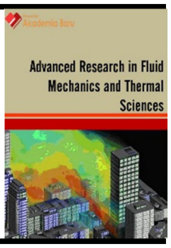
Penerbit
Akademia Baru

Journal of Advanced Research in Fluid
Mechanics and Thermal Sciences

Journal homepage: www.akademiabaru.com/arfmts.html
ISSN: 2289-7879



Enhanced *k*-Means Clustering Algorithm for Malaria Image Segmentation

Open Access

Aimi Salihah Abdul Nasir[1,2,*], Mohd Yusoff Mashor[2], Zeehaida Mohamed[3]

[1] Faculty of Engineering Technology, Universiti Malaysia Perlis, UniCITI Alam Campus, Sungai Chuchuh, 02100 Padang Besar, Perlis, Malaysia
[2] Electronic & Biomedical Intelligent Systems (EBItS) Research Group, School of Mechatronic Engineering, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia
[3] Department of Microbiology & Parasitology, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

ARTICLE INFO

Article history:
Received 2 December 2017
Received in revised form 30 January 2018
Accepted 10 February 2018
Available online 18 February 2018

ABSTRACT

Clustering is one of the most commonly used digital image segmentation technique in multifarious fields including medical image segmentation. In essence, this study proposes clustering algorithm to acquire good segmented images of *Plasmodium Vivax* malaria parasite species via unsupervised pixel segmentation. In this study, enhanced *k*-means (EKM) clustering algorithm which is an enhanced version of the conventional *k*-means (KM) clustering algorithm has been proposed for malaria slide image segmentation. In the proposed EKM clustering algorithm, 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. The satisfactory sensitivity together with the high specificity and accuracy values obtained from an average of 100 malaria images, indicates that the EKM algorithm has provided good segmentation performances as compared to *k*-means, fuzzy *c*-means and moving *k*-means clustering algorithms. Good segmented malaria parasite and clean segmented malaria image has been acquired using the proposed clustering algorithm. Hence, the proposed EKM clustering can be considered as an image segmentation tool for segmenting the malaria images.

Keywords:
Clustering, enhanced *k*-Means, image segmentation, Malaria

1. Introduction

Malaria is a major health problem in the world, particularly in tropical and subtropical regions, where mortality cases recorded are significantly higher than any other infectious diseases except Tuberculosis (TB). In 2015, it was estimated that out of 212 million cases of malaria, 429,000 deaths were reported globally, where 90% of the victims were from African Region [1]. Children aged less than five years and pregnant women are classified as the most vulnerable and the severely affected [1]. *Plasmodium Falciparum* and *Plasmodium Vivax* are the two malaria main species that have

---

[*] *Corresponding author.*
E-mail address: aimisalihah@unimap.edu.my (Aimi Salihah Abdul Nasir)

spawn the most malaria infections in worldwide. Each of the species undergoes specific life-cycle comprising of the ring (young trophozoite), mature trophozoite, schizont and gametocyte.

Manual microscopic examination of Giemsa-stained blood smears using a light microscope is still accepted as the gold standard for confirmation of malaria disease [2]. In blood analysis, parasites and infected RBCs are recognizable by their physical features. However, visual inspection of blood smears is a time-consuming task which could lead to misdiagnosis especially when microbiologists need to diagnose a large number of samples [3]. As a result, several image processing methods have been reported for computerized recognition of malaria parasites [4,5].

Mohana-Rao and Dempster [6] have suggested morphological area opening on distance transform for choosing the markers in order to avoid the under and over-segmentation of malaria image. Das *et al*., [7] applied gray world assumption in order to correct the illumination in malaria image. Then, the RBCs have been segmented using the marker controlled watershed algorithm. Preedanan *et al*., [8] applied local adaptive thresholding method on the grayscale image in order to segment the RBCs in malaria image. In a more recent study, Muda and Salam [9] used clustering technique to segment the blood cells that were infected with malaria parasites. A two-step image segmentation process using *k*-means, fuzzy *c*-means and mean-shift followed by Median-cut have been used to perform this task. The results indicate that k-means algorithm has performed better in segmenting the cell images.

Clustering is a data processing task that involves grouping a set of data patterns into groups with similar characteristics. For medical application, clustering has been applied for segmentation of region of interest in various types of images such as malaria [9,10], leukaemia [11], retinal [12] as well as dermoscopic images [13]. *K*-means (KM) clustering [14] is one of the most command methods applied for image segmentation because of its simplicity and easily implemented. During the KM clustering for image segmentation process, each pixel in the image is assigned to the closest cluster centre, where the distance between the pixel and the centre is measured by using the Euclidean distance [14]. Though KM algorithm is very popular for general clustering, this algorithm having some drawbacks like being sensitive to initialization of cluster centres, has the dead centre problem which results in empty clusters, as well as the possibility to converge to local minima [15,16]. All these weaknesses of KM algorithm could lead to poor representation of data. Hence, unacceptable segmentation results regardless of the resultant images being under-segmented or over-segmented are generated.

In order to overcome the aforesaid weaknesses, fuzzy *c*-means (FCM) clustering [17] which employs fuzzy concept has been introduced. In FCM clustering algorithm, 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, the FCM algorithm is still sensitive to initial centres, and may converge to local minima as well as consuming more time to converge [18].

Some advanced clustering algorithms have been proposed to overcome drawbacks of KM algorithm in solving the general clustering problems [19,20]. One of the proposed algorithms is moving *k*-means (MKM) clustering algorithm [19]. This study has highlighted the main contribution of MKM which is the algorithm is not sensitive to the initial cluster centres. Therefore, the MKM algorithm offers several advantages over the KM algorithm like its ability for avoiding the dead centres and centre redundancy. The fitness concept is proposed to ensuring that each cluster has a significant number of members and final fitness values before the new position of the centre is calculated. Although it becomes less sensitive to initialization, comparison of different clustering algorithms for image segmentation conducted by Yusoff and Isa [21] shows that MKM algorithm was unable to provide fully and smoother segmented cytoplasm region in cervical cell image compared with the results given by KM and 2D-KM algorithms.

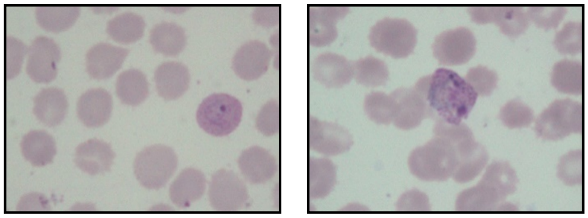
In this study, a modified clustering algorithm namely enhanced *k*-means (EKM) clustering is proposed as an enhanced version of the conventional KM clustering algorithm. The performance of EKM clustering will be compared with the other three popular clustering algorithms which are KM, FCM and MKM clustering algorithms.

