

Optimized Segmentation of Microscopic Images by Otsu' Method to Find White Blood Cells

Rangu.Srikanth¹, Dr.Kalagadda Bikshalu²

¹Kakatiya Institute of Technology and Science, Warangal, 506013, INDIA

²Kakatiya University, Warangal, 506009, INDIA

¹rangu.srikanth@gmail.com, ² kalagaddaashu@gmail.com

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract— In the medical field, microscopic image-based investigations are broadly used to examine cell morphology and for disease diagnosis. In this article white blood cells (WBC) are identified from microscopic images by image segmentation. One of the important segmentation methods is multilevel thresholding of an image, in this process pixels are grouped into different classes depends on thresholding levels. The selection of threshold levels affects the efficiency of segmentation. Otsu's method is a significant and important segmentation technique based on the multi-threshold of the histogram. Optimization techniques can be used to compute optimized threshold levels, Harmony Search Algorithm (HSA) is used for this purpose. From the color, microscopic image red, green, and blue components are extracted and segmentation to find white blood cells, and from the same images gray, hue, saturation, and intensity components also extracted pathological features. After segmentation Morphological opening and closing are applied for an efficient finding of white cells, from results the green component of microscopic images is giving efficient results. Results are described by using standard deviation (STDR), mean square error (MSE), Mean of fitness of algorithm (MOFIT), time for execution, and FITNESS function.

Keywords—Microscopic images, Image Segmentation, Otsu's technique, Optimization techniques, Harmonic Search Algorithm.

1. Introduction

In this paper, a method of detecting the WBC from microscopic images is projected as Otsu's method for multi-level threshold (th_i) and HSA for computing optimized threshold level on the histogram. From color images the HIS (Hue Saturation Intensity) color space can be extracted, and these are used given to clustering algorithm. For examinee overall condition of patients, physicians will use The Complete Blood Count (CBC). Identification of intensity of WBC is important in recognizing blood diseases early. Analysis of CBC gives a forecast of quantifying the Red Blood Cells (RBC), WBC, and platelet count, an automatic image processing technique essential to find the above-described targets.

Blood cell is required for the discovery and classification of sickle cell anemia [1]. From the microscopic image, segmentation WBC can be discovered, and detecting the occurrence of leukemia in a person's blood [2] as leukemia's major symptom is the huge size of nuclei of With Blood cells [3]. An image processing scheme is required to finding and categorization of hematologic diseases. The image segmentation consequences to extract desired medical details to categorize the examined blood sample into abnormal cells or normal. In literature, numerous segmentations developed to segment microscopic images to identify white cells.

Multilevel thresholding [4]classify the gray values of (pixel values)of the image into different classes, background and various objects can be identified with the triangle method or Zack algorithm [5].Diverse feature extraction methods and classifiers are used to get pathological information from microscopic images. Supported Vector Machine (SVM) [6] is utilized to categorize the WBC, Fractal dimension and shape features with SVM in [7], for segmentation Artificial Neural Network (ANN) is applied in [8] to sortout the leukemia cells, Geometric features also used for categorization of White Blood Cells in [9].

The minimum filter along with the Arithmetic operation is collectively used to the segment of the cell nucleus in [10]. Clustering can be used in [11] to segment white blood cells K- means. In this paper, Otsu's method used for multilevel thresholding(MT) with HSA for finding optimized threshold levels, and then morphological opening and closing were used to target WBC.

2. Otus method for segmentation

The MT is the method, the pixels are partitioning into classes depends on their pixels values levels (p) based on threshold level (th), like given below:

$$C1 \leftarrow p \quad \text{if} \quad 0 \leq p < th,$$

$$C2 \leftarrow \text{if } th \leq p < L - 1$$

For microscopic image pixel levels $p = 0$ to $L - 1$, where $L - 1$ is the maximum pixel value in the image, generally for gray-images it is 256. For bi-level thresholding $C1$ and $C2$ are the two classes in, and th is the threshold value, category of each pixel follows given in Equations (1), the threshold must satisfy the condition as given in Equation (8), for multiple sets given as below:

$$\begin{aligned} C1 &\leftarrow p \quad \text{if } 0 \leq p < th1 \\ C1 &\leftarrow p \quad \text{if } th1 \leq p < th2 \\ C_i &\leftarrow p \quad \text{if } th_i \leq p < th_{i+1} \\ C_n &\leftarrow p \quad \text{if } th_n \leq p < th_{n+1} \end{aligned} \tag{1}$$

Where $\{th1$ to $th_k\}$ are different threshold levels, computation of optimized threshold levels is a key agenda of this research paper. Along with Otsu’s method [13], HAS used to find optimized th values, the HAS algorithm proposed an objective function (given in Equation.8) it must be maximized inter-class variance for better results. Taking into account the maximum pixel levels $L-1$ from gray scale images or from red, green, and blue components from a color image, the probability distribution of the pixels is computed like given below:

$$Ph_c^i = \frac{h_c^i}{NP}, \sum_{i=1}^{NP} Ph_c^i = 1 \tag{2}$$

$$c = \begin{cases} 1,2,3 & \text{for RGB image} \\ 1, & \text{for Gray scale image} \end{cases}$$

