

Chapter 4

Leukocytic Cell Nucleus Identification Using Boundary Cell Detection Algorithm with Dilation and Erosion Based Morphometry



Ishfaq Majeed Sheikh and Manzoor Ahmad Chachoo

Contents

1	Introduction	79
2	Literature Review	85
3	Methodology	87
4	Result and Discussions	94
5	Performance Measures	98
6	Conclusion	100
	References	101

1 Introduction

The important component of the human body is the immune system. It is the biological defence system that protects our body from different types of vulnerabilities: virus, bacteria, fungi, parasites and other pathogens. The immune system performs its function by the following components, cells (RBC (Red Blood Cells), WBC'S, T and B lymphocytes), organs and lymphoid tissue. Figure 4.1 describes the structure of the different cell types present in the human body. The cells of the immune system that defends the body from foreign substances are the leukocytes. The WBC'S help in fighting infections, killing harmful bacteria and defend the body from irrelevant substances by creating antibodies. They originate from the soft fatty fluid of the bone called the bone marrow. This process of generating cells in the bone marrow is called hematopoiesis. These cells grow from a hematopoietic stem where

I. M. Sheikh (✉) · M. A. Chachoo
PG Department of Computer Science, University of Kashmir, Srinagar, India
e-mail: ishfaqmajeed.scholar@kashmiruniversity.net

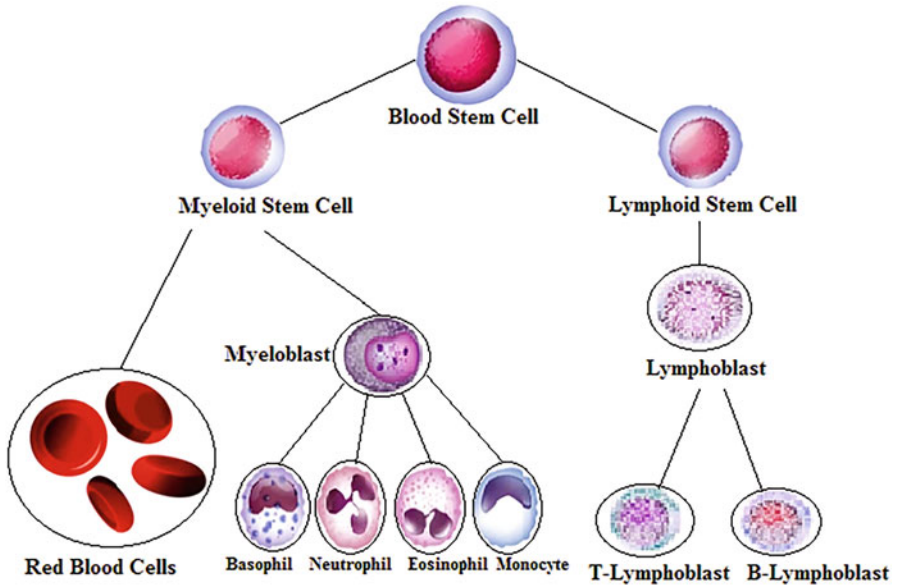


Fig. 4.1 Different types of cells in the human body

we discriminate them at the different stages, this is shown in Fig. 4.2. Different types of leukocytes perform several functions for maintaining the norm in the human body.

1.1 Types of WBC'S

- **Neutrophils:** About 50% of the WBC belongs to the Neutrophil class. The function of the neutrophil cell is to fight foreign invaders. It also communicates with the rest of the immune system to respond to the infection. They are the first in the group to perform their functions by sending biological signals. Around a hundred billion of these cells are released by the bone marrow per day but their normal life span is only 8 h.
- **Eosinophils:** They account for only 5% of the total WBC content in the human body but a major portion of these cells, are present in the digestive system. They also play an important role in fighting bacteria and attacking parasites. When the range of these cells goes out of limits, they may act as the cause for producing allergic symptoms.
- **Basophils:** The component corresponds to only 1% of the portion of leukocytes. The basophils can act as the source for controlling asthma and generates the response to pathogens. The stimulation of these cells releases histamine which may result in inflammation.

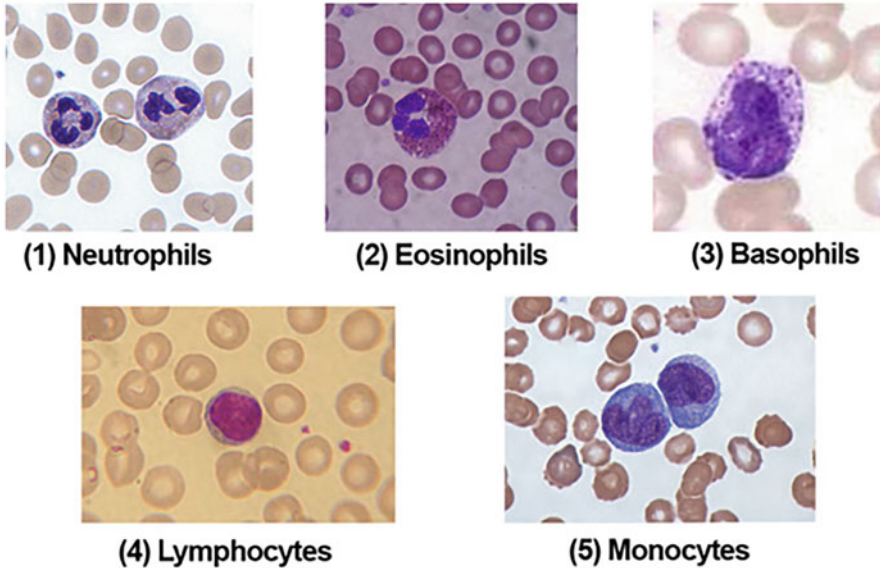


Fig. 4.2 Different types of WBC'S

- **Lymphocytes (B and T):** The lymphocytic cell system consists of two subtype cells B and T lymphocytes. The T lymphocyte plays the main role indirectly killing most of the foreign invaders whereas B lymphocytes control humoral immunity and is also used to know the efficacy of most of the vaccines.
- **Monocytes:** They comprise around 5–12% of the leukocytes in the human immune system. They perform the function of cleaning the dead cells, also named as garbage trucks of the human immune system.

The leukocytes have a certain range in which they remain beneficial to human health. Within their specified limits they perform the optimal function for the human immune system. If there is any deviation of the leukocytes from the normal limits, overproduction or underproduction of these cells, from the bone marrow. They affect the human body in a harmful way.

1. **Overproduction of WBC'S:** It is the case in which the bone marrow produces an enormous number of WBC'S that grow above the normal limits. This increased growth interrupts the function of the normal flow of the human body. There are multiple causes for this elevated blood cell growth either the bone marrow produces too many cells at the early stage or of the following symptoms.
 - (a) Cancer of type leukaemia, in which a large number of abnormal cells are made in the bone marrow.
 - (b) Infection.
 - (c) Asthma.

- (d) Stress may also lead to the increased growth of cells.
- (e) Autoimmune disorder.

2. **Under production of WBC'S:** This is the case in which a very smaller number of WBC'S, are produced by the bone marrow. The symptoms which may result in the undergrowth of leukocytes are as follows.

- (a) Autoimmune diseases.
- (b) Severe infections may include fever, cough, diarrhoea, pain etc.
- (c) Cancer of type metastatic.
- (d) Aplastic anaemia disorder.
- (e) Accumulation of WBC in the spleen.

