Classification of Plasmodium Skizon and Gametocytes Malaria Images Using Deep Learning

1st Yessi Jusman
Department of Electrical Engineering, Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia
yjusman@umy.ac.id

2nd Indah Monisa Firdiantika
Department of Electrical Engineering, Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

3rd Slamet Riyadi
Department of Electrical Engineering, Faculty of Engineering
Universitas Muhammadiyah Yogyakarta
Yogyakarta, Indonesia

4th Siti Nurul Aqmariah Mohd Kanafiah
School of Mechatronics Engineering, Universiti Malaysia
Perlis, Malaysia

5th Rosline Hassan
Department of Haematology, Universiti Sains Malaysia
Kelantan, Malaysia

6th Zeehaida Mohamed
Department of Microbiology and Parasitology, Universiti Sains Malaysia
Kelantan, Malaysia

Abstract— **Identification analysis of the malaria parasite cell infection, there is a possibility of human error factor done by paramedics because of the number of samples that must be analyzed. This case is because the human eye tends to be tired while working continuously, which can lead to misclassification and treatment that is not right. Therefore, it takes a computer-based system that facilitates medical expert or laboratory technician in identifying two types of parasite cells namely Plasmodium skizon and Plasmodium gametocytes to reduce instances of human error. This research will be conducted on computer-based identification by processing the image type of plasmodium malariae consists of two types, namely Plasmodium skizon and Plasmodium gametocytes levels using convolutional neural network with VGG-16 pre-trained model using 13 layers and 2 dense layers. This study applied 5-fold cross validation for datasets and the datasets are tested using 4 level epoch nodes. The results showed the success of the classification results which have highest training accuracy 90% as well as the results of the highest testing accuracy 100%. It showed the classification using CNN VGG-16 pre-trained model successfully classified the malaria type images.**

Keywords— ***Malaria, Skizon, Gametocytes, Plasmodium, Deep Learning***

I. INTRODUCTION

Malaria is a contagious infectious disease that is spread through mosquito bites. Malaria sufferers will complain of fever and chills. Although easily transmitted through mosquito bites, malaria can be completely cured if handled properly. However, if left untreated, this disease can be fatal, from causing severe anemia, kidney failure, to death. Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, and Plasmodium malaria are the four species of plasmodium that can infect human blood cells.

The parasite's species and development phase were identified using clinical diagnostic procedures. Identifying the parasite's genus and development phase [1]. There are still some differences at the time of identification; usually, small-scale differences in features tend to be overlooked by the human eye. To avoid this, computer-based automatic screening methods that examine abnormalities that fail to make decisions should be developed.

Much research has gone into developing intelligent computational tools known as Computer-Aided Design (CAD), which uses an image processing technology to improve the diagnostic quality of paramedics. Some study has been done on the use of the watershed thresholding approach for imaging in some circumstances, as demonstrated in [1], [2], [3], [4], [5].

Image that has been improved by preprocessing with red blood cell (RBC) count using Hough transform then segmented with Annular Ring Ratio (ARR), determined features using several methods with various characteristics ranging from size using dilation and erosion, shape using Hough transform, intensity using ARR transform, texture using co-occurrence matrix (contrast, entropy, energy), and statistical features using skewness, kurtosis, standard deviation. The results of the study reached more than 95% [6].

In 2015, researchers used the Global Contract Streatching technique in the preprocessing stage. There are 100 malaria images obtained from the Department of Microbiology and Parasitology, Universiti Sains Malaysia Hospital (HUSM) and Universiti Kebangsaan Malaysia Hospital (HUKM). Malaria images were segmented using Cascaded moving K-means and Fuzzy C-Means clustering methods. Researchers also used the R-Y, B-Y, modified B-Y and intensity methods to determine features for infection detection in the malaria images. As a result, the study showed very good accuracy, sensitivity, and specificity values, namely 99.37%, 88.22%, and 99.82%. For the next research can be developed on the modification of B-Y in order to produce better segmentation [2].

The same researcher developed the research results in 2018 by preprocessing images using the partial contrast stretching technique, median filter and seeded region growing area extraction method. Segmentation with MKM clustering and extraction using a region filling algorithm was also carried out. The results showed an increase in accuracy from previous studies to 99.49% [7].