2. Methodology

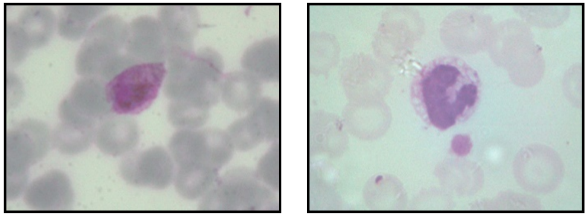
A. Data Acquisition of Malaria Images

The malaria blood slides were obtained from Department of Microbiology and Parasitology, Hospital Universiti Sains Malaysia (HUSM), Malaysia. For visualization of the cell components, each slide was stained using Giemsa staining. The malaria slides were analyzed using *Leica DLMA 1200* digital microscope under 100X oil immersion lens. A mounted *Luminera Infinity-2* digital camera was used to capture the image.

100 malaria images obtained from eight malaria slides of *P. vivax* sample were analysed. The images consist of the malaria parasites in ring, mature trophozoite and gametocyte stages. Figures 1(a)-(c) show examples of malaria images, while Figure 1(d) shows example of blood image with appearance of a WBC and platelet. As seen in these images, a malaria image comprises of malaria parasite, normal RBCs and background regions. From image (a), it can be seen that the normal RBCs are uniformly distributed, while the normal RBCs inside images (b), (c) and (d) are overlapped.



(a) Ring (b) Trophozoite



(c) Gametocyte (d) WBC and platelet

Fig. 1. Samples of the captured malaria images

B. The Proposed Enhanced k-Means Clustering Algorithm

KM clustering algorithm is based on concept of data assignation to their respective centre by the shortest Euclidean distance. Though KM algorithm is a very popular technique, its performance is still inadequate because of several shortcomings as described previously which may result on poor segmentation in the image. In the proposed EKM algorithm, 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 so that a good segmented image can be generated. Here, the variance calculation is applied for measuring the relationship between the centre and its members. During the clustering process, the variance between the members with their centre is constantly checked. If the centre does not meet a specified criterion, the data in that centre will be assigned to its proper centre based on the proposed data transferring process.

In order to describe the proposed EKM clustering algorithm, 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) ($x = 1, 2, \ldots, X$, $y = 1, 2, \ldots, Y$ and $j = 1, 2, \ldots, n_c$). The EKM algorithm that will be used for image segmentation is composed of the following steps:

1. Initialize the centres using:

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

   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 α_0 and α_b are small constant values.
3. Calculate the Euclidean distance, d using:

$$d = \left\| p(x,y) - c_j \right\| \tag{2}$$

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

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

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

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

   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 with the smallest and 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 closest 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 closest 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 closest 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. If $c_n < c_l$, assign the members of c_l to the closest centre if $p(x,y) < c_l$, where $x,y \in c_l$, and leave the rest of the members to c_l.

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

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

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

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 α_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 closest centre and recalculate the centre positions using Eq. 3.
10. Update α_a and α_b according to $\alpha_a = \alpha_0$ and $\alpha_b = \alpha_b - \alpha_b / n_c$ respectively, and repeat steps 5 to 9 until $\sigma^2(c_s) \geq \alpha_b \sigma^2(c_l)$.

For EKM algorithm, small constant value of $\alpha_0 = 0.05$ and $\alpha_b = 0.05$ are required for clustering process. To control the assignation of data to the proper centre in EKM algorithm, the variance is used in tandem with the Euclidean distance calculation. Therefore, after all data are assigned to the closest centre based upon Euclidean distance and the new position for each centre is computed using Eq. 3, the variance for each cluster is then calculated by using Eq. 4. The higher the cluster variance, the weaker the relationship between the cluster centre and its members. Therefore, the centres are sorted in ascending order, in order to find the centre with the largest value of variance between the members with their centre.

The proposed EKM algorithm has introduced a procedure of transferring process for clustered members based on certain condition. During the transferring process, the members of c_l which fulfil either the conditions of $c_n < c_l$ and $p(x,y) < c_l$, or the conditions of $c_n > c_l$ and $p(x,y) > c_l$, are assigning to the closest centre. On the other hand, the rest of c_l members will be kept in c_l. The transferring process of the c_l members could minimize the cluster variance by assigning up farther members to the appropriate cluster centre to produce good clustering.

C. Segmentation of Malaria Parasites

This study uses the same segmentation procedure as presented by Abdul-Nasir *et al.*, [10], in order to obtain the segmented malaria parasite. In Abdul-Nasir *et al.*, [10], the segmentation procedure was applied to compare four different colour components which are intensity, R-Y, B-Y and the proposed modified B-Y component in terms of providing good result in segmenting the malaria parasite. In this study, the segmentation procedure is applied to compare the performance of the proposed EKM clustering with KM, FCM and MKM clustering algorithms in segmenting the malaria parasite. A summary of the procedures for segmentation of malaria parasite are illustrated in Figure 2.

At first, the malaria image is enhanced by using modified global contrast stretching (MGCS) technique [22]. Then, the malaria image will be segmented by using clustering algorithm. This segmentation process is performed using the intensity colour component that is extracted from HSI (hue, saturation, intensity) colour models. In [10], cascaded moving *k*-means (MKM) and fuzzy *c*-means (FCM) clustering algorithm was used to segment the malaria image into parasite, normal RBCs and background regions. In this study, the cascaded moving *k*-means (MKM) and fuzzy *c*-means (FCM) clustering is substituted with four different clustering algorithms which are KM, FCM, MKM and the proposed EKM clustering algorithms. Comparisons among these four clustering algorithms have been made for measuring the performance of each clustering algorithm in image segmentation. During performing the clustering process, the centre initialization for the four clustering algorithms is standardized by using Eq. 1. After the segmented image has been obtained, 7×7 pixels median filter will be applied on the resultant image for smoothing the region of the segmented image. Finally, seeded region growing area extraction (SRGAE) algorithm is used to remove the artefacts, platelets and RBCs regions in the image.

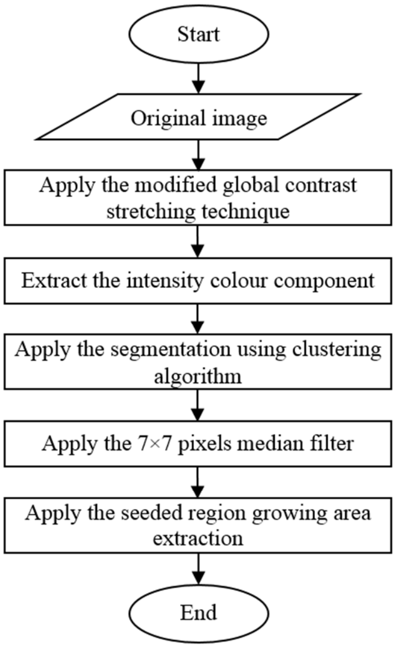


Fig. 2. The procedures for segmentation of malaria parasite

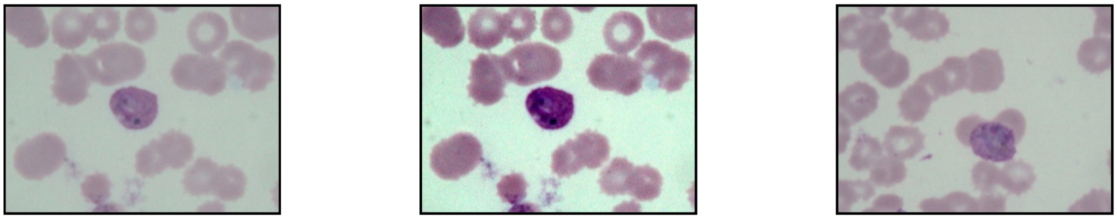
3. Results and Discussions

The unsupervised image segmentation approach using KM, FCM, MKM and the proposed EKM clustering algorithms has been performed on 100 malaria images. In this study, the primary criterion for assessing the performance of each clustering algorithm is based on its ability to produce fully segmented malaria parasite region as well as segmenting the malaria parasite from its blood cells background. The qualities of segmented images have been determined qualitatively and quantitatively.

