

Threshold Estimation Based on Local Minima for Nucleus and Cytoplasm Segmentation

Simeon Mayala (✉ mayalasimeon@gmail.com)

University of Bergen

Jonas Bull Haugsøen

University of Bergen

Research Article

Keywords: Segmentation, Local minima, Nucleus and Cytoplasm

Posted Date: November 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1017730/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Medical Imaging on April 26th, 2022.
See the published version at <https://doi.org/10.1186/s12880-022-00801-w>.

RESEARCH

Threshold Estimation Based on Local Minima for Nucleus and Cytoplasm Segmentation

Simeon Mayala^{1*} and Jonas Bull Haugsøen^{2,3}

*Correspondence:

simeon.mayala@uib.no

¹Department of Mathematics,

University of Bergen, Allégaten

41, 5007 Bergen, Norway

Full list of author information is available at the end of the article

Abstract

Background: Image segmentation is a process of partitioning the input image into its separate objects or regions. It is an essential step in image processing to segment the regions of interest (ROI) for further processing. We propose a method for segmenting nucleus and cytoplasm from the white blood cells (WBC).

Methods: Initially, the method computes an initial value based on the minimum and maximum values of the input image. Then, a histogram of the input image is computed and then approximated to obtain function values. The method searches for the first local maximum and local minimum from the approximated function values. We approximate the required threshold from the first local minimum and the computed initial value based on defined conditions. The threshold is applied to the input image to binarize it and then perform post-processing to obtain the final segmented nucleus. We segment the whole WBC before segmenting the cytoplasm depending on the complexity of the objects in the image. For WBCs which are well separated from the RBCs, n thresholds are generated and then produce n thresholded images. Then, a standard Otsu method is used to binarize the average of the produced images. Morphological operations are applied on the binarized image and then use a single-pixel point from the segmented nucleus to segment the WBCs. For images in which RBCs touch the WBCs, we segment the whole WBC using SLIC and watershed methods. The cytoplasm is obtained by subtracting the segmented nucleus from the segmented WBC.

Results: The method is tested on two different public data sets. Performance analysis is done and the results show that the proposed method segments well the nucleus and cytoplasm.

Conclusion: We propose a method for nuclei and cytoplasm segmentation based on the local minima of the approximated function values from the histogram of image. The method has demonstrated its utility in segmenting nuclei, WBCs and cytoplasm and the results are reasonable.

Keywords: Segmentation; Local minima; Nucleus and Cytoplasm

Background

Leukocytes (white blood cells) are cells of the immune system that take part in the body's defense against infectious disease and foreign material [1, 2]. Leukocytes can be categorized based on structure (granulocytes or agranulocytes) and cell lineage (myeloid or lymphoid cells). Broadly, there are five types of leukocytes; three types of granulocytes - neutrophils, eosinophils and basophils, and two types of agranulocytes - lymphocytes and monocytes. Granulocytes and monocytes are of myeloid lineage, whereas lymphocytes are, as the name implies, of lymphoid lineage.

In microscopic images of stained blood smears, leukocytes can be differentiated from erythrocytes (red blood cells) and platelets by having nuclei and their (in most cases) larger size. They also stain darker using common dyes such as hematoxylin and eosin (H&E). Three main characteristics are used to identify the different types of leukocytes - the shape of their nuclei, their granularity and their staining.

All the granulocytes are large, granular cells with lobulated nuclei [2]. Neutrophils stain neutrally and have nuclei with multiple (2-5) lobes. Eosinophils stain red and have nuclei with 2-4 lobes. Basophils stain blue and have nuclei with 2-3 lobes. The shapes of these lobes are characteristic for each cell type, as seen in figure 1. Monocytes are large agranular cells with kidney-shaped nuclei. Lymphocytes are also agranular and are smaller than the other leukocytes. Their nuclei are round, often eccentric, and stain dark blue. Examples of the different types of leukocytes can be seen in figure 1.

Figure 1 Normal leukocytes. The cells seen in the background are erythrocytes (red blood cells) identified by their red colour, thin (often almost transparent) center and lack of nuclei.

Different related studies that have worked on nucleus, cytoplasm and WBCs segmentation are presented here.

N. Ghane *et al.* [3], proposed a method whose main purpose was to segment WBCs from microscopic images. The method is a combination of thresholding, K-means clustering and a modified watershed algorithm. The segmentation is done by implementing the method in the following steps: segmentation of white blood cells (WBCs) from a microscopic image, extraction of nuclei from cell and then separation of overlapping cells and nuclei. The results obtained using the proposed method show high similarity measures, precision, and sensitivity for both nuclei and segmented cells. Also, the result shows a high similarity between manual segmentation and the results obtained using the proposed method.

N. Theera-Umpon [4] proposed an automatic segmentation method that segments each cell into the nucleus, cytoplasm, and background. In this paper, a fuzzy C-means (FCM) algorithm and mathematical morphological operation are used to segment white blood cells. The morphological granulometric are applied to extract the shape and size of the objects. The performance of the proposed method is evaluated by comparing its results with the cells segmentation that have been obtained manually by an expert.

J. Prinyakupt *et al.* [2] proposed a system that performs multitask. First, the system pre-processes the images by locating the white blood cells within microscopic blood smear images. Then, the located WBC are segmented into nucleus and cytoplasm regions. In the segmentation step, the method uses WBC's morphological properties and the calibrated size of a real cell relative to image resolution in which thresholding, morphological operation and ellipse curve fitting are combined. The method extract suitable features from the segmented images. Finally, the method performs feature selection and then classifies the cells into basophil, eosinophil, neutrophil, lymphocyte and monocyte. Then, linear and naïve Bayes classifiers are applied to the selected prominent features to perform a comparison. The segmented results obtained using the proposed method is consistent with the

manually segmented images by a haematologist.

Y. Li *et al.* [5] proposed a dual-threshold method for segmenting white blood cells. The method is proposed based on a strategic combination of RGB and HSV colour space. In the preprocessing step the method obtains a contrast-stretched grey image and an H component image from transformed HSV color space. Then, a dual-threshold method is applied to improve the conventional single-threshold approach. The method searches for the optimal threshold using a golden section search method. Morphological operations and median filter are used to perform the post-processing. The method was tested on lymphoblasts on a public Acute Lymphoblastic Leukemia (ALL) image dataset and the obtained results indicate that the proposed method outperforms the single-threshold approach performed in RGB and HSV color space independently.

H. Miao and C. Xiao [6] presented a marker-controlled watershed algorithm for segmenting leukocyte and erythrocyte. The proposed method simultaneously extracts the two types of blood cells to simplify the operation and computation time. First, the method segments the cell nucleus and blood cell. The segmented leukocytes and erythrocytes are obtained simultaneously by classification and the results show that the proposed method is efficient.

M. Kuse *et al.* [7] presents a method for automating the detection of lymphocytes in histopathological images. The proposed method segments cells from the extracellular matrix which is a two steps process. The process is carried out on the HSV-equivalent of the image using mean shift-based clustering for color approximation. Then, a thresholding step follows in the HSV space. Features are extracted from the cells and then used to train a support vector machine (SVM) classifier which is used to classify lymphocytes and non-lymphocytes. For the overlapping lymphocytes, a contour-based overlap resolution method is applied to resolve the problem.

F. Sadeghian *et al.* [8] proposed a framework for WBC's segmentation. The main purpose was to segment the WBCs into the nucleus and cytoplasm. The framework involves several digital image processing algorithms. Twenty microscopic blood images were tested, and the results obtained from the proposed framework indicate that the framework can extract the nucleus and cytoplasm region in a WBC image sample.

A. Ismail and M. Marhaban [9] proposed a simple approach for fast calculation of the threshold value for automatic image thresholding. The method searches for the best threshold based on gradient analysis of the image histogram. In this approach, a histogram of the input image is computed. Then, the histogram function is differentiated before gradient analysis. The method successfully separates the object from the background. The results obtained are satisfactory.

Image processing is computerized processing of images [10] with a target of extracting specific information. During the processing of the image, noise information is removed because it would affect the analysis. Image segmentation is an essential step in image processing because it aims at segmenting the regions of interest (ROI) for further processing. The segmentation process represents the image in a simple form that is easier to understand and simple to analyze.

Different techniques are used in image segmentation. Some methods used for image segmentation include thresholding, region growing, region splitting, region merging, detection of boundary discontinuities, watershed segmentation, active contours [10, 11] and graph-based image segmentation [12, 13, 14]. Thresholding methods are widely used compared to other methods because they are simple to use. They can be categorized into either global or local thresholding [15, 16]. The global thresholding methods are simple and suitable for binarizing grayscale images, especially for images whose foreground and background are well-separated [17]. If $f(x, y)$ is the image to be segmented using a global threshold T , mathematically the segmented image is represented as

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq T \\ 0 & \text{if } f(x, y) < T \end{cases}$$

In practice, global thresholding depends on the histogram of the image. It is challenging to use global thresholding if the histogram of the image does not have a clear separation of foreground and background features. It is also challenging if the image is noisy and if the background intensity varies significantly across the image. Local thresholding is used to overcome the complexity of binarizing an image whose illumination is not uniform. Different thresholds are estimated locally for each pixel based on the local feature in the image, which differs from its immediate neighbors [17].

Our contribution in this work relies on the way the threshold is estimated. The proposed method estimates threshold based on the local minima of the approximated function values from the histogram of the input image. We run the method to compute an initial value from the input image. A histogram of the input image is computed and then approximated to obtain the function values. The method searches for the first local maximum and local minimum from the approximated function values. Then the initial value and the first local minimum are used for the final decision of the threshold for segmenting the nucleus. The method can be applied over a range of many images because thresholds are approximated automatically.

We also develop a simple strategy of segmenting the whole WBC whenever it is well separated from the RBCs. We generated n thresholds and each of the n values is applied to the input image to produce a thresholded image. The produced images are combined by taking their average and then a standard Otsu method is used to binarize it. Morphological operation is applied on the binarized image and finally, use a single-pixel point from the nucleus to segment the WBCs from the RBCs and other backgrounds. We obtain the cytoplasm by subtracting the segmented nucleus from the segmented WBC.

For images in which the RBCs touch the WBCs, we opt for classical techniques to separate the separate the touching objects. We use Simple Linear Iterative Clustering (SLIC) approach based on superpixel [18]. Since the focus is on the WBCs, we utilize their strength of boundary adherence of the superpixel to segment it. We use a single pixel point from the segmented nucleus to separate the WBC from other backgrounds. Similarly, the cytoplasm is obtained by subtracting the nucleus

from the segmented WBC. Note that parameter tuning could differ depending on the WCB to be segmented because of the staining.

We also apply watershed transformation to segment the WBCs when the RBCs touch each other [19]. Since the criterion of segmenting the WBC based on the homogeneity of the grey values is not helpful, we make use of the shape of the WBC. We allow an automatic choice of number of local maximum and take it if the non interesting objects (RBCs) appear to be detached from the WBC. We perform a postprocessing to remain with the WBC. We also tried compact watershed by inputting gradient of the image and choose the markers and compactness manually. Both options works but the choice of the input in watershed can influence the quality of the results.

The rest of the paper is organized as follows: Section "Methods" presents a detailed description of methodological ideas for segmenting the nucleus and WBC. Section "Descriptions of Data" describes the data used and their availability. Section "Implementation of the proposed method and experimental results" gives a detailed description of the implementation procedures of the proposed method for both nucleus, WBC and then cytoplasm segmentation. Section "Quantitative analysis to Measure the Performance of the Proposed Method" presents a quantitative analysis of similarity indexes by using Jaccard index (JI) and Dice similarity coefficients (DSC). Section "Discussion" gives discussions and section "Conclusion" gives concluding remarks. Section "Supplementary material and other results" consists of more segmentation results and supplementary materials for methodological ideas.

Methods

In this section, we establish a method for segmenting the nucleus and cytoplasm based on the distribution of the intensity values of the input image. The technique estimates the threshold automatically based on the local minima of the input image.

