Concavity Point and Skeleton Analysis Algorithm for Detection and Quantization in Heavily Clumped Red Blood Cells

Izyani Ahmad[#1], Siti Norul Huda Sheikh Abdullah[#2], Raja Zahratul Azma Raja Sabudin[*]

[#]*Pattern Recognition Research Group, Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia Bangi, Malaysia*
E-mail: [#1]izyani7172@gmail.com; [#2]snhsabdullah@ukm.edu.my

[*]*Pathology Department, Universiti Kebangsaan Malaysia Medical Centre, Cheras, Malaysia.*
E-mail: zahratul@ppukm.ukm.edu.my

Abstract- In practice, most hospitals use light microscope to examine the smeared blood for blood quantification. This visual quantification is subjective, laborious and time-consuming. Although automating the process is a good solution, the available techniques are unable to count or ignore the clumpy red blood cells (RBC). Moreover, clumping cell can affect the whole counting process of RBC as well as their accuracy. This paper proposes a new quantization process called concavity point and skeleton analysis (CP-SA) for heavily clump RBC. The proposed methodology is based on induction approach, enhanced lime blood cell by using gamma correction to get the appropriate edges. Then, splitting the clump and single cells by calculating each object area in pixel. Later, the quantification of clumpy cells with the proposed CP-SA method is done. This algorithm has been tested on 556 clump RBC taken from thin blood smear images under light microscope. All dataset images are captured from Hematology Unit, UKM Medical Centre in Kuala Lumpur. On all tested images, the cells of interest are successfully detected and counted from those clump cells. A comparative study and analysis to evaluate the performance of the proposed algorithm in three levels of clump have been conducted. The first level was with two clumps, second level with three clumps and third level with four clumps. The counting number of clump cells has been analyzed using quantitative analysis, resulting in much better results compared to other recent algorithms. The comparison shows that the proposed method gives better precision result at all levels with respect to ground truth: two clump cells (92%), three clump cells (96%) and four clump cells (90%). The results prove that this study has successfully developed a new method to count heavily clump RBC more accurately in microscopic images. In addition, this can be considered as a low-cost solution for quantification in massive examination.

Keywords – red blood cells; Concavity Point-Skeleton Analysis (CP-SA); heavily clump red blood cells quantization; thin blood smear.

I. INTRODUCTION

Human red blood cells (RBC) possess a typical form of a biconcave disk, with a diameter of 8 microns and reddish in color (Fig. 1(a)). In clinical practice, experts are required to perform the blood smears in response to a clinical feature or to a previously abnormal complete blood count in patients. They also have to manually classify the clump cells which is tedious, time-consuming and involves qualitative process [1], [2]. In addition, the existing methods contribute to inaccuracy, inconsistency and poor reliability diagnosis that may lead to false diagnosis situation.

Many computer-aided systems have successfully been developed for counting single or individual cells. However, a disease could better be diagnosed by counting both single and clump cells. Therefore, it is vital to guarantee an achievement of overall vision task and overcome mistaken diagnosis. Moreover, limited number of experts to examined

the growth number of blood samples each year, make it crucial to automate this routine through image processing methods. Here, our motivation of research comes to assist hematologist to count the clump of RBC automatically and then to further identify the morphology of the clumped red cells.

