Implementation of Convolutional Neural Network in the classification of red blood cells have affected of malaria

Mawaddah Harahap[1]*, Jefferson[2] Surya Barti[3], Suprianto Samosir[4], Christi Andika Turnip[5]
[1)2)3)4)5)] Universitas Prima Indonesia, Indonesia
[5)] mawaddah@unprimdn.ac.id, [2)]*jeffersonfebruari2404@gmail.com, [3)] suryabarti29@gmail.com,
[4)]suprianto19samosir97@gmail.com, [5)] christiandika13@gmail.com

Abstract: Malaria is a disease caused by plasmodium which attacks red blood cells. Diagnosis of malaria can be made by examining the patient's red blood cells using a microscope. Convolutional Neural Network (CNN) is a deep learning method that is growing rapidly. CNN is often used in image classification. The CNN process usually requires considerable resources. This is one of the weaknesses of CNN. In this study, the CNN architecture used in the classification of red blood cell images is LeNet-5 and DRNet. The data used is a segmented image of red blood cells and is secondary data. Before conducting the data training, data pre-processing and data augmentation from the dataset was carried out. The number of layers of the LeNet-5 and DRNet models were 4 and 7. The test accuracy of the LeNet-5 and DrNet models was 95% and 97.3%, respectively. From the test results, it was found that the LeNet-5 model was more suitable in terms of red blood cell classification. By using the LeNet-5 architecture, the resources used to perform classification can be reduced compared to previous studies where the accuracy obtained is also the same because the number of layers is less, which is only 4 layers.

Keywords: Classification ,Convolutional Neural Network, Deep Learning, Malaria, Red Blood Cell

INTRODUCTION

Malaria is a disease caused by plasmodium. This disease is usually transmitted through Anopheles mosquito bites, blood transfusions or injections that are used interchangeably. This disease attacks the blood cells that cause fever, fatigue, mutilation, and headaches. If this disease is not treated quickly, you will experience complications such as seizures, impaired consciousness, shortness of breath and death. A malaria diagnosis can be done by examining the patient's blood in the laboratory. Examination of patient blood in the laboratory can take a long time due to the small size of the plasmodium and the number of other objects that are similar to malaria parasites (Autino et al. 2012).

One way to speed up blood examination is by image classification of red blood cells (Poostchi et al. 2018). Several methods that have been used to classify red blood cell images are K-Nearest Neighbor (KNN) (Nanoti et al. 2016), Learning Vector Quantization (LVQ) (Rahmanti et al. 2014), Support Vector Machine (SVM) (Dave 2018), Artificial Neural Network (Sankaran et al. 2017), Random Forest (Singh et al. 2020), and others. One of the classification methods currently being developed is the Convolutional Neural Network (CNN).

CNN is a Deep Learning method and is a development of the Multilayer Perceptron (MLP) method. The use of CNN has been frequently used in medicine to classify images of diseases such as: diabetic retinopathy (Ignatius et al. 2019), chest disease (Abiyev and Ma'aitah 2018), maize disease (Hidayat, Darusalam, and Irmawati 2019), and others (Yadav and Jadhav 2019). CNN has been shown to classify red blood cell images with sufficient accuracy compared to other image classification methods (Liang et al. 2017). In previous studies, the CNN architecture was used to classify red blood cells quite deeply (Yohannes, Devella, and Arianto 2020).

This study will implement a Convolutional Neural Network (CNN) to classify red blood cells affected by malaria using an architecture that is not deep enough. By conducting this research, it is hoped that the use of CNN can classify red blood cells that have parasites and do not have plasmodium parasites

*name of corresponding author

LITERATURE REVIEW

In the case of malaria, several methods have been used to solve the problem of image classification of red blood cells such as K-Nearest Neighbor (KNN) (Nanoti et al. 2016), Learning Vector Quantization (LVQ) (Rahmanti et al. 2014), Support Vector Machine (SVM) (Dave 2018), Artificial Neural Network (Rahmanti et al. 2014), Random Forest (Singh et al. 2020) and others. The CNN method has been widely used to solve this case. From the research conducted by Yadas SS and Jadhav SM, it was explained that the CNN method has better accuracy than other image classification methods (Liang et al. 2017).

