

A Robust Segmentation for Malaria Parasite Detection of Thick Blood Smear Microscopic Images

Umi Salamah[#], Riyanarto Sarno^{*}, Agus Zainal Arifin^{*}, Anto Satriyo Nugroho⁺, Ismail Ekoprayitno Rozi^{\$},
Puji Budi Setia Asih^{\$}

[#]Informatics Department, Universitas Sebelas Maret, Surakarta, 57126, Indonesia
Email : umisalamah@staff.uns.ac.id

^{*}Informatics Department, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
Email : riyanarto@if.its.ac.id; agusza@cs.its.ac.id

⁺Center for Information and Communication Technology Agency for the Assessment and Application of Technology (PTIK-BPPT),
Jakarta, 10340, Indonesia
Email: anto.satriyo@bppt.go.id

^{\$}Eijkman Institute for Molecular Biology, Jakarta, 10340, Indonesia
Email: eko@eijkman.go.id; puji@eijkman.go.id

Abstract—Parasite Detection on thick blood smears is a critical step in Malaria diagnosis. Most of the thick blood smear microscopic images have the following characteristics: high noise, a similar intensity between background and foreground, and the presence of artifacts. This situation makes the detection process becomes complicated. In this paper, we proposed a robust segmentation technique for malaria parasite detection of microscopic images obtained from various endemic places in Indonesia. The proposed method includes pre-processing, blood component segmentation using intensity slicing and morphological operation, blood component classification utilising rule based on properties of parasite candidates, and parasite candidate formation. The performance was evaluated on 30 thick blood smear microscopic images. The experimental results showed that the proposed segmentation method was robust to the different condition of image and histogram. It reduced the misclassification error and relative foreground error by 2.6% and 45.5%, respectively. Properties addition to blood component classification increased the system precision. Average of precision, recall, and F-measure of the proposed method were all 86%. It is proven that the proposed method is appropriate to be used for malaria parasites detection.

Keywords— detection; intensity slicing; malaria parasites; morphological operation; thick blood smear.

I. INTRODUCTION

Malaria is a serious infectious disease caused by a peripheral blood parasite of the genus *Plasmodium* transmitted by the bite of an infected female *Anopheles* mosquito. According to World Malaria report 2014, from 198 million cases of malaria were reported in 2013, recorded 584,000 deaths [1]. With a high mortality rate, malaria should be treated as soon as possible. A rapid and accurate diagnosis will facilitate the treatment of the disease with appropriate drugs and help to control the spread of the disease.

A proper diagnosis of malaria can be obtained by conducting a series of tests on blood samples. According to World Health Organization, manual microscopy examination of blood smears is the gold standard in the diagnosis of malaria. It should be observed at least 100 microscopic fields

of view with a high magnification [2]–[5]. This process is exhausting for health workers and requires special skills. Therefore, the accuracy of the observations will be highly dependent on the experience and expertise of medical personnel.

Detection and identification of malaria parasites are a critical step in the automation process of malaria diagnosis. Many studies have been conducted it using thin blood smear succeeded in identifying the presence or absence of malaria parasite [6]–[12], even some of the research has reached identification of malaria parasite stage [13], [14] and Malaria detection and enumeration [15]–[17]. Technically, processing in a thin blood smear image is easier than a thick blood smear image. It is due to a thin blood smear preserves RBC shapes and parasites, so it is more suitable for species identification and specification. Meanwhile, a thick smear

preparation process destroys RBCs and thus makes the identification of species difficult [3], [4], [14]. However, detection sensitivity using thin blood smear is not as good as using thick blood smear. Examination on a thick blood smear is performed on a larger volume of blood. Hence, it is more sensitive to detect parasites. It is in line with standard practice manual diagnosis, positive or negative-type decisions is conducted using a thick blood smears, while identifying species and life-stage is undertaken using a thin blood smears [3], [14].

In the detection of malaria parasites, only a few studies that use a thick blood smear to the detection process. Those studies showed that parasite segmentation played an essential role in the process. Kaemkamnerd et al. [18] succeeded to detect the malaria parasites using thick blood smear. Segmentation was performed using an adaptive threshold on the V-Value histogram of the HSV image. The used images were good quality because they were generated by the image acquisition unit that was designed to be easily mounted on most conventional light microscopes and automatically controls the movement of the microscope stage in 3-directional planes. It makes it easier to separate the foreground image component histogram from the background components. Elter et al. [19] detected the malaria parasites in the case of a low parasite density (fewer than six parasites per image). Segmentation was conducted using the black-top-hat morphological operator on comparisons of green and blue components of the smear. It used an initial of 174 features which were pruned to become a subset of 60 ones to distinguish the parasite from other blood components. Arco et al. [20] performed automatic enumeration of malaria parasites in thick blood smears. This study was success calculates the number of parasites with high accuracy compared to manual calculation. Segmentation was performed using adaptive thresholding on the gray level image. Components of thick blood smear were assumed only consist of parasites and white blood nucleus, so the other blood components, namely platelet and artifact considered absent. The used feature was the area of the blood component. A parasite core is easily distinguished from the white blood core because of its much different size, but it is difficult to distinguish it with platelet and especially with the artifact that has similar size.

In the studies above, they were obtained high segmentation accuracy because the used images have high quality too. Unfortunately, they did not publish the test results using a low quality of thick blood smear. It has following characteristics: high noise, foreground intensity which is similar to the background, and the possibility of the presence of artifacts. Also, the use of a large number of features is suitable to be applied to smear with low parasite density, but it will take a lot of time for the detection process in a smear with high parasite density. Hence, the results cannot be used as a basis for determining whether an object is a parasite or not. We need a detection method that has segmentation technique to handle the high noise and low difference intensity between background and foreground.

In this paper, we proposed a robust segmentation technique for detection of the malaria parasite from microscopic images obtained from various endemic places in Indonesia. This technique is expected to overcome enormous

noise on the detection process of thick blood smear. The result of this study will be used to support the diagnosis of malaria in Indonesia, especially in eastern Indonesia that has a high prevalence of the disease.

II. MATERIALS AND METHOD

This chapter describes the related work, the materials, and the proposed methods of this paper.