1.2 *Medical Staff Recommends CBC (Complete Blood Count) Test to Diagnose the Blood-Related Diseases*

To test a wide range of disorders, present in the human body doctors follow a commonly recommended test called CBC that evaluates overall health performance [1–6]. It measures several components RBC, WBC, haemoglobin and platelets. Usually, doctors prescribe this test to analyze one or the combination of the following disorders anaemia, infection and leukaemia. In Table 4.1 we have depicted the result of the CBC for the normal patient. Medical experts recommend this test for the following reasons.

Table 4.1 Normal blood count of the healthy person

Blood component	Range of values
Haemoglobin	Male: 13.2–16.6 g/dL ^a (132–166 g/L) Female: 11.6–15 g/dL (116–150 g/L)
Hematocrit	Male: 38.3–48.6% Female: 35.5–44.9%
White blood cell count	3.4–9.6 billion cells/L (3400–9600 cells/ μ L)
Red blood cell count	Male: 4.35–5.65 trillion cells/L ^b (4.35–5.65 million cells/ μ L ^c) Female: 3.92–5.13 trillion cells/L (3.92–5.13 million cells/ μ L)
Platelet count	Male: 135–317 billion/L (135,000–317,000/ μ L) Female: 157–371 billion/L (157,000–371,000/ μ L)

^adL = deciliter

^bL = liter

^c μ L = microliter

- **To review the health performance of the patient:** The doctors usually judge the health of the patient, by analyzing the range of different blood components present in the human body, so that we can get information regarding the different disorders (allergy, infection, anaemia and leukaemia). These measures can be collected using CBC to monitor the overall health information of the patient.
- **To diagnose a related disease:** If somebody experiences a feeling of fever, weakness, inflammation, bleeding and infection. The doctors advise the CBC to the patients for knowing the cause for the related symptoms.
- **Treatment for the disease.** It can also be used to monitor the effect of the desired treatment method. That is related to the counting of blood cells. By doing CBC doctors can get information regarding the perfect medication required for the particular type of symptom.

1.3 Automatic Disease Detection Approach

Automatic detection of the blood disorder can be achieved by using machine learning analysis of the microscopic blood sample images. That can characterize different components of the blood sample image efficiently at the earliest stages. This will aid the medical staff in the treatment of blood-related diseases. Artificial Intelligence has already revolutionized the modern world with the development of powerful machine learning models. For driving autonomous cars, weather forecast prediction, intelligent agents, natural language processing, vision, biometrics etc. know it is time to shift our efforts towards the healthcare system, by developing an automatic disease diagnosing system that will ease the process of the treatment methods. The data which is generated by the medical staff has too much complexity and volume in its nature. The source of information is heterogeneous. The data exist in different format of X-rays, chest images, knee images, blood images, bone marrow, DNA (deoxyribonucleic acid) microarray and RNA (ribonucleic acid) sequencing. In recent, the procedure of analyzing this volume of information has been relying solemnly on the expert of medical staff. Which is complex, time-consuming, and prone to errors. Also, a large portion of the information is not explored because of the negligence of the medical staff this requires the need for intelligent machine learning models to find the meaningful patterns in this vast amount of data.

In the past decade machine learning has been explored in the field of medical sciences to find automatic identification of most of the challenging diseases in the healthcare system like leukaemia, prostate cancer, colon cancer, cardiovascular disease, diabetes, Alzheimer's disease etc. For some of the diseases, automatic drugs have been discovered. That will improve the speed up of the treatment methods. Among the mentioned disease cancer has been observed as the deadliest disease with a maximum number of average deaths per year. Research data has shown that on average around 17 million cancer cases are diagnosed each year and there is the possibility that the average number of cases may go up, to the number of

27 million by the year 2040. Every type of cancer is born from the cells [5, 6]. Which originate from the soft fatty fluid from the bone. These abnormal cells show an unstoppable growth and they live beyond the normal period of life. Which results in the disruption to the flow of normal cells. Usually, cancers are treated with the following methods of treatment as transplantation of bone marrow, radiations, molecular therapies (imatinib, dasatinib and nilotinib), and immunotherapy. These modes of treatment are suggested by the medical expert, based on accurate disease identification. It has been observed in patients about the consequences of these modes of treatment in the form of hair loss, infertility and change of skin colour. The cause behind this was that the medical staff has not utilized the patient's data efficiently, accurate treatment of the disease is only possible, once we can perfectly identify the disease. This identification of the disease can be done by using a machine learning-based solution. It has the potential to explore a vast number of meaningful possibilities from the highly complex structure of the patient's gene data. For some of the diseases, intelligent models have been developed. Consider the case of cardiovascular disease factors that influences the patient's personal life were collected by the risk assessment approach. Then the suggestions of the treatment method were made at the next level. The genetic information of the patient (genes that could cause obesity) was utilized in the development of an intelligent system for type 1 diabetes.

In recent research, we have found most of the diseases are related to the abnormal growth of the different types of WBC'S. Considering this in mind we have picked our objective of analyzing the leukocytic cells from the blood image data content. Features of blood images can only be extracted optimally once; we can properly discriminate multiple components of the blood-related data content. This requires the need for a robust methodology at the segmentation phase because the other phases like the extraction of features and their classification depend upon the accurate segmentation of the multiple blood sub-structures. Keeping this in consideration we have relaxed our objective for sub-segmenting the blood smears using robust methodology with dilation and erosion-based morphology. Our work, will not only have a focus on finding and analyzing the challenges in recent literature but also address some of the challenges in the current methodology.

To analyse the WBC'S automatically, we need to understand the structure of the leukocytic cell system. This can be done, by first segmenting the blood image into several distinct components (WBC, RBC, Platelets and Background). Among the multiple components, present in the blood image only the leukocytes are responsible for different types of blood disorders. This simplifies our machine learning approach because we need to focus only on a single component from the number of blood image data components. Most of the information related to WBC'S are stored in the nucleus of its cell. Keeping this in consideration, we need to divide the cell pattern into two sub-components i.e., the nucleus and the cytoplasm. This has further simplified the direction of our work towards the extraction of cell nucleus pattern because it acts as the main reservoir of the information in the cell body.

2 Literature Review

There is vast literature available on the segmentation of medical blood images. Most of it has thrown light on the boundary cell detection-based approach. It has a focus on retaining the nucleus of different leukocytic cells. The approach extracts the nucleus by identifying discontinuities between different blood components. This is done by computing gradients that show a directional shift in the change of intensities. This procedure is simple. The pattern-based approach performs the segmentation of data points based on their distance to the nearest mean. The data points are grouped into k number of clusters. Each data point belongs to the group with the highest membership value. In [7] the authors have segmented different types of WBC'S from the fluorescent biomedical images. They have divided their work into three stages. In the initial, they have generated L^*a^*b colour space from the given blood images, as this colour space suits best with the vision and perception. Next, they have utilized an unsupervised machine learning algorithm (k -means) for clustering of the data points based on the Euclidian distance measures between the nearest data points. In the end, they have done morphological processing on the connected data components of the resulting binary image. Even though, the authors claim that the model has generated promising results with a sensitivity of 96.49% and a precision of 98.3584%. There were certain problems with the concerned methodology because the model has highlighted some of the non-WBC components in the leucocyte class.