A. Qualitative Analysis

Three original malaria images which are ring, trophozoite and gametocyte images are shown in Figures 3(a), (c) and (e), respectively. These malaria images are blurred with very low contrast. To overcome this problem, MGCS technique has been applied and the enhanced images can be seen in Figures 3(b), (d) and (f).

Then, KM, FCM, MKM and EKM algorithms have been used to segment the malaria image into malaria parasite, RBCs and background regions. In this study, segmentation using different number of clusters (segmented regions) has been performed to analyze the significance of each clustering algorithm for image segmentation. Figures 4 to 6 show the results of the four clustering algorithms. The images on the first column represent the intensity image. Meanwhile, the images on the second, third and fourth columns represent the results of KM, FCM, MKM and EKM algorithms with the number of clusters equal to 3, 6 and 9, respectively.


(a) Original Ring image (b) Enhanced Ring image (c) Original Trophozoite image

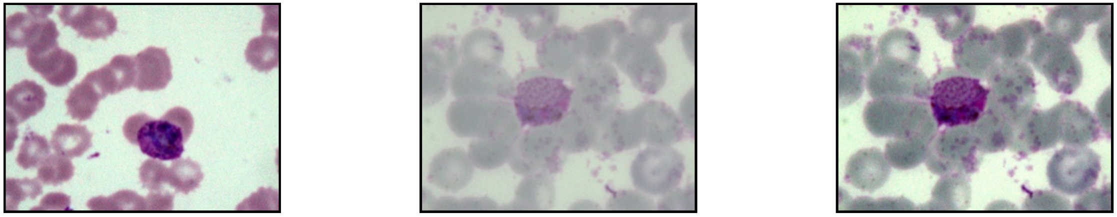
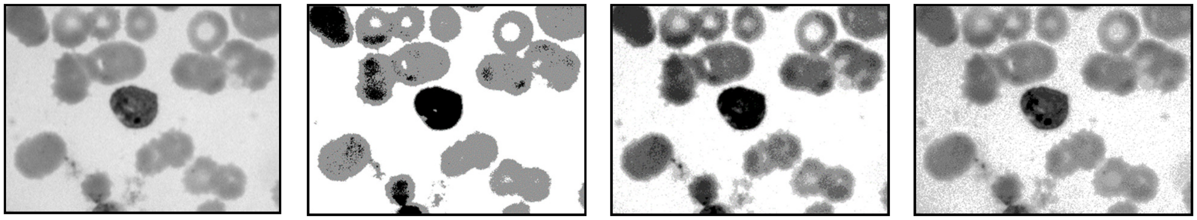
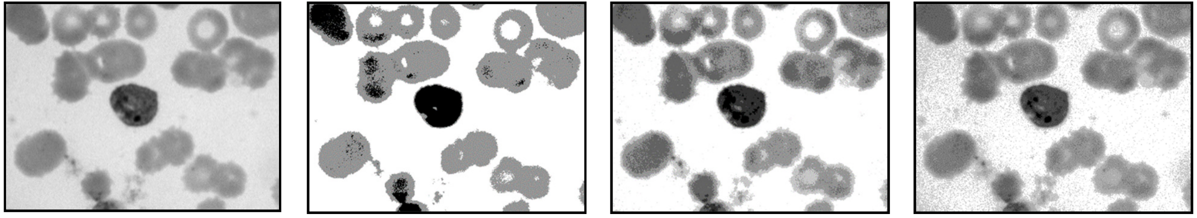
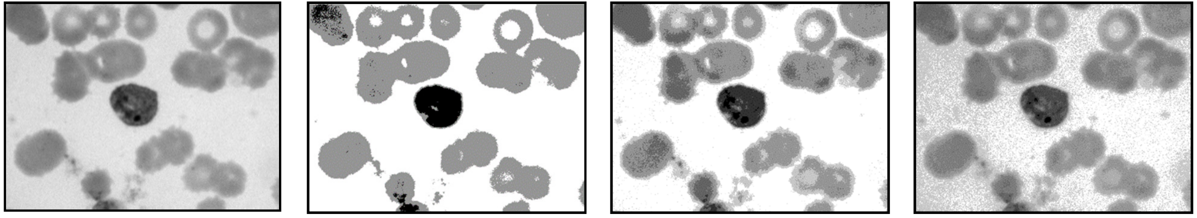

(d) Enhanced Trophozoite image (e) Original Gametocyte image (f) Enhanced Gametocyte image

Fig. 3. Original malaria images and results of images after applying the MGCS technique


