ANALYSIS OF LYMPHOCYTE IMAGES FOR DIAGNOSING RHEUMATOID ARTHRITIS USING THRESHOLD SEGMENTATION METHOD WITH SLIDER CONTROL

Chokkalingam Subramanian[1,*], Komathy Karupannan[2], BrahimBelhaouari Samir[3]

[1]Department of Information Technology, Saveetha School of Engineering, Saveetha University, Chennai.

[2]Department of Information Technology, Hindustan University, Chennai.

[3]Department of Mathematics, University of Sharjah, Sharjah
* e-mail:cho_mas@yahoo.com

Abstract: Medical images today help bring to divulge hitherto concealed knowledge about diseases that were, at one time, rarely subject to intense and prolonged scrutiny and analysis. In recent times, however, imaging has gone a long way in helping establish plausibly both the causes and behavioural patterns of a given disease. The objective of this paper is to constitute a series of techniques to detect accurately, in lymphocytes from blood smear images, the occurrence of Rheumatoid Arthritis (RA). This paper has used computer-aided diagnosis for accuracy and consistency, and the threshold segmentation method with slider control as a proposed segmentation method for lymphocyte extraction from blood smear images, precisely because it performs much better than existing segmentation methods. The paper discusses critical medical parameters - such as area, perimeter, circularity, roundness and solidity - from segmented lymphocytes, and also describes the ADTree method governing classification and decision rules. The final part of the paper deals with a case study on datasets of inflamed and non-inflamed types (for three different medical cases) using correlation and Analysis Of Variance (ANOVA) techniques to discover the homogeneity and relationships that exist between the critical parameters listed above for identifying the status of RA.

Keywords: Computer-Aided Diagnosis, Edge Detection, Histogram Smoothing, Rheumatoid Arthritis, Statistical Analysis

Introduction

Rheumatoid Arthritis (RA) is a lingering systemic inflammatory disease involving, predominantly, the peripheral synovial joints. Figure 1 includes two different shapes of lymphocyte extracted from a blood smear. The first image Figure 1 (a) represents a normal-shaped lymphocyte, while Figure 1 (b) represents an inflamed lymphocyte being considered for further RA analysis (Aman Kumar Sharma, 2011).

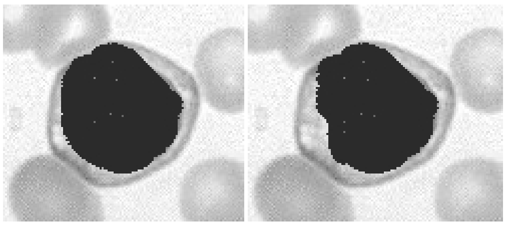


Figure 1(a) Normal Lymphocyte (b) Inflamed Lymphocyte

Image preprocessing is used to enhance the quality of the image obtained from various sources to ensure that all requirements are satisfied before further processing can take place. Image pre-processing techniques, like de-noising, can be used to remove noise from an acquired image (Bharanidharan, 2012). Noise in an image can be grouped, based on the underlying contents. Stray marks, marginal noises, and salt-and-pepper noises are independent of the size and location of the underlying content Noises naturally present in microscopic medical images can be removed using any of the following filters: mean, median, Gaussian, and Wiener (Chokkalingam , 2014). A set of 390 images, comprising 227 inflammatory and 163 non-inflammatory types (Chokkalingam, 2014) was taken for the purpose of analysing the condition of RA, Each noise-generating image has been filtered using different filters. A comparative study of the relative performance of each filter was made alongside, with the best filter discovered at the end of the process being studied at length. The filtered image is compared against the original image using the following quality measures: Parameter Peak Signal-To-Noise-Ratio (PSNR), Mean Square Error (MSE), Normalized Correlation (NC) and Normalized Absolute Error (NAE)(Krishnapuram, 1993; NiponTheera-Umpon, 2007). When compared to the Wiener filter, the median filter performs better. Consequently, the median filter is reckoned to be optimal for microscopic (electronic) blood smear images (Pavlova, 1996; Tabrizi, 2010).

Materials and Methods

Segmentation is an essential part of obtaining the Region of Interest (ROI), since the results obtained depend entirely only on the ROI. In this model, therefore, a new segmentation method proposed - the threshold segmentation method with slider control - has been introduced. In this method, values are assigned to the slider controls and images segmented based on the slider values. RGB images are converted into grey scale images, with image sizes in the form of a matrix that can be assigned rows and columns. The slider control is required to process upto*n* number of rows and columns, and checks each time whether the row and column values are greater than the slider values - and if they are, they can be converted into white pixels; otherwise they can continue to remain black.