Liu Y. et al. [8] has proposed an advanced approach for WBC segmentation. The approach has a two-phase process. In which the model first detects the WBC location by the multi-window and multiscale methods. It fits leukocytes with different sizes. Once the model has successfully identified different locations of the WBC, then the authors have done final segmentation of the identified regions by using iterative Grab Cut. The performance of the proposed methodology was evaluated, on the two challenging datasets CellaVision and jiashan. Even though the authors have reported around 98% precision and recall on the mentioned datasets, they were unable to find the optimal parameter values, thus they have manually set the values of the parameters for adjustment of different types of datasets. In [9] the authors have extended the Otsu thresholding algorithm to compute different threshold values for each of the fractal regions in the leukocytic blood image. This has been done to extend the approach to biological images, where a single global threshold value is not sufficient to distinguish different classes of pixel values. However, the model needs improvement in its methodology (inertia weighting, adaptability parameter) to perform robust discrimination of different subcomponents. In [10] NIOA (Nature Inspired Optimization algorithms) is utilized to prevent false positiveness of the segmented data points that belongs to different components of the cell image.

In [11], on blood leukaemia smear images optimal thresholding and morphological operations were performed to generate the texture features. Later supervised and unsupervised classification of the features was done by many of the supervised and unsupervised classifiers to accurately classify the data. Thresholding based methodology was adopted in [12], where the authors had dealt with the k number of

thresholds in the binary image. In which, the object is assigned to the positive region if its value is greater than k otherwise the object is treated as non-relevant content of the data. The approach does not work with variate types of illuminations in the input image. The region growing method is proposed in [13] that treats data objects as seeds. Depending upon the size of the seed the model decides whether pixels needed to be added or eroded. To extract cells from the blood image the authors have utilized features like illumination, shape and texture of the blood sample images. Erosion based morphological blood cell analysis was also performed in [14]. The authors have modelled the problem by combining region extraction with erosion-based limits to segment the overlapping leukocytic substructures. The region-based watershed segmentation was adopted on grey level intensities with graph-based manifold learning [15]. The concerned approach has many drawbacks of over or under segmentation of the data. Also, some of the blood cells were not identified in the blood image.

Vonn Vincent Quiñones et al. [16] applied basic grey level boundary cell operations on the blood image. The approach consists of the following steps conversion of RGB (Red Green Blue) blood image to HSV (Hue Saturation Value) and grayscale colour space, component differentiation, binarization and blob detection. The approach does not work appropriately on different types of blood images. Bilkis Jamal Firdosi et al. [17] has done pattern-based clustering on L^*a^*b colour space by using k -means. The model has not performed appropriately on different WBC types. Also, it has highlighted non-WBC blood components from the blood image. Rosyadi et al. [18] has retained the circularity feature of the blood images by applying Otsu for segmentation, and K -Means for clustering.

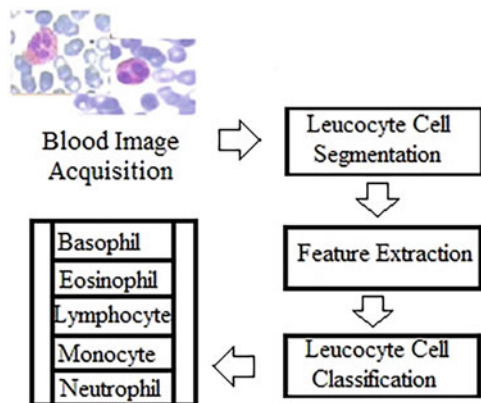
Garcia-Lamont et al. [19] has surveyed the most common recent approaches, for the segmentation of biomedical images. The colour of the biomedical image acts as the base for discrimination of sub-components. The authors have put their focus on the grey-level and colour based segmentation approach. The techniques that they have utilized are; boundary cell detection using thresholding, region growing, watershed segmentation for catchment basins supervised and unsupervised segmentation using clustering-based approaches, and some deep learning solutions. To identify leukocytes more appropriately. Anil-Kumar et al. [20] has done a review, on the recent segmentation methods used for the detection of leukaemia. The authors have highlighted most of the automatic disease detection-based approaches. The methods were categorized based on the procedure followed for the segmentation of biomedical images. Wang and Cao [21] propose a quick leukocyte nucleus segmentation method based on the component difference in RGB colour space (B-G), because this difference of the colour components (B and G) is very large in the leukocytes and platelets. Next, they retain the value of the WBC nucleus, by setting a filter that removes the content of platelets. In [22–27] the authors have modelled the problem, by applying the pattern-based segmentation methods on the characteristics of the biomedical images. The objective was to diagnose leukaemia. We have implemented many algorithms from the existing literature for segmentation but we have found some problems associated with ML (Machine Learning) methods that extract the nucleus from the blood image while removing the rest of the content as the image noise.

3 Methodology

Nowadays, in the advent of technology research efforts are being made for the development of computer-aided solutions that will do the automatic analysis on the biomedical image data. The automatic system can be utilized to aid the oncologist in diagnosing the different types of diseases, related to blood disorders. The CAD (Computer Aided Developed) solution, can predict the disease at a very fast rate with high accuracy. The solution comprises some set of steps from the image acquisition to the appropriate leukocytic cell identification. The layout of the CAD architecture is depicted in Fig. 4.3. To diagnose the disease, the oncologist has to only look towards the results where he can get the information about the type and number of WBC cells in the blood image. There are different types of WBC'S (Basophil, Eosinophil, Monocyte, Lymphocyte and Neutrophil) with little variation of parameters (perimeter, eccentricity, intensity etc.) between these cell types moreover, the blood components are congested in nature. This induces complexity in data thus requires the need for efficient machine learning methodology.

Appropriate discrimination of objects in the cell image can be performed, by understanding the structure of sub-components. This can be done by picking the robust method for segmenting the blood sample sub-components. Our work has a focus on the segmentation of microscopic blood image samples. There are two approaches through which we can perform the segmentation of the leukocytic cell nucleus. One is based on leukocytic cell boundary detection that deals with grey level intensities of different blood components, whereas the other has a focus on the pattern-based characterization of the blood samples that use different types of distance measures (Euclidian distance, Manhattan distance, Minkowis distance) to minimize the dissimilarity within the pattern of intraclass sub-segments. We have followed the boundary-based cell detection approach, with dilation and erosion of seed object. It is used to address some of the challenges in the recent literature. The pseudo-code of our model is depicted in Algorithm 4.1.

Fig. 4.3 Leucocyte subtype classification



3.1 Formulation of Our Developed Approach

- Conversion of the RGB blood image to the number of grey levels:** Conversion of the true colour blood image to the grayscale image. The grayscale image does not contain the complexity of large colour variation and it simplifies the differentiation of the leukocytic cell nucleus from the rest of the blood image content. The conversion retains only the luminance information to get a brighter nucleus spot. The result is depicted in Fig. 4.4b with reduced hue and saturation information. We represent the pixels of a given grayscale image in L grey levels $[1, 2, 3, \dots, L]$. Where each of the grey levels has a different set of pixel intensities. The number of pixels at level I is denoted by n_i and the total number of pixels $N = n_1 + n_2 + n_3 + \dots + n_n$.

Algorithm 4.1 Boundary Cell Detection Based Approach

Input: A microscopic blood sample image is read as input.

1. Convert the input image to the number of grey levels.
2. Contrast Enhancement of the number of grey levels for the robust discrimination of blood image substructures.
3. Adjust the selected blood image to the particular image histogram.
4. Choose a global threshold value t that separates Foreground class pixels from background class pixels.
5. Calculate the Variance of Foreground class pixels $V_f = \sigma^2$, Calculate the Variance of Background class pixels $V_b = \sigma^2$, Minimize the within-class variance and maximize the between-class variance.
6. Reduce the damage of the pre-processing steps by applying noise reduction techniques.
7. Segment the processed blood cell image into several regions.
8. Extract the WBC region by choosing the threshold that removes remaining unwanted portions (RBC, Platelets Background and Cytoplasm).
9. Apply dilation and erosion operation to retain the actual shape of the WBC.