From the above expressions, pixel intensity levels of an image are represented as $i (0 \leq i \leq L - 1)$, NP is the total pixel count of image i . Frequency gray values in an image of the histogram are represented as h_c^i , where the histogram is normalized within a probability distribution Ph_c^i , for bi-level thresholding

$$C1 = \frac{Ph_1^c}{w_0^c(th)}, \dots, \frac{Ph_{th}^c}{w_0^c(th)}, \quad C2 = \frac{Ph_{th+1}^c}{w_1^c(th)}, \dots, \frac{Ph_L^c}{w_1^c(th)} \tag{3}$$

Whereas $w_0(th)$ and $w_1(th)$ are probability distributions for two classes or regions $C1$ and $C2$, as it is shown below as

$$w_0^c(th) = \sum_{i=1}^{th} Ph_i^c, \quad w_1^c(th) = \sum_{i=th+1}^{th} Ph_i^c \tag{4}$$

It needs to find mean levels μ_0^c and μ_1^c that describe the classes using Equation (5). The variance between classes σ^{2c} is computed using Equation (6) as given below:

$$\mu_0^c = \sum_{i=1}^{th} \frac{iPh_i^c}{w_0^c(th)}, \quad \mu_1^c = \sum_{i=th+1}^L \frac{iPh_i^c}{w_1^c(th)} \tag{5}$$

$$\sigma^{2c} = \sigma_1^c + \sigma_2^c \tag{6}$$

From Equation (6), the left-hand term is Otsu’s variance operator. The variances of two regions or classes $C1$ and $C2$ given as σ_1^c and σ_2^c , articulated as

$$\sigma_1^c = w_0^c(\mu_0^c + \mu_T^c)^2, \quad \sigma_2^c = w_1^c(\mu_1^c + \mu_T^c)^2 \tag{7}$$

Where the $\mu_T^c = w_0^c\mu_0^c + w_1^c\mu_1^c$ and $w_0^c + w_1^c = 1$. Based on the values σ_1^c and σ_2^c , Equation (8) represents the fitness function or objective function:

$$J(th) = \max(\sigma^{2c}(th)), 0 \leq th \leq L - 1$$

where $\sigma^{2c}(th)$ is the variance for a specified th value. Here the optimization algorithm used to compute the threshold (th) that maximizes Equation (8). In the case of color images, apply a similar process for each it is needed to apply for each component of color images. For multi-level thresholding require $k-1$ number of thresholds on the histogram to separate an image into k regions or classes using Equation(2); subsequently, computation of k variances is required. The fitness or objective function $J(th)$ in Equation (8) modified as given in Eq.(9) for multiple thresholds as given below:

$$J(TH) = \max(\sigma^{2c}(TH)), 0 \leq th_i \leq L - 1, \quad \text{wherte} \quad i = 1,2 \dots k \tag{9}$$

Here the vector $\mathbf{TH} = [th_1, th_2 \dots th_{k-1}]$ is with multiple Threshold levels and the corresponding variances are estimated as below

$$\sigma^{2c} = \sum_{i=1}^k \sigma_i^c = \sum_{i=1}^k w_i^c(\mu_i^c - \mu_T^c)^2 \tag{10}$$

Here, w_i^c indicates the t_{th} class, w_i^c is histogram or probability and μ_j^c mean of these expressions given below

$$\begin{aligned}
 w_0^c(th) &= \sum_{i=1}^{th_1} Ph_i^c \\
 w_1^c(th) &= \sum_{i=th_1+1}^{th_1} Ph_i^c \\
 &\vdots \\
 w_{k-1}^c(th) &= \sum_{i=th_{k+1}}^{th_1} Ph_i^c
 \end{aligned}
 \tag{11}$$

The mean values computed as

$$\begin{aligned}
 \mu_0^c &= \sum_{i=1}^{th_1} \frac{iPh_i^c}{w_0^c(th_1)} \\
 \mu_1^c &= \sum_{i=th_1+1}^{th_2} \frac{iPh_i^c}{w_0^c(th_2)} \\
 &\vdots \\
 \mu_{k-1}^c &= \sum_{i=th_{k+1}}^L \frac{iPh_i^c}{w_1^c(th_k)}
 \end{aligned}
 \tag{12}$$

3. Harmonic search algorithm (hsa)

A. Introduction to HSA

In the HAS, each probable or possible set of solutions is named as a “harmony” and it n-dimensional vector with real values, at beginning randomly selected population taken as of harmony vectors. A new candidate (or next-generation/ next iteration) harmony by random reinitialization or by adjusting the pitch by considering elements in HM. Then, the HM is modified or updated by compare newly computed candidate harmony. The week or worst harmony vector updated by a newly generated candidate vector which gives better solutions, this procedures run several iterations until the termination criterion is satisfied(no change in object function) The mail phases of HSA are given as