Heavily clumped cells such as '*Rouleaux*' and Agglutination (Fig. 1(b),(c)) in blood images could be due to underlying medical illness such as Plasma cell myeloma of Chronic lymphocytic leukemia but it could also be due to many other factors such as variety of blood smearing skill of the laboratory technician, blood storage and temperature. In a worse scenario, a repeated blood sample is impossible to be obtained from the patients (e.g. premature babies and death). It is a big challenge to the hematologist, in order to identify morphology of clump RBC accurately.
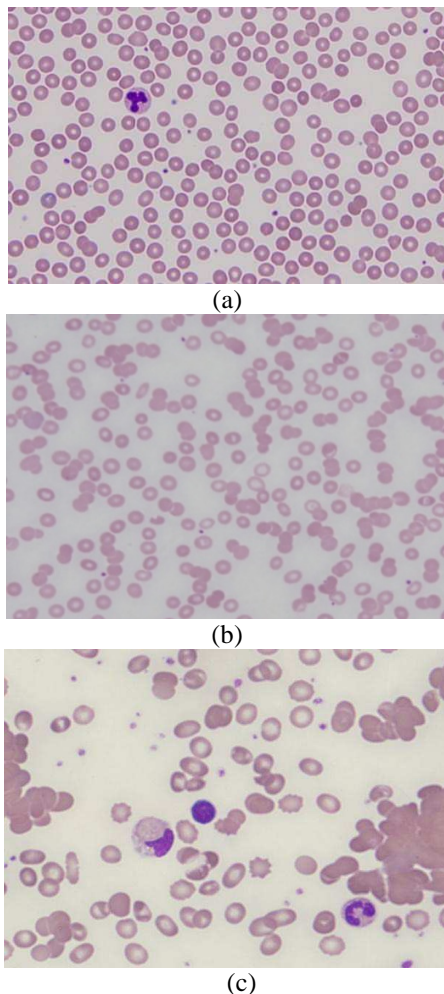
Fig. 1(a) Normal, (b)Rouleaux, (c) Agglutination. Source: UKM Medical Centre

Several methods have been proposed to predict the clumped cells regions. Some of them are based on Watershed and Distance transform [3], morphological [4]–[6], Appearance model [7], distance information [8] and Hough transform [9]. One recently common method for segmenting clumped cells is concavity analysis. The concavity point algorithm (CP) [10] is one of the techniques to segment and separate clump cells. Zafari used this method for the segmentation of partially overlapping nanoparticles with a convex shape in silhouette images. This method comprises of two main stages: contour evidence extraction and contour estimation. Contour evidence extraction starts with contour segmentation which recovered from a binarized image by detecting concave points. Then, contour segments are sought and grouped by utilizing the properties of fitted ellipses. Finally, the contour estimation is implemented through a non-linear ellipse fitting. This method relies only on edge information and can be applied to any segmentation problems where the objects are partially overlapped. However, some of the gradient values in the clump regions do not show a significant difference to other areas of the cells. Therefore, segmenting them according to the concavity and convexity of heavily clump cells are impossible because of the very narrow angles.

The Iterative Randomized Irregular Circular algorithm (IRIC) [11] is considered as a geometrical feature approach

as it was inspired by searching possible circle using non-collinear equation and distance criteria. At first, the image is divided into several partitions and IRIC algorithm is activated by randomly selecting four edge pixels prior to determine the potential circle. However, IRIC has some limitations; IRIC is less efficient when dealing with huge image size consisting of a high number of clump cells.

Our objective is to propose an algorithm to analyses heavily clump hematology images to automate detection and counting clump cells. Two experts perform gold standard analysis of the images. Then, evaluation and quantitative analysis are conducted from the proposed algorithm and its performance is being compared with several selected methods. The rest of this paper is arranged as follows: the proposed CP-SA algorithm for clump cells in blood smear images in Section II, experimental results and analyses in Section III, and the conclusion of this work in Section IV.

II. MATERIAL AND METHOD

We proposed a counting method for heavily clump cells to overcome the limitation of concavity-point methods. Two or more RBC clumps in various forms resulting in one or more concavities. In some of the clump cells, the gradient values in the clump regions do not show a remarkable difference compared to the other areas of the cells. Therefore, splitting them according to the concavity and convexity of heavy clump cells is impossible because of the very narrow angles. In the case of clumping area, more than fifty percent of concavity points cannot be established. As an alternative, this paper will introduce the framework to detect and count the multiple clumping cells based on the combination of concavity-point algorithm and skeleton analysis. Fig. 2 shows the general methodology of the proposed CP-SA algorithm.
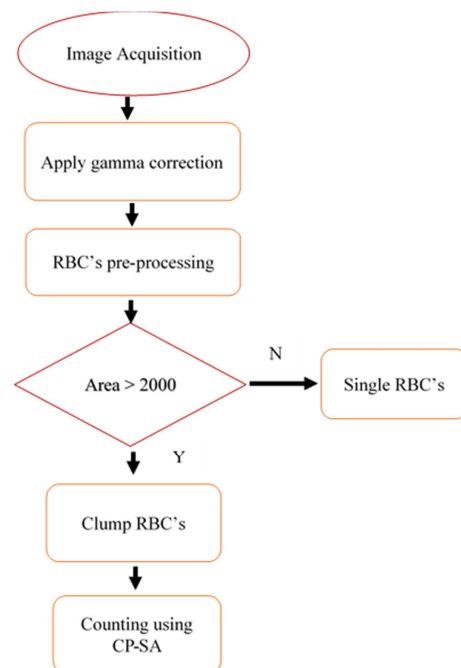

