Performance Analysis of Color Cascading Framework on Two Different Classifiers in Malaria Detection

Cucun Very Angkoso
Departement of Informatics
University of Trunojoyo Madura
Bangkalan, Indonesia
cucunvery@trunojoyo.ac.id

Yonathan Ferry Hendrawan
Departement of Informatics
University of Trunojoyo Madura
Bangkalan, Indonesia
yonathan.hendrawan@trunojoyo.ac.id

Ari Kusumaningsih
Departement of Informatics
University of Trunojoyo Madura
Bangkalan, Indonesia
ari.kusumaningsih@trunojoyo.ac.id

Rima Tri Wahyuningrum
Departement of Informatics
University of Trunojoyo Madura
Bangkalan, Indonesia
rimatriwahyuningrum@trunojoyo.ac.id

***Abstract*— Malaria, as a dangerous disease globally, can be reduced its number of victims by finding a method of infection detection that is fast and reliable. Computer-based detection methods make it easier to identify the presence of plasmodium in blood smear images. This kind of methods is suitable for use in locations far from the availability of health experts. This study explores the use of two methods of machine learning on Color Cascading Framework, i.e. Backpropagation Neural Network and Support Vector Machine. Both methods were used as classifier in detecting malaria infection. From the experimental results it was found that Color Cascading Framework improved the classifier performance for both in Support Vector Machine and Backpropagation Neural Network.**

Keywords*— Malaria, classifier, Color Cascading Framework, Backpropagation Neural Network, Support Vector Machine

I. INTRODUCTION

Malaria is a deadly disease. Based on World Malaria Report 2017 organized by WHO, it is stated that 91 countries reported an aggregate of 216 million cases of malaria in 2016. This number increased five million cases compared to the previous year's number. Globally, deaths from malaria reached 445,000 deaths. The number is approximately the same as reported by WHO in 2015. Although the incidence of malaria cases has declined globally since 2010, the rate of decline has stalled in several areas since 2014. The mortality rate follows a similar pattern [1]. The disease is caused by Plasmodium species, which present in the host's blood cells (erythrocytes). Some types of Plasmodium (P) are P. Malariae, P. Falciparum, P. Ovale, and P. Vivax. Each type of Plasmodium experiences different phases during their development cycle within 48 hours. Each stages has a different visual appearance that can be detected using the microscope. It has several phases of life in the human body, including ring, trophozoites, schizonts, and gametocytes [2], [3], [4]. The rapid parasite development cycle and the slow process of malaria diagnosis, resulting in high mortality due to malaria. The sooner malaria diagnosis is expected to decrease the mortality rate.

There are many methods available to diagnose malaria. The gold standard method of malaria detection in blood smear is the method where technicians or pathologists manually examine thin or/and thick blood smear under a microscope [5], [6], [7]. Other than that, malaria proteins detection [8], host/human antibodies for fighting malaria proteins detection [9], malaria gene based detection [10], [11], and microscopic cell image detection [12] was also used to diagnose malaria. Each method has its advantages and disadvantages. The gold standard method is cheap, but require experienced microscopist and certain set of tools which might be hard to find in remote areas [7]. Human inconsistencies and the difficulty of detecting low infection levels also make it worse [13], [14]. Malaria proteins detection can differentiate species infection, but it needs higher number of parasites/microliter blood sample (above 100) to be reliable compared to only four parasites/microliter for the gold standard detection performed by a skilled medical expert [5], [15]. Host antibodies detection can also discriminate different malarial infection, but the antibodies remain in the blood system after the infection has been cured; making it unreliable if used to test otherwise healthy people [15], [16]. Malaria gene detection can detect mixed infection between more than one malaria species, but needs elaborate sample preparation [10], [17], [18]. Microscopic image detection is fast and cheap method, but unable to detect drug resistance developed by the malaria and unable to capture and use malaria genomic information [7]. In this paper, we explore the detection of Malaria infections using two supervised machine learning: Backpropagation Neural Network (BPNN) and Support Vector Machine (SVM), comparing in accuracy, sensitivity and specificity. Furthermore, we investigate the effect of Color Cascading Framework (CCD-Framework) on both classifier machine.

This paper is organized as follows. Section I provides general information regarding malaria and its current detection method. Section II describes the datasets we used

The authors thank to Balai Besar Laboratorium Kesehatan (Center for Health Laboratory) Indonesian Ministry of Health Surabaya Indonesia (BBLK-dataset). We also thank to The Ministry of Research, Technology, and Higher Education of Republic Indonesia, Directorate of Research and Community Service, who funded this research through the basic research program 2018.