Output: Segmented WBC nucleus.

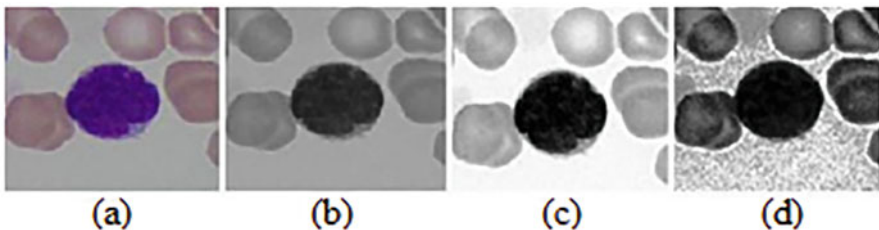


Fig. 4.4 Model results on a test data where labels indicate (a) Input blood sample image, (b) Discriminated Blood components, (c) Contrast adjusted blood image, (d) Contrast enhancement by matching c to a specified histogram

- **Blood Image Enhancement:** Quality of the blood image is an important factor for leukocytic cell nucleus segmentation, it depends upon the lighting conditions of the surrounding, resolution of the electron microscopic camera, the angle of capturing the blood smear image etc. The received input image at most of the times is of low to medium quality which makes difficulties in discriminating sub-components, thus requires a transformation in quality from the input blood sample quality to the desired level. This is done by the transformation equation that does the enhancement of the blood image, results are depicted in Fig. 4.4c, d.
 - Let x be an input image of dimension $m_i * m_j$ with intensity values between 0 to $L - 1$. Consider the case of analyzing the blood image, its quality usually falls in the intensity values between 100 and 150 that is enhanced by modifying the values of L to the maximum value i.e., 255. Its objective function is depicted in (4.1).

$$g(x) = \left(\left(\frac{f - 100}{150} \right) * 255 \right), \forall 100 \leq f \leq 150 \quad (4.1)$$

- Next, we normalize the values of histogram p to the desired intensity levels of L so that the extracted pixel values have a uniform distribution, which is computed from (4.2).

$$p_k = \frac{\text{Number of pixels with intensity } k}{\text{Total number of pixels}} \text{ where } k = 0, 1, 2, 3, \dots, L - 1, \quad (4.2)$$

- After applying histogram equalization, we need to round of L to the nearest positive integer by the use of floor functions depicted in (4.3) and (4.4).

$$g_{i,j} = \text{floor} \left((L - 1) * \sum_{n=0}^{f_{ij}} p_n \right) \quad (4.3)$$

$$T(K) = \text{floor} \left((L - 1) * \sum_{n=0}^k p_n \right) \quad (4.4)$$

- **Global Optimal Threshold t selection:** To do robust subcomponent segmentation we have divided the blood cell image components into two classes of pixels, Foreground (WBC Nucleus), the important component and Background (RBC Platelets Cytoplasm and Noise), non-required part. These two classes of pixel levels were separated by computing the optimal threshold levels, using the Otsu threshold algorithm. It is the global optimal threshold algorithm. It works by,

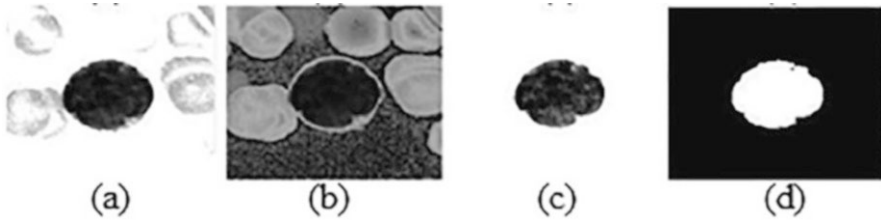


Fig. 4.5 Model results on test data during the nucleus extraction process where labels indicate (a) Brightened blood components (RBC, Platelets and Background), (b) Highlighted WBC nucleus with some noise, (c) Extracting WBC cell with minimum distortion to the nucleus, (d) The output of the shape-based Thresholding algorithm where between-class variance is high and within-class variance is low

maximizing the variation between the irrelevant class values and minimize the variation within the relevant pixel class values. Its pseudo-code is depicted in Algorithm 4.2, and results are depicted in Fig. 4.5a. We have represented foreground class pixels by V_f and background class pixels by V_b . These two classes are separated by an optimal global pixel intensity threshold value t where levels $V_b = 0, 1, 2, \dots, t$ and $V_f = t + 1, t + 2, \dots, L - 1$.

- **Restoration:** Due to the pre-processing done on blood cell images we are causing some form of damage to the blood image subcomponents, in the form of multiple sources of noise. We need to reverse the effect of the noise by setting filters that will reduce the impact on the final segmented blood components. Noise arises due to the pattern of low-intensity pixel values that are found around the edges of WBC during the extraction of the nucleus. It can be reduced by keeping a filter that removes abnormalities having pixel value less than the particular t . There are various sources of noise found in the processing of blood cell image data; such as salt-and-pepper noise, Gaussian noise or periodic noise. Results of the operation are depicted in Fig. 4.5a.
- **Region separation:** Segmentation involves dividing a blood cell image into several segments or regions. Microscopic blood cell mainly consists of the segments like WBC, RBC, Platelets and Background. Each segment has a particular structure like Size, Eccentricity, Major-Arc and Minor-Arc. Among the given segments we are interested in extracting the WBC portion for the detection of leukaemia. The result of separated components is depicted in Fig. 4.5b.
- **Thresholding:** It involves converting blood cell image into binary image. It reduces complexity and simplifies component recognition and their classification. The binary version contains essential information about the shape and position of the extracted foreground part i.e., WBC Nucleus. Normally only those black and white pixels of blood cell are retained having pixel intensity that matches the intensity level of WBC nucleus remaining unnecessary portion is discarded. Results are depicted in Fig. 4.5c.

- **Edge Detection:** To retain the shape of the WBC cell nucleus both the Sobel Edge Detection and Canny Edge Detection techniques were applied. Edge detection is the most useful approach for detecting discontinuities in a set of connected pixels that lie on the boundary between different regions of the blood cell image, where a gradient is a directional change in the intensity. Sobel detection computes the approximate absolute gradient magnitude at each point by computing 3×3 filters. Canny edge detection first removes noise by applying a low pass filter and then it picks out the best pixel values for edges among the multiple pixel values in a local neighborhood. Results are depicted in Fig. 4.5d.
- Let the convolution mask g_x estimates the gradient in the x-direction and the g_y estimates the gradient in the y-direction, A is the binarized image and $*$ represents the convolution operator. It is depicted in (4.5).

$$g_x = \begin{bmatrix} -2 & 0 & +2 \\ -1 & 0 & +1 \\ -2 & 0 & +2 \end{bmatrix} * A \text{ and } g_y = \begin{bmatrix} -2 & -1 & -2 \\ 0 & 0 & 0 \\ +2 & +1 & +2 \end{bmatrix} * A \quad (4.5)$$

- The resulting gradient expression for the Sobel filter is given in (4.6).

$$\sqrt{G_x^2 + G_y^2} \quad (4.6)$$

using this information gradient direction is calculated as,

$$\theta = \arctan \left(\frac{g_y}{g_x} \right)$$

Algorithm 4.2 Computing Global Optimal Thresholding Value

Input: Contrast-enhanced blood sample image

1. Compute the pixel intensity weights for background class:

(a) Represent the weight of background class by W_b .

