferation of primitive cells and progenitor cells [1, 2]. Hematopoiesis is the process responsible for the formation of blood cells inside the bone marrow, from a cell mother called stem cell. This pluripotent cell originates from the different blood cell lines: erythrocytes, leukocytes and platelets. This phenomenon is dependent on a series of actions and responses that lead to the processes of duplication, differentiation and maturation, resulting in the production and release of mature cells into the bloodstream [3, 4].

As a result, human blood consists of approximately 55% plasma and 45% blood cells, with their total volume represented by approximately 7% of the individual's body weight. Its function consists in the conduction of gases and nutrients to the tissues, aid in the excretion of metabolites, hormonal distribution, passage of chemical messages between distant organs, regulation and maintenance of body temperature, basic acid balance and osmotic equilibrium [5]. Plasma is an aqueous solution formed by molecules of different molecular weights, being responsible for the transport of substances such as water, plasma proteins, inorganic salts, amino acids, hormones, glucose, albumin, immunoglobulins (IG), components of the blood coagulation cascade, among others. For its part, the blood cells are responsible for the homeostasis of the organism, transport of gases, defense of the organism against distinct etiological agents, tissue repair and blood coagulation. However, such observations could only occur through the invention of instruments capable of visualizing the cells [6].

The erythrocytes are also called red blood cells. They are anucleated, biconcave disc-shaped cells, formed by a tetramer (two α chains and two β chains) responsible for the transport of oxygen through four iron molecules [7]. When erythrocytes reach the lungs, oxygen molecules are attached to the iron molecule, generating oxyhemoglobin. This binding is disrupted only when the erythrocytes reach the tissues, where the oxygen pressure is lower. In tissues, erythrocytes deposit oxygen and withdraw carbon dioxide through carbo-hemoglobin binding. Thus, carbon dioxide is either taken directly to the lungs or is dissolved in plasma [3].

Leukocytes are the defense cells of the body, constituting the innate immune response and the adaptive immune response. They are also called white cells, are classified into 2 distinct classes: leukocytes granulocytes and leukocytes agranulocytes. Granulocytic leukocytes are the defense cells that have granules dispersed in the cytoplasm, which have anti-inflammatory and antimicrobial action. In addition to the presence of cytoplasmic granules, these cells present variable numbers of lobes, being called polymorphonuclear cells (neutrophils, eosinophils and basophils). Agranulocyte leukocytes are those that do not have granules visible in optical microscopy, and

have only one lobe, being denominated monomorphonuclear. Agranulocyte leukocytes are monocytes and lymphocytes [3, 8].

During hematopoiesis, cells destined to form the granulocytic series (neutrophils, eosinophils and basophils) synthesize proteins and cytoplasmic granules. The primary granules have azurophilic staining and are responsible for the conversion of precursor cells called myeloblasts into promyelocytes. Subsequently, there is the appearance of specific granules, responsible for the progression to myelocytes of the type: neutrophils, eosinophils and basophils. After determining the granulocytic lineage, the following cells are indivisible by mitosis, being characterized by the presence of segmented nucleus, ability of motility, phagocytosis and microbial destruction. Depending on the lineage, mature leukocytes are able to adhere and traverse the wall of the venules, for the purpose of performing tissue defense and reconstruction [3, 9].

Given the importance of blood cells, the analysis morphologically, quantitatively qualitatively, is an important task, since they can indicate anemia, leukemias, virus infections, bacteria or parasites, thromboses, allergies, among other diseases. This analysis is performed through an examination called complete blood count test (CBC), which is a highly requested medical examination in the medical routine, as it provides the diagnosis directly or is indicative of several diseases. This test consists of the erythrogram, leukogram, and platelet, which evaluate the quantity and morphologies of red blood cells, leukocytes and platelets, respectively [10].

2 Problematic

Currently, the complete blood count test (CBC) can be performed through two methodologies: the manual and the automated. The manual methodology is totally dependent on the human performance combined with the use of non-automated equipment. It is a cheaper but more time consuming and less reliable exam because it depends on the counts and calculations performed by health professionals. It is considered a good alternative for small laboratories, where the demand for exams is small and the cost with the acquisition with hematological equipment and reagents does not match the cost-benefit of the process [10].

