



Research article

Red and white blood cell classification using Artificial Neural Networks

Simge Çelebi ¹ and Mert Burkay Çöteli ^{2,*}

¹ Histology and Embryology Department Doctoral Program, Yıldırım Beyazıt University

² Graduate School of Informatics, Middle East Technical University

* **Correspondence:** Email: mert.coteli@metu.edu.tr

Abstract: Blood cell classification is a recent topic for scientists working on the diagnosis of blood cell related illnesses. As the number of computer vision (CV) applications is increasing to improve quality of human life, it spreads in the areas of autonomous drive, surveillance, robotic applications, telecommunications and etc. The number of CV applications also increases in the medical sector due the decreasing value of doctors per patient ratio (DPPR) in urban and suburban areas. A doctor working in such areas sometimes would have to interpret thousands of patients' test results in a day. This condition would result disadvantages such as false diagnosis on patients and break on working motivations for doctors. Some of the tests would probably be interpreted using an application developed by Artificial Neural Networks (ANN). Tests related to blood cells are examined for the patients as a starting point of diagnosis and information obtained about their abnormalities give doctors a preliminary idea about the illnesses. This article issues generation of a CV application that would be used as an assistant of doctors who have domain expertise. The article issues segmentation of blood cells, classification of red and white blood cells containing 6 types such as erythrocyte, lymphocyte, platelets, neutrophil, monocytes and eosinophils using the segmentation results. It also discusses about a method for detection of abnormalities on red blood cells (erythrocyte).

Keywords: Blood cell classification; Artificial Neural Networks; peripheral blood smear; blood cell abnormalities

1. Introduction

Computer Vision (CV) is a wide working area on telecommunications, robotics, autonomous drives and medicine. Its main ability is providing profit on human work and assisting human to improve the efficiency of work. Blood cell classification is a recent topic for scientists working on the diagnosis of blood cell related illnesses. Classification of blood cells would only be performed manually by doctors who have domain expertise. It is important to classify blood cells from vision automatically without doctors due to the decreasing value of doctors per patient ratio (DPPR) in urban and suburban areas. Decreasing DPPR and harder to find a hematologist causes it to be a bio-engineering problem that has to be dealt with researchers working CV. Blood Cell analysis can be divided into two categories. First is the usage of Complete Blood Count (CBC) test [1]. It is required to find the percentage of Red and White Blood Cells from the patient's blood. These statistics are the health map of the patient. Each cell type has a role on human body. An increase or decrease on the number of blood cell type would cause a type of an illness. This test is the first analysis of the patient while entering the hospital and it is performed by flow cytometers. The next test is peripheral blood smears (PBS) [2]. It is a laboratory work that involves cytology of peripheral blood cells smeared on the plate. It requires pre-processing steps such as smearing, painting and washing of the microscope slide which contains a drop of blood. Then, it requires the analysis step of slide under a microscope by an expert. This study illustrates a method for the analysis of PBS using CV. Firstly, it suggests a classification task of 6 types of blood cells including erythrocyte, lymphocyte, platelets, neutrophil, monocytes and eosinophils without examining whether it is abnormal or not and a Monte-Carlo [3] type cell counting algorithm as an alternative to CBC test with statistics. There are also subjective and objective comparison sections. Next, it provides a method for detection of abnormalities using ANN. It is a post-processing step of the first classification stage. The remainder of this paper is organized as follows. Section 2 illustrates studies about segmentation methods and blood cell classification's literature. The proposed algorithm's classification steps and abnormality detection are issued at Section 3. This section also expresses the statistics and comparison of the work with doctor's vision as a subjective evaluation and CBC Test as an objective evaluation. Section 4 concludes the work.

2. Literature

In the literature, there are studies related to classification of blood cells and they are mostly concerned about the illnesses types. Studies on the diagnosis of Malaria infection are mostly related to moving object detection on PBS and bone marrow slides [12,15,21]. Other study is related to color and edge features of blood cells for the classification. If smear slide has an information about colors used in PBS, K-means separation algorithm works on RGB images and it is used as background subtraction or white blood cell region identification. However, PBS is a test which is not standardized for color types. Therefore, using different color features on the preparation of smear sample would cause a failure on the separation of blood cells using K-means clustering. Edge features are also used to extract cytoplasm of each blood cell from others [10,11]. Thresholding is another method to separate red and white blood cells from the slide [20]. Most of the works do not consider the cell types in detail. Instead, they use a separation algorithm for only the feature of red and white blood cell formats [9,17,18]. Detection of acute leukemia from PBS is another interest of

the works in the literature [8,19,20]. Morphological operations such as erosion and dilation are also illustrated to these studies for simplifying the separation process.

