U.S. National Library of Medicine

72nd Meeting of the
LHNCBC Board of Scientific Counselors
September 6-7, 2018

U.S. National Library of Medicine

Lister Hill National Center for
Biomedical Communications
An Intramural Research Division of the U.S. National Library of Medicine

A Report to the Board of Scientific Counselors
September 2018

Malaria Screening:
Research into Image Analysis and Deep Learning

Stefan R. Jaeger, PhD

Sameer Antani, PhD
Mahdieh Poostchi, PhD
Sivaramakrishnan Rajaraman, PhD
Feng Yang, PhD
Hang Yu

Communications Engineering Branch

U.S. National Library of Medicine, LHNCBC
8600 Rockville Pike, Building 38A
Bethesda, MD 20894

U.S. National Library of Medicine

Contents

1. Background

With more than 200 million cases worldwide, and more than 400,000 deaths per year, malaria is a major burden on global health [1, 16]. Malaria is caused by parasites that are transmitted through the bites of infected mosquitoes. When malaria parasites enter the blood stream, they infect and destroy the red blood cells. Typical symptoms of malaria include fever, fatigue, headaches, and in severe cases seizures, coma, and death. Most deaths occur among children in Africa, where a child dies from malaria every other minute of every day, and where malaria is a leading cause of childhood neuro-disability.

While existing drugs make malaria a curable disease, inadequate diagnostics and emerging drug resistance are major barriers to successful mortality reduction. The current standard method for malaria diagnosis in the field is light microscopy of blood films [17, 18, 19]. Microscopists examine millions of blood films every year for malaria. This involves manual counting of parasites or infected red blood cells, which is a labor-intensive and error-prone process, especially if patients have to be tested several times a day. However, accurate counts are essential to diagnosing malaria accurately, and are an important part of testing for drug-effectiveness, drug-resistance, and estimating disease severity [20].

Unfortunately, microscopic diagnostics 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 in place that can ensure the maintenance of their skills and thus diagnostic quality. This leads to incorrect diagnostic decisions in the field. For false negative cases, this leads to unnecessary use of antibiotics, a second consultation, lost days of work, and in some cases progression into severe malaria. For false positive cases, this means unnecessary use of anti-malaria drugs and suffering from their potential side effects, such as nausea, abdominal pain, diarrhea, and even more severe complications.

2. Project Objectives

To improve malaria diagnostics, the objective is to perform research towards an automated system that can assist microscopists in screening blood smears for malaria. For this purpose, the Lister Hill National Center for Biomedical Communications (LHC), in collaboration with the National Institute of Allergy and Infectious Diseases (NIAID) and Mahidol-Oxford University, is investigating new image processing and machine learning methods to detect and count parasites and infected blood cells in digitized images of blood smears. The research goal is to find the best machine learning methods, in particular recent deep learning models, which can learn the typical shape and visual appearance of parasites and infected cells based on a large repository of manually annotated training images. These methods will allow machines to detect parasites, perform the counting, and discriminate between infected and uninfected cells.

Furthermore, our goal is to use smartphone technology to process images of blood smears for validating and testing our algorithms in the field. The idea is to attach a smartphone to the eyepiece of a microscope by means of an adapter. This setup would allow taking pictures of blood smears with the smartphone's built-in camera and processing these pictures directly on the phone, using our algorithms. The microscopist can then simply view the parasite counts on the smartphone display, without the need for manual counting. For this to be successfully implemented in practice, research is required as to whether powerful image analysis and deep learning methods can be run on a smartphone platform, given the limitations in processing speed, memory, and other system constraints. Therefore, LHC is investigating the complexity of problems

like parasite detection and cell segmentation and classification to find efficient algorithms that can solve these problems on resource-constrained hardware.

3. Project Significance

The development of fast and reliable diagnostics is one of the most promising ways of fighting malaria, together with better treatment, development of new malaria vaccines, and mosquito control. Automatic parasite counting has several advantages compared to manual counting: a) it provides a more reliable and standardized interpretation of blood films, b) it allows more patients to be served by reducing the workload of malaria field workers, and c) it reduces diagnostic costs. The use of inexpensive, common light microscopy equipment makes this project well suited for resource-poor settings, where malaria is often prevalent, and where field workers typically rely on this type of microscopy. Likewise, the use of relatively inexpensive smartphone technology that is often already in the possession of field workers makes this project ideal for malaria-prone regions in countries with limited resources. The use of digital blood smear images in combination with highly portable smartphones, and the use of intelligent software based on image analysis and machine learning, is a very promising approach to fighting malaria and potentially a key step towards eradication of this disease. Therefore, investigating the potential contribution of artificial intelligence, in particular deep learning, to the diagnosis of a major infectious disease like malaria is a worthwhile endeavor that can help to improve global health, a major goal of NIH.

4. Methods and Procedures

We investigate computational methods to identify and quantify malaria parasites in blood smear images. This section gives an overview of our research toward optimal techniques for cell and parasite detection, segmentation, and classification.