For this, it is necessary to make a blood smear, by sliding a few microliters of blood on a glass slide and subsequent use of dyes, which allow the visualization of cellular structures. These dyes have the function of staining the nuclear and cytoplasmic structures of blood cells. Only the final portion of the slide is used to perform the counts, since the anterior portions have clustered and/or overlapping cells, preventing a reliable count. This blood smear is used to analyze the morphology and staining of red blood cells, leucocytes and platelets [10, 11].

The automation of the hemogram implies a greater agility in the accomplishment of the exams and in the release of the reports, however, they are a more expensive methodology when compared to the manual methodology. In the 1950s, Coulter Electronic, Inc. introduced the impedance principle for cell counts. The principle

of impedance is based on the fact that the electrically conductive cells are diluted in a conducting solution of electricity. This cell suspension is weighed through an orifice with a diameter of about $100\ \mu\text{m}$, where there is an electric current passing through. This electric current originates from two electrodes: one located on the inner side of the hole and positively charged, and another located on the outer side of the hole, negatively charged. In this way, each time the cell passes through the hole it interrupts the electric current and there is a change in the conductance, consequently, each interruption is counted as a particle [10, 11].

The impedance principle, over the years, was enabled with counters capable of measuring cell volume. Such evolution was the result of the correlation of the proportionality of the magnitude of the interruption of the electric current (pulses) as the cellular volume. Thus, it was observed that small pulses correspond to small volumes, whereas large pulses result from larger volumes. From this correlation between the magnitude of the electric current and the cellular volume, a new concept was created called the threshold concept. The threshold concept is responsible for classifying cells according to their volume, thus allowing the detection of globular volume. The globular volume corresponds to the hematocrit performed in the manual blood count, however, it receives this name because it is performed without the need for microcentrifugation. Both the impedance principle and the threshold concept are responsible for the introduction of multi-parameter devices on the market. These devices are able to perform simultaneous cell counts using separate channels for the counts [10, 11].

In the 1970s, laser light scatters and hydrodynamic fluid techniques were introduced. Both techniques preserve nuclei and granulation of leukocytes, retracting only the cytoplasmic membrane. The techniques are based on the principles of diffraction, refraction and reflection of the light emitted. However, in these techniques, the erythrocytes are undetectable, as a solution the erythrocytes are counted by means of flow cytometry and hydrodynamic focus, where the erythrocytes are counted one by one through an extremely fine capillary. These cells are subjected to a laser beam, where the light scattering is analyzed at different angles of deviation, where at zero degree is indicated the cell size, the ten-degree indication of the internal structure and the 90-degree indication of leukocytes and their characteristics of looseness and granulation content [10, 11].

Over the years, interest in digital imaging methods has increased due to its two main areas of application: improvement of pictorial information for human interpretation and processing of image data for storage, transmission and representation for autonomous machine perception. This way, new technology has been developed in the engineering branch, being of great utility in the medical areas, such as x-ray and tomography. These examinations are based on the capture of images of the patient's body, thus facilitating the diagnostic imaging of various diseases [12]. The aim of the present study was to develop a blood smear segmentation algorithm capable of detecting and counting erythrocytes and leukocytes accurately and with high precision, sensitivity and specificity.

3 Proposal

The experiments were conducted through digital images acquired in hematology databases. The blood smear is made by depositing a few microliters of blood on a glass slide. The fields of the blood smear contained erythrocytes in different sizes, colors, and quantities [13]. It is important to emphasize that only the final portion of the slide was subjected to analysis and image capture, because in this place the cells do not show overlap and/or hyper coloring by the use of dyes. The images were obtained in jpeg, jpg e png format and were transferred to Matlab® software simulation environment, version 8.3 of 64 bits (2014a). In this environment, image segmentation algorithm was developed for counting and detection of erythrocytes. This process was performed through the union of the Watershed Transform image processing techniques and morphological operations, which originated the WT-MO methodology for the detection and counting of erythrocytes and leukocytes. The logic of the algorithm is shown in Fig. 1.