CV is a research area which contains a group of connected image processing or Artificial Intelligence (AI) blocks. Object detection is specifying the detected boundaries for the object and detection of the related class. There are methods used for object detection in the literature. Early algorithms focused on face and pedestrians detection [23,24]. Convolutional Neural Networks (CNN) are used for the extraction of object features in applications [22]. ANN contains a group of CNN connected each other with a constraint defined for each neuron. This method would be used for human, drone, face, car and blood cell detection [13]. The performance of ANN is related to the number training and testing samples and its structure. Success criteria for object detection is specification of object boundaries and object type. Classification of images having more than one object is the current deep learning working topic. Deep Residual Learning for Image Recognition [26] and YOLO (You only look at once) [25] are studies related to multiple object detections in the context. This case also has a number of solutions such as offering object regions and passing this region to ANN or a full ANN structure that does not have a pre-processing step. It has offerings on regions named as R-CNN [22]. This region offerings depend on the Intersection over Union (IOU) between other regions.

White blood cell identification is a task that is concerned with deep learning-based systems. First, a deep learning network is offered in this type of studies. Then, a post processing stage is required for multiple object detection. The article says that it results in 96.5% accuracy in 5 different blood cell types [26]. Most of the papers deal with white blood cell classification, some studies are concentrated on erythrocyte detection. They use similar ANN structures with the state of the art methods [27]. The next section illustrates the overall algorithm containing segmentation, classification step for 6 blood cell types and a region illustration for the output using Intersection over Union (IOU).

3. Algorithm

This paper illustrates an ANN algorithm to classify blood cells with its pre and post processing steps. Pre-processing steps include a segmentation rule to provide best fitting rectangles to ANN. Each rectangle is evaluated individually using ANN. At the post processing steps, we provide a region elimination method to detect overlapping regions and provide an accurate estimation for blood cell counting. This algorithm contains an information for a single magnified image. Monte Carlo principle [3], gathering images using a sampling method over the blood smeared slide would give an estimation about the complete blood cell count. The last step is the detection example of abnormalities on erythrocytes using a single binary classifier which outputs whether there is an abnormality or not. The algorithm flow is seen on Figure 1.

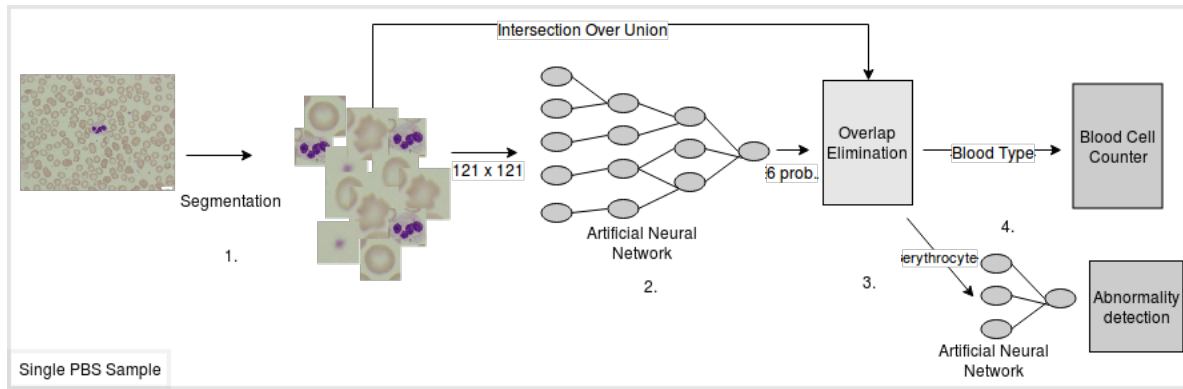


Figure 1. The Algorithm flow: 1. Segmentation, 2. Artificial Neural Network, 3. Overlap Elimination, 4. Abnormality Detection.

3.1. PBS image segmentation

Main features for the detection of cell objects from a basic magnified image are their circular feature and color. Possible rectangle regions are predicted using these features. Color feature is used to find center positions of the blood cells over the image. Then, cytoplasm edges are used for region boundaries. Here, we start with the conversion of RGB image to Grayscale format. Grayscale image is thought as a 2D data distribution over (x, y) points. The next step is thresholding the image with Otsu thresholding [5]. It is binarization of the image with representation of only black and white components. This data is used for two distinguishing parts. First one is finding the edges to determine the region boundaries. The other one is finding Euclidean distances [6] for center detection. In CV applications, Sobel operator is used to find the data distribution's first order derivative in x, y coordinates by convolution of a Sobel filter with a 2D data distribution [7]. It is seen on Figure 2.

-1	0	+1		+1	+2	+1
-2	0	+2		0	0	0
-1	0	+1		-1	-2	-1
G_x				G_y		

Figure 2. Sobel filters to find edges in x and y dimensions.