In a study conducted by Zhaohui Liang et al, prior to training, researchers performed image preprocessing such as increasing local brightness and contrast, whitening the entire image using the eigenvalues decomposition (EVD) operation on the covariance matrix. The CNN architecture used has a convolutional layer and 3 hidden layers. Architectural testing of the test data yields an accuracy of 97% (Liang et al. 2017).

In a study conducted by Yohannes et al, a CNN with a deep architecture was used. Researchers also combine the saliency method to perform classification. The accuracy obtained was 95% (Yohannes, Devella, and Arianto 2020).

In a study conducted by Soner Can Kalkan and Ozgur Koray Sahingoz, researchers used a self-constructed CNN architecture to classify red blood cells. The architecture used has 3 convolitional layers and 1 hidden layer. Architectural testing of the test data yields an accuracy of 95% (Kalkan and Sahingoz 2019).

In a study conducted by Aimon Rahman et al, researchers conducted preprocessing and data augmentation. The preprocessing of data is image normalization, rescaling, standardization, while data augmentation is done by reversing the image horizontally, vertically, performing filtering, rotation, and so on. The architecture used is VGG16 and a custom built architecture consisting of 8 convulsion layers and 2 hidden layers. Architectural testing of the test data yields a fairly good accuracy, namely above 90% (Rahman et al. 2019).

In a study conducted by Jaspreet Singh Chima et al, researchers conducted preprocesses such as dividing data, changing image sizes, and changing categories of image classes. The architecture used in this study is Resnet-50. Architectural testing of the test data yields a fairly good accuracy that is above 95% (Chima et al. 2020).

In a study conducted by Nwankpa C et al, researchers used several architectures such as Alex-net, VGG-16, Resnet-50, Xception, DenseNet-121, and self-created architectures. The results of this study Resnet-50 has a fairly good performance.

METHOD

In this study, the research procedures are as follows:

Data Collection

The data used in this study is secondary data that can be accessed publicly. The red blood cell data can be downloaded via the website at https://ceb.nlm.nih.gov/repositories/malaria- datasets / (Rajaraman et al. 2018). The data used in this study consisted of 27,558 images of red blood cells. From these data, there are 13,775 images of red blood cells infected with plasmodium parasites and 13,813 images of red blood cells that are not infected with plasmodium parasites.
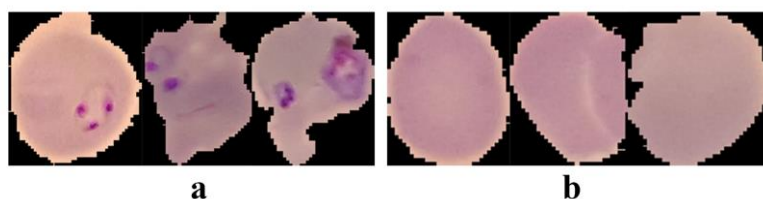

Fig. 1 Normal Red Blood Cells (a) Malaria Red Blood Cells (b)

Preprocessing and Augmentation

Before doing the training process, the image from the red blood cell data will go through preprocessing and augmentation. The preprocessing that will be carried out on the image is as follows:

a. Change the image size

In this process, the entire image will be converted into the same size. The size that is changed will match the size of the input in the Convolutional Neural Network architecture.

b. Rescaling.

In this process, each pixel of the image will be divided by 255 to produce a value between 0 and 1. This process aims to accelerate convergence during training (Rahman et al. 2019).

c. Grouping Data

In this process, the data will be grouped into 2 groups, namely training and testing. The ratio used to separate this data is 8: 2.

Meanwhile, the augmentation process carried out in this research is rotation, reversing horizontally and vertically, and cutting the image.