In the image preprocessing and color conversion step, image quality is being improved. This occurs because most of the images of this type can find illumination problems. Thus, it passes through processes of filtering, image enhancement, color conversion and the segmentation process itself for this digital quality of these images [14, 15]. In the process of image segmentation, the morphological technique is the most used, because the mathematical morphology offers a powerful tool for the segmentation of images, being used to describe the format of the region, such as limits, skeleton and texture.

The segmentation process consists in the act of fractionating an image into groups of spatially connected pixels homogeneously [14]. Based on these characteristics the Watershed transform was chosen as an image segmentation tool, responsible for the detection and counts of blood cells present in the digital images of blood smears.

The segmentation of images by the Watershed transform is employed as a process of pixel labeling, where all pixels belonging to the same homogenous region are marked with the same label. The definition of homogeneity of a region of the image

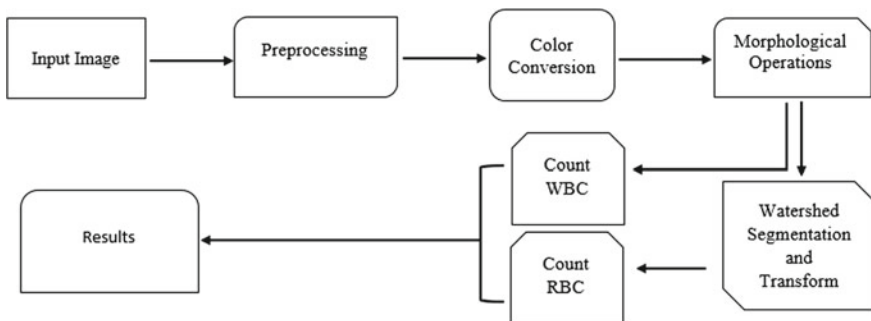


Fig. 1 Diagram demonstrating the logic used in the development of the WT-MO methodology

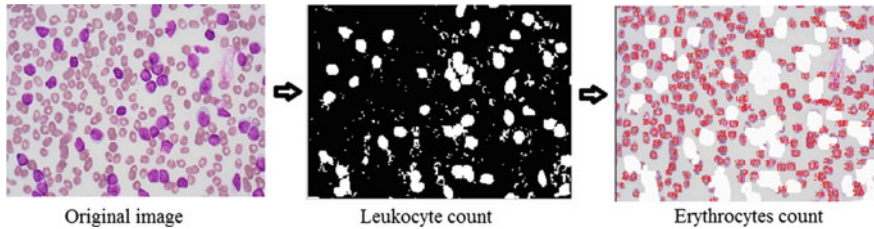


Fig. 2 Detection and counting of erythrocytes and leukocytes simultaneously by the WT-MO methodology

presents a particular concept of the segmentation process, since each image presents its own pattern of pixels [16, 17]. Considering that erythrocytes in a healthy state have color and homogeneous sizes, the Watershed transform presents a good performance in the labeling of sets of pixels representing these blood cells.

For the labeling of the red cells the pixel is assumed as a variable directly related to the morphological characteristics of the cells under analysis. In that case, the labeling should take into account the size and color of the erythrocytes. As the pixels are grouped according to their characteristics, the algorithm counts and labels each cell in the image. The labeling consists of inserting a number on top of each cell and according to the counting order established by the WT-MO methodology, as shown in Fig. 2.

Morphological operations constitute a broad set of image processing operations based on the shape of binary images [18]. Removal of leukocytes from the image is an important process for the detection and counting of blood cells. Considering the morphological similarity of leukocytes and erythrocytes, referring to the rounded form, it is possible that the algorithm performed an erroneous count. Another factor that could cause a failure to count would be the size of the blood cells, since an erythrocyte measures about $8\ \mu\text{m}$ of diameter, while a newly released lymphocyte (leukocyte type) by the bone marrow measures about $10\ \mu\text{m}$.