Algorithm : Threshold segmentation method with slider control

Input : Microscopic blood smear image of size m x n

Output : Segmented lymphocyte image

1. Let the input image be I (m,n).

2. The RGB image is converted into a greyscale image.

3. Calculate a mean value of input image , say T_m

4. Let the pixel values be $p_0, p_1, p_2, \ldots, p_N$; g denotes grey scale value; N be the maximum pixel value of the image.

5. An initial threshold guess at t, where t is the median value of the image

$$\sum_{g=0}^{t} p_g >= \frac{n^2}{2} > \sum_{g=0}^{t-1} p_g$$

where p_g is the sum of pixel value of grey scale image and n^2 is the number of pixel in the image

6. Threshold value t is assigned to slider control. Calculate mean pixel value under two conditions namely, less than and greater than the threshold value separately

(a) for values less than or equal to t, T_{low} is calculated as

$$a = \sum_{g=0}^{t} g p_g \quad b = \sum_{g=0}^{t} p_g$$

$$T_{low} = a/b$$

(b) and for values greater than or equal to t, T_{high} is calculated as

$$c = \sum_{g=t+1}^{N} gp_g \quad d = \sum_{g=t+1}^{N} p_g$$

$$T_{high} = c/d$$

(c) Re-estimate t between T_{low} and T_{high}

$$T_{new} = \frac{T_{low} + T_{high}}{2}$$

Repeat steps 6(a), (b) and (c) until T_{new} converges to a single value.

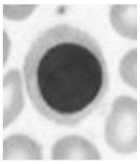
7. Assign T_{new} to slider control

8. If g >T_{new}then the image pixel values are converted into white color (255) else convert to black (0).

Figure 2 (a) shows segmented image using the gradient method, Figure 2 (b) displays segmented image using the proposed threshold segmentation method with slider control and Figure 2 (c) illustrates the segmented image using the global threshold segmentation method. Table 1 illustrates the existing method is evaluated by means of comparisons with other segmentation methods such as the gradient method and the global threshold method. It is found that the dice similarity measurement is very high - while the white count is low and the black count is high - for the threshold segmentation method with slider control, compared with the existing segmentation methods.

S.NO.	Gradient Method			Global Threshold Method			Proposed Method		
	DSM	WC	BC	DSM	WC	BC	DSM	WC	BC
1	0.9038	0.1755	0.8245	1.1213	0.1689	0.8311	1.9121	0.1629	0.8371
2	0.8692	0.2314	0.7686	1.0012	0.2058	0.7942	1.8982	0.1348	0.8652
3	0.9428	0.1083	0.8917	1.2134	0.1118	0.8882	1.9423	0.1033	0.8967
4	0.8710	0.2285	0.7715	1.0034	0.2252	0.7748	1.8994	0.2081	0.7919
5	0.8989	0.1837	0.8163	1.0097	0.1897	0.8103	1.9014	0.1755	0.8245

Table 1 Comparison of Threshold Segmentation Method with Slider Control and Existing Methods



Lymphocyte with tiny particles

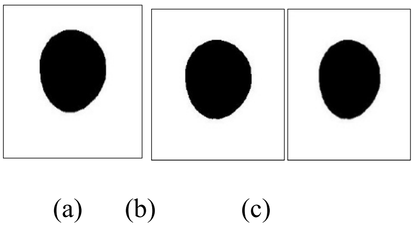


Figure 2 (a) Segmented image using the gradient method (b) Segmented image using the proposed threshold segmentation method with slider control (c) Segmented image using the global threshold segmentation method

Feature Extraction

Features like area, perimeter, circularity, roundness, and solidity are calculated and the sample values shown in Table 2.The area, determined by counting the total number of non-zero pixels within the image region, is given by

$$\text{Area} = \frac{(\text{Perimeter})^2}{\text{Compactness}} \qquad (1)$$

The perimeter, estimated by calculating the distance between successive boundary pixels, is given by

$$\text{Perimeter} = \sqrt{(\text{Compactness} * \text{Area})} \qquad (2)$$

A circularity ratio measures the compactness of a shape.

$$\text{Circularity} = \frac{4\Pi * (\text{area})}{(\text{perimeter})\ 2} \qquad (3)$$

