

# A weakly supervised deep learning approach for detecting malaria and sickle cells in blood films

Petru Manescu<sup>1,\*</sup>, Christopher Bendkowski<sup>1</sup>, Remy Claveau<sup>1</sup>, Muna Elmi<sup>1</sup>, Biobele J. Brown<sup>2</sup>, Vijay Pawar<sup>1</sup>, Mike J. Shaw<sup>1</sup> and Delmiro Fernandez-Reyes<sup>1,2,\*</sup>

<sup>1</sup> Department of Computer Science, Faculty of Engineering Sciences, University College London, London, United Kingdom

p.manescu@ucl.ac.uk

<sup>2</sup> Department of Pediatrics, College of Medicine University of Ibadan, University College Hospital, Ibadan, Nigeria

delmiro.fernandez-reyes@ucl.ac.uk

\* Equal contribution

**Abstract.** Machine vision analysis of blood films imaged under a brightfield microscope could provide scalable malaria diagnosis solutions in resource constrained endemic urban settings. The major bottleneck in successfully analyzing blood films with deep learning vision techniques is a lack of object-level annotations of disease markers such as parasites or abnormal red blood cells. To overcome this challenge, this work proposes a novel deep learning supervised approach that leverages weak labels readily available from routine clinical microscopy to diagnose malaria in thick blood film microscopy. This approach is based on aggregating the convolutional features of multiple objects present in one hundred high resolution image fields. We show that this method not only achieves expert-level malaria diagnostic accuracy without any hard object-level labels but can also identify individual malaria parasites in digitized thick blood films, which is useful in assessing disease severity and response to treatment. We demonstrate another application scenario where our approach is able to detect sickle cells in thin blood films. We discuss the wider applicability of the approach in automated analysis of thick blood films for the diagnosis of other blood disorders.

**Keywords:** Weak Supervision, Malaria, Sickle Cells, Blood Films, Microscopy

## 1 Introduction

Assessment of the ubiquitous Giemsa-stained Thick Blood Film (TBF) by a trained microscopist remains the gold-standard diagnostic tool in regions of the world where malaria is endemic [1]. The TBF is often accompanied by a Blood Film Smear (BFS), also called Thin Blood Film, which is not always fully examined. Assessment of both

TBF and BFS provides valuable information that facilitate malaria and hematology care pathways in resource constrained settings. Despite being treatable, malaria remains a major global health challenge with over 219 million cases worldwide leading to almost half-million deaths annually [1], mostly among children. Fast and reliable diagnostic testing scalable to large urban endemic regions is urgently needed. However, human visual inspection of the TBF by a malaria pathologist is time-consuming (particularly in low-parasitemia or negative cases), subject to errors associated with human-fatigue and cognitive load in busy clinical microscopy services and relies on the availability of trained personnel. Advances in digital pathology could provide automated or assisted solutions to facilitate malaria and hematology clinical pathways in large urban malaria regions as well as supporting peri-urban and rural regions lacking pathologists.

In recent years, computer vision techniques have been developed in an attempt to automatically detect malaria parasites in digitized microscopic images of TBFs based on image processing and supervised machine learning [2], [3]. Convolutional Neural Networks (ConvNets) have shown promise in identifying malaria parasites in TBF [4], [5], [6]. Nevertheless, the challenges in using these techniques are two-fold: Firstly, they rely on thousands, if not millions, of object-level manual annotations to train deep learning models to differentiate objects of interest (parasites) from distractors (staining artefacts, platelets, etc.). These annotations are difficult to obtain since they require the expertise of teams of trained microscopists to assess thousands of digitized Fields of View (FoVs). Secondly, automated diagnosis, which requires inspection of up to a hundred FoVs, is unreliable in cases with a low parasite count due to the inherent false positive rate of supervised deep learning detection models. Recently, weakly supervised approaches have been successfully applied to identify cancerous regions in digital histopathology images [7], [8], [9], [10]. More specifically, they classify and segment microscopy images using only whole image level annotations [11]. However, in the context of high-magnification blood film analysis, image level annotations are not easily available. Nevertheless, sample-level annotations, which translate into a set of images (FoVs) associated with a single label (e.g. clinical diagnosis of malaria confirmed by microscopy) are far easier to obtain.

