Image Processing Techniques to Identify Red Blood Cells

Nicoleta Safca
Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest
Bucharest, Romania
nicoleta.safca@gmail.com

Dan Popescu
Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest
Bucharest, Romania
dan.popescu@upb.ro

Loretta Ichim
Faculty of Automatic Control and Computer Science, UPB
"St. S. Nicolau" Institute of Virology
Bucharest, Romania
loretta.ichim@upb.ro

Hassan Elkhatib
Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest
Bucharest, Romania
free2hasi@yahoo.com

Oana Chenaru
Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest
Bucharest, Romania
oana.chenaru@gmail.com

***Abstract*—This paper presents a method for the automatic identification and classification of red cells in different classes of interest for diagnosis using microscopic images of blood smear. The whole system uses different image processing techniques such as binarization, contrast enhancement, noise elimination, morphological operations (dilatation, erosion), labeling and extraction of some features of interest (area, perimeter, diameter). Using this information, some factors (form factor, circularity factor, and deviation factor) involved in the classification of red cells are calculated. The classification process has two phases: the first separates red cells in normal and abnormal type and the second classifies the abnormal in three subclasses. This system does not aim to replace the pathologist, but to assist him / her and to improve the execution time of these types of analyzes.**

Keywords—red blood cells, image processing, circularity factor, deviation factor, red blood cells classification

I. INTRODUCTION

Early diagnosis of blood diseases is extremely important in their rapid and effective treatment. Therefore, the identification and counting of the red cells in microscopic images of blood smear can lead to early diagnosis of blood diseases such as malaria, anemia, heavy metal intoxication, and other types of diseases whose rapid diagnosis is very important for controlling disease evolution, treatment evaluation and, for complete healing. In this moment, the conventional procedure consists in visual examination of blood samples by the pathologist using a microscope. This method is laborious, requires time and a long training of the examiner, and there can be some factors related to the human nature that can interfere, such as fatigue or lack of attention.

Automatic systems for identifying and classifying abnormal blood cell types have always been a field of interest to researchers around the world, but no method has been developed yet to replace or improve the classical one.

Erythrocytes are the most numerous elements found in peripheral blood. Their morphological examination should include information about size, shape, color (palate) and the presence of cellular inclusions. Normal red blood cells have a diameter of 7 μm to 8 μm, they have no nucleus and they are biconcave discs, with a flexible membrane that allows them to move through blood vessels of different sizes (arteries, veins and capillaries). Approximately one third of the red cells should have a central pallor [1].

Some red cells can have variations in morphological properties, which is the cause of abnormal red cell formation. There can be three types of defects that can occur to a red cell:

- Variations in size: Anisocytosis is the general term for size defects. Depending on the size, the cells can be normocyte (the normal red cell), macrocyte (larger than the normal red cell, with diameter bigger than 8.2 μm) and microcyte (smaller than the normal red cell, with diameter less than 7.2 μm).
- Variations in color or hemoglobin content: hemoglobin is a very important protein that gives the color of the red cells. When the quantity of hemoglobin is normal (the central pallor is less than 1/3 of cell), the red cell is a normochromic. If the central pallor is more than 1/3, than the cell is hypochromic.
- Variations in shape: there are a lot of shape defects that can occur to a red cell. The term for this kind of variation is Poikilocytosis. Some of most important types of abnormal shape red cells are: elliptocyte (an elongate oval cell), sickle cell (crescent shape), echinocytes (spin cell).

The presence with any of these cells can indicate the existence of some blood diseases such as [1]: anemia, thalassemia, lead poisoning, liver disease, kidney disease, etc.

In the last years, the researchers have shown a high interest for the automatic detection of the red cells abnormalities and for the illnesses caused by them, such that the actual stage of the research in this field is truly advanced. Thus, Hirimutugoda

978-1-5386-4444-7/18/$31.00 ©2018 IEEE

and Wijayarathna [2] studied a rapid automatic method for diagnosing disorders caused by red cell abnormalities. In the described method, they used imaging techniques and artificial neural networks to identify malaria and thalassemia in blood samples. Also, the authors in [3] proposed an automatic system, based on a SVM binary classifier for detection and counting the erythrocytes infected by malaria.

