# Detection and Segmentation of Blood Cells Based on Supervised Learning

#### Sufola Das Chagas Silva Araujo<sup>1</sup>, Meenakshi Sundaram Karuppaswamy<sup>2</sup> and V. S. Malemath<sup>3</sup>

<sup>1</sup>Department of Computer Science Engineering, Visvesvaraya Technology University, Machhe - 590018, Belgaum, Karnataka, India; sufolachagas100@rediffmail.com

<sup>2</sup>Department of Engineering, Faculty of Engineering and Applied Sciences, Botho University, Gaborone, Botswana; meenaksji@gmail.com

<sup>3</sup>Department of Computer Science Engineering, KLE Dr MSS College of Engg. & Tech., Udyambag - 590008, Belgaum, Karnataka, India; veeru\_sm@yahoo.com

#### **Abstract**

Cells in our blood can be categorized. Pathologist detects these blood cells and diagnoses different disease. This paper describes the way in which we can detect these cells in body by using the concept of image processing. The manual way of detecting these cells and diagnosing the disease is very time consuming and has more chances of making mistakes. This method is developing to identify the cell when any cell image is given as an input. The methodology used is, image acquisition, image segmentation, trained images, test images, feature extraction of cells, feature vector dataset, similarity of cells. This will detect the different cells.

Keywords: Feature Extraction, Image Segmentation, Red Blood Cells (RBC). White Blood Cells (WBC)

## 1. Introduction

Different diseases found in human blood cell were some can be very harmful and can also cause death of the patient. Detecting these cells can help the pathologists in diagnosing diseases like AIDS, Leukemia and blood cancer1 In order to control such harmful disease many scientist have tried different method for identifying the different cells in human blood<sup>3</sup>. Human Blood contains Red Blood Cells, White Blood Cells or Leukocytes and platelets3. Red blood cells are erythrocytes. These are most common in blood. In human blood there is 20-30 trillion RBC. WBC is fewer in human blood. In human blood there is 4,000-10,000 ( $\mu m$ ) WBC $^{10}$ . Platelets are smallest cells in our blood. In human blood the size of platelets is 2-3µm in diameter4. These cells are detected by the pathologist and diseases are diagnosed. This paper focuses on WBC cells. White Blood Cells builds our immune system which protect the body from contagious disease<sup>3</sup>. The count of WBC indicates the diseases in human body16. The WBC cells are different from other

blood cells because of their shape, color and size<sup>3</sup>. WBC has five different types based on the shape and color<sup>3</sup>. The five type of WBC are neutrophil, eosinophils, monocytes, basophils and lymphocytes<sup>2</sup>. These five types of WBC are classified in two categories Granulocytes and A granulocytes.

Granulocytes are the cells with several nuclei lobes. Granulocytes have following types of cells Basophils, Eosinophils and Neutrophils. A granulocytes area cells with no granules. Granulocytes have following types of cells Monocytes and Lymphocytes. The neutrophil cells contains nucleus divided in two to five lobes. Diameter of neutrophil cells is 10-12 ( $\mu$ m). The eosinophil contains nucleus divided in two lobes. Diameter of eosinophil cells is 10-12 ( $\mu$ m). The basophils contains nucleus of bi or tri lobed but it is difficult to see these lobes because coarse granules hide it. Diameter of these cells 12-15( $\mu$ m). The lymphocyte contains one large nucleus which covers entire cytoplasm. Diameter of these cells 7-8( $\mu$ m) for large and 12-15( $\mu$ m) is small. The monocyte contains one nucleus of kidney shape. Diameter of

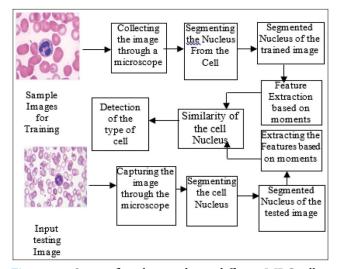
<sup>\*</sup>Author for correspondence

these cells  $15-30(\mu m)$ . The the amount of WBC cells in human body are Neutrophil-50-70%, Eosinophil-1-4%, Basophil-1%, Monocyte-6% and Lymphocyte-20-40% 4.16. The pathologists are detecting these cells manually which is more prone to human errors and time consuming also.