Threshold Estimation for Nucleus Segmentation

Assuming that a gray-scale image has intensity values that can be classified into several gray levels. Let a gray-scale image A be a matrix of $M \times N$ dimension such that $f(x, y)$ gives an intensity value at position (x, y) . We define,

$$A = \left[f(x, y) \right]_{M \times N} \text{ for } x = 0, 1, \dots, N - 1; y = 0, 1, \dots, M - 1 \quad (1)$$

Considering the input image A in equation (1), we compute a value T_{nc0} that will be used for searching a threshold value for segmenting the nucleus. The T_{nc0} value is computed from the input image by

$$T_{nc0}(n_1, n_2) = \frac{Max(A)}{n_1} + \frac{Min(A)}{n_2} \quad (2)$$

where $n_1, n_2 \in \mathbb{Z}$ greater than 0 (a simple analysis on how to choose n_1 and n_2 is presented in the subsection for supplementary material and other results). Assuming that T_{nc0} is an intensity value on the histogram of the input image A . We want to estimate T_{nc0} such that it is not one of the extreme values in the input image. So,

T_{nc0} is within the range of minimum and maximum intensity values in the input image. That is, $Min(A) < T_{nc0} < Max(A)$.

So, we represent the input image A in form that increases monotonically to locate the value T_{nc0} . Let \tilde{A} be a vectorized matrix of A , then

$$\tilde{A} = \left[f(x, y)_i \right]_{k \times 1}^T \quad \text{for } k = M * N; \quad i = 0, 1, \dots, k - 1; \quad (3)$$

$T = \text{transpose}$

k is a total number of pixels in the image A . The intensity values $f(x, y)$ at each position (x, y) in the vectorized matrix are

$$\tilde{A} = [f(x, y)_0, f(x, y)_1, f(x, y)_2, \dots, f(x, y)_{k-2}, f(x, y)_{k-1}]^T \quad (4)$$

We construct a non-decreasing sequence of values in \tilde{A} by using the intensity values $f(x, y)$ at each position (x, y) in \tilde{A} . Then, the sequence will be

$$\tilde{A}_{seq} = [f(x, y)_i \leq f(x, y)_{i+1} \leq \dots \leq f(x, y)_{k-2} \leq f(x, y)_{k-1}] \quad (5)$$

The constructed sequence of intensity values in \tilde{A}_{seq} are ordered sets and preserve a non-decreasing order, so it takes the form of a monotonically increasing function. Actually the form has been changed from the function of two variables $f(x, y)$ in image A into the function of one variable in \tilde{A}_{seq} . For each $f(x, y) \in \tilde{A}_{seq}$ we can write $f(x, y)_i \rightarrow I_i$ and

$$\tilde{A}_{seq} = [I_i \leq I_{i+1} \leq \dots \leq I_{k-2} \leq I_{k-1}] \quad (6)$$

Therefore, T_{nc0} is not necessarily one of the values in \tilde{A}_{seq} but it is within the range of \tilde{A}_{seq} . That is,

$$\tilde{A}_{seq} = [I_i \dots, T_{nc0}, \dots, I_{k-1}] \quad (7)$$

The threshold $\tilde{\epsilon}_t$ will be estimated after representing the intensity values of the input image in the form of a histogram. T_{nc0} together with some conditions will be used to approximate the required threshold value.

Approximating threshold $\tilde{\epsilon}_t$

Let $h(x)$ be the function representing the histogram of an image A , where $x =$ intensity value I , and $h(x) =$ frequency of the intensity values.

Figure 2 Histogram. Histogram of the input image represented by the function $h(x)$.

Let $\tilde{h}(x)$ be the approximation of $h(x)$. We approximate $h(x)$ without distorting the general tendency of the histogram function. We can decompose $\tilde{h}(x)$ into

$$\tilde{h}(x) = \tilde{h}_1(x) + \tilde{h}_2(x) \quad (8)$$

Figure 3 Estimating histogram. The smooth line represents $\tilde{h}_1(x)$, and the unsmooth line represents the original data points $h_1(x)$.

Figure 4 Estimated histogram. Figure showing the approximation $\tilde{h}_1(x)$ after removing the long flat tail. It shows the turning point (local minimum and maximum of interest).

Computation of Local Minima and Maxima

Given points $(x_i, \tilde{h}_1(x_i))_{i=1}^m$ from the approximated histogram, we need to compute a spline g such that

$$g(x_i) = \tilde{h}_1(x_i), \text{ for } i = 1, \dots, m \quad (9)$$

The spline concept and basis spline (B-spline) is introduced in the subsection "Supplementary Material and Results". The concept is referred from [20, 21]. B-Spline has a local propagation and gives us the freedom to choose any degree we want. We are interested in the number of turning points produced from the spline $g(x)$. The spline is assumed to be smooth. So, a $g(x)$ polynomial of n degrees gives $n - 1$ turning points.

Setting d be the degree of $g(x)$, the turning points will be $(x_1, g(x_1)), \dots, (x_{d-1}, g(x_{d-1}))$. Let $(x_c, g(x_c)), (x_{c+1}, g(x_{c+1}))$ be first local maximum and minimum of the function $g(x)$ respectively.

Based on the filtered data points of figure 4, the threshold $\tilde{\varepsilon}_t$ is expected to be closer to the first local minimum of the function $g(x)$. Also, the threshold $\tilde{\varepsilon}_t$ must appear after the first local maximum of the function $g(x)$ in the order of increase of the knots.

Also, from equation 2, simple sensitivity analysis of T_{nc0} in the supplementary materials guarantees that the value of T_{nc0} will not be very far from $\frac{1}{3}$. We introduce a parameter Er which is a small value chosen to control T_{cn0} and the local minimum value during threshold estimation. Then, the threshold $\tilde{\varepsilon}_t$ is obtained by checking the following conditions:

- 1 $x_c < T_{nc0}$.
- 2 If $|T_{nc0} - x_{c+1}| < Er$ then,

$$\tilde{\varepsilon}_t = \frac{T_{nc0} + x_{c+1}}{2}.$$

- 3 If $|T_{nc0} - x_{c+1}| > Er$ and $T_{nc0} > x_{c+1}$ then,

$$\tilde{\varepsilon}_t = \frac{T_{nc0} + x_{c+1} + Er}{2}$$

- 4 If $|T_{nc0} - x_{c+1}| > Er$ and $T_{nc0} < x_{c+1}$ then,

$$\tilde{\varepsilon}_t = \frac{T_{nc0} + x_{c+1} - Er}{2}.$$

The threshold $\tilde{\epsilon}_t$ is applied on the input image A to produce image A_{recons} . We call it a reconstructed image. The reconstructed image is obtained mathematically by

$$A_{recons} = \begin{cases} \text{Min}(A) & \text{if } f(x, y) \leq \tilde{\epsilon}_t \\ f(x, y) & \text{elsewhere} \end{cases}$$

The reconstructed image A_{recons} is binarized to segment the nucleus of WBC. Then post-processing operations are applied to the binarized image to obtain the final segmented nucleus.

Threshold Estimation for WBC Segmentation

We develop a strategy for segmenting the WBC after preprocessing the input image. The input image is preprocessed by averaging n thresholded images. It is preprocessed to reduce the variability of intensity values in the WBC region. The n images are obtained using different thresholds produces in a defined range by using

$$T_{wbc}^i = \frac{\text{Max}(A) + T_{nc0}}{2} + er_i \quad (10)$$

where er_i is a step size from one threshold to another, $i = 0, 1, 2, \dots, n - 1$. It is a control parameter that is computed to separate the thresholds at a constant step within a defined interval. It is introduced to govern the indistinguishable color in the cytoplasm and the RBC.

Let a_u be an upper limit of the intensity values, a_l be a lower limit of the intensity values in the chosen interval. Let er_i be denoted by Δh . Let Δh be a step size between two consecutive thresholds T_{wbc}^i and T_{wbc}^{i+1} in the interval a_l and a_u . To generate n equal sub-intervals in the defined interval we can use

$$n = \frac{a_u - a_l}{\Delta h} \quad (11)$$

as $\Delta h \rightarrow 0$, $n \rightarrow \infty$ and $(a_u - a_l)$ is fixed. T_{wbc}^i will be generated at the interval Δh . So, T_{wbc}^i values at different i indices can be generated as follows

$$T_{wbc} = [a_l, a_l + \Delta h, a_l + 2 \Delta h, a_l + 3 \Delta h, \dots, a_l + (n - 1) \Delta h] \quad (12)$$

where $a_l + (n - 1) \Delta h = a_u$. So, for the interval $a_l \leq er_i \leq a_u$ divided into $n - 1$ equal subintervals is represented as

$$a_l = T_{wbc}^0 < T_{wbc}^1 < T_{wbc}^2 < \dots < T_{wbc}^{n-1} = a_u \quad (13)$$

where $T_{wbc}^i = a_l + i \Delta h$, for $i = 0, 1, 2, \dots, n - 1$ and $\Delta h = \frac{a_u - a_l}{n}$. Each T_{wbc}^i is used to produce one thresholded image from the input image A by

$$A_j = \begin{cases} T_{wbc}^i & \text{if } f(x, y) \leq T_{wbc}^i \\ f(x, y) & \text{elsewhere} \end{cases}$$

Since i varies from 0 to $n - 1$, then the reconstructed (preprocessed) image A_{recons} is produced by averaging the n images obtained using the n thresholds.

$$A_{recons} = \frac{1}{n} [A_0 + A_1 + A_2 + \cdots + A_{n-1}] \quad (14)$$

Equation 14 can be rewritten in a more compact way to obtain

$$A_{recons} = \frac{1}{n} \sum_{j=0}^{n-1} A_j \quad (15)$$

The reconstructed image reduces the contrast of the intensity values in the WBC region and the plasma membrane of the RBC. The reconstructed image is binarized and then segment WBC from the RBC and other backgrounds. Note that this strategy works well for images whose WBCs are well separated from the RBCs. It also depends on the choice of n . For images in which the WBCs touch the RBCs we opt for classical techniques to segment the WBCs. In both cases the cytoplasm is obtained by subtracting the segmented nucleus from the segmented WBC. The summarized steps for nucleus, WBC and cytoplasm segmentation are in figure 5.

Figure 5 Schematic diagram. A block-diagram summarizing the steps for nucleus, white blood cells (WBC) and cytoplasm segmentation.

Descriptions of Data

The proposed method is tested on two different WBC images datasets. The two datasets are significantly different from each other in terms of the colors of cytoplasm and background. The first image dataset contains a total of 17,092 images of individual normal cells, which were acquired using the analyzer CellaVision DM96 in the Core Laboratory at the Hospital Clinic of Barcelona [22]. The dataset is organized into eight different groups. It includes neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes, erythroblasts and platelets (thrombocytes). The immature granulocytes include (promyelocytes, myelocytes, and metamyelocytes). The datasets were collected from different individuals. At the time of blood collection, the individuals were free from any pharmacologic treatment and they were without infection, hematologic or oncologic disease. As described in [22], it is a high-quality labelled dataset that can be used for benchmarking, training and testing models. They are JPG images with a size of (360×363) pixels each image. For more information about this dataset refer [22].

The second image dataset used in [23] contains 300 images together with the manual segmented images. As stated in the article [23], the dataset was originally obtained from iangxi Tecom Science Corporation, China. The images are manually segmented by domain experts. The nuclei, cytoplasm and background including red blood cells are marked in white, gray and black respectively. The images were taken by a Motic Moticam Pro 252A optical microscope camera with a N800-D motorized auto-focus microscope, and the blood smears were processed with a newly-developed hematology reagent for rapid WBC staining [23]. They are 120×120 images of WBCs.

Implementation of the proposed method and experimental results

The proposed method was implemented by writing scripts in the python programming language. All the thresholds estimations and segmentation were in groups based on the two data sets used in this study.

