

# The Cascaded Enhanced $k$ -Means and Fuzzy $c$ -Means Clustering Algorithms for Automated Segmentation of Malaria Parasites

Aimi Salihah Abdul Nasir<sup>1,\*</sup>, Haryati Jaafar<sup>1</sup>, Wan Azani Wan Mustafa<sup>1</sup> and Zeehaida Mohamed<sup>2</sup>

<sup>1</sup>Faculty of Engineering Technology, Universiti Malaysia Perlis, UniCITI Alam Campus, Sungai Chuchuh, 02100 Padang Besar, Perlis, Malaysia

<sup>2</sup>Department of Microbiology & Parasitology, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

**Abstract.** Malaria continues to be one of the leading causes of death in the world, despite the massive efforts put forth by World Health Organization (WHO) in eradicating it, worldwide. Efficient control and proper treatment of this disease requires early detection and accurate diagnosis due to the large number of cases reported yearly. To achieve this aim, this paper proposes a malaria parasite segmentation approach via cascaded clustering algorithms to automate the malaria diagnosis process. The comparisons among the cascaded clustering algorithms have been made by considering the accuracy, sensitivity and specificity of the segmented malaria images. Based on the qualitative and quantitative findings, the results show that by using the final centres that have been generated by enhanced  $k$ -means (EKM) clustering as the initial centres for fuzzy  $c$ -means (FCM) clustering, has led to the production of good segmented malaria image. The proposed cascaded EKM and FCM clustering has successfully segmented 100 malaria images of *Plasmodium Vivax* species with average segmentation accuracy, sensitivity and specificity values of 99.22%, 88.84% and 99.56%, respectively. Therefore, the EKM algorithm has given the best performance compared to  $k$ -means (KM) and moving  $k$ -means (MKM) algorithms when all the three clustering algorithms are cascaded with FCM algorithm.

## 1 Introduction

Malaria is an infectious disease caused by the Plasmodium blood parasite, with high prevalence in tropical and subtropical regions. World Health Organization (WHO) has given the analysis report for the year 2015 as around 214 million peoples on earth suffers from malaria, with 90% of the victims were from African Region amassing a total of 438,000 death cases [1]. Across the globe, it was recorded that 306,000 children under the age of five lost their lives due to malaria, with 292,000 of them are from Africa [1].

Malaria develops to become life-threatening without immediate action. Currently, microscopy-based diagnosis is the most widely used approach to diagnose malaria. The manual microscopic diagnosis is often characterized by its high sensitivity and accuracy. However, its time-consuming and labour intensive characteristics seem to outweigh benefits as it consumes approximately 20 to 60 minutes to examine each slide [2]. The final diagnosis ultimately depends on the ability and experience of the experts, and may be prone to human error [3].

Recognition of malaria parasite requires the segmentation of the malaria parasite or infected red blood cells (RBCs) in malaria image. However, the segmentation process is difficult since most of the

parasites are often a part of the RBCs that they have infected; presented in a complex and irregular visual aspect. As a result, various applications of image processing techniques for segmenting the malaria parasites have been reported to necessitate the malaria diagnosis process. These include the use of thresholding [4-5], marker controlled watershed with morphological approach [6-7], fuzzy rule base segmentation [8], active contour base segmentation [9] and clustering approach [10-11] as the latest trend.

Clustering is an unsupervised procedure that aims to group a set of data into meaningful groups, based on the distances between data points.  $K$ -means, fuzzy  $c$ -means and mean shift are among the popular clustering algorithms that have been utilized to segment the region of interest in various types of medical images.  $K$ -means (KM) clustering [12] employs the concept of data allocation to the nearest cluster centre based on the shortest Euclidean distance measurements.

Though KM algorithm is very popular for general clustering, this algorithm is sensitive to cluster centres initialization and has the dead centre problem [13]. The limitations of KM algorithm could lead to poor representation of data. As a result, several clustering algorithms such as fuzzy  $c$ -means and enhanced  $k$ -means algorithms have been proposed to overcome the limitations of KM algorithm.