In 2017, researchers improved their malariae images using the Laplace'equation and Laplacian values

preprocessing methods for image denoising. The segmentation process was also carried out using the adaptive histogram thresholding technique on thin smear and thick smear images. The feature extraction process is carried out using the unsupervised learning technique method. With the deviated calculation obtained 2.17% and 10.12% with the algorithm for 5.23 s and 4.78 s [8].

Other research performed the feature extraction stage using the Concatenated Feature of Color Histogram (H), Local Area Gran Ulometry (G), and Shape Measurements Vector (R) methods. The classification results show a very small error of 0.01% [9]. In 2019, researchers classified malaria images using a feature extraction method by combining texture and color features [10].

In this study, it is hoped that it can be developed even better by having the right feature extraction and adding more types of malariae images so that the deviated results are better. Based on the background and references, this study aims to identify the malaria parasite that infects red blood cells using a deep learning-based classification method with the species plasmodium falciparum, vivax, malaria, and as research material because this species can cause death if not properly handled.

II. METHODOLOGY

A. Research Concept

In image classification research, the system will be designed using a deep learning method. The research flow are started by read the original images of microscopic images which are obtained from Universiti Sains Malaysia Hospital. The images are validated by the pathologist as two classes namely scizo and gamet images. The images are captured by cytologist and pathologist with resolution size of 1288x966.

The architecture used is the Convolutional Neural Network with pre-trained VGG-16 model. This model is often used and to perform good classification results for many types of datasets. This system is built using Jupyter Notebook with Keras as the main library and python programming language. This method will be used to compare the accuracy values against five different datasets.

B. Hardware and Software Requirements

In making this system, hardware and software are needed. Hardware is used to build systems and test systems that have been built using a computer. The software consists of integrated development systems (IDE). The computer specifications used to create the system can be seen in Table 1. The software used can be seen in Table 2.

TABLE I.   HARDWARE SPECIFICATIONS

Processor	Intel Core i7 4600U
Speed	3.3 GHz
VGA	Intel HD Graphics 4400
Screen size	14 inch
RAM	12 GB DDR3

For hardware specifications, we used processor intel core i7 4600U, speed 3,3 GHz, and intel HD graphics 4400. The used software of the system is window 10 64 bit operating system.

TABLE II.   SOFTWARE

Operating system	Windows 10 64 bit
Language compiler	Phyton
Compilers	IDLE Python version 3.6.8
Library	Keras
IDE	Jupyter Notebook

C. Image Data

This research used image data of malaria parasite (gamet and scizo). The data obtained were 180 images consisting of 90 gamet images and 90 scizo images. The data used was divided into two folders, namely training data and test data with a ratio of 8:2, respectively.

