Comparative study of classification method on diagnosis of *plasmodium* phase

I M D Maysanjaya

Department of Information System, Faculty of Engineering and Vocational, Universitas Pendidikan Ganesha, Bali, Indonesia – 81116

E-mail: dendi.ms@undiksha.ac.id

Abstract. Malaria is a plague in humans induced by the Plasmodium parasitoid transferred through a single bite of female Anopheles mosquitoes. Once reaching human blood passage, this parasitoid bears asexual proliferation which is classified toward three phases (trophozoite, schizont, and gametocyte). To discover the phases, the paramedic will analyse the blood specimen of the subject within the microscope. However, the aforementioned approach has the potential to misdiagnosis. Many CAD-based investigations have been carried out to reduce the diagnosis errors that have formed. This research sought to develop a CAD-based design that can help paramedics in diagnosing the parasitoid Plasmodium, including trying to get a classification algorithm that is able to analyze the Plasmodium phase precisely. Based on the experiment outcome, Naïve Bayes was mostly useful applied as a classifier that attained the accuracy of 97.29%, the sensitivity, and specificity of 97.30% toward the P.vivax case. In the case of P.falciparum, it emerged in the accuracy, sensitivity, and specificity of 98.36%, 98.40%, and 98.40% apiece. Meantime, Perceptron was a useless algorithm applied as a classifier that realized the accuracy up to 81.08%, sensitivity and specificity of 93.80% on each. In the case of P.vivax, the accuracy, sensitivity, and specificity realized were 80.33%, 92.50%, and 92.50% respectively.

1. Introduction

Malaria is one of the plagues in humans which may lead to death if it is not treated properly. *Plasmodium* is the main cause of malaria, dispatched over the sting of female *Anopheles* mosquitoes. *Plasmodium* consists of four types of parasitoids [1], i.e. (1) *Plasmodium falciparum* (*P.falciparum*)–it can lead the patient to suffer from cerebral malaria, severe anemia, renal failure, and asphyxia. This parasite is the most significant factor claiming the human death compared to other forms of *Plasmodium*; (2) *Plasmodium vivax* (*P.vivax*)–the second most dangerous *Plasmodium* parasite (contributing of about 43% in human [2]), causing tertian malaria (malaria with fever once in three days); (3) *Plasmodium malariae* (*P.malariae*) – the root of quartan malaria (fever once in every four days), and contributing 7% of malaria cases in the world; and (4) *Plasmodium ovale* (*P.ovale*) – a type of parasite causing pernicious malaria (rarely to be found). *Plasmodium* firstly will infect the human liver followed with the asexual propagation phase comprising three phases – trophozoite, schizont, and gametocyte[2–4].

In handling a patient with malaria, the paramedic will give a diagnosis based upon the blood sample examination through a microscope. The result of the diagnosis is highly determined by the condition of the microscope being used and the level of competence and expertise of the paramedic. The result could
have relatively high potency of misdiagnosis; thus, through the current advance technology, it is expected to be able to minimize the misdiagnosis.

A digital blood smear is initiated by combining the conventional microscope with the Optilab camera. The result of this acquisition would later be processed under the method in digital image processing, and then the type of Plasmodium and its phase is identified. Much researches have been conducted to develop the model and method to detect and identify the Plasmodium in the erythrocyte as conducted by[3,5–8]. However, there are still some weaknesses found in those researches. The weaknesses could be caused by some factors, among others is the selection of the classification algorithm that might be less compatible.

Based on the explanation above, this present research aimed to observe the classification algorithm that might result in the optimal accuracy – particularly to identify the incubation phase of parasite *P. falciparum* and *P. vivax*. Both parasites were selected because are the type of Plasmodium that are most frequently found – particularly for the malaria case occurring in Indonesia. It expected that the result of the research could bring a contribution to the computer-based aided diagnosis (CAD) development in assisting the paramedic to diagnose malaria.

