Segmentation and Feature Extraction of Lymphocytes WBC using Microscopic Images

Dr. Mazin Z. Othman
Department of Computer and Information Engineering
Mosul's University, Electronic Engineering College

Alaa B. Ali
Department of Computer and Information Engineering
Mosul's University, Electronic Engineering College

Abstract—**This paper presents a proposition for an algorithmic procedure to isolate and count the lymphocytes White Blood Cell (WBC) form microscopic Images. The process involves segmentation of cells, scanning algorithm ,feature extraction, and recognition of lymphocyte cells. The scanning algorithm returns the number and location of candidate area in WBC images. For feature extraction, a combination of the shape feature moment invariants and the roundness are found to have an excellent recognition accuracy for identifying the lymphocyte cells from other WBC types. The whole work has been developed using MATLAB environment, in particular the Image Processing Toolbox.**

Keywords—Segmentation, White Blood Cells(WBC), Feature Extraction

I. INTRODUCTION

The WBCs, or leukocytes are cells of the immune system involved in defending the body against both infectious disease and foreign materials[1]. Human blood consists of five types of white blood cells, namely, Neutrophils (40-60%), Lymphocytes (20-40%), Monocytes (2-8%), Eosinophils (1-4%), and Basophils (0.5-1%) as shown in Figure (1). Recognizing the WBC type and their number are of important issue for medical diagnosis and useful measure, which indicates the health status of the body. For example, excess of lymphocytes (above 30000 cells per micro-liter) may indicate a Lymphocytic Leukemia.
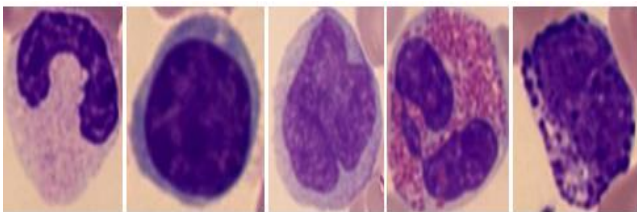

Fig. 1. Mature WBCs, Left to right: Neutrophil, Lymphocyte,

Monocyte ,Eosinophil, and Basophil.

However, manual classification of blood cells is time-consuming ,cumbersome process and susceptible to error due to the different morphological features of the cells [2]. Therefore, a fast and automated method for identifying the different blood cells is required. Researchers had applied many algorithm to segment (WBC) from blood smear image. Among those on can mention the followings:-

Joshi, M, D. et al. investigated the problem of blood slide image segmentation followed by feature extraction to detect leukemia. Shape features of nucleus such as area, perimeter, circularity etc. are considered for better accuracy of detection [2]. However, their technique gave only 93% recognition accuracy. Ghosh et al. presented a marker controlled watershed segmentation embedded with morphological operator to segment WBC and its nuclei from light microscopic image of blood samples . They used only four statistical significant features known as Naïve Bayes classifier for pattern identification [3]. The proposed idea reflected 83.2% overall classification accuracy. Madhloom et al. suggested an automatic algorithm for the detection and classification of leukocytes [4] . This approach utilized a combination of automatic contrast stretching supported by image arithmetic operation, minimum filter and global threshold techniques. Results showed that the proposed method managed to obtain accuracy between 85-98% . However, this algorithm has a major drawback as its accuracy is highly and critically depends on the initial contrast of the image. Putzu,L and Ruberto,C. used the complete and fully automatic method for white blood cells identification from microscopic images. The proposed method firstly individuates WBC from which, subsequently, nucleus and cytoplasm are extracted [5]. The performances of the proposed method are excellent in most cases. The worst results achieved from images that showed significant overlapping between leukocytes . Generally the average accuracy almost 92%. Mohamed, M and Far, B. presented a new white blood cell nucleus segmentation algorithm is an adaptation of the one presented by Madhloom et al. [4], Small false entities were removed by the utilization of minimum size constraint of 50% RBC size[6] In this scheme, the performance of segmentation technique was found to be superior. Fatichah, C et al. presented a system architecture of WBC type classification. The first process was a manually segmented of WBC images. The second process was the feature extraction. The third process was the support vector machine (SVM) classification method [7]. In their approach, four textural feature extraction methods i.e. granulometry ,gray level co-occurrence matrix ,local binary patterns and local fuzzy patterns were presented. The best accuracy of all feature extraction methods was obtained on polynomial 2 kernel function of support vector machine. This paper focus on recognition of lymphocyte cells among WBCs by segmentation and feature extraction .This paper organized as follows:- section 2 discusses image preparation. Hu's moment invariants and roundness were used as feature extractors, including scanning algorithm to check candidate area and segmentation. This is illustrated in section 3. Section 4 reports the lymphocyte recognition scheme, while Section 5 highlights some concluding points of our study.