Nucleus Segmentation

The inputs were converted into grayscale images from two groups of data. For each gray scale image, a value T_{nc0} is computed automatically using a equation 21 in the supplementary material. Then, a *numpy function histogram* is used for computing a histogram of the input image [24]. The histogram is initially decomposed into two equal parts and then run a script on the first part of decomposition to check and remove the long tail at the beginning. Then, a SciPy function *savgol filter* from [25] is applied to obtain the approximated function values. A *Bspline* (3^{rd} order polynomial) interpolation function is applied on the approximated function values and then use a *argrextrema* function to compute the local minima and maxima, both from from SciPy [25]. Note that if the local minima and maxima are not found in the first run, the length of decomposition will be incremented and start searching again until the first local maximum and minimum are found. The value of the first local maximum and first local minimum together with the value T_{nc0} are used for computing the threshold $\tilde{\epsilon}_t$ based on the given conditions in the methodology subsections. The threshold $\tilde{\epsilon}_t$ is applied on the input image to preprocess it by assigning a value equal to a minimum value in the input image if it is less than the threshold value. The preprocessed image is binarized to segment the nucleus. The binarized image is post-processed by applying the morphological operation to remove small objects and close holes if they exist. The result is the segmented nucleus of the input image. Some results for the implementation steps are visualized in figure 6.

Figure 6 Nucleus segmentation steps. First row: Original image of eosinophils, lymphocyte, monocyte and neutrophils images. Second row: First part of the decomposed histograms T_{nc0} and the local minimum in dotted lines. The full vertical line indicates the estimated threshold $\tilde{\epsilon}_t$ after applying the conditions presented in subsection "Approximating threshold $\tilde{\epsilon}_t$ ". The third row are binarized image using estimated threshold $\tilde{\epsilon}_t$. Fourth row: Segmented nucleus after removing small objects and then filling holes. The results were obtained by setting $Er = 0.07$.

WBC and Cytoplasm Segmentation

For each input image A , the computed value T_{nc0} is also used for estimating the threshold values for preprocessing the input image before WBC segmentation. We generate n values automatically in the interval $[Min(A), \frac{Max(A)}{3}]$ which are referred as control parameters er_i in equation 10. The thresholds are spaced equally in a chosen interval. Each threshold is applied to the input image to generate a preprocessed image by thresholding.

The n preprocessed images are averaged to obtain a reconstructed image A_{recons} . The reconstructed image is binarized using a standard Otsu thresholding method [26] and then morphological operations are applied to separate the WBC from the background and RBCs. From the segmented nucleus (explained in the previous section), we take a single pixel value and use it to extract the whole WBC from the

binarized image. The cytoplasm is obtained by subtracting the segmented nucleus from the WBC. Figure 7 visualizes some images summarizing the steps for nucleus and WBC segmentation.

Figure 7 Cytoplasm and cell segmentation steps. From the left, First column: input images, Second column: reconstructed images and then binarized using estimated threshold $\tilde{\epsilon}_t$ for nucleus segmentation. Third column: Segmented nucleus obtained after removing small objects and then filling holes. Fourth column: reconstructed images by averaging different thresholded images using T_{wbc}^n thresholds in equation 10. Fifth column: Assigning new labels to the objects and remove some objects. Sixth column: Segmented WBC (separated from other objects using a single-pixel value from the nucleus). Seventh column: Segmented nucleus and cytoplasm. From top to down: the first row is the neutrophil, the second row is the eosinophil, the third row is the monocyte, the fourth row is the lymphocytes.

Note that the goal in this subsection is to segment the whole WBC which is then used to segment the cytoplasm. Results presented in Figure 7 show that the method is effective whenever the RBCs do not touch the WBCs. So, the binarization of the A_{recons} will definitely not work as intended. For images in which RBCs touch the WBC we opt for two classical approaches and use them to segment the WBC.

We apply Simple Linear Iterative Clustering (SLIC) approach based on superpixel [18] for segmenting the WBC when the RBCs touch the WBCs. The general considered condition for superpixel algorithm include: *partition* (the superpixels should be disjoint and each pixel should be assigned to a superpixel), *connectivity* - each superpixel should represent a connected set of pixels, *boundary adherence* - the superpixels should preserve object boundaries present in the original image, *compactness* - where no boundaries are present in an image, the superpixels should be compact, placed regularly and have smooth boundaries and the number of superpixels should be controllable by the user [27]. To achieve our goal we ignore some conditions and focus on boundaries of ROI. Since the algorithm adheres to boundaries, we focus on identifying the WBCs boundaries whenever the RBCs touch the WBCs. We allow under segmentation to segment well the WBCs. The results are visulized in Figure 8 fifth column. The summary of SLIC is found in the subsection of "Supplementary Material and other results".

We also apply watershed transformation for segmenting the WBCs when the RBCs touch the WBCs. The approach is a transformation on grayscale images [19]. It aims at segmenting the image when the edges of two objects of interest touch each other. In this case the criterion of segmenting the WBC based on the homogeneity of the grey values of is not helpful. We make use of the shape of the WBC. We concentrated in the marker construction step where it is easier to control it. We allow an automatic choice of number of local maximum and take it if the the RBCs are detached from the WBC. In this segmentation step we do not care much about the over segmentation of the ROI. We only care on detouching the non interesting objects from the object of interest (WBC).

Table 1 Summary for 976 sample images (251 eosinophils, 201 neutrophils, 208 monocytes, and 316 lymphocytes).

	Eosinophils	Neutrophils	Monocyte	Lympocytes
Thresholds mean for Nucleus segmentation	0.338991314	0.356037820	0.359911376	0.362329243
Thresholds mean for WBC segmentation	0.894988532	0.887716863	0.892829014	0.890228957
Contrast Mean between nucleus and cytoplasm	0.555997217	0.531679043	0.532917638	0.527899714

Figure 8 Cell segmentation using SLIC and marker-controlled watershed. From the left, First column: input images, Second column: converted into gray-scale and then binarized using Otsu thresholding method. Third column: marker-controlled watershed segmentation using a distance transformation. Fourth column: Segmented WBC after preprocessing the watershed results. Fifth column: segmentation of input image using SLIC. Sixth column: Segmented WBC after preprocessing results from SLIC. Seventh column: Segmented nucleus and cytoplasm.

Figure 8 gives some results of segmented WBCs using watershed technique and SLIC. To segment the whole WBC using watershed, it depends on the input. In this part we tested two different prepared inputs in the watershed, that is image gradient and Otsu binarized images. Both options work but the challenging part is to control the markers. Looking at the third column in figure 8, it takes time to identify the labels and then remove the unwanted ones.

The results in the fourth column figure 8 were obtained using SLIC. For our purposed SLIC seems to be straight because it identifies the boundaries of WBCs. We use a single pixel point from the segmented nucleus to segment the WBC from SLIC results. In comparison to table 1, we observed that parameter tuning in SLIC to control the WBC boundaries is easy for eosinophils. Possibly because the intensity values in the nuclei of eosinophils WBCs are very low compared to the nuclei in Neutrophils, Monocyte and Lymphocytes. This can be easily identified even from their staining by looking at their original images. For monocyte, lymphocyte, and neutrophils, it is a bit challenging to tune parameters to segment the whole WBC without dividing it.

Quantitative analysis to Measure the Performance of the Proposed Method

We present a simple analysis to assess the efficiency and effectiveness of the proposed method. The performance evaluation is done by a quantitative measure of the segmentation results. We analyze the accuracy by evaluating the similarity between the segmentation results obtained using the proposed method and the ground truth results which were obtained using manual segmentation.

For each segmented image, we compute Jaccard Index (JI) and Dice similarity coefficients (DSC) to check the level of similarity between the ground truth and the segmentation from the proposed method. We compute the Jaccard Index by dividing the number of labels that intersect by the size of the union of the two sets of labels. The Jaccard Index (JI) is defined by

$$JI(L_M, L_P) = \frac{|L_M(x, y) \cap L_P(x, y)|}{|L_M(x, y) \cup L_P(x, y)|} \quad (16)$$

Also, we compute Dice Similarity coefficient (DSC) to assess the accuracy of the proposed method by checking the level of similarity between the manual segmentation results and the one obtained using the proposed method. The Dice Similarity Coefficient is defined by

$$DSC(L_M, L_P) = \frac{2|L_M(x, y) \cap L_P(x, y)|}{|L_M(x, y)| + |L_P(x, y)|} \quad (17)$$

where $|L_M(x, y)|$ is the number of labels in the ROI obtained using manual segmentation. $|L_P(x, y)|$ is the number of labels in ROI obtained using the proposed

method. $|L_M(x, y) \cap L_P(x, y)|$ is the number of labels appearing in the ROI from the proposed and the manual segmentation methods.

In the analysis performed in this part, we use the second data set used in [23]. In this section, we use this data set because the first data set used in [22] does not have the ground truth images. In the analysis, some images were removed because of different reasons. Some of the images not considered in this section include the ones visualized in figure 9 showing the original and the manually segmented images.

Figure 9 Sample of the excluded images. From the left: (a), (c) and (e) are input images, whereas (b), (d) and (f) are their corresponding Manual segmentations of the original images. Images of this kind were not considered in the quantitative analysis. Each input image seems to have two WBC (one is showing only half of it) but the ground truth considered one cell which is at the centre. Since the suggested approach is based on the intensity values it segments both WBCs. These images were segmented together with other images but were not considered in the quantitative analysis.

A quantitative evaluation was performed by comparing the level of similarity between the manual segmentation and the results obtained using the proposed method. The JI and DSC metrics range from 0 to 1. 0 indicate that there is no overlap between the segmented region from both methods and 1 shows that there was perfect overlap between manually segmented regions and the one obtained using the proposed method.

Table 2 Summary for the Jaccard Index (JI) and Dice similarity coefficients (DSC) for nucleus segmentation from 211 images.

$0.8 < JI \leq 0.9$	$JI > 0.9$	$0.8 < DSC \leq 0.9$	$DSC > 0.9$	JI(mean)	DSC(mean)
30(14.2%)	181(85.8%)	19(9.01%)	192 (90.99%)	0.937601707	0.94807970

Table 3 Summary for the Jaccard Index (JI) and Dice similarity coefficients (DSC) for WBCs segmentation from 211 images.

$0.8 < JI \leq 0.9$	$JI > 0.9$	$0.8 < DSC \leq 0.9$	$DSC > 0.9$	JI(mean)	DSC(mean)
34(16.1%)	177(83.88%)	6(2.84%)	205 (97.15%)	0.93349731	0.9701763

We also present graphs in figure 10 and 11 to show the correlation of the size of the segmented area (nucleus and WBC) for the manual segmentation results and the one obtained using the proposed method.

Figure 10 Correlation of the nuclei size. The figure shows the correlation for the size of the segmented nuclei using the proposed method to the size of the nuclei segmented manually.

Figure 11 Correlation of the WBCs size. The figure shows the correlation for the size of the segmented WBCs using the proposed method to the size of the WBCs segmented manually.

In the performance analysis, the implementation was done by writing scripts in the python programming language, and it was run on a PC processor (Core i7-8650U CPU @ 1.90GHz × 8). To process 40 images of 360 × 363 size it takes 13.06518197059 seconds, which is approximately 0.326629549 seconds are used to process one image.

Discussion

We proposed a method for segmenting the nucleus and cytoplasm of the WBCs based on the local minima. The nucleus and the WBC were segmented first and then the cytoplasm. The method estimates a threshold for binarizing the input image. Some post-processing techniques are applied to the binarized image to remove small objects and holes filling. The result obtained is the segmented nucleus. The uniqueness of our approach is based on the threshold estimation. The threshold estimation depends partially on the local minima and maxima concepts. It also depends on a value computed using an expression developed in this paper. We develop some conditions to allow a wide range of checking for a good approximation. The conditions suggest a better estimation even for images whose nuclei are not well separated from the background. The thresholds are computed automatically from the input image

Before segmenting the cytoplasm, we first segment the WBC. We know that the indistinguishable color of the cytoplasm of the WBC from the RBC causes a challenge of segmenting the cytoplasm [28]. For images in which the WBCs are well separated from the RBCs we use a simple numerical technique to reduce the contrast between the intensity values in the WBC region and the plasma membrane of the RBC. This step simplifies the segmentation of the whole WBC. The method identifies and segments the WBC independent of the nucleus. After the WBC segmentation, the approach uses a single pixel from the nucleus to extract the whole WBC from other backgrounds. The cytoplasm segmentation is obtained by subtracting the nucleus from the segmented WBC. For images in which the RBCs touch the WBCs we use classical techniques to segment the WBCs and then obtain cytoplasm.