- HM initialization
 - improvisation of New Harmony vectors
 - updating the HM
- The subsequent sections describe each step

B. Problem Definition and the Algorithm Parameters

Usually, any optimization algorithm can be concise below:

$$\begin{aligned}
 &\text{minimize or maximization of function } f(x), \text{ where } x = (x(1), x(2), \dots \dots x(n)) \in R^n \text{ and :} \\
 &x(j) \in [l(j), u(j)] \text{ where } j = 1 \text{ to } n
 \end{aligned}
 \tag{13}$$

Whereas $f(x)$ is a fitness function or objective function, $x = x(i), i = 1$ to $n, x(i)$ is design variable, lower and upper limits of $x(j)$ are $l(j)$ & $u(j)$ respectively. The required various parameters or variables for HSA (i) HM (ii) the harmony-memory consideration rate (HMCR) (iii) PAR (iv) BW and (v) NI -number of iterations. $x_i(j) = l(j) + (u(j) - l(j)) * rand(0,1)$ for $j = 1,2,3 \dots n$ and $i = 1,2,3 \dots HMS$

C. Harmony Memory Initialization

In this phase, an initial component in HM, that is, HMS vectors are configured. Let $x_i = \{x_i(1), x_i(2) \dots x_i(n)\}$ epitomize the i_{th} randomly-computed harmony vector: $x_i(j) = l(j) + (u(j) - l(j)) * rand(0,1)$ for $j = 1,2,3 \dots n$ and $i = 1,2,3 \dots HMS$, the value $rand(0,1)$ is given as a uniform random, (j) is lower and $u(j)$ limits of threshold levels (search space). Then, the HM matrix is given below

$$HM = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{HMS} \end{bmatrix}$$

(14)

D.Updating of New harmonic Vector

For updating the tern x_{new} .New Harmony vector, three operations required, firstly memory consideration, secondly random reinitialization, and finally pitch adjustment. By the process of improvisation,the New Harmony can be generated to the previous harmony. In the memory consideration phase, the term, first variable $x_{new}(1)$ for the new vector is obtained randomly from any of the values which are already existing in the current HM, it means from the set of $\{x_1(1), x_2(1), . . . , x_{HMS}(1)\}$. This process accomplished by selecting a homogeneous random number $r_1[0, 1]$ and if it is less than HMCR, the decision variable $x_{new}(1)$ is produced during the process of memory considerations; otherwise, the term $x_{new}(1)$ is taken from a random reinitialization between the search space limits $[l(1), u(1)]$, similarly, the other variables are $x_{new}(2), x_{new}(3), . . . , x_{new}(n)$ are computed. The two operations (i) memory consideration, and (ii) random reinitialization, given as follows:

$$x_{new}(j) \text{ is given } \begin{cases} x_i(j) \in \{x_1(j) \dots \dots \dots x_{HMS}(j)\} \\ l(j) + (u(j) - l(j)).rand(0,1) \\ \text{with probability of } 1 - HMCR \end{cases} \quad (15)$$

Each new component computes and tested for whether it needs pitch adjustment or not. For this process, the PAR frequency updating, and the BW for find new value in solutions space of the HM. Then, the pitch-adjusting is required to new harmonies by modifying the original one. Given new value as $x_{new}(j)$ can be computed by $x_{new}(j) \pm rand(0,1) * BW$ with probability PAR and value will not change with probability $1 - PAR$. In the pitch adjusting phase, new potential (or budding) harmonies generation by faintly adjusting original positions. Thus, the decision variable is either concerned by a random number between 0 and BW or left unaltered. To pitch adjusting course, the points must be reassigned which are outside the range $[l, u]$.

At the final stage, the term x_{new} , a New Harmony vector is by updating HM with which win the survival of the fitness war or competition among x_{new} and the x_w in the HM, consequently x_{new} will replace x_w , where x_w is worst harmony vector. In this article, HSA is used for maximizing inter-class variance [15].

4. Multilevel thresholding by has

A.HSA Implementation

In this article segmentation by multi-level thresholding is considered by using the fitness function (objective functions) of Otsu’s algorithm as given in Eq. (9). The HSA is united with the Otsu functions, producing diverse segmentation algorithms [14].The HSA is powerfully affected by assigned parameters (i) HM (ii) HMCR (iii) PAR (iii) BW, and the (iv) NI, by using HAS we will find the best results(threshold values to satisfy Eq.(9).The universal procedure is to find the best parameters, initially fix parameters randomly within the limits, and then apply HAS is executed, assign new values of parameters to HSA if results are not satisfactory, and implement algorithm again, to find best parameters many trials required.

B. Selection of parameter values

The HM value is taken as 30, HMCR is 0.5, PAR 0.2, BW is 0.1, NI is taken as 3500. The number of threshold values taken as 4 to 6.

C.Opening and Closing

The Morphological operation is a significant image processing step to extract objectives from the binary image. There are two important operations in the morphological process (i) Opening (ii) Closing, these two operations are based on erosion and dilation operations [16]. The symbol of opening \circ and \bullet is the symbol of closing.