978-1-5386-8402-3/18/$31.00 ©2018 IEEE

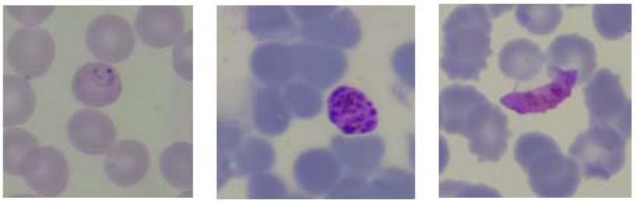
and details of the proposed methods. Next, the experimental results and analysis are presented in Section III. The conclusion of this study in Section IV.

II. MATERIAL AND METHOD

In this section, we explain the material, proposed method for detecting malaria disease on microscopic blood smear by using two supervised machine learning methods: Backpropagation Neural Networks and SVM.

A. The Material

Our study used a dataset which composed of 574 microscopy images, consisting of 287 infected images and 287 non-malaria-infected images. The dataset also consisting four types of Plasmodium (P) specifically P. Falciparum, P. Malariae, P. Ovale, and P. Vivax. Each type of Plasmodium has several phases of life including trophozoites, schizonts and gametocytes. All microscopic images of the dataset are using Red, Green, Blue (RGB) colour-space as its baseline colour and resized its dimension to 256 x 256 pixels. The sample data used for the experiment as shown in Figure 1. The data were provided by Balai Besar Laboratorium Kesehatan (Center for Health Laboratory) Indonesian Ministry of Health Surabaya Indonesia (BBLK-dataset).



(a) (b) (c)

Figure 1. The sample data of Plasmodium Falciparum, (a) trophozoites (b) schizonts (c) gametocytes

B. Method

We used Color Cascading Framework by Hendrawan et al. as a preprocessing method in preparing the images [19]. The framework involves several processing stages in a sequential step; starting from RGB normalization process, to gamma correction, noise reduction, exposure compensation, edge enhancement, Fuzzy C Means (FCM), and lastly morphological process. In this paper we only utilized the first step (RGB normalization) up to fifth step (edge enhancement). The output from that last step is used as the input for the Neural Network and the Support Vector Machine (SVM). The system's block diagram is shown in Figure 2.

Neural Network is an intelligent system composed of a collection of neurons and the most popular algorithm that has been used by researcher to solve pattern recognition problems [20], [21]. The characteristics method is to minimize the error on the output which generated by the network. The concept of the backpropagation is finding the gradient errors of the network against modified network weights. This gradient error will be used to find the weight value that will minimize the error. In general, neurons are arranged in layers of: input, hidden, and output. In our system the number of input neuron was adjusted to the dimension/size of the images as shown in Figure 3. Since the dimension of our images is 256 x 256 pixels, the input layer has 256 x 256 = 65,536 neurons. For the hidden layer, we varied the number of neurons: 10, 50, and 100 neurons. For the output screen, we used a single neuron to determine whether the image is infected with malaria or not.

SVM is an intelligent system based on supervised learning. In the training process SVM builds a model that separates each element of input data into different categories. The SVM tries to find the best classifier / hyperplane function to separate two objects. The best hyperplane is a hyperplane located halfway between two sets of objects of two classes. Finding the best hyperplane is equivalent to maximizing margins or the distance between two sets of objects from different classes. In this research we use binary classification system because the problem we want to solve only have two different classes: infected by malaria and not infected by malaria. We used the quadratic programming types of SVM.

The proposed method includes the Color Cascade Framework proposed by Hendrawan et al. which consists of: RGB normalization process, gamma correction, noise reduction, and edge enhancement [19]. Our research focus is on observing the effect of Color Cascade Framework on two different supervised machine learning: Backpropagation Neural Network (BPNN) and Support Vector Machine. The input images are the microscopic images of the blood samples.