The problem of red cell segmentation is the first step for blood cell analysis. To this end, Miswan et al. [4] presented a method for segmenting red cells. The main idea of the study was to use morphological operations and masks to remove unwanted objects from images in order to improve the image quality. The number of red blood cells in a blood sample is evaluated in [5] by an image processing method based on Hough Transform and other features. On the other hand, the threshold based segmentation is proposed in [6] to separate the red blood cells from white blood cells.

In 2013, Taherisadr et al. [7] made a classification of abnormal red cells in 12 classes using image processing techniques. The study focused primarily on calculating the actual cell dimension characteristics (real diameter, area, and perimeter). Also, the authors in [8] studied the classification of abnormal red cells by calculating the shape factor of the identified cells. Sickle Cells Anemia becomes more and more common and can be detected from blood samples analysis [9].

The goal of this paper is to present a method for the automatic identification and classification of red cells in different classes of interest for diagnosis based on microscopic images of blood smear. The novelty of the paper consists in the way of combination of all the factors involved in classifications. Until now, these factors were not used in this particularly method.

II. Materials and Methods

The classification process of red cells has two phases: the first separates red cells in normal and abnormal type and the second classifies the abnormal in three subclasses. The algorithm proposed for red cell classification in normal and abnormal is presented in Fig. 1.

It has three main modules: Preprocessing module, Calculation module, and Classification module (Classification I and II).

A. PreProcessing

The first step in the proposed method is the preprocessing of the images. This phase of the algorithm is extremely important for the subsequent correct features extraction. Several sequential processes have been applied:

- Convert RGB images to grayscale images and contrast enhancement;
- Binary converting;
- Morphological operations;
- Labeling and marking of Regions of Interest.

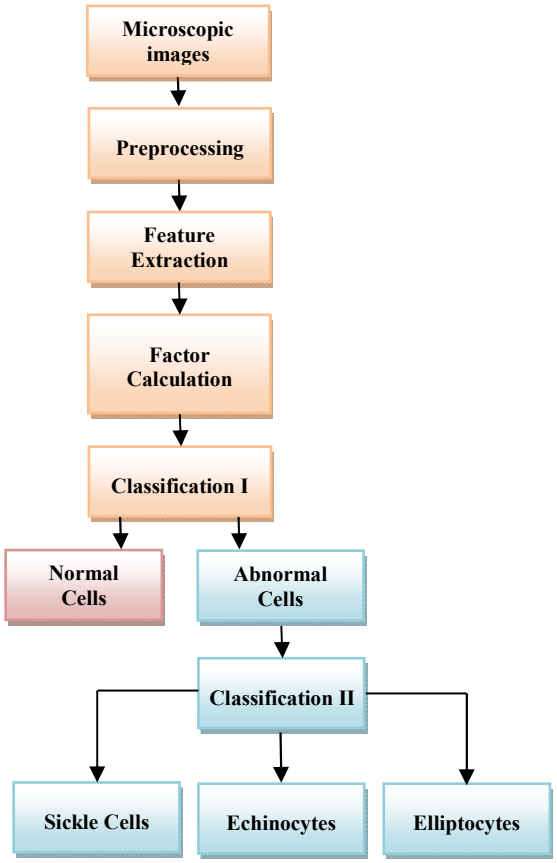


Fig. 1. Proposed method for red cell classification.

B. Features Extraction

The extraction of the features for each labeled region is done with the "regionprops" function [10], and all the extracted properties are used to calculate the factors that are parts of the red cell classification algorithm in the two classes mentioned above.

- *Area* is the number of pixels for each identified region. This variation is used to identify abnormalities associated with changes in red cell dimensions.
- *Perimeter* is the sum of boundary pixels for each cell.
- Diameter (d) is defined as the ratio (1):

$$d = \frac{4 \cdot Area}{Perimeter} \tag{1}$$

- *Major Axis* and *Minor Axis*: major axis represents the length (in pixels) of the major axis of the ellipse that has the same second-moments as the region and the minor axis represents the length (in pixels) of the minor axis of the ellipse that has the same second-moments as the region [9].