(a) Intensity (b) KM 3 clusters (c) KM 6 clusters (d) KM 9 clusters


(e) Intensity (f) FCM 3 clusters (g) FCM 6 clusters (h) FCM 9 clusters


(i) Intensity (j) MKM 3 clusters (k) MKM 6 clusters (l) MKM 9 clusters

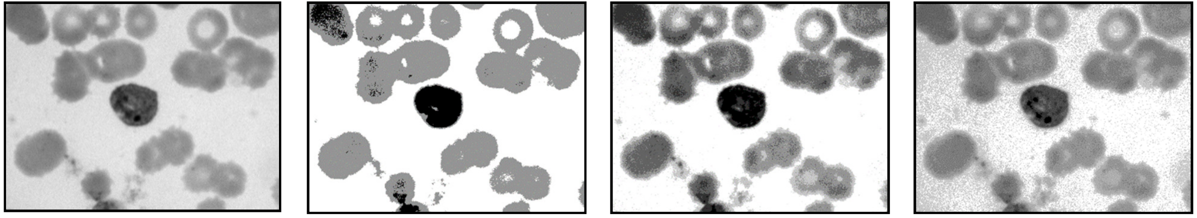
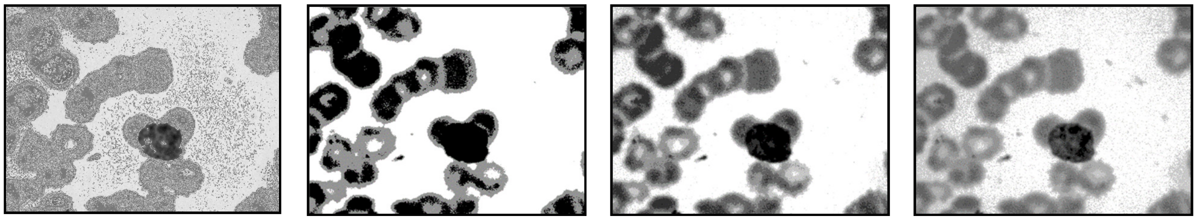
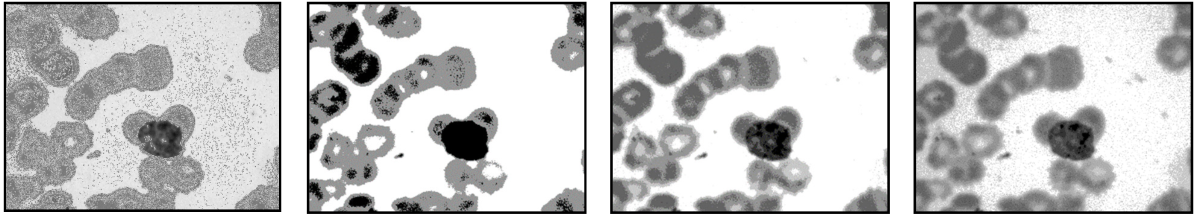
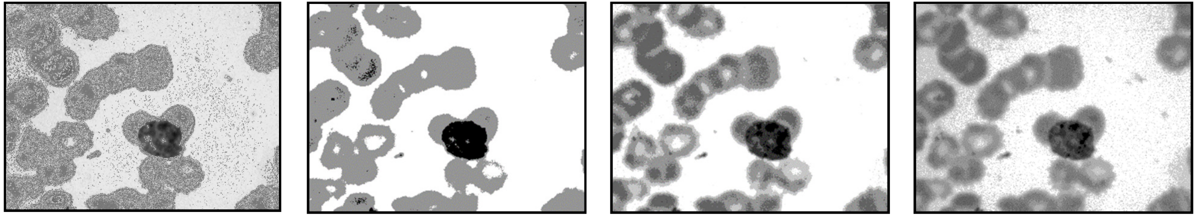

(m) Intensity (n) EKM 3 clusters (o) EKM 6 clusters (p) EKM 9 clusters

Fig. 4. Results for ring image after applying the four clustering algorithms

By referring to the results of segmentation using 3 clusters as demonstrated in Figures 5(b) and (f), as well as Figures 6(b) and (f), both KM and FCM algorithms are unable to segment the malaria parasite properly, as the malaria parasite and some of the RBCs components have been clustered into a single region. As for MKM and EKM algorithms, the results clearly show that both clustering algorithms have successfully segmented the malaria parasite from the RBCs and background regions in the three malaria images compared with the results given by KM and FCM algorithms. As the number of clusters increase to 6, the segmented regions provided by the four clustering algorithms are different for each algorithm. However, as the number of clusters increase to 9, the proposed and the other clustering algorithms have produced almost similar results especially on the malaria parasite region. Based on different clusters representation for the malaria parasite region among the four clustering algorithms, a proper selection for the number of clusters is also important to segment the malaria parasite properly.


(a) Intensity (b) KM 3 clusters (c) KM 6 clusters (d) KM 9 clusters


(e) Intensity (f) FCM 3 clusters (g) FCM 6 clusters (h) FCM 9 clusters


(i) Intensity (j) MKM 3 clusters (k) MKM 6 clusters (l) MKM 9 clusters

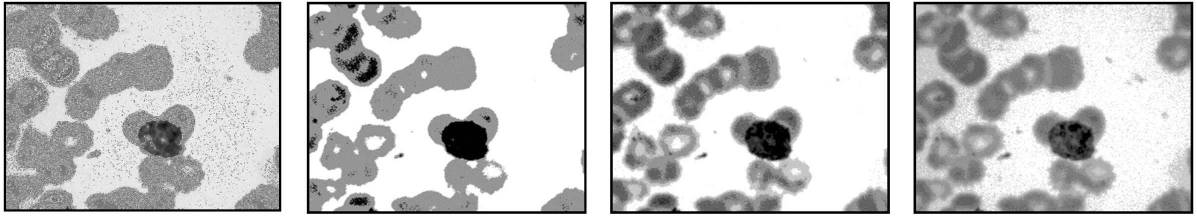

(m) Intensity (n) EKM 3 clusters (o) EKM 6 clusters (p) EKM 9 clusters

Fig. 5. Results for trophozoite image after applying the four clustering algorithms

Afterwards, median filter has been applied to smoothen the area of the segmented malaria parasite. Regions outside the ranges of 5000 to 30000 pixels of the images are unwanted. Accordingly, SRGAE algorithm has been used to remove these unwanted regions. Figures 7 to 9 show the final segmented images after the filtering and region growing process. However, the RBCs still appeared on the image due to its similarity in size to the malaria parasite.

Based on the final segmented images of 3 clusters for the three malaria images, segmentation using MKM and EKM algorithms has given better results compared with the results given by KM and FCM algorithms. Both clustering algorithms have produced a clean segmented malaria image without any appearance of RBCs region. As for the results of KM and FCM, both clustering algorithms are unable to segment the malaria parasite from the RBCs region properly. The problem occurs because some pixels in malaria parasite region have the similar intensity values with the pixels in RBCs especially in the case of overlapping RBCs. As a result, an over-segmented image has been produced as several segmented RBCs are appeared inside the segmented image. Therefore, these clustering algorithms are unable to segment the malaria parasite properly, especially for overlapping RBCs. By comparing the results between MKM and EKM algorithms, as seen in Figures 8 and 9, the EKM algorithm shows a better segmentation result by segmenting the malaria parasite region homogeneously as compared to MKM algorithm.