Fig. 2 General methodology for the proposed method

In this stage, gamma correction was used to make sure the cell edges to emerge clearly. This step was critical and

imposed an accurate counting cell at the end. Next, we applied Otsu's threshold method [12] for extracting the foreground from the background followed by morphological dilation and erosion methods in order to form a solid foreground pixel map and free from noise. This pre-processing step is widely used in computer vision involving other applications such as textiles[13] and agriculture [14]. Later, the connected regions of the foreground map were grouped together to identify the clump and non-clump regions based on the area. For the clump regions, further analysis was performed to predict the number of connected cells. Next, the white blood cells (WBC) images were extracted from RBC images by using thresholding; we studied the histogram of 20 sample grayscale images, and the best thresholding value to extract RBC was found at 0.7. This proposed method for cell segmentation worked with smooth edge images. Fig. 3 shows the overall pre-processing steps for the RBC. First, the gamma correction image was converted into a grayscale image by eliminating the Hue and saturation information while retaining its luminance.
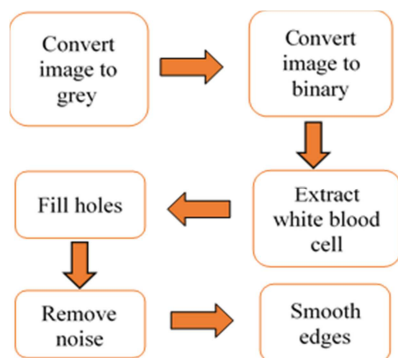

Fig. 3 Pre-processing step for RBC

Then, the image was converted into binary using the thresholding value of 0.7, to visualize all the RBC and WBC, as shown in Fig. 4(a). To remove the WBC from the image, the complementary white cells image was taken as in Fig. 4(b) and subtracted from the first image to obtain only the RBC, as shown in Fig. 4(c). When the image was converted to binary and the WBC were removed, some undesired pixels appeared after edge detection. These pixels represented either platelets or noise, and it subsequently affected the segmentation process. Therefore, these pixels were removed using a morphology operator on the binary image. An area was calculated for each cell to separate the overlapped and single RBC as shown in Fig. 6(d). From 100 images of normal single RBC, the average area obtained was at 2000 pixels. After these pre-processing steps, the image was prepared for input to our proposed Concavity Point with Skeleton Algorithm.
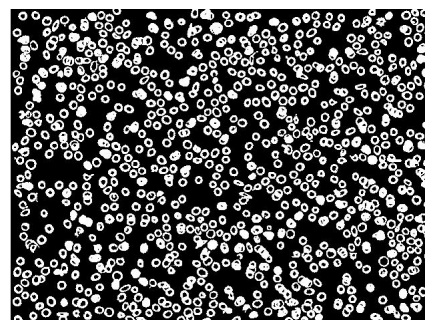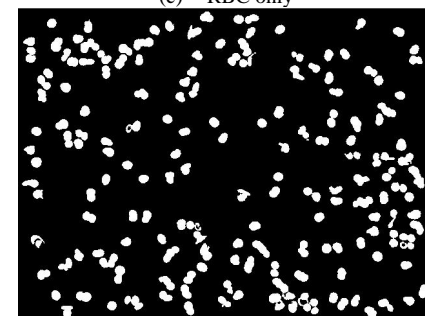