\* Corresponding author: [aimisalihah@unimap.edu.my](mailto:aimisalihah@unimap.edu.my)

In fuzzy *c*-means (FCM) clustering [14], each data is assigned a membership grade indicating the degree of belonging of the data to each centre rather than assigning each data to one centre as in KM algorithm. The centre is an average of all data points in the cluster, weighted by their degree of belonging of the data to the centre. Substantially, these attributes could minimize the dead centre problem in FCM algorithm. However, this algorithm is still sensitive to cluster centres initialization [15]. The limitations of FCM algorithm could lead to poor segmentation result regardless of the resultant cell images being under-segmented or over-segmented as shown in [16-17].

Another type of clustering algorithm that has been proven to produce good segmentation performance on medical image is enhanced *k*-means (EKM) clustering [18] algorithm. In EKM clustering, the concept of variance and a new version of transferring process for clustered members are used to assist the assignation of data to the proper centre during the process of clustering, so that good segmented image can be generated. In addition, comparisons of different image segmentation techniques show that EKM algorithm has successfully segmented the malaria parasite from the overlapping RBCs compared to the results of KM and FCM clustering [18].

Based on the advantages of EKM algorithm in image segmentation, the current study will utilize its potential by cascading the EKM with FCM algorithm. In addition, the other two popular clustering algorithms which are KM and moving *k*-means (MKM) [19] will also be cascaded with FCM algorithm to compare the performance of segmentation among these three cascaded algorithms.

## 2 Methodology

### 2.1 Image Acquisition

The malaria thin blood smear samples were obtained from Department of Microbiology and Parasitology, Hospital Universiti Sains Malaysia (HUSM), Malaysia. The malaria images were acquired by using a Luminera Infinity-2 digital camera mounted to the *Leica DLMA 1200* digital microscope with 100X magnifications. Images were captured in the bitmap (\*.bmp) format at the resolution of 800×600 pixels. This study have analysed a total of 100 images obtained from eight malarial *Plasmodium Vivax* slides. The images consist of malaria parasites in ring, trophozoite, schizont and gametocyte stages.

### 2.2 The Proposed Cascaded Enhanced *k*-Means and Fuzzy *c*-Means Clustering Algorithms

As previously mentioned, FCM clustering algorithm was introduced as a fuzzy version of KM algorithm for reducing the weaknesses of KM. However, its performance is depending on the choice of the initial cluster centre. If a good initial cluster centre that is close to the actual final cluster centre is used, the FCM

algorithm will converge faster and the processing time can be deducted. Without a proper initialization, a set of poor final cluster centres would be generated by this algorithm hence producing poor clustering results. Even though FCM is sensitive to cluster centre initialization, the concepts of fuzziness makes this algorithm becomes more flexible in producing the best possible centre.

Thus, to take advantage of EKM and FCM clustering algorithms, the cascaded EKM and FCM clustering algorithm is proposed. Technically, the cascaded clustering algorithm combines the fundamental theory of EKM algorithm which employs the variance concept for data allocation to its cluster centre, with the concept of FCM algorithm that allows each data to be simultaneously assigned to multiple clusters by different degrees of membership. Thus, the final centres generated by EKM clustering will become the initial centres for FCM clustering. Besides, fuzzy membership function is used along with Euclidean distance and variance calculations, in order to assist the assignation of the members to the proper centre during the clustering process.