2. Related Works

To diagnose the Plasmodium, examining the blood sample utilizing a microscope can be conducted. In general, there are two blood samples used in diagnosis; those are blood samples for the thick blood smear and thin blood smear. In the thick blood smear, the erythrocytes will be lysed to make only parasitic cells visible. The thick blood smear was applied to define the number of which can be used to define the severity level of the patient[9]. Meanwhile, for the thin blood smear, the erythrocytes are not lysed; thus, the paramedic can classify the variety or the stage of Plasmodium. Figure 1 illustrates the thick blood smear and the thin blood smear.

![Figure 1. Plasmodium parasite on (a) thick blood smear, (b) thin blood smear.](image)

As explained previously, the result of the diagnosis was determined by the level of the experience and the expertise of the paramedic. This certainly has made the subjectivity on the diagnosis result is highly influential. Hence, it is necessary to minimize the level of error in the diagnosis and subjectivity. A number of researches has been conducted and developed based upon the CAD technique and collaborated with the digital image processing.

Nasir *et al*[3] used the image of *P. vivax* as the research object purposely to compare the classification result between segmentation on the saturation band and the intensity band of HIS color space. This segmentation method was combined with Moving K-Means (MKM) Clustering and SRGAE algorithm. Khan *et al*[6] developed a segmentation method using *b* band from LAB color space combined by K-Means algorithm to gain the tissue of *P. vivax*. Ghosh *et al*[10] compared the result of the Plasmodium parasite segmentation in the color space of HIS, YCbCr, LAB, and CMYK, using Fuzzy Divergence-based Thresholding method. Based on that research, the result of the segmentation has been found in the cyan band from the CMYK color space resulting in the best of the segmentation.

In addition to the three researches previously explained, there are other researches focused on Plasmodium based upon the digital image processing such as the ones conducted by [5,7,8,11,12].
3. Methods

This research focused on the test of classification method that can achieve in the best level of accuracy to classify the phase of Plasmodium, particularly the phase of *P. falciparum* and *P. vivax*. The details of methods are explained as follows.

3.1. Classification

The classification process refers to a phase to learn the independent variables that have the highest correlation towards the dependent variable, in which the independent variables are the object features and the dependent variables are the object class. To do the classification process, the artificial neural network (ANN) algorithm is frequently used, defined as an algorithm imitating the work process of the human brain in processing the knowledge, or in simple word ANN is a method to treat the input vector to be output one[13].

In this research, there were five types of classifier algorithm tested to classify the phase of *Plasmodium*, particularly the phase of *P. falciparum* and *P. vivax*. A brief explanation about those five classifier algorithms is presented as follows:

3.1.1. Radial Based Function (RBF). RBF consists of three layers i.e. the input, hidden, and output layer. The training process was conducted in the hidden layer through the supervised learning algorithm. RBF has main excellence that is the hybrid learning. The weight of the node commonly will be sought from the smallest in the network when the learning process is in progress. Each unit of input $X_i$ receives the input signal and forwards it to the hidden layer, known as the radial centre ($c$). This centre is delimited based on the input vectors set. RBF could give a genuine effect if the input data resembles the training data, and vice versa[14,15].

In the hidden layer, the formula commonly applied is Gaussian function as formulated in Equation (1) beneath.

$$\varphi_k(||X - C_k||) = \frac{1}{\sigma^2} e^{-||X - C_k||^2}$$  \hspace{1cm} (1)

where $C_k$ denotes the center of RBF, while $k$ is a number of neuron.

3.1.2. Multi-layer Perceptron (MLP). In the determination of the new weight and bias, MLP uses the gradient to be rapidly adjusted into the weight of the network. MLP is one of neural network (NN) methods comprising many layers that commonly consists of at least three hidden layers in addition to the input and output layers. MLP was developed from Perceptron, which by adding hidden layer. To its learning mechanism, MLP uses the algorithm of back-propagation. MLP is broadly utilized because of its strength to maintain parallel execution, ability for generalization, error limit, and an adequate learning algorithm. Nevertheless, MLP has a weakness particularly when it is implemented in digital image processing. This predicament befalls as the formation of topology from the input pattern is ignored, and the input of MLP is considered as the one dimension vector, while the image is two-dimension vector [16].