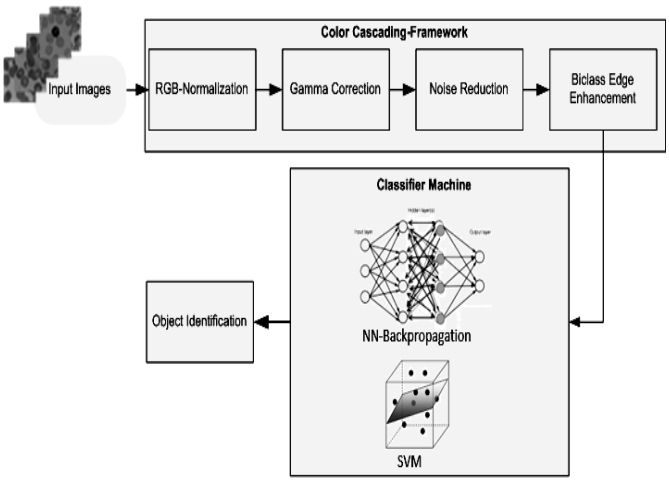


Figure 2. The block diagram of the research

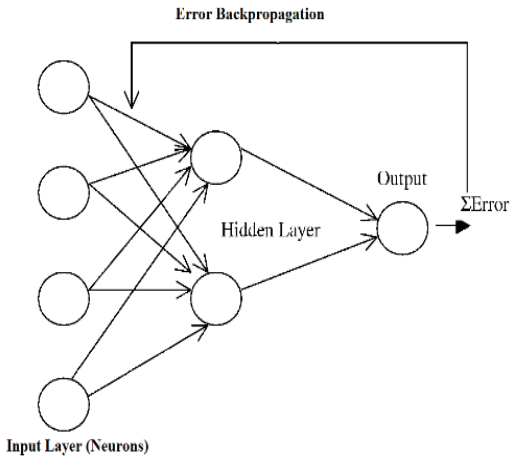


Figure 3. The illustration of BPNN

III. EXPERIMENT RESULT AND ANALYSIS

The tests were conducted on 574 images in the dataset used as training, validation, and testing data. We made two experiments. The first is using BPNN as its classifier method, and the second is using Support Vector Machine SVM).

A. Backpropagation Neural Networks

We choose Backpropagation since it is a popular training method for multi-layer neural networks and has the advantages of accuracy and versatility compared to other methods, such as perceptron. In our experiment, we used several parameters as follows:

a. Data division : Random pick
b. Training method : Scaled Conjugate Gradient
c. Training Performance measurement : Cross-Entropy
d. Maximum Epoch : 1000
e. Error target : 10^{-6}
f. Pruning strategy : None

The images were divided into three groups: 70% data as training, 15% data as validation, and 15% data as testing. In the experiment, we used two scenarios: with and without color cascading framework. We also varied the number of hidden layer neurons: 10, 50, and 100. The experiment result is shown on Table 1.

TABLE I. BACKPROPAGATION NEURAL NETWORK CLASSIFIER RESULT

Hidden Layer	Without CCD-Framework			With CCD- Framework		
	A	B	C	A	B	C
100	83.00	80.00	86.60	87.60	80.20	100.00
50	82.40	74.00	100.00	93.20	89.00	98.40
10	87.10	82.20	93.80	83.40	75.10	100.00
AVG	84.17	78.73	93.47	88.07	81.43	99.47

A= Accuracy (%), B= Sensitivity (%), C= Specificity (%)

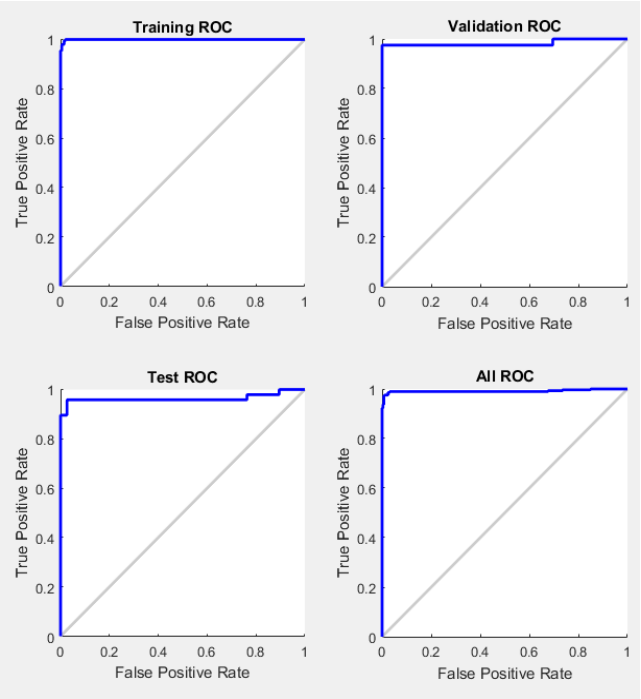


Figure 4. The receiver operating characteristic curve

From the results, it is clear that the use of color cascading framework increased the system's accuracy, sensitivity and specificity. The number of hidden layers did not affect the accuracy, sensitivity, and specificity of the system.