Here we propose a weakly-supervised deep learning approach entitled Multiple Objects Features Fusion (MOFF) to diagnose malaria in one hundred high resolution TBF FoVs. Our method does not necessitate any hard object-level annotations for training (full supervision) but relies only the clinical-microscopy diagnosis of the sample (weak supervision with labels provided by routine clinical-microscopy). The method is also able to detect individual parasites in images, which is useful for estimating the severity of the disease and the response to treatment. To show the versatility of our approach, we also present an extension of the MOFF method for detection of sickle Red Blood Cells (RBCs) in the BFS, illustrating its broader value for extraction of clinically important data from the ubiquitous blood film. Since MOFF does not rely on a complex annotation system or availability of human-expert pathologists to annotate at the object-level, our approach has significant potential for deployment in resource constrained settings for a variety of diagnostic applications.

## 2 Multiple Objects Feature Fusion (MOFF) learning

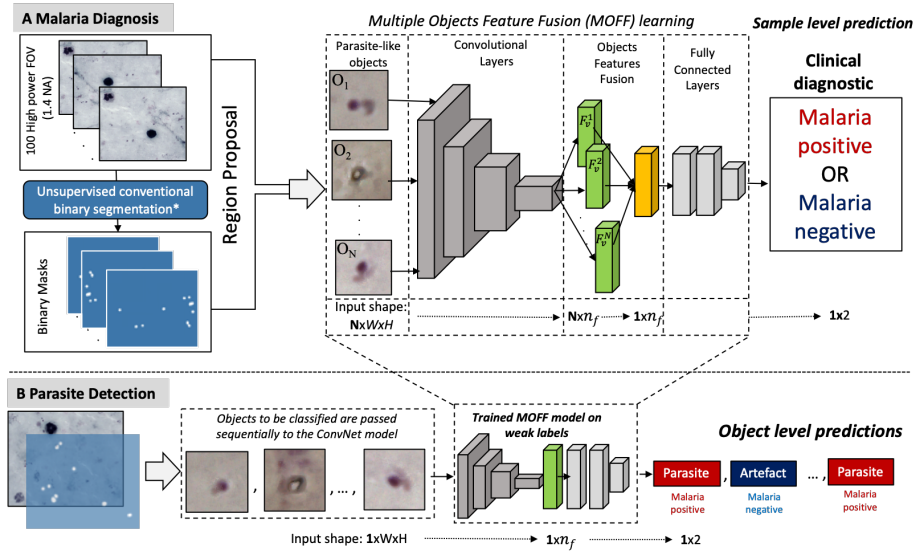
Potential objects of interest ( $O_i$ ) (e.g. parasites, distractors) are identified and extracted from all captured FoVs (Figure 1). For this region proposal phase, we chose a segmentation approach based on binary thresholding and morphological operations to extract regions of interest (ROIs) around objects of 2-4  $\mu\text{m}$  in diameter (parasite-like objects) in TBF FoV [2]. Depending on the task, other modalities such as selective search [12] or cell segmentation techniques [13] (Supplementary material) can be employed during this phase. These  $O_i$  are next passed as the input of a ConvNet classification model whose output is the diagnosis of the sample provided by the clinical microscopist (e.g. malaria positive). The flattened feature vectors ( $F_v^i$ ) output by the convolutional layers (*conv*) of the model are fused into one single feature vector which is next passed through the fully connected (FC) layers to predict the clinical diagnosis ( $C_d$ ):

$$C_d(O_1, O_2, \dots, O_N) = \sigma_L(\dots(\sigma_1(f_{fusion}(F_v^1, F_v^2, \dots, F_v^N) \cdot W_1 + b_1) \cdot \dots) \cdot W_L + b_L) \quad (1)$$

where  $N$  is the number of input images,  $L$  is the number of FC layers,  $W_k, b_k, \sigma_k$  are the corresponding weights, biases and activations of each FC layer,  $F_v^i = \text{conv}(O_i)$  and  $f_{fusion}$  is the feature fusion rule. More specifically:

$$f_{fusion}^{i=1\dots N}(F_v^i) = \text{agg}_{1 \leq i \leq N}(\zeta_j^i) \quad (2)$$

With  $n_f$ , the number of individual features ( $\zeta_j^i$ ) in each  $F_v^i$ , and *agg*, the aggregation modality (for example *average* or *max*).