(a)   After thresholding
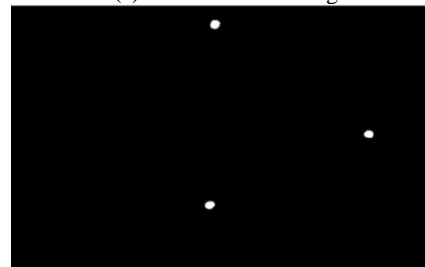
(b)   WBC only
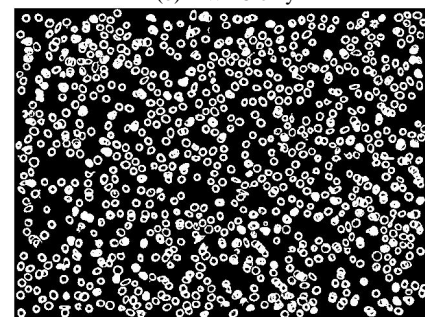
(c)   RBC only

(d)   Clump cells

Fig. 4 (a) –(d): Step in pre-processing RBC's.

Here, a part of algorithm in [10] was used to get the concave point along with skeleton algorithm [15]. As proven in previous research, the skeleton algorithm are best performed in binary images [16], [17] to get the backbone of the object. Our proposed method is shown below to outperform earlier methods in the task of heavily clump RBC based on detection rate and quantization accuracy. Here, we used rad=22.

Algorithm 1: Pseudo code to determine the distance of each concavity point

---

Input: Binary image of concavity point
Output: Array of distance for each concavity point
1. Check for concavity point denoted as *C (xci, yci)* and total number of concavity point is denoted as *N*.
   *Note: i=1,2 …,N

2. Determine distance of each concavity point using Equation (1).

$$\text{Distance} = \sqrt{(xc_i - xc_{i+1})^2 + (yc_i - yc_{i+1})^2} \quad (1)$$

3. Determine median of the distance as *median*.
4. Determine maximum distance as *maxDistance* and minimum distance *as minDistance*.

Algorithm 2: Pseudocode for detection and quantization of red blood cell

Input: Binary image of skeleton and concavity point
Output: Detected red blood cell

5. Obtain pixels located at the skeleton. These pixels referred as skeleton pixel, *SkelPix(i)* and the total number of skeleton pixel is denoted as *M*.
6. Check for concavity point denoted as *C (xci, yci)* and a total number of concavity point is denoted as *N*.

       *Note: i=1,2 …,N

7. **For** *Each Object* **do**
8.   **if** $N = 0$ AND $M <= 10$ **then**
9.   **for** each circle **do**
                Estimate circle =1
                Increment = radian-1
                Draw circle
            End for

10.   **if** $N < 2$ AND $M <= 10$ **then**
11.   **for** each circle **do**
                Estimate circle =1
                Increment = radian-1
                Draw circle
            End for

12.   **if** $N == 2$ AND ($M < 44$ AND $M > 10$) **then**
            Rad2 = M /2
13.   **for** each circle **do**
                Estimate circle =2
                Increment = rad2
                Draw circle
            End for

14.   **else**
15.   **if** *median* < rad+5 **then**
                i. D=rad+4
16.   **if** rad+7 > *median* >= rad+5 **then**
                i. D=rad+7
17.   **if** rad+10 > *median* >= rad+7 **then**
                i. D=rad+9
18.   **if** rad+40 > *median* >= rad+10 **then**
                i. D=rad+15
19.   **if** *median* >= rad+40 OR
            (*maxDistance – minDistance* >=15) **then**
                i. D=rad
                for each circle **do**
                    Increment = D
                    estimate circle =estimate circle+1
                    Draw circle
                End for
        End for

Based on the model in Fig. 5(a), with the assumption of the normal radius of RBC is at 22 pixels and clump between cells is more than 50% to resemble heavily clump RBC (i.e. Roulaeux), Fig. 5(b) shows our proposed method output. This model resembles the real cases of heavily clump cell in blood smear images. The first and second row (as simple model) show that the proposed method has successfully detected all the clump even without a complete number of concavity points. This situation normally happens in blood smear images. The third and fourth row represents a complex clump that often occurred.