The morphological opening operation on a binary image is denoted by erosion of I by B, followed by the process of dilation of the eroded image with B. Mathematically it is given in Eq. (16)

$$[I \circ B] = (I \ominus B) \oplus B \quad (16)$$

Where the symbol \ominus is erosion and \oplus is a symbol of dilation, the structuring element I is taken a 10x10 matrix with all elements of 1(ones), and B represents a binary image.

The process of closing operation on a binary image is defined as dilation of I by B, followed by erosion of dilated image with B. Its expression is given as:

$$[I \bullet B] = (I \oplus B) \ominus B \quad (17)$$

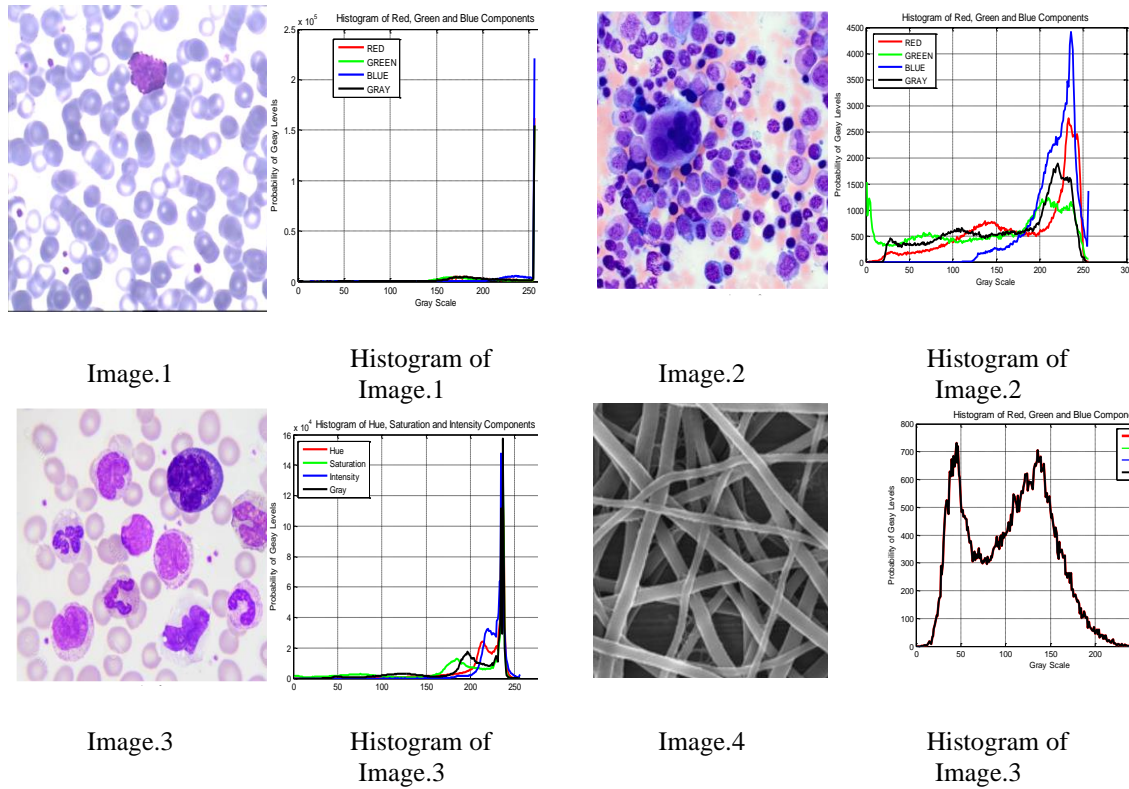


Fig.1: The images in the first Column are input images for experimentation, Second Column indicates histograms of Red(R), Green(G), Blue (B), and gray level components of images from 1 to 4.

5. Experimental results

For experimentation of our method for finding WBC implementing our algorithm three images (Images related to getting information of white blood cells from).

The MATLAB programs written for this research article executed on the processor Intel CORE, model i5-8250u, and 8th generation, with clock speed 1.60GHz and with 8GB RAM.

After getting the segmented images to apply the opening and closing forgetting interested area only. The standard deviation (STDR) is used for estimation of stability and consistency, it is representation of how the data is dispersed, and the algorithm becomes more if this value increases [13]. The resemblance between segmented image and the original image is measured by mean square error (MSE), the Mean of fitness of algorithm (MOFIT) is a parameter of the optimization algorithm, time (seconds) taken for each experiment is taken for comparison, and FITNESS is a parameter of final fitness of algorithm.