For training data, 144 images are consisting of 72 gamete images and 72 scizo images. While the test data, 36 images are consisting of 18 gamet images and 18 scizo images. The image data used is in JPG file format with a size of 1288x966, which is reduced to 600x450 pixels. Image data is in the form of RGB images. Gamet and scizo images and seen in Figures 1 and 2. The division of the dataset is randomized into five datasets which can be seen in Table 3.
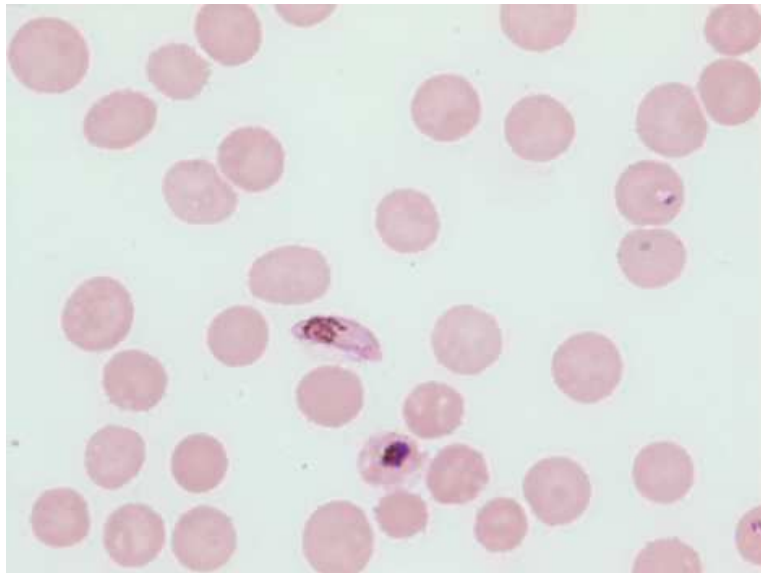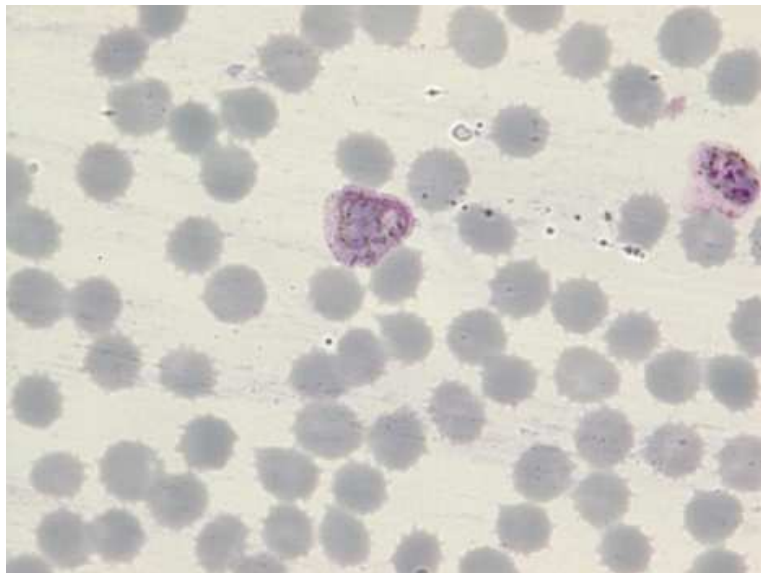
Fig. 1. Gamet Image

Fig. 2. Scizo Image

TABLE III. MANAGEMENT OF DATASET

Data set	Training	Testing
Data set 1	data 1, data 2, data 3, data 4	data 5
Data set 2	data 1, data 2, data 3, data 5	data 4
Data set 3	data 1, data 2, data 4, data 5	data 3
Data set 4	data 1, data 3, data 4, data 5	data 2
Data set 5	data 2, data 3, data 4, data 5	data 1

D. Methods

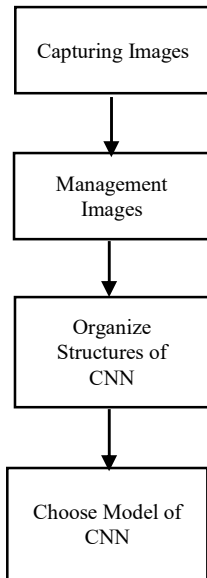In this research, several stages are needed in conducting research which can be seen in Figure 3.

Fig. 3. The Flow of Research

The image data collection or dataset used is the malaria parasite image of gametes and scizo. The dataset is divided into two folders with the names of the train and test folders. The system design uses Jupyter Notebook, Python Programming Language, and Keras Library. The architectural model used is the VGG-16 pre-trained model. After completing the design, the next stage is testing. Testing is seen from the accuracy value obtained during the training and testing process to obtain the results of a 1-5 dataset comparison.

E. Analysis Methods

The method of analysis used in this research is quantitative analysis, which is carried out by analyzing the classification results of the malaria parasite image system in the form of accuracy values. The resulting classification outputs are in the form of gamet and scizo image grouping.

F. Architecture Model

The deep learning method uses a pre-trained model or what is called transfer learning. The architecture used is a convolutional neural network with a pre-trained VGG-16 model. The VGG-16 is a pre-trained model who has been trained on ImageNet dataset.
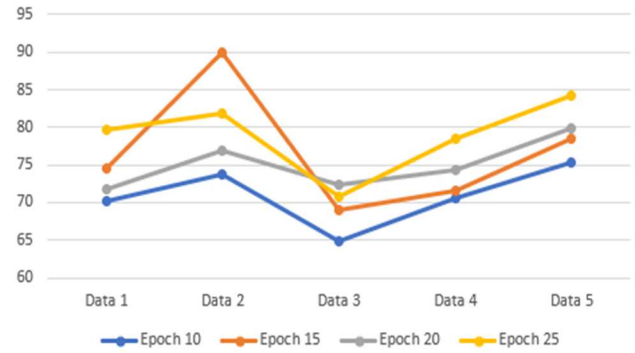

Fig. 5. Training Accuracy

The transfer learning method uses the pre-trained VGG-16 model, which has 16 layers. This architectural model uses 13 convolutional layers from VGG-16 and 3 dense layers frozen and replaced with two dense layers.