C. Calculating Factors

Using the features extracted, the next step in the proposed method is calculating each factor necessary for the algorithm:

- Circularity Factor (*CF*): is the proportion of major and minor axis. If circularity factor is smaller than 1.2 then the cells has a circle shape, otherwise the cells have elongation.

$$CF = \frac{MajorAxis}{MinorAxis} \tag{2}$$

- Form Factor (*FF*): For a perfect circle, the value of the Form Factor is equal to 1; otherwise the Form Factor has a different value.

$$FF = \frac{4 \cdot \pi \cdot Area}{Perimeter^2} \tag{3}$$

- Deviation Factor (*DF*): This factor evaluates the deviation of cell area from peripheral oval area.

$$DF = \frac{CF}{Area} \tag{4}$$

If the cell has no variation in shape, such as spins, the cell area and the peripheral oval area have similar values, and the deviation factor is smaller than 0.2. Otherwise, if deviation factor is bigger than 0.2, the cell has shape abnormalities [4].

- Semilunarity Factor (*SF*): This factor is used exclusively to identify sickle or semilunar cells and it is obtained:

$$SF = \frac{MajorAxis}{MajorAxis - MinorAxis} \tag{5}$$

If its value is less than 1.78, then the cell in question is a sickle cell [7].

D. Classification

Regarding the red cell classification process, this was done using the factors described above. In the first phase of classification process we classified cells in two classes: normal red blood cells and abnormal red blood cells. After image preprocessing, extraction of the characteristics and calculation of factors, a set of conditions were set for the values of factors calculated. The thresholds used are based on biological information, past experiences and literature review. The classification process can be seen in the Fig. 2. In the second part of the classification process, the abnormal red cells were classified in three different classes according with their shape defects: sickle cells, echinocytes and elliptocytes. This classification was also done using image processing techniques and thresholding.

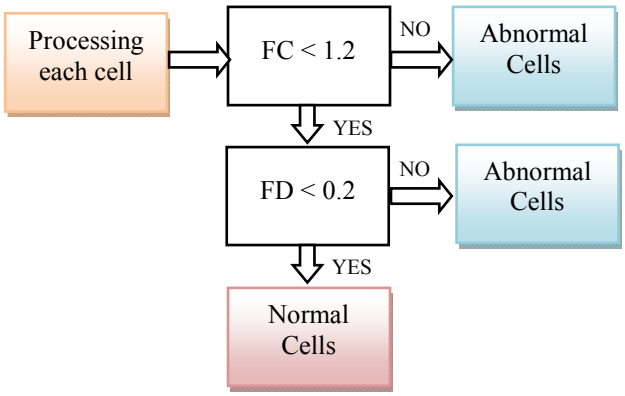


Fig. 2. Classification process I.

III. Experimental Results

All the operations are done in MATLAB [10-12] and the results for the obtained image preprocessing can be seen in Fig. 3, which represents the first step in preprocessing applied to a microscopic image. First, the original image is converted into a grayscale image because the entire system has been designed for 2D images, and then a contrast enhancement, preparing the next phase of preprocessing, binarization. The images used were taken from the American Society of Hematology Image Bank, an online image library that offers a comprehensive collection of reference images for a wide range of hematological subjects.

The gray level image is a matrix containing intensities with values ranging from 0-255, and the binarization process consists in converting this range into {0, 1} by replacing all values higher than the set threshold with "1" (white) and the smaller than this threshold with "0" (black) (Fig. 4).

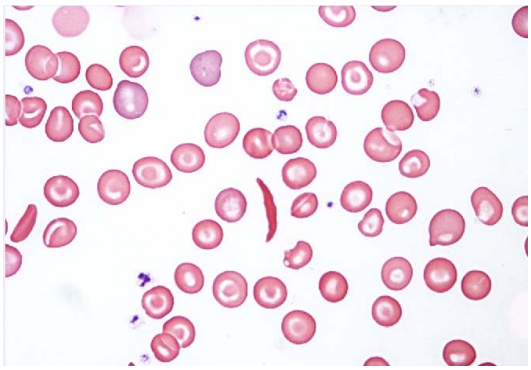
The function used for this process is "im2bw" [12], and if the threshold is not specified, the function uses the value of 0.5, being a mean value between black and white. In this case, for the calculation of the binarization threshold we used the "graythresh" function that sets the threshold so that the variation between the pixel values (black and white) is minimal.