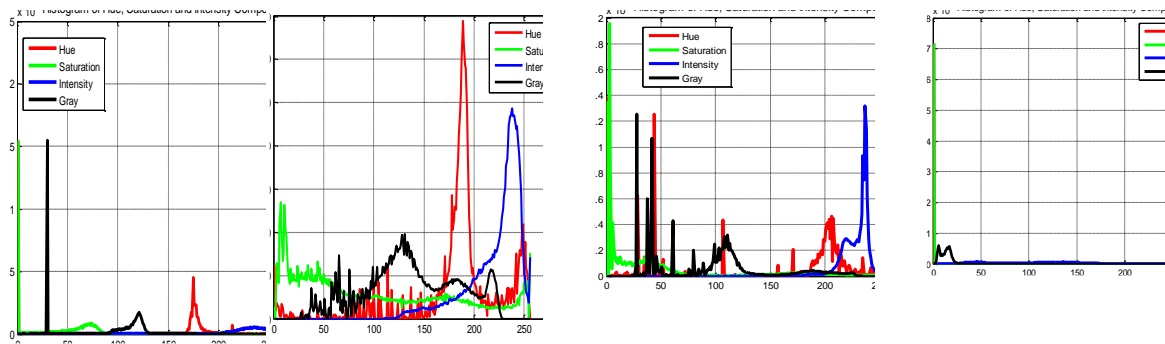


Fig.2: Histogram of Hue, saturation and Intensity Components of Images from Image.1 to Image.4

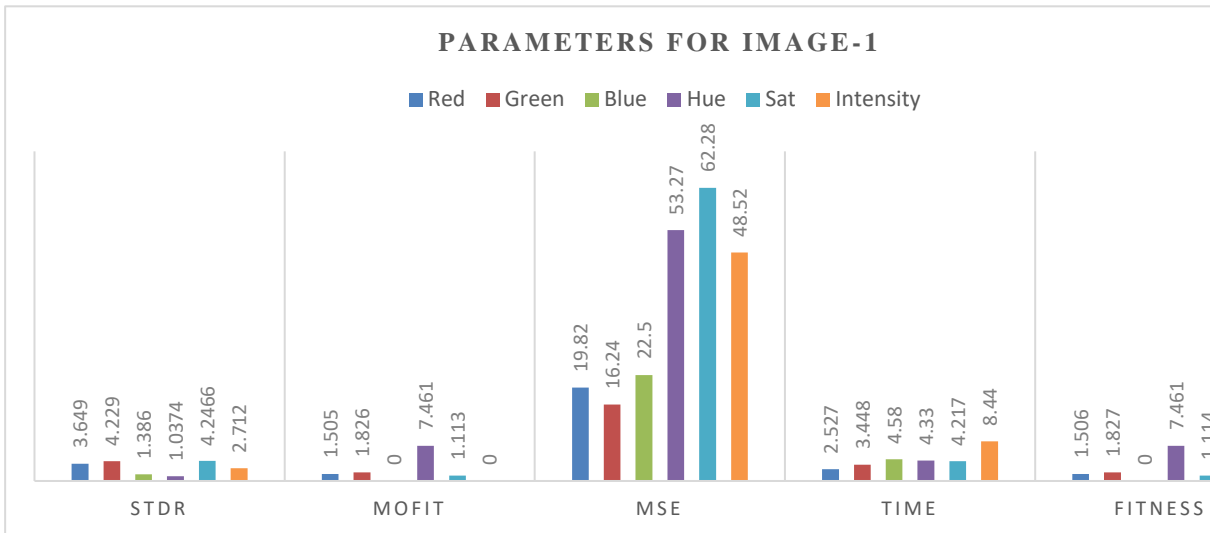


Fig.3: Compare the parameters STDR, MOFIT, MSE, TIME and FITNESS values of images taken by considering Red, Green, Blue, Hue, Saturation and Intensity components of Image.1 (MOFIT and FITNESS values are normalized by dividing with 1000)