Manual detection is more troublesome because the human eyes have to detect the cells based on shape of the cells. The system proposed in this paper is to detect the different cells. This will give less error prone result within less amount of time. The pathologists have to detect the count of the cells manually . This can be time consuming and can have human errors. Manual detection is more troublesome because the human eyes have to detect the cells based on shape of the cells. The proposed method will detect the different cells of different shapes of the white blood cells. This will give less error prone result within less amount of time.

# 2. System Proposed

The system developed is automated system where it detects the different cells in human blood. This process has different steps. These steps are shown in the following diagram (Figure 1). The input is image of blood smear which has only one cell of any type of WBC cells. The output of the system will identify the cell detected and the file path where it is stored is also registered.



System flowchart to detect different WBC cells.

# 3. Methodology

The methodology of the proposed system to detect different cells is as follows.

#### 3.1 Image Acquisition

The images are obtained for the pathology lab. Pathologist prepares the blood smear and places under 100x resolution lens of microscope. Now the camera is placed on the eyepiece of the microscope and the image is obtained. All of these images are stored in \*.jpg format in folder.

#### 3.2 Image Segmentation

These images obtained are given as input to segmentation phase. Segmentation of an image is clustering similar property pixels into one cluster. The segmentation is done so that the image is represented in more meaningful way so that it becomes easier to analyse. This system uses image segmentation since we will be referring to different cells in blood smear. Using segmentation we will be separating the image background and the cell so that we could only get the cell nucleus in the foreground of the image 14,16. Image segmentation will also help to distinguish between the shapes of cell nucleus. Here the region of interest (nucleus) is set to its original color and background is set to white.

# 3.3 Trained Images

It is a data set of images which will have only one cell which can be any of the five types of WBC cell. These will be used to create the data set of feature vector.

# 3.4 Test Images

It is a data set of images which will have only one cell which can be any of the five types of WBC cell. The Feature vector of these images will be used to compare against the standard feature vector in the feature vector dataset.

#### 3.5 Feature Extraction

This is very important phase of this project. Feature extraction includes morphological operations. The features are based on shape, colour and texture feature<sup>13</sup>. It extracts some important information of the object of interest. In this system we are trying to find the shape features of the cell's nucleus12.

The feature vector of the trained image and the feature vector obtained from the test image which is blood smear which has only one cell of any type of WBC are compared. The shape features are obtained by the calculating the moments of each cell. Following are steps to calculate feature vector:

- 1. Dividing the image into regions.
- 2. Calculating area of each region.
- 3. Calculate the number of x-coordinate pixels and y-coordinate pixels of each region.
- 4. Calculate x centroid and y centroid of each region.
- 5. Calculate x and y centroid of entire image.
- 6. Calculating the feature vector

#### 1. Dividing the image into regions

After segmenting the image the image is divided into regions. This step is needed since it becomes easy to get the region of interest. It also becomes easier to get the area and centroid of the region of interest

#### 2. Calculating area of each regions

Here we count the number of pixels which satisfies some condition.

We use the following formula for it:

area = 
$$\sum f(x, y)$$
   
  $\begin{cases} +1 & \text{if } (g < 90 \&\& r > 5 \&\& b > 5) \\ +0 & \text{if } (g > 90 \&\& r < 5 \&\& b < 5) \end{cases}$ 

Where,

f(x, y) gives the value of the pixel at coordinates x, x, is the height and y is the width of each block made, r, g, and b are the red, green, and blue values of pixel of image respectively3.

## 3. Calculate the number of x-coordinate pixels and y coordinate pixels

Here we are summing up all the values of x-coordinate pixel and y-coordinate pixel which satisfies above condition.

# 4. Calculating x-centroid and y-centroid of each region

Here it calculates the x-centroid and y-centroid of each region.

The following is the formula to calculate x-centroid of each region:

$$xcentroid = \frac{sumx}{area}$$

sumx is the number of pixel in x-coordinate area is the area of each region The following is the formula to calculate y-centroid of each region:

$$ycentroid = \frac{sumy}{area}$$

sumy is the number of pixel in y-coordinate area is the area of each region

## 5. Calculating the x-centroid and y-centroid of entire image

Here it calculates the x-centroid and y-centroid of entire image region.

#### 6. Calculating the feature vector

Here we are calculating the feature vector. Feature vector is vector of moments

Moments:

Image moment is average or moment of image pixel's intensities or moment function which is usually has some properties related to image. These properties can be area, centroid and so on. Moments are suitable in shape learning. Zero to third order moments are applied for shape learning and orientation<sup>18</sup>.