We segmented individual normal cells acquired using the analyzer CellaVision DM96 in the Core Laboratory at the Hospital Clinic of Barcelona [22]. We also segmented images dataset obtained from Jiangxi Tecom Science Corporation China [23]. The two data sets differ significantly but the results obtained from both data sets indicate that the proposed method can be applied to different image data sets because it is simple and flexible.

We performed a visual representation of segmentation results obtained using the proposed method. A quantitative measure of the segmented results is performed to assess the efficiency and effectiveness of the proposed method. It has demonstrated its effectiveness in segmenting high quality and poor quality images. For images in which the RBCs do not touch the WBCs the method works well even if the cytoplasm color is indistinguishable from the RBCs color. For images in which the RBCs touch the WBCs, the method works well when the cytoplasm colour is distinguishable from the RBCs. Otherwise, other techniques for separating touching objects can be used whenever the cytoplasm color is indistinguishable from the RBCs colour and they touch each other. For images with very low contrast between the cytoplasm and background, the method can segment well the nucleus but not the cytoplasm.

Conclusion

The method segments WBCs after a special preprocessing of the input images. The proposed method has demonstrated its utility in segmenting nuclei, WBCs and

cytoplasm. The performance analysis indicates that the method gives reasonable results when compared to the manually segmented images.

Supplementary Material and Other Results

Sensitivity analysis of T_{nc0}

The concept of sensitivity analysis and parameter tuning for image segmentation has been used in [29]. We use the formula for normalized forward sensitivity index [30] of the variable T_{nc0} to a parameter n which is defined as

$$\Upsilon_n^{T_{nc0}} = \frac{\partial T_{nc0}}{\partial n} \times \frac{n}{T_{nc0}} \quad (18)$$

From equation 2

$$T_{nc0}(n_1, n_2) = \frac{Max(A)}{n_1} + \frac{Min(A)}{n_2}$$

For n_1

$$\Upsilon_{n_1}^{T_{nc0}} = \frac{\partial T_{nc0}}{\partial n_1} \times \frac{n_1}{T_{nc0}} = -\frac{Max(A)}{n_1^2} \times \frac{n_1}{T_{nc0}} \quad (19)$$

For n_2

$$\Upsilon_{n_2}^{T_{nc0}} = \frac{\partial T_{nc0}}{\partial n_2} \times \frac{n_2}{T_{nc0}} = -\frac{Min(A)}{n_2^2} \times \frac{n_2}{T_{nc0}} \quad (20)$$

By considering a series of images we investigate the convergence of $\Upsilon_{n_1}^{T_{nc0}}$ and $\Upsilon_{n_2}^{T_{nc0}}$ as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$ respectively.

Figure 12 Convergence of $\Upsilon_{n_1}^{T_{nc0}}$. Figure indicates that the values of $\Upsilon_{n_1}^{T_{nc0}}$ start stabilizing near $n_1 = 3$. The number of images tested is 8960.

Figure 13 Convergence of $\Upsilon_{n_2}^{T_{nc0}}$. Figure show that the values of $\Upsilon_{n_2}^{T_{nc0}}$ stabilizes at near $n_2 = 3$. The number of images tested is 8960.

For $n_1 = n_2 = 3$, equation 2 becomes

$$T_{nc0} \approx \frac{Max(A)}{3} + \frac{Min(A)}{3} \quad (21)$$

From equation 21, for images whose intensity values are in the intervals $[Min(A), Max(A)] = [0, 1]$, $T_{nc0} = \frac{1}{3}$.

For images whose intensity values are in the interval of either $(Min(A), Max(A)) = (0, 1)$, $(Min(A), Max(A)) = (0, 1]$ or $[Min(A), Max(A)) = [0, 1)$, either $T_{nc0} \leq \frac{1}{3}$ or $T_{nc0} \geq \frac{1}{3}$.

Looking at figures 12 and 13, the values of $\Upsilon_{n_2}^{T_{nc0}}$ and $\Upsilon_{n_1}^{T_{nc0}}$ show that n_1 is more sensitive compared to n_2 . So, n_1 needs to be chosen carefully.

Figure 14 Eosinophil images. Segmented nucleus and cytoplasm of white blood cells.

Figure 15 Neutrophils images. Segmented nucleus and cytoplasm of the white blood cells.

Figure 16 Lymphocytes images. Segmented nucleus and cytoplasm of white blood cells.

Figure 17 Monocyte images. Segmented nucleus and cytoplasm of the white blood cells.

B-Spline interpolation of the data points

A spline is a piecewise polynomial function that is constructed as a means to interpolate a set of data points. Each piece of the function is confined between knots $t_i < t_{i+1}$ with strict continuity conditions between each piece at the knots. These polynomials are strung together at the knots to form the completed curve. Now, B-splines (basis splines) are a set of piecewise functions which form a basis for the set of all possible spline functions. Any spline function can be described as a linear combination of B-splines [20, 21].

Let U be a set $m + 1$ non-decreasing numbers $t_0 < t_1 < t_2 < \dots < t_m$ over the interval which we want to interpolate the data. The t_i s are called knots and the set U is called knot vector. Then, the i^{th} B-spline function of degree 0 is given by:

$$B_i^0(x) = \begin{cases} 1 & \text{if } t_i < x < t_{i+1} \\ 0 & \text{elsewhere} \end{cases}$$

The subsequent higher level B-splines are constructed from the recurrence relation:

$$B_i^d(x) = \frac{x - t_i}{t_{i+d} - t_i} B_i^{d-1}(x) + \frac{t_{i+d+1} - x}{t_{i+d+1} - t_{i+1}} B_i^{d-1}(x) \quad (22)$$

Equation 22 creates the basis functions for a d degree spline. So, a linear combination of B-Splines or spline function is a combination of form

$$g(x) = \sum_i^n c_i^d B_i^d(x) \quad (23)$$

where c_i s are the coefficients of the basis functions. The computation of these coefficients can be found in the reference sources [20, 21]. A linear combination of these functions in equation 23 can be used to interpolate a data set. In order to interpolate the data, the B-spline curve must span the interval of the data points to be interpolated.

Summary of SLIC Algorithm

Mainly, the algorithm involves three steps which include initialization, assignment and enforce-connectivity. It performs K-means segmentation using 5 dimension

space of color information and image location. The RGB color space is transformed into *CIELAB* color space [31] and then each of the pixel images is mapped into (l, a, b, x, y) . The algorithm starts by initializing a k number of clusters which are sampled on a regular grid. Then in the assignment step each pixel p_i is associated with the nearest cluster centre. Different from the standard k-means, SLIC updates the cluster centres by searching in a certain neighbourhood and not on the whole image. But the update step depends on the convergence of the error term which is obtained by computing the residual error between the old and the new centres. In case there are disjoint pixels from the previous steps then enforcing connectivity is applied so as to assign them a nearest superpixel [18].

Declarations

Acknowledgements

Not applicable

Funding

This research work is funded by the university of Bergen, Norway.

Abbreviations

WBC: White Blood Cells, RBC: Red Blood Cells, SLIC: Simple Linear Iterative Clustering, H and E: hematoxylin and eosin, FCM: fuzzy C- means algorithm, ALL: Acute Lymphoblastic Leukemia, RGB: Red, Green and Blue, HSV: Hue Saturation Value, SVM: support vector machine, ROI: region of interest, JI: Jaccard index, DSC: Dice similarity coefficients, CIELAB: implies $L^*a^*b^*$ color space, L^* for perceptual lightness, a^* and b^* for the four unique colors of human vision.

Availability of data and materials

All data analysed during this study are available online. Links can be found in their corresponding published articles [22] and [23].

Ethics approval and consent to participate

Not applicable

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Authors' contributions

SM implemented the proposed method and run the analysis to generate the results. JBH reviewed the biological concepts and clarity of the data set used. Interpretation of the results was done by all authors. Drafting, writing, and proof-reading to approve the manuscript was done by all authors.

Author details

¹Department of Mathematics, University of Bergen, Allégaten 41, 5007 Bergen, Norway. ²Department of Clinical Medicine, Neuro-SysMed, University of Bergen, Jonas Lies veg 87, N-5021 Bergen, Norway. ³Department of Neurology, Neuro-SysMed, Haukeland University Hospital, Jonas Lies vei 71, 5053 Bergen, Norway.

References

1. Mathur, A., Tripathi, A.S., Kuse, M.: Scalable system for classification of white blood cells from leishman stained blood stain images. *Journal of pathology informatics* **4**(Suppl) (2013)
2. Prinyakupt, J., Pluempitiwiriyaew, C.: Segmentation of white blood cells and comparison of cell morphology by linear and naïve bayes classifiers. *Biomedical engineering online* **14**(1), 63 (2015)
3. Ghane, N., Vard, A., Talebi, A., Nematollahy, P.: Segmentation of white blood cells from microscopic images using a novel combination of k-means clustering and modified watershed algorithm. *Journal of medical signals and sensors* **7**(2), 92 (2017)
4. Theera-Umporn, N.: White blood cell segmentation and classification in microscopic bone marrow images. In: *International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 787–796 (2005). Springer
5. Li, Y., Zhu, R., Mi, L., Cao, Y., Yao, D.: Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method. *Computational and mathematical methods in medicine* **2016** (2016)
6. Miao, H., Xiao, C.: Simultaneous segmentation of leukocyte and erythrocyte in microscopic images using a marker-controlled watershed algorithm. *Computational and mathematical methods in medicine* **2018** (2018)
7. Kuse, M., Sharma, T., Gupta, S.: A classification scheme for lymphocyte segmentation in h&e stained histology images. In: *International Conference on Pattern Recognition*, pp. 235–243 (2010). Springer
8. Sadeghian, F., Seman, Z., Ramli, A.R., Kahar, B.H.A., Saripan, M.-I.: A framework for white blood cell segmentation in microscopic blood images using digital image processing. *Biological procedures online* **11**(1), 196 (2009)

9. Ismail, A., Marhaban, M.: A simple approach to determine the best threshold value for automatic image thresholding. In: 2009 IEEE International Conference on Signal and Image Processing Applications, pp. 162–166 (2009). IEEE
10. Hemalatha, R., Thamizhvani, T., Dhivya, A.J.A., Joseph, J.E., Babu, B., Chandrasekaran, R.: Active contour based segmentation techniques for medical image analysis. *Med Biol. Image Anal* **17** (2018)
11. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing* vol. Fourth Edition. 330 Hudson Street, New York, NY 10013, ??? (2008)
12. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *International journal of computer vision* **59**(2), 167–181 (2004)
13. Shi, J., Malik, J.: Normalized cuts and image segmentation. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 731–737 (1997). IEEE
14. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* **22**(8), 888–905 (2000)
15. Lazar, M., Hladnik, A.: Implementation of global and local thresholding algorithms in image segmentation of coloured prints. In: *35th International Research Conference IARIGAI*, vol. 35 (2008)
16. Singh, T.R., Roy, S., Singh, O.I., Sinam, T., Singh, K., et al.: A new local adaptive thresholding technique in binarization. *arXiv preprint arXiv:1201.5227* (2012)
17. Firdousi, R., Parveen, S.: Local thresholding techniques in image binarization. *International Journal Of Engineering And Computer Science* **3**(03) (2014)
18. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence* **34**(11), 2274–2282 (2012)
19. Beucher, S., et al.: The watershed transformation applied to image segmentation. *Scanning microscopy-supplement-*, 299–299 (1992)
20. Prautzsch, H., Boehm, W., Paluszny, M.: *Bézier and B-spline Techniques*. Springer, ??? (2002)
21. Lyche, T., Morken, K.: *Spline methods draft*. Department of Informatics, Center of Mathematics for Applications, University of Oslo, Oslo, 3–8 (2008)
22. Acevedo, A., Merino, A., Alférez, S., Molina, Á., Boldú, L., Rodellar, J.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data in Brief*, 105474 (2020)
23. Zheng, X., Wang, Y., Wang, G., Liu, J.: Fast and robust segmentation of white blood cell images by self-supervised learning. *Micron* **107**
24. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E.: Array programming with NumPy. *Nature* **585**, 357–362 (2020). doi:10.1038/s41586-020-2649-2
25. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. *Nature Methods* **17**, 261–272 (2020). doi:10.1038/s41592-019-0686-2
26. Liao, P.-S., Chen, T.-S., Chung, P.-C., et al.: A fast algorithm for multilevel thresholding. *J. Inf. Sci. Eng.* **17**(5), 713–727 (2001)
27. Stutz, D., Hermans, A., Leibe, B.: Superpixels: An evaluation of the state-of-the-art. *Computer Vision and Image Understanding* **166**, 1–27 (2018)
28. Zhang, C., Xiao, X., Li, X., Chen, Y.-J., Zhen, W., Chang, J., Zheng, C., Liu, Z.: White blood cell segmentation by color-space-based k-means clustering. *Sensors* **14**(9), 16128–16147 (2014)
29. Teodoro, G., Kurç, T.M., Taveira, L.F., Melo, A.C., Gao, Y., Kong, J., Saltz, J.H.: Algorithm sensitivity analysis and parameter tuning for tissue image segmentation pipelines. *Bioinformatics* **33**(7), 1064–1072 (2017)
30. Chitnis, N., Hyman, J.M., Cushing, J.M.: Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. *Bulletin of mathematical biology* **70**(5), 1272 (2008)
31. Pub, C.: *Technical Report 15: 2004: Colorimetry*. Vienna: CIE Central Bureau (2004)