**Fig. 1.** Multiple Objects Features Fusion (MOFF) learning approach applied to malaria diagnosis and parasite detection. \*For details, see supplementary material.

## 2.1 Sample-level prediction

Given the large number of images which must to be analyzed per sample, the object proposal phase outputs a large number (thousands) of proposed object images. Due to GPU memory constraints, during the training phase, for each iteration  $N_o$  a randomly selected subset of object images from each sample are passed through the network. This works as an augmentation technique since a different subset of object images are selected from each sample during each epoch. The architecture allows a variable number of input object images because the ConvNet feature vectors are fused. During the test (prediction) phase, all the extracted object images for each sample are passed through the network in subsets of  $N_T$  ( $N_T \gg N_o$ ) and a final predicted  $C_d$  is reached following a majority voting rule. The feature vector fusion before the FC layers has the additional benefit of reducing the GPU memory requirements and computational effort. This allows a large enough number of object images to be passed through the network during both training ( $N_o > 100$ ) and testing ( $N_T > 500$ ) which is essential in cases where parasites (or any other object of interest) are sparsely distributed in the samples.

## 2.2 Object-level detection

The MOFF model is trained to distinguish between those ROIs which contain parasites, when assessing TBF or sickle red blood cells when analyzing BFS, and those ROIs which do not, without specifically relying on object-level annotations. Once a sample has been classified as positive, the individual objects extracted by the segmentation algorithms are passed *one-by-one* through the trained model. The feature fusion step does not affect the inference since the number of  $F_v$  is equal to one (Figure 1 B). Consequently, an object image gets classified as malaria positive (or sickle RBC positive) if it contains a malaria parasite (or morphologically abnormal RBCs).

# 3 Experiments

## 3.1 Data sets

**Clinical malaria microscopy.** Malaria parasites (MPs) were detected and counted by human-expert microscopists in thick blood films stained with Giemsa at our clinics at the University College Hospital (UCH) in the city of Ibadan, Nigeria. A patient was declared malaria negative if no parasites were detected in 100 high magnification (100x) TBF FoVs. The corresponding films were next digitized (Section 3.2) and used to train and evaluate our MOFF model (Table 1).

**Sickle cell disease diagnosis.** Hemoglobin electrophoresis [14] is routinely used to obtain the haemoglobin phenotype ( $C_d$  labels) and test patients for Sickle Cell Disease (SCD) in our clinics. Giemsa-stained thin blood smears were prepared and digitized for a group of patients (Table 1).

**Ethical Statement.** The internationally recognized ethics committee at the Institute for Advanced Medical Research and Training (IAMRAT) of the College of Medicine, University of Ibadan (COMUI) approved this research with permit numbers: UI/EC/10/0130, UI/EC/19/0110. Parents and/or guardians of study participants gave informed written consent in accordance with the World Medical Association ethical principles for research involving human subjects.

### 3.2 Image acquisition and pre-processing

Images were captured with custom built brightfield microscope fitted with a 100X/1.4NA objective lens, a motorized x-y sample positioning stage and a color camera. For each thick blood film sample, we used random sampling to automatically select and capture 100 non-overlapping FoVs, each covering an area of  $166 \mu\text{m} \times 142 \mu\text{m}$ . For blood film smears, between 10 and 20 FoVs containing 1500 - 4000 RBCs were manually captured. Given the limited depth of field of the high numerical aperture objective lens, a focal series (z-stack) of 14 planes with a separation of  $0.5 \mu\text{m}$  was acquired to capture the entire thickness of the blood film (typically  $\sim 3$  to  $6 \mu\text{m}$ ) for each FoV. Each z-stacks was then projected onto a single (xy) plane using a wavelet-based Extended Depth of Field (EDoF) algorithm [15].