3.1.3. Naïve Bayes. Naïve Bayes is the algorithm realized from Bayes theory, which owns several excellences such as approximately quick on the training process, capable to manage real and discrete data, not being influenced by the unrelated features, as a tough presumption towards the independence of each features [17–19]. In simple, the equation for Naïve Bayes explained through Equation (2).

$$\text{posterior} = \frac{\text{prior} \ast \text{likelihood}}{\text{evidence}}$$  \hspace{1cm} (2)

*Posterior* is a possibility of the appearance of the sample with a definite character in a class. It is accomplished by procreation among the possibility of the class appearance (*Prior*) and the possibility...
of the appearance of sample characteristics in class (Likelihood) and then separated by the possibility of the sample characteristic appearance in global (Evidence).

3.1.4. Back Propagation Algorithm (BPA). BPA is the type of multi-layered feed-forward NN, which is formed based upon the back-propagation algorithm. Its gets the input vector and relinquishes it into the network, discriminates as necessitated, provides output based on the input signal, set the weight with the differential of fault based on the weight of learning rate[14,15].

3.1.5. Perceptron. Perceptron is the mildest type of ANN utilized to classify the linearly separable patterns. Perceptron is one kind of single-layer NN since it only has an input layer and output layer (without hidden layer). Perceptron uses a rigid limitation as the activation function or transfer function (f). This function possesses two sets of output signals, namely 0 and 1[15]. The structure of this Perceptron is displayed in Figure 2.

![Figure 2. A Perceptron Neuron Structure[20].](image)

The value of net and f(net) can be defined by Equation (3) and Equation (4) sequentially.

\[
net = \sum x_i \cdot w_i + b
\]

\[
f(\text{net}) = \begin{cases} 1, & jika \text{ net} > \theta \\ 0, & jika -\theta \leq \text{net} \leq \theta \\ -1, & jika \text{ net} < -\theta \end{cases}
\]

where \(x\) denotes input, \(w\) is a weight value, bias is denoted by \(b\).

3.2. Measure for Performance Evaluation

3.2.1. Accuracy, Sensitivity, and Specificity. Accuracy, sensitivity, and specification were used to assess the experiment. Each formula is shown in Equation (5), Equation (6), and Equation (7) apiece. Accuracy denotes an analysis to examine the amount of completion of the classification is executed. Sensitivity indicates the capability of classifier in divining the positive class as the True Positive (TP), while specificity relates to the ability classifier in identifying the negative class as the True Negative (TN).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

where \(TP + TN + FP + FN\) is total number of data.
3.2.2. K-Folds Cross-Validation. To reach more reliable classification results, k-folds cross-validation was employed. This approach has numerous benefits, such as decreasing bias given while the training process by using random sampling[21]. This scheme is begun by breaking all data randomly into the k group, known as a fold, with each fold is expected to have an identical number of members. The classification algorithm has to perform the training and testing k times. In per process, one-fold is enrolled as a data test while the other folds work as training data. Every process will perform various results. These values are then equalized and the average value is accepted as the accuracy value of the corresponding algorithm.

4. Result and Discussion
In this work, the digital thin blood smear dataset was obtained from Department of Parasitology, Faculty of Medicine, Universitas Gadjah Mada, as many as 180 data for \( P.\text{vivax} \) and \( P.\text{falciparum} \). The pre-processing to segmentation stage used the combination of method developed by[22–24]. The stage of feature extraction used 10 shape feature algorithms (perimeter, area, roundness, trimness, dispersion, convexity, solidity, 1\(^{st}\) – 3\(^{rd}\) invariant moment), and 11 texture feature algorithms (mean of intensity, contrast_histogram, skewness, energy, entropy_histogram, smoothness, ASM, IDM, contrast_GLCM, entropy_GLCM, correlation). Once the features of shape and texture were acquired, the following stage was feature selection using the \textit{Wrapper} method. Based on the result of feature selection, six features of \( P.\text{vivax} \) were obtained (roundness, solidity, skewness, entropy_histogram, IDM, and entropy_GLCM), while five features were found for \( P.\text{falciparum} \) (area, trimness, mean of intensity, correlation, and contrast_GLCM). These features are used as the objects on the classification stage which the results are shown in Table 1 and Table 2.