Solidity is the ratio of the actual area and convex hull area, and is also an essential feature for blast cell classification. It is estimated as

$$\text{Solidity} = \frac{\text{Area}}{\text{Convex Area}} \qquad (4)$$

Roundness is a measure of the sharpness of a particle’s edges and corners, calculated as

$$\text{Roundness} = \frac{4 * \text{Area}}{(\pi * \text{Major axis})^2} \qquad (5)$$

Table 2. Features extracted from lymphocytes for the analysis of RA

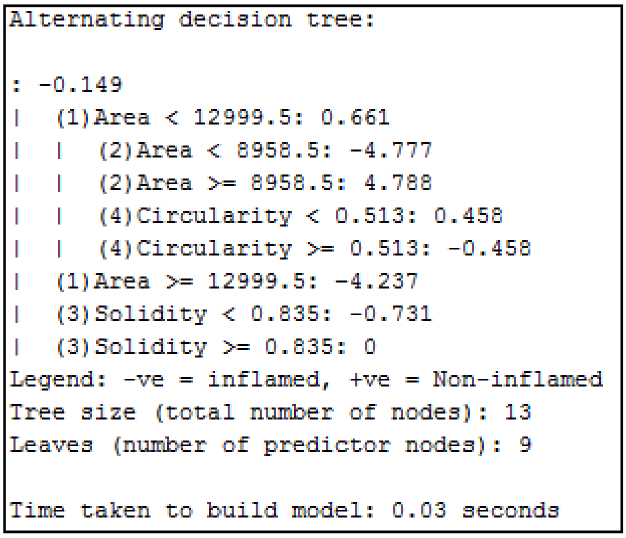
Image	Area	Perimeter	Circularity	Solidity	Roundness	Result	Level
1	6581	359	0.6413	0.9167	0.0585	Inflamed	High
2	7583	398	0.6013	0.904	0.067	Inflamed	Moderate
3	7679	405	0.5880	0.9028	0.0672	Inflamed	Moderate
4	9036	471	0.5116	0.8866	0.0794	Non-inflamed	Level 0
5	9069	473	0.5091	0.8852	0.0792	Non-inflamed	Level 0
6	13479	640	0.4133	0.8294	0.1109	Inflamed	Low
7	13519	641	0.4133	0.8289	0.1149	Inflamed	Low
8	14876	674	0.4113	0.8117	0.142	Inflamed	Moderate
9	14957	676	0.4111	0.8107	0.1252	Inflamed	Moderate
10	24753	864	0.4165	0.6854	0.1919	Inflamed	High

Results and Discussion

Classification

For the purpose of detecting RA, lymphocytes are classified into two categories: inflammatory and non-inflammatory, and rules generated for the classification process.

The ADT automatically generates positive and negative values for all parameters, and all attribute values are interlinked; based on which decision trees are classified. The clustering results for inflammatory and non-inflammatory types of images are given in Figure 3. The clustering is performed by considering all the five attributes, where the difference shows the classification between the inflammatory and non-inflammatory types The Minitab tool (http://www.minitab.com/en-us/products/minitab) is used to plot contour interaction and study its main effects, also enabling improved quality and providing the confident assurance that the right predictions have been made.



```
Alternating decision tree:

: -0.149
|  (1)Area < 12999.5: 0.661
|  |  (2)Area < 8958.5: -4.777
|  |  (2)Area >= 8958.5: 4.788
|  |  (4)Circularity < 0.513: 0.458
|  |  (4)Circularity >= 0.513: -0.458
|  (1)Area >= 12999.5: -4.237
|  (3)Solidity < 0.835: -0.731
|  (3)Solidity >= 0.835: 0
Legend: -ve = inflamed, +ve = Non-inflamed
Tree size (total number of nodes): 13
Leaves (number of predictor nodes): 9

Time taken to build model: 0.03 seconds
```

Figure 3 ADT for inflammatory and non-inflammatory types of images

Roundness, skewness and median are taken as significant attributes in this model. It can be deduced from these attributes whether or not the lymphocytes are inflamed, taking into consideration the constraints that follow. In the ADTree, the negative represents the inflamed and the positive for the non-inflamed. The tree size has been calculated based on the total number of nodes and predictor nodes, and is used to define the leaf node.