In order to describe the cascaded EKM and FCM clustering algorithms, consider an image with resolution of  $X \times Y$  pixels to be clustered into  $n_c$  regions. Let  $p(x,y)$  as an input pixel to be clustered and  $c_j$  is the  $j$ -th centre (cluster) ( $x = 1, 2, \dots, X, y = 1, 2, \dots, Y$  and  $j = 1, 2, \dots, n_c$ ). The cascaded EKM and FCM clustering algorithms for image segmentation is described as follows:

1. Initialize the centres using:

$$c_j = \min_{p(x,y)} + (2j - 1) \left( \frac{\max_{p(x,y)} - \min_{p(x,y)}}{2n_c} \right) \quad (1)$$

where  $\min_{p(x,y)}$  and  $\max_{p(x,y)}$  are the minimum and maximum pixel intensity in the image.

2. Set the values of  $\alpha_0 = 0.05$ ,  $\alpha_b = 0.05$  and  $\alpha_a = \alpha_0$ , where  $\alpha_0$  and  $\alpha_b$  are small constant values.
3. For each pixel of an image, calculate the Euclidean distance,  $d$  using:

$$d = \|p(x,y) - c_j\| \quad (2)$$

4. Assign all pixels to the nearest centre based on  $d$  and calculate the centre positions using:

$$c_j = \frac{1}{n_j} \sum_{y \in c_j} \sum_{x \in c_j} p(x,y) \quad (3)$$

5. Check the variance between the members with their centre using:

$$\sigma^2(c_j) = \frac{1}{n_j} \sum_{y \in c_j} \sum_{x \in c_j} (p(x,y) - c_j)^2 \quad (4)$$

Note: The variance calculation is applied for measuring the relationship between the centre and its members.

6. Find  $c_s$  and  $c_l$ , the centre that has the smallest and the largest value of variance between the members with their centre.
7. If  $\sigma^2(c_s) < \alpha_a \sigma^2(c_l)$ ,
  - a. Find  $c_n$ , the nearest centre to  $c_l$  by calculating the Euclidean distance between the  $c_l$  and other centres.
  - b. If  $c_n < c_l$ , assign the members of  $c_l$  to the nearest centre if  $p(x,y) < c_l$ , where  $x,y \in c_l$ , and leave the rest of the members to  $c_l$ .
  - c. Else if  $c_n > c_l$ , assign the members of  $c_l$  to the nearest centre if  $p(x,y) > c_l$ , where  $x,y \in c_l$ , and leave the rest of the members to  $c_l$ .

Note: The transferring process of  $c_l$  members in (b) and (c) is performed in order to reduce the variance of  $c_l$  by giving up members for  $p(x,y) < c_l$  and  $p(x,y) > c_l$ , respectively to the appropriate cluster centre so that the variance of  $c_l$  will be kept at a reasonable value.

- d. Recalculate the positions of  $c_l$  and  $c_n$  according to:

$$c_l = \frac{1}{n_l} \sum_{y \in c_l} \sum_{x \in c_l} p(x, y) \quad (5)$$

$$c_n = \frac{1}{n_n} \sum_{y \in c_n} \sum_{x \in c_n} p(x, y) \quad (6)$$

Note:  $n_l$  and  $n_n$  in (5) and (6) are the number of the new members of  $c_l$  and  $c_n$  respectively, after the reassigning process in step 7d.

8. Update  $\alpha_a$  according to  $\alpha_a = \alpha_a - \alpha_a / n_c$  and repeat steps 6 and 7 until  $\sigma^2(c_s) \geq \alpha_a \sigma^2(c_l)$ .
9. Reassign all pixels to the nearest centre and recalculate the centre positions using Eq. 3.
10. Update  $\alpha_a$  and  $\alpha_b$  according to  $\alpha_a = \alpha_0$  and  $\alpha_b = \alpha_b - \alpha_b / n_c$  respectively, and repeat steps 5 to 9 until  $\sigma^2(c_s) \geq \alpha_b \sigma^2(c_l)$ .
11. Calculate membership function,  $M_{jp(x,y)}$  according to Eq. 7 to 9.