Table 1. The Comparison Result of Classification Methods for \( P.\text{vivax} \) Parasite Phase.

| Classifier | Parameter | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|------------|-----------|--------------|----------------|-----------------|
| RBF        | Accracy   | 94.59        | 94.60          | 94.60           |
| MLP        | Accracy   | 89.19        | 89.20          | 89.20           |
| Naive Bayes| Accracy   | 97.29        | 97.30          | 97.30           |
| BPA        | Accracy   | 94.59        | 94.60          | 94.60           |
| Perceptron | Accracy   | 81.08        | 93.80          | 93.80           |

Table 2. The Comparison Result of Classification Methods for \( P.\text{falciparum} \) Parasite Phase.

| Classifier | Parameter | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|------------|-----------|--------------|----------------|-----------------|
| RBF        | Accracy   | 96.72        | 96.70          | 96.70           |
| MLP        | Accracy   | 96.72        | 96.70          | 96.70           |
| Naive Bayes| Accracy   | 98.36        | 98.40          | 98.40           |
| BPA        | Accracy   | 88.52        | 88.50          | 88.50           |
| Perceptron | Accracy   | 80.33        | 92.50          | 92.50           |

Based on the outcomes presented in Table 1 and Table 2, it can be noticed that Naive Bayes gained the highest value in accuracy, sensitivity, and specificity of other classifiers. Meanwhile, the Perceptron has given the lowest level of classification. Naive Bayes could achieve the highest level of classification because it has some advantages, particularly not being affected by the irrelevant features and making a
strong assumption towards independence in each feature. Yet, for Perceptron, it resulted in the lowest classification for its simple structure making it less capable of handling any complexity in the classified feature data. As the effect, in the case of\textit{P. falciparum}, 8 data could not be classified and in the case of \textit{P. vivax}, there were 5 data unclassified. Certainly, this is the one affecting the accuracy value resulted from Perceptron.

The result of Naïve Bayes had a difference of 1.07\% when classifying the phase of \textit{P. falciparum} and \textit{P. vivax}, caused by the difference in the characteristics of \textit{P. falciparum} and \textit{P. vivax}. In \textit{P. falciparum} the prominent feature was the texture; while, in \textit{P. vivax} it was in the shape feature. Surely, this difference would bring an effect on the classification result. In general, from the result in this work, it can be stated that the Naïve Bayes algorithm can gain the best result of classification in comparison with the other four classification algorithms.

5. Conclusion and Future Work

Based on the result of the test obtained, it can be concluded that among five ANN algorithms, Naïve Bayes had the highest accuracy (98.36\% for the case of \textit{P. falciparum}, and 97.29\% for the \textit{P. vivax} case). Meanwhile, Perceptron had the lowest value (80.33\% and 81.08\% for the case of \textit{P. falciparum} and \textit{P. vivax} on each). Perceptron had the lowest accuracy caused by some data that were not classified. On the other hand, Naïve Bayes resulted in the greatest accuracy because of the ability to fault sensitivity.

From the result of this study in the future, it is expected that it can be used as a reference in selecting the classification algorithm which has a good result of accuracy. Besides, for the research development, it is expected to add the number of data used to give a more convincing result.