Besides the shape and diameter, the amount of cells produced and released by the bone marrow also justifies the choice of Morphological Operations as an image segmentation tool. The erythrocytes are expressed on the order of 10^6 , while the leukocytes are expressed in 10^3 [4]. The lower amount of leukocytes facilitates the action of morphological operations in their removal of the digital image. Consequently, this action interferes positively with the results of precision.

Morphological operations are applied to the structural element at the input of an image creating an output of the same size [16]. In this way, the value of the pixels in the image output is based on the comparison of pixels corresponding to the image pixels and adjacent pixels. The number of pixels added or removed from objects in an image depends on the size and format of the structural element used to render the image. The morphological operations were used to segment erythrocytes. This process is characterized in the concepts of size, shape, structure, and connectivity of objects in the image, involving erosion, dilation, opening, closing and reconstruction.

The dilation consists of adding pixels to the edges of objects in an image, aiming to repair breaks in the image. The erosion consists of removing the pixels from the edges of the image, being used to divide objects. The opening is applied to smooth out contours. The closing is used to merge intervals and fill spaces. The reconstruction is responsible for extracting relevant information from the image [18].

Therefore, this entire process involves 2 images and a structural element. The two images: one acts as a starting point for transformation and another act as a mask that restricts the transformation of the image. Finally, morphological reconstruction aims at restoring the original forms of the object that remains after the erosion process [19, 20]. This process is applied in the detection and counting of leukocytes by the WT-MO methodology, as shown in Fig. 2.

4 Results and Discuss

Were selected 30 images of microscopy fields, containing 6453 red cells and images of microscopy fields containing 187 leukocytes. Both fields were submitted to manual counts, with subsequent counting by the algorithm of detection and counting of blood cells. The values are then compared with each other to determine the accuracy of the proposed methodology. To obtain the results of this study, more than 300 simulations were performed.

The digital images were first submitted to a traditional manual counting methodology. Afterward, these images were sent to the Matlab software, where the red cells were detected and counted by the Watershed transform and the leucocytes were quantified through the morphological operations technique. The results obtained were compared with the manual methodologies, as shown in Figs. 3 and 4.

The development of systems, methods or tests involving the detection, diagnosis or prediction of results, presents the need to validate their results in order to quantify their discriminative power and to identify a procedure or method as appropriate or

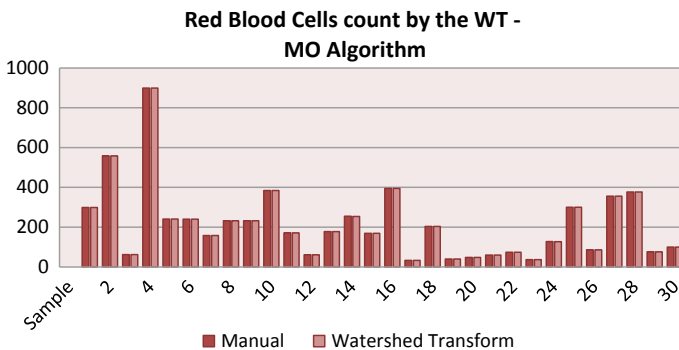


Fig. 3 Comparison of manual counting of erythrocytes x counting by the WT-MO methodology

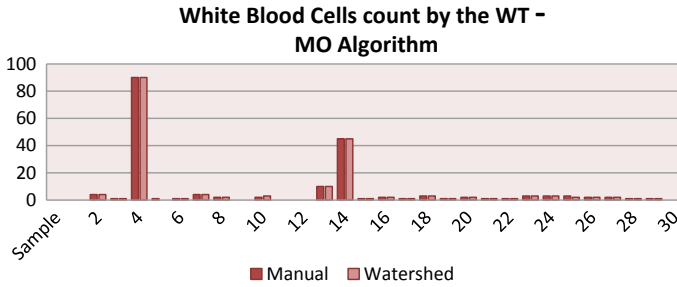


Fig. 4 Comparison of the manual count of leukocytes x count by the WT-MO methodology