$$M_{jp(x,y)} = 1 / \left( \sum_{k=1}^{n_c} (d_{jp(x,y)} / d_{kp(x,y)})^2 \right) \quad (7)$$

where,

$$d_{jp(x,y)} = \|p(x, y) - c_j\| \text{ and } d_{kp(x,y)} = \|p(x, y) - c_k\| \quad (8)$$

$$\left. \begin{matrix} M_{kp(x,y)} = 1 \\ M_{jp(x,y)} = 0; \text{ for } p(x, y) \neq k \end{matrix} \right\} \text{ if } d_{kp(x,y)} = 0 \quad (9)$$

12. Calculate the centre positions using:

$$c_j = \frac{\sum_{y \in c_j} \sum_{x \in c_j} M_{jp(x,y)}^2 p(x, y)}{\sum_{y \in c_j} \sum_{x \in c_j} M_{jp(x,y)}^2} \quad (10)$$

In order to ensure a better clustering process, the basic concept of EKM algorithm suggests that each cluster must have strong relationship strength between the members with their centre so that the differences between them could be minimized. After the final centres produced by EKM algorithm are obtained, these centres will be fed as initial values to FCM algorithm for further clustering process. In FCM algorithm, the fuzziness concept is used for allowing each data to assign a membership grade indicating the degree of belonging of the data to each centre simultaneously. However, in the proposed cascaded EKM and FCM clustering algorithms, steps 11 and 12 are used once and not be repeated as in the conventional FCM algorithm. This is done in order to test the capability of each single clustering algorithm in producing good initial cluster centre to FCM algorithm. If a good initial cluster centre is used, hence a set of good final cluster centres would be generated by FCM algorithm after performing steps 11 and 12. Therefore, the proposed cascaded EKM and FCM clustering algorithms combines the characteristics of both clustering algorithms to produce the best clustering process.

### 2.3 Segmentation of Malaria Parasites

In order to obtain the segmented malaria parasites, the image segmentation process has been conducted based on the procedures as proposed in [20]. In [20], these procedures were used for comparing the performance of malaria parasite segmentation using different colour components. Meanwhile, this study utilizes the similar segmentation procedures for comparing the performance of three cascaded clustering algorithms in segmenting the malaria image. The three cascaded clustering algorithms are obtained by cascading KM, MKM and EKM with FCM clustering algorithm. The procedures for segmentation of malaria parasite are illustrated in Fig. 1.

Initially, the pre-processing step is conducted by applying the modified global contrast stretching (MGCS) technique to generate high contrast image. Then, segmentation process using the three cascaded clustering algorithms is performed on green component image to isolate the RBCs and background regions, followed by  $7 \times 7$  pixels median filter for noise reduction. Finally, seeded region growing area extraction (SRGAE) algorithm is used to remove the large unwanted regions such as artefacts, platelets and RBCs in the image.