The next step is finding the center position of each blood cell. The method used here is Euclidean distance transform (EDT) [6]. It is measurement of distances from white pixels to black pixels by counting white pixels. The data distribution obtained after EDT will be used to find local maximum points. Second derivative of the data distribution is formed as gradient function.

$$\nabla f(x, y) = \langle f_{xx}, f_{yy} \rangle \quad (1)$$

A second derivative test is applied for each point. It is used for local maximum and minimum position finding mathematical expression. If $H > 0$ and $f_{xx}(x_0, y_0) > 0$ conditions occur at checking point, it is taken as a local maximum point.

$$H = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}(x_0, y_0)^2 \quad (2)$$

The pixel found as a local maximum is the center of the blood cell. The last step is the illustration of blood cell edges and its center using the algorithm defined in the literature is watershed [4,23]. Then, we would define a region that contains a blood cell. However, we do not have an information about its type. It is a feasible segmentation method to define the regions that will be given as input to ANN because we do not have to look whole possible blood cell regions and make eliminations from these regions using ANN. We handle it in image processing domain using this segmentation rule and computational complexity decreases without the need of region elimination. We know there is a region containing a cell, but we do not have a prediction about its type. The overall segmentation progress is seen on Figure 3.

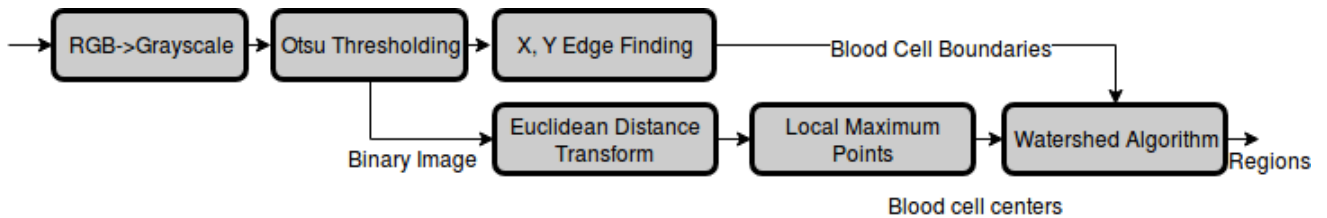


Figure 3. Segmentation algorithm flow.

3.2. Region classification using Artificial Neural Network

This part contains the details about ANN, training and testing samples. ANN used for this case contains 5 CNNs connected on behalf of each other. There is a pooling and local response normalization (LRN) layers between each CNN. At the end, there are inner product layers and a Soft Max classifier. It is generated for the classification of 6 types of cells (erythrocyte, lymphocyte, platelets, neutrophil, monocytes and eosinophils) using the offered region from the basic image. ANN structure is seen from Figure 4.

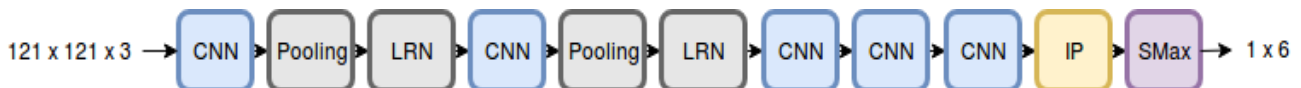


Figure 4. Artificial Neural Network structure.

Labelling of each blood cell is handled using doctors who have domain expertise. A blood cell database is created using 11869 PBS images captured from different patient's PBS tests using an Olympus BX51 series microscope with a magnification level of 100. It contains 19330 blood cell images with (121px) x (121px) in RGB. 3686 images are used for validation in a percentage of 0.19. 15644 images are used for training with a percentage of 0.81. Training and validation is an iterative

process. As a third-party framework, Caffe is used for training and testing the ANN [14]. Preparation of the image samples and labelling are handled with software used from the machine of Mantiscope [16].

3.3. Overlap elimination and Monte-Carlo sampling

PBS is a test that is not standardized if it is manually performed. There are devices designed to automate the manual process of PBS. [16] Mantiscope is a device that we have used for the preparation of microscope slides from blood tubes to make it analyze ready microscope slides. The quality on smearing, painting and washing would affect the PBS efficiency and accuracy of results. We would have overlapping regions due to insufficient smearing of blood drop. At this part, we illustrate the elimination of overlapping cells algorithm for accurate counting of blood cells.

Intersection over Union (IOU) for each region is defined as an intersection ratio according to other regions using the segmentation algorithm. It gives a decision rule for each classified region. The IOU rule for elimination is seen on Figure 5. It is used for some types of cells which would have nucleus more than once. IOU rule threshold is checked for different types of cells whether they have a single nucleus or not. In other words, it is Red (erythrocyte), White Blood Cell (lymphocyte, neutrophil, monocytes and eosinophils) or Platelets.