Figures

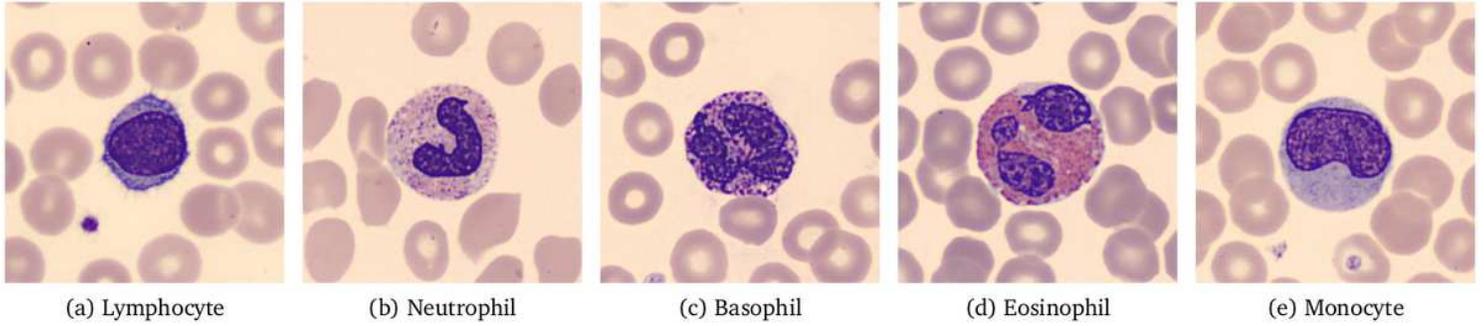


Figure 1

Normal leukocytes. The cells seen in the background are erythrocytes (red blood cells) identified by their red colour, thin (often almost transparent) center and lack of nuclei.

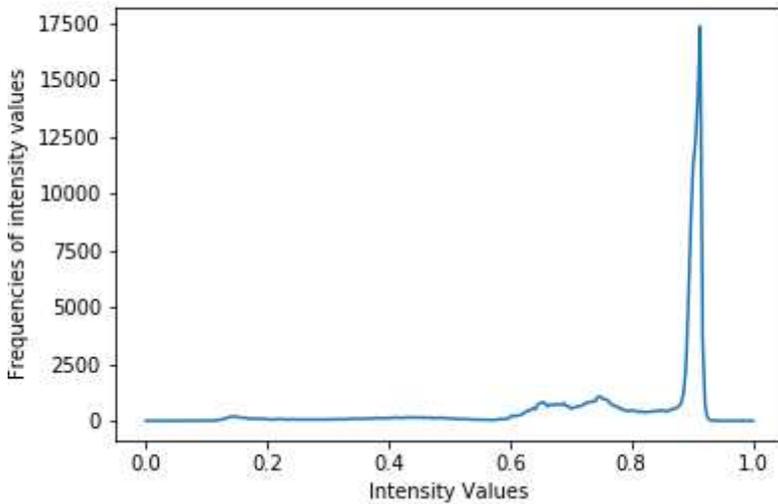


Figure 2

Histogram. Histogram of the input image represented by the function $h(x)$

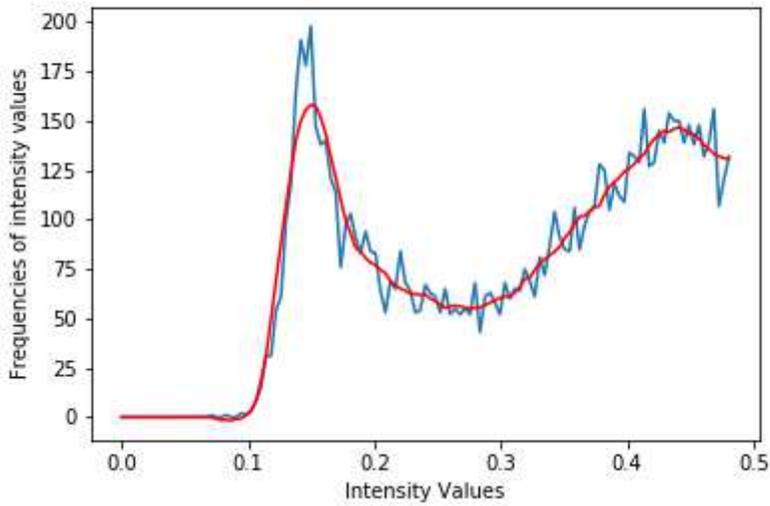


Figure 3

Estimating histogram. The smooth line represents $h^{-1}(x)$, and the unsmooth line represents the original data points $h_1(x)$.

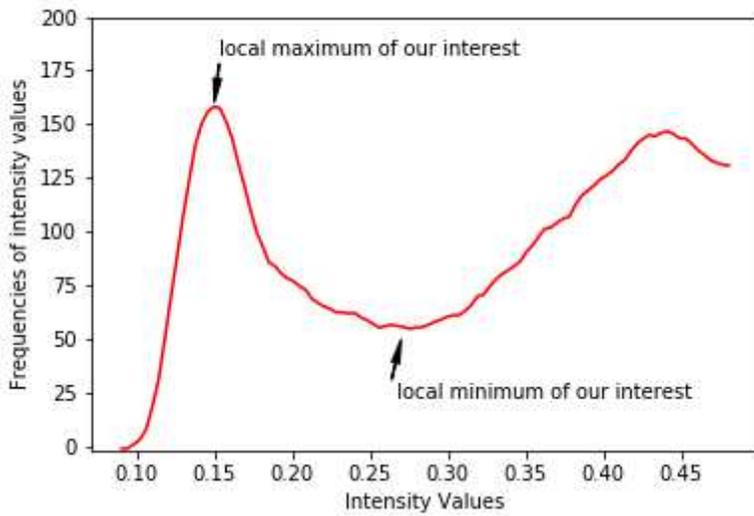


Figure 4

Estimated histogram. Figure showing the approximation $h^{-1}(x)$ after removing the long flat tail. It shows the turning point (local minimum and maximum of interest).

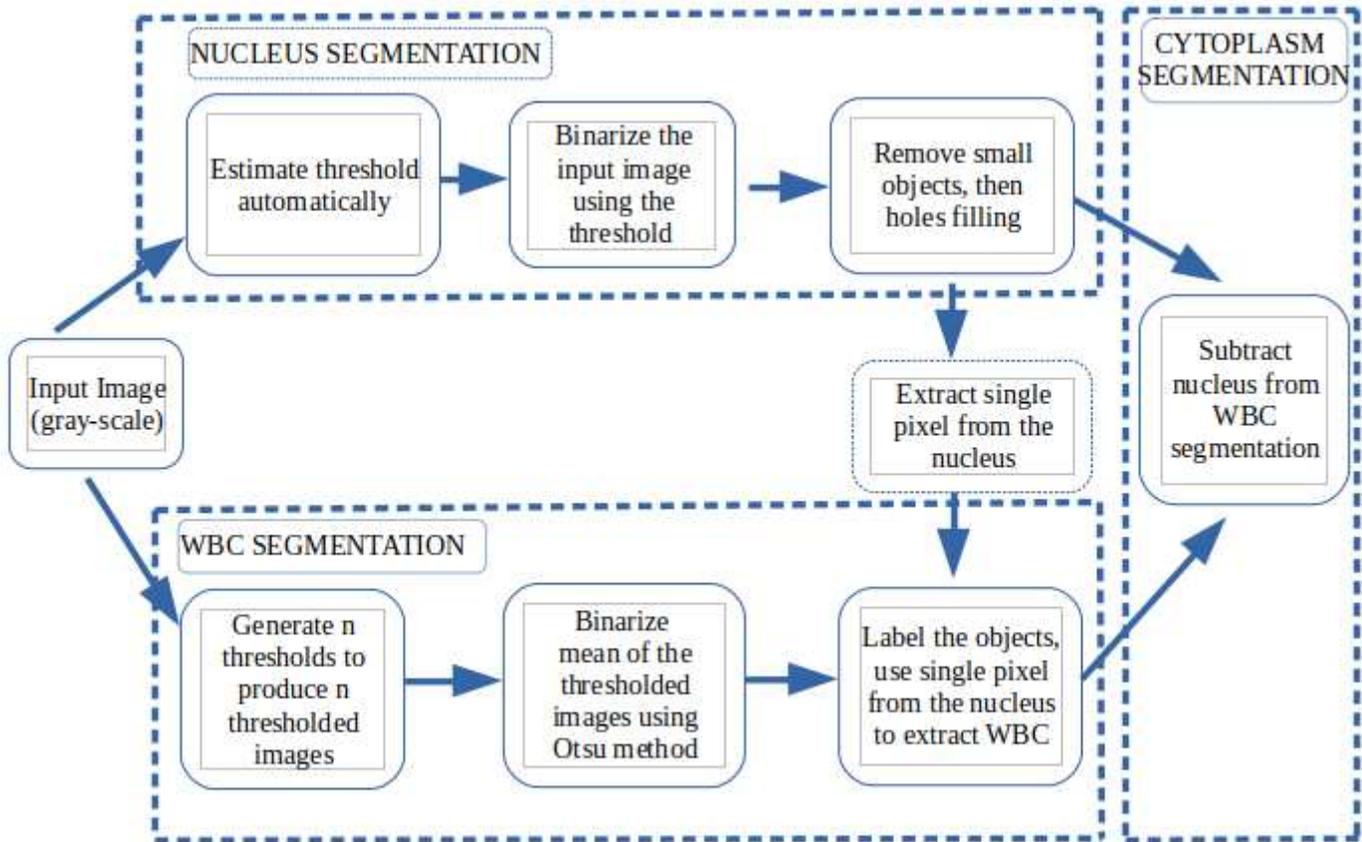


Figure 5

Schematic diagram. A block-diagram summarizing the steps for nucleus, white blood cells (WBC) and cytoplasm segmentation.