Morphological imaging operations generally affect the shape or structure of an object in the image. This type of processing can only be done on a binary image and aims to filter and eliminate the protuberances.

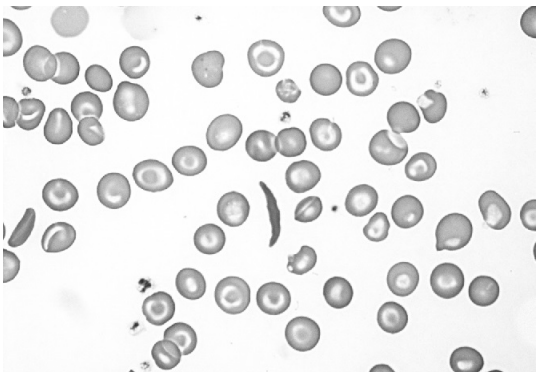
The "imfill" function was used to fill the holes (Fig. 5.a). In this case, a hole is a set of background pixels that represent the complement of pixels of interest and cannot be accessed by filling the background at the edges of the image. To remove incomplete objects from the edges of the image, the "imclearborder" function (Fig. 5.b) is used to erase any element that interacts with the image border, and incomplete objects are removed by the "bwareaopen" function (Fig. 5.c).

By performing these morphological operations, we removed all the small objects or those that touch the edges and filled the holes inside the cells. Thus, we have eliminated one

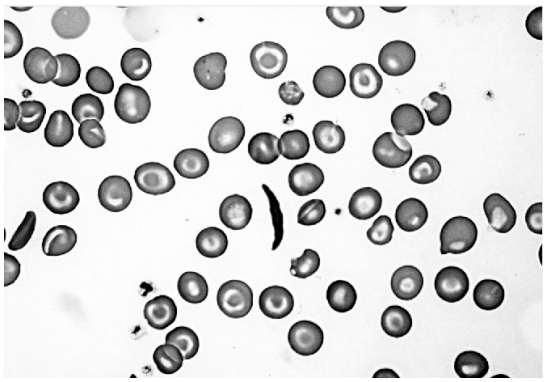
of the causes of errors, but in the same time it can be a problem, because remoted items can carry important information that is lost through these operations. Also, another element that could be an error generator is overlapping cells.



a)



b)



c)

Fig. 3. First step of preprocessing: a) original image, b) gray scale image, c) contrast enhancement.

In this paper, the overlapping cells were not removed, being considered important for the analysis that was performed. The labeling of the objects from images is done using the "bwlabel" function. In order to develop the algorithm, each object in the image must be identified and labeled so that for each cell identified in the image it is possible to analyze the characteristics of interest (Fig. 6). This is the process by which red cells are separated from the background of the image.

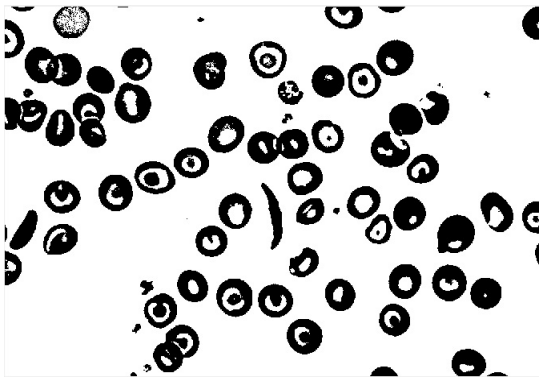
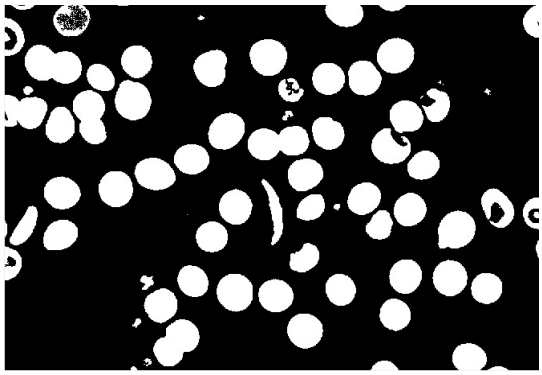
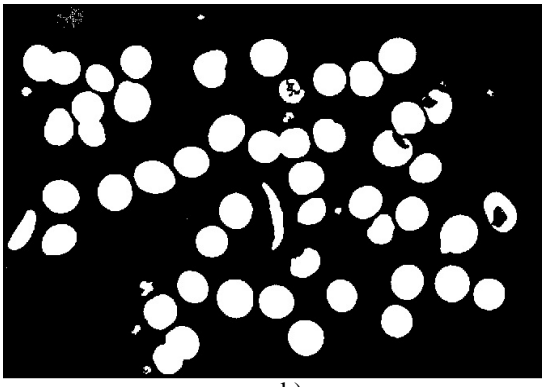


