

Feature analysis for stage identification of *Plasmodium vivax* based on digital microscopic image

Hanung Adi Nugroho¹, I Md. Dendi Maysanjaya², Noor Akhmad Setiawan³

E. Elsa Herdiana Murhandarwati⁴, Widhia K.Z Oktoeberza⁵

^{1,2,3,5}Department of Electrical and Information Technology, Faculty of Engineering,
Universitas Gadjah Mada, Yogyakarta, Indonesia

²Department of Informatics Education, Faculty of Engineering and Vocation, Universitas Pendidikan Ganesha, Indonesia

⁴Department of Parasitology, Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

Article history:

Received Oct 9, 2018

Revised Nov 19, 2018

Accepted Nov 28, 2018

Keywords:

Feature analysis

Stages of plasmodium vivax

Thin blood film

Wrapper feature selection

ABSTRACT

Plasmodium parasite is identified to confirm malaria disease. Paramedics need to observe the presence of this parasite prepared on thick and thin blood films under microscope. However, false identification still occurs which is caused by human factor during the examination. Thus, malaria identification based on digital image processing has been widely developed to overcome the error possibility. This paper proposes a scheme to identify and classify the stages of *Plasmodium vivax* parasite on digital microscopic image of thin blood films based on feature analysis. Shape and texture features are extracted from segmented parasite objects. Feature selection based on wrapper method is then conducted to obtain relevant features which may contribute in improving the classification result. The classification process is conducted based on Naïve Bayes classifier. The performance of proposed method is evaluated using 73 digital microscopic images of *P-vivax* parasite on thin blood films comprising of 29 trophozoites, 10 schizonts and 34 gametocytes stages. By using six selected features including perimeter, dispersion, mean of intensity, ASM, contrast GLCM and entropy GLCM, the proposed scheme achieves the best classification rate with the accuracy, sensitivity and specificity of 97.29%, 97.30% and 97.30%, respectively. This indicates that the proposed scheme has a potential to be implemented in the development of a computerised aided malaria diagnosis system for assisting the paramedics.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Hanung Adi Nugroho,

Department of Electrical Engineering and Information Technology,

Faculty of Engineering, Universitas Gadjah Mada,

Jl. Grafika 2 Kampus UGM, Yogyakarta 55281, Indonesia.

Email: adinugroho@ugm.ac.id

1. INTRODUCTION

Malaria is a disease caused by *Plasmodium* parasite which is transmitted to humans through the bite of female Anopheles mosquitos. As reported by World Health Organisation (WHO), this disease are transmitted in more than 90 countries and put about 3.2 billion people at risk of malaria with mainly morbidity occur in Africa, South-East Asia, Latin America and the Middle East[1]. *Plasmodium* is divided into five species, i.e. *Plasmodium falciparum* (*P. falciparum*), *Plasmodium vivax* (*P. vivax*), *Plasmodium ovale* (*P. ovale*), *Plasmodium malariae* (*P. malariae*) and *Plasmodium knowlesi* (*P. knowlesi*). The greatest threat of malaria causes comes from *P. falciparum* and *P. vivax*. [1].

The *Plasmodium* undergoes two phases during the infection process of the human body, namely exoerythrocytic phase in the liver and intraerythrocytic phase in blood stream circulation. In the bloodstream circulation, it will go through other three stages, i.e. trophozoites, schizonts and gametocytes stages [2].

Figure 1 shows the life cycle of malaria. When *Plasmodium* infection is suspected, thick and thin blood films preparation will be made. Examination on thick blood aims to detect the presence of *Plasmodium* parasites while the thin blood film examination is to identify what species of *Plasmodium* causing the disease.