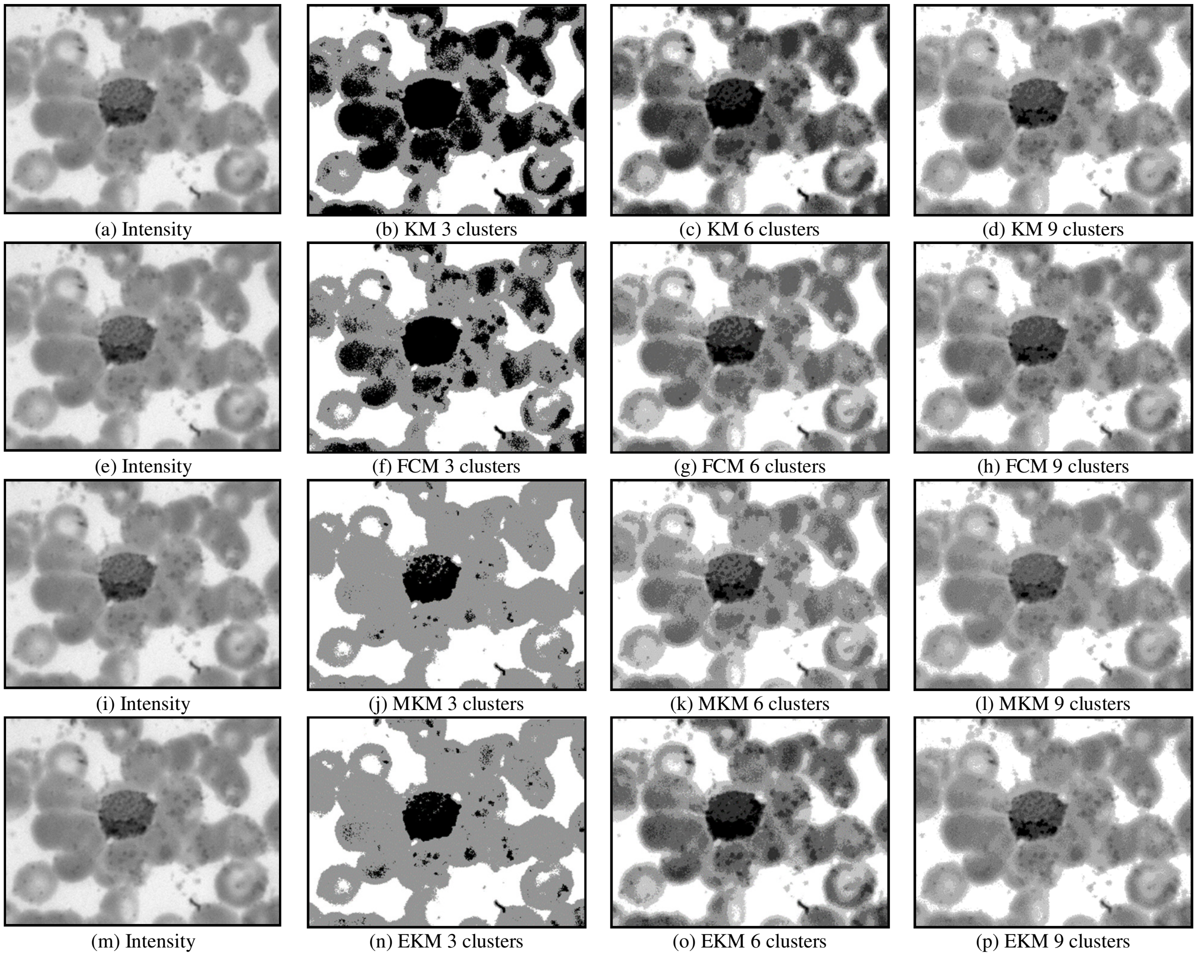
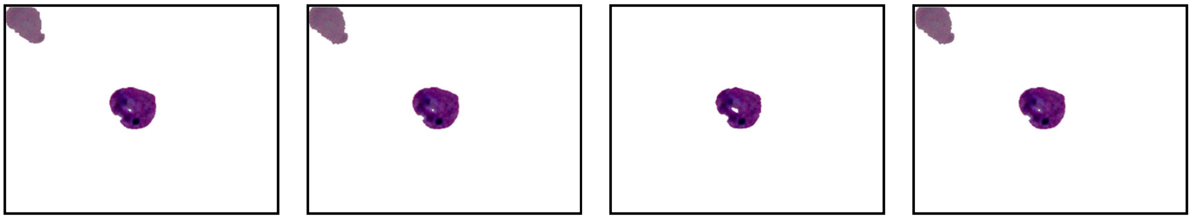
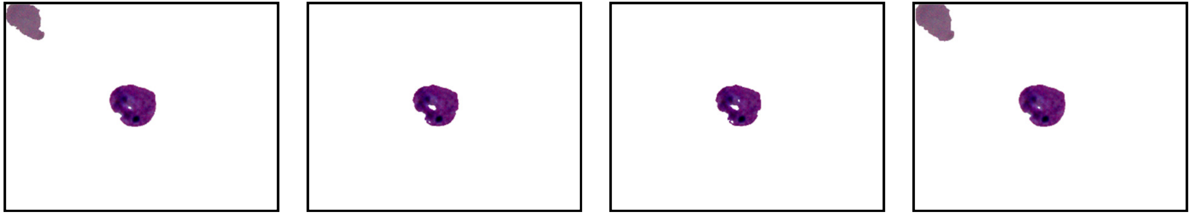

(a) Intensity (b) KM 3 clusters (c) KM 6 clusters (d) KM 9 clusters
(e) Intensity (f) FCM 3 clusters (g) FCM 6 clusters (h) FCM 9 clusters
(i) Intensity (j) MKM 3 clusters (k) MKM 6 clusters (l) MKM 9 clusters
(m) Intensity (n) EKM 3 clusters (o) EKM 6 clusters (p) EKM 9 clusters

Fig. 6. Results for gametocyte image after applying the four clustering algorithms

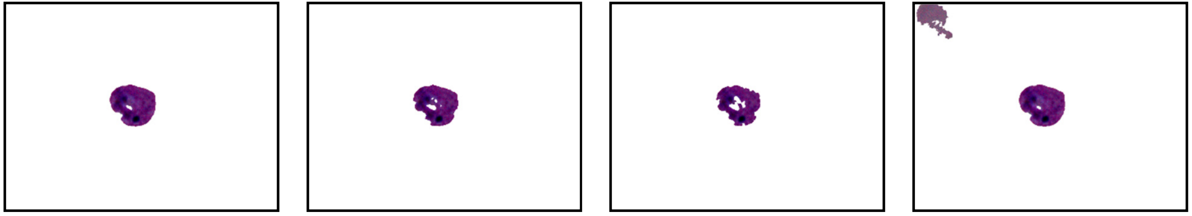
As the number of clusters increase to 6, the EKM algorithm is still able to produce good segmentation result as compared to the other clustering algorithms. EKM algorithm has produced a proper segmented malaria parasite and clean segmented malaria image on trophozoite and gametocyte images. As for FCM and MKM algorithms, segmentation using both algorithms has produced under-segmented images. As the number of clusters increase to 9, there is a tendency for the RBCs region to be clustered as a part of malaria parasite region. Based on the resultant images of segmentation using 9 clusters (read 3 regions) for the four clustering algorithms, this selection of image segmentation has produced under-segmented images for all malaria images. Meanwhile, segmentation using 9 clusters (read 4 regions) has produced over-segmented images in ring and trophozoite images.



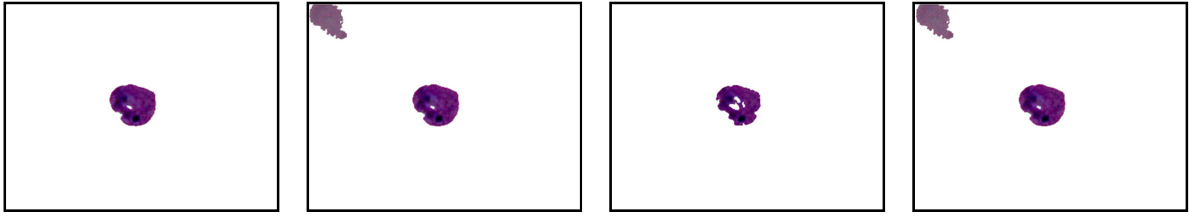
(a) KM 3 clusters (Read 1 region)
(b) KM 6 clusters (Read 2 regions)
(c) KM 9 clusters (Read 3 regions)
(d) KM 9 clusters (Read 4 regions)



(e) FCM 3 clusters (Read 1 region)
(f) FCM 6 clusters (Read 2 regions)
(g) FCM 9 clusters (Read 3 regions)
(h) FCM 9 clusters (Read 4 regions)