II.MMICROSCOPIC IMAGES PREPARATION

In this practical session, the slide with the stained blood smear is already in place in the slide stage of your microscope, and the 100x oil immersion lens is in focus. One should be very careful when using this high power lens, since it leaves extremely low gap between the front of the lens and the glass slide surface. As a thumb of the rule – never touch the macro-focusing knob for any focus-improving action. Never change the objective lens to any other magnification as the immersion oil on the blood film may contaminate the other lenses and reduce their optical quality. Scan the blood smear in a meandering way, as shown in Figure (2), starting from the thinner edge of the blood film.
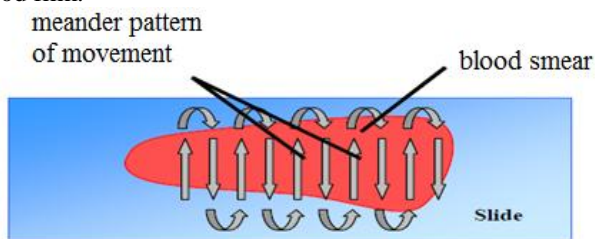

Fig. 2. Search pattern on a blood smear.

III.THE PROPOSED PROCEDURE FOR ISOLATION AND COUNTING LYMPHOCYTE

The algorithmic procedure is applied to each image and the resultant segmentation and counting of lymphocyte WBC is then achieved. The proposed algorithmic procedure includes three primary steps . The first step is image segmentation. The second step is the scanning algorithm that returns the number and location of each WBC . The third step is extracting descriptive feature measured from the segmented cells. Figure 3 illustrate these three steps of the proposed algorithmic procedure.
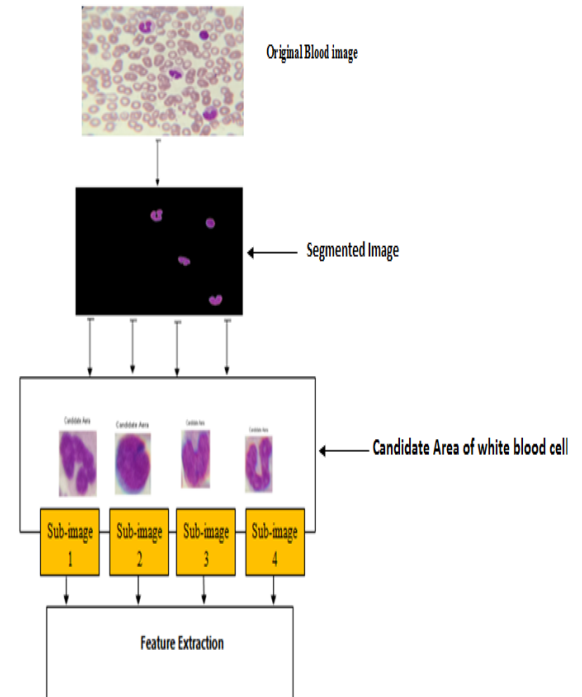

Fig. 3 . Block diagram of a proposed algorithmic procedure.

A. Segmentation

In the segmentation step , the nucleus is separated using threshold as explained in [6] . One can summarized the main steps for segmentation as follows:-

Step 1: Read the color blood slide image to the system

Step 2: Convert the color image into gray scale image.

Step 3: Enhance contrast of the gray scale image by histogram equalization (result A).

Step 4: To adjust image intensity level apply linear contrast stretching to gray scale image (result B).

Step 5: Obtain the image I1=result A+ result B to brighten all other image components except cell nucleus.

Step 6: Obtain the image I2=I1- result A to highlight the entire image objects along with cell nucleus.

Step 7: Obtain the image I3=I1+I2 to remove all other components of blood with minimum effect of distortion over nucleus.

Step 8: To reduce noise, preserve edges and increase the darkness of the nuclei implement 3-by-3 minimum filter on the image I3.

Step 9: Apply a global threshold Otsu's method on image I3.