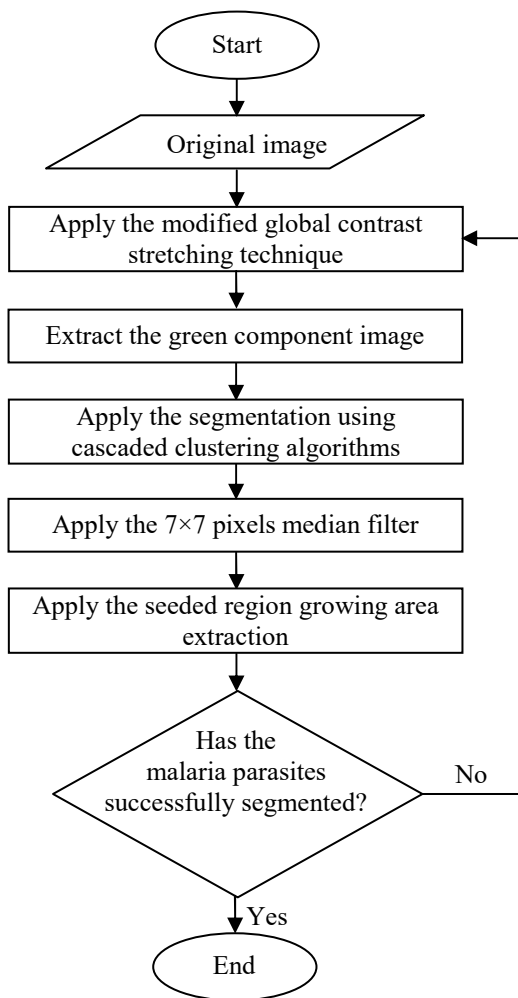


Fig. 1. The procedures for segmentation of malaria parasite.

### 3 Results and Discussions

The three cascaded clustering algorithms have been proposed in order to combine the advantages of KM, MKM and EKM with the FCM clustering. Therefore, comparisons among these three cascaded clustering algorithms have been made in order to measure the performance of each cascaded clustering algorithm for image segmentation. Figs. 2(a) and (b) show the original malaria images of Schizont and Gametocyte, respectively. The results obtained after applying the MGCS technique with  $min_p=0.5\%$  and  $max_p=10\%$  on the two malaria images are shown in images (c) and (d).

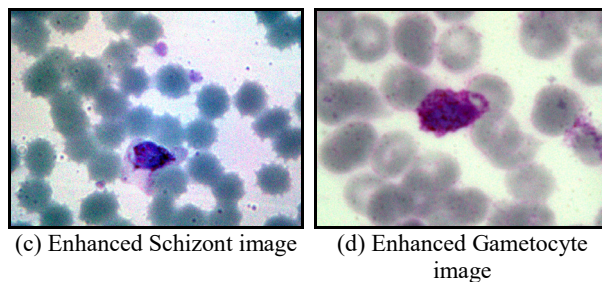
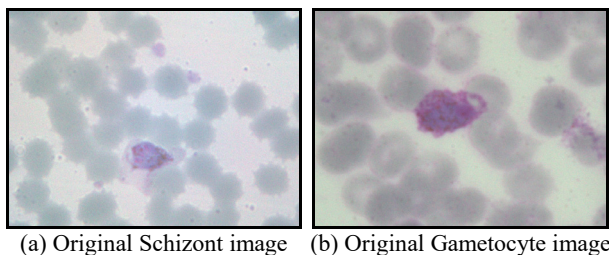
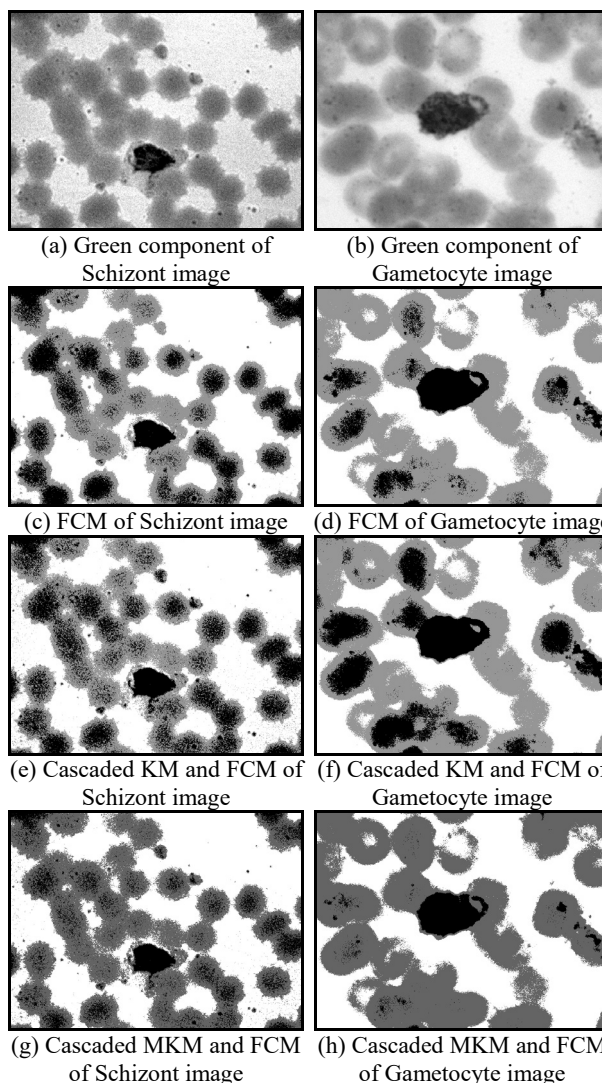
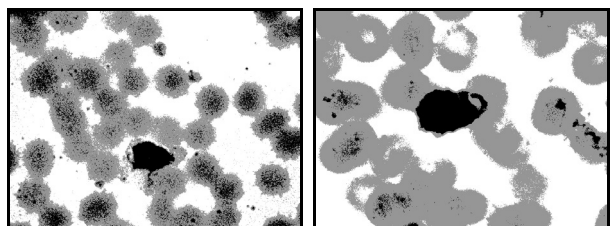