(b) Compute the weight of background pixels, $W_b = \sum_{i=1}^t \frac{n_i}{N}$

(c) Mean background pixel weight μ_b , $W_b = \frac{\sum_{i=1}^t i * n_i}{\sum_{i=1}^t n_i}$

• The variance of the background class, $\sigma_b^2 = \frac{\sum_{i=1}^t (i - \mu_b)^2 * n_i}{\sum_{i=1}^t n_i}$

2. Computation of foreground, pixel intensity weights:

(a) Represent the weight of foreground class by W_f .

(b) Compute the weight of foreground pixels, $W_f = \sum_{i=t+1}^L \frac{n_i}{N}$

(c) Mean foreground pixel weight μ_f , $W_f = \frac{\sum_{i=t+1}^L i * n_i}{\sum_{i=t+1}^L n_i}$

(d) The variance of the foreground class, $\sigma_f^2 = \frac{\sum_{i=t+1}^L (i - \mu_f)^2 * n_i}{\sum_{i=t+1}^L n_i}$

3. Compute the within-class variance $\sigma_w^2 = W_b \sigma_b^2 + W_f \sigma_f^2$, by taking the summation of foreground and background variances, multiplied by their associated weights.

Output: a global threshold that maximizes between-class variance and minimizes within-class variance.

- Morphological Processing:** Mathematical morphological operation is more suitable for retaining the shape of WBC cells in images. The two main processes that can be used in the retention of WBC structure are dilation and erosion. These involve combining two sets of pixels, where one set consists of an input blood sample image which is to be processed and the other is Structuring Element (SE) which is a small matrix of pixels each with a value of zero or one. The SE is said to fit the image if for each of its entries matches the corresponding entries of image pixels. It is set to hit if at least one of its entries matches the corresponding pixel entries of the blood image. Results are shown in Fig. 4.6.

In dilation, every point in the cell image is superimposed by the SE, with its surrounding pixels. It is also called thickening operation where the operation is controlled by the shape of the SE.

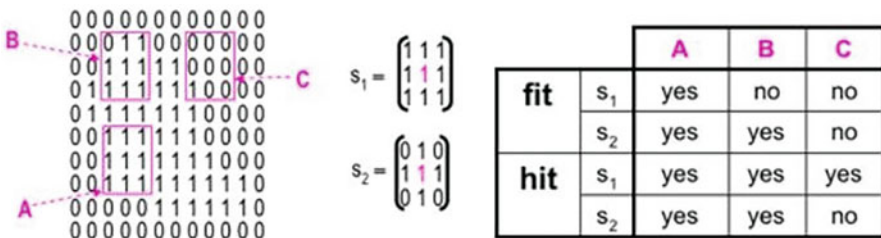


Fig. 4.6 Structuring element fit and hit results

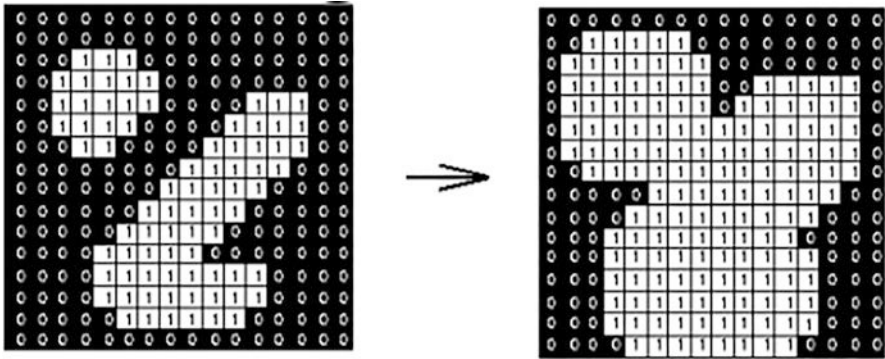


Fig. 4.7 Result of Dilation process on binary Matrix

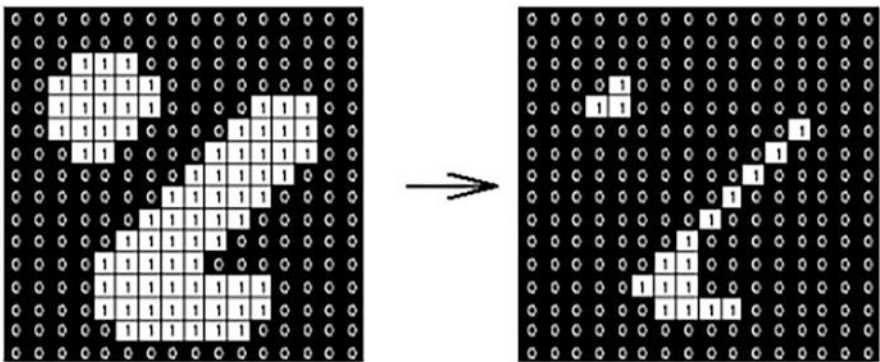


Fig. 4.8 Result of Erosion process on 0,1 Matrix

- The dilation of an image im by a structuring element SE (denoted as $im \oplus SE$) produces a new binary image $g = im \oplus SE$ with ones in all locations (x, y) of blood image at which the SE hits the blood image i.e., $g(x, y) = 1$, if SE hits or fits im , otherwise $g(x, y) = 0$ repeating for all pixel coordinates of $g(x, y)$. The effect of dilation is depicted in Fig. 4.7.
- The erosion (opposite of dilation) thins a microscopic cell image im with a structuring element SE (denoted as $im \ominus SE$) produces a new binary image $g = im \ominus SE$ places ones in all locations (x, y) of blood image if and only if the SE fits the blood image i.e., $g(x, y) = 1$ otherwise $g(x, y) = 0$ repeating for all pixel coordinates of $g(x, y)$. The effect of erosion is depicted in Fig. 4.8.

4 Result and Discussions

To do the optimal analysis, on the performance of the adopted methodology, we have used different types of datasets for evaluation. The data on which we have evaluated our model consists of the number of publicly accessible digital libraries ALL-IDB (Acute Lymphoblastic Leukaemia International Database), Kaggle, LISC (Leucocyte Images for Segmentation and Classification). ALL-IDB is the widely used dataset. It contains samples from the leukaemia and non-leukaemia classes in which the dimension of each blood sample image is resampled to the size of 1200*1200 pixels plotted horizontally and vertically. The dataset comprises the two subsets of segmented and non-segmented microscopic blood samples. The segmented set contains around 260 blood images that require only classification and the non-segmented set contains 108 microscopic blood images thus requires segmentation and classification. We have used lymphocytes from both of the sub-classes for extraction of WBC cell nucleus and computing their important features for the classification of leukaemia and non-leukaemia cells.

Another public access dataset on which we have evaluated our methodology is the LISC. This dataset contains four different types of leukocytes (Basophil, Eosinophil, Monocyte, Lymphocyte and Neutrophil) the dimension of each blood sample 720*576 with a magnification of 100. This dataset contains nearly 250 colour blood images, collected from the Hematology and BMT (Bone Marrow Transplantation) research centre of the hospital in Tehran, Iran. One more dataset that we have collected for evaluation purpose belongs to the Kaggle. This dataset also contains several leukocytes from different types of WBC cells.

Our main target was to develop an efficient machine learning methodology to analyze biomedical images that can be used to aid the medical expert in diagnosing blood-related disorders. Multiple approaches are used for the identification of WBC cell. The method that extracts these features with minimum distortion is desirable. We have adopted robust methodology with dilation and erosion based morphological operations, to extract the complete leukocytic cell. The work has successfully extracted leukocytic cells, counting their number and computing their basic features. We have depicted results in Fig. 4.9.