A. Related Work

Arco et al. [20] detected parasite on thick blood smears using a method that consists of four main parts: pre-processing, segmentation, morphological operations, and connected component analysis. In the pre-processing stage, Gaussian low-pass filter, adaptive histogram equalisation, and h-minima transform were applied sequentially to get a clearer image. Segmentation stage utilised local characteristics of the image called adaptive thresholding where the segmented image was produced from two processes: calculated the average image using convolved with a mean filter (15×15 mask) and compared the average image with the value of each pixel in the input image. A pixel in the original image was considered as the background if the value was higher than T% of the pixels in the average image. Conversely, it was considered an object. In the morphological operation stage, dilation, erosion, and holes filling were applied to extract the desired components of an image. In the connected component analysis stage, there were labeling process of a formed connected components and measuring the size. A connected component which had smaller size than the average size of white blood nucleus was a parasites nucleus.

B. Materials

Microscopic images of the blood smears were prepared in Eijkman Institute for Molecular Biology, Indonesia. There were 30 microscopic images of thick blood smear from malaria patient on ring or trophozoite stage. In this study, the images were generated using 1.2-megapixel resolution with 10×100 magnifications. The size of images were 1280×960 pixels in the JPEG format. An example of data is presented in Fig. 1.

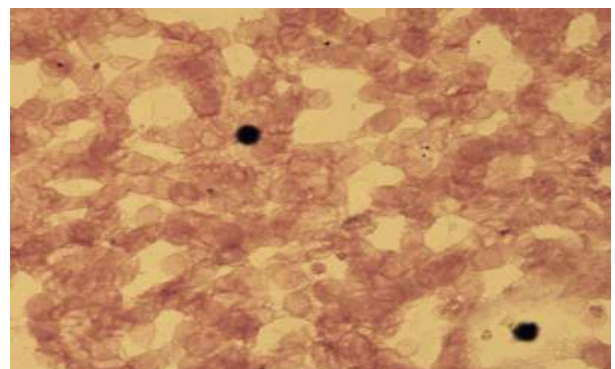


Fig. 1 A specimen of thick blood smear microscopic image

C. Method

The proposed method includes pre-processing, blood component segmentation, blood component classification, and parasite candidate formation (Fig. 2).

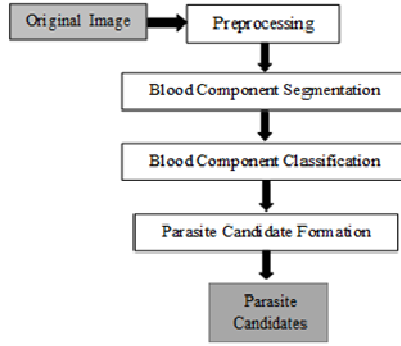


Fig. 2 Schematic diagram of malaria parasite detection method

1) *Pre-processing*: A digital image is a two-dimensional array or a matrix whose elements are expressed the intensity level of picture elements or pixels. Pixel intensities on low-contrast image gather in one area, so it has low-intensity range. Conversely, pixels intensities on a high-contrast image spread all over the area, so it has high-intensity range. On malaria detection, a high-contrast condition is indispensable because it can highlight the parasite from the background. Microscopic image of a thick blood smear used in this study is an image that has been converted into gray level image. Each smear had varying intensity, so the intensity range is also different. To facilitate the process of determining the threshold of the system, we need to equalize the intensity range. Therefore, the image is enhanced by contrast stretching to raise the contrast and equalize the intensity range of the used data. The form of contrast stretching [21] is defined as

$$I'_{(i,k)} = \frac{(\max - \min)}{(I_{\max} - I_{\min})} (I_{(i,k)} - I_{\min}) + \min \quad (1)$$

where $I'_{(i,k)}$ and $I_{(i,k)}$ is the intensity of the output and input pixels, I_{\max} and I_{\min} are the maximum and minimum intensity of input image, max and min are the maximum and minimum intensity that is desired respectively. This step purpose is to map the minimum of the array to 0 and the maximum of the array to 255. Moreover, it can decrease the effect of differences in illumination. The output of this phase is indicated by using image histogram before and after pre-processing (Fig. 3).

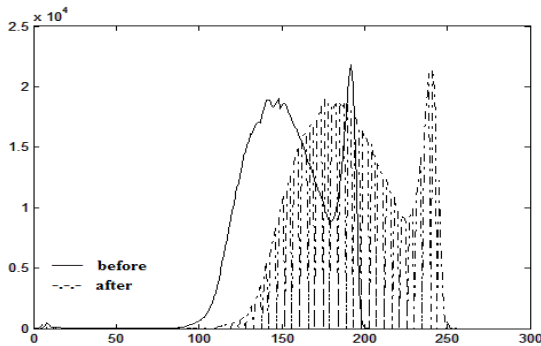


Fig. 3 Histogram image before and after pre-processing

2) *Blood Component Segmentation*: The first step is to highlight the location of parasite candidates. In the thick blood smear microscopic image, parasites are shown with an object whose intensity is darker than its surroundings.

Intensity slicing is carried out with a certain threshold, so objects with pixel intensity below the threshold can be obtained. This method is defined by,

$$I'_{(i,k)} = \begin{cases} I_{(i,k)}, & I_{(i,k)} \leq Th \\ 255, & I_{(i,k)} > Th \end{cases} \quad (2)$$

where $I'_{(i,k)}$ and $I_{(i,k)}$ are output and input pixel intensity, respectively, and Th is given threshold.

The result of intensity slicing transformation forms areas with a white background (Fig. 4). These areas are formed because the image is transformed using a certain threshold so that the pixels with intensity below the threshold value forming area or object, otherwise it will be the white background. With this process, objects becomes more obvious.

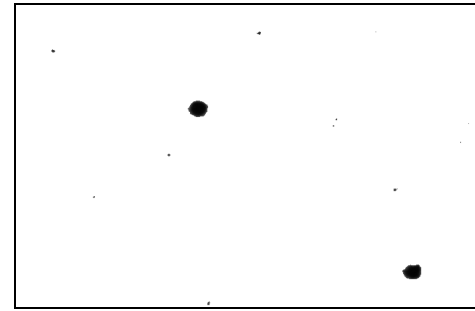


Fig. 4 Result of intensity slicing transformation

The second step, the initial segmentation was conducted using canny edge detection to detect the edge of formed contour. Further, it performed a morphological closing operation: dilation to clarify the contour obtained, holes filling to eliminate little holes and fills the gaps in the contour, and erosion to remove the noise resulting from thresholding and the canny edge detection steps. The form of dilation and erosion [20] respectively are given by

$$A \oplus B = \{z \mid (B)_z \cap A \subseteq A\} \quad (3)$$

$$A \ominus B = \{z \mid (B), z \subseteq A\} \quad (4)$$

where A is input image and B is a structuring element. The dilation of A and B is the set of all displacements, z , such that B and A overlap in at least one element. The erosion of A and B is the set of all the pixels z such that B has to be contained in A . In this study, the dilation uses two flat linear structuring element with 3 in length and angles are 90° and 0° . This composition of two flat structuring elements is conducted by dilating the scalar value 1 with both structuring elements in sequence. The erosion uses a flat diamond-shaped structuring element with r , distance from the structuring element origin to the points of the diamond, is 1.