Fig. 2. The original and enhanced malaria images.

In order to easily isolate the malaria parasite from the RBCs and background regions, the green component information has been extracted from the enhanced RGB image. Then, FCM and the three cascaded clustering algorithms have been applied for segmenting the malaria image. Fig. 3 shows the results of images after applying the clustering algorithms on the green component image.

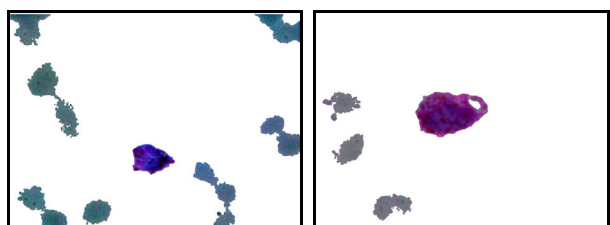




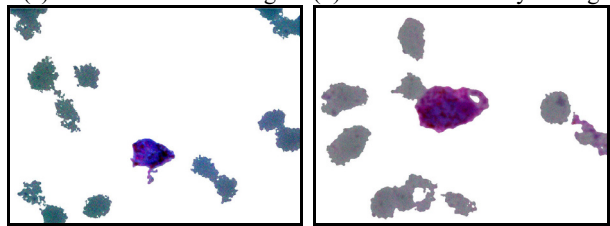
(i) Cascaded EKM and FCM of Schizont image (j) Cascaded EKM and FCM of Gametocyte image

**Fig. 3.** Results of images for the four clustering algorithms.

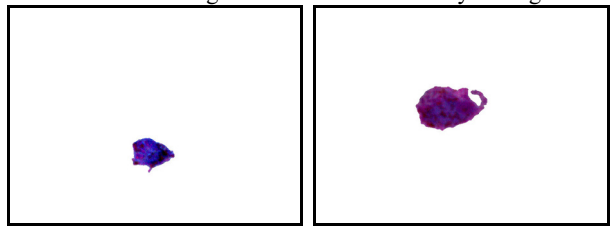
Later, the resultant images will undergo filtering process using  $7 \times 7$  pixels median filter for noise reduction and image smoothing. Finally, SRGAE algorithm has been applied to remove any unwanted regions which are outside the ranges of 3500 to 30000 pixels from the image and the results are shown in Fig. 4. Based on the final segmented images, segmentation using the cascaded KM and FCM clustering is unable to produce a clean segmented malaria image, as the malaria parasite and some of the RBCs component have been clustered into a single region. This over-segmentation is due to pixel similarity of green component between the malaria parasite and RBCs regions especially in the case of overlapping RBCs. In contrast, the cascaded MKM and FCM, as well as the cascaded EKM and FCM clustering algorithms are capable to produce a clean segmented malaria image without or less appearance of RBCs region in which can be seen in the resultant images of cascaded KM and FCM clustering.