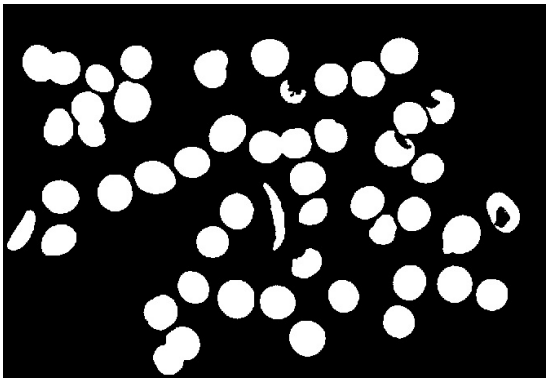
Fig. 4. Binary image.



a)



b)



c)

Fig. 5. Morphological operations: a) filled holes, b) clear borders, c) removing small objects.

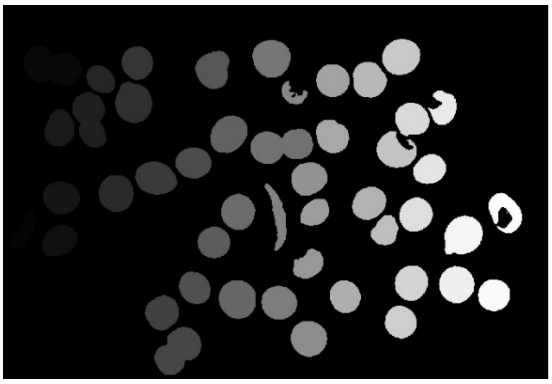


Fig. 6. Cell labeling.

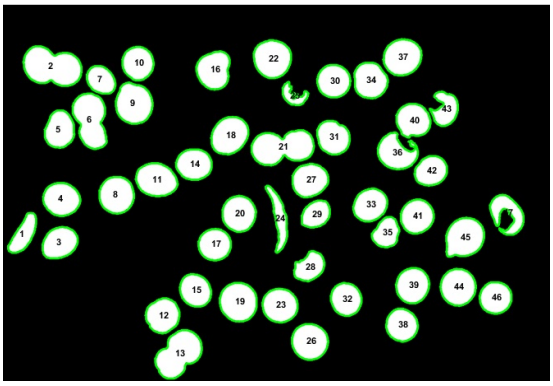


Fig. 7. Borders and cells counting.

The "bwboundaries" function only applies to a binary image where pixels with the value "1" represent the objects, and the background is made up of pixels with the value "0". As can be seen in Fig. 7, all red cells that do not touch the edges of the image or are incomplete have not been selected and numbered because they could lead to erroneous data analysis. They were excluded by applying morphological operations.

The results obtained for factors by applying the formulas mentioned above for a processed image are presented in Table I. As it can be seen, the factors have been calculated just for the marked cells and then used in classification process.

Fig. 8 is the result of all the processes applied within the MATLAB work environment for a microscopic image. As it can be seen, the developed algorithm is able to perform a two-class classification: the normal red cell class, and the abnormal red cell class.

The results of second part of the study can be seen in Fig. 9, Fig. 10, and Fig. 11. First class of cells that was identified, was the one of sickle cells. The algorithm was tested using 34 different types of images, pre-classified by a specialist. In this case, for this class, the yield of the algorithm was 100%, and all sickle cells were identified.

