

The Effect of Regularization

by Kkmms 1

Submission date: 08-Feb-2022 10:34PM (UTC-0600)

Submission ID: 1758265711

File name: The_Effect_of_Regularization.pdf (915.46K)

Word count: 2563

Character count: 14376

The Effect of Regularization on Deep Learning Methods For Detection of Malaria Infection

7 Windra Swastika
Informatics Engineering
Universitas Ma Chung
Malang, Indonesia

windra.swastika@machung.ac.id

12 Romy Budhi Widodo
Informatics Engineering
Universitas Ma Chung
Malang, Indonesia

romy.budhi@machung.ac.id

Ginza Alfarizha Balqis
Informatics Engineering
Universitas Ma Chung
Malang, Indonesia

311910028@machung.ac.id

Rehmadenta Sitepu
Program of Pharmacy
Universitas Ma Chung
Malang, Indonesia

rehmadanta.sitepu@machung.ac.id

Abstract— Malaria is an infectious disease caused by peripheral blood parasites of the genus *Plasmodium*. The estimated global malaria cases reached 229 million cases in 2019, of which 250,644 cases occurred in Indonesia. The large number of cases makes the early and accurate diagnosis of malaria very important because it can reduce the severity and prevent death. The most widely used method of diagnosis by far is examining a thin blood smear under a microscope and looking for infected cells. This research examines the effect of regularization applied to several Convolutional Neural Network (CNN) architectures to obtain the best accuracy of malaria parasite detection on thin blood smear images. The regularization techniques used are dropout layer, L2 regularization, and data augmentation. The results show that the use of BaselineNet without regularization achieved 94.92% accuracy. The use of regularization on ResNet-50, MicroVGGNet, BaselineNet-1 obtained 97.12%, 95.64% and 96.28% accuracy respectively.

Keywords—Deep Learning; Malaria Infection Detection; Regularization

I. INTRODUCTION

Malaria is one of the deadliest diseases in the world. The process of early and accurate diagnosis of malaria is very important because it can reduce the severity and prevent death. The most widely used method of diagnosis by far is examining a thin blood smear under a microscope and looking for infected cells. The patient's blood is smeared on the slide and given contrasting agents to more easily identify the infected parasite on the red blood cells. Medical personnel will manually count the number of parasitic red blood cells, sometimes in the thousands according to the WHO protocol. The process of diagnosing malaria can be accelerated and its accuracy improved with the help of a branch of Deep Learning (DL), namely the Convolutional Neural Network (CNN) [1].

CNN as one of the DL models was chosen because it can quickly identify malaria parasites on thin blood smear images. The promising performance of CNN is complemented by the ability to process large amounts of data annotations. As proposed in [2] the performance of pre-trained and customized CNN models as feature extractors was evaluated for classifying infected and uninfected cells. Furthermore, models with good performance will be cross-validated at the patient level (direct diagnosis) with the aim of reducing bias and generalization errors. The evaluation results are then used again to determine the optimal layer for pre-trained and customized CNN models. CNN models that have been used in previous studies to identify malaria parasites in thin blood

smear images include VGGNet, ResNet-50, AlexNet, DenseNet, Xception, and custom architecture.

In addition to optimizing the CNN model, improving the accuracy and time of CNN training can be achieved by techniques such as preprocessing the digital image before the image is processed. Examples of preprocessing that can be done on thin blood smear images are color normalization, histogram equalization, unsharp mask, and illumination equalization [3]. The addition of this image preprocessing technique aims to make identification easier because the image has been added with filters that clarify the presence of malaria parasites.

Another optimization technique is to regularize the learning algorithm to improve the performance of the model. Examples of regularization that can be done are dropout (removing some of the training data), image augmentation (changing the shape of the image), L1 and L2 regularization, and early stopping which stops the model training process when the model validation deteriorates. Hence, the purpose of this study is to design a model to detect malaria infection and evaluate the effects of regularization on accuracy, sensitivity and specificity.

II. RELATED WORKS

A. Convolution Neural Network (CNN)

CNN is a feed-forward neural network consisting of convolution operations. CNN has fewer parameters than other artificial neural networks, so it is suitable for processing large numbers of images. It is also able to recognize whether the image has a close relationship between adjacent pixels [4]. After the input layer and before the output layer, CNN generally consists of three layers, namely convolutional layer, pooling layer, and fully-connected layer.

There are many CNN architectures that can be used, including VGGNet, ResNet, DenseNet, Xception, or standard CNN baseline modeling which is often referred as BaselineNet. An example of a baseline architecture model is shown in Fig. 1. This architecture uses an input shape of 64x64 pixels, four convolution layers, four pooling layers, two fully-connected layers, and the ReLU activation function.