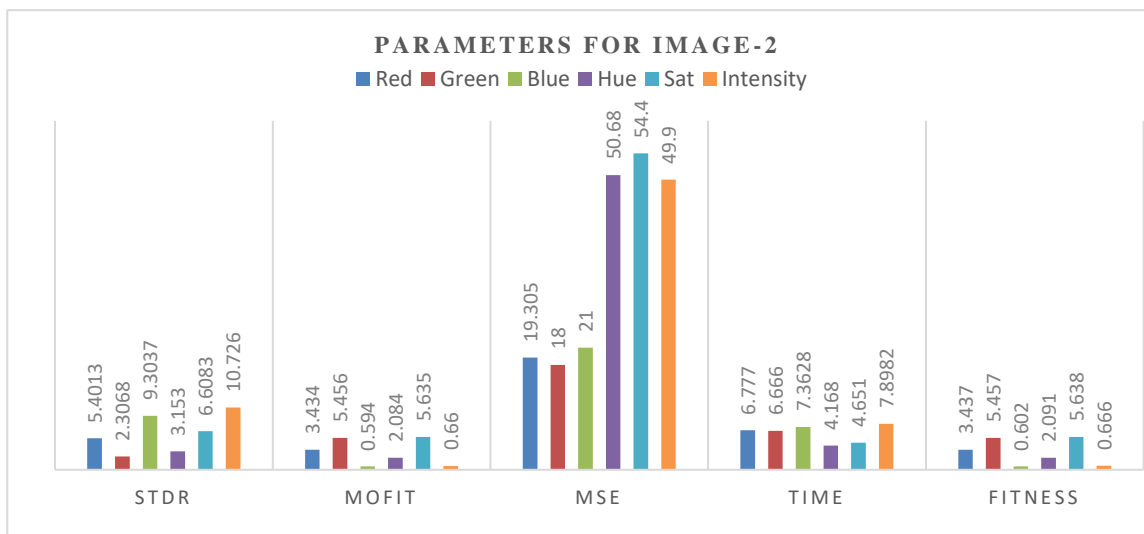


Fig.4: Compare the parameters STDR, MOFIT, MSE, TIME and FITNESS values of images taken by considering Red, Green, Blue, Hue, Saturation and Intensity components of Image.2 (MOFIT and FITNESS values are normalized by dividing with 1000)

From Fig.1 it very clear that the histogram of only the green component is spared over the total grayscale range, it is essentially required to segment the image into various regions. The purple color region indicating in the image.1 to image.3 is related to WBC, we can find the shape and type of WBC from the images. The image.4 is an image with a carbon nanotube, the histograms of R, G, B, and gray levels are overlapped, clearly shown in Fig.1. By segmenting the images taken from Green components are giving good results for getting details of white blood cells. Segmented results are compared among R G and B components of each image. The images with the green component from each image can give more details as compared to images of Hue, Saturation, and Intensity components, results are given in the next section. From the above figure, segmenting the image by multilevel thresholding taken from Hue, Saturation, and Intensity components will not give good results, and not possible to partition an image into various regions.

From figures Fig.3 and Fig.4, it is very clear that mean square error (MSE) between the segmented image and input image is less for the images which are taken green component of image for segmentation. By using other than green components we cannot get white blood cell information from segmentation images by multi thresholding method. By considering the green component only will give a clear threshold level to segment the

image to extract the required information from the color image. After applying Otsu’s method of segmentation with HAS to find thresholding levels, for Image.1, the threshold levels for an image with Red Component are 117, 175, 203, 236, for Green component 123, 170, 198, 234, for Blue component 87,169, 224, 244, for Hue component 59, 146, 180, 196, for Saturation component 17, 44, 65, 105 and for Intensity component 93, 174, 225, 245. Similarly for Image.2, the threshold levels for image with Red Component are 78, 127, 170, 213, for Green component 47, 105,161, 207, for Blue component 175, 202, 219, 231, for Hue component 57, 134, 179, 215, for Saturation component 38, 85, 141, 203 and for Intensity component 168, 191, 223, 240.

The figures Fig.5 and Fig.6 illustrate the possibility of extracting the required information of WBC from only images with green component only, it is proved in both cases of Image.1 and Image.2, for images Red component can also give good results. From the above quantitative and qualitative comparisons of various components of images, only segmentation of the Green component is considered in this paper.

From Fig.7 to Fig.9 shown segmented images, then thresholding the applied, generally, the threshold is the second-lowest threshold level calculated by Otsu’s method, if the pixel is below the threshold, indicated with the white area(pixels), that is nothing but WBC area, and above the threshold, pixels are indicated as a black area. In Fig.7 the white large area indicates white cell area and small dots are platelets in the blood sample.

For fixing the final threshold values by Harmony Search optimization algorithm for Otsu’s multi-threshold method, the variance should be maximum, variance computation and for each iteration is also illustrated in above images. The data image data set is taken from Leukocyte Images for Segmentation and Classification [17].

