

Reducing the Diagnostic Burden of Malaria Using Microscopy Image Analysis and Machine Learning in the Field

Stefan Jaeger,^{1*} Kamolrat Silamut,² Hang Yu,¹ Mahdiah Poostchi,³ Ilker Ersoy,⁴ Andrew Powell,⁵ Zhaohui Liang,⁶ Md Amir Hossain,⁷ Sameer Antani,¹ Kannappan Palaniappan,³ Richard J Maude,^{2,8} George Thoma¹



Abstract

Microscopy remains the main technique for diagnosing malaria, despite the availability of Rapid Diagnostic Tests. Hundreds of millions of blood films are examined using microscopy every year for diagnosing malaria and quantifying parasite burdens. Processing this large number of slides consumes scarce resources. Microscopy technicians who read these slides in the field may be inadequately trained or overwhelmed with the volume of slides to process, leading to missed and incorrect diagnoses. To ease the burden for microscopists and improve diagnostic and quantitative accuracy, we have developed a smartphone application that can assist field microscopists in diagnosis of malaria. The software runs on a standard Android smartphone that is attached to a microscope by a low cost adapter. Images of thin-film microscope slides are acquired through the eyepiece of the microscope using the smartphone's built-in camera. The smartphone application assists microscopist in detecting parasites and estimating the parasitaemia. For each microscope field, the image processing software identifies infected and uninfected cells, and reports the parasite count per microliter of blood. The software was trained with more than 200,000 red blood cells from slides acquired at Chittagong Medical College Hospital in Bangladesh from patients with and without *P. falciparum* infection. These were manually annotated by an experienced professional slide reader. This is one of the largest labeled malaria slide image collections, enabling the application of new machine learning techniques such as deep learning. For each field-of-view image taken, an image processing pipeline is applied first to detect and segment cells before computing color and texture features for automatic machine classification to discriminate between infected and uninfected cells and other objects in the slide. Initial experiments show that our software correlates highly with human experts and flow cytometry.

¹Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, National Institutes of Health, USA

²Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand

³Computer Science Department, University of Missouri, Columbia, MO 65211, USA

⁴School of Medicine, University of Missouri, Columbia, MO 65212, USA

⁵Computer Science Department, Swarthmore College, Swarthmore, PA 19081, USA

⁶School of Information Technology, York University, Toronto, ON, M3J1P3, Canada

⁷Chittagong Medical College Hospital, Chittagong, Bangladesh

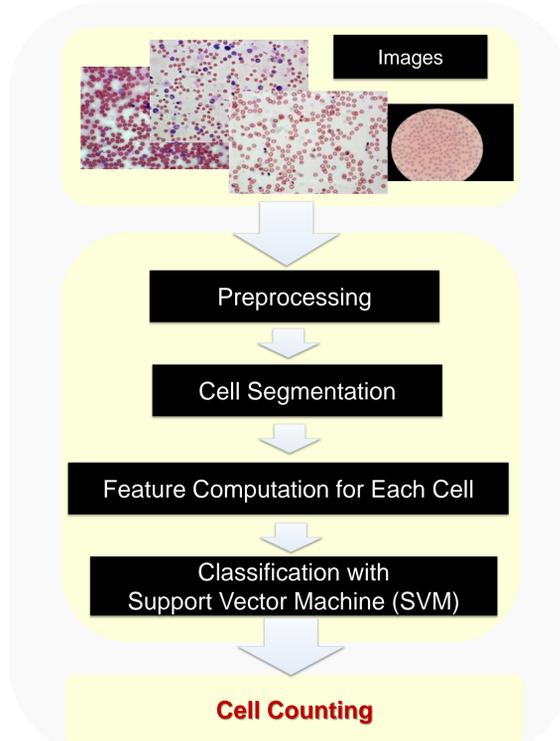
⁸Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 7FZ, UK