Table 1 Matrix of confusion

		True value (observed value)	
		Y = 1	Y = 0
Predictive value	Y = 1	TP (True positive)	FP (False positive)
	Y = 0	FN (False negative)	TN (True negative)

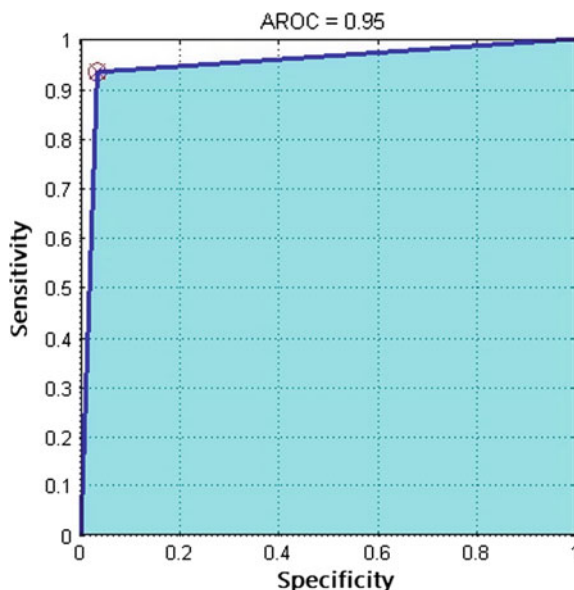
inappropriate for a particular type of analysis. However, the simple quantification of hits in a test group does not adequately reflect the efficiency of a system, since the quantification is dependent on the quality and distribution of the data in this test group [21].

Thus, when the test presents binary response variables (1 if the individual is an event and 0 otherwise), it is necessary to choose a prediction rule that determines where the individual should be framed (cutoff point). In the case of this research, the classification is based on the values obtained during the manual counts of red blood cells and leukocytes. A widely used way to determine the cutoff point is through the Receiver Operating Characteristic Curve (ROC) curve. The ROC curve plots sensitivity versus specificity for all possible cut-off points between 0 and 1 [22, 23].

After determining the cut-off point, it is necessary to evaluate the discriminating power of the model, that is, to discriminate between the events of the non-events. For that, the metrics were created: Accuracy, Sensitivity, Specificity, True Predictive Positive and True Predictive Negative. All these metrics are dependent on the confusion matrix, which is represented by the contingency Table 1 in which the expected value is in the line, and in the column, the observed value (true value) [21–23].

- True Positive: the test is positive in diseased patients, in the case of this research the cells are counted correctly by the algorithms in comparison to the manual counts;
- True Negative: the test is negative in healthy patients, in the case of this research are leukocyte-free images in both manual counts and counts by algorithms;
- False Positive: the test is positive in healthy patients, in this case, they are counted more cells in relation to manual counting;

Fig. 5 Roc curve referring to red cell count by the WT-MO methodology



- False Negative: the test is negative in diseased patients, in this case, they are leukocytes counted by the algorithms, and in manual counting these leukocytes are non-existent.

Thus, 2 binary matrices were created: (1) matrix for erythrocytes quantified by the Watershed transform; and (2) matrix for leukocytes quantified by morphological operations. Later, these matrices were transferred to Matlab software, where the values of: Sensitivity, Specificity, AROC (area of the ROC curve), VPP—True Predictive Positive (the probability of an individual evaluated and with a result positive) and VPN—True Negative Predictive—(probability that an evaluated individual with negative result is actually normal). The ROC curve of the Watershed Transform is shown in Figs. 5 and 6, with your parameters showed in the Tables 2 and 3. All values were given by function used in software Matlab.

Thus, it is possible to note that the WT MO methodology can be seen as a tool with high accuracy in erythrocyte detection (95%), high sensitivity (93%) and high specificity (96%). Regarding leukocyte detection, the algorithm also presented highly satisfactory values of accuracy (98,3%), sensitivity (99%) and specificity (96%). The highest accuracy, sensitivity and specificity in white blood cell counts is due to human physiology, as the white blood cells are synthesized by the bone marrow in a smaller quantity when compared to the red blood cells.

