An Automated Malaria Cells Detection from Thin Blood Smear Images using Yolov4

Dhevisha Sukumarran[1], Khairunissa Hasikin[2*], Anis Salwa Mohd Khairuddin[3], Romano Ngui[4], Wan Yusoff Wan Sulaiman[5], Indra Vythilingam[6], Paul C.S Divis[7]

[1,2]Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia

[3]Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia

[4,5,6]Department of Parasitology, Faculty of Medicine, Universiti Malaya, Kuala Lumpur 50603, Malaysia

[7]Malaria Research Centre, Faculty of Medicine and Health Sciences, Universiti Malaysia Sarawak, Sarawak 94300, Malaysia

[*]Corresponding Author

doi: https://doi.org/10.21467/proceedings.141.19

ABSTRACT

Malaria is a severe global health problem, with an estimated 241 million malaria infections and 627,000 malaria deaths globally in 2020. Hundreds of millions of blood films are examined annually for malaria, which includes manually counting parasites and infected red blood cells by a trained microscopist. Segmented red blood cells play an important role in applying deep learning for malaria diagnosis. However, traditional segmentation and separation of single red blood cells is challenging and requires much human intervention. Therefore, instead of segmented red blood cells, the performance of deep learning models can be studied using bounded cell images. Various object detection architectures are studied in detecting red blood cells from thin blood smear images. However, there is a lack of study on the performance of Yolov4 to detect infected cells in thin blood smear images. This study aims to evaluate the performance of Yolov4 in detecting red blood cells infected by four types of malaria species and integrate a separate algorithm to automatically crop the infected cells. Different types of malaria images are used to study if the model can still detect cells infected by various malaria parasites and from multiple stages of infection despite their morphology differences. The MP-IDB malaria datasets were used in the experiments. The performance of the Yolov4 model was evaluated by partitioning the train and test dataset by 90/10 and 80/20. The partitioning was done on datasets with and without augmentations. The results show that upon training Yolov4 model can detect infected cells despite their morphological differences. Model 4 with 80/20 dataset partition and augmentation is chosen as the best model with the best mAP of 93.43%.

Keywords: malaria, Yolov4

1 Introduction

Malaria is a severe and sometimes fatal disease caused by a *Plasmodium* parasite that commonly infects the Anopheles mosquito, feeding on humans. Malaria is an ultimate global health problem, with an estimated 241 million malaria infections and 627,000 malaria deaths globally in 2020 (WHO, 2021). Accurate parasite counts are essential for malaria diagnosis and testing for drug resistance, measuring drug effectiveness, and classifying disease severity. However, microscopic diagnostics is not standardized and depends heavily on the experience and skill of the microscopist. It is common for microscopists in low-resource settings to work in isolation, with no rigorous system to ensure their skills and, thus, diagnostic quality. This leads to incorrect diagnostic decisions in the field, prompting efforts to perform malaria diagnoses automatically. Since the wide acceptance of deep learning, the importance of large, annotated data image repositories for

Proceedings DOI: 10.21467/proceedings.141; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-957605-4-1

training is now widely understood, leading to the incredible support of data acquisition efforts. Given these developments, automated microscopy is in the race toward a cheap, simple, and reliable method for diagnosing malaria.

Among the previous works, such as Liang et al. (2017) used active contour to segment the cells, Dong et al. (2017) segmented cells using thresholding techniques and morphological operations, while their cell separation depends upon Hough Circle transform, and Rajaraman et al. (2019) applied level-set algorithm for RBC segmentation. Morphological operations for cell segmentation have been widely used (Arshad et al., 2022; Dong et al., 2017; G.Gopakumar, 2018; Molina et al., 2021). The typical steps involve applying histogram equalization for contrast improvement, Otsu's thresholding to binarize and separate the cells from the background and mathematical morphologies to remove noise and unwanted pixels. The watershed algorithm is famously used to segment each cell from a clumped of cells.

Traditional segmentation methods require much human intervention, and the parameters of these approaches may change in various stages of malaria infection as the morphology of the species varies in different stages. However, instead of traditional methods, deep learning object detection architectures can be used to detect the cells of interest upon training automatically. Such as, Zhao et al. (2020) used a deep learning object detection algorithm known as SSD300 to detect all the red blood cells from thin blood smear images. They trained their model using segmented NIH cells but cross-validated it with the bounded cell images. Alternatively, a deep learning model could be trained on bounded images. So far, the performance of object detection algorithms such as Faster R-CNN [5], Mask R-CNN [6], Yolov2[10] and SSD300[11] have been studied on thin smear.