Step 10: Using the threshold value in above step convert I3 to binary image.

Step 11: To remove small pixel groups use morphological opening.

Step 12: To form objects connect the neighbouring pixels.

Step 13: By applying the size test removal of all objects that are less than 50% of average RBC area is done. It is observed that this method of segmentation yields better results than that of previous methods [6] .

The resultant image from the segmentation step is shown in Fig. (4)

Fig. 4 . The segmented nucleus region.

B. Proposed scanning algorithm

The purpose of the scanning algorithm is to check candidate area of white blood cell . Before applying the proposed method to an image, the image is pre-processed and segmentation to become a binary image. An image is scanned vertically and horizontally to locate four values (top, bottom, left, and right) building a bounding box which is a minimal rectangle that covers the cell region. Let "0"denote a background pixel and "1"be a nucleus pixel. In scanning the image vertically from top to bottom , the summation of row greater than one is indexed to locate the top most pixel of the cell region. After that scanning is continue until summation of all row would equal to zero indicating the bottom end of the cell . The algorithm repeats similar scanning horizontally, forming a closed structure. This process optimizes the algorithm and increases the processing efficiency. This algorithm returns the number and the location of candidate area in WBC images. Applying such vertical and horizontal scanning procedure to an image containing , for example four different WBCs, ends up with four sub-images bounded as follows:-

Sub-image 1 location = Top: 162   Bottom: 294
Left: 552   Right: 680
Sub-image 2 location = Top: 700   Bottom: 799
Left: 845   Right: 975
Sub-image 3 location = Top: 265   Bottom: 372
Left: 1142   Right: 1239
Sub-image 4 location = Top: 1138   Bottom: 1252
Left: 1175   Right: 1316

The proposed scanning algorithm can be summarized in the following steps:

Step 1: Set the four locations (Top, Bottom, Left and Right) to initial value 0.
Step 2: Scan the image vertically from top to bottom pixel of the cell region
Step 3: Check if summation of all row greater than one then record index which
represented the top location of WBC cell.
Step 4: Scanning continues until summation of all row equal to zero . Draw a
Vertical line along this pixel representing the bottom of the blood image.
Step 5: Start horizontal scan from the left side of the image to the right most pixel of the cell region.

Step 6: Check if summation of all column greater than one then record index which represented the left location of WB cell.
Step7: Continue with scanning to the next pixel until summation of all column equal to zero. Store the column index representing the right side of the blood image.
Step 8 : Stop.

C. Feature Extraction

The choice of image features immensely affects the classifier performance. For a robust classification, the features must characterize each WBC subtype and must be independent of each other for better judgment and comparison.

All calculated image features can be grouped into (i) shape features (areas of cell , nucleus cell perimeter, compactness, roundness, moment…etc). (ii) intensity features histogram (mean, standard deviation, skewness) ,and (iii) texture features (contrast, homogeneity, correlation, entropy) [8]. In this work , the shape feature of the object is adopted since it is found to be effective in distinguishing the lymphocytes from other types of the WBCs . Usually, there are two types of shape descriptors; the contour-based shape descriptors and the region-based shape descriptors [10].

1) Moment invariants(contour-based shape)

The regular moment invariants are one of the most popular and widely used contour-based shape which is derived by Hu [9] who derived a set of invariants using algebraic invariants. The two-dimensional moments of a digitally sampled $M \times M$ image that has gray function f (x, y), (x, y = 0, . . . M − 1) as is given by
Eq. (1)

$$\mu_{pq} = \sum_{x=1}^{x} \sum_{y=1}^{y} (x - \bar{x})^p (y - \bar{y})^q f(x, y)$$
(1)

where$(\bar{X}, \bar{Y})$ are $\bar{X} = m_{10}/m_{00}$ and $\bar{Y} = m_{01}/m_{00}$ which are the centered of the image. Therefore, the central moments are invariant to image translations .

η_{pq} is normalized central moment

$$\eta_{pq} = \mu_{pq}/(\mu_{00})^{\gamma}$$ (2)

$$\gamma = \frac{p+q}{2} + 1$$ (3)

$$\phi_1 = \eta_{20} + \eta_{02}$$ (4)

$$\phi2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}$$ (5)

$$\phi_3 = (\eta_{30} - \eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$ (5)

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$ (6)

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]$$
$$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$
(7)

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2$$
$$+ (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

A set of seven moment invariant features $\phi = [\phi_1, \cdots \phi_7]$ is calculated for the white blood cell images. Hu's seven moment invariants have the desirable properties of being invariant under image scaling, translation, and rotation. It was found that to compute the higher order of Hu's seven moment invariants is quite complex, and to recover the image from the results is deemed to be difficult [9].