Training

In this process, data from the training group will be trained using a Convolutional Neural Network architecture. Due to limited resources, the architecture used in this study will not be very deep. The following is the architecture that will be used in this study:

a. LeNet-5

The architecture of this model consists of 2 convolution layers followed by an Average Pooling layer, and 2 hidden layers. The architecture uses an input size of 224 pixels x 224 pixels (Ignatius et al. 2019).
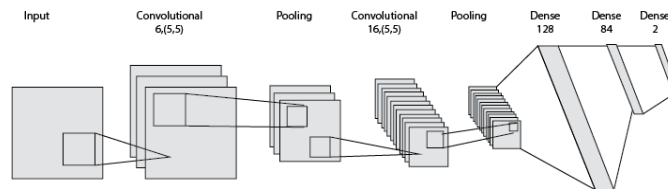
Fig. 2Lenet-5 Architecture

b. DRNet

The architecture of this model consists of 4 layers of convolution followed by the Average Pooling layer, and 2 hidden layers [8]. The architecture uses an input size of 128 pixels x 128 pixels (Ignatius et al. 2019).
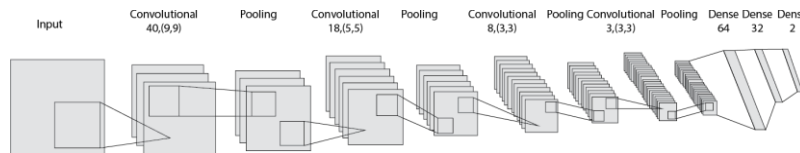
Fig. 3 DrNet Architecture

1. Testing

In this process, the trained model will be tested on the test data group. The test has to analyze whether the trained model is good enough in classifying the cases that are being resolved.

RESULT

Preprocesssing and Augmentation

Before conducting data training, preprocessing and augmentation were carried out on the dataset. Some of the preprocessing methods used in this study are as follows:

a. Change Image Size

In this process the various image sizes are converted into one size, namely 224 x 224pixel and 128 x 128 pixels. The following is an example of the result of resizing the image where the original image has a size of 127 x151 pixels which is converted to 128 x 128 pixels.
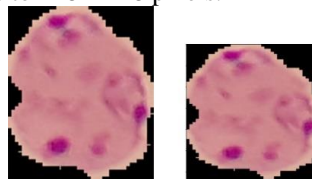

Fig. 4 Original Image and Changed Image Size

b. Rescaling

In this process, the entire image is normalized where each pixel of the image is divided by 255. The value of each image pixel is between 0-1.

c. Splitting data

In this process, the dataset is divided into 2 groups, namely the training and testing group. With the use of a ratio of 8: 2, the training group will use 17,638 data and the test group will use 4,408 data.

Several augmentation methods used in this study are rotation, horizontal and vertical flipping, and image cropping. The rotation carried out in this study was 45 degrees.
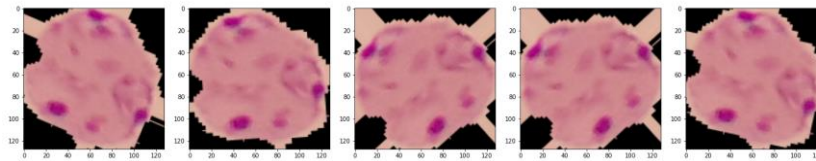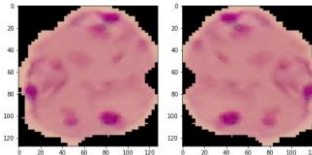
Fig. 5 Image Rotation
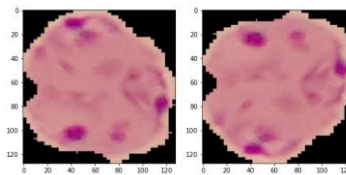

Fig. 6 Horizontal Flipping

Fig. 7 Vertical Flipping

Training

The number of parameters trained in the Lenet-5 model is 1,628,046 while in the DrNet model is 63,959. The training is planned to be carried out for 100 iterations. Training will be stopped when the model has overfitting. Here is a graph of accuracy and loss in the training process.
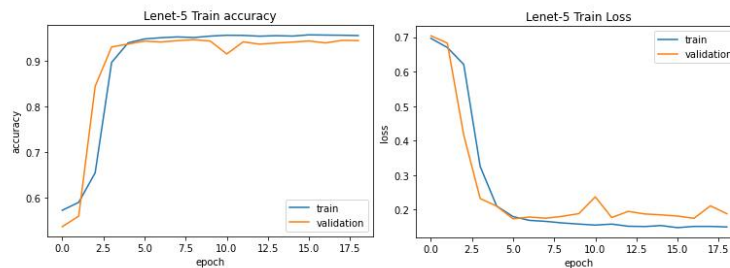