Yolov4 architecture was mainly fine-tuned in previous studies to detect malaria parasites from thick blood smear images (Abdurahman et al., 2021). Yolov4 was introduced in 2020, surpassing the other object detectors in speed and accuracy. It is yet to be employed on thin blood smear images.

In this study, we investigated Yolov4's ability to detect only infected cells from thin blood smear images. In contrast to the prior research, the images used in this study depict infections caused by all malaria species at various stages of infection. This is to determine if the object detection architecture can recognize infected cells despite morphological changes across species and the infection phase. In addition, a different algorithm will be incorporated into the final model to automatically crop and store the infected cells that have been detected. The crop-infected cells can be employed in future applications, such as forwarding them to a second deep learning model to identify them based on their malarial species or infection stage.

2 Materials and Methods

2.1 Data acquisition

The publicly available dataset of 210 thin blood smear images from MP-IDB (https://github.com/andrealoddo/MP-IDB-The-Malaria-Parasite-Image-Database-for-Image-Processing-and-Analysis) was used in this study. The following dataset provides thin smear images infected by all malaria species from various stages of infection. Data annotation was performed only on the infected cells based on the ground truth provided in the form of binary masks.

2.2 Yolov4 hyperparameters setting

A custom model configuration of Yolov4 was performed according to the number of classes to be detected. The hyperparameters used were the batch size of 64, subdivisions 16, maximum batch 6000, steps of 4800,5400, a learning rate of 0.001, and filter of 18 in three convolutional layers before the Yolo layers. In

this experiment, four models were trained and tested, models 1 and 2 at 90/10 data partition and models 3 and 4 at 80/20 data partition. For models 2 and 4, data augmentation was performed where the training set was augmented to 1000 images.

2.3 Cropping algorithm integration

The final model was saved in the form of TensorFlow model files. A separate algorithm was then implemented in python to automatically crop and save only the cells detected by the model.

2.4 Performance Evaluation

The results were evaluated with four performance metrics: recall rate, precision, F1-score, and mean average precision (mAP). Recall rate is the intolerance of the model toward the false negative (1). Precision is the intolerance of the model toward false positives (2)

$$\frac{TP}{TP+FN} \quad (1) \qquad \frac{TP}{TP+FP} \quad (2)$$

F1-score represents the harmonic mean of recall and precision. It gives equal weight to recall and precision (3). Mean average precision is the mean of the average precision of all the classes detected by the object detection model (4).

$$2.\frac{precision(class=1).recall(class=1)}{precision(class=1)+recall(class=1)} \quad (3) \qquad mAP = \frac{1}{n}\sum_{k=1}^{k=n} APk \quad (4)$$

APk= AP of class k, n= number of classes

3 Results and Discussion

Table 1: *Performance metrics of model 1 at 90/10 data partition without augmentation*

Epochs	Recall (%)	Precision (%)	F1-score (%)	mAP (%)
1000	92	79	85	87.25
2000	93	88	90	94.37
3000	**92**	**93**	**93**	**93.68**
4000	89	95	92	92.74
5000	91	94	92	93.86
6000	91	94	92	93.42

Table 2: *Performance metrics of model 2 at 90/10 data partition with augmentation*

Epochs	Recall (%)	Precision (%)	F1-score (%)	mAP (%)
1000	80	82	81	83.50
2000	82	86	84	87.82

Proceedings DOI: 10.21467/proceedings.141
ISBN: 978-81-957605-4-1

Series: AIJR Proceedings
ISSN: 2582-3922

3000	83	85	84	87.34
4000	**89**	**86**	**88**	**89.80**
5000	82	87	84	87.93
6000	86	84	85	87.14

Table 3: *Performance metrics of model 3 at 80/20 data partition without augmentation*

Epochs	Recall (%)	Precision (%)	F1-score (%)	mAP (%)
1000	93	74	83	88.93
2000	90	77	83	87.48
3000	87	82	85	87.31
4000	84	84	84	87.34
Best	**88**	**82**	**85**	**89.93**

Table 4: *Performance metrics of model 4 at 80/20 data partition with augmentation*

Epochs	Recall (%)	Precision (%)	F1-score (%)	mAP (%)
1000	92	74	82	89.08
2000	93	82	88	92.08
3000	87	85	86	87.63
4000	**95**	**83**	**89**	**93.43**

Various performance metric was obtained to evaluate the performance of the models. A high recall rate is crucial to ensure that the model has a lesser chance of letting slip the infected cells as it is important to flag them. Since only one class is being detected, the mean average precision will be the average precision of the detected class. Like F1-score, average precision incorporates precision and recall, but average precision is finding the area under the precision-recall curve. Average precision summarizes the precision-recall curve. Therefore, the mAP will be used to evaluate the performance of the models.