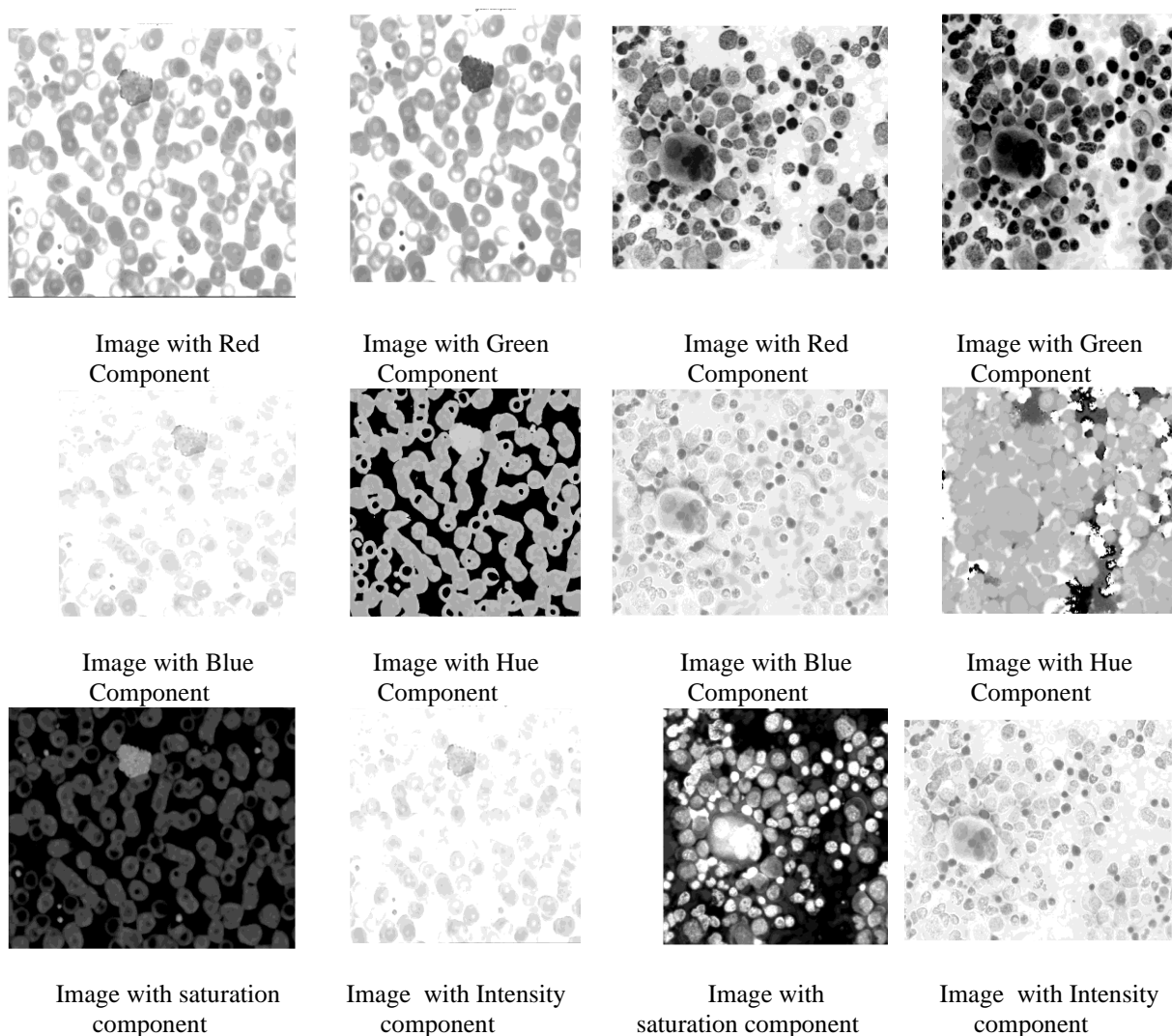


Fig.5: Images are taken by different Components of Image.1

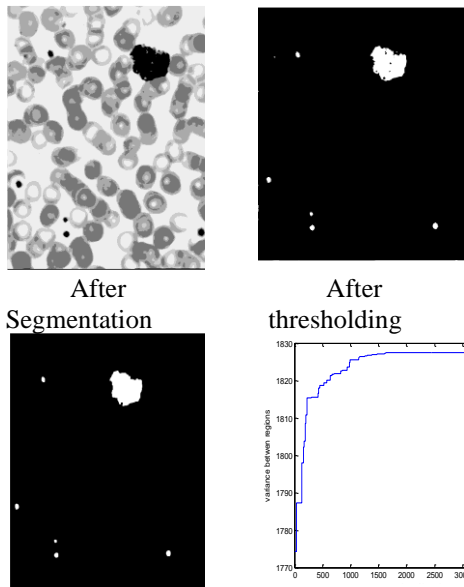
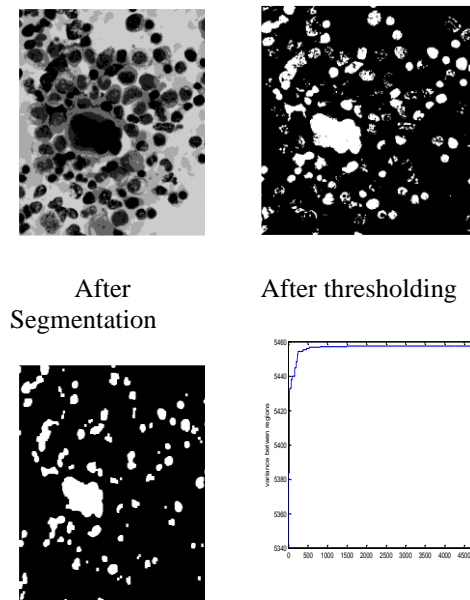
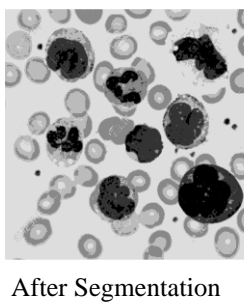


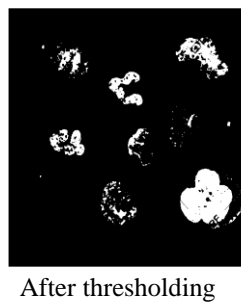
Fig.6: Images are taken by different Components of Image.2