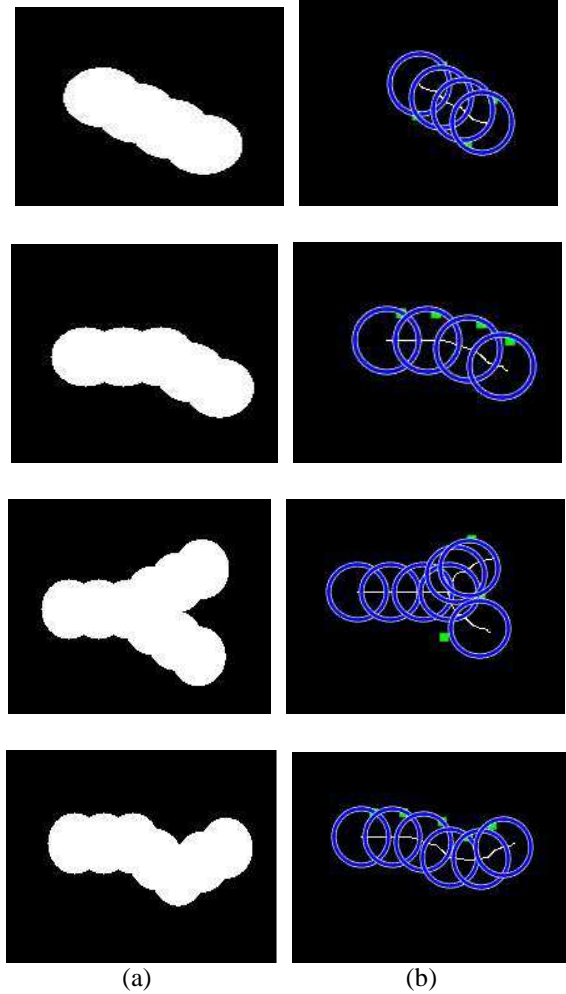
| (a) | (b) |

Fig. 5 (a) Simulation model, (b) Output of proposed method

These situations happen in a bad blood smear process or in some special diseases. As depicted, they are stacked in many angles and branches imposing a greater challenge to be detected and counted. Here, our proposed algorithm can successfully detect almost all the cases. This will give better accuracy in the counting of the whole number of cells in an image.

III. RESULTS AND DISCUSSION

The main objective of the CP-SA algorithm is to detect and count the RBC precisely in microscopic images of heavily clump cells.

1903

A. Experimental Setup

In this study, 556 clump RBC were cropped from three microscopic images of 'Rouleaux' patient's in RGB color. All patient's dataset was collected from the Hematology Unit, UKM Medical Centre. This process has been done under the supervision of the experts. This image was acquired under the light microscope with effective magnification at 40 times objective or equal to 400 magnifications.

In this study, we introduced three levels of clump namely simple, moderate and complex. The first level contained two clump cells as shown in Fig. 6(a). The second level contained three clump cells as in Fig. 6(b). While the third level with four clump cells as in Fig. 6(c). The images were tested based on the levels of clump using our proposed method, IRIC, and concavity point method.

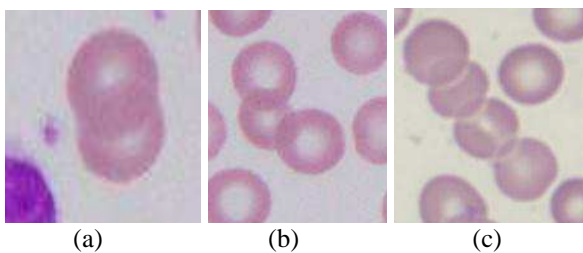| (a) | (b) | (c) |

Fig. 6 Zoomed image samples of three levels of clumping cells. (a) simple (b) moderate and (c) complex. Source: UKM Medical Centre

As a comparison, we conducted the experimentation based on the same dataset and pre-processing step with the chosen parameters as below:

CP_SA (proposed):	IRIC parameters:	CP
radian = 22	$T_a = 20$	$k = 10$
	$T_d = 5$	$t_1 = 23$
	$T_r = 0.4$	$t_2 = 45$
	$r_{min} = 10$	
	$r_{max} = 50$	

As mentioned previously, one of the most important criteria to diagnose a blood-based disease is the capability to correctly recognize and count the number of red blood cells including the clump cells accurately. Here, the average of true positive (TP), false negative (FN) and false positive (FP) values were determined. It is a condition when either there is an agreement between the expert and the method to detect the clump RBC, when the method is unable to detect the clump RBC but the expert is able to detect the clump, or when the method is able to detect the clump RBC but the expert is not. All the related formulas are:

$$\text{Precision, } PR = \frac{TP}{TP+FP}; \tag{2}$$

$$\text{Recall, } RC = \frac{TP}{TP+FN}; \tag{3}$$

$$\text{F-measure, } FM = \frac{2}{(\frac{1}{PR}+\frac{1}{RC})} \tag{4}$$

B. Quantitative Result

Table 1 summarizes the results of the RBC detection and count at each level of clump cell using three different methods. The best results obtained are in bold. Almost all levels of RBC segmented by IRIC gave the results below 77% for both accuracy and recall. The worst results were obtained for the four clumps with 42% and 37% for accuracy and recall. The overall clump mean for this method is under 65%. This finding supported that IRIC has an issue in segmenting and detecting cells in more than 3 clumps.

The concavity point algorithm (CP) [10] produced better results compared to the IRIC. It produced a higher mean percentage of accuracy and recall at 89% and 85% respectively. This finding supported the results as in [11] that IRIC is best segmenting and detecting cells under 3 clump. The proposed CP-SA algorithm produced the highest accuracy of mean at 91% and 96% for accuracy and recall respectively. Based on this result, there was a significant increment at 7% to 12% for accuracy and 11% to 18% for recall as compared to CP in each level of the clump. This finding proved that the proposed CP-SA algorithm has a high capability in overcoming the very narrow-angle in the concavity point algorithm, as mentioned before.

TABLE I
RESULTS FOR RBC DETECTION WITH THREE LEVELS OF CLUMP

METHOD	CLUMP	PR (%)	RC (%)	FM (%)
CP_SA (Proposed)	2	**92.0**	**98.0**	**95.0**
	3	**96.0**	**94.0**	**95.0**
	4	**90.0**	**94.0**	**92.0**
	MEAN	**91.0**	**96.0**	**94.0**
IRIC	2	75.0	77.0	76.0
	3	56.0	53.0	54.0
	4	42.0	37.0	39.0
	MEAN	65.0	62.0	63.0
CP	2	85.0	87.0	86.0
	3	89.0	90.0	89.0
	4	78.0	76.0	77.0
	MEAN	89.0	85.0	87.0

Based on Fig. 7, the number of undetected cells for the concavity point algorithm (CP) is around 70 cells and half of them is at 2 clump. It tends to detect this case as 1 cell. This finding is parallel with the earlier result showing that CP is less effective to detect almost fully overlapped cells. The IRIC has the biggest number of undetected cells with more than 100. This is due to IRIC suffers from under-segmentation and over-segmentation problems in heavily clumping cells. Based on [1], this algorithm is best suited for detecting RBC in blood smear images of a normal person. The proposed algorithm has successfully overcome this problem with the minimal number of undetected cells at less than 20. This finding is important to make sure that the whole quantization cell process is accurate and reliable. Therefore, this leads to a more accurate diagnosis of a disease by the experts and helps the patients in getting the right medication.
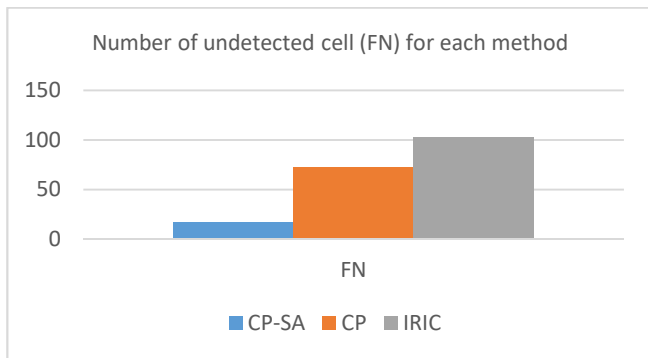
Fig. 7 Number of undetected cell (FN) by each algorithm. (1-CPSA, 2-CP, 3-IRIC)