Although we have optimally utilized our methodology, to do an automatic analysis of the biomedical image data. Our method mainly concentrates on the retention of the leukocytic cell nucleus, as done by most of the related methods. There exist several new challenges that have been overlooked, by the majority of the recent works. We have observed several new problems while doing experiments in Matlab. To some extent, we have addressed the issues by adopting dilation and erosion-based morphometry. It still needs attention towards the more advanced machine learning models. For that, we are looking for a vision-based solution that will focus on the retention of the complete leukocytic cell, which will do the best characterization of the blood image components. So that, we can easily differentiate various types of leukocytic cells. The challenges are discussed as follows.

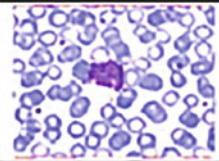
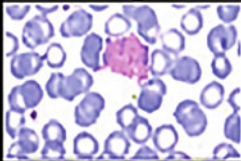

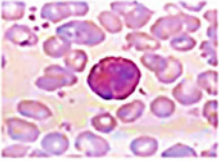
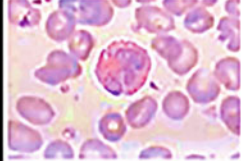

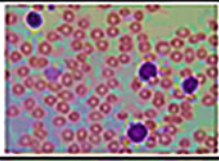
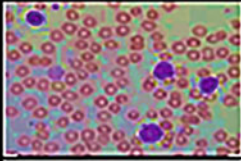

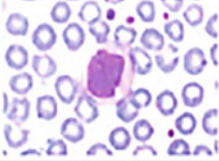
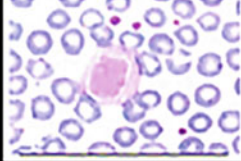

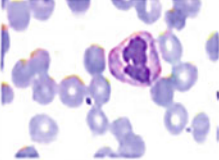
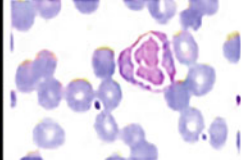

<i>Types of WBC</i>	<i>Input Blood Image</i>	<i>Highlighted WBC cell nucleus</i>	<i>Automatic Segmented WBC nucleus</i>
<i>Basophil</i>			
<i>Eosinophil</i>			
<i>Lymphocytes</i>			
<i>Monocytes</i>			
<i>Neutrophil</i>			

Fig. 4.9 Result of leukocytic cell nucleus extraction on different types of WBC'S

1. **Identify different types of WBC cells:** Every type of leukocytic cell has some specific properties like Eccentricity, size, perimeter, Major-Arc, Minor-Arc etc. that differentiates it from the other types of blood components. All these properties are associated with complete leukocytic cell including cytoplasm. Results have shown that the nucleus extraction process manipulates the size parameter and thus we face difficulty in identifying the type of leukocytic cell. In Fig. 4.10 we have depicted the results.
2. **Counting the leukocytic cells at different stages of maturity:** Leukocytes have been classified into granulocytes and non-granulocytes. The nucleus of the granulocytes gets divided into some segments at the mature stage. Segmenting

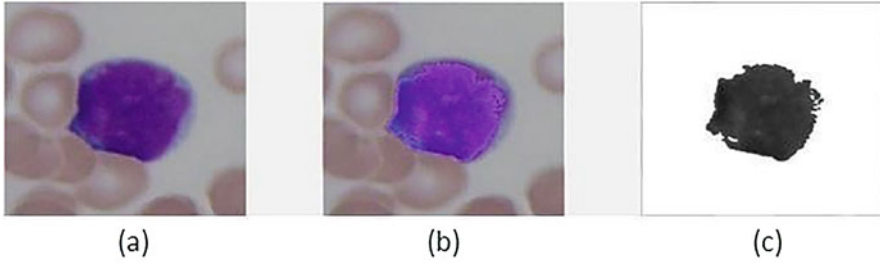


Fig. 4.10 (a) Input Blood Sample Image of lymphocytic class. (b) Blood Image with Highlighted Nucleus. (c) Output WBC cell with size variation in cell size (Neutrophil class)

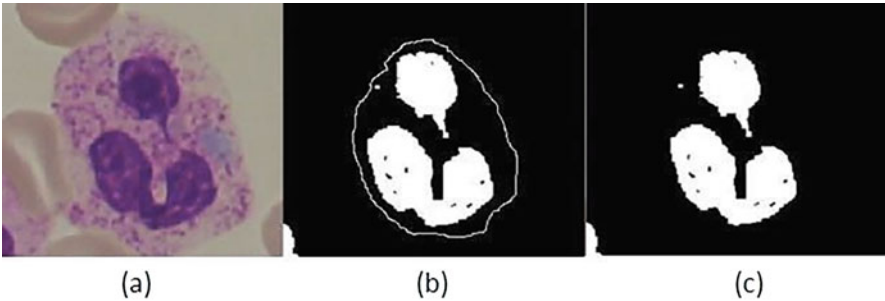


Fig. 4.11 (a) Blood image with one object single WBC cell. (b) Expected count should be single object. (c) Model counted three objects for the single nucleus

the nucleus of the granulocytic cell at the mature stage gives the wrong count of the number of objects in the blood image. Result is depicted in Fig. 4.11.

- (a) **Challenge addressed by incorporating morphological process in the methodology:** We have addressed the multiple count problem by utilizing mathematical morphological operations on the divided segments of the single leukocytic cell nucleus. This is an important machine learning tool for analyzing the different types of binary structures in the biomedical image content. This approach can be used for the extraction of leukocytic cell patterns based on the circular structure of the cell and the filtering gradients. There are four operations in the morphological process dilation, erosion, opening and closing. Dilation and closing extend the boundary cell region where ever it finds the hit in the data content of the blood image, we have used these operations for region growing of the cells. Whereas erosion and opening are the reverses of dilation, it shrinks the blood image data content. We have used these two operations to remove the unwanted or irrelevant content (noise) from the blood image. Figure 4.12 shows the results of the morphological processing on the divided or segmented WBC granules. Where, the initial experimental results have shown, the count of three cells, for the single

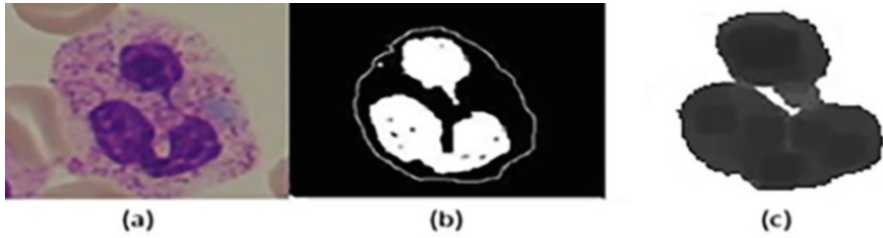


Fig. 4.12 Result of dilation on the microscopic blood sample image, where (a) Blood image with one object single WBC cell (b) Expected count should be a single object (c) Automatic count of a single object

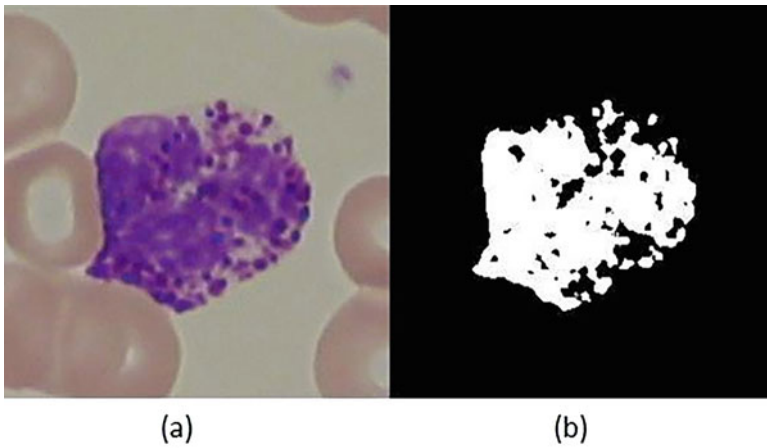


Fig. 4.13 (a) Input blood sample image. (b) Result of distorted WBC cell nucleus after removing noise from the blood sample image

WBC object cell. After applying morphological processing. We have filled the holes between the granules. It is done by dilating the region, to get the actual result of the single object cell.