When compared these results with the works [24–27], the WT-MO methodology also presents a high accuracy. In this way, the WT-MO methodology developed in this research is a viable alternative for the counting of erythrocytes and leukocytes in laboratories of clinical analysis, since the precision and reliability of the tests are indispensable criteria.

Fig. 6 Roc curve referring to white cell count by the WT-MO methodology

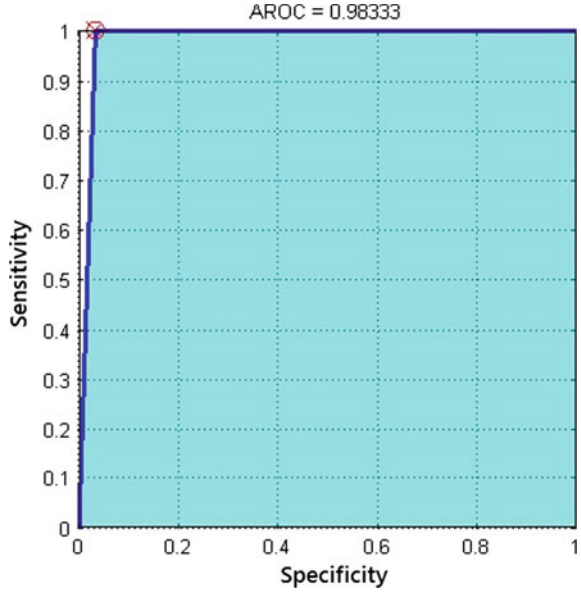


Table 2 Parameter analysis of the ROC curve for erythrocyte counting

Parameters of the ROC curve	
RBC— <i>Watershed Transform</i>	
Distance	0.074
Threshold	0.5
Sensitivity	0.93
Specificity	0.96
AROC	0.95
Accuracy	95.00%
PPV	96.60%
NPV	93.54%

Table 3 Parameter analysis of the ROC curve for erythrocyte counting

Parameters of the ROC curve	
WBC— <i>Morphological Operations</i>	
Distance	0.033
Threshold	0.5
Sensitivity	0.99
Specificity	0.96
AROC	0.98
Accuracy	98.33%
PPV	96.77%
NPV	100%

5 Conclusions

The confirmation of the diagnosis or even the accomplishment of laboratory tests through algorithms, provides greater reliability of the results to both health professionals and patients, because the algorithms reduce the chances of human failures. Due to its high sensitivity, accuracy and specificity in the simultaneous counting of erythrocytes and leukocytes, the WT-MO methodology can be seen as a future tool of medical laboratory routine, and can be used to confirm altered reports in laboratories with higher purchasing power, or It can be used with a method of substitution of the manual methodology in smaller laboratories, where the demand for examinations per day does not entail the acquisition of a high-cost hematological equipment.

It is important to emphasize that techniques result in the reduction of the cost during the production/supply of products/services, it directly impacts the final value. High values of products and services are important factors for the exclusion or removal of populations from access to health. Thus, the use of blood cell detection algorithms can make the blood count a more accessible test for less favored populations around the world.