III. METHODOLOGY

At this stage, several tests will be discussed to classify gamet and scizos using a convolutional neural network with a pre-trained VGG-16 model. The percentage of training accuracy can be seen in Table 4.

TABLE IV. TRAINING ACCURACY

	Epoch 10	Epoch 15	Epoch 20	Epoch 25
Data 1	70.12	74.56	71.79	79.63
Data 2	73.67	73.86	76.82	81.86
Data 3	64.82	69.02	72.42	70.78
Data 4	70.56	71.64	74.29	78.41
Data 5	75.25	78.46	79.81	84.2

The information in table 4 shows the results of various training accuracy on data 1, data 2, data 3, data 4, data 5 on epoch 10, epoch 15, epoch 20, epoch 25. In data 2, data 4, and data 5 can be seen that the more epochs, the higher the training accuracy value is obtained. However, in data 1 and data 3, the value of training accuracy fluctuates against the number of epochs, meaning that the number of epochs used does not have a major effect on the accuracy value. The graph of the training accuracy value can be seen in Figure 5.

The test accuracy is obtained based on the percentage of the number of data classified correctly to all the number of test data. The test accuracy can be seen in Table 5.
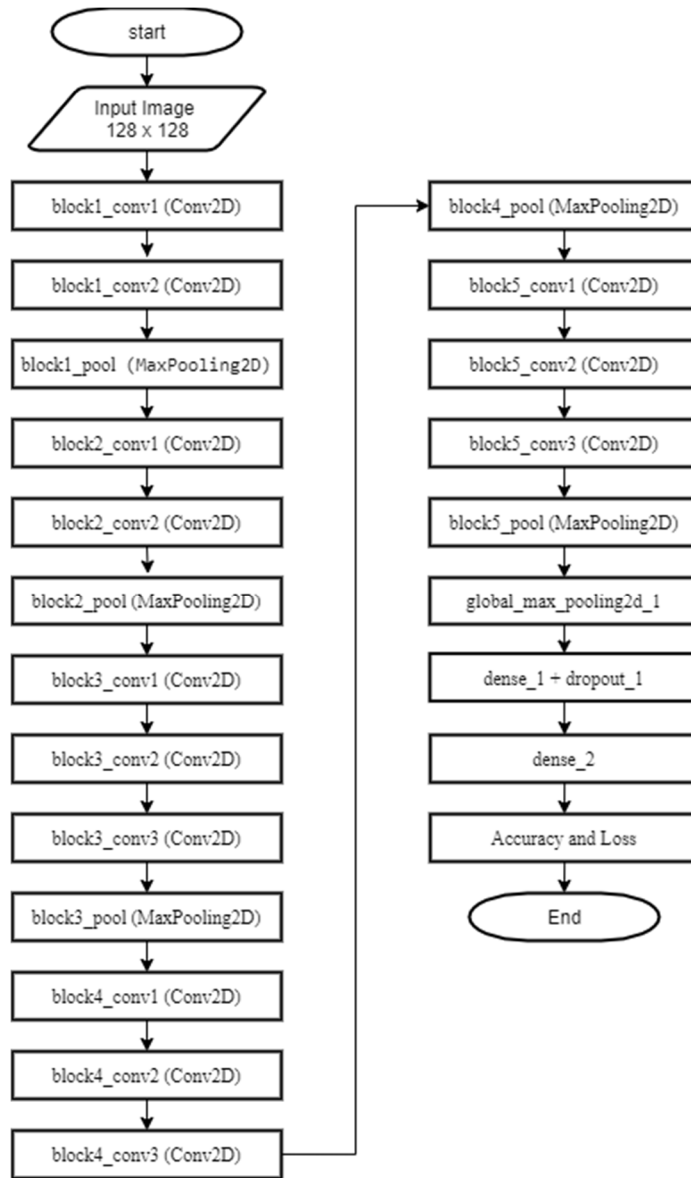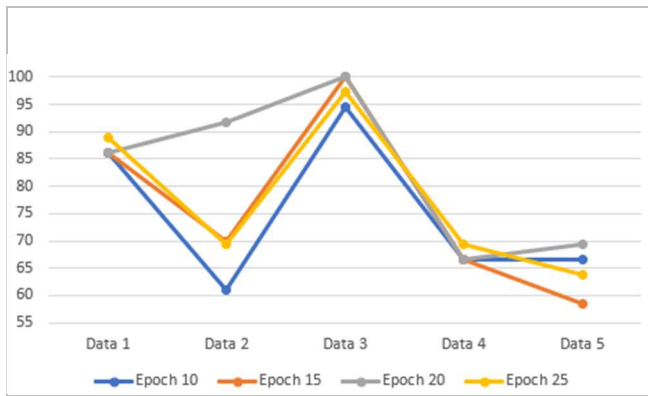
Fig. 4. Convolutional Neural Network Architecture