**Table 1.** Number of patient samples used to train and evaluate our approach.

	Malaria, Thick Blood Film (TBF) <sup>1</sup>		Sickle Cell, Blood Film Smear (BFS) <sup>2</sup>	
	Negative	Positive	Negative	Positive
Train	85	84	34	37
Test	70	60	35	35
Total	155	144	69	72

### 3.3 Model training

The weights of the convolutional layers of the MOFF model were initialized with weights from a VGG-19 model [16] pre-trained on the ImageNet dataset [17]. A window of  $64 \times 64$  pixels ( $4.2 \mu\text{m} \times 4.2 \mu\text{m}$ ) around the centroid of each segmented object was chosen for generating the object images for the malaria problem. Images of individual RBCs corresponding to their bounding boxes as output by the segmentation step were scaled to  $128 \times 128$  pixels and used as input object images for the sickle cell anemia diagnosis. Geometrical transformation such as random rotations and random flips were applied to the object images from each sample in the training set.  $N_o$  was set to 200 whereas  $N_T$  to 1000. Stochastic gradient descent with a learning rate of 0.0003 and a cross entropy loss function were applied to optimize the model weights during

<sup>1</sup> TBF dataset available at <https://doi.org/10.5522/04/12173568>

<sup>2</sup> BFS dataset available at <https://doi.org/10.5522/04/12407567>

100 epochs. No object-level labels were used. We compared two feature fusion strategies: *max* and *average*<sup>3</sup>.

## 4 Results

We evaluated our diagnostic prediction for malaria and SCD against the clinical diagnostic on the test samples in Table 1. The object-level detection of the MOFF models, for both parasites and sickle cells, was evaluated on a subset of images manually annotated by expert clinical microscopists.

### 4.1 Malaria diagnosis and parasite detection

We compared the accuracy of our weakly supervised malaria diagnostic approach with an existing supervised approach [4]. This latter approach relies on an object detector to detect individual parasites in high power FoVs as suggested in [18] and diagnoses a sample as positive if the number of detected parasites exceeds a certain threshold determined using the negative samples in the training set as described in [4]. The object detector was trained on a separate annotated set of similar FoVs [19].

**Table 2.** Evaluation of automated malaria diagnosis in thick blood films. Parasitemia levels in (MP/100 FoVs) stratified as *Low*: up to 10; *Medium*: >10 to 100; *High*: >100

Method	Type	Specificity	Sensitivity			
			<i>Low</i>	<i>Medium</i>	<i>High</i>	<i>Overall</i>
Mehanian <i>et al</i> [4] (Reported)	Full Supervision	>.90	-	-	-	.87
Torres <i>et al</i> [5] (Reported)	Full Supervision	.85	-	-	-	.72
Mehanian <i>et al</i> [4] (Our test set) <sup>1</sup>	Full Supervision	<b>.95</b>	0	.23	.81	.66
<b>MOFF-avg</b> (Our test set)	Weak Supervision	<b>.95</b>	.50	.62	.80	.75
<b>MOFF-max</b> (Our test set)	Weak Supervision	.93	<b>.75</b>	<b>.92</b>	<b>.94</b>	<b>.93</b>

Our MOFF learning approach outperforms the Mehanian *et al.* diagnosis method<sup>4</sup> [4] (which is based on supervised parasite detection) on our test dataset in terms sensitivity for a specificity higher than .90 (Table 2), especially in samples with a low parasite count (as reported by the pathologists on the field). Both, Mehanian *et al.* [4] or Torres *et al.* [5] have not provided publicly accessible images for us to be able to test our MOFF method on their datasets. However, both the specificity and sensitivity of our weakly supervised method are equal or higher than those reported by these works