The segmentation processes form connected components which are blood components in thick blood smear, i.e. white blood nucleus, parasite nucleus, platelet, and artifact (Fig. 5). It can be seen that the size of white blood nucleus are very distinct from the other blood components, while there is a partly parasite nucleus sized similarly to platelet and artifact. Therefore in the parasite detection process, classification is

necessary to distinguish the parasite with other blood components.

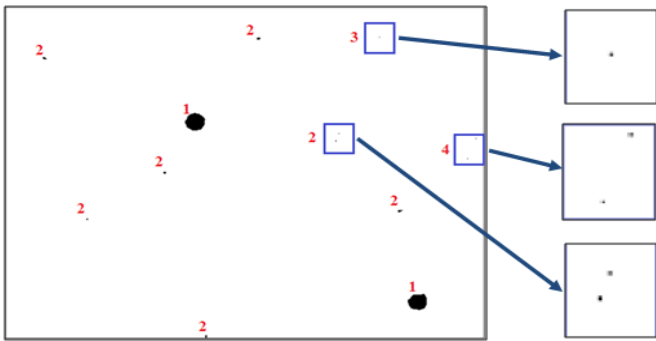


Fig. 5 Result of segmentation: 1. white blood nucleus, 2. parasite nucleus, 3. platelet, 4. artifact.

3) *Blood Components Classification*: A parasite nucleus is described as chromatin dot because normally the shape is round. In fact, the shape is not too round. Moreover, the thick blood smear preparation process often destroys red blood cells, so more precise shape is oval with a certain degree. As shown in Fig. 6, the nucleus area is characterised by red area.



Fig. 6 Thick film malaria parasite in trophozoite stage [4]

In several studies [18], [20], the classification of blood component are based only on its area. A malaria parasite would be easily distinguished by the white blood nucleus or platelet because its area is different. However, recognizing the parasites and artifacts are not easy because an artifact can be derived from mould, dirt that occurs in the process of manufacture/storage smears, and other bacteria that vary in size and may have an area equal to the parasite. So, some properties of the connected component that can distinguish a parasite with the artifact should be added.

Feature	Cor	Abs(Cor)
Solidity	0.33	0.33
Area	-0.33	0.33
MaxAxis	-0.30	0.30
MinAxis	-0.28	0.28
Ovalness	0.20	0.20
Max_I	-0.16	0.16
Min_I	-0.09	0.09
Eccent	-0.07	0.07
Mean_I	-0.05	0.05

a.

$\alpha = 0.4$

Solidity	Area	MaxAxis	MinAxis	Ovalness	Max_I	Min_I
No_Rem	No_Rem	No_Rem	No_Rem	No_Rem	No_Rem	No_Rem
	No_Rem	Rem	Rem	No_Rem	Rem	Rem
		No_Rem	Rem	No_Rem	No_Rem	Rem
		No_Rem	No_Rem	Rem	Rem	Rem
			No_Rem	No_Rem	No_Rem	No_Rem
				No_Rem	No_Rem	No_Rem

c.

Solidity	Area	MaxAxis	MinAxis	Ovalness	Max_I	Min_I
1	0.36	0.20	0.22	0.33	0.19	0.05
	1	0.96	0.97	0.11	0.42	0.64
		1	0.98	0.04	0.38	0.61
			1	0.04	0.42	0.66
				1	0.06	0.03
					1	0.20
						1

b.

Fig. 8 Selection features process: a. select significant features, b. correlation between input features, c. removing process of redundant feature with $\alpha=0.4$

The classification stage starts by labeling the connected component produced in the previous process and specify some properties that are Area, Maximum Axis Length (MaxAxis), Minimum Axis Length (MinAxis) as in Fig. 7, the proportion between the maximum and minimum axis length (Ovalness), the distance between the center of the ellipse and its maximum axis length (Eccentricity), the proportion of actual cell area to convex hull area (Solidity), Maximum of Intensity (Max_I), Minimum of Intensity (Min_I), and Mean of Intensity (Mean_I).



Fig. 7 A connected component and the properties

The feature selection process uses 280 connected components in the form of parasites and non-parasites derived from 15 samples. The class label of connected components is given by experts. Some studies used correlation measure for feature selection [22], [23]. In this paper, we use the algorithm from Jiang and Wang [23] as follows.

- Compute correlation between every input features with the class label.
- Select significant features using the t-test. Order features in descending correlation value and select its value $> t$.
- Compute the correlation between input features.
- Remove redundant features that the correlation value $> \alpha$. Selected features are features that have no remove status.

From Two-tail test of Pearson correlation table with a significant level of 0.05 for 280 data, the t value is 0.087. Selected features are features that have correlation values with class label greater than 0.087 (Fig. 8a). The next processes are calculate the correlation between the input features (Fig. 8a b) and remove the redundant feature by comparing it with α . In this paper, the used α values are 0.4, 0.5, and 0.6. One example of removal process was shown in Fig. 8c that uses $\alpha = 0.4$.

Based on Fig. 8c and removing process with $\alpha=0.5$ and $\alpha=0.6$, three features were selected, i.e. Area, Solidity, and Ovalness. These features have no remove status (Rem). Further, the classification of the blood components is based on the rule of properties of the connected component (Fig. 9).