Based on the results, at a partition of 90/10, the best mAP obtained is 93.68% at 3000 epochs. Following the 3000 epochs, the model starts to overfit as it struggles to give more than 1 percent of the current mAP. Data augmentation was performed to study whether the model's performance improved and whether overfitting could be prevented or delayed. The augmentation was done so that there is an equal number of images in all stages of infection to reduce the model's bias. The augmentations were flipping, rotation at various angles, zoom in and out and distortions. Based on Table 2, it can be seen the mAP in every epoch decreases as compared to model 1, and the best mAP is 89.80%. However, the overfitting was eventually

delayed. The augmentation was performed on the whole blood smear image, which also contained healthy red blood cells and other cells and stains. Multiplying the images multiplies all the other cells, stains, and information in the images, such as noise. This may have influenced the features the model learns.

The models achieved satisfactory results from 90/10 partitions with and without data augmentation. Figure 1 shows a few examples of results on the test images. Although model 1 achieves higher mAP than model 2, both models fairly detect the infected cells but with different confidence. Based on the detections checked individually on the test images, the models fairly detect most of the cells infected from the schizont, trophozoite, and gametocyte stages. Typically, there are more infected cells in the ring stage in a single blood smear image; although the model does not detect all those infected cells, they fairly detect most of them (Figure 1).

For better confidence in the model's performance, training was again conducted at the partition of 80/20. From the results in Table 3, model 3 starts to overfit faster than the previous models; therefore, the training was halted after 4000 epochs. It should be considered that when a partition of 80/20 is done, the number of training images decreases compared to the 90/10 partition, but the number of test images increases. This may have eventually caused an early overfit. Nevertheless, the model achieved the best mAP of 89.93% in more test images with fewer train images. To obtain a more robust model, the training was again conducted using augmented images at a partition of 80/20.

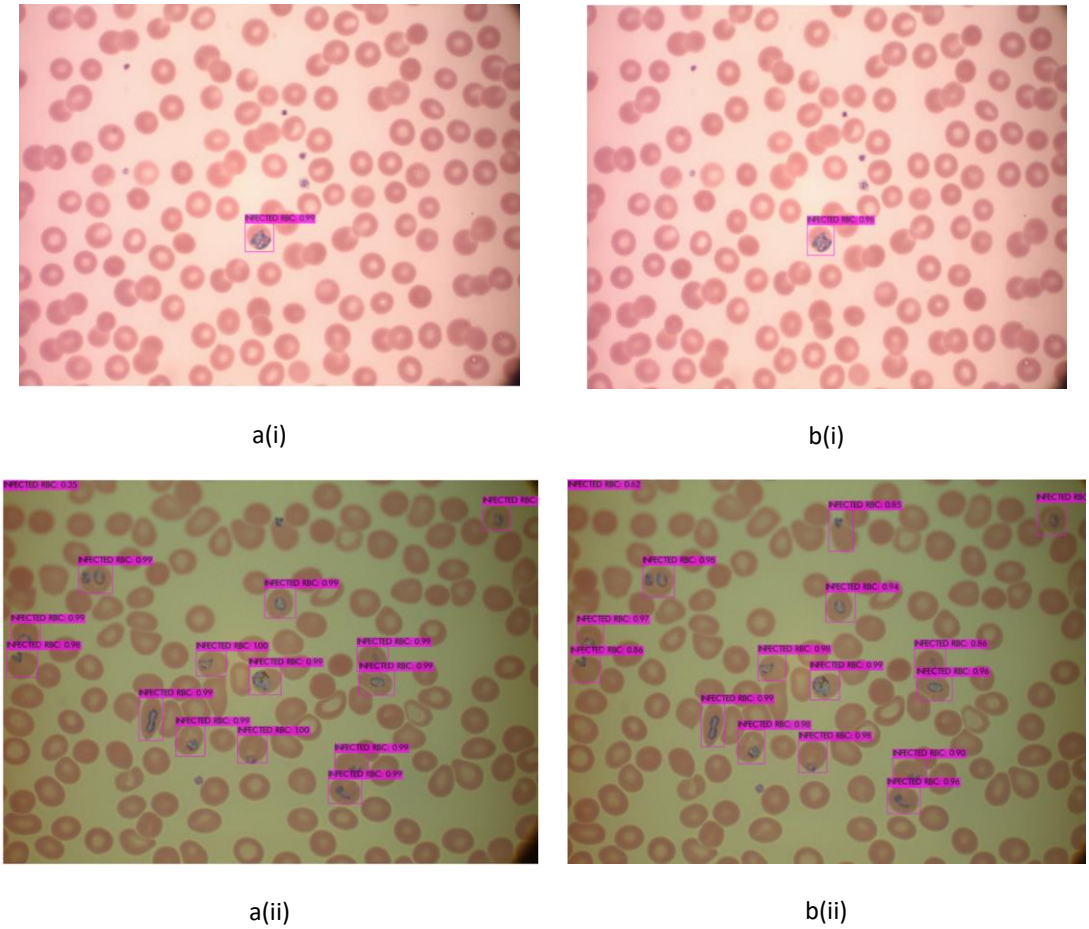
A few adjustments to the annotations were made for better results with the augmented data. Such as, the cells that are at the edge of the image and were cut off are not included in the annotations. Besides that, the images from the same patient were separated to either be in the train or the test images to prevent stain leakage [9] and make the model more robust.