Formula to calculate moment is as follows:

$$\mathbf{M}_{pq} = \sum_{x} \sum_{y} x^{p} y^{q}$$

Here,

$$p, q = 0,1,2$$

x = (number of x-coordinate pixel) - (xloc)

y = (number of y-coordinate pixel) - (yloc)

Zeroth order moment gives the information of the area in the foreground or it counts the total number of pixel in the region of interest.

 $M_{10}$  gives the first order moment along the x-axis<sup>8,15</sup>.

M<sub>01</sub> gives the first order moment along y-axis<sup>8, 15</sup>.

 $M_{20}$  gives the second order moment along the x-axis<sup>17</sup>.

 $M_{02}$  gives the second order moment along the y-axis<sup>17</sup>.

#### 3.6 Feature Vector Data Set

It is the data set which will have the name of the cell, the file path was the cell is stored and the feature vector of the cell.

#### 3.7 Similarity of Cells

The features of the image stored in the feature vectors are compared to obtain a match. The feature vector of the test image is associated with the feature vector of the different WBC cells in the trained images feature data set. To relate the two feature vector we calculate the Coefficient of Correlation (CoC) of the two feature vectors.

### 3.7.1 Coefficient of Correlation

Correlation helps in judging resemblances between the two measured vector quantities which analysis whether the two quantities are identical or they are completely different. Pearson's correlation coefficient is denoted as r. It was developed by Karl Pearson. It is widely used in pattern Learning and computer vision<sup>2</sup>.

Steps to calculate coefficient correlation:

Considering two feature vector u, v.

To find average of the two feature vectors u,v.

Calculating the difference vector of u,v.

Calculating unit vector.

Calculating Correlation (Similarity) of the two unit Vector.

To find average of the two feature vectors u,v.

Determine the average of the feature vector by using,

$$\bar{u} = \frac{\sum_{i} u_{i}}{n}$$

Where 'n' is the number of values in the vector u, i = I, 2, 3...n

$$\bar{v} = \frac{\sum_{i} v_{i}}{n}$$

n is the number of values in the vector v i=1,2,3...n

#### 3.7.2 Calculating the Difference Vector of u,v.

Here we subtract each element of the feature vector by the average of the feature vector.

Formula to find difference vector for feature vector u:

$$\{u\}$$
  $-\overline{u}$ 

Formula to find difference vector for feature vector v,

$$\{v\} - \bar{v}$$

Where,

{u} is the feature vector.

{v} is the feature vector.

 $\overline{u}$  is the average of feature vector  $\mathbf{u}^{\underline{o}}$ .

 $\overline{v}$  is the average of feature vector  $\mathbf{v}^{\underline{6}}$ .

# 3.7.3 Calculating Unit Vector

Here we calculate the unit vector which is the difference vector divided by the length.

Formula for calculating the length of the vector u:

$$|u| = \sqrt{u_1^2 + u_2^2 + \dots + u_n^2}$$

Where

|u| is the length of the u vector.

 $u_1, u_2, \dots, u_n$ , is the element of the vector u. Formula for calculating the length of the vector v:

$$\overline{v} = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

Where,

|v| is the length of the v vector.

 $v_1, v_2, \dots, v_n$ , is the element of the vector v.

Formula to calculate the unit vector of the feature vector u:

$$\frac{\{\{u\}-\overline{u}\}}{|u|}$$

Where,

 $\{\{u\} - u\}$  is the difference vector, of feature vector u. |u| is the length of the feature vector u.

To evaluate the unit vector of the vector v,

$$\frac{\{\{v\} - \overline{v}\}}{|v|}$$

 $\{\{v\} - \overline{v}\}\$  is the difference vector, of feature vector v. |v| is the length of the feature vector  $v^{16}$ .

# 3.7.4 Calculating Correlation (Similarity) of the Two Unit Vector

Here we find the dot product of the two unit vector. Formula to calculate the dot product:

$$\frac{\{\{u\} - \bar{u}\}}{|u|} \bullet \frac{\{\{v\} - \bar{v}\}}{|v|}$$

 $\{\{u\} - \overline{u}\}$  is the difference vector of feature vector u. |u| is the length of the feature vector  $\mathbf{u}^{16}$ .

 $\{\{v\} - \overline{v}\}$  is the difference vector of feature vector v.

|v| is the length of the feature vector v.

If the coefficient of correlation is calculated as value equal to 1 than the two images are absolutely identical<sup>2,16</sup>.

If the coefficient of correlation is calculated as value *equal to* 0 than the two images are completely uncorrelated $\frac{1}{2}$ .