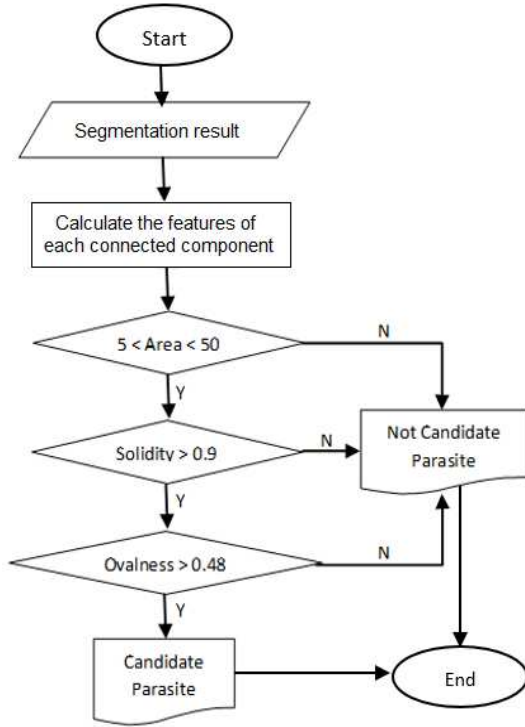


Fig. 9 Rule-based of the classification process

The value limits for selected features are based on data of connected component that parasite labelled. As consideration, the average size of a white blood nucleus in this study was 2260 pixels.

4) Parasite Candidate Formation

Malaria parasite in the thick blood smear is characterised by the presence of parasites nucleus followed by cytoplasm [3], [4]. The previous process only produces nucleus of parasites candidates with the position. All parts of a parasite candidate are obtained by crop thick blood smear in a small area of size 35×35 based on centroid of the connected components (Fig. 10).

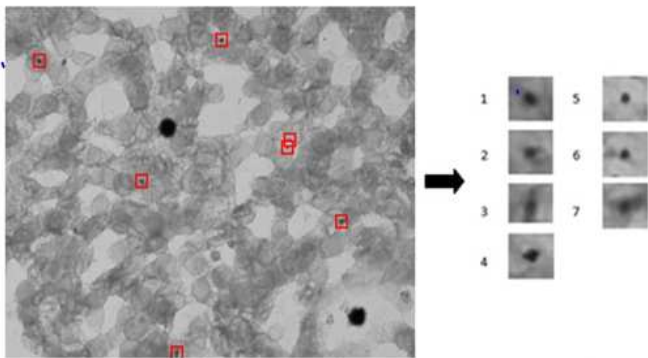


Fig. 10 Parasite candidate formation

This measure is obtained by taking four times of the average parasite nucleus diameter to the up, the down, the right, and the left of the parasite's nucleus. The addition of area around the nucleus aims to ensure the cytoplasm of parasite exists in the area to be cropped.

D. Performance Measurement

Performance measurement of detection is conducted by comparing the ground truth and the result of detection. Ground truth was created manually based on instructions that are given by experts, and the results have been validated by a microscopist. From the process described in the methodologies, the system only determines whether an object in the thick blood smear is a parasite candidate or not. The result is called true positive (TP) if the system predicts as parasites and ground truth is also expressed as a parasite, false positive (FP) if the system predicts as parasites were ground truth stated as not parasites, and false negatives (FN) if the system predicts as not parasites were ground truth expressed as parasites. Performance measurement utilising these values are recall (R) or true positive rate, precision (P) or positive predictive value, and F-measure (F). The form of the performances are given by

$$P = \frac{TP}{FP + TP} \quad (5)$$

$$R = \frac{TP}{FN + TP} \quad (6)$$

$$F = \frac{2RP}{R + P} \quad (7)$$

The performance measure of the segmentation uses misclassification error (ME) that calculates the proportion of object pixels that misclassified as background, and vice versa. Moreover, it also uses relative foreground area error (RAE) that calculates the proportion of the difference between the object's area in ground truth image and the segmentation result. Assume, F_T and F_R are the object pixels in the ground truth image and the segmentation result. B_T and B_R are the background pixels in the ground truth image and the segmentation result. A_T and A_R are area of the object in the ground truth image and segmentation result. ME and RAE can define as Equation (8) and Equation (9), respectively.

$$ME = 1 - \left(\frac{(|B_T \cap B_R| + |F_T \cap F_R|)}{|B_T| + |F_T|} \right) \quad (8)$$

$$RAE = \begin{cases} \frac{A_T - A_R}{A_T} & \text{if } A_R < A_T \\ \frac{A_R - A_T}{A_R} & \text{if } A_R \geq A_T \end{cases} \quad (9)$$

The range of ME and RAE values are [0, 1]. If the value of ME is 0, then the segmentation result is exactly the same as the ground truth image. If the value of RAE is 0, then the

object in ground truth image has same area with an object in the segmentation result.

III. RESULTS AND DISCUSSION

In this study, we used 30 samples of thick blood smear microscopic image. Fifteen samples were used to parameter estimation. The proposed method was evaluated on the rest of thick blood smear microscopic images. The experiment results were described in detail as follows.

Experiments were conducted using variations of threshold 55 to 75 at 15 thick blood smear microscopic images. Fig. 11a was a result of system detection on sample No. 8 with a threshold intensity slicing value of 57 that compared to the ground truth (Fig. 11b). The thick blood smears condition in the experimental in Fig. 11 is a case where a thick smear has similar intensity between the background and blood components. False positives occur due to the formation of a contour from the background that the intensity value is below the threshold (candidate 9th, 11th, and 15th) or the presence of artifacts (candidate 16th) that the area and features are matching to the parasite. False negatives occur because the intensity of parasite candidate is still above the threshold or because of the parasite candidates located around the background that have similar intensity.

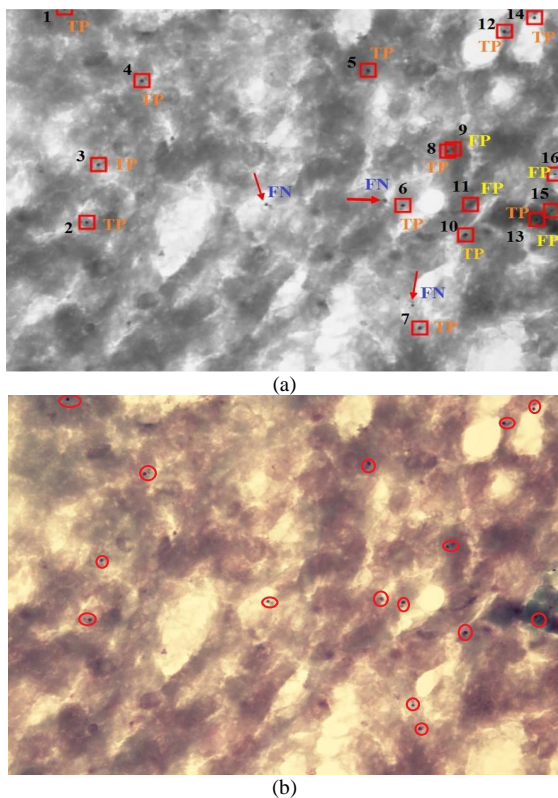


Fig. 11 Comparison of the system result and the ground truth: a. system result, b. groundtruth

The values of TP, FP, and FN in the experiment (Fig. 11a) are 12, 4, and 3, respectively. Using (5), (6), and (7), the value precision = 0.75 and recall = 0.8, so the F-measure is 0.774. The overall performance measures of sample no. 8 with threshold variation from 55 to 75 are presented in Fig. 12.