(i) MKM 3 clusters (Read 1 region)
(j) MKM 6 clusters (Read 2 regions)
(k) MKM 9 clusters (Read 3 regions)
(l) MKM 9 clusters (Read 4 regions)



(m) EKM 3 clusters (Read 1 region)
(n) EKM 6 clusters (Read 2 regions)
(o) EKM 9 clusters (Read 3 regions)
(p) EKM 9 clusters (Read 4 regions)

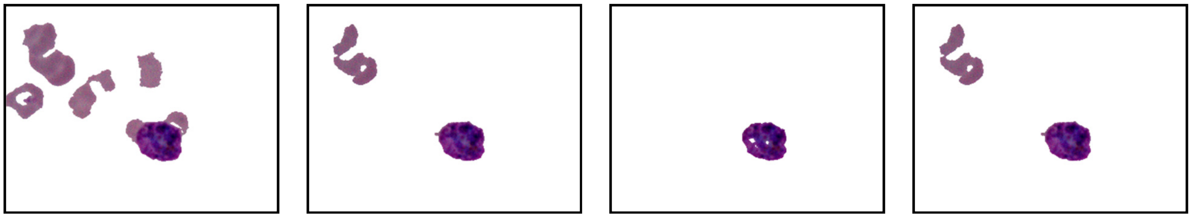
Fig. 7. Final segmented images of ring

Overall, the proposed EKM clustering algorithm has the capability to segment the malaria images with better segmentation result compared with the results given by KM, FCM and MKM clustering algorithms. The EKM algorithm has some advantages over the KM and FCM algorithms, especially when applying this algorithm on malaria image with overlapping RBCs. As can be seen in the final segmented malaria images, the EKM algorithm is able to segment the malaria parasite from the overlapped RBCs. In addition, the EKM algorithm is able to produce a smoother and fully segmented malaria parasite region as compared to MKM algorithm.

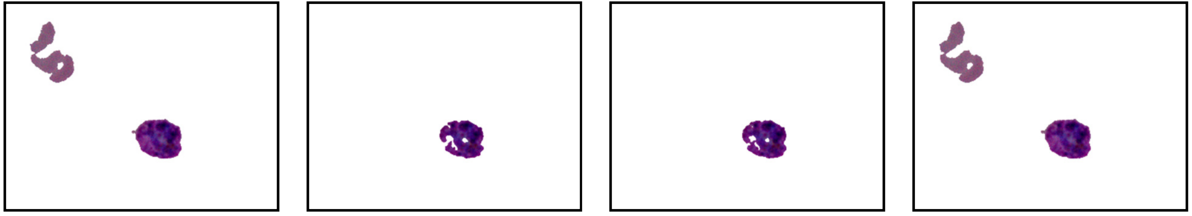
B. Quantitative Analysis

The final segmented images are compared to manually segmented images to quantify the segmentation performance for each algorithm. Higher pixel similarity indicates better segmentation performance. The results of sensitivity, specificity and accuracy of the three malaria images are presented in Tables 1, 2 and 3, respectively. From Table 1, segmentation using KM algorithm has produced the highest sensitivity results for almost all images, except for gametocyte image that has been segmented into 9 clusters. Even though high sensitivity values have been obtained, segmentation using KM algorithm has produced over-segmented images, especially when applying this algorithm on malaria images with overlapping RBCs as can be seen in Figures 8(a)-(b) and 9(a)-(b).

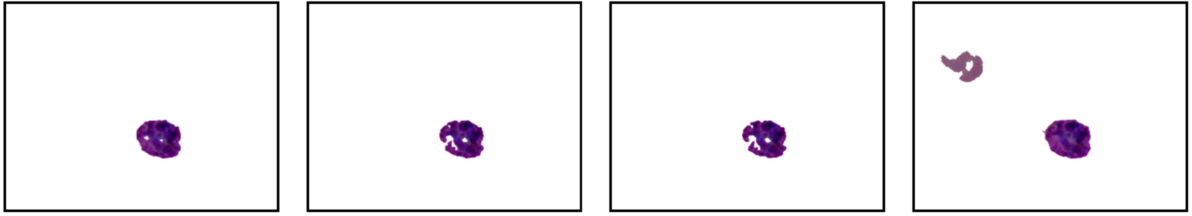
From Table 2, segmentation using MKM algorithm has produced the highest specificity results for almost all images, except for trophozoite image that has been segmented into 9 clusters (read 4 regions). However, segmentation using MKM algorithm has produced under-segmented images on the three malaria images that have been segmented into 6 and 9 clusters (read 3 regions). From Table 3, segmentation using EKM algorithm has produced the highest segmentation accuracy for most images that have been segmented into 3, 6 and 9 clusters (read 4 regions). Meanwhile, segmentation using KM algorithm has produced the highest segmentation accuracy for all images that have been segmented into 9 clusters (read 3 regions).



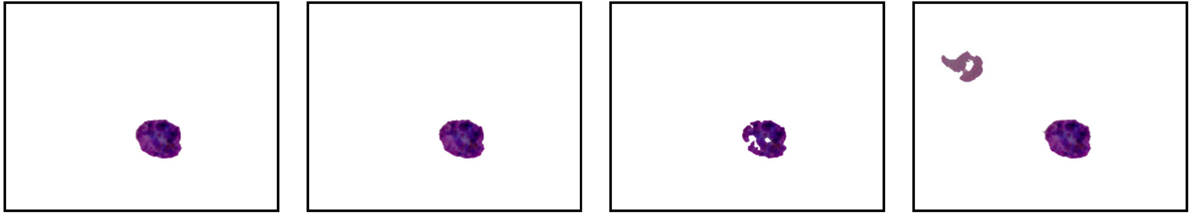
(a) KM 3 clusters (Read 1 region)
(b) KM 6 clusters (Read 2 regions)
(c) KM 9 clusters (Read 3 regions)
(d) KM 9 clusters (Read 4 regions)



(e) FCM 3 clusters (Read 1 region)
(f) FCM 6 clusters (Read 2 regions)
(g) FCM 9 clusters (Read 3 regions)
(h) FCM 9 clusters (Read 4 regions)



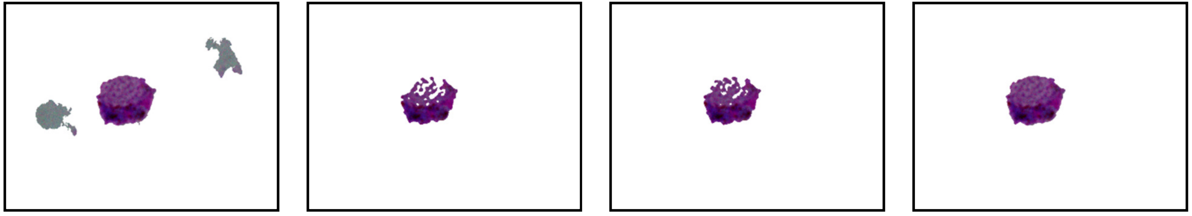
(i) MKM 3 clusters (Read 1 region)
(j) MKM 6 clusters (Read 2 regions)
(k) MKM 9 clusters (Read 3 regions)
(l) MKM 9 clusters (Read 4 regions)