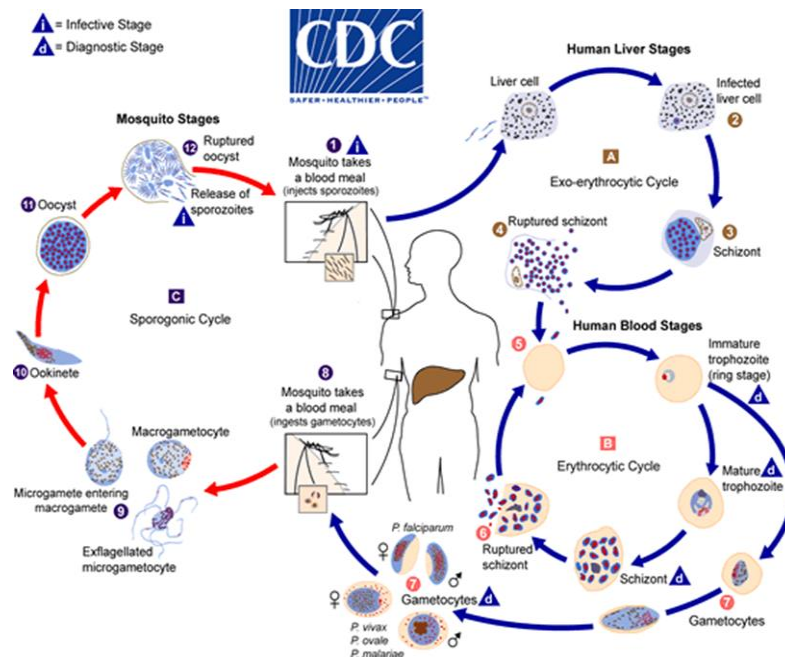


Figure 1. The life cycle of malaria [2]

A false diagnosis on thin blood film examination can be affected by some factors particularly the expertise level of paramedics, the blood film preparation method, the staining method and the quality of microscope used. Hence, several studies have been conducted to develop computer-aided malaria diagnosis based on digital image processing to reduce the error possibility.

Khan *et al.* [3] applied k -means on b channel of the L^*a^*b colour model to segment *P. vivax* parasite. However, the k value was determined manually, and the visual quality of segmentation result was poor. Nasir *et al.* [2] employed the combination of moving k -means clustering (MKM) and seeded region growing area extraction (SRGAE) methods to identify *P. vivax*. Their study proved that the use of saturation (S) band of HSI colour model was able to obtain better segmentation result than that of intensity (I) band. Dian *et al.* [4] detected the blood cell component in red thin blood smear by applying global thresholding and connected component labelling (CCL). Ruberto *et al.* proposed the combination of automatic thresholding and morphological approach to detect and classify malaria parasites [5]. Furthermore, Akbar *et al.* [6] introduced combination of k -means clustering and morphological operation methods on HSV colour model to segment *P. falciparum* on the thin blood films. Then, several shape and texture features were extracted and classified by using MLP classifier to classify *P. falciparum* stage into three classes, i.e. trophozoites, schizonts and gametocytes. However, the determination of the cluster number in k -mean was still manual and the obtained features were still too many.

To complete the identification study of *Plasmodium* parasite, this paper proposes a scheme to classify *P. vivax* parasite on digital microscopic image of thin blood films. The classification is categorised into three stages, i.e. trophozoites, schizonts and gametocytes. The main purpose of this study is to obtain the significant features for improving the classification result based on wrapper subset evaluation. The structure of this paper is organised as follows. Section II illustrates the experimental set up. The results and discussion are presented in Section III followed by conclusion in Section IV

2. APPROACH

The methodology consists of five main processes, namely pre-processing, segmentation, feature extraction, feature selection and classification as depicted in Figure 2. The first two processes, i.e. pre-processing and segmentation, are conducted by adopting the proposed scheme in our previous work [7]. Firstly,

the RoI image with the resolution of 250x250 pixels is cropped from the original image. The red and saturation bands are used in this study. Then, contrast stretching, and median filter are applied to enhance the quality of RoI image. Furthermore, Otsu thresholding and morphological operations are conducted to segment *P. vivax*.