* Contact: stefan.jaeger@nih.gov

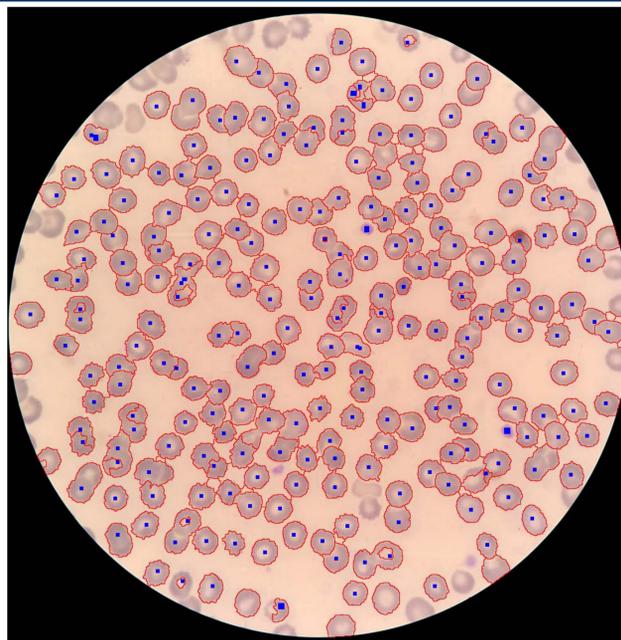
Acknowledgments

This research is supported by the HHS Ventures Fund and the Intramural Research Program of NIH, NLM, and Lister Hill National Center for Biomedical Communications. We would like to acknowledge all the patients in Bangladesh. Mahidol-Oxford Tropical Medicine Research Unit is funded by the Wellcome Trust of Great Britain.

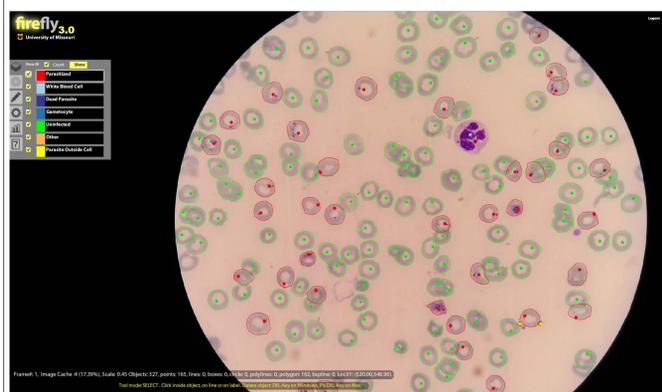
Processing Pipeline (SVM)



Level-Set Cell Segmentation



Training Data



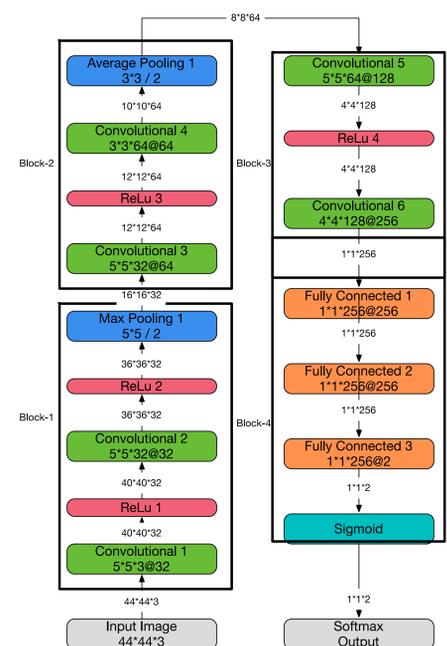
- Firefly annotation tool (firefly.cs.missouri.edu)
- 150 infected patients, 50 healthy patients
- 2500 slide images
- 250,000 manually annotated red blood cells

Smartphone Application



Deep Learning with CNN

Convolutional Neural Network (CNN):



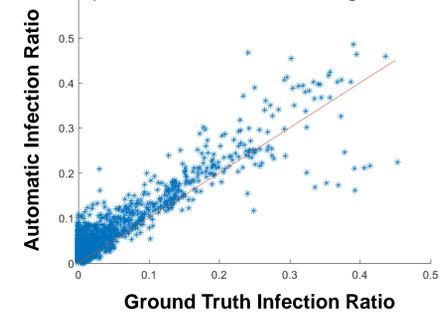
Results

Single Cell Classification Performance (CNN)

Measure	CNN Model	Transfer Learning
Accuracy	97.37	91.99
Sensitivity	96.99	89.00
Specificity	97.75	94.98
Precision	97.73	95.12
F1 Score	97.36	90.24
Matthews correlation coefficient	94.75	85.25

Pipeline Performance per Slide (SVM)

Ground Truth vs Automatic Infection Ratio*
(nRGB & LBP Features; avg. diff. = 0.017)



$$*Infection\ Ratio = \frac{\text{Number of infected red blood cells}}{\text{Total number of red blood cells}}$$