<sup>3</sup> Code available at <https://github.com/UCL/FAST-MAL-MOFF>

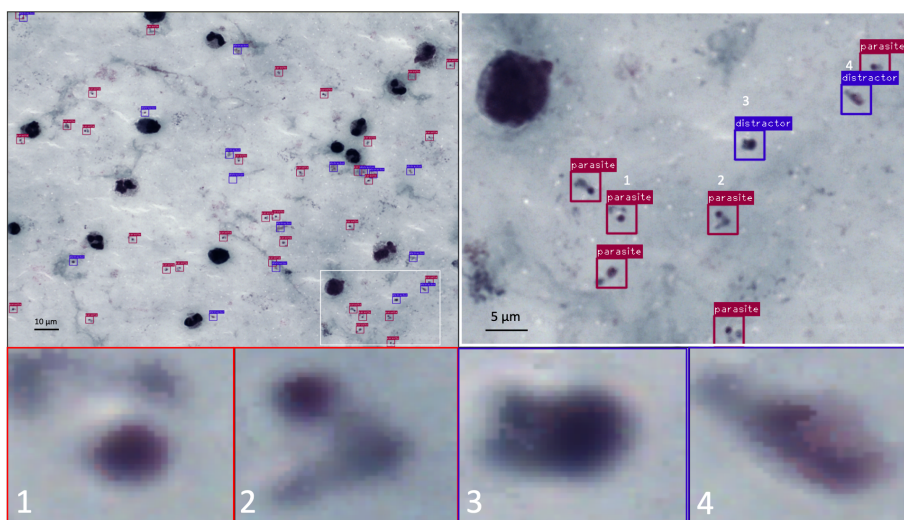
<sup>4</sup> Implemented slightly different than in the original paper as original code was not available

(Table 2). The *max* feature fusion strategy outputs a better sensitivity (.93) compared to the *average* one (.75) at the expense of a slightly lower specificity (.93 to .95).

**Table 3.** Malaria parasite (object-level) detection accuracy

Method	Type	Precision	Recall	F1-Score
R-FCN	Full Supervision	.92	.57	.70
MOFF-avg	Weak Supervision	.89	.52	.66
Faster R-CNN	Full Supervision	.70	.58	.64
MOFF-max	Weak Supervision	.64	.65	.64

The trained MOFF model’s ability to detect individual objects was evaluated on a test dataset of 33 FoV containing 300 manually annotated parasites by an expert clinical microscopist. We compared the outcome with that of two state-of-the-art fully supervised object detectors [20], [21] trained to identify parasites on a separate dataset (159 FoV with 2287 manually annotated parasites<sup>1</sup>). Surprisingly, without being trained on any object-level annotations, the performance of our MOFF model in terms of parasite detection is close to that of the fully supervised object detectors (Table 3). Figure 2 shows an example of parasites (red) and distractors (blue) identified by the weakly supervised MOFF in a test FoV from a malaria positive sample.



**Fig. 2.** MOFF parasite detection on a test high magnification FoV (100x 1.4NA). Red boxes indicate objects classified as parasites. Blue boxes indicate objects classified as distractors.