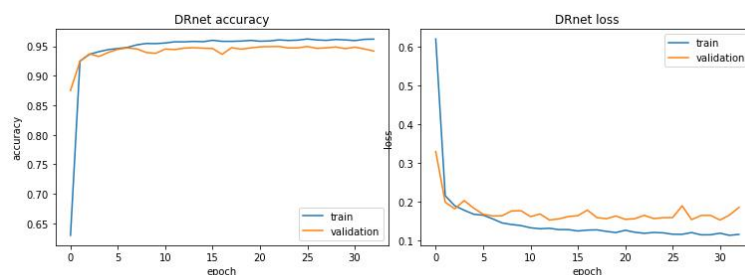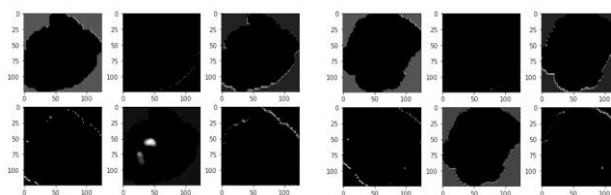
Fig. 8 Lenet-5's Training Accuracy and Loss

Fig. 9 DRNet's Training Accuracy and Loss

The following is a visualization of the feature extraction of red blood cell images obtained in the convolution layer during training in the convolution layer:

    a.   LeNet-5

Terinfeksi Malaria   Tidak Terinfeksi Malaria

Fig. 10 Feature Extraction from Blood Cell Image in 1st Convolution Layer
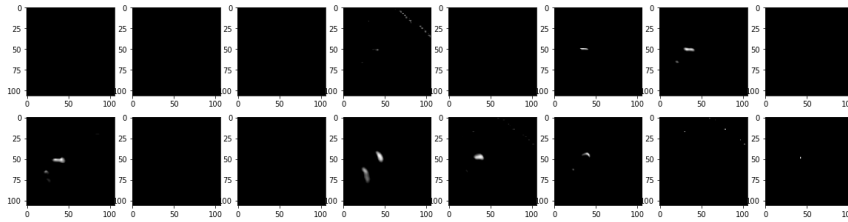

Fig. 11 Feature Extraction on Malaria Blood Cell Image in 2nd Convolution Layer.
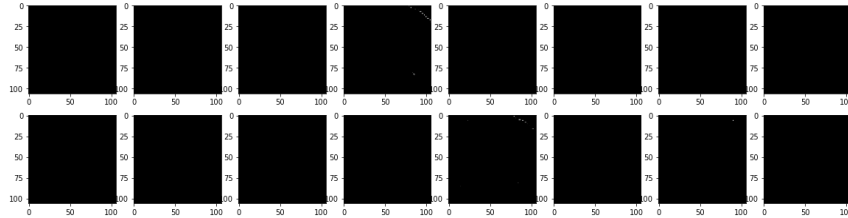

Fig. 12 Feature Extraction on Normal Blood Cell Image in 2nd Convolution Layer.

Based on the figure above, the features that distinguish red blood cells in feature extraction that occur in the LeNet model convolution layer are white spots. Where the red blood cells infected with malaria have white spots, it is seen in Figures 10a and 11 whereas in Figures 10b and 11 they do not have white spots.
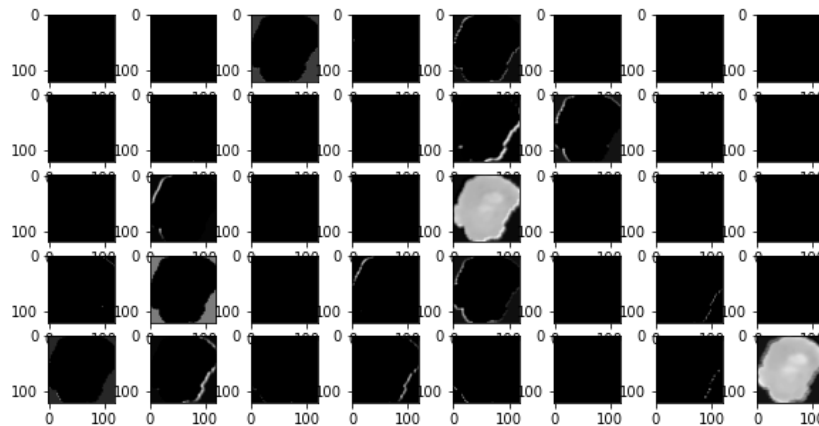
b. DRnet