Next, the echinocytes were identified. In this case, a problem has been identified due to the non-specific form of this type of cells. The algorithm's performance was not 100%, being omitted at least 2 cells for each processed image. The results of a processed image are shown in Fig. 10. As it can be seen in Fig. 10, there are at least 5 omitted cells. This part of the algorithm was tested using 23 different types of image, pre-classified before by a specialist, and the results show that no image was entirely correct diagnosed.

TABLE I. INTEREST FACTORS COMPUTED IN MATLAB

Cell number	Circularity Factor	Deviation Factor
1	2.68	1.29E-03
2	1.76	3.04E-04
3	1.29	4.41E-04
4	1.13	3.49E-04
5	1.30	4.45E-04
6	1.94	4.17E-04
7	1.30	6.23E-04
8	1.06	3.08E-04
9	1.12	2.81E-04
10	1.07	3.85E-04
11	1.29	3.56E-04
12	1.06	3.48E-04
13	1.60	3.21E-04
14	1.17	3.95E-04
15	1.15	4.24E-04
16	1.16	3.43E-04
17	1.04	3.77E-04
18	1.22	3.34E-04
19	1.06	2.73E-04
20	1.07	3.34E-04
21	2.14	3.91E-04
22	1.02	2.69E-04
23	1.06	3.32E-04
24	4.98	2.53E-03
25	1.71	1.73E-03
26	1.03	2.98E-04
27	1.13	3.56E-04
28	1.40	6.45E-04
29	1.34	6.58E-04

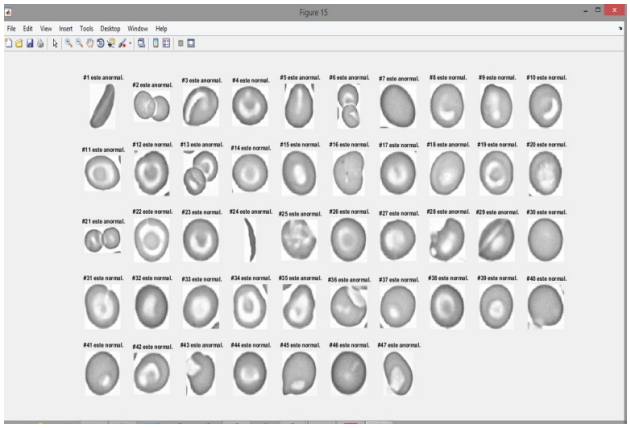


Fig. 8. Classification results.

In the last part of the study elliptocytes were identified. For this classification the form factor and the circularity factor were used and the thresholds were imposed so that the shape was as close as possible to the shape of an ellipse. The algorithm was tested on a set of 10 different images, also pre-classified by a specialist. One of the processed images can be seen in Fig. 11. The algorithm's performance for this

classification is around 96% in terms of the average accuracy (Table II).

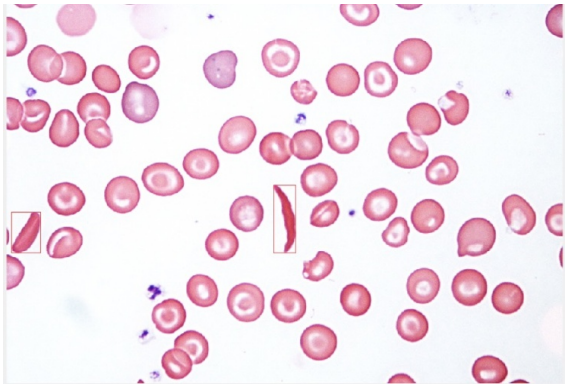


Fig. 9. Detection of Sickle cells.

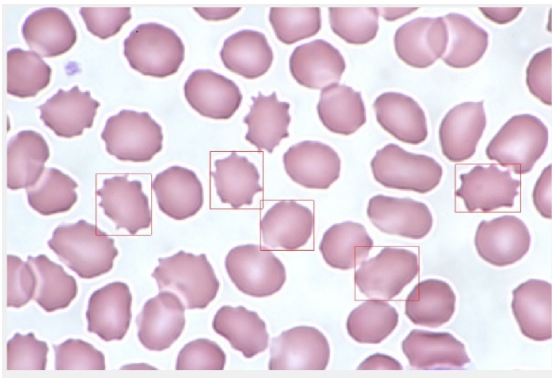


Fig. 10. Detection of Echinocytes.