The distortion augmentation method was removed as it might change the cell morphologies and complicate learning. Moreover, the size of the bounding boxes was adjusted by a few pixels to avoid fitting on the cells of interest and for a better intersection of the ground truth and predicted bounding boxes. Based on Table 4, model 4 manages to achieve the best and highest mAP of 93.43%. From Figure 2, model 4 can detect a few of the cells not detected by model 3. This can explain the highest recall rate of 95% achieved by model 4. Model 4 is selected as the best and final model as a partition of 80/20 gives better confidence in the model. With data augmentation, it is trained with more training sets, increasing its robustness and reducing its bias.

To our knowledge, no comparable literature evaluated the performance of Yolov4 on thin blood smear images. The final Yolov4 model achieves a better result than other object detection algorithms, such as the SSD300 in a study by Zhao et al. (2020) and the Yolov2 model (Yang et al., 2020) on thin blood smear images. In most previous studies, deep learning models are trained and tested on images of the same malaria species. Although this study uses an object detection algorithm, based on the results, it can be determined that upon training, other deep learning models can also classify cells as infected despite the differences in features respective to malaria species and stages of infection.

Proceedings DOI: 10.21467/proceedings.141
ISBN: 978-81-957605-4-1

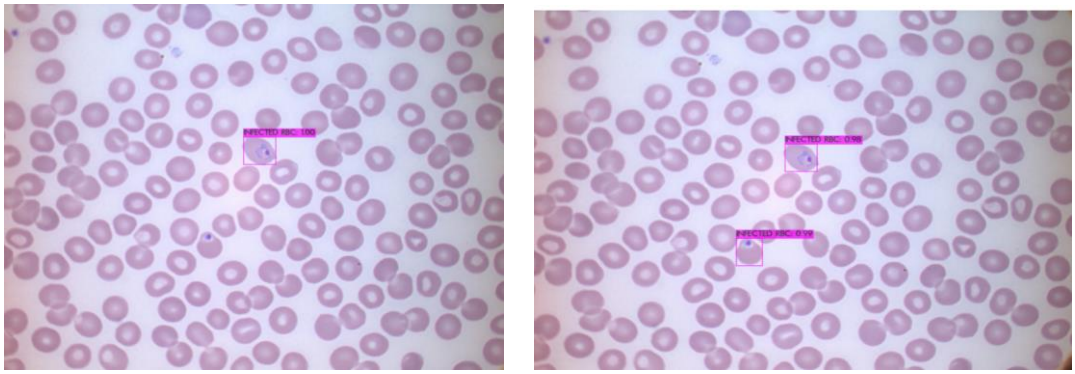
Series: AIJR Proceedings
ISSN: 2582-3922



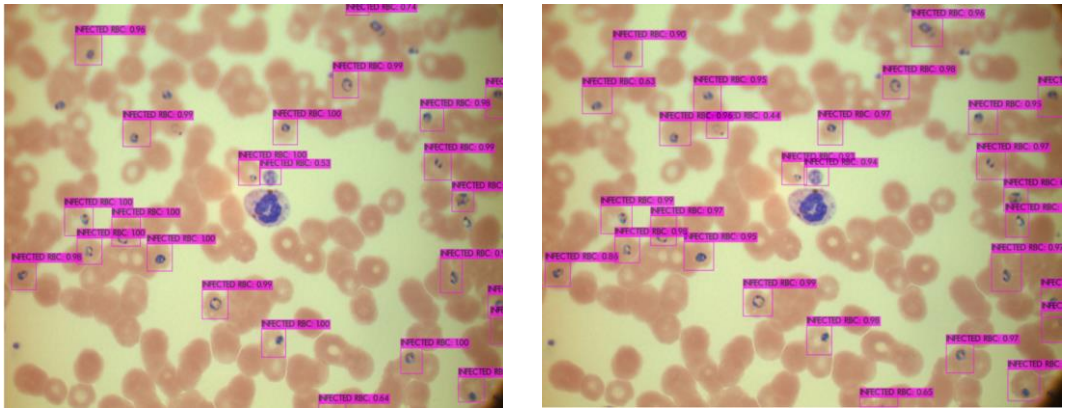
a(i) b(i)