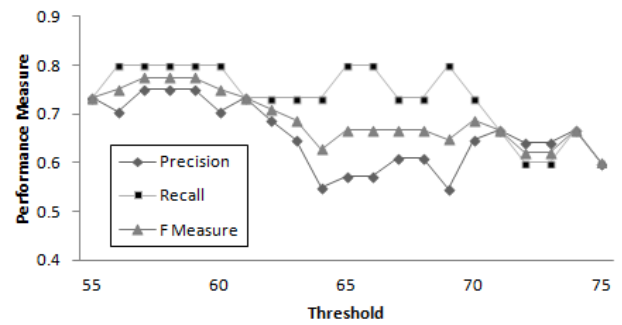


Fig. 12 Precision and Recall graph of sample No. 8

The determination threshold value of intensity slicing highly affects the value of precision and recall. Improper threshold determination could increase the error detection of parasite candidates. An optimal threshold value is required to obtain proper parasite (TP) as much as possible and error detection (FP) to a minimum. In other words, it achieves the highest F-measure. For example, the optimal threshold of sample No. 8 is obtained at 55, 56, and 57. The optimal threshold for other microscopic image samples is presented in Fig. 13.

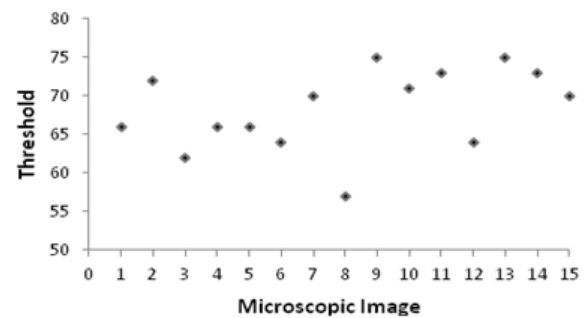


Fig. 13 Optimal threshold per microscopic image

It can be seen from Fig. 13 that the optimal threshold indicates the different values for each image microscopic samples and do not form a pattern. The use of each optimal threshold is not possible on a system. It is necessary to set a fixed threshold to provide the best results. Average F-measure per threshold recalculates to all testing data (Fig. 14). The best average F-measure value is met at given threshold 71 with a value of 0.86.

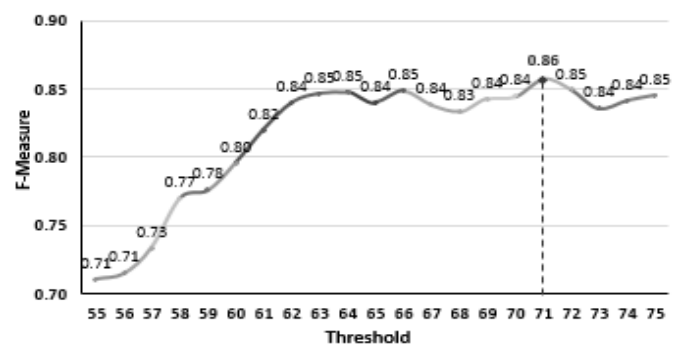


Fig. 14 Average F measure per Threshold

Evaluation is conducted by comparing the proposed method with Arco's method [20]. There are differences in assumptions related components on the thick blood smear of the two methods. Arco et al. determining thick blood components are white blood nucleus and parasites. While,

the proposed method uses whole blood components that are white blood nucleus, parasites, platelets, and the possibility of artifacts. So, this paper will compare the results of both methods just up the process of image segmentation as shown in Figure 15.