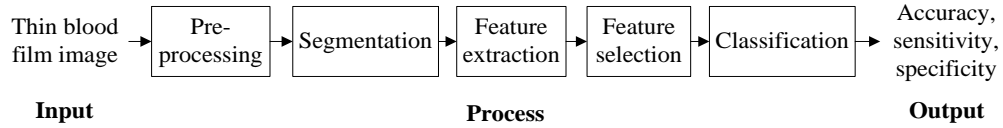


Figure 2. Block diagram of the approach

2.1. Feature extraction

The segmented image subsequently undergoes feature extraction process based on the shape and texture features. The shape feature comprising the contour-based and invariant moment features. For the texture feature, histogram-based and GLCM features are extracted. There are seven contour-based features including perimeter, area, roundness, slimness, convexity, solidity and dispersion. Perimeter represents the edge length of an object as formulated in (1). The object with 4-adjacency obtain better result of perimeter than that of 8-adjacency. Here, N_e is an even number of codes and N_o is an odd number of codes. Area is the total of pixels object as calculated in (2). The notation of R and ∂R represent the object area and edge of the object, respectively.

$$P = N_e + N_o\sqrt{2} \tag{1}$$

$$A = \iint_R dx dy = \int_{\partial R} y(t) \frac{dx(t)}{dt} dt - \int_{\partial R} x(t) \frac{dy(t)}{dt} dt \tag{2}$$

Roundness is the ratio between the object area and quadratic perimeter while slimness is the ratio between the width and the length of the object. Roundness and slimness are expressed in (3) and (4), respectively.

$$R = \frac{4\pi \times Area}{perimeter^2} \tag{3}$$

$$Slimness = \frac{width}{length} \tag{4}$$

Convexity is the ratio between convex perimeter and object perimeter as declared in (5) and solidity is the ratio between the object and convex areas as formulated in (6). Dispersion feature expresses the irregularity of the object which is calculated using (7) as the ratio between the lengths of main cord to the object area.

$$Convexity = \frac{Convex\ perimeter}{Object\ perimeter} \tag{5}$$

$$Solidity = \frac{Object\ area}{Convex\ area} \tag{6}$$

$$(S) = \frac{\pi \max(\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2})}{A(S)} \tag{7}$$

here, (\bar{x}, \bar{y}) is the centre point of the mass area $A(S)$ while $A(S)$ is the object area.

The invariant moment known as Hu moment is calculated based on normalised centre moments [8]. The moment values do not depend on translation, scaling and rotation. There are seven features of the invariant moment but only three features used in this study, i.e. moment 1, moment 2 and moment 3 as mathematically formulated in (8) to (10). Normalised moment is declared by η_{ij} while ij is the moment order.

$$\emptyset_1 = (\eta_{20} + \eta_{02}) \tag{8}$$

$$\emptyset_2 = (\eta_{20} + \eta_{02})^2 + (2\eta_{02})^2 \quad (9)$$

$$\emptyset_3 = (\eta_{30} + 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2 \quad (10)$$

Texture is the basic feature related to roughness, granulation and regularity of pixels structure and as the repetition of basic pixels is called as *texel* (texture element) [9]. The two kinds of texture feature based on the statistical order used include histogram-based and grey level co-occurrence matrices (GLCM) features. The histogram-based feature is the first-order statistical which comprises of six features, i.e. mean of intensity, deviation standard, skewness, energy, entropy and smoothness. They are formulated in (11) to (16).

$$m = \sum_{i=0}^{L-1} i \cdot p(i) \quad (11)$$

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (i - m)^2 p(i)} \quad (12)$$

$$skewness = \sum_{i=0}^{L-1} (i - m)^3 p(i) \quad (13)$$

$$energy = \sum_{i=0}^{L-1} [p(i)]^2 \quad (14)$$

$$entropy = - \sum_{i=0}^{L-1} p(i) \log_2(p(i)) \quad (15)$$

$$R = 1 - \frac{1}{1 + \sigma^2} \quad (16)$$

The second-order statistical method is conducted by calculating the probability of adjacency relationship between two pixels at a certain distance and angular orientation (0, 45, 90 and 135 degrees) [10]. Five GLCM features extracted are angular second moment (ASM), inverse difference moment (IDM), entropy, contrast and correlation.