a(ii) b(ii)

Figure 1: *Example results on the same test images from model 1(a) and model 2(b)*



c(i) d(i)



c(ii) d(ii)

Figure 2: *Examples of results on the same test images of 80/20 partition with and without augmentation, model 3(c) and model 4(d)*

4 Conclusions

In this study, Yolov4 was implemented to identify the infected cells from thin blood smear images. The study shows that upon training, the model can detect infected cells despite the species and stages of infection, although they have different morphologies. The object detection architecture with an integrated cropping algorithm provides a faster automatic detection and cropping of the infected cells. The cropped bounded cell images can facilitate further applications of deep learning in malaria diagnosis, such as for classification according to species or stages of infection. The limitation of the study is the limited availability of datasets from all species and stages of infection.

5 Declarations

5.1 Study Limitations

Lack of thin blood smear images.

5.2 Competing Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

5.3 Publisher's Note

AIJR remains neutral with regard to jurisdiction claims in published maps and institutional affiliations.

References

[1] Abdurahman, F., Fante, K. A., & Aliy, M. (2021). Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models. *BMC Bioinformatics*, *22*(1), Article 112. https://doi.org/10.1186/s12859-021-04036-4

[2] Arshad, Q. A., Ali, M., Hassan, S. U., Chen, C., Imran, A., Rasul, G., & Sultani, W. (2022). A dataset and benchmark for malaria life-cycle classification in thin blood smear images. *Neural Computing and Applications*, *34*(6), 4473-4485. https://doi.org/10.1007/s00521-021-06602-6

[3] Dong, Y., Jiang, Z., Shen, H., David Pan, W., Williams, L. A., Reddy, V. V. B., Benjamin, W. H., & Bryan, A. W. (2017). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. 2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017,

[4] G.Gopakumar, S., M., Sai Siva Gorthi, Gorthi. R. K. Sai Subrahmanyam. (2018). CNN based Malaria Diagnosis from Focus-stack of Blood Smear Images Acquired using Custom-built Slide Scanner. *Journal of biophotonics, 11(3).*

[5] Hung, J., & Carpenter, A. (2017). Applying Faster R-CNN for Object Detection on Malaria Images. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops

[6] Loh, D. R., Yong, W. X., Yapeter, J., Subburaj, K., & Chandramohanadas, R. (2021). A deep learning approach to the screening of malaria infection: Automated and rapid cell counting, object detection and instance segmentation using Mask R-CNN. *Computerized*

medical imaging and graphics: the official journal of the Computerized Medical Imaging Society, *88*, 101845. https://doi.org/10.1016/j.compmedimag.2020.101845

[7] Molina, A., Rodellar, J., Boldú, L., Acevedo, A., Alférez, S., & Merino, A. (2021). Automatic identification of malaria and other red blood cell inclusions using convolutional neural networks. *Computers in biology and medicine*, *136*, Article 104680. https://doi.org/10.1016/j.compbiomed.2021.104680

[8] 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*, Article e6977. https://doi.org/10.7717/PEERJ.6977

[9] Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., Jaeger, S., & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, *2018*(4), Article e4568. https://doi.org/10.7717/peerj.4568

[10] Yang, F., Quizon, N., Yu, H., Silamut, K., Maude, R. J., Jaeger, S., & Antani, S. (2020). Cascading YOLO: Automated malaria parasite detection for Plasmodium vivax in thin blood smears. Progress in Biomedical Optics and Imaging - Proceedings of SPIE,

[11] Zhao, O. S., Kolluri, N., Anand, A., Chu, N., Bhavaraju, R., Ojha, A., Tiku, S., Nguyen, D., Chen, R., Morales, A., Valliappan, D., Patel, J. P., & Nguyen, K. (2020). Convolutional neural networks to automate the screening of malaria in low-resource countries. *PeerJ*, *8*, Article 9674. https://doi.org/10.7717/peerj.9674