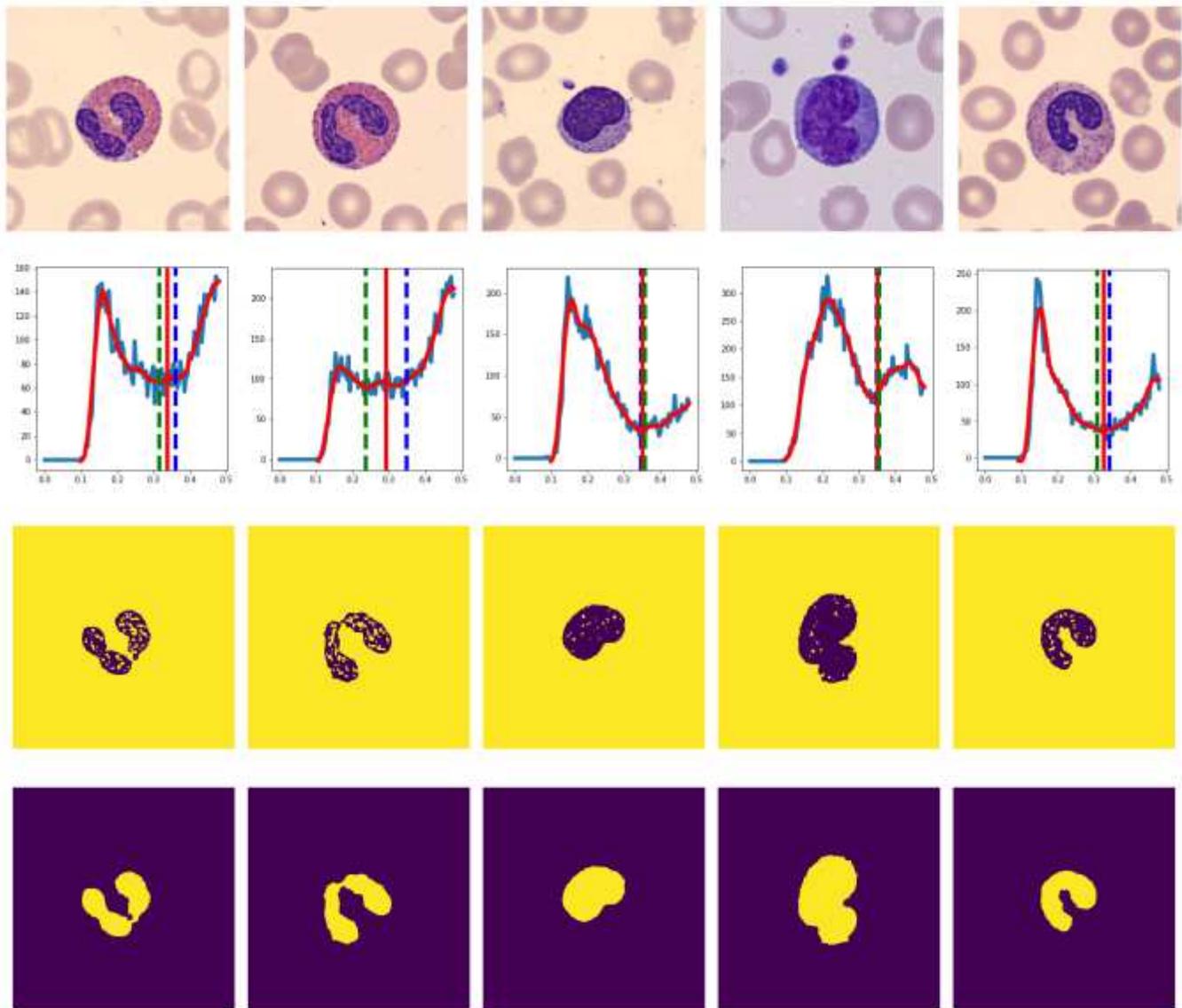


Figure 6

Please see the Manuscript PDF file for the complete figure caption

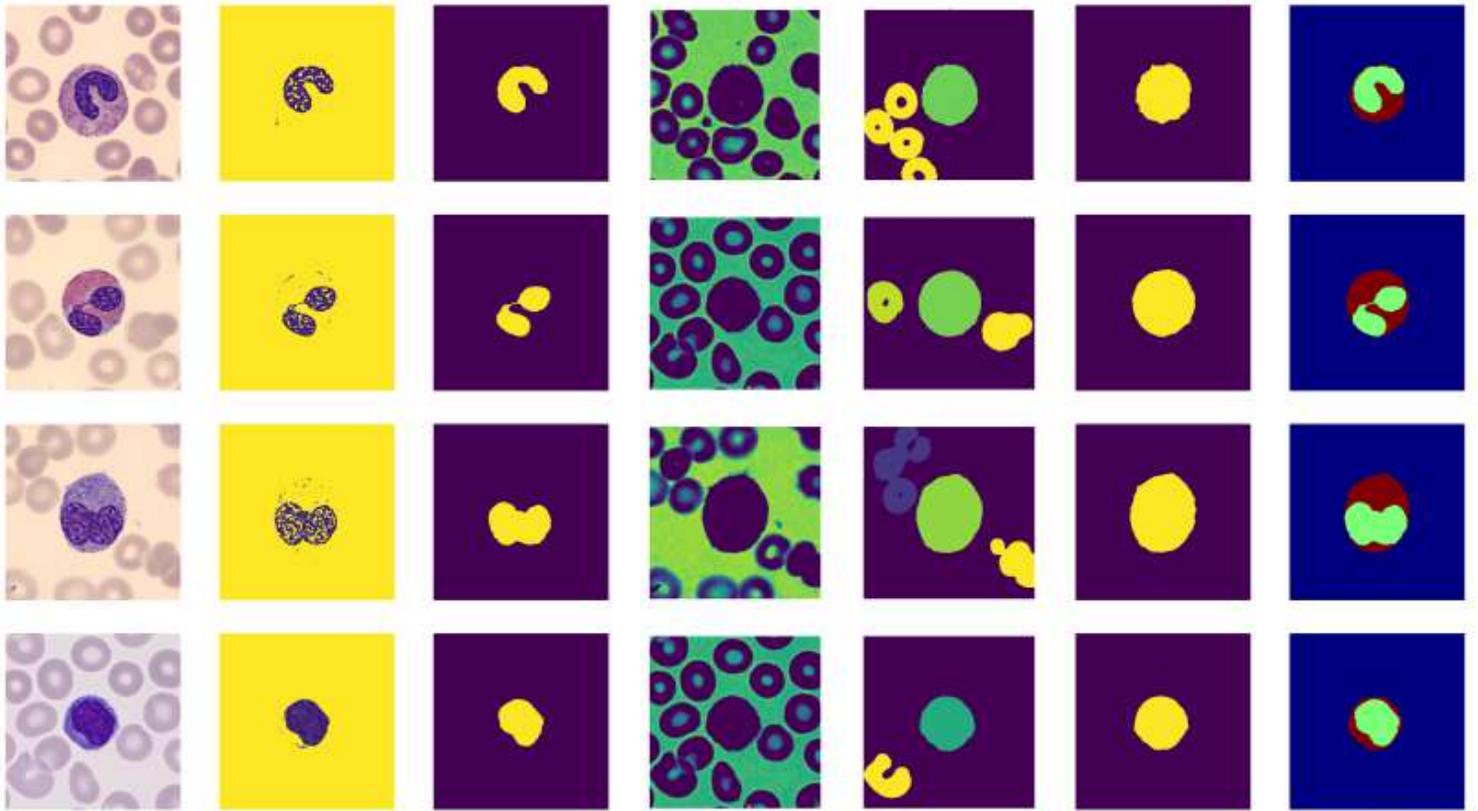


Figure 7

Please see the Manuscript PDF file for the complete figure caption

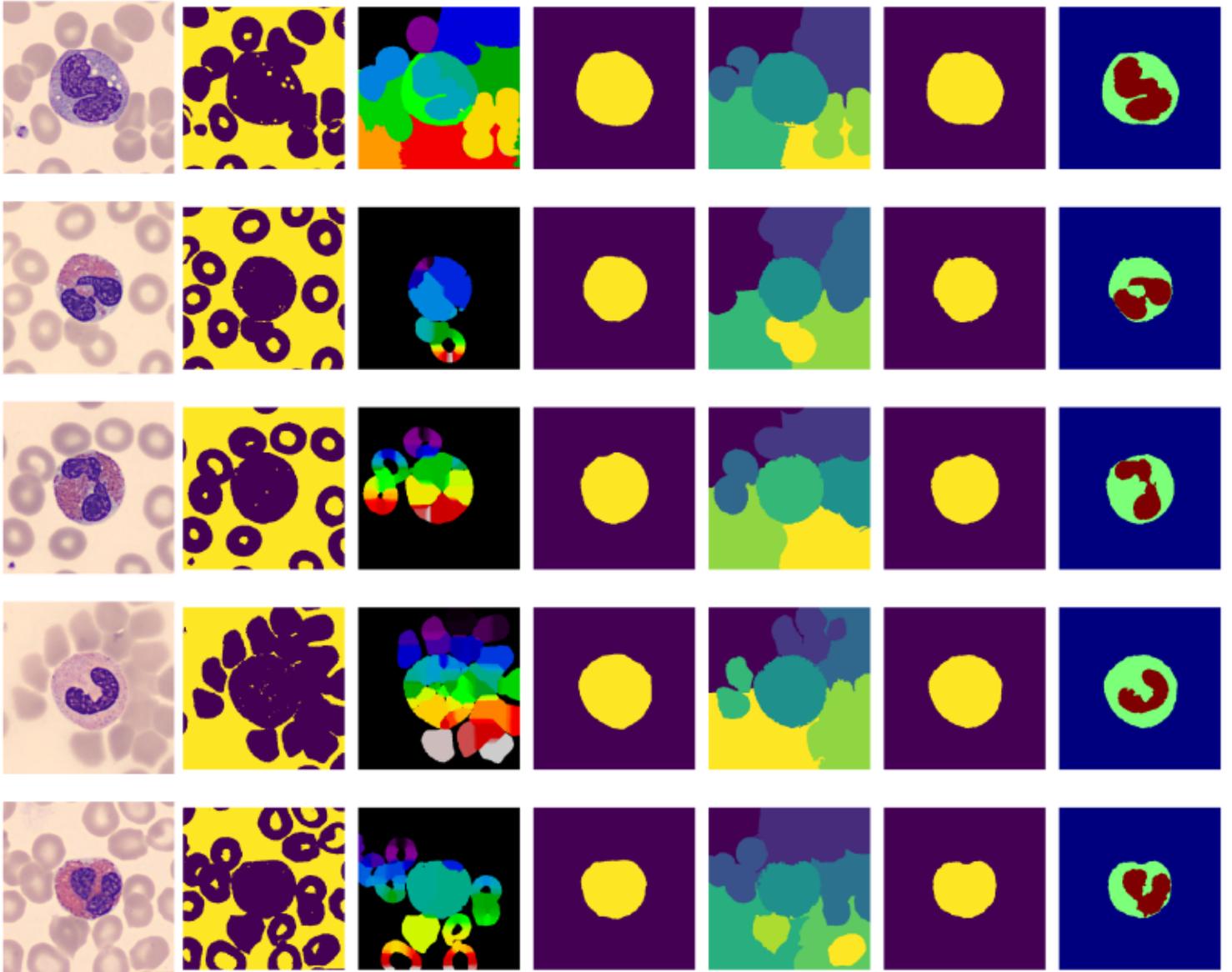
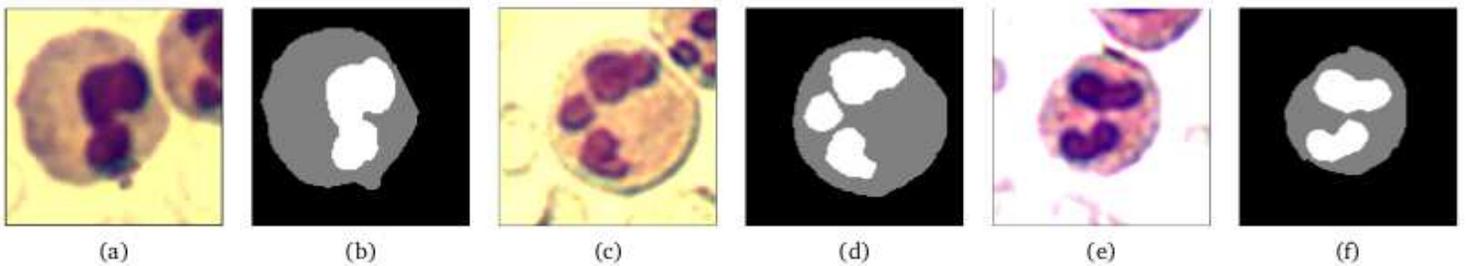


Figure 8

Cell segmentation using SLIC and marker-controlled watershed. From the left, First column: input images, Second column: converted into gray-scale and then binarized using Otsu thresholding method. Third column: marker-controlled watershed segmentation using a distance transformation. Fourth column: Segmented WBC after preprocessing the watershed results. Fifth column: segmentation of input image using SLIC. Sixth column: Segmented WBC after preprocessing results from SLIC. Seventh column: Segmented nucleus and cytoplasm.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 9