2) geometrical shapes(region-based descriptors)

The geometrical shapes descriptor is characterized by the following parameters :

(i) The area - the number of pixels on the interior of the cell.

$$Area = \sum_{i=1}^{n} \sum_{j=1}^{m} b_{ij} \quad (9)$$

Where b_{ij} - is the value of binary image (0,1)at the pixel coordinate(i, j) within an m × n image.

(ii) The perimeter : It is an important feature of an object. Contour based features which ignore the interior of a shape, depend on finding the perimeter or boundary points of the object [10]. If $x_1, \cdots, x_n$ is a boundary coordinate list, then the object perimeter is given by:

$$perimeter = \sum_{i=1}^{N} d_i = \sum_{i=1}^{N} |x_i - x_{(i+1)}| \quad (10)$$

(iii) The roundness: Area-to-perimeter ratio is the measure of roundness or circularity [10]. But local irregularities are not reflected by this feature. It is defined as:

$$Roudness = 4 \pi \, Area / Perimeter \quad (11)$$

A circle gets the value of 1, while objects with bumpy boundaries get lower values.

D. Lymphocytes recognition from microscope images

To demonstrate our proposed procedure for isolating and counting the lymphocytes WBC, a 25 microscopic images are shot from a stained blood smear. Then , the WBCs are segmented from those images. The total sub-images that indicate the whole WBCs are found to be 47 sub-images. In order to classify the lymphocytes WBCs only, the shape features mentioned earlier (the moments invariants and the roundness) are calculated for these 47 sub-images of different types of WBC cells . Afterwards, the following recognition scheme is used :-

If Max(Ø1....Ø7)=Ø7 OR (the roundness≥0.9)
Then the sub image is recognized as a lymphocytes
else

the sub image is recognized as other type of WBC
.

