Blood Components Separation From Microscopic Blood Smear Images Using Fuzzy Inference System

Mrinal Kanti dhar [1], Shahriar Malik[2], Ahmed Al Mashrif Rim[3], M. Jahirul Islam[4]
[1-3] Department of Electrical and Electronic Engineering
[1-3] Leading University, Sylhet, Bangladesh
[4]Department of Computer Science & Engineering
[4]Shahjalal University of Science and Technology, Sylhet, Bangladesh
Email: mrinal054@lus.ac.bd[1], shahriarnayeem8@gmail.com[2], ahmedalmashrifrim@gmail.com[3], jahir-cse@sust.edu[4]

***Abstract*—The conventional method in hematology for detecting blood components is based on human inspection. In this paper, fuzzy inference system is applied for automatic classification and detection of blood components. Histogram peaks in the green channel of the RGB image and distances between leukocyte nuclei centroid and the remaining pixels are used to calculate fuzzy sets automatically during pre-processing. In the processing stage, the intensity levels found from the histogram based on gray level of every pixel and the vicinity of the pixels nearest to the leukocyte nucleus centroid are fuzzified using pertinence functions. Fuzzy rules are then applied, and followed by the defuzzification of the image. As a result, it is categorized into four areas: blood plasma, leukocyte nuclei, erythrocytes and leukocyte cytoplasm. During post-processing, false positives are removed and false plasmas are filled accordingly. Performance of the proposed method is evaluated using confusion matrix.**

Keywords—Hematology, FIS, Histogram, Morphological operation, Confusion matrix.

I. INTRODUCTION

In hematology, segmentation of blood cell allows the similar ones to make a common process together and separate data to process separately. Blood creates antibodies to fight infectious diseases, carries oxygen and creates blood clot for wounded areas, carries nutrients to body parts and removes waste for excretion. So, the study of blood and its components do help us to detect diseases such as: Hemophilia, Sickle-cell anemia, Thalassemia, Leukemia, and Myeloma etc. [1]. Ko et al. [2] proposed a WBC segmentation algorithm used stepwise merging rules based on mean-shift clustering and boundary removal rules based on a GVF (Gradient Vector Flow) snake. Ramesh et al. [3] developed a method using the S channel of the HSV color model for the segmentation of leukocytes. Rezatofghi and Soltanian-Zadeh [4] developed a method for the segmentation of the nucleus and cytoplasm of leukocytes using a snake algorithm and Gram–Schmidt orthogonalization. The data of leukocyte cytoplasm is gained from the vicinity of leukocyte nucleus mass and its closest erythrocyte area [5]. Mohapatra [6] applied clustering for white blood cells segmentation and extracted the features like shape, color, texture, fractal, Fourier descriptors and contour. The system was trained to recognize Leukemia. Zheng [7] produced a ranked structure for localization and segmentation of leukocyte. Shitong [8] developed a technique merging the threshold segmentation, fuzzy and some mathematical morphology. The problem with this technique is that it is not separating the nucleus and cytoplasm properly. Doriniet et al. [9] proposed a scheme for the nucleus extraction. The watershed transform has been used in this scheme which is based on the image forest transform. This scheme does not work well if the cytoplasm isn't round. P. Maji et al [10] have designed a way to calculate and indicate red blood cell. They also used hole filling morphological operation. Our proposed methodology is to have an automatic system which can segment and classify Leukocytes Nuclei, Leukocytes Cytoplasm, Plasma and Erythrocytes. The image will be microscopic images of blood cell with randomly selected 100 examples.

II. METHODOLGY

In this paper, we create an automatic system to separate blood cell in components such as: leukocyte nuclei, leukocyte cytoplasm, erythrocytes and blood plasma. Our proposed method consists of 3 parts – pre-processing, processing and post-processing.

A. PRE-PROCESSING

In pre-processing, we calculate intensity and vicinity. Intensity of an image means the difference in contrast and pixel color quality. Vicinity is defined as a distance from one point to another fixed point.

Intensity can be found in 4 steps which are discussed detail in [12]. In Step 1, green channel image is extracted from the RGB image as G-channel is easier for intensity conversion and most sensitive to the human eye.