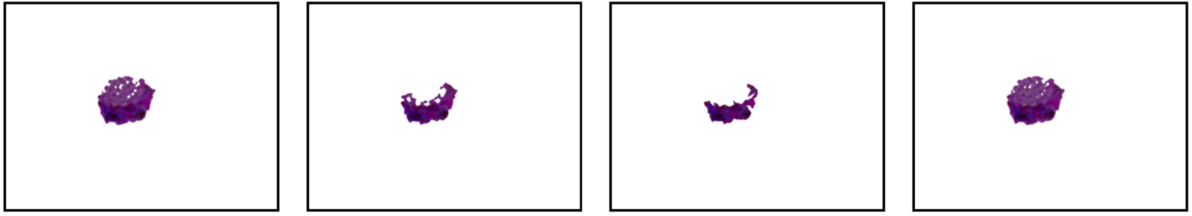
(m) EKM 3 clusters (Read 1 region)
(n) EKM 6 clusters (Read 2 regions)
(o) EKM 9 clusters (Read 3 regions)
(p) EKM 9 clusters (Read 4 regions)

Fig. 8. Final segmented images of trophozoite

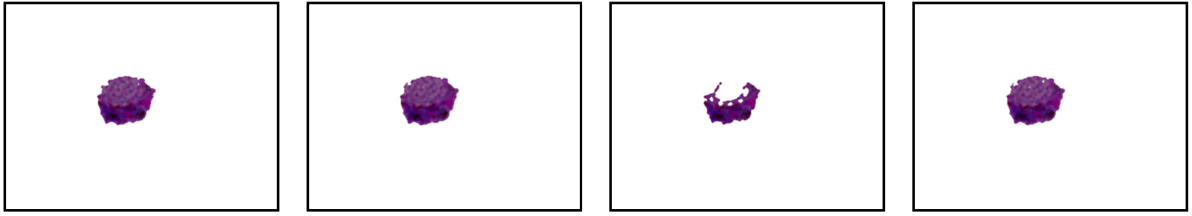
(a) KM 3 clusters (Read 1 region) (b) KM 6 clusters (Read 2 regions) (c) KM 9 clusters (Read 3 regions) (d) KM 9 clusters (Read 4 regions)



(e) FCM 3 clusters (Read 1 region) (f) FCM 6 clusters (Read 2 regions) (g) FCM 9 clusters (Read 3 regions) (h) FCM 9 clusters (Read 4 regions)



(i) MKM 3 clusters (Read 1 region) (j) MKM 6 clusters (Read 2 regions) (k) MKM 9 clusters (Read 3 regions) (l) MKM 9 clusters (Read 4 regions)



(m) EKM 3 clusters (Read 1 region) (n) EKM 6 clusters (Read 2 regions) (o) EKM 9 clusters (Read 3 regions) (p) EKM 9 clusters (Read 4 regions)

Fig. 9. Final segmented images of gametocyte

Table 1
Sensitivity for the Segmented Images

Number of Clusters	Image	Sensitivity for Clustering Algorithm (%)			
		KM	FCM	MKM	EKM
3 clusters (Read 1 region)	Ring	**92.90**	92.20	88.72	90.14
	Trophozoite	**98.88**	97.23	92.12	95.42
	Gametocyte	**96.66**	93.48	81.67	87.69
6 clusters (Read 2 regions)	Ring	**93.53**	85.09	80.29	91.47
	Trophozoite	**97.31**	81.67	82.99	94.56
	Gametocyte	**93.07**	67.66	47.81	89.74
9 clusters (Read 3 regions)	Ring	**86.27**	82.94	71.07	74.85
	Trophozoite	**90.46**	86.17	80.78	81.67
	Gametocyte	73.94	**75.27**	35.39	51.03
9 clusters (Read 4 regions)	Ring	**93.53**	93.05	90.56	91.47
	Trophozoite	**97.31**	**97.31**	96.51	96.38
	Gametocyte	91.25	**92.14**	84.26	88.40

Table 2
Specificity for the Segmented Images

Number of Clusters	Image	Specificity for Clustering Algorithm (%)			
		KM	FCM	MKM	EKM
3 clusters (Read 1 region)	Ring	98.15	98.28	**100**	**100**
	Trophozoite	90.94	97.68	**99.99**	99.97
	Gametocyte	77.13	97.21	**100**	**100**
6 clusters (Read 2 regions)	Ring	98.06	**100**	**100**	98.43
	Trophozoite	97.60	**99.99**	**99.99**	99.98
	Gametocyte	97.56	**100**	**100**	99.99
9 clusters (Read 3 regions)	Ring	**100**	**100**	**100**	**100**
	Trophozoite	**100**	**100**	**100**	**100**
	Gametocyte	**100**	**100**	**100**	**100**
9 clusters (Read 4 regions)	Ring	98.06	98.11	**98.80**	98.43
	Trophozoite	97.60	97.60	98.67	**98.73**
	Gametocyte	99.99	99.99	**100**	**100**

Table 3
Accuracy for the Segmented Images

Number of Clusters	Image	Accuracy for Clustering Algorithm (%)			
		KM	FCM	MKM	EKM
3 clusters (Read 1 region)	Ring	97.99	98.11	99.67	**99.72**
	Trophozoite	91.14	97.67	99.79	**99.85**
	Gametocyte	77.97	97.05	99.21	**99.47**
6 clusters (Read 2 regions)	Ring	97.93	**99.57**	99.43	98.23
	Trophozoite	97.59	99.52	99.55	**99.84**
	Gametocyte	97.37	98.60	97.75	**99.55**
9 clusters (Read 3 regions)	Ring	**99.60**	99.50	99.16	99.27
	Trophozoite	**99.75**	99.64	99.49	99.52
	Gametocyte	**98.87**	98.93	97.21	97.88
9 clusters (Read 4 regions)	Ring	97.93	97.97	**98.56**	98.23
	Trophozoite	97.59	97.59	98.61	**98.66**
	Gametocyte	99.50	99.61	99.32	**99.65**

Table 4 summarizes the average segmentation performances of 100 malaria images. Here, segmentation using KM algorithm shows good result in segmenting the malaria parasite region with the highest sensitivity of 89.52% obtained from the images that have been segmented into 9 clusters (read 4 regions). Segmentation using MKM algorithm shows good result in producing a clean segmented malaria image with the highest specificity of 99.94% obtained from the images that have been segmented into 9 clusters (read 3 regions). However, the specificity results for KM as well as the sensitivity results for MKM that have been obtained are low when compared to EKM algorithm. This is due to over-segmented and under-segmented images produce by KM and MKM algorithms, respectively. As for segmentation accuracy, segmentation using EKM algorithm shows good result in segmenting the entire area in malaria image with the highest accuracy of 98.93% obtained from the images that have been segmented into 3 clusters.

By considering primary criterion for assessing the performance of each clustering algorithm, the proposed EKM algorithm has produced better segmentation results. This is reflected by the good sensitivity, high specificity and high accuracy values compared to the segmentation results produced by KM, FCM and MKM algorithms. Segmentation using EKM algorithm does not give the best sensitivity result as similar to KM. However, it is still able to produce smoother and well defined form of the malaria parasite region.