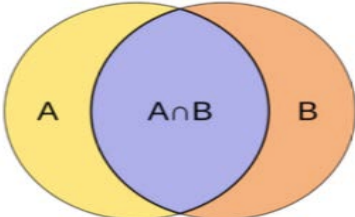
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$


Figure 5. IOU Rule for intersecting regions.

First, region boundaries are classified, and threshold is used for merging the rectangles would give a corrected region for boundary identification. Two different threshold values (for Red and White Blood Cells) are used after first classification stage and they are used to give the decision about region merging. Then, reclassification is done for defined region boundaries. The effects of defining region boundaries for different thresholds on a real work example is seen on Figure 6.

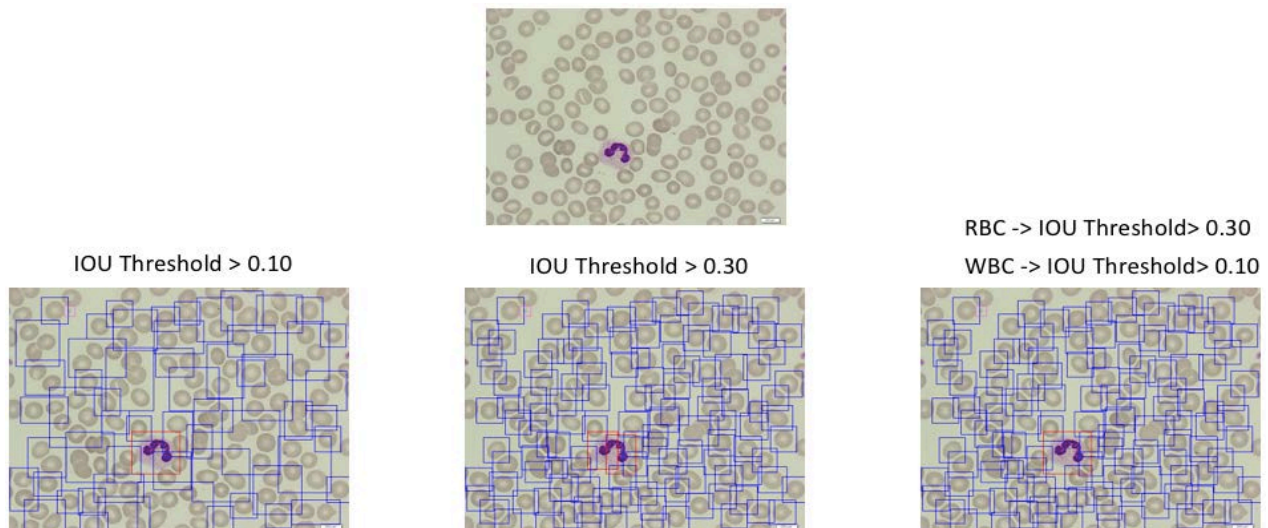


Figure 6. IOU Threshold rule for boundary region determination and its effect.

Blood Cell counting and comparison with CBC results is the estimation method used here named as Monte Carlo principle. In the ideal case, if each position of the microscope slide captured as an image sample is analyzed, we would have an accurate estimate about cell counting. However, it is a time consuming and overhead task. A prediction would already be done with a number of sampling. On the evaluation section, we will make a comparison of PBS Blood Cell counting method using 50 samples with CBC results. Each test is examined from the same blood tube of patient.

3.4. Abnormality detection

This part is a post processing step that is used to detect abnormalities of blood cells. Each cell is categorized up to this stage using Segmentation, ANN and Overlap Elimination algorithm. For each cell type, we have trained a single ANN with labeling the cells having shapes whether they are abnormal or not. Two classes are labelled for each cell. If we know the type of cell, we would take a decision to use the appropriate ANN repository. ANN structure for determination of abnormal and normal cells is shown in Figure 7.

This case is tested for abnormality detection of erythrocytes. A total number of 2319 erythrocyte images are labeled for abnormal and normal cases by 3 different doctors who have domain expertise. 0.2 percentage is used for evaluation, 0.1 percentage is used for testing and 0.7 percentage is used for training ANN. There is not a real-world scenario to evaluate the results, but after the training, we evaluated it with testing samples that are labeled by another doctor who isn't used in the training case. At the end, we have obtained an accuracy of 0.79 from the labeled test images. The network acquired an accuracy of 0.95 at the training step. It is a fact that diagnosis or defining the abnormalities of blood cell samples differs from one doctor to another. Since, this method is a subjective evaluation that is related to the expertise of the doctor, it doesn't give the final result that is enough to define diagnosis results. It would be an elimination tool to be used by a hematology expert during the early diagnosis.