Figure 4 shows the classifier performance for the 50 hidden-layer neuron CCD-Framework experiment. Since the characteristic of ROC curve states that the closer the curve to the main diagonal indicates the less accurate the result, our system performance is favorable since the curve is plotted nearing the (0,1) coordinate.

B. Support Vector Machine (SVM)

In the SVM experiment, we divided the data randomly into two parts: 70% data as training and 30% data as testing. As in the previous experiment, we also used two scenarios: with and without color cascading framework. We conducted three tests in this experiment. The SVM experiment result is shown on Table 2.

TABLE II. SUPPORT VECTOR MACHINE CLASSIFIER RESULT

SVM	Without CCD-Framework			With CCD- Framework		
	A	B	C	A	B	C
Test-1	91.70	92.50	90.00	97.70	95.50	100.00
Test -2	90.90	90.30	87.60	95.90	94.30	97.60
Test -3	91.50	92.70	87.50	96.50	95.70	97.50
AVG	91.37	91.83	88.37	96.70	95.17	98.37

A= Accuracy (%), B= Sensitivity (%), C= Specificity (%)

From the result, it is also clear that the use of color cascading framework raised the system's accuracy, sensitivity, and specificity of the SVM classifier.

Figure 5 illustrates the two experiments' results. The graph presents the average accuracy, sensitivity, and specificity of 10, 50, and 100 hidden layer neurons from the BPNN experiment. In table 1, the numbers are located in the last row. It also displays the result of the second experiment in the form of average results of the three tests. As in the previous, the numbers are situated in the last row of Table 2.

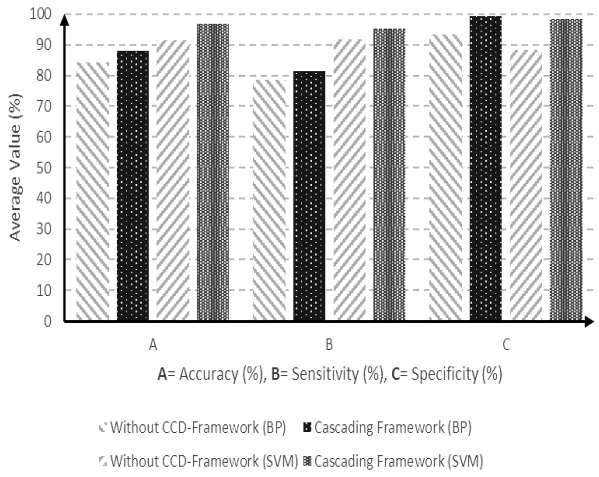


Figure 5. The comparison of experimental result: with and without using Color cascading framework

The system's average value of performance on all classification methods increases with the addition of CCD-Framework. From the results, we can also observe that SVM classifier generates higher performance numbers compared to BPNN classifier in both with and without using CCD-Framework. SVM classifier also has the highest accuracy value (96.70%). It is obtained in the SVM classifier + CCD-Framework test.

In [19], Hendrawan et al used Fuzzy C-Means (FCM) as the CCD Framework's classifier. They achieved their highest results of 98.26% for Accuracy, 97.91% for Specificity, and 98.61% for Sensitivity when they used CCD Framework. Without the framework, they acquired significantly worse results. Whereas for this research, the best results are attained using SVM classifier with CCD framework, that are 96.70%, 95.17%, and 98.37% for accuracy, sensitivity, and specificity. If we compare them, FCM's results are better for all criterions. However, the difference is not large: less than 3% for each measurement component.

IV. CONCLUSIONS

The use of CCD-framework has been shown to provide better classifier's performance results. The highest classification results were obtained when applying CCD-Framework with SVM classifier. Its accuracy, sensitivity, and specificity are 96.70%, 95.17%, and 98.37% respectively. So it can be concluded that CCD-Framework is more suitable to be applied in conjunction with SVM classifier in detecting malaria infection in blood smear images. These results are slightly less (<3% for each of the three criterions) than the results of previous study which used Fuzzy C-Means as the classifier. For future studies, we plan to classify the malaria infection into its development phases and species. Correctly identifying the infecting plasmodium species is crucial in determining the appropriate treatment, since it can increase the healing rate.