C. Qualitative Result

The results of the segmentation and detection of the clump cells are shown in Fig. 8. Each column represents each level of clump. The first column shows level of detection for 2 clump, second column for 3 clump, and a third column for 4 clump. The detected cells are highlighted inside the boxes.
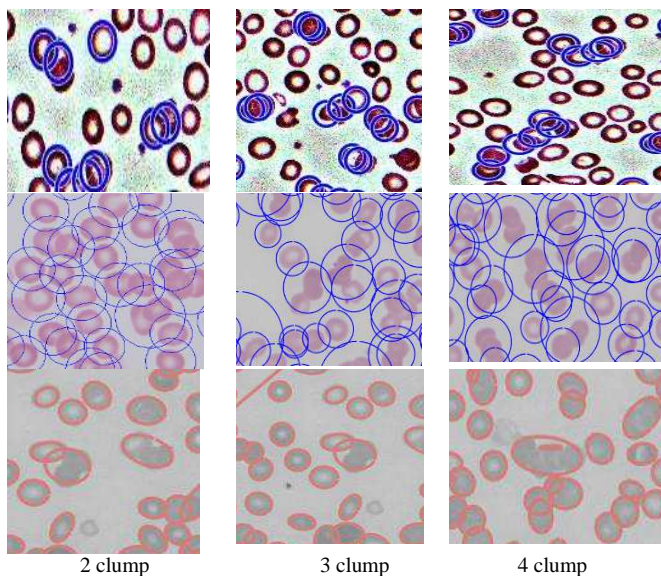

Fig. 8 RBC segmentation by CP-SA (first row), IRIC (second row) and CP (last row). Source: UKM Medical Centre

As clearly observable in Fig. 8, the second row produced by IRIC, the over-segmented and under-segmented occurrences were noted in 4 clump and 3 clump levels. This condition supports the earlier findings stating that IRIC has a limitation in detecting beyond 1 clump and 2 clump of RBC. On the other hand, CP algorithm produced a better detection as Fig. 8 (third row) compared to IRIC. However, it can be observed that CP has a poor detection of 2 clump RBC where it wrongly detected the clump as a single cell. The same problem occurred in 3 clump and 4 clump of almost totally overlapped cells. This result proved that CP has a problem in detecting a very narrow-angle that leads to non-concavity point detection. The proposed CP-SA algorithm overcomes this problem by combining the concavity point and skeleton analysis algorithms. It can assume and detect RBC correctly even at clump that almost fully overlapped as in Fig. 8 (first row). Thus, from these findings, the proposed

CP-SA algorithm produced the best results as compared to the IRIC and CP algorithm.

IV. CONCLUSION

As a sum up, we have examined the achievement of CP-SA, IRIC and CP algorithms on three levels of clump cell in blood smear images. The capability of the proposed method to segment and detect in heavily clumped cells had been analyzed on 556 clump RBC from thin blood smear images. Here, the qualitative outcomes of the detected circles in each case have been shown. We also presented a comparison of the quantitative table to show the result variations-SA used skeleton algorithm in getting the backbone of each clump RBC and analyzed it with the concavity point pixel to detect and count the RBC. Here, we found that CP-SA successfully caught the clumping RBC precisely, especially in cases of heavily clumpy blood smear images. However, IRIC tended to over-count the cells and CP suffered from an unrecognized concavity point. As a future reference, it is suggested for the proposed CP-SA algorithm to be modified for all medical and non-medical images to improve its robustness.