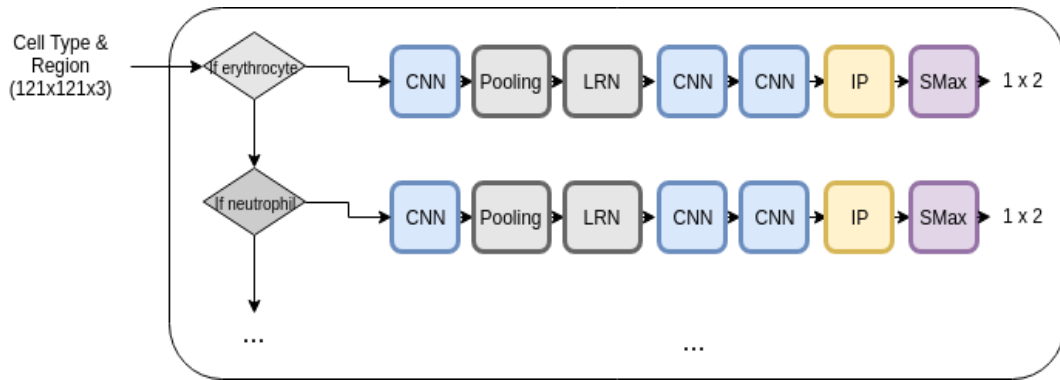


Figure 7. Abnormality detection by cell type specific ANN, Each ANN results whether there is an abnormality or not.

3.5 Evaluation

This section contains evaluation of two different scenarios. First one is randomly selection of 50 PBS Images and from these, we have used two measurements. They are manually counting total number of blood cells by a hematology expert and automatically counting of blood cells using the algorithm. Hematology Expert Vision's decision is accepted as true as a hundred percent to compute accuracy. The results are given on Table 1. White Blood Cell identification accuracy is calculated as 0.92 by including the platelets.

Table 1. Comparison of Hematology Expert and Computer Vision decisions.

Cell Type	Hematology Expert Vision	Computer Vision	Precision
erythrocyte	5800	5156	0.88
lymphocyte	22	21	0.95
platelets	156	142	0.91
neutrophil	48	43	0.93
monocyte	8	7	0.87
eosinophil	12	11	0.91

Second scenario is comparison of our Monte-Carlo Blood Cell counting algorithm with CBC Test results. These two tests are evaluated for the same patient's blood tube. Here, [16] Mantiscope is used for PBS Slide preparation and automatic slide scanning with its autofocus ability. Horiba DX Nexus device is used for CBC Test. For 6 cell types and 6 patients, we have illustrated graphs that show the comparison of CBC and PBS results. The sampling number for Monte-Carlo is 50. Figure 8 is the percentage of White Blood Cells (WBC) such as lymphocyte, neutrophil, monocytes and eosinophils. It makes a comparison between CBC and PBS. It is shown that the percentages of WBC resemble to each other with small errors.

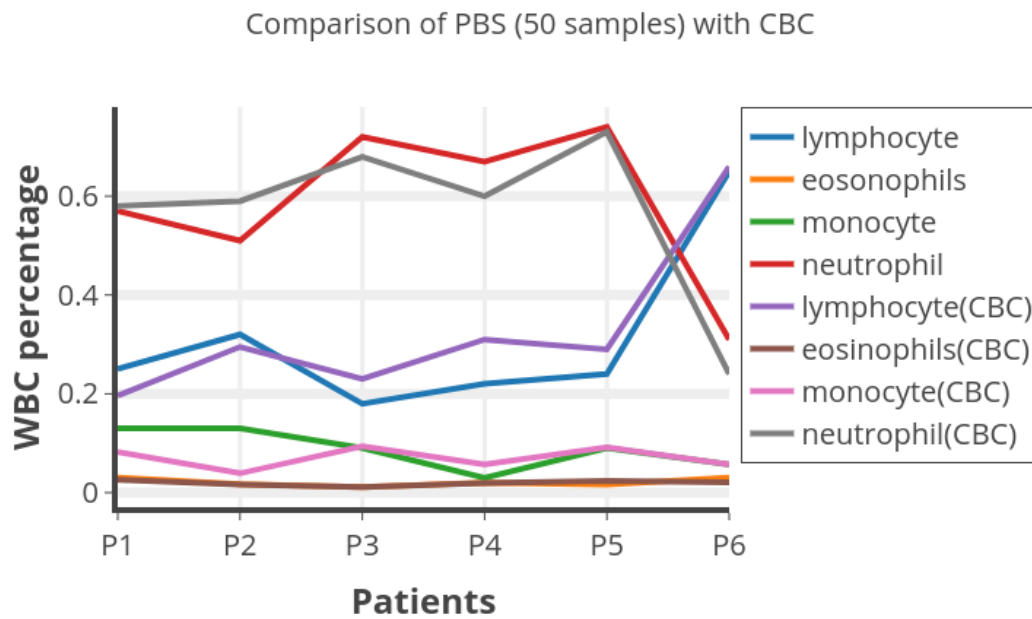


Figure 8. White Blood Cell (lymphocyte, neutrophil, monocytes, and eosinophils) percentage comparison between PBS and CBC.