REFERENCES

[1] "World Health Organization, World Malaria Report 2017." [Online]. Available: http://www.who.int/malaria/publications/world-malaria-report-2017/en.

[2] H. Hisaeda, K. Yasutomo, and K. Himeno, "Malaria: immune evasion by parasites," The International Journal of Biochemistry & Cell Biology, vol. 37, no. 4, pp. 700–706, 2005.

[3] M. Poostchi, K. Silamut, R. Maude, S. Jaeger, and G. Thoma, "Image analysis and machine learning for detecting malaria," Translational Research, 2018.

[4] R. T. Wahyuningrum and A. K. Indrawan, "A hybrid automatic method for parasite detection and identification of plasmodium falciparum in thin blood images.," International Journal of Academic Research, vol. 4, no. 6, 2012.

[5] M. M. Kettelhut, P. L. Chiodini, H. Edwards, and A. Moody, "External quality assessment schemes raise standards: evidence from the UKNEQAS parasitology subschemes," Journal of clinical pathology, vol. 56, no. 12, pp. 927–932, 2003.

[6] 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.

[7] P. Gascoyne, J. Satayavivad, and M. Ruchirawat, "Microfluidic approaches to malaria detection," Acta tropica, vol. 89, no. 3, pp. 357–369, 2004.

[8] M. T. Makler et al., "Parasite lactate dehydrogenase as an assay for Plasmodium falciparum drug sensitivity," The American Journal of Tropical Medicine and Hygiene, vol. 48, no. 6, pp. 739–741, 1993.

[9] O. Silvie et al., "Potential value of Plasmodium falciparum-associated antigen and antibody detection for screening of blood donors to prevent transfusion transmitted malaria," Transfusion, vol. 42, no. 3, pp. 357–362, 2002.

[10] M.-A. Lee et al., "Real-time fluorescence-based PCR for detection of malaria parasites," Journal of clinical microbiology, vol. 40, no. 11, pp. 4343–4345, 2002.

[11] K. Ganesan, L. Jiang, and P. K. Rathod, "Stochastic versus stable transcriptional differences on Plasmodium falciparum DNA microarrays." Elsevier, 2002.

[12] J. Somasekar and B. Eswara Reddy, "Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging," Computers and Electrical Engineering, vol. 45, pp. 336–351, 2015.

[13] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Analysis of infected blood cell images using morphological operators," Image and vision computing, vol. 20, no. 2, pp. 133–146, 2002.

[14] G. Diaz, 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.

[15] J. Iqbal, N. Khalid, and P. R. Hira, "Performance of rapid malaria Pf antigen test for the diagnosis of malaria and false-reactivity with autoantibodies," in Tropical Diseases, Springer, 2003, pp. 135–148.

[16] A. Irion, H.-P. Beck, and T. Smith, "Assessment of positivity in immuno-assays with variability in background measurements: a new approach applied to the antibody response to Plasmodium falciparum MSP2," Journal of immunological methods, vol. 259, no. 1–2, pp. 111–118, 2002.

[17] P. L. Blair, A. Witney, J. D. Haynes, J. K. Moch, D. J. Carucci, and J. H. Adams, "Transcripts of developmentally regulated Plasmodium falciparum genes quantified by real-time RT--PCR," Nucleic acids research, vol. 30, no. 10, pp. 2224–2231, 2002.

[18] C. H. Lai, S. S. Yu, H. Y. Tseng, and M. H. Tsai, "A protozoan parasite extraction scheme for digital microscopic images," Computerized Medical Imaging and Graphics, vol. 34, no. 2, pp. 122–130, 2010.

[19] Y. F. Hendrawan, C. V. Angkoso, and R. T. Wahyuningrum, "Colour image segmentation for malaria parasites detection using cascading method," in Sustainable Information Engineering and Technology (SIET), 2017 International Conference on, 2017, pp. 83–87.

[20] J. Tarigan, Nadia, R. Diedan, and Y. Suryana, "Plate Recognition Using Backpropagation Neural Network and Genetic Algorithm," 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI, vol. 116, pp. 365–372, 2017.

[21] Z. Chiba, N. Abghour, K. Moussaid, A. El Omri, and M. Rida, "A novel architecture combined with optimal parameters for back propagation neural networks applied to anomaly network intrusion detection," Computers & Security, vol. 75, pp. 36–58, 2018.