Table 1. The segmentation and the recognition
results

	Sub-Img count	$\phi1$	$\phi2$	$\phi3$	$\phi4$	$\phi5$	$\phi6$	$\phi7$	roundness	Recognition result
1	1	1.4497	0.9270	2.6149	2.5156	5.0027	2.2704	6.0795	0.8857	Yes
	2	0.1128	7.3067	5.5366	6.4185	14.9777	7.7871	13.4538	0.3997	No
	3	0.1193	5.3674	3.4308	5.2575	11.5731	8.1837	10.7919	0.9341	Yes
	4	0.8402	1.6775	0.1805	0.7448	0.5436	1.1792	0.0970	0.6385	No
2	1	0.0854	5.9011	4.5705	4.9148	10.4029	8.1240	10.6617	0.7229	Yes
	2	0.9972	0.4114	1.9319	0.9610	2.1718	0.6106	0.2839	0. 3102	No
3	non									
4	1	0.1876	4.0345	5.2868	6.3730	12.6645	10.9500	13.8727	0.9015	Yes
5	1	0.2092	8.2780	6.4122	6.3019	13.5875	12.6330	14.4038	0.8708	Yes
	2	0.2092	8.2780	6.4122	6.3019	13.5875	12.6330	14.4038	0.9115	Yes
	3	0.0204	10.8056	6.1861	7.3524	11.3381	-8.0363	10.8787	0.5410	No
6	1	0.0239	2.2482	3.4489	3.8404	9.3398	7.5105	9.7268	0.8861	Yes
	2	0.5868	1.8167	0.3806	3.1346	5.0399	6.6635	8.9127	0.731	No
7	1	1.0559	0.1723	0.4098	0.7576	1.5391	0.9200	3.1833	0.9018	Yes
	2	0.2537	3.3727	3.6595	3.6815	7.6037	5.9812	7.9824	0.8906	Yes
	3	0.0416	3.3115	3.2725	5.6927	10.7570	7.8767	11.9434	0.9021	Yes
	4	0.3110	2.1898	4.9940	5.0694	10.2631	6.2775	10.3455	0.8513	Yes
8	1	0.4838	11.6145	1.7223	4.5410	7.9992	11.8733	12.8848	0.7410	Yes
	2	0.6839	0.5509	1.3435	1.5550	2.9271	1.2310	6.4707	0.7633	Yes
	3	0.3474	0.5170	3.5722	3.4700	10.808	5.6363	7.4426	0.3257	No
9	1	0.0239	2.2482	3.4489	3.8404	9.3398	7.5105	9.7268	0.8570	Yes
	2	0.1876	4.0345	5.2868	6.3730	12.6645	10.9500	13.8727	0.7286	Yes
10	1	0.0123	5.5249	3.0753	7.0619	10.3000	7.7157	10.8488	0.9194	Yes
	2	0.0740	4.7789	5.5427	7.6650	14.6744	11.3829	16.3777	0.6310	Yes
	3	0.5012	0.2746	0.5265	2.2684	3.8530	2.7389	3.5685	0.2741	No
11	1	0.2750	4.9965	5.3197	4.7788	11.9136	9.0241	13.8968	0.9195	Yes
12	1	0.0268	1.9420	5.0137	6.5279	14.2361	10.0307	16.5578	0.9320	Yes
	2	0.0740	4.7789	5.5427	7.6650	14.6744	11.3829	16.3777	0.8034	Yes
13	1	0.4894	2.6328	3.5891	8.0097	13.5768	9.3562	14.5070	0.9054	Yes
	2	0.5512	4.4573	1.5159	4.1541	7.3809	8.6954	6.4656	0.9021	Yes
	1	0.4894	2.6328	3.5891	8.0097	13.5768	9.3562	14.5070	0.9054	Yes
14	1	0.1698	3.5923	2.8195	5.6820	10.8350	9.1040	10.6881	0.9317	Yes
	2	0.1122	6.3841	3.1113	2.3689	5.2589	4.0785	7.4105	0.9430	Yes
15	1	0.2039	5.2083	4.3121	8.1351	13.6027	9.2599	13.7986	0.6429	Yes
	2	0.2268	4.8344	0.6668	3.6010	5.4855	5.3169	7.0612	0.1193	No
16	non									
17	1	0.9431	3.4304	2.6912	1.3362	3.7419	4.9886	3.8874	0.9478	Yes
18	2	0.1198	4.5997	5.6029	7.1177	13.5706	0.6129	17.4709	0.9543	Yes
	1	0.4476	7.2090	7.4917	7.7535	15.4512	11.3089	17.2798	0.8410	Yes
19	1	0.2092	8.2780	6.4122	6.3019	13.5875	12.6330	14.4038	0.9476	Yes
	2	0.0740	4.7789	5.5427	7.6650	14.6744	11.3829	16.3777	0.8742	Yes
	3	0.4876	3.7962	4.9811	2.4132	8.1743	6.0007	4.3193	0.2621	No
20	1	0.0740	4.7789	5.5427	7.6650	14.6744	11.3829	16.3777	0.8742	Yes
	2	0.2750	4.9965	5.3197	4.7788	11.9136	9.0241	13.8968	0.9010	Yes
21	1	0.2039	5.2083	4.3121	8.1351	13.6027	9.2599	13.7986	0.9652	Yes
22	1	0.2951	3.4035	6.2860	5.8847	8.9408	6.5163	8.0183	0.7184	No
	2	0.4751	2.1688	0.0109	1.0587	3.2945	2.4344	1.7848	0.6320	No
23	1	0.3021	5.3104	6.8165	8.6468	17.1775	13.6261	18.4547	0.7901	Yes
24	1	0.4410	5.2328	2.5471	3.0038	5.7958	5.9932	6.7558	0.9440	Yes
25	1	0.1128	7.3067	5.5366	6.4185	14.9777	7.7871	13.4538	0.6014	No