B. MicroVGGNet

MicroVGGNet is a simplified CNN VGGNet architecture. Both VGG-16 and VGG-19 are included in deep neural networks that have many layers and large model sizes. Fig. 2 shows the classic architecture of the VGG-16.

III. METHODS

A. Image Dataset

In this study, the thin blood smear image dataset was obtained from the Malaria Screener study conducted by the Communications Engineering Branch (malaria datasets n.d.), part of the Lister Hill National Center for Biomedical Communications, an R&D division of the U.S. National Library of Medicine. These images were acquired by photographing under a light microscope of thin blood smears that had been stained with Giemsa. These thin blood smear samples were from 50 healthy patients and 150 infected patients. Each red blood cell in each image is then segmented and classified. The result is a total of 27,588 images, with 13,779 samples of uninfected/healthy blood cells and 13,779 samples of blood cells containing malaria parasites. Fig. 6 shows an image sample of uninfected and infected malaria parasites.

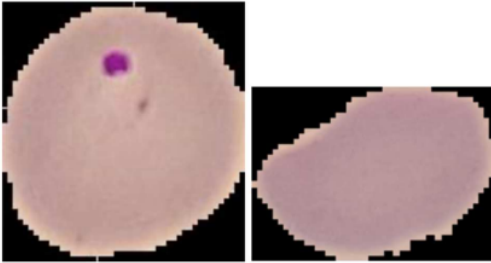


Fig. 6. Example of training data for malaria-infected blood cells (left) and healthy blood cells (right)

B. CNN Architecture

Fig. 7 shows the flow of the classification process in this study. First of all, image acquisition is carried out to obtain input images that are available in a data set with two types of classes (infected and uninfected). The input image can then be applied to be trained with BaselineNet-1, MicroVGG Net, ResNet-50 and Rajaraman architectures.

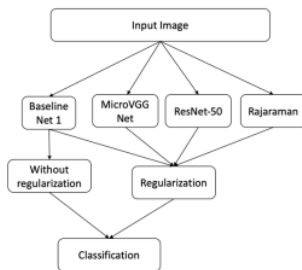


Fig. 7. Flow of the classification process

The regularization used in this study are data augmentation, dropout, and L2 regularization data which applied to BaselineNet-1.

Finally, the effectiveness of the combination CNN architecture and regularization is measured using confusion matrix (Table 1). The accuracy, sensitivity, and specificity are also measured.

TABLE I. CONFUSION MATRIX

		Actual Value	
		+	-
Predicted Value	+	True Positive (TP)	False Positive (FP)
	-	False Negative (FN)	True Negative (TN)

If the infected image is detected as infected by CNN, the identification is true positive (TP). If the infected image is considered as an uninfected image by CNN, then the identification is false negative (FN). If an uninfected image is considered an infected image, then the identification result is a false positive (FP). Then, when an uninfected image during identification is considered as an uninfected image, then the value of the identification result is true negative (TN).

The calculation of the level of accuracy (2) shows the success of CNN to identify the status of infection with malaria parasites from thin blood smear images. The result of sensitivity (3) calculation shows how well the identification of malaria parasites is. The results of the specificity (4) calculation show how well CNN can distinguish between infected and uninfected images.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

IV. RESULTS

A. BaselineNet-1 Without Regularization

BaselineNet-1 architecture is a simple CNN architecture that is used as a comparison. It was trained without regularization and five repetitions of training were carried out with 20 epochs of each training. The evaluation results based on the confusion matrix value are as shown in Table 1.

TABLE II. RESULTS OF ACCURACY, SENSITIVITY, SPECIFICITY AND TEST LOSS OF BASELINE NET-1 WITHOUT REGULARIZATION

#	Accuracy	Sensitivity	Specificity
1	0.9470	0.9552	0.9390
2	0.9534	0.9596	0.9472
3	0.9507	0.9545	0.9469
4	0.9458	0.9303	0.9609
5	0.9494	0.9574	0.9415
Avg	0.9492	0.9514	0.9471

The average of accuracy, sensitivity, and specificity are 94.92%, 95.14% and 94.71% respectively.

B. CNN Architectures With Regularization

Like baselinenet-1 architecture, architectures using regularization are trained in 20 epochs and repeated 5 times. The average results obtained for each architecture are shown in Table III.

TABLE III. AVERAGE OF ACCURACY, SENSITIVITY, SPECIFICITY AND TEST LOSS OF BASELINE NET-1, MICROVGGNET, RESNET-50 AND RAJARAMAN WITH REGULARIZATION

Architecture	Accuracy	Sensitivity	Specificity
BaselineNet-1	0,9628	0,9707	0,9551
MicroVGGNet	0,9564	0,9793	0,9444
ResNet-50	0,9712	0,9795	0,9660
Rajaraman	0,9625	0,9783	0,9510

As shown in Table III, all CNN architectures with regularization outperformed in the accuracy, sensitivity and specificity compared to BaselineNet-1. The best accuracy, sensitivity and specificity were obtained by ResNet-50 with 97.12%, 97.95% and 96.60% respectively.