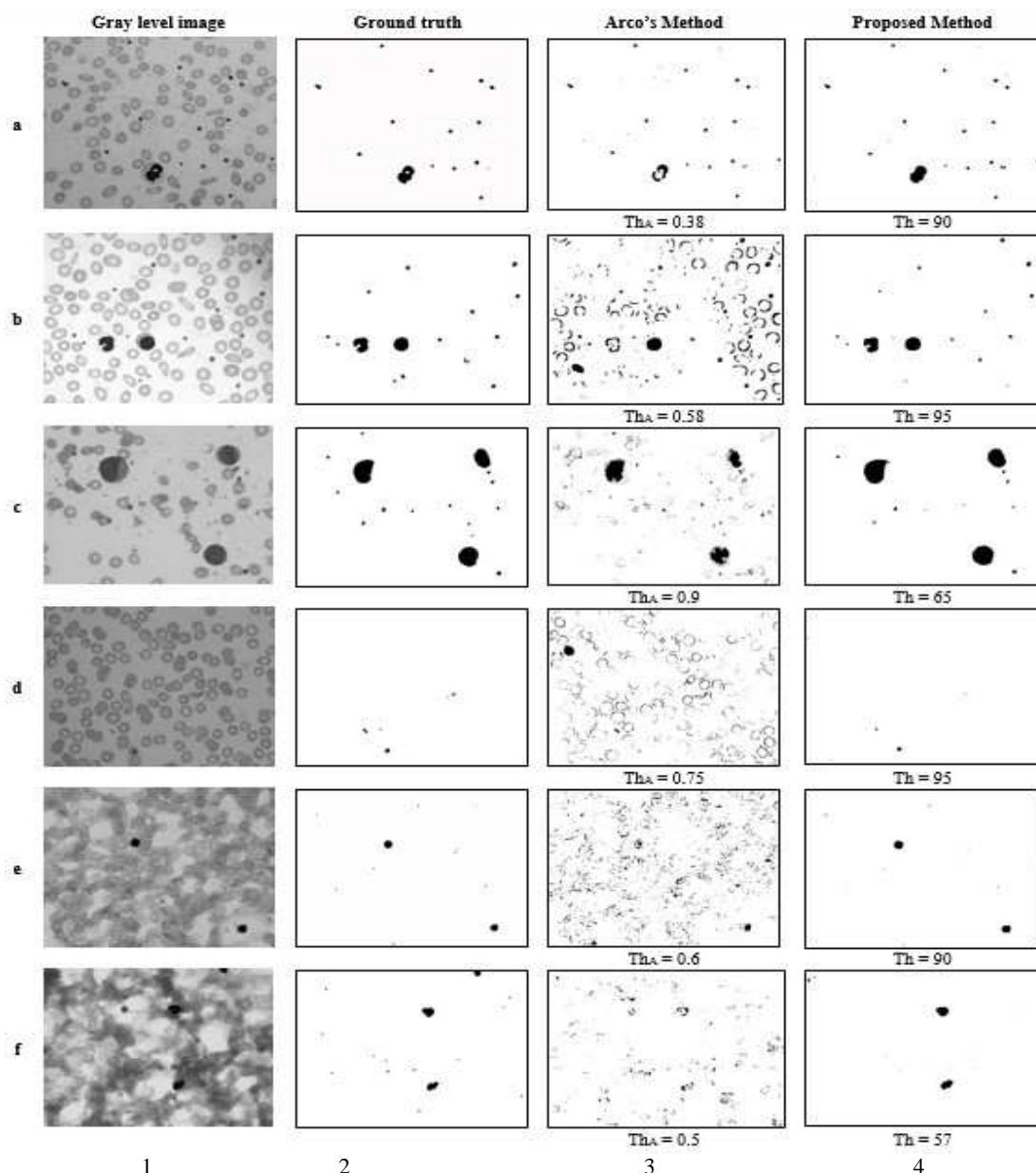


Fig. 15 Result of segmentation: 1. graylevel image, 2. ground truth, 3. segmentation result with Arco's Method, 4. segmentation result with the proposed method

The used data comparisons (Fig. 15.1) is obtained from the Muwardi Surakarta Hospital, Karanganyar District Hospital, and the Eijkman Institute for Molecular Biology. The data show several images of blood smear containing parasites or nucleated cells in several diseases. Fig. 15a–d are images of thin blood smear on iron deficiency diseases, Leukemia, and Malaria. Fig. 15e and 15f are images of thick blood smear on Malaria. The aim of the experiment is extracting the parasite nucleus and another nucleus of blood components from the smear image.

There are three conditions of smear image from Fig. 15.1. Firstly (Fig. 15a–b), components nucleus are visible than the background. Secondly (Fig. 15c–d), component nucleus are visible, but the intensity is similar to the surrounding intensity. The thirdly (Fig. 15e–f), component nucleus are camouflaged among the blood artifacts.

Results of the segmentation process of both methods are shown in Fig. 15.3 and Fig. 15.4. Threshold segmentation of the two methods is adapted to each condition image to obtain the best possible result that can extract blood components nucleus on the image. The threshold of Arco's Method and

proposed methods are Th_A and Th , respectively. Th_A is a percentage of the pixels in the average image with a mean filter (15×15 mask), while Th is intensity value from 0 to 255.

From Fig. 15.3 and Fig. 15.4 are seen that proposed method successfully extract the blood components nucleus of the smears image on all conditions (Fig. 15a-f). While, Arco's Method also successfully extract the blood components nucleus on the first condition (Fig 15a). On the second sample (Fig. 15b), there are many extracted non-nucleus object. In the second condition (Fig 15c-d), not all the nucleus are extracted, and many non-nucleus objects are extracted. In the third condition (Fig. 15e-f), objects that are non-nucleus are extracted more.

In the second condition image, the threshold required to extract the nucleus of Arco's method is high that are 0.9 and 0.75. It means that the intensity of the nucleus is not too different than the surrounding intensity with similarity 90% and 75%. The number of a non-nucleus object extracted can be reduced by decreasing the segmentation threshold, but it has resulted in extracted part of nucleus reduced too. If the segmentation threshold is increased, part of the nucleus will be more extracted but non-nucleus object more extracted too. Failure Arco's method in the third condition, mainly due to the use of adaptive histogram equalisation in pre-processing. In the first condition image, the process can increase the contrast of the image, so difference of the object from the background is more prominent. (Fig 16.1).

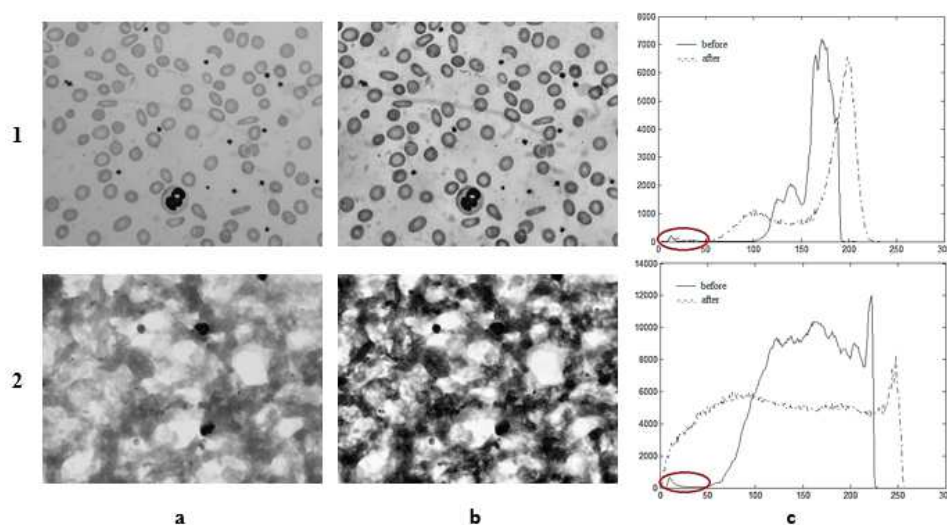


Fig. 16 Graylevel image before and after the process of adaptive equalisation histogram: a. before, b. after, c. histogram before and after the process

However, in the third condition image, the process makes many pixels move to low intensity, i.e. intensity corresponding to the blood components intensity (marked with red circles on the histogram in Fig 16.2.c). The images are darker than previous and many nucleuses are missing because they are covered by the blood artifacts, especially the small nucleus. (Fig. 16.2.b).