3. **Retaining the geometric structure of the leukocytic cells with minimal distortion to the nucleus:** At times noise removal filters distorts the geometric pattern of WBC'S. The filter is a threshold value, and what falls below the value of the filter is removed as noise from the cell image. In Fig. 4.13 the data points of class noise have the same intensity, that matches the pixel intensity of the cell nucleus. Setting the filter of that value removes also the data points of the nucleus. This approach distorts the geometric shape of the cell nucleus.
4. **Handling multiple sources of noise in the data:** Preprocessing leads to small data points, in the form of pixels plotted horizontally and vertically across the blood image. To get a true count of the number of leukocytes noise removal filters are adopted to get rid of the irrelevant content. Figure 4.14 depict its result.



Fig. 4.14 (a) Input blood sample image. (b) Result of WBC cells

5 Performance Measures

For every successful execution of the concerned methodology, we have calculated the following performance metrics Accuracy, Specificity, Error rate, Precision and Prevalence. The performance of the model on the segmentation of the leukocytic cell nucleus was computed by the following machine learning metrics TP, FP, TN and FN. Where TP (True Positive) is the number of correct predictions. It indicates, the number of leukocytes that are predicted by the model and the same number is present in the blood image. It represents the correct segmentation results by the model. FP (False positive) is the wrong count of the number of leukocytes that are not present in the blood sample image but are counted by our method. FN (False Negative) is the reverse of FP. In this metric, the model does not count some of the leukocytes that are truly present in the blood image. TN (True Negative) is the metric in which the model has not identified the region as the leukocytic cell, and the same region does not belong to that object when counted manually. To evaluate the performance of our method, we have calculated the following performance measures accuracy, sensitivity and precision. The formulation of these measures is depicted in Eqs. (4.7–4.9).

$$Accuracy = \frac{TP + TN}{Total\ data} \quad (4.7)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4.8)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.9)$$

The developed approach was tested on some few hundred (250) medical blood images collected from the mentioned datasets in random proportion. We have evaluated the model on different combinations of blood sample images. Figures 4.15 and 4.16 depicts the results of the experiment on the individual and combined

subtypes of WBC'S. These blood images were categorized into five different classes (basophil, eosinophil, monocyte, lymphocyte and neutrophil). Table 4.2 represent the result of the confusion metric that is used to evaluate the automatic and manual counting of leukocytic cells.

The approach that we have adopted in our work, has shown comparatively higher performance as compared to some of the proposed methods in the literature. The model has performed best on the segmentation of the lymphocytes, with an average accuracy of 93.1%, which is very high as compared to the segmentation accuracy achieved by Hiren and Velani [24]. They have reported accuracy of 91.07% with an over-segmentation and under-segmentation rate of 6.50% and 6.04%. The over-segmentation rate has been expressed as the region, which must be included in the background class, the objects were placed in the foreground class of cells. That gives the wrong count of the number of leukocytes. Whereas, the under-segmentation rate is the reverse of over-segmentation, in which multiple cells are counted as a single object. The work of Khalid A. et al. [13] has outperformed our boundary-based cell detection approach, they have achieved the mean best accuracy of 95% on the segmentation of lymphocytes, and 92% on the neutrophils. The model was evaluated

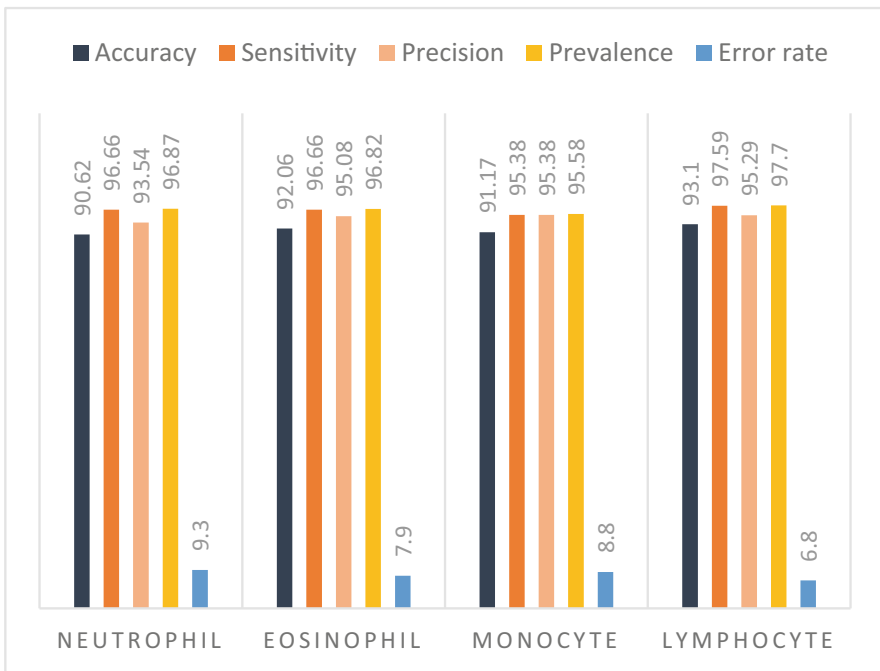


Fig. 4.15 Segmentation performance of the model on different leukocytic cell types

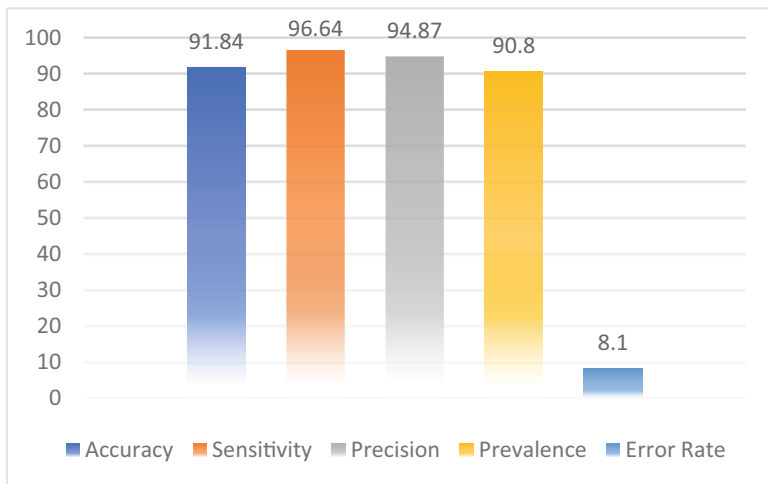


Fig. 4.16 Overall segmentation performance of the model on the combined subtypes of WBC

Table 4.2 Manual and Automatic counting of leukocytes

WBC	Total images	Automatic count	Manual count	TP	FP	FN
Lymphocyte	70	86	87	81	4	2
Monocyte	60	65	68	62	3	3
Eosinophil	60	67	63	58	3	2
Neutrophil	60	50	64	58	4	2
Total	250	268	282	259	14	9

only on the set of 100 samples of microscopic WBC images, which is very less in comparison to our approach, which we have evaluated on 250 blood cell images.