ASM is used to calculate the homogeneity of image using (17) with the number of levels for computation expressed as L . The measurement variation of grey level pixels image known as contrast is formulated in (18). Whilst, IDM is used to measure homogeneity as formulated in (19).

$$ASM = \sum_{i=1}^L \sum_{j=1}^L GLCM(i, j)^2 \quad (17)$$

$$contrast = \sum_{n=1}^L n^2 \left\{ \sum_{|i-j|=n} GLCM(i, j) \right\} \quad (18)$$

$$IDM = \sum_{i=1}^L \sum_{j=1}^L \frac{GLCM(i, j)}{1 + (i - j)^2} \quad (19)$$

Entropy describes the irregularity of grey level image. If elements of GLCM are relative the same, high entropy value would be obtained. Low entropy value is achieved if the elements of GLCM near 0 or 1. Correlation features is used to measure the linear dependence of grey level value of the image. Entropy and correlation are denoted in (20) and (21).

$$entropy = - \sum_{i=1}^L \sum_{j=1}^L GLCM(i, j) \log(GLCM(i, j)) \quad (20)$$

$$correlation = \frac{\sum_{i=1}^L \sum_{j=1}^L (ij)(GLCM(i, j)) - \mu_i \mu_j}{\sigma_i \sigma_j} \quad (21)$$

2.2. Feature selection

Feature selection is conducted to obtain the significant extracted features for improving the accuracy and reducing the computation time during classification process [11]. Wrapper subset evaluation-based method used in this study since it uses a learning algorithm and k folds cross-validation as part of the evaluation function while searching the features [12]. Iteratively, wrapper will preserve the relevant features and eliminate the irrelevant features.

2.3. Classification

The classification process aims to determine independent variable (features) that has the highest correlation to dependent variable (class of the object). Naïve Bayes classifier is used in this study since its relatively fast in training, able to handle the real and discrete data and unaffected by irrelevant features [13], [14].

3. RESULTS AND DISCUSSION

A total of 73 digital microscopic images of *P. vivax* parasite on thin blood films taken from the Department of Parasitology, Faculty of Medicine, Universitas Gadjah Mada, were used in this study. The dataset consists of three stages images, namely 29 images of trophozoites, 10 images of schizonts and 34 images of gametocytes stages, in BMP format with the resolution of 1600x1200 pixels.

Firstly, original image is cropped into 250x250 pixels in ROI of parasite area as depicted in Figure 3. Then, contrast stretching is applied to enhance the quality of ROI image. For segmentation process, R-band from the RGB colour model and S-band from HSV colour model are chosen since they have the best quality of intensity. Afterward, each of them is filtered by median filter and combined. To obtain the parasite object, Otsu thresholding followed by morphological operation are conducted to filtered image. The sample of segmentation result is presented in Figure 4. For the detail process has been explained in [7].

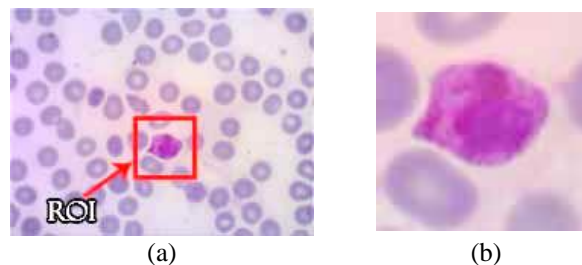


Figure 3. (a) Original image (b) RoI image [7]

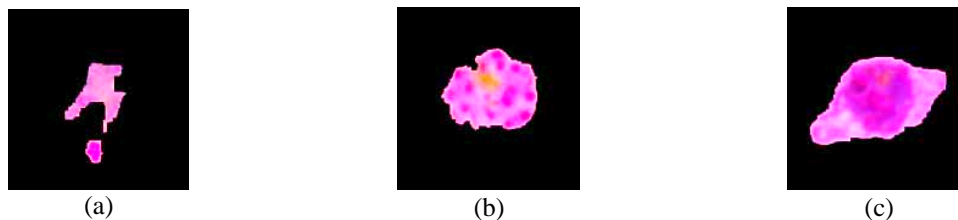


Figure 4. The segmentation result of (a) trophozoites (b) schizonts and (c) gametocytes stages [7]

Having obtained the parasite object, the shape-based and texture-based feature extraction are then conducted. A total of 10 shape-based features are extracted which comprises of seven contour-based features and three features of invariant moment. There are seven features of invariant moment but only three features used since the four others obtain 0 value. The value of moment 1 represents the centre of gravity, the value of moment 2 denotes the smoothness and the 3-moment value represents the asymmetry of intensity. For the texture-based features, a total of 11 features are extracted consisting of six histogram-based features and five features of GLCM. The summary of 21 extracted features is described in Table 1. Furthermore, feature selection is conducted to obtain the significant features based on Wrapper method. Six selected features are perimeter, dispersion, mean of intensity, ASM, contrast GLCM and entropy GLCM. These extracted features are then classified by using Naïve Bayes classifier based on 10-folds cross validation.