Next two figures are blood cell counting results for two types of cells (erythrocytes and platelets). Monte-Carlo Cell counting would not give an accurate result for evaluation of the algorithm with CBC due to small sampling number (50), but it is seen from results that changes in CBC resembles changes in PBS. Since there is an interval determining the required number of cells that shows the patient's health, an interval criterion would be determined for the evaluation of PBS results and it would be used instead of CBC Test. It will be handled as a future work. Figure 9 shows the comparison.

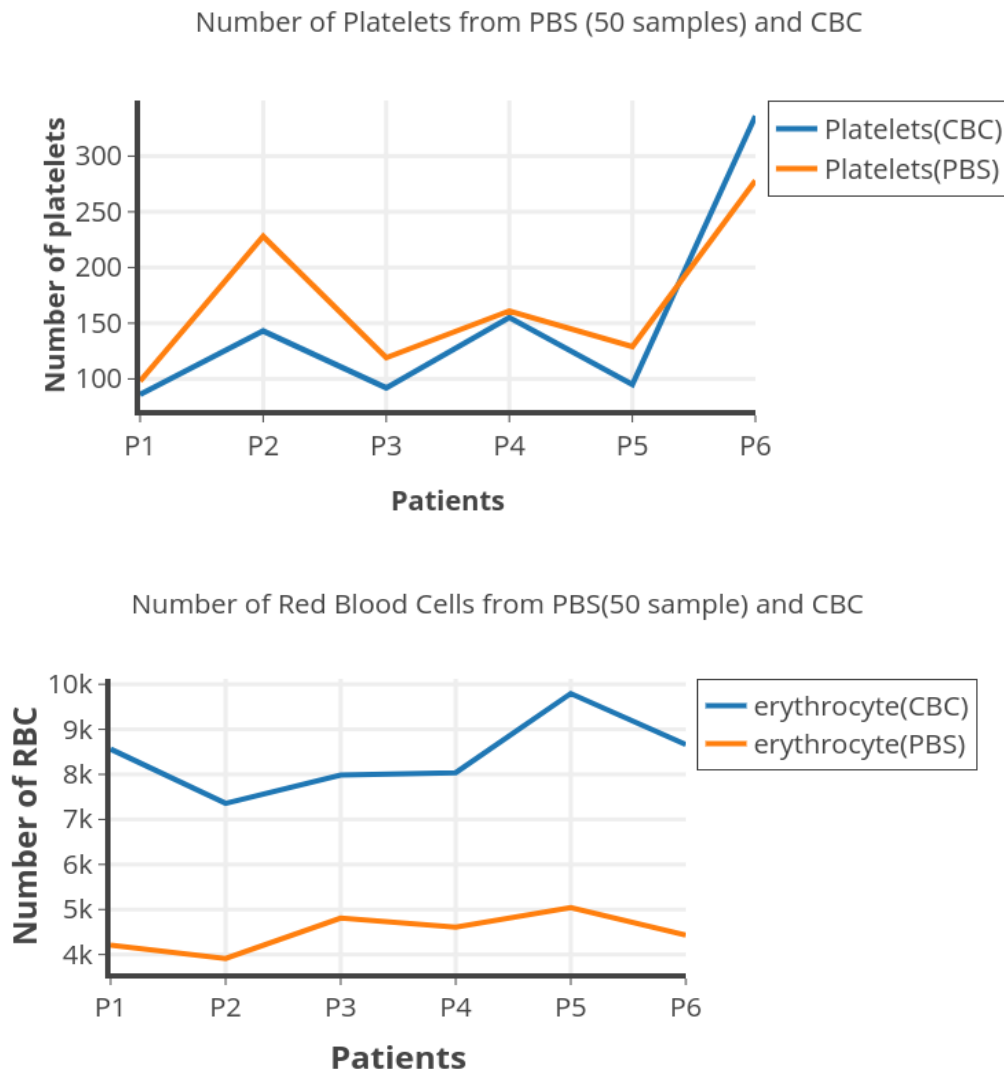


Figure 9. Number of erythrocytes and platelets from CBC ($\times 10^3/\mu\text{L}$) and PBS ($\times 1/50$ samples) Algorithm.