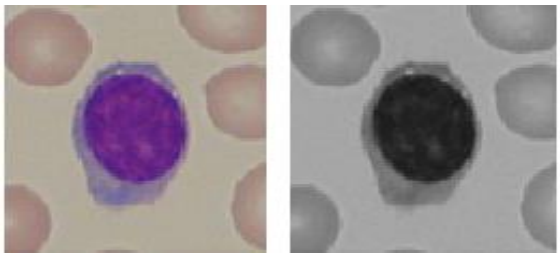

Fig.1. (Left) Original and (Right) G-channel

In step 2, we calculate three regions based on intensity level such as low, intermediate and crest. To find them, we plot histogram of the G-channel image by assigning the histogram in 256 position vectors.

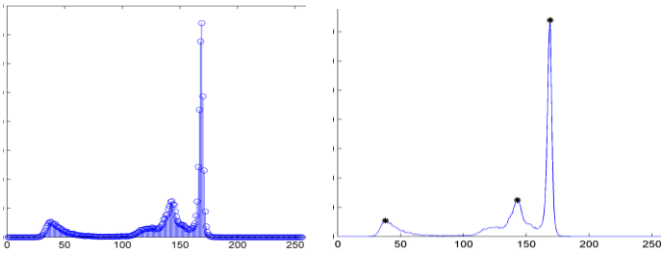

Fig. 2. (Left) G-channel histogram and (Right) three peak points

In step 3, we find the three valley points in the histogram (Fig. 2). Every position is evaluated starting from the zero position in the vector. If frequency of the present position is greater than the frequency of the next position, the value of the previous position replaces the next position. As a result, the reiteration of maximum frequency indicates that the value of the frequency in histogram curve is reducing. This will reveal the presence of a valley. These valleys will be categorized with several points.

In step 4, from 0 to Low point is replaced by 0, from Low to Crest point is presented by 126, and from Crest to 255 position is represented by 255. Pre-classified area is drawn (Fig. 3). A pre-classified area is an image which contains clearly distinguished and separate regions. Here it separates leukocyte nucleus, cytoplasm and erythrocyte. Image is eroded followed by dilation for better efficiency. Adjacent disconnected area is filled. In here, white indicates plasma, black indicates leukocyte nucleus and gray indicate both cytoplasm and erythrocyte.

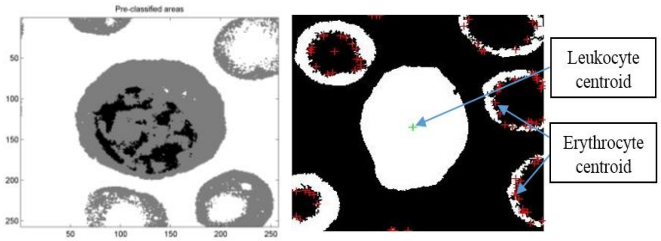

Fig. 3. (Left) Pre-classified area and (Right) leukocyte (nucleus + cytoplasm)

Vicinity is found by calculating High_Vicinity and Low_vicinity. High_Vicinity is an approximate distance between nucleus centroid and cytoplasm. Low_vicinity is an approximate distance between nucleus centroid and erythrocytes. Vicinity calculation is performed in 3 steps.

In step 1, centroid of leukocyte nucleus is calculated. A morphological structuring operation is used to connect the disconnected black regions of previous image. After filling operation in these areas, center of mass (centroid) is found by measuring properties of image region and superimposing the location of the image.

In step 2, centroid of leukocyte (nucleus + cytoplasm) and erythrocyte is calculated. By using the same method as previous step, we find the centroid of leukocyte (nucleus + cytoplasm) (Fig. 3).Centroids of leukocyte nearest to leukocyte nucleus centroids are calculated. Those with the minimum leukocyte centroids are stored and used to draw circles corresponding leukocyte centroids.

In step 3, calculating High_Vicinity and Low_Vicinity zone using major and minor axis (Fig. 4). Major and minor axis is used to find the diameter from centroid of nucleus to centroid of leukocyte (nucleus + cytoplasm) and erythrocyte. This variable is stored in matrix, YG_{kxl} .The least value of matrix YG is called High_Vicinity, specifying a leukocyte nucleus nearest to the blood plasma. Low_vicinity is defined as being ⅓ of the Euclidean distance between high-vicinity and the maximum value of matrix *YG*. Low_Vicinity is calculated from the following expression-

$$Low\_Vicinity = High\_Vicinity\ (D_{max} - High\_Vicinity)/3.$$

Where, D_{max} = Maximum distance between nucleus centroid and pixel.