To evaluate the proposed scheme, some statistical parameters are involved including accuracy, sensitivity and specificity which are mathematically formulated from (22) to (24). Accuracy expresses the successful rate of classification process. Sensitivity is a capability of classifier to predict positive class as positive while specificity is a capability of classifier to predict negative class as negative.

In this work, four types of classification based on extracted features are conducted. They are shape features, texture features, shape and texture features and selected features. Table 2 presents the comparison of

classification result of these features. As depicted in Table 2, eleven texture features yield the low classification rate with the accuracy, sensitivity and specificity of 94.59%, 94.6% and 94.6%, respectively. The better evaluation rate is gained by using 10 shape features with the accuracy of 97.29%, sensitivity of 97.30% and specificity of 97.30%. The same result is not only achieved by the 21 full features of combination shape and texture features but also is achieved by six selected features.

Table 1. The result of feature extraction

Shape features	Texture features
Perimeter	Mean of intensity
Area	Contrast
Roundness	Skewness
Slimness	Energy
Convexity	Entropy
Solidity	Smoothness
Dispersion	ASM
Moment 1	IDM
Moment 2	Contrast GLCM
Moment 3	Entropy GLCM
	Correlation

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (22)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \quad (23)$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \quad (24)$$

Table 2. The comparison evaluation result of extracted features

Extracted features	Accuracy (%)	Sensitivity (%)	Specificity (%)
Shape features (10)	97.29	97.30	97.30
Texture features (11)	94.59	94.60	94.60
Shape and texture features (21)	97.29	97.30	97.30
Selected features (6)	97.29	97.30	97.30

Although they produced the same value, the evaluation rate by using six features is better than that of the full features. It indicates that not all of the 21 full features may significantly contribute in the classification process. Moreover, by using a small number of features, the proposed scheme is still able to gain the high accuracy, sensitivity and specificity even may reduce the computation time. This result indicates that the proposed scheme successfully obtains the significant features for identifying and classifying the stage of *P. vivax* parasite on the digital microscopic image of thin blood films.

4. CONCLUSION AND FUTURE WORK

This study proposes a scheme to classify *P. vivax* parasite on digital microscopic image of thin blood films into three stages, namely trophozoites, schizonts and gametocytes. A total of 10 shape-based features and 11 texture-based features are extracted to facilitate the classification process. Feature selection based on wrapper method is conducted to gain the relevant features which may contribute to improve the rate of classification result.

Six selected features consisting of perimeter, dispersion, mean of intensity, ASM, contrast GLCM and entropy GLCM achieve the best evaluation rate with the accuracy of 97.29%, sensitivity of 97.30% and specificity of 97.30%. The proposed scheme is able to identify and classify the stage of *P. vivax* parasite by using only significant selected features resulting in the more efficient computation time during the process. Hence, the proposed scheme has a potential to be implemented as part of the computerised aided malaria diagnosis system for assisting the paramedics.

In the next investigation, the authors consider more data with the balanced proportion in each class and feature usage in order to gain the higher accuracy. Thus, the performance of proposed scheme can be more convincing.

ACKNOWLEDGEMENTS

This project is funded by Directorate General of Higher Education, Ministry of Research, Technology and Higher Education, Republic of Indonesia. The authors would like to thank the Intelligence Systems research group for sharing meaningful knowledge and inspiring discussion.

REFERENCES

[1] W. H. Organisation. "Malaria: Key Facts," 2017. Available: <http://www.who.int/mediacentre/factsheets/fs094/en/>

[2] A. A. Nasir, M. Mashor, and Z. Mohamed, "Segmentation based approach for detection of malaria parasites using moving k-means clustering," in *2012 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 653-658, 2012.