6 Conclusion

We have utilized a robust methodology to perform the segmentation of blood images. Since the nucleus of leukocytic cells, plays the main role in diagnosing several diseases (leukaemia, allergic infections, AIDS). We have picked our objective of extracting the leukocytic cell nucleus, by adopting shape-based morphological processing, which extracts the shape of the cell nucleus with less distortion to its structure. While implementing the methodology in the work-plan, we have observed, that there exist several challenges that need to be addressed. Although, we have perfectly retained the shape of WBC'S appropriately. There still exist the challenges that we have observed during the evaluation of the experiment. So, we come with some new suggestions that needs to be addressed in future.

- For the task of medical blood image analysis, there is a need for a robust method, that should differentiate multiple substructures appropriately.
- While removing other blood components (Cytoplasm, RBC, Platelets, and Background). The methodology for leukocytic cell extraction must focus to retain the geometric structure (cell shape) of the WBC cell.
- To prevent several counts of single WBC. The nucleus extraction process must encompass all the granules of divided WBC inside the curvature cytoplasm, as a single object.

References

1. Vezzoli M, Bonardelli S, Peroni M, Ravanelli M, Garrafa E (2017) A simple blood test, such as complete blood count, can predict calcification grade of abdominal aortic aneurysm. *J Vasc Med.* <https://doi.org/10.1155/2017/1370751>
2. Escott-Stump S (2015) Hematology: anemias and blood disorders. *Nutr Diagn Relat Care* 3(2): 703–748
3. Ferrari D, Motta A, Strollo M, Banfi G, Locatelli M (2020) Routine blood tests as a potential diagnostic tool for COVID-19. *Clin Chem Lab Med* 58(7):1095–1099. <https://doi.org/10.1515/cclm-2020-0398>
4. Iv FT, Raess PW, Ang D, Andal JJ, Press R, Tran N (2021) Somatic mutational analysis using next generation sequencing in predicting disease behavior of cytogenetically normal myelodysplastic syndromes. *Blood Res Disord.* <https://doi.org/10.23937/2469-5696/1410063>
5. The Leukemia & Lymphoma Society, New York. <https://www.ils.org/facts-anstatistics>. Accessed 16 Nov 2019
6. Cancer Research UK. <http://www.cancerresearchuk.org>. Accessed 16 Nov 2019
7. Ferdosi BJ, Nowshin S, Sabera FA, Habiba (2019) White blood cell detection and segmentation from fluorescent images with an improved algorithm using K-means clustering and morphological operators. In: 4th International conference on electrical engineering and information and communication technology, ICEEICT 2018, pp 566–570. <https://doi.org/10.1109/CEEICT.2018.8628068>
8. Liu Y, Cao F, Zhao J, Chu J (2017) Segmentation of white blood cells image using adaptive location and iteration. *IEEE J Biomed Health Inform* 21(6):1644–1655. <https://doi.org/10.1109/JBHI.2016.2623421>
9. Mandyartha EP, Anggraeny FT, Muttaqin F, Akbar FA (2020) Global and adaptive thresholding technique for white blood cell image segmentation. *J Phys Conf Ser* 1569(2). <https://doi.org/10.1088/1742-6596/1569/2/022054>
10. Dhal KG, Gálvez J, Ray S, Das A, Das S (2020) Acute lymphoblastic leukaemia image segmentation is driven by stochastic fractal search. *Multimed Tools Appl* 79(17–18): 12227–12255. <https://doi.org/10.1007/s11042-019-08417-z>
11. Rawat J, Singh A, Bhadauria HS, Virmani J, Devgun JS (2017) Classification of acute lymphoblastic leukaemia using hybrid hierarchical classifiers. *Multimed Tools Appl* 76(18): 19057–19085. <https://doi.org/10.1007/s11042-017-4478-3>
12. Bhavnanian LA, Jaliyab UK, Joshib MJ (2015) Blood cell segmentation and counting: a survey. *Int J Innov Emerg Res Eng* 2(11):21–24
13. Abuhasel KA, Faticah C, Iliyasu AM (2016) A commixed modified gram-Schmidt and region growing mechanism for white blood cell image segmentation. *IEEE*, New York
14. Fan J, Zhang Y, Wang R, Li S (2013) A separating algorithm for overlapping cell images. *J Softw Eng Appl* 2013:179–183

15. Sarosh P, Parah SA, Bhat GM (2021) Fast image encryption framework for medical images. In: 2021 2nd International conference on intelligent engineering and management (ICIEM). IEEE, New York, pp 149–154
16. Quiñones VV, Macawile MJ, Ballado A Jr, Dela Cruz J, Caya MV Leukocyte segmentation and counting based on microscopic blood images using HSV saturation component with blob analysis. In: 2018 3rd International conference on control and robotics engineering
17. Umamaheswari D, Geetha S (2018) Segmentation and classification of acute lymphoblastic leukemia cells tooled with digital image processing and ML techniques. In: Proceedings of the second international conference on intelligent computing and control systems (ICICCS 2018)
18. Rosyadi T, Arif A, Nopriadi, Achmad B, Faridah (2016) Classification of leukocyte images using K-means clustering based on geometry features. In: 6th International annual engineering seminar (InAES), Yogyakarta, Indonesia
19. Garcia-Lamont F, Cervantes J, López A, Rodriguez L (2018) Segmentation of images by color features: a survey. *Neurocomputing* 292:1–27
20. Anilkumar KK, Manoj VJ, Sagi TM (2020) A survey on image segmentation of blood and bone marrow smear images with emphasis to automated detection of leukemia. *Biocybernet Biomed Eng* 40(4):1406–1420. <https://doi.org/10.1016/j.bbe.2020.08.010>
21. Wang Y, Cao Y (2019) Quick leukocyte nucleus segmentation in leukocyte counting. *Comput Math Methods Med*. <https://doi.org/10.1155/2019/3072498>
22. Muntasa A, Yusuf M (2019) Modeling of the acute lymphoblastic leukemia detection based on the principal object characteristics of the color image. *Proc Comput Sci* 157:87–98. <https://doi.org/10.1016/j.procs.2019.08.145>
23. Acharya V, Kumar P (2019) Detection of acute lymphoblastic leukaemia using image segmentation and data mining algorithms. *Med Biol Eng Comput* 57(8):1783–1811. <https://doi.org/10.1007/s11517-019-01984-1>
24. Hiren DV, Velani N (2017) An automated blood cell segmentation using fuzzy based system. In: 8th International conference on computing, communications and networking technologies, ICCCNT 2017. <https://doi.org/10.1109/ICCCNT.2017.8203971>
25. Gull S, Mansour RF, Aljehane NO, Parah SA (2021) A self-embedding technique for tamper detection and localization of medical images for smart-health. *Multimed Tools Appl* 80(19): 29939–29964
26. Showkat S, Parah SA, Gull S (2021) Embedding in medical images with contrast enhancement and tamper detection capability. *Multimed Tools Appl* 80(2):2009–2030
27. Ghayvat H, Pandya SN, Bhattacharya P, Zuhair M, Rashid M, Hakak S, Dev K (2021) CP-BDHCA: Blockchain-based Confidentiality-Privacy preserving Big Data scheme for healthcare clouds and applications. *IEEE J Biomed Health Inform*. <https://doi.org/10.1109/JBHI.2021.3097237>