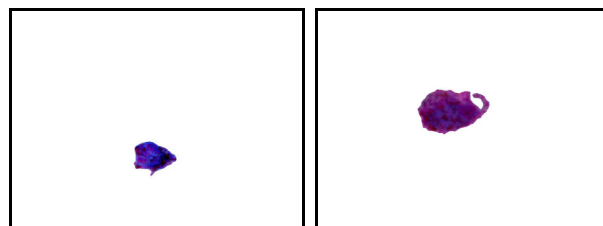
(a) FCM of Schizont image (b) FCM of Gametocyte image



(c) Cascaded KM and FCM of Schizont image (d) Cascaded KM and FCM of Gametocyte image



(e) Cascaded MKM and FCM of Schizont image (f) Cascaded MKM and FCM of Gametocyte image



(g) Cascaded EKM and FCM of Schizont image (h) Cascaded EKM and FCM of Gametocyte image

**Fig. 4.** Results of final segmented images for Schizont and Gametocyte.

In order to ascertain segmentation performances provided in the previous results, the quantitative readings of 100 malaria images are collected. Table 1 summarizes the average segmentation performances based on sensitivity, specificity and accuracy that have been collected from 100 malaria images. The final segmented images are compared to manually segmented images to quantify the segmentation performance for clustering algorithm. The accuracy of segmentation can be obtained by computing the percentage of pixels that are correctly segmented as malaria parasite or background region in the image. The sensitivity is obtained by computing the percentage of pixels that are correctly segmented as positive region, while the specificity is obtained by calculating the percentage of pixels that are correctly segmented as negative region. Higher pixel similarity indicates better segmentation performance.

Based on Table 1, segmentation using the cascaded KM and FCM clustering has proven to be the best in obtaining the fully segmented malaria parasite with sensitivity of 89.9816%. Even though the sensitivity result that has been achieved is high, segmentation using this algorithm is unable to produce a clean segmented malaria image. In contrast, segmentation using the cascaded MKM and FCM clustering has proven to be the best in producing a clean segmented malaria image with specificity of 99.5603%, which is slightly higher than cascaded EKM and FCM clustering (99.5560%).

As for segmentation accuracy, segmentation using the cascaded EKM and FCM clustering has proved to be slightly better than the cascaded MKM and FCM clustering in segmenting the entire area in malaria image with accuracy of 99.2196%. In this study, the main criterion for evaluating the performance of each clustering algorithm is based on its ability to produce fully segmented malaria parasite region as well as separating the malaria parasite from the overlapping RBCs. Thus, the cascaded EKM and FCM clustering algorithm has given the best performance by providing good sensitivity as well as high specificity and accuracy values compared with the results provided by other two cascaded clustering algorithms. This result also shows that by using the final centres that are generated by EKM clustering as the initial centres for FCM clustering, has resulted in good segmentation of malaria images.

**Table 1.** Segmentation performances for the segmented malaria images.

Analysis	Segmentation Technique	Schizont	Gametocyte	Average of 100 Images
Sensitivity (%)	FCM	95.6705	92.1226	89.4300
	Cascaded KM and FCM	<b>98.6198</b>	<b>95.5749</b>	<b>89.9816</b>
	Cascaded MKM and FCM	96.9926	87.3575	87.6128
	Cascaded EKM and FCM	97.4575	88.6024	88.8437
Specificity (%)	FCM	90.4158	97.1253	98.2454
	Cascaded KM and FCM	89.8885	88.1643	97.3182
	Cascaded MKM and FCM	<b>99.9242</b>	<b>99.9835</b>	<b>99.5603</b>
	Cascaded EKM and FCM	99.9172	99.9750	99.5560
Accuracy (%)	FCM	90.4894	96.9306	97.9489
	Cascaded KM and FCM	90.0109	88.4526	97.0682
	Cascaded MKM and FCM	99.8812	99.4921	99.1770
	Cascaded EKM and FCM	<b>99.8828</b>	<b>99.5324</b>	<b>99.2196</b>