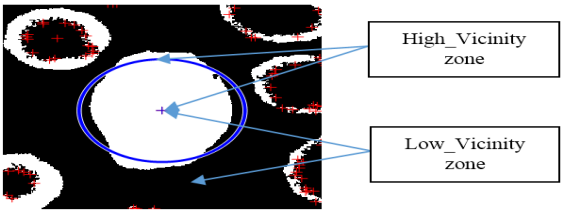

Fig. 4. High_Vicinity and Low_Vicinity

B. PROCESSING

The Fuzzy Inference System is used for the processing of image. Data obtained from the pre-processing stage are used to generate fuzzy sets using three linguistic terms, two input variables (intensity and vicinity) and one output variable (class),membership functions. It is called fuzzification. Inference is made based on four fuzzy rules where image is separated into four areas i.e. leukocyte nuclei, leukocyte cytoplasm, erythrocytes and blood plasma. Processing stage is performed in 4 steps [11].

Step 1: Set fuzzy rules.

- Rule 1: Classified region is leukocyte nucleus when intensity level is low;
- Rule 2: Classified region is plasma when intensity level is crest;
- Rule 3: Classified region is erythrocyte when intensity level is intermediate and vicinity is low;
- Rule 4: Classified region is cytoplasm when intensity level is intermediate and vicinity is high.

Step 2: Define two input membership functions (mFs) – one for intensity and another for vicinity.

For intensity mF, there are three types of variables: low, intermediate and crest. For vicinity mF, there are two types of variables: High_Vicinity and Low_Vicinity. Both input mFs type are trapezoidal (Fig 5).

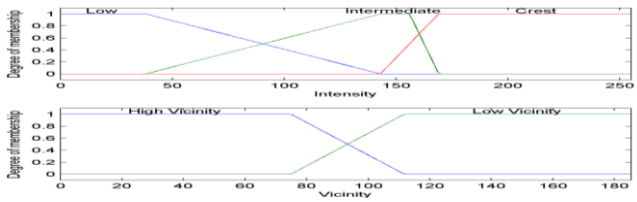

Fig. 5. Input membership function

Step 3: Define four output membership function.

The output mFs are – nucleus, cytoplasm, erythrocyte and plasma. The output mFs are of triangular shape. (Fig.6).

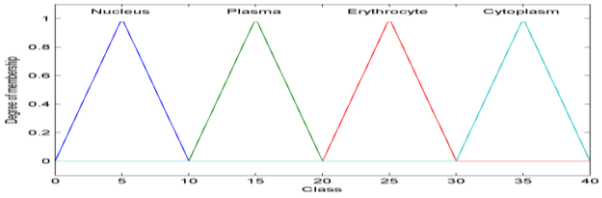

Fig. 6. Output membership function

Step 4: Defuzzify the image.

Here implication method is *minimum* and aggregation is *maximum*. The mean of maximum (MoM) method is used for defuzzification of the result. After defuzzification, image is classified into blood plasma, leukocyte nuclei, erythrocytes and leukocyte cytoplasm. Color mapping is used to draw classified image (Fig 7).

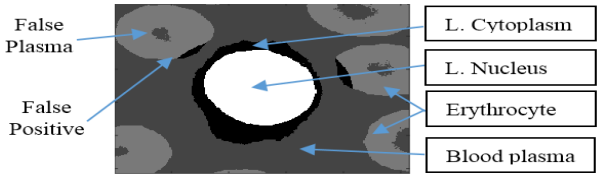

Fig. 7. Classified image

C. *POST-PROCESSING*

From (Fig 7), it is seen that some areas are falsely detected as blood plasma or leukocyte cytoplasm. Post-processing involves removing falsely detected blood plasma and falsely detected leukocyte cytoplasm. First, we remove false positive areas from the erythrocyte. Observation shows that only a fraction of leukocyte cytoplasm may exists in the erythrocyte region. For this, first we take the output of class 4 found from the FIS. Here, class 4 represents cytoplasm zone. As we see from (Fig. 8), there exists some small black areas which indeed are parts of leukocyte cytoplasm. By performing a sequence of morphological operations (imfill, bwareaopen, labelling etc.), we replace them by erythocyte color as shown in (Fig. 8).