Conditions for inflamed:

a). If solidity is less than 0.835, area less than 8995, and perimeter lies between 494 and 502, then the image is of the inflammatory type.

b). If solidity is less than 0.835 and area less than 11868, then the image is of the inflamed type.

c). If solidity is less than 0.835 and roundness less than 0.079, then the image is of the inflamed type.

d). If solidity is less than 0.835 and roundness less than 0.089, then the image is of the inflamed type.

e). If an area is greater than or equal to 12999, it clearly denotes that the images are of the inflammatory type.

f). If an area is less than 12999, it clearly defines the images as being of the inflammatory type.

Conditions for non-inflamed.

g). If solidity is greater than or equal to 0.835, area greater than or equal to 8995, perimeter less than or equal to 494, and perimeter greater than or equal to 502, then the image is of the non-inflamed type.

h). If solidity is greater than or equal to 0.835 and area greater than or equal to 11868, then the image is of the non-inflamed type.

i). If solidity is greater than or equal to 0.835 and roundness greater than or equal to 0.079, then the image is of the non-inflamed type.

j). If solidity is greater than or equal to 0.835 and roundness greater than or equal to 0.089, then the image is of the non-inflamed type.

Case Studies

The case study should be treated as a process that includes patients who are representatives of the target population so as to evaluate the degree to which the diagnosed dataset meets specific parameters - such as perimeter, area, roundness, solidity and circularity. The following section discusses in detail data profiles, significant levels and statistical analysis of the three types of datasets.

Data Profile

As part of this study, many doctors, clinics and labs have been approached and data certified, inclusive of

- Certified image data from rheumatologists;
- Portability to other tools like Excel & Database, and
- Precise clinical data alone to be taken into consideration.

The datasets are primarily divided into two types:inflamed and non-inflamed. The inflamed type has been further divided into low, moderate and high. The following are the two major types of datasets analysed as part of this case study:

- high-normal dataset
- moderate-moderate dataset

The high-normal dataset has sets of image data covering multiple visits of a patient aged about 32. This particular dataset contains all the required clinical parameters for the badly-affected arthritic patient in question who, having undergone a course of treatment spread over 2 years and 3 months, has made significant progress.

The moderate-moderate dataset contains all the required clinical parameters for a female patient, 63 years old, with a history of moderate RA symptoms. She had undergone medical treatment for a year and 8 months and noticed no improvement whatsoever, continuing instead to be subject to two hours of morning stiffness, flares involving the hands, complaints of swollen, painful joints and fatigue.

Significant Levels

Four significant levels have been identified and proposed for study as part of this research, including classifying and clustering the datasets appropriately so as to better analyse the sample effectively. The clinical dataset contains five levels of measures - such as area, perimeter, circularity, solidity, and roundness - related to RA. The dataset is classified into two types: non-inflamed and inflamed. The non-inflamed are denoted as Level 0 and the inflamed are further classified into low, moderate and high The following presents guidelines, representing images in a specific group, for the range of values for the critical parameters concerned.

The following presents the lower and upper range values for the moderate inflamed type:

Area (P1)	:	9036-12989
Perimeter (P2) :	471-630	
Circularity (P3)	:	0.4084- 0.5116
Solidity (P4)	:	0.8356-0.8866
Roundness (P5)	:	0.79-0.1123

The following presents the upper limit and lower limit values found in the image set corresponding to the low inflamed type:

Area (P1)	:	8639-8970 & 13010 - 13757
Perimeter (P2) :	453-469 & 631 -652	
Circularity (P3)	:	0.4065-0.4133 & 0.5122-0.5290
Solidity (P4)	:	0.8259-0.8352 & 0.8865 – 0.8906
Roundness (P5)	:	0.0748-0.0778 & 0.1084 – 0.1257

The following presents the lower and upper range values for the moderate inflamed type:

Area (P1)	:	7583 – 8599 & 13861 - 14957
Perimeter (P2)	:	398-451 & 652 -676

Circularity (P3)	:	0.4054-0.4148 & 0.5301- 0.6013
Solidity (P4)	:	0.8107 – 0.8245 & 0.8912 – 0.9040
Roundness (P5)	:	0.0670 – 0788 & 0.1148 - 0.1420

The following presents the lower and upper range values for the high-level inflamed type:

Area (P1)	:	6581- 7491 & 15106 - 26090
Perimeter (P2)	:	359-395 & 677- 894
Circularity (P3)	:	0.3955-0.4165 & 0.6030 – 0.6413
Solidity (P4)	:	0.6535 - 0.8088 & 0.9052-0.9167
Roundness (P5)	:	0.0585 - 0.0654 & 0.1269 – 0.2012

Analysis of datasets

The goal for this case study is to identify similarities, homogeneity and relationships between parameters such as area, perimeter, circularity, solidity, and roundness. Analysis Of Variance (ANOVA) tools have been used for performing a statistical analysis. Table 3 contains the medical visits of the patients and normalized values shown in Table 4.