This recognition scheme is illustrated in Fig.(5). Moreover, Table (1) lists the features of the 25 images taken from a stained blood smear slide. The table illustrates the number of sub-images that are segmented from each image, for example image one contains four segmented sub-images and image two contains two segmented sub-images and so on. For each sub-image the seven moments invariants features and the roundness feature are calculated. The results obtained from the recognition scheme is pointed as (YES) for the sub-image of a lymphocyte WBC and (NO) for other types of leukocytes .One can clearly discover that using the above recognition scheme , the lymphocytes WBC can be 100% recognized. The total number of lymphocytes are found to be 35 WBCs and the other 12 WBCs are recognized as non- lymphocytes type. To verify these results, the same slide of a stained blood smear was examined by several expert biologists and they proved our obtained results.
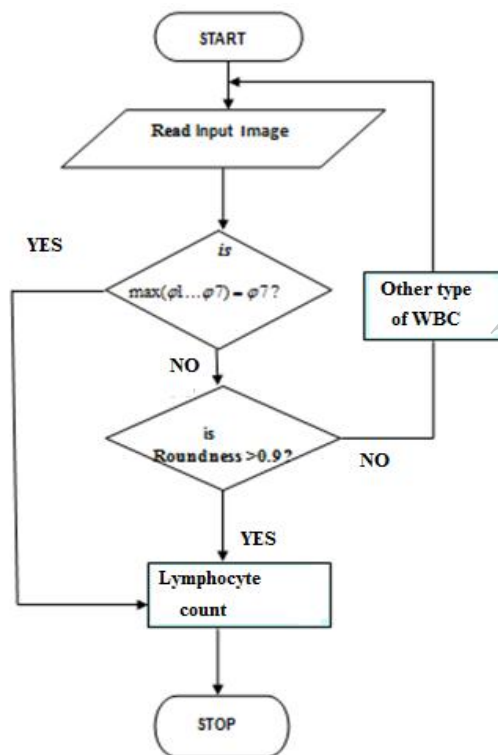

Fig.5. Flow chart of the proposed lymphocyte cells recognition scheme.

ACKNOWLEDGMENT

A. CONCLUSIONS

This paper presents an efficient method to isolate leukocytes or WBCs from macroscopic images. Here, we mainly focused on feature extraction of lymphocyte type ,in order to isolate and count this type of WBC. The proposed method was tested on approximately 25 images from a blood smear slide that contain different type of WBC cells. Experimental results show that the proposed procedure successfully works in segmented the WBC images. Furthermore, the combination of the moment invariants and the roundness gave high significant features that reflects an excellent accuracy in recognizing Lymphocytes from other type of WBCs.

REFERENCES

1. Mathur, A., Tripathi,S.A., and Kuse, M, (2013). Scalable system for classification of white blood cells from Leishman stained blood stain. *J Path Informatics, No*. IP:95.170.211.180.
2. Joshi1,M. D., Karode,A.H., Suralkar,S.R. (2013). White Blood Cells Segmentation and Classification to Detect Acute Leukemia. *International Journal of Emerging Trends & Technology in Computer Science* (IJETTCS),Volume 2, Issue 3.
3. Ghosh,M., Das,D., Mandal,S., Chakraborty and C., Pal,M. (2010).Statistical Pattern Analysis of White Blood Cell Nuclei Morphometry. *IEEE Students Technology Symposium3-4 April*.
4. Madhloom,H.T.,Kareem,S.A.,Ariffin,H.,Zaidan,A.A.,Alanazi,H.O.,Zaidan,B.B.(2010). An Automated White Blood cell Nucleus Localization and Segmentation using Image Arithmetic and Automatic threshold .*journal of Applied Sciences No* .10(11):959-966.
5. Putzu,L.,and Ruberto C. Di.(2013). White Blood Cells Identification and Counting from Microscopic Blood Image, *World Academy of Science, Engineering and Technology* .
6. Mohamed,M.and Far B.(2012) An Enhanced Threshold Based Technique for White Blood Cells Nuclei Automatic Segmentation. *IEEE 14th international Conference on e-Health Networking, Applications and Services Healthcom*.
7. Fatichah,C., Tangel, M. L., Widyanto, M. R., DONG,F., Hirota,K.(2012).Parameter Optimization of Local Fuzzy Patterns for Extracting White Blood Cell Texture Feature. *Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII)* , *No*. 16(3):412-419.
8. Phapale, V., Hajare ,V. D., and Patil ,H. T. (2013).Feature Extraction And Segmentation Of White Blood Cells . *International Journal of Engineering Research & Technology (IJERT)*Vol. 2 Issue 6.
9. Chenl,Q., Petriul,E., Yang, X.(2004). A Comparative Study of Fourier Descriptors and Hu's Seven Moment Invariants for Image Recognition.*IEEE, Canadian Conference on* Volume:1.
10. Thesis by Olsson,B., Wisskirchen,P and Jungmann,M.(2006). Shape Analysis and Measurement for the HeLa cell classification of cultured cells in high throughput screening.School of Humanities & Informatics University of Skövde, Sweden.