References

[1] Wahab I H A 2015 Deteksi Parasit Malaria dalam Sampel Darah Untuk Menunjang Keperluan Diagnosis Berbasis Data Ciri Tekstur Warna Disertasi (Yogyakarta: Universitas Gadjah Mada)
[2] Prevention C for D C and 2013 \textit{Plasmodium vivax}
[3] Nasir A S A, Mashor M Y and Mohamed Z 2012 Segmentation based approach for detection of malaria parasites using moving k-means clustering 2012 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) pp 653–8
[4] K. M, BJ. S, P. M and R. C 2001 The neglected burden of \textit{Plasmodium vivax} malaria Am. J. Trop. Med. Hyg. 64 97–106
[5] Tek F B, Dempster A and Kale I 2009 Computer vision for microscopy diagnosis of malaria Malar. J. 8 153
[6] Khan N A, Pervaz H, Latif A K, Musharraf A and Saniya 2014 Unsupervised identification of malaria parasites using computer vision 2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE) pp 263–7
[7] Diaz G, Gonzalez F A and Romero E 2009 A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images J. Biomed. Inform. 42 296–307
[8] Gatc J, Masiipyanti F, Sarwinda D and Aymurmyth A M 2013 \textit{Plasmodium} parasite detection on Red Blood Cell image for the diagnosis of malaria using double thresholding 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS) pp 381–5
[9] Datau E A, Gunawan S, Gunawan C A, Harijanto P N, Kristanto D, Laihad F J, Langi J, Nugroho A, Purwaningsih S, Rampengan T H, Richie T L, Sutjipto, Tambajong E H, Tjitra E and Tumewu-Wagey M 2000 \textit{Malaria: Epidemiologi, Patogenesi, Manifestasi Klinis, dan Penanganan} (Jakarta: Penerbit Buku Kedokteran EGC)
[10] Ghosh M, Das D, Chakraborty C and Ray A K 2011 \textit{Plasmodium vivax} segmentation using modified fuzzy divergence 2011 International Conference on Image Information Processing (ICIIP) pp 1–5
[11] Elter M, Hasslmeyer E and Zerfass T 2011 Detection of malaria parasites in thick blood films 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC pp 5140–4

[12] Le M-T, Bretschneider T R, Kuss C and Preiser P R 2008 A novel semi-automatic image processing approach to determine plasmodium falciparum parasitemia in giemsa-stained thin blood smears BMC Cell Biol 9 15

[13] Putra D 2010 Pengolahan Citra Digital (Yogyakarta: Penerbit Andi)

[14] Shukla A, Tiwari R, Kaur P and Janghel R R 2009 Diagnosis of Thyroid Disorders using Artificial Neural Networks Advance Computing Conference, 2009. IACC 2009. IEEE International pp 1016–20

[15] Rouhani M and Mansouri K 2009 Comparison of Several ANN Architectures on the Thyroid Diseases Grades Diagnosis International Association of Computer Science and Information Technology - Spring Conference, 2009. IACSITSC ’09 pp 526–8

[16] Isa I S, Fauzi N A, Sharif J M, Baharudin R and Abbas M H  2012 Comparisons of MLP transfer functions for different classification classes 2012 IEEE International Conference on Control System, Computing and Engineering (ICCSCE) pp 110–4

[17] Natalius S 2010 Metoda Naive Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen

[18] Bishop C M 2006 Pattern Recognition and Machine Learning (New York: Springer)

[19] Duda R O, Hart P E and Stork D G 2001 Pattern Classification and Scene Analysis (New York: John Wiley & Sons, Inc.)

[20] Ali S D and Ahuja R 2016 The Evolution and Core Concepts of Deep Learning & Neural Networks

[21] Polat K, Şahan S and Güneş S 2007 A novel hybrid method based on artificial immune recognition system ({AIRS}) with fuzzy weighted pre-processing for thyroid disease diagnosis Expert Syst. Appl. 32 1141–7

[22] Akbar S A, Nugroho H A, Wahyunggoro O and Murhandarwati E E H 2014 Segmentasi Sel Parasit Malaria dalam Sel Darah Manusia pada Ruang Warna HSV dengan K-Means Clustering Seminar Nasional Ilmu Komputer 2014 (Seminasik) vol 333–338 (Yogyakarta: Indonesian Computer Electronics and Instrumentation Support Society (IndoCEISS))

[23] Maysanjaya I M D, Nugroho H A, Setiawan N A and Murhandarwati E E H 2016 Segmentation of Plasmodium vivax phase on digital microscopic images of thin blood films using colour channel combination and Otsu method AIP Conference Proceedings vol 1755

[24] Maysanjaya I M D, Wibawa M S and Putra I M A W 2017 Pengaruh Seleksi Fitur Terhadap Hasil Klasifikasi Fase Plasmodium vivax Pada Citra Mikroskopis Digital Sediaan Darah Tipis Seminar Nasional Pendidikan Teknik Informatika (SENAPATI) Ke-8 (Denpasar: Jurusan Pendidikan Teknik Informatika) pp 96–101