Sample of the excluded images. From the left: (a), (c) and (e) are input images, whereas (b), (d) and (f) are their corresponding Manual segmentations of the original images. Images of this kind were not considered in the quantitative analysis. Each input image seems to have two WBC (one is showing only half of it) but the ground truth considered one cell which is at the centre. Since the suggested approach is based on the intensity values it segments both WBCs. These images were segmented together with other images but were not considered in the quantitative analysis

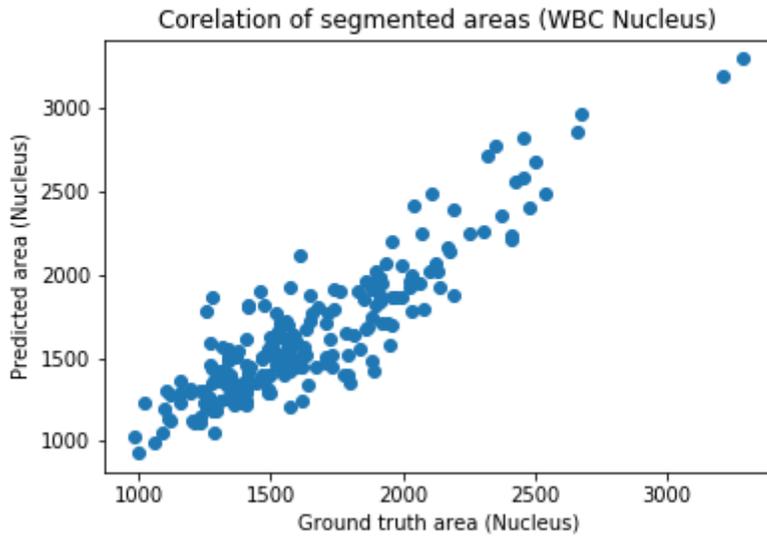


Figure 10

Correlation of the nuclei size. The figure shows the correlation for the size of the segmented nuclei using the proposed method to the size of the nuclei segmented manually.

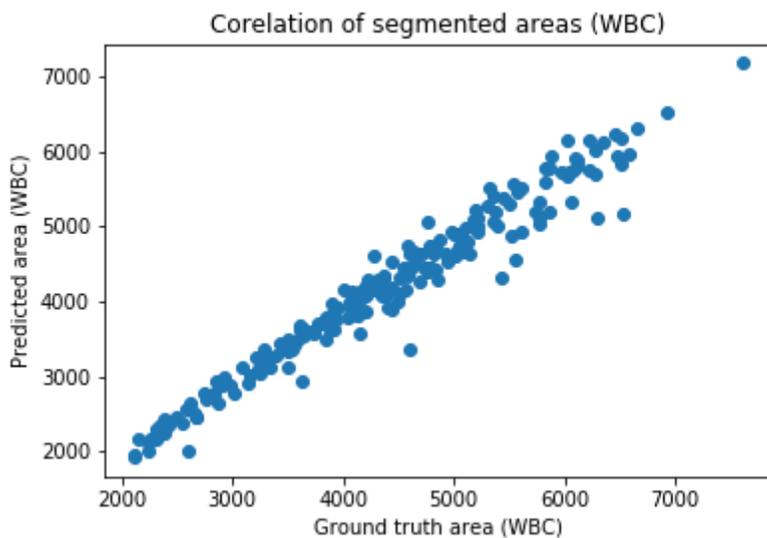


Figure 11

Correlation of the WBCs size. The figure shows the correlation for the size of the segmented WBCs using the proposed method to the size of the WBCs segmented manually.

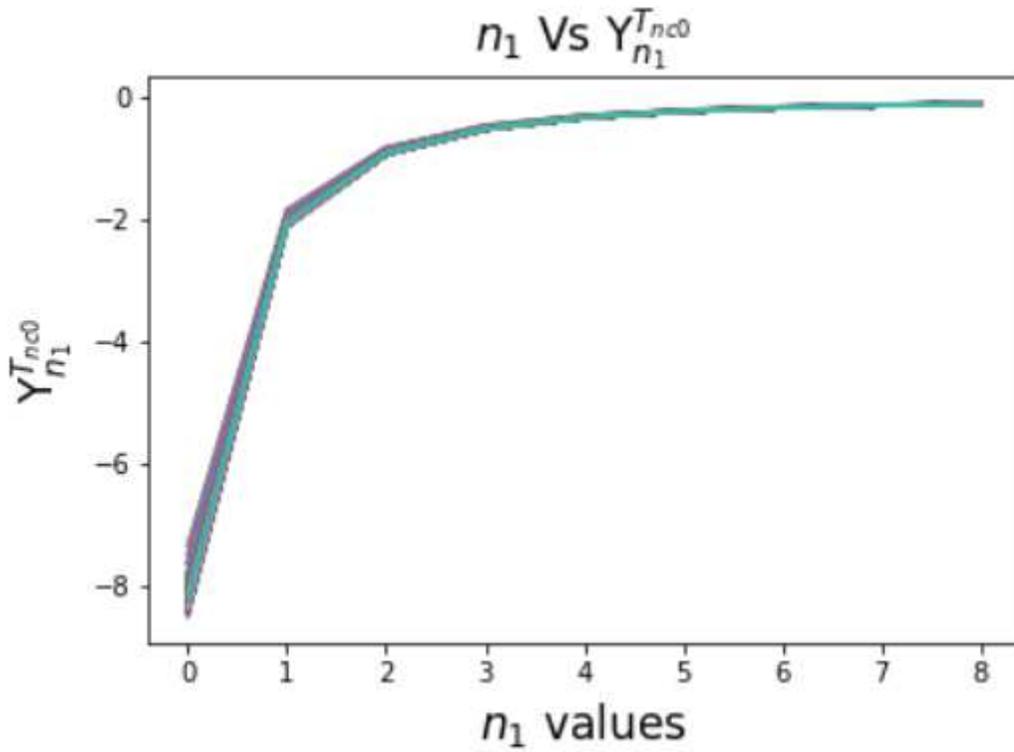


Figure 12

Please see the Manuscript PDF file for the complete figure caption

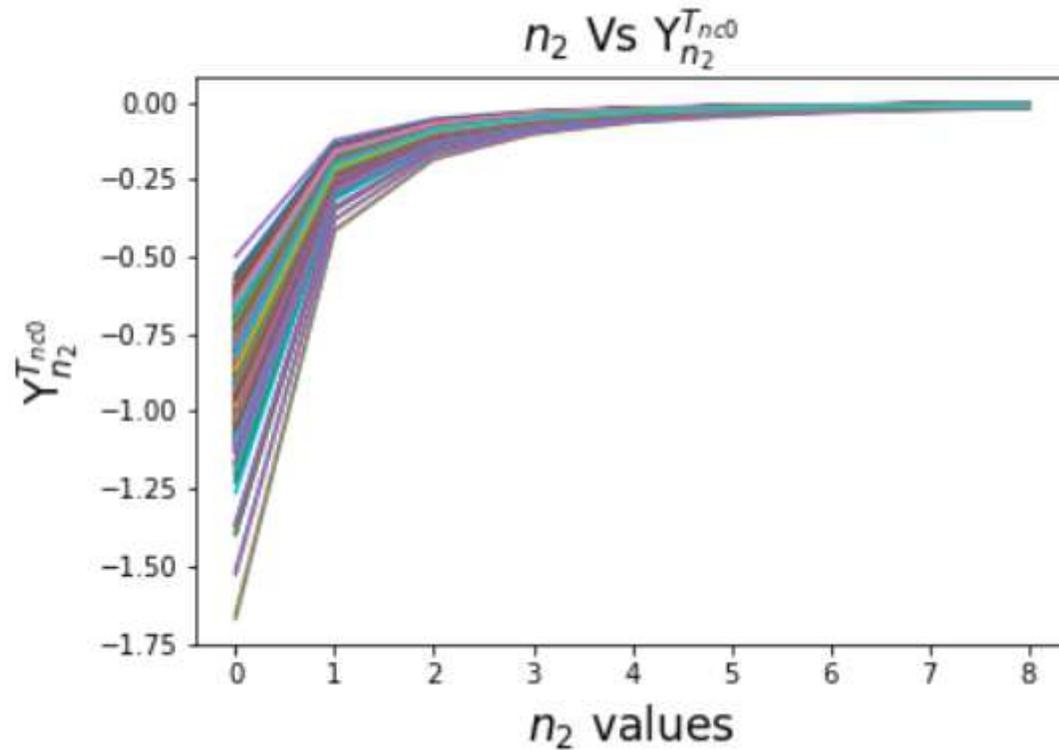


Figure 13

Please see the Manuscript PDF file for the complete figure caption

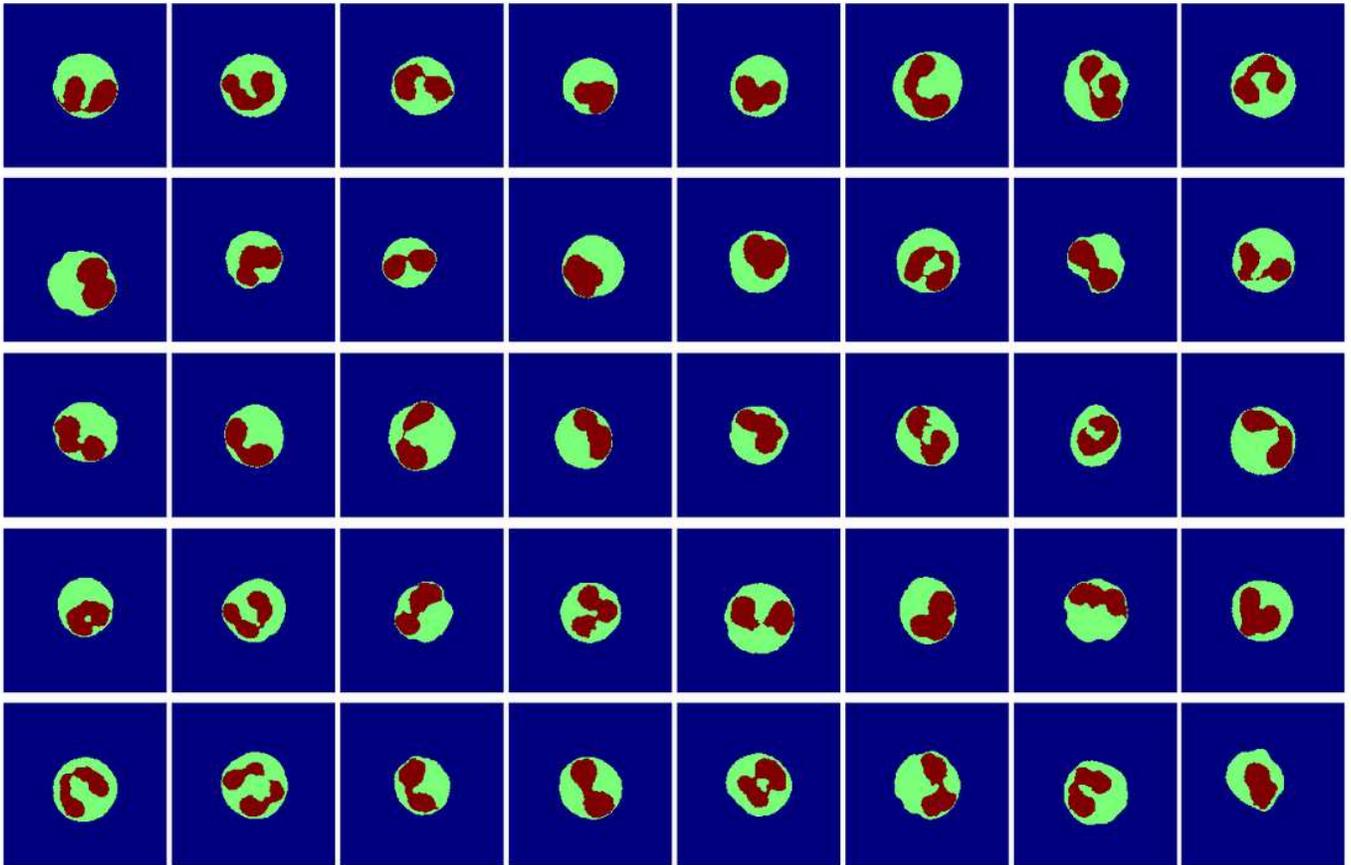


Figure 14

Eosinophil images. Segmented nucleus and cytoplasm of white blood cells.