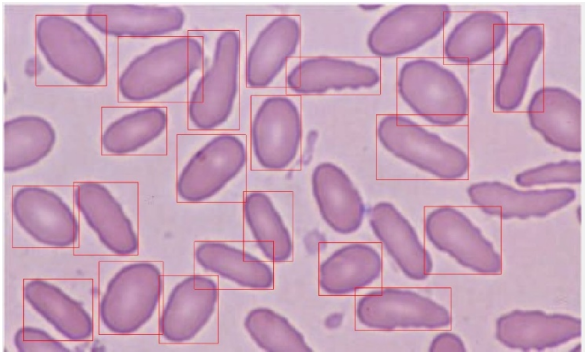


Fig. 11. Detection of Elliptocytes.

TABLE II. RED BLOOD CELLS DETECTION PERFORMANCE

Method	[5]	[6]	our
Accuracy [%]	96	95.3	96

IV. CONCLUSION

In this paper is presented an approach of classification of blood cells in two classes, normal and abnormal. The entire algorithm is based on image processing techniques applied in MATLAB. The first part of the study used 34 different types of images (format png or jpg), already classified by specialists. Regarding the obtained results, all the cells from images were identified and well categorized.

For the second part of this paper, 3 different types of abnormal red cells were identified. The performance of the algorithm is not 100% for all the classes, but it is an improvement to the existing methods of classifying using image processing techniques. This system can help the pathologist specialist to decide if a cell is normal or abnormal when he/she has doubts and also can reduce the necessary time for this kind of analyzes.

ACKNOWLEDGMENT

The work has been funded by UPB, Program GEX2017, project SET 22/2017 and project CAMIA 25/2017.

REFERENCES

[1] H. Theml, H. Diem, and T. Haferlach, Color Atlas of Hematology, Thieme, Second edition. 2004.

[2] Y. Hirimutugoda, and G. Wijayarathna, “Image analysis system for detection of red cell disorders using artificial neural networks,” Sri Lanka Journal of Bio-Medical Informatics, vol. 1, pp. 35–42, 2010.

[3] S.S.Savkarea, and S.P.Narote, “Automatic System for Classification of Erythrocytes Infected with malaria and identification of parasite’s life stage”, 2nd International Conference on Communication, Computing & Security, Procedia Technology 6, pp. 405 – 410, 2012

[4] S. Miswan, N. Salam, and A. Jamil, 2012 “Red blood cell segmentation using masking and watershed algorithm: A preliminary study,” International Conference in Biomedical Engineering, Penang, Malaysia, pp. 1–5, 27-28 February 2012.

[5] N. H. Mahmood, and M. A. Mansor, “Red blood cell estimation using hough transform technique,” Int. J. Signal and Image Processing (SIPIJ), vol. 3, no. 2, April 2012.

[6] Y. Alomari, S.N.H.S. Abdullah, R.Z. Azma, and K. Omar, “Automatic detection and quantification of WBCS and RBCS using iterative structured circle detection algorithm”, Comput Math Methods Med. 2014

[7] M. Taherisadr, M. Nasirzonouzi, B. Baradaran, and A. Mehdizade, “New approch to red blood cell classification using morphological image processing,” Shiraz E-Medical Journal, vol. 14, no. 1, pp. 44–53, 2013.

[8] D. N. Patil, and U. P. Khot, “Image processing based abnormal blood cells detection”, International Journal of Technical Research and Applications, no. 31, PP. 37-43, Sept. 2015.

[9] M. A. Fadhel, A. J. Humaidi, and S. R. Oleiwi, “Image processing-based diagnosis of sickle cell anemia in erythrocytes,” Annual Conference on New Trends in Information & Communications Technology Applications (NTICT), 7-9 March, 2017.

[10] https://www.mathworks.com/help/images/ref/regionprops.html

[11] www.mathwork.com.

[12] R. C. Gonzalez, R. E. Wood, and S. L. Eddins, “Digital Image Processing Using Matlab,” 2010.