Table 4
Average Segmentation Performances of 100 Malaria Images

Analysis	Clustering Algorithm	Segmentation Performance for Different Number of Clusters (%)			
		3 clusters (Read 1 region)	6 clusters (Read 2 regions)	9 clusters (Read 3 regions)	9 clusters (Read 4 regions)
Sensitivity	KM	**87.96**	**85.26**	**74.33**	**89.52**
	FCM	86.55	70.59	68.98	88.86
	MKM	78.47	59.31	47.92	81.58
	EKM	85.61	83.47	70.28	87.90
Specificity	KM	94.43	98.81	99.84	97.92
	FCM	97.18	99.85	99.87	98.28
	MKM	**99.64**	**99.91**	**99.94**	**99.11**
	EKM	99.13	99.40	99.87	98.59
Accuracy	KM	94.16	98.28	**98.85**	97.59
	FCM	96.71	97.72	98.66	98.10
	MKM	98.81	97.33	96.92	98.42
	EKM	**98.93**	**98.65**	98.71	**98.48**

4. Conclusions

This paper has proposed a clustering algorithm for segmentation of malaria images namely enhanced k-means clustering. The proposed EKM clustering algorithm has been tested on 100 malaria images quantitatively. Four clustering algorithms have been compared to quantify their performance in segmenting the malaria images. Qualitatively, the EKM algorithm has performed well in segmenting the malaria images with uniformly distributed and overlapping RBCs into malaria parasite, RBCs and background regions. Quantitatively, the experimental results demonstrate that EKM algorithm yields good sensitivity as well as high specificity and accuracy values compared with the segmentation results given by KM, FCM and MKM clustering algorithms. Here, the proposed EKM clustering algorithm has the capability to segment 100 malaria images into 3 clusters with average segmentation accuracy, sensitivity and specificity values of 98.93%, 85.61% and 99.13%, respectively. Thus, this study has shown that the proposed EKM clustering algorithm has a better potential in application of malaria image segmentation as compared to other clustering algorithms.

Acknowledgement

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] Azikiwe, C. C. A., C. C. Ifezulike, I. M. Siminialayi, L. U. Amazu, J. C. Enye, and O. E. Nwakwunite. "A comparative laboratory diagnosis of malaria: microscopy versus rapid diagnostic test kits." *Asian Pacific journal of tropical biomedicine* 2, no. 4 (2012): 307-310.

[3] Sio, Selena WS, Weiling Sun, Saravana Kumar, Wong Zeng Bin, Soon Shan Tan, Sim Heng Ong, Haruhisa Kikuchi, Yoshiteru Oshima, and Kevin SW Tan. "MalariaCount: an image analysis-based program for the accurate determination of parasitemia." *Journal of microbiological methods* 68, no. 1 (2007): 11-18.

[4] Somasekar, J., and B. Eswara Reddy. "Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging." *Computers & Electrical Engineering* 45 (2015): 336-351.

[5] Das, D. K., R. Mukherjee, and C. Chakraborty. "Computational microscopic imaging for malaria parasite detection: a systematic review." *Journal of microscopy* 260, no. 1 (2015): 1-19.

[6] Rao, KNR Mohana, and A. G. Dempster. "Modification on distance transform to avoid over-segmentation and under-segmentation." In *Video/Image Processing and Multimedia Communications 4th EURASIP-IEEE Region 8 International Symposium on VIPromCom*, pp. 295-301. IEEE, 2002.

[7] Das, Dev Kumar, Madhumala Ghosh, Mallika Pal, Asok K. Maiti, and Chandan Chakraborty. "Machine learning approach for automated screening of malaria parasite using light microscopic images." *Micron* 45 (2013): 97-106.

[8] Preedanan, Wongsakorn, Montri Phothisonothai, Wongwit Senavongse, and Suchada Tantisatirapong. "Automated detection of plasmodium falciparum from Giemsa-stained thin blood films." In *Knowledge and Smart Technology (KST), 2016 8th International Conference on*, pp. 215-218. IEEE, 2016.

[9] Muda, T. Zalizam T., and Rosalina Abdul Salam. "Comparative Analysis on Blood Cell Image Segmentation." In *2nd International Symposium on Computer, Communication, Control and Automation*, pp. 474-477. Atlantis Press, 2013.

[10] Abdul-Nasir, Aimi Salihah, Mohd Yusoff Mashor, and Zeehaida Mohamed. "Segmentation of Malaria Parasite Based on Stained Blood Cells Detection." In *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, vol. 24, pp. 43-55. Trans Tech Publications, 2015.

[11] Harun, Nor Hazlyna, Mohd Yusof Mashor, and Rosline Hassan. "Segmentation Technique for Acute Leukemia Blood Cells Images using Saturation Component and Moving K-Mean Clustering Procedures."

[12] Kaur, Ishmeet, and Lalit Mann Singh. *A method of disease detection and segmentation of retinal blood vessels using fuzzy c-means and neutrosophic approach*. Infinite Study, 2016.

[13] Kockara, Sinan, Mutlu Mete, Bernard Chen, and Kemal Aydin. "Analysis of density based and fuzzy c-means clustering methods on lesion border extraction in dermoscopy images." In *BMC bioinformatics*, vol. 11, no. 6, p. S26. BioMed Central, 2010.

[14] MacQueen, James. "Some methods for classification and analysis of multivariate observations." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, pp. 281-297. 1967.

[15] Khan, Shehroz S., and Amir Ahmad. "Cluster center initialization algorithm for K-means clustering." *Pattern recognition letters* 25, no. 11 (2004): 1293-1302.

[16] D. Li, J. Chen, and L. Qian, “A new k-means algorithm for data clustering, Int. J. Sensing,” Computing & Control, pp. 38–48, 2011.

[17] J Bezdek, James C. "Objective Function Clustering." In *Pattern recognition with fuzzy objective function algorithms*, pp. 43-93. Springer, Boston, MA, 1981.

[18] Hung, Ming-Chuan, and Don-Lin Yang. "An efficient fuzzy c-means clustering algorithm." In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pp. 225-232. IEEE, 2001.

[19] Mashor, Mohd Yusoff. "Hybrid training algorithm for RBF network." *International Journal of the computer, the Internet and Management* 8, no. 2 (2000): 50-65.

[20] Bagirov, Adil M. "Modified global k-means algorithm for minimum sum-of-squares clustering problems." *Pattern Recognition* 41, no. 10 (2008): 3192-3199.

[21] Yusoff, Intan Aidha, and Nor Ashidi Mat Isa. "Two-dimensional clustering algorithms for image segmentation." *WSEAS transactions on computers* 10, no. 10 (2011): 332-342.

[22] Abdul-Nasir, Aimi Salihah, Mohd Yusoff Mashor, and Zeehaida Mohamed. "Modified global and modified linear contrast stretching algorithms: New colour contrast enhancement techniques for microscopic analysis of malaria slide images." *Computational and mathematical methods in medicine* 2012 (2012).