The result of the experiment in Fig. 15 shows that the proposed segmentation method is visually more robust with

different condition's image and histogram than Arco's method. The performance of these methods was measured using number of misclassification pixel (MP), ME, and RAE and it was summarised on Table I. The results indicated that the proposed method outperformed in all performance of all images compared Arco's result. The use of the proposed method reduce ME and RAE significantly. The average reduction of ME and RAE are 2.6% and 45.5%, respectively.

TABLE I
COMPARISON RESULT WITH ARCO'S METHOD

Image	Image Dimension	Arco's Method			Proposed Method		
		MP	ME (%)	RAE (%)	MP	ME (%)	RAE (%)
a	560×450	3,384	1.34	30.58	555	0.22	9.19
b	560×450	11,185	4.44	75.48	849	0.34	12.29
c	2560×1920	85,684	3.49	12.72	19,279	0.05	4.22
d	1280×960	42,939	1.74	97.16	637	0.39	42.47
e	1280×960	53,276	4.34	92.96	932	0.08	9.31
f	1280×960	18,503	1.51	57.52	2,681	0.22	15.90
Average		35,829	2.81	61.07	4,156	0.22	15.56

Here would be seen the effect of adding ovalness and solidity in the classification process of the connected component. Experiments are conducted on sample No. 8 by

comparing the two conditions, which are utilising area, ovalness, and solidity (Fig. 17a) and only area (Fig. 17b).

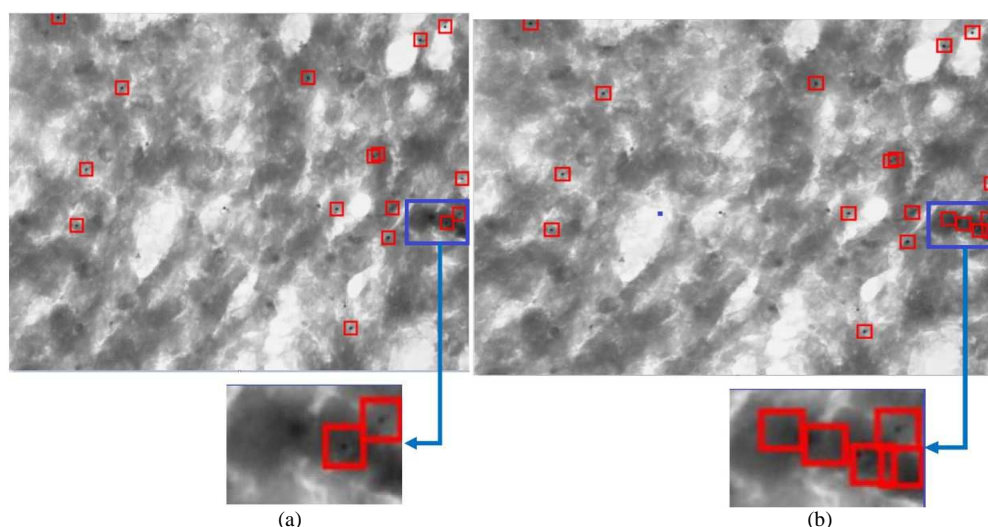


Fig. 17 Comparison system results: a. classification process use area, ovalness, and solidity, b. classification process uses area

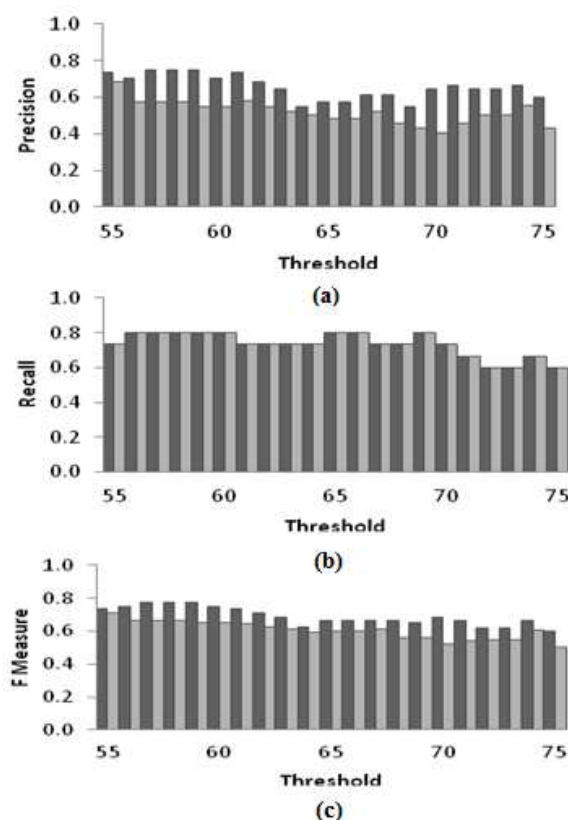


Fig. 18 Comparison of system performance: a. precision, b. recall, c. F-measure

□ = the classification use area, ovalness, and solidity
 ■ = the classification uses area

The result system utilised the area, ovalness, and solidity had similar results to systems that only used area. The different result was marked with a blue rectangle. There are additional results whose values were false positive (FP) at the system that only used area. It means that systems utilised area, ovalness, and solidity can reduce false positive results from the system. Conversely, false negative results are not

reduced. It means that the addition of properties can improve the precision system, but it does not change the recall system.

The threshold is increased from 55 to 75, and the effect of these properties on sample No.8 was shown at Fig. 18. It appears that the additional properties increase F-measure (Fig. 18c). However, the effect only occurs in the precision (Fig. 18a) while the recall unchanged (Fig. 18b). It means that the recall depends on the determination of the intensity slicing threshold, while the addition of properties reduces error detection (FP) but does not improve the correct detection of the parasite (TP).

CONCLUSION

In this study, we develop malaria parasite detection from thick blood smear microscopic images obtained from various endemic places in Indonesia. The detection is performed on data containing the actual peripheral blood that is white blood, platelet, parasite, and artifact. The process includes pre-processing, segmentation using blood component intensity slicing and morphological operation, blood component classification utilising rule based on properties of parasite candidates, and parasite candidate formation. Experiments were conducted by varying the threshold from 55 to 75 using 15 examples of each training and testing data. We also investigate the effect of adding properties during the classification process. The result shows that the proposed method is effective to detect malaria parasite from the thick blood smear microscopic image. The segmentation method is robust to the different condition of image and histogram. It reduced the misclassification error and relative foreground error by 2.6% and 45.5%, respectively. The addition of ovalness and solidity properties during the classification process improve precision of the system. From used sample data and given threshold intensity slicing 71, average precision, recall, and F measure are all 86%.