Next, to remove falsely detected blood plasma, at first, we make all region white except for erythrocyte where only erythrocyte remains black to highlight false plasma (Fig. 10). Then, we make all areas black except for false plasma which remains white (Fig. 9).

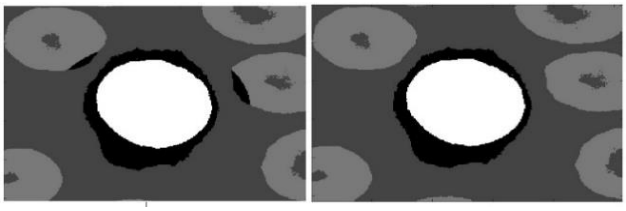

Fig. 8. (Left) before and (Right) after removing false

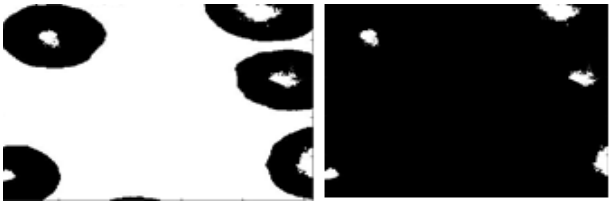

Fig. 9. (Left) Indicating erythrocyte and (Right) Indicating false plasma

Finally, false plasma is replaced by erythrocyte color. The image is then superimposed with image found from (Fig. 9) to get the final image which is without any false positive or false area regions (Fig. 10).

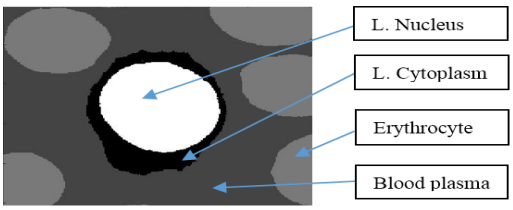

Fig. 10. Final image

III. RESULT & DISCUSSION

To evaluate the system, we collect 100 images from different sources. As discussed in pre-processing stage, the histogram should show three peaks, in some pictures we find more than three peaks, which are resolved into three convenient peaks by tracking and comparing them once from left to right of the histogram and then performing same operation in the opposite direction shown in (Fig. 11).

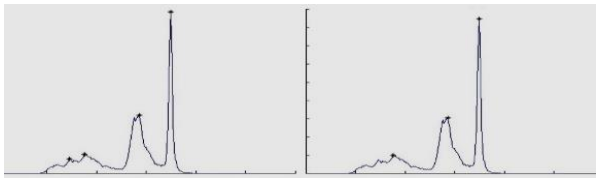

Fig. 11. (Left) 4 peak points and (Right) 3 peak points

The percentages of blood components extracted from a randomly chosen smear image are shown in Table I.

High true positive (TP) and true negative (TN) rates represent the success in detecting the presence or absence of

each component in the samples. False positive (FP) and false negative (FN) rates indicate the errors in identifying of a component in the samples. Results obtained from the automatic segmentation results of TP, TN, FP, FN and accuracy rates (AC) are shown in Table II.

TABLE I. BLOOD PERCENTAGE

Class	% amount
L. nucleus	12.8889
L. cytoplasm	7.5944
Erythrocyte	29.8294
Plasma	49.6874

TABLE II. SEGMENTATION RESULT

Category	TP (%)	TN (%)	FP (%)	FN (%)	AC (%)
L. Nucleus	90.9	90.91	5.45	3.63	95.2
L. Cytoplasm	83.6	83.64	16.36	1.82	90.2
Erythrocyte	83.6	89.09	1.81	14.54	91.4
Plasma	92.7	92.73	5.45	1.82	96.2

In the leukocyte nucleus, FP rate comes from the correlation in staining of the leukocyte nucleus and leukocyte cytoplasm in some samples. The inability to determine he corners of leukocyte nucleus also raises FN rate. In the leukocyte cytoplasm, FP occurs when the erythrocyte near cytoplasm is falsely identified as cytoplasm. FN error comes from the staining process. In the erythrocyte, if a filling operation takes place for removing the false positive area then it is FN. FP occurs when blood plasma is falsely identified in the center of erythrocyte. In case of blood plasma, if it is identified in the erythrocyte as false area then it is FN. Accuracy rate is calculated using the following equation:

$$AC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

In order to validate the proposed segmentation method, we compared the results with few existing segmentation methods present in the literature. Table III demonstrates the effectiveness of the proposed approach.