V. CONCLUSION

This study shows how regularization on deep learning methods for detection of malaria infection can increase the accuracy, sensitivity and specificity. It is also can be concluded that ResNet-50 with regularization has the best accuracy, sensitivity and specificity among others architectures.

REFERENCES

- [1] Dong, Y., Jiang, Z., Shen, H., Pan, W. D., Williams, L. A., Reddy, V. V., ... & Bryan, A. W. (2017, February). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In *2017 IEEE EMBS international conference on biomedical & health informatics (BHI)* (pp. 101-104). IEEE.
- [2] Rajaraman, S., Jaeger, S., & Antani, S. K. (2019). Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ*, 7, e6977.
- [3] Swastika, W., Kristianti, G. M., & Widodo, R. B. (2021, April). Effective preprocessed thin blood smear images to improve malaria parasite detection using deep learning. In *Journal of Physics: Conference Series* (Vol. 1869, No. 1, p. 012092). IOP Publishing.
- [4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [5] Zhiqi, Y. (2020, June). Gesture recognition based on improved VGGNET convolutional neural network. In *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOECE)* (pp. 1736-1739). IEEE.
- [6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [7] Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., ... & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, 6, e4568.
- [8] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.

The Effect of Regularization

ORIGINALITY REPORT

17%

SIMILARITY INDEX

13%

INTERNET SOURCES

12%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1

peerj.com

Internet Source

1%

2

Muhammad Yusup, Suci Dwijayanti, Hera Hikmarika, Zaenal Husin, Bhakti Yudho Suprpto. "Gyroscope Control System in Wave Power Plant Using PID Controller", 2021 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE), 2021

Publication

1%

3

Noer Soedjarwanto, Vina Widiawati, Nining Purwasih, Gigih Forda Nama. "Developing a Prototype of Solar Tracking for Solar Cell Energy Optimization with Internet of Things (IoT) Technology", 2021 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE), 2021

Publication

1%

4

Submitted to University of Strathclyde

Student Paper

1%

5	journal.unnes.ac.id Internet Source	1 %
6	ceb.nlm.nih.gov Internet Source	1 %
7	Kestrilia R. Prilianti, Samuel P. Yuwono, Marcelinus A. S. Adhiwibawa, Monika N. P. Prihastyanti et al. "Automatic leaf color level determination for need based fertilizer using fuzzy logic on mobile application: A model for soybean leaves", 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), 2014 Publication	1 %
8	ebin.pub Internet Source	1 %
9	www.ecmlpkdd2013.org Internet Source	1 %
10	dokumen.pub Internet Source	1 %
11	deepai.org Internet Source	1 %
12	Hartawan Sugihono, Romy Budhi Widodo, Oesman Hendra Kelana. "Study of the Android and ANN-based Upper-arm Mouse", 2018 5th International Conference on	1 %

Electrical Engineering, Computer Science and Informatics (EECSI), 2018

Publication

13 "Machine Learning for Intelligent Decision Science", Springer Science and Business Media LLC, 2020 1 %

Publication

14 Maria Delgado-Ortet, Angel Molina, Santiago Alférez, José Rodellar, Anna Merino. "A Deep Learning Approach for Segmentation of Red Blood Cell Images and Malaria Detection", Entropy, 2020 1 %

Publication

15 Submitted to School of Business and Management ITB 1 %

Student Paper

16 ijsrcseit.com 1 %

Internet Source

17 Young Bong Seo. "Overlapping decentralized EA control design for an active suspension system of a full car model", SICE 2000 Proceedings of the 39th SICE Annual Conference International Session Papers (IEEE Cat No 00TH8545) SICE-00, 2000 <1 %

Publication

18 www.frontiersin.org <1 %

Internet Source

19	cse.buffalo.edu Internet Source	<1 %
20	Y Ekawati, A A Hapsari. "Taguchi experimental design to determine the taste quality characteristic of candied carrot", IOP Conference Series: Materials Science and Engineering, 2018 Publication	<1 %
21	avatto.com Internet Source	<1 %
22	lup.lub.lu.se Internet Source	<1 %
23	www.ijsta.com Internet Source	<1 %
24	www.science.gov Internet Source	<1 %
25	Hassan, Rohail. "Active Stability Monitoring and Stability Control of DC Microgrids Using Incremental Continuous Injection.", Utah State University, 2021 Publication	<1 %
26	Lecture Notes in Computer Science, 2015. Publication	<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On

The Effect of Regularization

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4