4. Conclusion

This study illustrates methods such as Segmentation of Blood Cells from a PBS Image, Classification of Blood Cells into 6 types and abnormality detection using cell type specific ANN. First two of the methods (Segmentation and Classification) are used with the help of Monte-Carlo sampling through the whole microscope slide. It would be used as a counting mechanism that is used instead of CBC. We have shown the comparison of CBC and our algorithm on Figures 8 and 9. The main efficiency criteria is the PBS preparation technique. Smearing quality and markers used for PBS would affect the efficiency. During these algorithm development process and Monte-Carlo sampling over slide, we have used the device named as Mantiscope which has an automatic PBS preparation and slide scanner ability [16]. Artificial Intelligence is a growing topic for medical imaging. True diagnosis is the most important activity for patient's health. Therefore, a doctor would use an AI assistant to interpret the patient's test results by increasing the accuracy of true diagnosis.

The last part of the study is the abnormality detection of blood cells for the diagnosis of blood cell related illnesses. The ratio of the abnormality detection accuracy is calculated using fewer number of images, but the AI assistant's interpretation ratio would become higher with training more samples obtained after abnormality labeling by hematology experts. The effect of using higher number of doctors is also an important factor to increase the accuracy. Example AI processing results are seen from the following Figures 10.

As a future work, the abnormality detection case will be studied using more doctors and samples. Then it is used to relate the abnormalities to the illnesses using ANN.

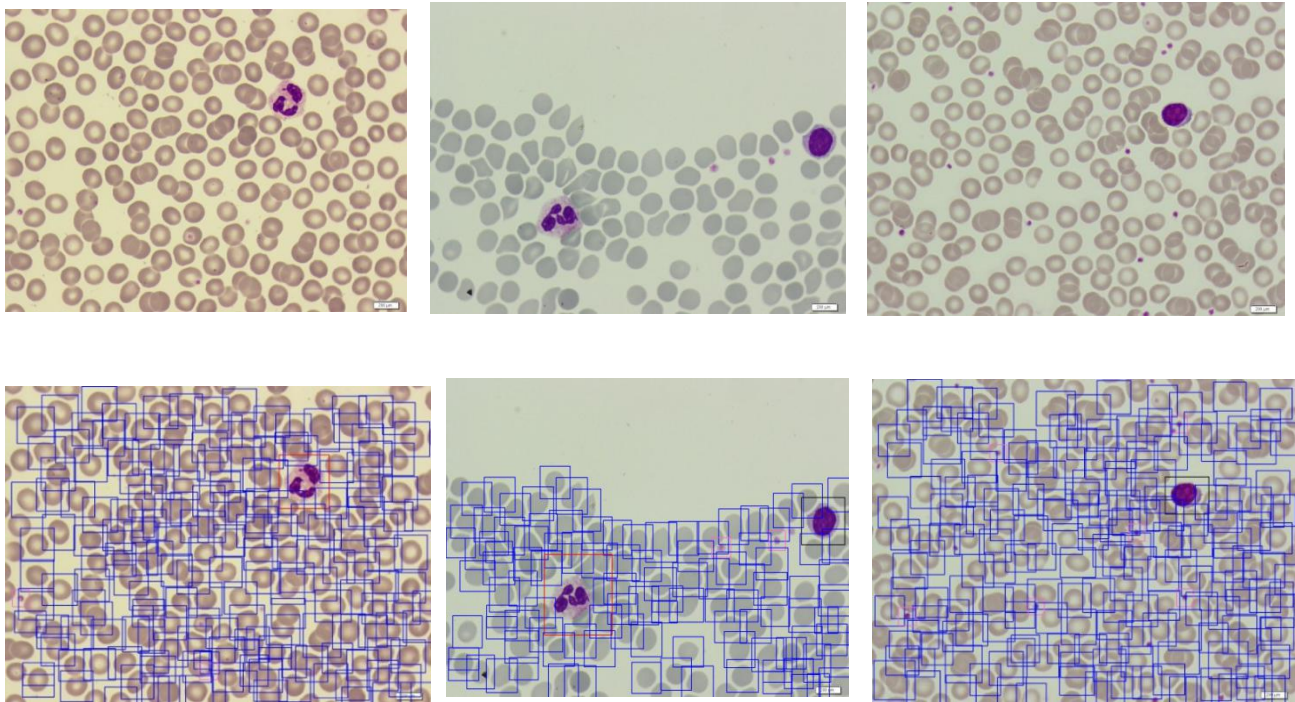


Figure 10. PBS AI results handled in different color features; black-> lymphocyte, blue -> erythrocytes, pink -> platelets, red -> neutrophils. Soft Max classifier threshold: 0.9.

Acknowledgements

The work reported in this paper is supported by Turkish Scientific and Technological Research Council (TÜBİTAK). The authors would like to thank Assoc. Dr. Emre Tekgündüz, Dr. Nurgül Özcan, Dr. Süheyla Çötel, Dr. Cemile Özsürekeçi, Şükrü Özsürekeçi for their support to archive and label the blood cells and Prof. Dr. Bülent Sankur for his guidance on the development of this work.