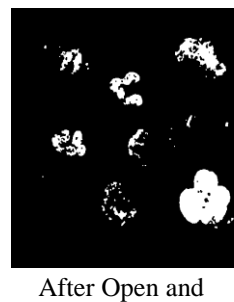
After Open and Close
 Fig.7: Segmentation Results of Image.1 by considering the green component of the Image.1



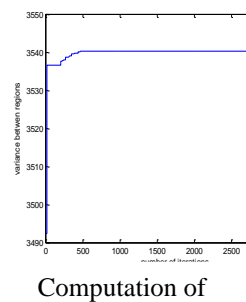
Computation of Variance



After Open and Close
 Fig.8: Segmentation Results of Image.2 by considering the green component of the Image.2



Computation of Variance



After Segmentation

After thresholding

After Open and Close

Computation of Variance

Fig.9: Segmentation Results of Image.4 by considering the green component of the Image.3

6. Conclusion

Identification of intensity of WBC is important in recognizing blood diseases early, to find, the huge size of nuclei of With Blood cells is important in predicting leukemia. In this article a method of identifying WBC is devised using multi-level segmentation with Otsu’s method along with HSA to find optimized threshold levels finally morphological operations, opening and closing applied. Images are split into Red, Green, Blue, Hue, saturation, and Intensity components, the algorithm has experimented on each component, and finally, the image with the green component is giving better results to the segment of microscopic images for WBC.

References

1. M. Veluchamy, K. Perumal, and T. Ponuchamy, “Feature extraction and classification of blood cells using artificial neural network,” American Journal of Applied Science, vol. 9, no. 5, pp. 615–619,2012
2. F. Kasimn, A. S. Prabuwno, and A. Abdullah, “Detection of leukemia in human blood sample based on microscopic images: a study,” Journal of Theoretical and Applied Information

- Technology, vol. 46, no. 2, 2012.
3. R. Soltanzadeh, H. Rabbani, and A. Talebi, "Extraction of nucleolus candidate zone in white blood cells of peripheral blood smear images using curvelet transform," *Computational and Mathematical Methods in Medicine*, Hindawi Publishing Corporation, Volume 2012, Article ID574184.
 4. L. PutzuandC. Di Ruberto, "White blood cells identification and counting from microscopic blood image," *World Academy of Science, Engineering and Technology*, vol. 73, pp. 363–370, 2013.
 5. G. Zack, W. Rogers, and S. Latt, "Automatic measurement of sister chromatid exchange frequency," 1977.
 6. N. Guo, L. Zeng, Q. Wu, "A method based on multispectral imaging technique for white blood cell segmentation," *Computers in Biology and Medicine*, vol. 37, pp. 70–76, 2006.
 7. S. Mohapatra, D. Patra, and S. Satpathy, "Unsupervised blood microscopic image segmentation and leukemia detection using color based clustering," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp. 477–485, 2012.
 8. A. Abdul Nasir, M.Y. Mashor, and R. Hassan, "Classification of acute Leukaemia cells using multilayer perceptron and simplified fuzzy ARTMAP neural networks," *The International Arab Journal of Information Technology*, vol. 10, no. 4, July 2013.
 9. P. S. Hiremath, P. Bannigidad, and S. Geeta, "Automated identification and classification of white blood cells (leukocytes) in digital microscopic images," *IJCA special issue on "Recent Trends in Image Processing and Pattern Recognition"*, 2010
 10. H. T. Madhloom, A. S. Kareem, H. Asiffin, A. A. Zaidan, H. O. Alanazi and B. B. Zaidan, "An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold," *Journal of Applied Sciences*, vol. 10, no. 11, pp. 959–966, 2010.
 11. N. M. Salem, "K4 . Segmentation of White Blood Cells from Microscopic Images using K-means Clustering A . Color Spaces," pp. 371–376, 2014.
 12. M. C. R. Trujillo, T. E. Alarcón, O. S. Dalmau, and A. Zamudio Ojeda, "Segmentation of carbon nanotube images through an artificial neural network," *Soft Comput.*, vol. 21, no. 3, pp. 611–625, 2017.
 13. D. Oliva, E. Cuevas, G. Pajares, D. Zaldivar, and M. Perez-Cisneros, "Multilevel thresholding segmentation based on harmony search optimization," *J. Appl. Math.*, vol. 2013, 2013.
 14. D. Oliva, E. Cuevas, G. Pajares, D. Zaldivar, and M. Perez-Cisneros, "Multilevel thresholding segmentation based on harmony search optimization," *J. Appl. Math.*, vol. 2013, 2013.
 15. Z.W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: harmony search," *Simulation*, vol. 76, no. 2, pp. 60–68, 2001
 16. R. Srisha and a M. Khan, "Morphological Operations for Image Processing : Understanding and its Applications Morphological Operations for Image Processing : Understanding and its Applications," *NCVSComs-13*, no. February 2015, pp. 17–19, 2013.
 17. Rezatofighi, S.H., Soltanian-Zadeh, H.: Automatic recognition of five types of white blood cells in peripheral blood. *Computerized Medical Imaging and Graphics* 35(4) (2011) 333--343