TABLE V.   TESTING ACCURACY

	Epoch 10	Epoch 15	Epoch 20	Epoch 25
Data 1	86.11	86.11	86.11	88.89
Data 2	61	70	91.67	69.44
Data 3	94.44	100	100	97.22
Data 4	66.67	66.67	66.67	69.44
Data 5	66.67	58.33	69.44	63.88

The information in Table 5 shows different accuracy results, data 1 and data 3 get high accuracy values for all the number of epochs. While data 4 and data 5 get almost the same test accuracy value in each epoch and data 2 experience fluctuating test accuracy values for several epochs. The test accuracy graph can be seen in Figure 6.

IV. CONCLUSSIONS

This research was conducted on computer-based identification by processing the image type of plasmodium malariae consists of two types, namely Plasmodium skizon and Plasmodium gametocytes levels using convolutional neural network with VGG-16 pre-trained model using 13 layers and 2 dense layers. This study applied 5-fold cross validation for datasets and the datasets are tested using 4 level epoch nodes. The results showed the success of the classification results which have highest training accuracy 90% as well as the results of the highest testing accuracy 100%. It showed the classification using CNN VGG-16 pre-trained model successfully classified the malaria type images.

REFERENCES

[1] Loddo, A., C. Di Ruberto, and M. Kocher, Recent Advances of Malaria Parasites Detection Systems Based on Mathematical Morphology. Sensors, 2018. **18**(2): p. 513.

[2] Abdul-Nasir, A.S., M.Y. Mashor, and Z. Mohamed. Segmentation of Malaria Parasite Based on Stained Blood Cells Detection. in Journal of Biomimetics, Biomaterials and Biomedical Engineering. 2015. Trans Tech Publ.

[3] Nasir, A.S.A., et al. The Cascaded Enhanced k-Means and Fuzzy c-Means Clustering Algorithms for Automated Segmentation of Malaria Parasites. in MATEC Web of Conferences. 2018. EDP Sciences.

[4] Abdul-Nasir, A.S., M.Y. Mashor, and Z. Mohamed. Clustering approach for unsupervised segmentation of malarial Plasmodium vivax parasite. in AIP Conference Proceedings. 2017. AIP Publishing.

[5] Jusman, Y., et al., Comparison of Malaria Parasite Image Segmentation Algorithm Using Thresholding and Watershed Method. Journal of Physics: Conference Series, 2021. **1783**: p. 012092.

[6] Charpe, K.C. and V. Bairagi. Automated malaria parasite and there stage detection in microscopic blood images. in 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO). 2015. IEEE.

[7] Nasir, A.S.A., M.Y. Mashor, and Z. Mohamed, *Enhanced k-Means Clustering Algorithm for Malaria Image Segmentation.* Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 2018. **42**(1): p. 1-15.

[8] Dave, I.R. and K.P. Upla. Computer aided diagnosis of Malaria disease for thin and thick blood smear microscopic images. in 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN). 2017.

[9] Tek, F.B., A.G. Dempster, and İ. Kale, *Parasite detection and identification for automated thin blood film malaria diagnosis.* Computer Vision and Image Understanding, 2010. **114**(1): p. 21-32.

[10] Bashar, M.K. Improved Classification of Malaria Parasite Stages with Support Vector Machine Using Combined Color and Texture Features. in 2019 IEEE Healthcare Innovations and Point of Care Technologies,(HI-POCT). 2019. IEEE.