Fig. 13 Feature Extraction on Normal Blood Cell Image in 1st Convolution Layer.
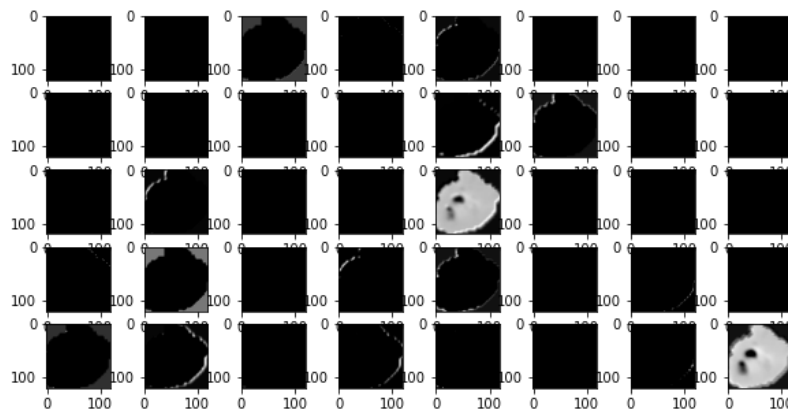

Fig. 14 Feature Extraction on Malaria Blood Cell Image in 2nd Convolution Layer.
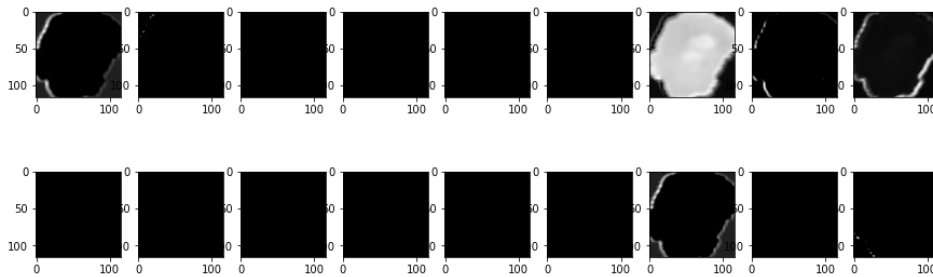
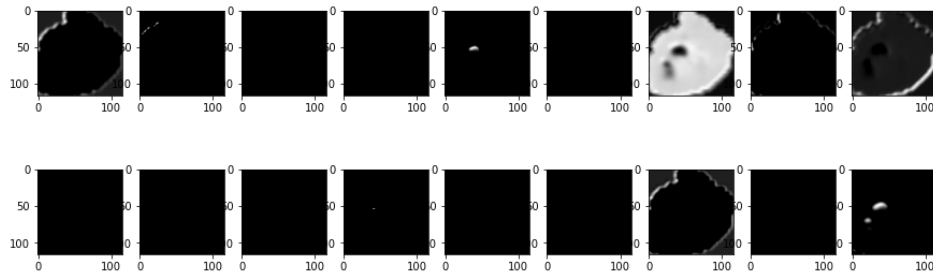Fig. 15 Feature Extraction on Normal Blood Cell Image in 2nd Convolution Layer.

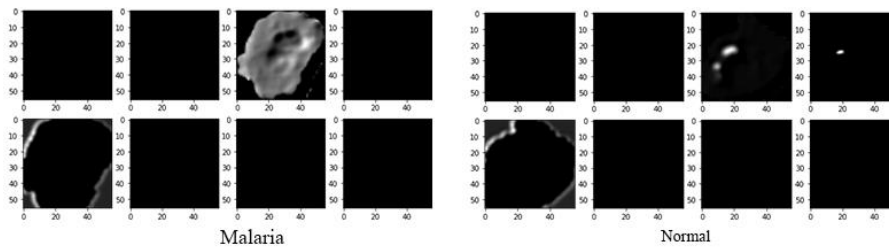Fig. 16 Feature Extraction on Malaria Blood Cell Image in 2nd Convolution Layer.