REFERENCES

[1] Y. M. Alomari, S. N. H. Sheikh Abdullah, R. Zaharatul Azma, and K. Omar, "Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm.," *Comput. Math. Methods Med.*, vol. 2014, p. 979302, 2014.

[2] Albashish, D., Sahran, S., Abdullah, A., Alweshah, M., & A., Adam. (2018). A hierarchical classifier for multiclass prostate histopathology image gleason grading.," vol. 2, no. 2, pp. 323–346, 2018.

[3] J. M. Sharif, M. F. Miswan, M. A. Ngadi, M. S. H. Salam, and M. M. bin A. Jamil, "Red blood cell segmentation using masking and watershed algorithm: A preliminary study," *Biomed. Eng. (ICoBE), 2012 Int. Conf.*, no. February, pp. 258–262, 2012.

[4] S. Chen, M. Zhao, G. Wu, C. Yao, and J. Zhang, "Recent Advances in Morphological Cell Image Analysis," *Comput. Math. Methods Med.*, vol. 2012, pp. 1–10, 2012.

[5] S. Khan, A. Khan, F. Saleh Khattak, and A. Naseem, "An Accurate and Cost Effective Approach to Blood Cell Count," *Int. J. Comput. Appl.*, vol. 50, no. 1, pp. 18–24, 2012.

[6] V. V. Panchbhai and L. B. Damahe, "RBCs and Parasites Segmentation from Thin Smear Blood Cell Images," *I.J. Image, Graph. Signal Process.*, vol. 10, pp. 54–60, 2012.

[7] R. Cai, Q. Wu, R. Zhang, L. Fan, and C. Ruan, "Red Blood Cell Segmentation Using Active Appearance Model," pp. 0–3.

[8] M. Hamghalam and A. Ayatollahi, "Automatic counting of leukocytes in giemsa-stained images of peripheral blood smear," *Proc. - 2009 Int. Conf. Digit. Image Process. ICDIP 2009*, pp. 13–16, 2009.

[9] M. Maitra, R. Kumar Gupta, and M. Mukherjee, "Detection and Counting of Red Blood Cells in Blood Cell Images using Hough Transform," *Int. J. Comput. Appl.*, vol. 53, no. 16, pp. 13–17, 2012.

[10] Zafari S., Eerola T., Sampo J., Kälviäinen H., Haario H. (2015) Segmentation of Partially Overlapping Nanoparticles Using Concave Points. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2015. Lecture Notes in Computer Science, vol 9474. Springer, Cham.

[11] Y. M. Alomari, S. N. H. S. Abdullah, R. R. M. Zin, and K. Omar, "Iterative randomized irregular circular algorithm for proliferation

rate estimation in brain tumor Ki-67 histology images," *Expert Syst. Appl.*, vol. 48, pp. 111–129, 2016.

[12]   N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[13]   S. Madenda and E. P. Wibowo, "Object Feature Extraction of Songket Image Using Chain Code Algorithm," vol. 7, no. 1, pp. 235–241, 2017.

[14]   R. Ruslan *et al.*, "Extraction of Morphological Features of Malaysian Rice Seed Varieties Using Flatbed Scanner Extraction of Morphological Features of Malaysian Rice Seed Varieties Using Flatbed Scanner," vol. 8, no. February, pp. 93–98, 2018.

[15]   W. Chen, L. Sui, Z. Xu, and Y. Lang, "Improved Zhang-Suen thinning algorithm in binary line drawing applications," *2012 Int. Conf. Syst. Informatics, ICSAI 2012*, no. Icsai, pp. 1947–1950, 2012.

[16]   W. Abu-Ain, S. N. H. Sheikh Abdullah, and K. Omar, "A simple iterative thinning algorithm for text and shape binary images," *J. Theor. Appl. Inf. Technol.*, vol. 63, no. 2, pp. 274–281, 2014.

[17]   W. Abu-Ain, S. N. H. S. Abdullah, B. Bataineh, T. Abu-Ain, and K. Omar, "Skeletonization Algorithm for Binary Images," *Procedia Technol.*, vol. 11, no. Iceei, pp. 704–709, 2013.