The future works of this study is modelling an automatic intensity slicing threshold that produces appropriate threshold with the condition of each thick blood smear

image to improve the recall of the system. Furthermore, advanced identification of the obtained parasite candidates will be conducted to reduce false positives of the system result. So, it can improve the precision value.

ACKNOWLEDGMENT

Datasets of this research are supported by Eijkman Institute of Molecular Biology Indonesia and Miss. Esti Suryani. This research is a part of research collaboration among Universitas Sebelas Maret, Center for Information and Communication Technology Agency for the Assessment and Application of Technology (PTIK-BPPT), and Eijkman Institute for Molecular Biology.

REFERENCES

- [1] WHO (2014) World Malaria Report 2014. [Online]. Available: http://www.who.int/malaria/publications/world_malaria_report_2014/en/
- [2] I. Hammami, A. Garcia, and G. Nuel, "Evidence for overdispersion in the distribution of malaria parasites and leukocytes in thick blood smears," *Malaria Journal*, vol. 12, pp. 1–15, 2013.
- [3] D. Syafruddin, P. B. Asih, I. E. Rozi, K. Chand, and S. Wangsamuda, *Diagnosis mikroskopik malaria*, 1st ed. Lembaga Biologi Molekuler Eijkman, 2010.
- [4] WHO, *Basic Malaria Microscopy*, 2nd ed. Switzerland: WHO Press, 2010.
- [5] N. Linder, R. Turkki, M. Willander, A. Mårtensson, V. Diwan, E. Rahtu, M. Pletikäinen, M. Lundin, J. Lundin, "A Malaria Diagnostic Tool Based on Computer Vision Screening and Visualization of Plasmodium falciparum Candidate Areas in Digitized Blood Smears," *PLOS ONE*, vol. 9, no. 8, pp. 1–12, 2014.
- [6] D. Anggraini, A. S. Nugroho, C. Pratama, I. E. Rozi, V. Pragesjvara, and M. Gunawan, "Automated status identification of microscopic images obtained from malaria thin blood smears using bayes decision: A study case in plasmodium falciparum," in *Proc. International Conference on Advanced Computer Science and Information Systems*, 2011, pp. 347–352.
- [7] D. K. Das, M. Ghosh, M. Pal, A. K. Maiti, and C. Chakraborty, "Machine learning approach for automated screening of malaria parasite using light microscopic images," *Micron*, vol. 45, pp. 97–106, 2013.
- [8] E. Dekel, A. Rivkin, M. Heidenreich, Y. Nadav, Y. Ofir-Birin, Z. Porat, N. Regev-Rudzki, "Identification and classification of the malaria parasite blood developmental stages, using imaging flow cytometry," *Methods*, vol. 112, pp. 157–166, 2016.
- [9] G. Díaz, F. A. González, and E. Romero, "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images," *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 296–307, 2009.
- [10] Z. May and M. Aziz, "Automated quantification and classification of malaria parasites in thin blood smears," in *Proc. International Conference on Signal and Image Processing Applications*, 2013, pp. 369–373.
- [11] M. I. Razzak, "Automatic Detection and Classification of Malarial Parasite," *International Journal of Biometrics and Bioinformatics*, vol. 9, pp. 1–12, 2015.
- [12] S. S. Savkare and S. P. Narote, "Automatic System for Classification of Erythrocytes Infected," in *Proc. 2nd International Conference on Communication, Computing & Security*, 2012, vol. 6, pp. 405–410.
- [13] V. V. Makkapati and R. M. Rao, "Ontology-based malaria parasite stage and species identification from peripheral blood smear images," in *Proc. International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6138–6141, 2011.
- [14] F. B. Tek, A. G. Dempster, and I. Kale, "Parasite detection and identification for automated thin blood film malaria diagnosis," *Computer Vision and Image Understanding*, vol. 114, pp. 21–32, 2010.
- [15] K. Bhowmik and P. Rakshit, "Detection of the presence of Parasites in Human RBC In Case of Diagnosing Malaria," in *Proc. Second International Conference on Image Information Processing*, 2013, pp. 329–334.
- [16] D. Mas, B. Ferrer, D. Cojoc, S. Finaurini, V. Mico, and J. Garcia, "Novel image processing approach to detect malaria," *Optics Communications*, vol. 350, pp. 13–18, 2015.
- [17] M. Le, T. R. Bretschneider, C. Kuss, and P. R. Preiser, "A Novel semi-automatic image processing approach to Determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears," *BMC Cell Biology*, vol. 12, pp. 1–12, 2008.
- [18] S. Kaewkamnerd, A. Intarapanich, M. Pannarat, S. Chaotheing, C. Uthaiipibull, and S. Tongshima, "Detection and Classification Device for Malaria Parasites in Thick-Blood Films," in *Proc. The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 2011, pp. 435–438.
- [19] M. Elter, E. Haßlmeyer, and T. Zerfaß, "Detection of malaria parasites in thick blood films," in *Proc. 33rd Annual International Conference of the IEEE EMBS*, 2011, pp. 5140–5144.
- [20] J. E. Arco, J. M. Górriz, J. Ramírez, I. Álvarez, and C. G. Puntonet, "Digital image analysis for automatic enumeration of malaria parasites using morphological operations," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3041–3047, 2015.
- [21] J. Kaur and A. Choudhary, "Comparison of Several Contrast Stretching Techniques on Acute Leukemia Images," *International Journal Engineering Innovation Technology*, vol. 2, pp. 332–335, 2012.
- [22] R. E. Putri, A. Yahya, N. M. Adam, and S. A. Aziz, "Correlation of Moisture Content to Selected Mechanical Properties of Rice Grain Sample," *International Journal on Advanced Science, Engineering & Information Technology*, vol. 5, no. 5, pp. 264–267, 2015.
- [23] S. Y. Jiang and L. X. Wang, "Efficient feature selection based on correlation measure between continuous and discrete features," *Information Processing Letters*, vol. 116, no. 2, pp. 203–215, 2016.