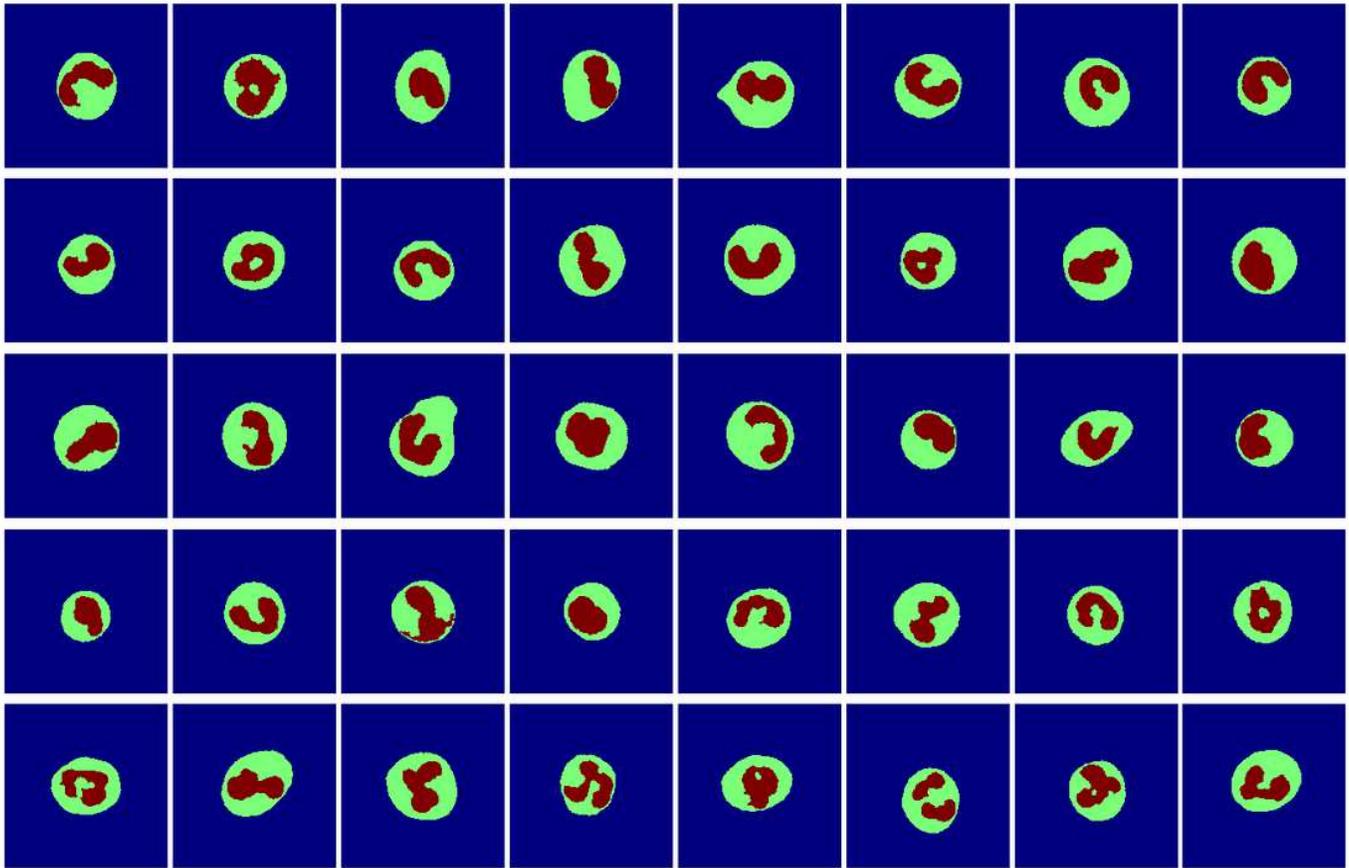


Figure 15

Neutrophils images. Segmented nucleus and cytoplasm of the white blood cells.

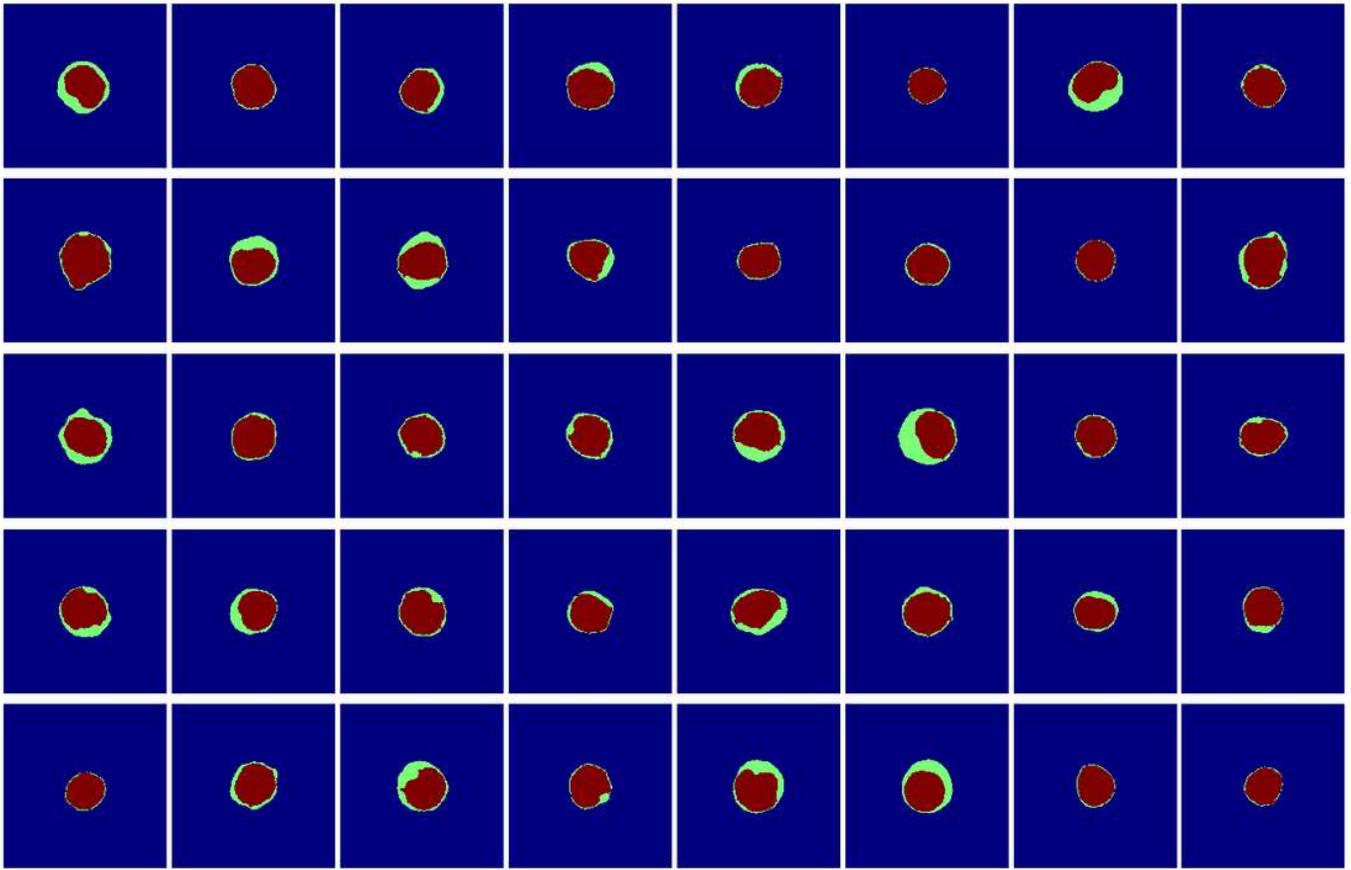


Figure 16

Lymphocytes images. Segmented nucleus and cytoplasm of white blood cells.

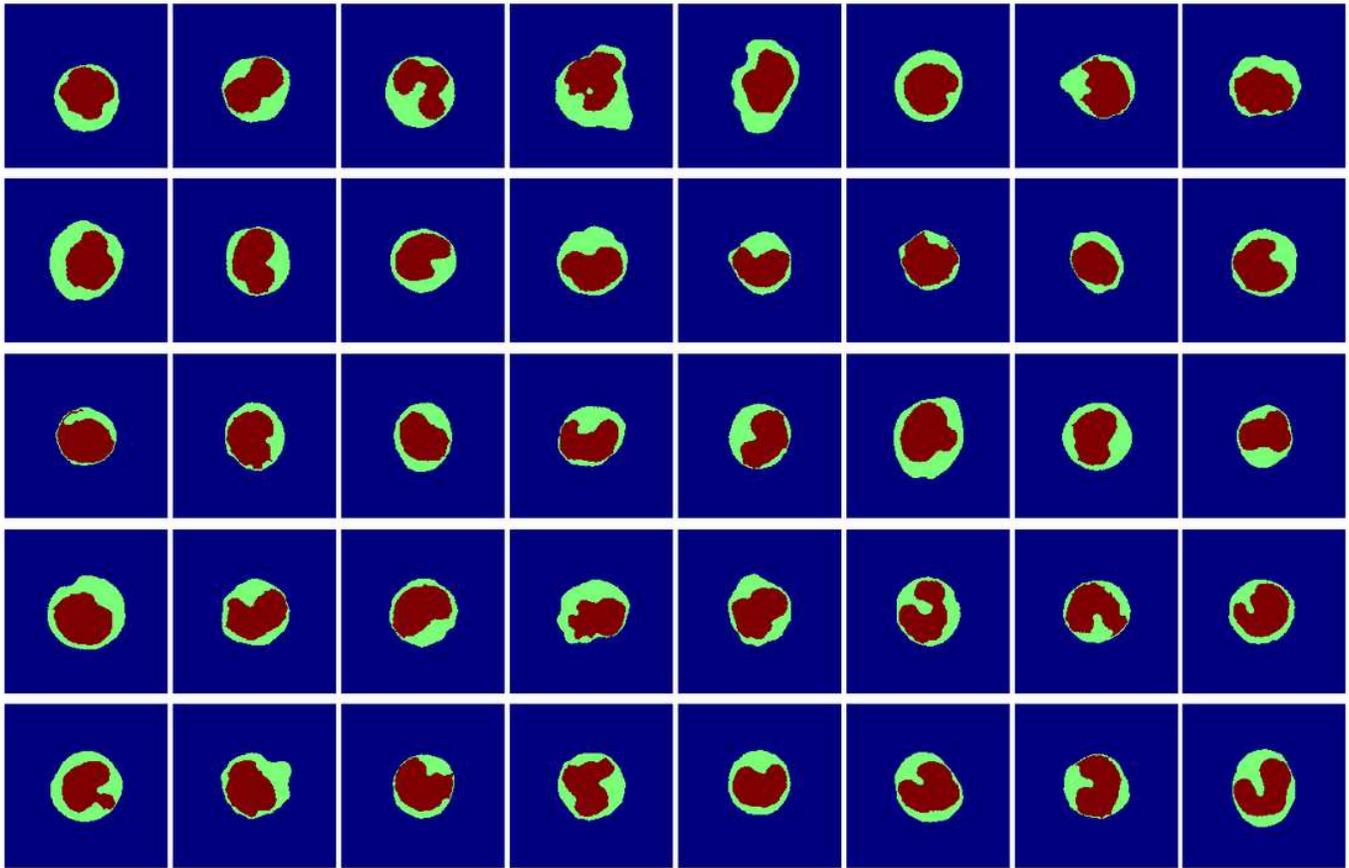


Figure 17

Monocyte images. Segmented nucleus and cytoplasm of the white blood cells.