If the coefficient of correlation is calculated as value *equal* is -1 than the two images are completely anti-correlated<sup>2,16</sup>.

# 4. Experimental Results and Comparisons

The images of the different types of white blood cells before and after segmentation are displayed<sup>2,9</sup> (Figure 2a,b).

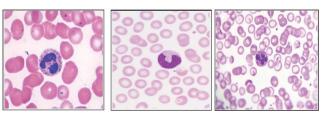
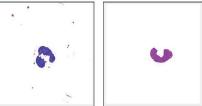


Figure 2(a): Images before segmentation.



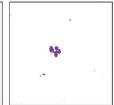


Figure 2(b): Images after segmentation.

Table1 shows the Feature vector extraction based on moments to check for similarity of nucleus of the cell image.

**Table 1.** Moment-Features extracted of the nucleus images

Index	Moment- Features extracted							
	Neutrophil	Eosinophil	Neutrophil					
0	-34.8620241438804	359.949983476756	330.750800119743					
1	-19.8785861245466	87.9896148284539	209.780970184697					
2	-34.8620241438804	359.949983476756	330.750800119743					
3	-9.67821454052283	-6.87800699246756	100.655958444535					
4	-8.73614806499537	-44.4491320613914	11.9955971595661					
5	-9.67821454052283	-6.87800699246756	100.655958444535					
6	-2.36431396935704	68.8598205413294	51.5465136155408					
7	-1.97650444496769	25.2227938912886	20.3024006230087					
8	-2.36431396935704	68.8598205413294	51.5465136155408					
9	12.7203111490997	-428.030726137753	-15.9587519800862					
10	9.52572587796256	-194.210826886912	-48.253805145794					
11	12.7203111490997	-428.030726137753	-15.9587519800862					
12	5.62660954295376	451.116627518597	82.5740964091076					
13	4.36955812338282	150.004307942766	34.7233425198393					
14	5.62660954295376	451.116627518597	82.5740964091076					
15	0.511968051063357	-106.652820130347	8.91689161299849					
16	0.492900144549501	-34.4764378788584	-3.24696187831646					
1/	0.511968051063357	-106.652820130347	8.91689161299849					
18	-3.07171711206663	227.8791075915	70.9409818003828					
19	-0.20533098141163	33.9439614565961	51.2295634421453					
20	-3.07171711206663	227.8791075915	70.9409818003828					
21	-0.47926883744345	-34.464554495706	15.7346850218682					
22	-0.19096228769876	-13.4143493440435	-5.86875020256506					
23	-0.47926883744345	-34.464554495706	15.7346850218682					
24	-0.11564355205388	45.5981942666388	10.2923102017382					
25	-0.05397835611633	8.10035243970682	4.00706440902942					
26	-0.11564355205388	45.5981942666388	10.2923102017382					

Table 2 shows the result class in which the test fall into. The result of testing is done on different images.

**Table 2.** Result of testing

Techniques	True positive	False negative	False positive	True negati ve	Sensitivit y (%)	Specificity (%)	Accuracy (%)
Proposed Technique	35	15	15	20	70	58	65

# 5. Conclusion

The Automatic Detection of WBC Cell Count is useful for the pathologist for detecting the cells. This system will reduce the errors which can occur during detection of cells manually. By comparing the feature vectors of the testing image feature data set and the trained image feature vector dataset we can analyse the relation between the vectors. The system finds the correlation between the two vectors and based on the similarity value it classifies the cell type<sup>4</sup>.

This system would be useful in the pathology lab for fast detection of the cells.

# 6. References

- Amruta Pandit, Shrikrishna U Kolhar, Pragati Patil. Survey on automatic RBC detection and counting, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. Jan 2015; 4(1). https://www.researchgate.net/publication/276248973\_ Survey\_on\_Automatic\_RBC\_Detection\_and\_Counting.
- Avneet Kaur, Lakhwinder Kaur, Savita Gupta. Image recognition using coefficient of correlation and structural SIMilarity index in uncontrolled environment, International Journal of Computer Applications (0975 - 8887). Dec 2012; 59(5). https://doi.org/10.5120/9546-3999.
- Rashmi Mothkur, Poornima Km. Machine Learning wills Transfigure Medical Sector: A Survey, International Conference on Current Trends towards Converging Technologies (ICCTCT); 2018. https://doi.org/10.1109/ ICCTCT.2018.8551134.
- Pramit Ghosh, Debotosh Bhattacharjee, Mita Nasipuri. Blood Smear Analyzer for White Blood Cell Counting: A Hybrid Microscopic Image Analyzing Technique, Applied Soft Computing. 2016.
- 5. Deepak Gupta, Jatin Arora, Utkarsh Agrawal, Ashish Khanna, Victor Hugo C. de Albuquerque. Optimized binary bat algorithm for classification of white blood