Conflict of Interest

All authors declare no conflicts of interest in this paper.

References

1. Shapiro MF, Sheldon G (1987) The Complete Blood Count and Leukocyte Differential Count: An Approach to Their Rational Application. *J Emerg Med* 106: 65–74.
2. Lynch E C. (1990) Peripheral blood smear. Butterworths, Boston: Pubmed, 90: 1373–1377.
3. Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57: 97–109.
4. Ng HP, Ong SH, Foong KWC, et al. (2006) Medical image segmentation using k-means clustering and improved watershed algorithm. *IEEE Southwest Symp Image Anal Interpret* 106: 61–65.
5. Kurita T, Otsu N, Abdelmalek N (1992) Maximum likelihood thresholding based on population mixture models. *Pattern Recogn* 25: 1231–1240.
6. Danielsson PE (1980) Euclidean distance mapping. *Comput Graph Image Process* 14: 227–248.
7. Sobel I (1990) An isotropic 3×3 image gradient operator. *Mach vision three dimensional scenes*, 376–379.
8. Scotti F (2005) Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images. *IEEE Int Conf Computl Intelli Meas Syst Appl CIMSA*. 25: 96–101
9. Sadeghian F, Seman Z, Ramli AR, et al. (2009) A framework for white blood cell segmentation in microscopic blood images using digital image processing. *Biol Proced Online* 11: 196.
10. Jesus A, Georges F (2003) Automated detection of working area of peripheral blood smears using mathematical morphology. *Anal Cell pathol* 25: 37–49.
11. Goswami R, Pi D, Pal J, et al. (2015) Performance evaluation of a dynamic telepathology system (Panoptiq™) in the morphologic assessment of peripheral blood film abnormalities. *Int J Lab hematol* 37: 365–371.
12. D'Ambrosio MV, Bakalar M, Bennuru S, et al. (2015) Point-of-care quantification of blood-borne filarial parasites with a mobile phone microscope. *Sci Transl Med* 7: 286re4.
13. Manik S, Saini LM, Vadera N (2017) Counting and classification of White blood cell using Artificial Neural Network (ANN). *IEEE Int Con Power Electron Intell Control Energy Syst* 2017:1–5
14. Jia YQ, Shelhamer E, Donahue J, et al.(2014) Caffe: Convolutional architecture for fast feature embedding. *Proc 22nd ACM Int Conf Multimedia*, 675–678
15. Das DK, Maiti AK, Chakraborty C (2015) Automated system for characterization and classification of malaria infected stages using light microscopic images of thin blood smears. *J Microsc-Oxford* 257: 238–252.
16. Automatic Peripheral Blood Smear and Slide Scanner Device. Available from: <http://www.mantiscopes.com>
17. Devi S, Singha J, Sharma M, et al. (2016) Erythrocyte segmentation for quantification in microscopic images of thin blood smears. *J Intell Fuzzy Syst* 4: 2847–2856.
18. Lee H, Chen YPP (2014) Cell morphology based classification for red cells in blood smear images. *Pattern Recogn Lett* 49: 155–161.
19. Amin MM, Kermani S, Talebi A, et al. (2015) Recognition of acute lymphoblastic leukemia cells in microscopic images using K-means clustering and support vector machine classifier. *J Med Signal Sensor* 5: 49.

20. Li Y, Zhu R, Mi L, et al. (2016) Segmentation of white blood cell from acute Lymphoblastic Leukemia images using dual-threshold method. *Comput MathMethod M* 2016: 9514707.
21. Linder N, Tukki R, Walliander M, et al. (2014) A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears. *PLoS One* 9: e104855.
22. Zhu C, Zheng Y, Luu K, et al. (2016) CMS-RCNN: contextual multi-scale region-based CNN for unconstrained face detection. In Bhanu B, Kumar A, *Deep Learning for Biometrics*, 3 Eds., Switzerland: Springer, 57–79.
23. Beucher S, Mathmatique CDM (1991) The watershed transformation applied to image segmentation. *Scanning Microsc Suppl* 6: 299–314
24. Dollar P, Wojek C, Schiele B, et al. (2012) Pedestrian detection: An evaluation of the state of the art. *IEEE T Pattern Anal* 34: 743.
25. Redmon J, Divvala S, Girshick R, et al. (2016) You only look once: Unified, real-time object detection. *Comput Vision Pattern Recognit* 2016: 779–788.
26. He K, Zhang X, Ren S, et al. (2016) Deep residual learning for image recognition. IEEE conference on *Compute Vison and Pattern Recogn*, 770–778.
27. Govind D, Lutnick B, Tomaszewski JE, et al. (2018). Automated erythrocyte detection and classification from whole slide images. *J Med Imag* 5: 027501.



AIMS Press

© 2018 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)