Table 3 Parameters and Values for CASE I

Image	P1	P2	P3	P4	P5	Level
1	6971	373	.6293	.9118	.0617	High
2	2490	868	.4152	.6847	.2009	High
3	1525	682	.4120	.807	.1253	High
4	1578	695	.4105	.8002	.1269	High
5	1266	622	.4111	.8395	.1071	Level 0
6	1275	624	.4111	.8385	.1075	Level 0

Table 4 Normalized Values for CASE I

Image	P1	P2	P3	P4	P5	Level
1	4	4	3	1	2	High
2	184	170	46	51	58	High
3	87	112	47	24	27	High
4	94	113	47	26	28	High
5	61	88	46	19	20	Level 0
6	62	89	46	18	20	Level 0

NULL HYPOTHESIS (H_0) - There is no difference between the rheumatoid arthritis clinical status for sample values with respect to the five parameters (area, perimeter, circularity, solidity, and roundness) for every visit.
ALTERNATE HYPOTHESIS (H_1) - There is a difference between clinical values on these parameters with every visit. Both H_0 and H_1 are validated for all three different cases.

The following shows the application of ANOVA with the mapped dataset.

$$Q = \sum\sum xij^2 - T^2 / N \qquad (8)$$

$\sum\sum xij^2$= 63638 and T2 / N= 1.2

By applying the equation (8), Q is calculated as 63636.8.

$$Q1= \sum (Ti2 / ni) – T2/N \qquad (9)$$

$\sum$ (Ti2 / ni) = 25437.2

By applying the equation (9), Q1 is calculated as 25436.

Q2 = Q – Q1; Q2=38200.8

Once all the necessary values are calculated by applying ANOVA, the results need to be filled in. In Table 5, the first column has the Source of Variation (SV), the second column has the Sum of Squares (SS), the third column contains the Degree of Freedom (df), the fourth column has the Mean Square (MS), and the last column has the Variance Ratio (F_0). The values for Q1, Q2 and Q are filled in the SS column. The df value for between visits is calculated by subtracting 1 from the number of visits, i.e., 6 -1. The df value for within clinical parameters is calculated by subtracting the number of rows from the total number of elements The stats table can be referred to by using the df value. Referring to the statistical table for F 5%, the corresponding F value for (5, 24) can be obtained by referring to V1 as 5 and V2 as 24. The statistical table value for F (5, 24) is 2.62. Based on the reference with the statistical table, the calculated Variance Ratio (F_0 = 3.19) is greater than the statistical table value F 5% (2.62). Hence the null hypothesis H_0, namely, that the means of the real-time values of the five attributes are homogeneous is rejected and an alternate hypothesis accepted. It shows that the status is likely to change with every visit of the patient.

Table 5 ANOVA Table for CASE I

SV	SS	Df	MS	F_0
Between the patient's visits	25436 (Q1)	5	5087.2	3.19
Within critical parameters	38200.8 (Q2)	24	1591.7	
Total	63636.8 (Q)	29		

Similar to the mapping methodology followed for Case 1, the actual values were mapped for each parameter (starting from 1 and ending at 196) found in Table 6 and Table 7.

Table 6 Parameters and Values for CASE II

Image	P1	P2	P3	P4	P5	Level
1	7583	398	.6013	.904	.067	Moderate
2	7892	410	.5897	.9001	.068	Moderate
3	7921	412	.5861	.8997	.072	Moderate
4	7998	418	.5749	.8988	.076	Moderate
5	8075	422	.5695	.8978	.070	Moderate
6	8090	423	.5679	.8976	0.07	Moderate

Table 7 Normalized Values for CASE II

Image	P1	P2	P3	P4	P5	Level
1	10	14	9	4	4	Moderate
2	14	18	11	4	5	Moderate
3	15	18	11	4	6	Moderate
4	15	20	13	4	8	Moderate
5	15	22	15	5	5	Moderate
6	16	22	15	6	6	Moderate

From the statistics F-tables, F5% (V1=5, V2=24) = 2.62. From Table 8, it is clear that F0 < F5%, hence the null hypothesis $_{H0,}$ - namely, that the means of the real-time values of the five attributes are homogeneous, is accepted. In other words, since the five attributes of rheumatoid arthritis do not differ significantly, H_0 is accepted. This analysis shows that the patient's status still remains unchanged.