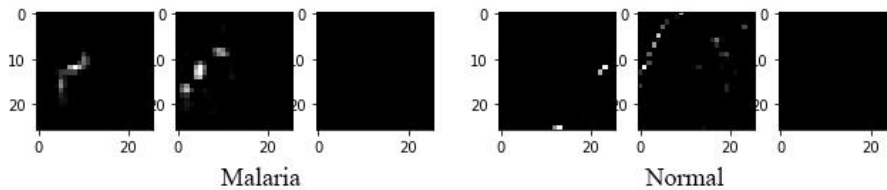Fig. 17 Feature Extraction on Blood Cell Image in 3nd Convolution Layer.
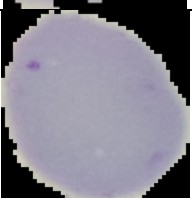
Fig. 18 Feature Extraction on Normal Blood Cell Image in 4nd Convolution Layer.

Based on the figure above, the feature that distinguishes red blood cells in feature extraction that occurs in the LeNet model convolution layer is the presence of holes in the blood cells. Where the red blood cells infected with malaria have holes in the blood cells, seen in Figures 14, 16 ,17a and 18a while figures 13, 15,17b and 18b do not have holes in the blood cells.

Testing

Below are the test results of a model that has been trained on the test data, where the test results with a value of 1 are normal and 0 is malaria.

No	Figure	Actual	Lenet Predict	DrNet Predict
1.		Malaria	Malaria	Malaria

2.		Malaria	Malaria	Malaria
3.		Malaria	Malaria	Malaria
4.		Malaria	Normal	Normal
…	…	…	…	…
4405.		Normal	Malaria	Normal
4406.		Normal	Malaria	Malaria
4407.		Normal	Normal	Normal
4408.		Normal	Normal	Normal

The results of the model testing on the test data will be presented in a configuration matrix. The following is a configuration matrix of the results of model testing against test data

Table 2
Lenet-5's Testing Confussion Matrix

		Predict	
		Malaria	Normal
Actual	Malaria	**2631**	72
	Normal	166	**2643**

.

Table 3
DrNet Testing Confussion Matrix

		Predict	
		Malaria	Normal
Actual	Malaria	**2650**	147
	Normal	128	**2587**

DISCUSSIONS

Based on the test results, the Lenet-5 model and the DRnet model managed to classify correctly with an accuracy value above 80%. But when compared to the two models, the Lenet-5 model is more precise in the case of classification of red blood cells affected by malaria. This can be seen where the evaluation value of the Lenet-5 model is higher than the DRnet model.

The following is a comparison of some of the evaluation values of Lenet-5 and DRnet which are presented in the form of table and chart charts.

Table 4
Comparison of Evaluation Score

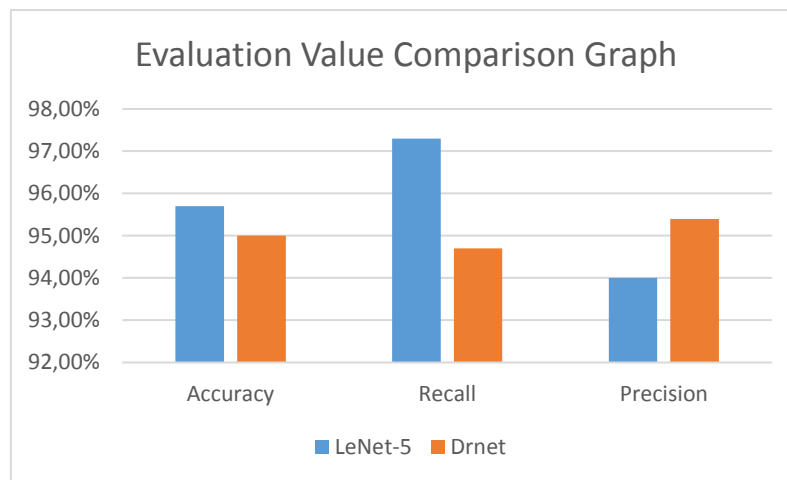| Model Name	Accuracy	Recall	Precision
LeNet-5	**95,7%**	**97,3%**	94%
Drnet	95%	94,7%	**95,4%**

Fig. 20 Comparison of Evaluation Values Graph

CONCLUSION

Based on the experimental results, the CNN method with the Lenet-5 and DRnet models can classify blood cells well where the accuracy obtained based on the Lenet-5 and Drnet models is 95.7% and 95%, respectively. For the case of red blood cell classification, Lenet-5 is more appropriate where the evaluation value of Lenet-5 is higher than DRnet.