References

1. Molinaro, E., Caputo, L., Amendoeira, R.: Métodos Para a Formação de Profissionais em Laboratórios de Saude. Fio Cruz, Rio de Janeiro (2009)
2. Verrastro, T., Lorenzi, T.F., Neto, S.W.: Hematologia e Hemoterapia: Fundamentos de Morfologia, Patologia e Clínica. Atheneu, Rio de Janeiro (2005)
3. Abbot laboratório de México S.A. Atlas com interpretação Histogramas y Escatergramas. Buenos Aires: E.G (2002)
4. Fischbach, F., Dunning, M.B.A.: Manual of Laboratory an Diagnostic Tests. Lippincott Williams e Wilkins, Pennsylvania (2009)
5. Junqueira, L.C., Carneiro, J.: Biologia Celular e Molecular. Guanabara Koogan, São Paulo (2012)
6. Junqueira, L.C., Carneiro, J.: Histologia Básica. Guanabara, São Paulo (2004)
7. Monteiro, A.C.B et al.: Sickle Cell Anemia, a genetic disorder characterized by the change in shape of red blood cells. Saúde em Foco, Edição nº: 07/Ano (2015)
8. Abbas, A.K., Lichtman, A.H., Pillai, S.: Cellular and Molecular Immunology. 7th edn. Elsevier (2012)
9. Loffler, H., Rastetter, J., Hafferlach, T.: Atlas of clinical hematology.6th edn. Springer, Berlin (2005)
10. Failace, R et al. Hemograma, manual de interpretação. 5ª. Artemed, Porto Alegre (2009)
11. Bernadette, F., Rodak, G.A., Fristma, K.D.: Hematology—Clinical principles and aplications. Elsevier, New York (2015)
12. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Using Matlab. 2nd. Gatesmark (2009)
13. Hoofman, R., Benz, E.J., Silberstein, L.E., Heslop, H., Weitz, J., Anastasi, J.: Hematology: Basic Principles and Practice, Canada, 6th edn. Elsevier (2013)
14. Comaniciu, D., Meer, P.: Robust analysis of feature spaces: color image segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern, pp. 750–755 (1997)
15. Deng, Y., Kenney, C., Moore, M.S., Manjunath, B.S.: Peer group filtering and perceptual color image quantization. Proceedings of IEEE International Symposium on Circuits and Systems, vol. 4, pp. 21–24 (1999)

16. Mogra, M., Srivastava, V.: A Comprehensive review of analysis of counting blood cells using different image processing algorithms. *Int. J. Eng. Sci. Invention*. **3**(6), 29–31 (2014). www.ijesi.org. ISSN (Online): 2319–6734, ISSN (Print): 2319–6726
17. Mogra, M., Srivastava, V.: A comprehensive review of analysis of counting blood cells using different image processing algorithms. *Int. J. Eng. Sci. Invention*. ISSN (Online): 2319–6734, ISSN (Print): 2319–6726 (2014)
18. Soltanzadeh, R.: Classification of three types of red blood cells in peripheral blood smear based on morphology. In: *Proceedings of ICSP* (2010)
19. Chourasiya, S., Rani, G.U.: Automatic red blood cell counting using watershed segmentation. (*IJCSIT*) *Int. J. Comput. Sci. Inf. Technol.* **5**(4), 4834–4838 (2014)
20. Hemant, T., Sexena, S., Vadak, N.: Segmentation using morfological watershed transformation counting blood cells. *Int. J. Comput. Appl. Inf. Technol.* **2**(3) (2013)
21. Kawamura, T.: Interpretação de um teste sob a visão epidemiológica: eficiência de um teste. *Arq. Bras. Cardiol., São Paulo*, **79**(4), 437–441 (2002)
22. Sabattini, R.M.E.: Um Programa para o Cálculo da Acurácia, Especificidade e Sensibilidade de Testes Médicos. *Revista Informédica* **2**(12), 19–21 (1995)
23. Chow, S.C.: *Statistical Design and Analysis of Stability Studies*. CRC Press, Boca Raton (2007)
24. Arivu, S.K., Sathiy, M.: Analyzing blood cell images to differentiate WBC and counting of linear & non-linear overlapping RBC based on morphological features. *Elixir Comp. Sci. Engg.* **48**, 9410–9413 (2012)
25. Berge, H., Taylor, D., Krishnan, S., Douglas, T.S.: Improved red blood cell counting in thin blood smears. In: *Proceedings of ISBI*, pp. 204–207 (2011)
26. How, K.B., Bin, A.S.K., Bin, A.S.K., Siong, N.T., Soo, K.K.: Red blood cell segmentation utilizing various image segmentation techniques. In: *Proceeding of International Conference on Man-Machine Systems*, September 15–16, 2006, Langkawi, Malaysia (2006)
27. Zack, G.W., Rogers, W.E., Latt, S.A.: Automatic-measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **25**, 741–753 (1977)