A set of 390 images was taken for analysing the condition of RA. Of these images, 227 are inflammatory and 163 non-inflammatory by computational values and, by medical records, 225 are inflammatory and 165 non-inflammatory. The application of the J48 and ADTree classification algorithms from Weka tool shows that J48 is classified at 99% and ADTree at 98.9% for the computational dataset and at 99.49% for the medical dataset Similarly/Likewise, other decision parameters listed in Table 9 show that the medical record dataset and computational classification dataset are matched perfectly – by as much as 99.49%.

Table 8 ANOVA Table for CASE II

SV	SS	DF	MS	F_0
Between the patient's visits	75.47	5	15.09	0.39
Within critical parameters	914	24	38.08	
Total	989.47	29		

Table 9 Relationship between all parameters

Decision-making attributes / Algorithms	Dataset with computational values		Dataset with medical values (ESR)	
	J48	ADTree	J48	ADTree
Correctly classified	99.49	98.97	99.49	99.49
Incorrectly classified	0.51	0.53	0.51	0.51
Kappa statistic	0.99	0.98	0.99	0.99
Mean absolute error	0.01	0.02	0.01	0.01
Root_mean_squared_error	0.02	0.06	0.03	0.04
Relative_absolute_error	1.05	3.14	1.05	2.23
Root_relative_squared_error	4.57	11.21	6.48	7.45

Conclusions

This paper presents a robust segmentation scheme for automatic lymphocytic cell analysis and the probability of detecting the disease, rheumatoid arthritis. Since the results obtained are entirely dependent on the segmented lymphocytes (ROI); the segmentation of lymphocytes from the blood smear image, consequently, is vital. The new edge threshold segmentation with slider control algorithm is used to segment lymphocytes dynamically and features like area, perimeter, circularity, solidity, and roundness found. The ADTree is used to classify features and, based on this classification, decision rules are generated. Our model classifies the given image as inflamed or non-inflamed, using the decision rules generated. Statistical tools like Karl Pearson's correlation coefficient correlation technique and Analysis Of Variance (ANOVA) have been used to performing statistical analyses to discover homogeneity, similarities and relationships between the parameters involved.

Acknowledgements

My sincere thanks to Dr.V.VKausik, MRCP (UK), CCT (Rheumatology), FRCP (London) Arthritis & Rheumatism Centre, Nungambakkam, Chennai for providing the necessary information about Rheumatoid Arthritis , Microscopic and Digital X-ray images without which this research would not have been possible.

References

Aman Kumar Sharma ,SuruchiSahni,"A Comparative Study of Classification Algorithms for Spam Email Data Analysis", *International Journal of Computer Science& Engineering (IJCSE),***2011**, 3, 1890-1895.

Bharanidharan T, Ghosh DK, "A Two Dimensional Image classification Neural Network for Medical Images", *European Journal of Science and Research*, **2012**, 74, 286-291.

Chokkalingam SP, Komathy K , "Classification and Segregation of Abnormal Lymphocytes through Image Mining for Diagnosing Rheumatoid Arthritis Using Min-max Algorithm",*Research Journal of Applied Science and Engineeringand Technology*, **2014**, 7, 3926-3934.

Chokkalingam SP, Komathy K, "Performance analysis of bone images using various Edge detection algorithms and denoising Filters", *International journal of Pharma and BioScience*,**2014**, 5, 943-954.

Krishnapuram R, Keller JM,"A possibilistic approach to clustering",*IEEE Transaction on Fuzzy System,***1993**, 2, 98-110.

NiponTheera-Umpon, SompongDhompongsa. Morphological Granulometric, Features of Nucleus in Automatic Bone Marrow White Blood Cell Classification, *IEEE Transaction on InformationTechnology in Bio medicine* **2007**, 11, 353-359.

Pavlova PE, Cyrrilov KP, Moumdjiev, IN. Application of HSV colour system in the identification by colour of biological objects on the basis of microscopic images. *Computational Medical Imaging and Graphics*,**1996**, 20, 357-64.

Tabrizi PR, Rezatifighi SH, Yazdanpanah MJ. Using PCA and LVQ Neural Network for Automatic Recognition of Five Types of White Blood Cells,*International Conference of the IEEE Engineering in Medicine& Biology Society (EMBC)*,**2010**, 5593-5596.