REFERENCES

Abiyev, Rahib H., and Mohammad Khaleel Sallam Ma'aitah. 2018. "Deep Convolutional Neural Networks for Chest Diseases Detection." *Journal of Healthcare Engineering* 2018.

Autino, Beatrice, Alice Noris, Rosario Russo, and Francesco Castelli. 2012. "Epidemiology of Malaria in Endemic Areas." *Mediterranean Journal of Hematology and Infectious Diseases* 4(1).

Chima, Jaspreet Singh, Abhishek Shah, Karan Shah, and Rekha Ramesh. 2020. "Malaria Cell Image Classification Using Deep Learning." *International Journal of Recent Technology and Engineering* 8(6): 5553–59.

Dave, Ishan R. 2018. "Image Analysis for Malaria Parasite Detection from Microscopic Images of Thick Blood Smear." *Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2017* 2018-Janua: 1303–7.

Hidayat, Ardi, Ucuk Darusalam, and Irmawati Irmawati. 2019. "Detection of Disease on Corn Plants Using Convolutional Neural Network Methods." *Jurnal Ilmu Komputer dan Informasi* 12(1): 51.

Ignatius, Hartanto, Ricky Chandra, Nicholas Bohdan, and Abdi Dharma. 2019. "COMPARISON OF CONVOLUTIONAL NEURAL NETWORK MODEL IN CLASSIFICATION OF DIABETIC RETINOPATHY." *Jurnal Penelitian Pos dan Informatika* 9(2): 141–50.

Kalkan, Soner Can, and Ozgur Koray Sahingoz. 2019. "Deep Learning Based Classification of Malaria from Slide Images." *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*: 1–4.

Liang, Zhaohui et al. 2017. "CNN-Based Image Analysis for Malaria Diagnosis." *Proceedings - 2016 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016*: 493–96.

Nanoti, Akshay, Sparsh Jain, Chetan Gupta, and Garima Vyas. 2016. "Detection of Malaria Parasite Species and

Life Cycle Stages Using Microscopic Images of Thin Blood Smear." *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016* 1.

Poostchi, Mahdieh et al. 2018. "Image Analysis and Machine Learning for Detecting Malaria." *Translational Research* 194(2018): 36–55. https://doi.org/10.1016/j.trsl.2017.12.004.

Rahman, Aimon et al. 2019. "Improving Malaria Parasite Detection from Red Blood Cell Using Deep Convolutional Neural Networks." *Https://Arxiv.Org/Ftp/Arxiv/Papers/1907/1907.10418.Pdf*: 1–33.

Rahmanti, Farah Zakiyah et al. 2014. "LVQ (Learning Vector Quantization) Method for Identification of Plasmodium Vivax in Thick Blood Film." *Icbeta* (October 2015).

Rajaraman, Sivaramakrishnan et al. 2018. "Pre-Trained Convolutional Neural Networks as Feature Extractors toward Improved Malaria Parasite Detection in Thin Blood Smear Images." *PeerJ* 2018(4): 1–17.

Sankaran, S et al. 2017. "Detection and Classification of Malaria Parasites Using Digital Image Processing." *international Research Journal of Engineering and Technology (IRJET)* 04(05): 87–89.

Singh, Amartya et al. 2020. "Malaria Detection Using Contour Detection And Random Forest Classifier." 29(3): 503–13.

Yadav, Samir S., and Shivajirao M. Jadhav. 2019. "Deep Convolutional Neural Network Based Medical Image Classification for Disease Diagnosis." *Journal of Big Data* 6(1). https://doi.org/10.1186/s40537-019-0276-2.

Yohannes, Yohannes, Siska Devella, and Kelvin Arianto. 2020. "Deteksi Penyakit Malaria Menggunakan Convolutional Neural Network Berbasis Saliency." *JUITA: Jurnal Informatika* 8(1): 37.