- cells, Measurement. 2019. https://doi.org/10.1016/j. measurement.2019.01.002.
- Ming-Ho Hsiao. Searching the video: An efficient indexing method for video retrieval in peer to peer network, Lecture Notes in Computer Science. 2007. https://doi. org/10.1007/978-3-540-69429-8\_18.
- Min Wang, Rong Chu. A Novel White Blood Cell Detection Method Based on Boundary Support Vectors, IEEE International Conference on Systems, Man and Cybernetics; 2009. https://doi.org/10.1109/ICSMC.2009.5346736. PMCid: PMC2773724.
- Alisha Pradhan, Deepak BBVL. Design of intangible interface for mouseless computer handling using hand gestures, Procedia Computer Science. 2016. https://doi. org/10.1016/j.procs.2016.03.037.
- Jung, Hye-Won, Lee Sang-Heon, Martin Donnelley, David Parsons, Ivan Lee. Automated detection of circular marker particles in synchrotron phase contrast X-ray images of live mouse nasal airways for mucociliary transit assessment, Expert Systems with Applications. 2017. https://doi. org/10.1016/j.eswa.2016.12.026.
- Hao Zhou, Yajian Huang, Sai Huang, Kai Chen, Yifan Zhang, Zhiyong Feng. Comprehensive Time-Frequency-Spatial Spectrum Measurement and Analysis of TV Band in Beijing, IEEE Wireless Communications and Networking Conference (WCNC); 2015. https://doi.org/10.1109/ WCNC.2015.7127810.
- Leeson JJ. Novel Feature Vector for Image Authentication, International Conference on Multimedia and Expo ICME 03 Proceedings (Cat No 03TH8698) ICME-03; 2003.
- Lee, Howard, Yi-Ping Phoebe Chen. Cell morphology based classification for red cells in blood smear images, Pattern Recognition Letters. 2014. https://doi.org/10.1016/j. patrec.2014.06.010.
- Saraswat, Mukesh, Arya KV. Automated microscopic image analysis for leukocytes identification: A survey, Micron. 2014. https://doi.org/10.1016/j.micron.2014.04.001. PMid: 25041828.
- 14. Rakibull Ahasan, Ahasan Ulla Ratul, Bakibillah ASM. White Blood Cells Nucleus Segmentation from Microscopic Images of Strained Peripheral Blood Film During Leukemia and Normal Condition, 5th International Conference on Informatics, Electronics and Vision; 2016. https://doi.org/10.1109/ICIEV.2016.7760026.
- Rokhsana Titlee, Ashfaq Ur Rahman, Hasan U. Zaman, Hafiz Abdur Rahman. A Novel Design of an Intangible Hand Gesture Controlled Computer Mouse using Vision Based Image Processing, 3rd International Conference on Electrical Information and Communication Technology (EICT); 2017. https://doi.org/10.1109/EICT.2017.8275171.

- 16. Sapna S, Renuka A. Techniques for Segmentation and Classification of Leukocytes in Blood Smear Images - A Review, IEEE International Conference on Computational Intelligence and Computing Research (ICCIC); 2017. https://doi.org/10.1109/ICCIC.2017.8524465.
- 17. Stephen Adams, Graham Crannell, Ann Bolcavage, Roy McIntyre, Peter A. Beling. A Condition Monitoring System for Low Vacuum Plasma Spray using Computer Vision, IEEE International Conference on Prognostics and Health Management (ICPHM); 2018. https://doi.org/10.1109/ ICPHM.2018.8448464. PMCid: PMC6099332.
- 18. Jacobs IS, Bean CP. Fine Particles, Thin Films And Exchange Anisotropy, in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic; 1963. p. 271-350. https://doi. org/10.1016/B978-0-12-575303-6.50013-0.
- 19. Sriharsha KV, Alphonse PJA. A Single Tier Homogeneous Visual Sensor Network for Real Time face Recognition, IEEE International Conference on Computational Intelligence and Computing Research; 2017. https://doi. org/10.1109/ICCIC.2017.8524274.