## 4.2 Sick cell detection in blood film smears

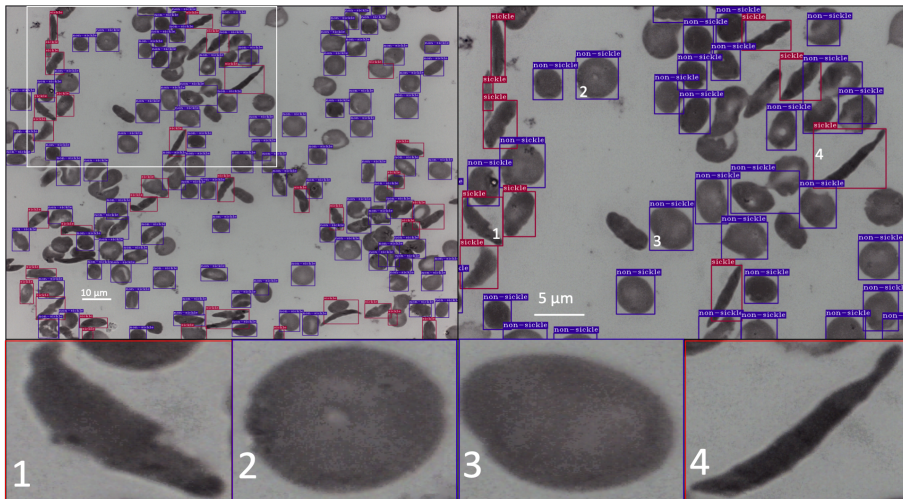
Recent approaches for automatic analysis of BFS [22], [23] including sickle cell disease (SCD) detection [24] still rely on fully supervised machine learning approaches which require large amounts of cell-level manual annotations. In contrast, our weakly supervised MOFF approach was able to provide a diagnosis (Table 4) together with an abnormal RBC count (Figure 3) without having to rely on any individual cell labels. As for malaria diagnosis, the *max* fusion strategy outputs a higher accuracy than the average one in detecting SCD (Table 4). The *max* fusion strategy was also significantly more precise in detecting abnormal RBCs associated with sickle cell anemia than the *average* fusion strategy at the expense of a slight decrease in recall (Table 5, test set of 5 FoVs containing 233 manually labelled sickle RBCs).

**Table 4.** MOFF Sickle Cell Disease detection accuracy

Method	Specificity	Sensitivity	Accuracy
MOFF-max	.86	.97	.91
MOFF-avg	.89	.74	.80

**Table 5.** Abnormal RBC detection accuracy using MOFF

Method	Precision	Recall	F1-Score
MOFF-max	.86	.69	.77
MOFF-avg	.56	.74	.64



**Fig. 3.** Abnormal RBC detection using MOFF. Red boxes indicate RBCs identified as sickle cells. Blue boxes indicate RBCs identified as non-sickle



Figure 3 shows that the MOFF model trained with weak labels is able to differentiate between sickle RBCs (red boxes) and normal RBCs (blue boxes) in BFS.

## 5 Discussion

We have demonstrated a weakly supervised deep learning method that is able to diagnose malaria in TBF and detect sickle cells in BFS by analyzing multiple FoVs. The method, which is based on the fusion of the convolutional features of multiple objects, was successful in identifying individual objects associated with the two disorders (malaria parasites and sickle cells). The MOFF approach has successfully overcome the challenge that the markers of the conditions (malaria parasites or abnormal RBCs) are often sparse (not present in every image). Results indicate that the maximum fusion strategy performs better which can be explained by the fact that, in most of the samples, the number of malaria parasites (or sickle RBCs) is much lower than the number of artefacts (or normal RBCs) and therefore a max-fusion would better signal the presence of a small number of markers in a large pool of candidate objects. Whereas most weakly supervised approaches use image-level labels to detect specific objects inside the image our approach uses sample-level labels associated with multiple image fields knowing that the malaria parasites, or sickle RBCs might not be present in every image of the sample. Although we have restricted our analysis to binary classification ('diseased' or 'healthy'), our method could readily be adapted to multiple classes and extended to identify other types of blood disorders.

## References

1. World Health Organization, World Malaria Report, (2018).
2. Arco J., Górriz J., Ramírez J., Álvarez I. and Puntonet C., Digital image analysis for automatic enumeration of malaria parasites using morphological operations, *Expert Systems with Applications*, no. 42, pp. 3041-3047(2015).
3. Rosado L., Da Costa J., Elias D. and Cardoso J., Automated detection of malaria parasites on thick blood smears via mobile devices, in *Procedia Computer Science*, Loughborough (2016).
4. Mehanian C., Jaiswal M., Delahunt C., Thompson C., Horning M., Hu L., ... and Wilson B., Computer-automated malaria diagnosis and quantitation using convolutional neural networks, in *Proceedings of the IEEE International Conference on Computer Vision* (2017).
5. Torres K., Bachman C., Delahunt C., Baldeon J., Alava F., Vilela D., Proux S., Mehanian C., McGuire S., Thompson C. and Ostbye T., Automated microscopy for routine malaria diagnosis: a field comparison on Giemsa-stained blood films in Peru, *Malaria journal*, vol. 17 (2018).