TABLE III. EFFECTIVENESS OF THE ALGORITHMS

Leukocyte segmentation method	Accuracy (%)
Ko et al. [2]	88.75 (L. Nucleus) 69.05 (L. Cytoplasm)
Ramesh et al. [3]	93.08
Rezatofghi et al. [4]	93.09
Zheng et al. [7]	91.94
Proposed Method	**93.25**

Finally, we measure execution time in computer have Core i5, 3.3GHz processor and 8GB RAM. MATLAB 2017a is used for simulation. We run an image 5 times and then average them before tabulating. Table IV shows that almost 90% time is spent during the processing.

TABLE IV. EXECUTION TIME

Avg. execution time	18.224 sec
% Pre-processing time	≈08
% Processing time	≈90
% Post-processing time	≈02

IV. CONCLUSION

This research involves detection and separation of blood cell components from microscopic blood sample images. Automatic processing is achieved by filtering the microscopic image via color, cell sizes and shapes. Fuzzy inference system is used for classification, while different morphological operation based techniques are used for segmentation. Experiments show the system is efficient and reliable, while faster in operation as well. Information extracted from microscopic images of blood samples can benefit us by predicting, solving and treating blood related diseases of a patient. The system is analyzed using a comparatively small dataset. In future, we would try to analyze it taking a large dataset with distorted images.

REFERENCES

[1] Markiewicz T, Moszczynski. "Analysis of Features for Blood Cell Recognition", VI International Workshop Computational Problems of Electrical Engineering, Poland (2004).

[2] Ko, B.C.; Gim, J.W.; Nam, J.Y. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. *Micron* 2011, *42*, 695–705.

[3] Ramesh N, Dangott B, Salama ME, Tasdizen T. Isolation and two-step classifcation of normal white blood cells in peripheral blood smears. Journal of Pathology Informatics. 2012; 3:13. PMid:22530181 PMCid:PMC3327044. http:// dx.doi.org/10.4103/2153-3539.93895.

[4] Rezatofghi SH, Soltanian-Zadeh H. Automatic recognition of fve types of white blood cells in peripheral blood. Computerized Medical Imaging and Graphics. 2011; 35:333-43. PMid: 21300521. http://dx.doi.org/10.1016/j. compmedimag.2011.01.003.

[5] Hamghalam M, Ayatollahi A. "Automatic counting of leukocytes in Giemsa-stained images of peripheral blood smear", ICDIP 2009: Proceedings of the International Conference on Digital Image Processing (2009 Mar 7-9); Bangkok, TH. 2009. p. 13-6.

[6] Mohapatra S, Patra D, Satpathy S. "An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images", J NeuralComput Appl (2014); 24:1887–904.

[7] Zheng X, Wang Y, Wang G, Chen Z. "A novel algorithm based on visual saliency attention for localization andsegmentation in rapidly-stained leukocyte images", Micron (2014); 56:17-28. PMid: 24148877.

[8] W. Shitong and W. Min. "A new detection algorithm based on fuzzy cellular neural networks for white blood cell detection", IEEE Transactions on Information Technology in Biomedicine, 10(1):5 –10, (January 2006).

[9] L.B. Dorini, R. Minetto, and N.J. Leite. "White blood cell segmentation using morphological operators and scale-space analysis", In Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing, pages 294–304, (October 2007).

[10] P. Maji, A. Mandal, M. Ganguly, and S. Saha, "An automated method for counting and characterizing red blood cells using mathematical morphology," in 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), 2015, pp. 1–6.

[11] Vale, Alessandra Mendes Pacheco Guerra et al. "Automatic segmentation and classification of blood components in microscopic images using a fuzzy approach", Rev. Bras. Eng. Bioméd. [online] (2014), vol.30, n.4, pp.341-354. ISSN 1517-3151. http://dx.doi.org/10.1590/1517-3151.0626.