## 4 Conclusions

This paper has proposed an image segmentation technique for malaria images, by cascading two clustering algorithms which are EKM and FCM algorithms. Overall, the EKM algorithm has given the best performance compared to KM and MKM algorithms when all the three algorithms are cascaded with FCM algorithm. In addition, a better segmentation result has been achieved by cascading two clustering algorithms as compared of using a single clustering algorithm for image segmentation. Here, the segmentation performances using 3 clusters for the cascaded EKM and FCM algorithms is found to be slightly effective than the EKM algorithm with average segmentation accuracy, sensitivity and specificity values of 99.22%, 88.84% and 99.56%, respectively. Overall, these findings prove that the proposed cascaded EKM and FCM clustering algorithms have good potential in the application of malaria image segmentation as compared to the other segmentation techniques.

The authors gratefully acknowledge and thank Hospital Universiti Sains Malaysia (HUSM) for providing the malaria blood samples and validating the results. We also would like to acknowledge Universiti Malaysia Perlis (UniMAP) for providing the financial support under Short Term Grant Scheme (Grant No. 9001-00514).

## References

1. *World malaria report 2015* (World Health Organization, Switzerland, 2016)
2. B.S. Kakkilaya, *Lab Medicine* **34**, 602 (2003)
3. K. Mitiku, G. Mengistu, B. Gelaw, *Ethiopian Journal of Health Development* **17**, 197 (2003)
4. V.V. Panchbhai, L.B. Damahe, A.V. Nagpure, P.N. Chopkar, *International Journal of Image, Graphics and Signal Processing* **4**, 54 (2012)
5. J.E. Arco, J.M. Górriz, J. Ramírez, I. Álvarez, C.G. Puntonet, *Expert Systems with Applications* **42**, 3041 (2015)
6. J. Soni, *Int J Eng Sci Technol* **3**, 5260 (2011)
7. D.K. Das, M. Ghosh, M. Pal, A.K. Maiti, C. Chakraborty, *Micron* **45**, 97 (2013)
8. C. Dallet, S. Kareem, I. Kale, *2014 IEEE International Symposium on Circuits and Systems (ISCAS)* (2014)
9. X. Gual-Arnau, S. Herold-García, A. Simó, *Med Biol Eng Comput* **53**, 623 (2015)
10. J. Somasekar, B.E. Reddy, *Computers & Electrical Engineering* **45**, 336 (2015)
11. S.S. Savkare, A.S. Narote, S.P. Narote, *Proceedings of the Third International Symposium on Computer Vision and the Internet* (2016)
12. J. MacQueen, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* (1967)
13. S.S. Khan, A. Ahmad, *Pattern Recognition Letters*, **25**, 1293 (2004)
14. J.C. Bezdek, *Pattern recognition with fuzzy objective function algorithms* (Plenum Press, New York, 1981)
15. M.C. Hung, D.L. Yang, *Proceedings 2001 IEEE International Conference on Data Mining* (2001)
16. N.A.M. Isa, M.Y. Mashor, N.H. Othman, *Proceeding of International Conference on Robotics, Vision, Information and Signal Processing* (2003)
17. N.H. Harun, M.Y. Mashor, H. Rosline, *IFMBE Proceedings* (2011)
18. A.S. Abdul Nasir, M.Y. Mashor, Z. Mohamed, *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* (to be published)
19. M.Y. Mashor, *International Journal of the Computer, The Internet and Management* **8**, 50 (2000)
20. A.S. Abdul-Nasir, M.Y. Mashor, Z. Mohamed, *Journal of Biomimetics, Biomaterials and Biomedical Engineering* **24**, 43 (2015)