6. Yang F., Poostchi M., Yu H. et al., Deep Learning for Smartphone-based Malaria Parasite Detection in Thick Blood Smears, *IEEE journal of biomedical and health informatics* (2019).
7. Couture H. D., Marron J. S., Perou C. M., Troester M. A. and Niethammer M., Multiple Instance Learning for Heterogeneous Images: Training a CNN for Histopathology, in *MICCAI* (2018).
8. Jia Z., Huang X., Eric I., Chang C. and Xu Y., Constrained deep weak supervision for histopathology image segmentation, *IEEE transactions on medical imaging* vol. 36, no. 11 (2017).
9. Courtiol P., Maussion C., Moarii M., Pronier E., Pilcer S., Sefta M., ... and Girard N., Deep learning-based classification of mesothelioma improves prediction of patient outcome, vol. 25, no. 10 (2019).
10. Campanella G., Hanna M., Geneslaw L. et al., Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, *Nature medicine*, vol. 25, pp. 1301–1309 (2019).
11. Kraus O. Z., Ba J. L. and Frey B. J., Classifying and segmenting microscopy images with deep multiple instance learning, vol. 32, no. 12 (2016).
12. Uijlings, J.R.R., Van De Sande K.E.A., Gevers, T. and Smeulders, A.W.M. Selective search for object recognition, vol. 104, no. 2, (2013).
13. Das D. K., Mukherjee R. and C. Chakraborty, Computational microscopic imaging for malaria parasite detection: A systematic review, *Journal of Microscopy*, no. 1, pp. 1-19 (2015).
14. Naik R. P. and Haywood C. Jr, Sick cell trait diagnosis: clinical and social implications. *The American Society of Hematology Education Program Book*, no. 1, pp. 160-167 (2015).
15. Forster B., Van De Ville D., Berent J., Sage D. and Unser M., Complex Wavelets for Extended Depth-of-Field: a new method for the fusion of multichannel microscopy images, *Microscopy research and technique*, no. 65, pp. 33-42 (2004).
16. Simonyan K. and Zisserman A., Very deep convolutional networks for large-scale image recognition. (2015)
17. Deng J., Dong W., Socher R. et al., ImageNet: A large-scale hierarchical image database, *CVPR* (2009)
18. Yang F., Yu H. et al, Parasite Detection in Thick Blood Smears Based on Customized Faster-RCNN on Smartphones, *Lister Hill National Center for Biomedical Communications* (2019).
19. Manescu P., Shaw M. et al, Giemsa Stained Thick Blood Films for Clinical Microscopy Malaria Diagnosis with Deep Neural Networks Dataset.. University College London. Dataset. <https://doi.org/10.5522/04/12173568.v1> (2020).

20. Dai J., Li Y., He K. and Sun J., R-FCN: Object detection via region-based fully convolutional networks, *Adv. Neural Inf. Process. Syst.*, pp. 379–387, (2016).
21. Ren S., He K., Girshick R. and Sun J., Faster R - CNN: Towards Real - Time Object Detection with Region Proposal Networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137 – 114 (2017).
22. Mundhra D., Cheluvvaraju B., Rampure J. and Dastidar T. R., Analyzing microscopic images of peripheral blood smear using deep learning, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (2017).
23. Sadafi A., Koehler N., Makhro A., Bogdanova A., Navab N., Marr C. and Peng T., Multiclass Deep Active Learning for Detecting Red Blood Cell Subtypes in Brightfield Microscopy., in *International Conference on Medical Image Computing and Computer Assisted Intervention*, (2019).
24. Xu M., Papageorgiou D. P., Abidi S. Z., Dao M., Zhao H. and Karniadakis G. E., A deep convolutional neural network for classification of red blood cells in sickle cell anemia, *PLoS computational biology*, vol. 13, no. 10 (2017).