[3] N. A. Khan, H. Pervaz, A. K. Latif, and A. Musharraf, "Unsupervised identification of malaria parasites using computer vision," in *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 263-267, 2014.

[4] 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 *2011 International Conference on Advanced Computer Science and Information System (ICACSIS)*, pp. 347-352, 2011.

[5] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Morphological image processing for evaluating malaria disease," in *Visual Form 2001*, ed: Springer, pp. 739-748, 2001.

[6] H. A. Nugroho, S. A. Akbar, and E. E. H. Murhandarwati, "Feature extraction and classification for detection malaria parasites in thin blood smear," in *2015 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, pp. 197-201, 2015.

[7] I. M. D. Maysanjaya, H. A. Nugroho, N. A. Setiawan, E. E. H. Murhandarwati, T. R. Nuringtyas, R. Roto, et al., "Segmentation of Plasmodium vivax phase on digital microscopic images of thin blood films using colour channel combination and Otsu method," in *AIP Conference Proceedings*, pp. 160002, 2016.

[8] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE transactions on information theory*, vol. 8, pp. 179-187, 1962.

[9] B. Chaudhuri, N. Sarkar, and P. Kundu, "Improved fractal geometry based texture segmentation technique," *IEE Proceedings E (Computers and Digital Techniques)*, vol. 140, pp. 233-242, 1993.

[10] R. M. Haralick and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. SMC-3, pp. 610-621, 1973.



[11] M. A. Hall, "Correlation-based feature selection for machine learning," The University of Waikato, 1999.




[12] G. Naqvi, "A Hybrid Filter-Wrapper Approach for Feature Selection," Master Thesis, *Department of Technology*, Örebro University, Sweden, 2012.

[13] C. M. Bishop, "Pattern recognition and machine learning," New York: Springer, 2006.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification and Scene Analysis Part 1: Pattern Classification," Wiley, Chichester, 2000.

BIOGRAPHIES OF AUTHORS

	<p>H.A. Nugroho completed his bachelor's degree in electrical engineering (2001) from the Universitas Gadjah Mada, Yogyakarta, Indonesia and master's degree in biomedical engineering (2005) from The Univ. of Queensland, Brisbane, Australia. In 2012, he received his PhD in Electrical and Electronic Engineering from the Universiti Teknologi Petronas (UTP), Malaysia. Currently, he is an Assistant Professor and also a Vice Head of Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada (UGM). His current research interests include biomedical signal and image processing and analysis, computer vision, medical instrumentation and pattern recognition.</p>
	<p>I Md. Dendi Maysanjaya completed his Bachelor degree in Educational of informatics engineering, Universitas Pendidikan Ganesha (2012) and his Master degree in the Department of Electrical Engineering and Information Technology (2015) from the Universitas Gadjah Mada, Yogyakarta, Indonesia. Currently, he is a lecturer at the Department of Informatics Education, Faculty of Engineering and Vocation, Universitas Pendidikan Ganesha, Indonesia.</p>

	<p>N.A. Setiawan completed his bachelor's degree and master's degree in electrical engineering (1998) from the Universitas Gadjah Mada, Yogyakarta, Indonesia. In 2009, he received his PhD in Electrical and Electronic Engineering from the Universiti Teknologi Petronas (UTP), Malaysia. Currently, he is an Assistant Professor at the Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada (UGM). His current research interests include computational intelligence, soft computing machine learning, electrical engineering, and biomedical engineering.</p>
	<p>E.E.H. Murhandarwati completed her bachelor's degree (1993) and master's degree (1996) in Faculty of Medicine, Universitas Gadjah Mada. In 2011, she received her PhD in Medical Science Monash University. Currently, she is a Lecturer in the Department of Parasitology, Faculty of Medicine, Universitas Gadjah Mada (UGM). Her research interest is tropical medicine.</p>
	<p>W.K.Z. Oktoeberza completed her bachelor's degree in informatics (2012) from the Universitas Bengkulu and her master's degree in electrical engineering (2015) from the Universitas Gadjah Mada, Yogyakarta, Indonesia. Currently, she works as a Lecturer at the Department of Informatics, Faculty of Engineering, Universitas Bengkulu, Indonesia.</p>