To diagnose malaria under a microscope, a drop of blood is applied to a glass slide to create a blood smear, which is then immersed in a staining solution to make parasites more easily visible under a conventional light microscope, usually with a 100x objective. Two different types of blood smears are typically prepared for malaria diagnosis: thin and thick smears. Both thin and thick smear analyses are complementary ways to screen for malaria. Thin smears, which are the result of spreading the drop of blood across the glass slide, allow the examiner to identify malaria species and recognize parasite stages more easily. On thin smears, parasite numbers per microscopy field are lower; however, individual parasites are more clearly distinguishable from the background allowing a more precise quantification of parasites and distinction between different parasite species and development stages. On the other hand, a thick smear is used to detect the presence of parasites in a drop of blood. Thick smears allow a more efficient detection of parasites than thin smears, with a much higher sensitivity. Thick smears are mainly used for rapid initial identification of malaria infection but it can be challenging to quantify parasites when the parasitemia is high and to determine species. Note that in thin smears, infected red blood cells containing parasites are counted to quantify the degree of malaria infection (usually per 1000 red blood cells), whereas in thick smears, parasites are counted directly, without counting cells (usually per 200 or 500 white blood cells).

Image Acquisition and Annotations

In the last four years, we have acquired several image sets for machine training, including manual annotations of individual cells and parasites. In collaboration with University of Missouri, we have used the Firefly annotation tool to allow remote annotations by experts in Thailand and elsewhere. Firefly is a web-based annotation tool for visualization, segmentation, and tracking. It allows labeling and manual segmentation of objects in an image.

We have acquired images from Chittagong Medical College Hospital in Chittagong, Bangladesh, where our collaborators have been preparing and photographing Giemsa-stained blood smears for us. These images have become the main set for our research. We used the built-in camera of a smartphone, which we attached to a microscope by means of an adapter, to take pictures of each microscopic field of view through the microscope's eyepiece.

Using the Firefly annotation tool, an expert manually annotated the images at the Mahidol-Oxford Tropical Medicine Research Unit in Bangkok, Thailand. The expert identified uninfected and infected red blood cells, as well as other categories in thin smears, for example white blood cells or platelets. In addition, the expert outlined cell boundaries for some cells. For thick smears, the expert annotated parasites and white blood cells. Figure 1 shows an example of an annotated blood smear image, with uninfected and infected cells outlined in green and red, respectively.

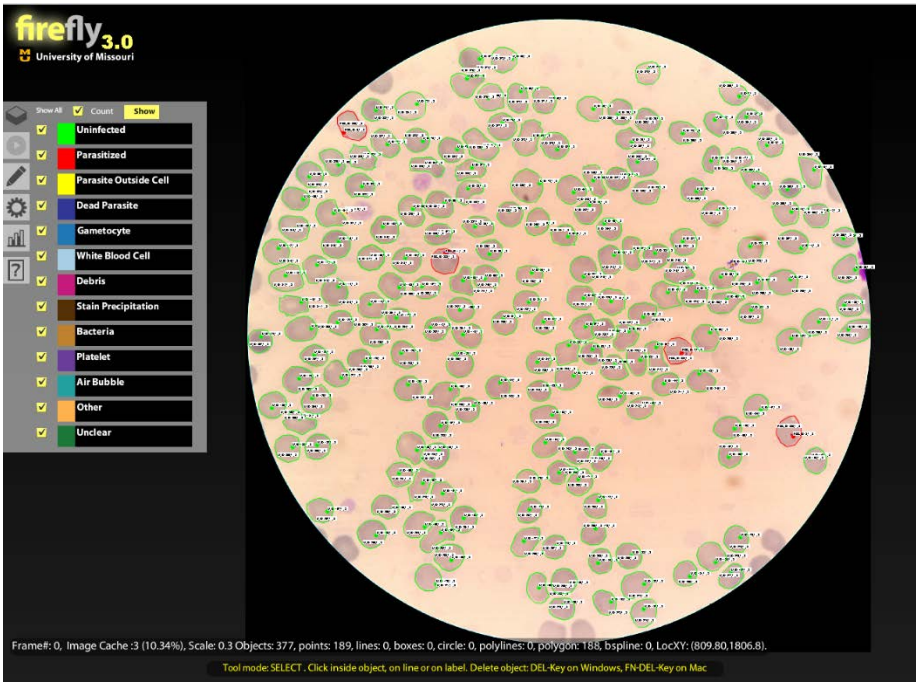


Figure 1. Labeling blood smears using the Firefly annotation tool (http://firefly.cs.missouri.edu/).

All de-identified images and annotations are archived at NLM (IRB#12972). In total, we have acquired the following images and annotations from Bangladesh:

Thin blood smears: We have acquired around **2500 blood smear images**, among which 1300 have cell annotations. We acquired these images from **200 patients** at Chittagong Medical College Hospital (150

infected patients and 50 normal patients). In total, our expert annotated almost 200,000 individual cells in these images. We have divided these images into a test and training set. For the test set, which contains images from 33 patients, we manually outlined the boundary of more than 34,000 cells for evaluation of cell segmentation methods, including white blood cells. The training set contains more than **160,000 manually identified cells** from 160 patients.

Thick blood smears: We have acquired 1819 thick smear images from 150 infected patients, including 84,961 annotated parasites and 35,036 annotated white blood cells. From 50 normal patients, we acquired 1142 images, including 27,112 annotated white blood cells. In total, we acquired and annotated **2961 images** from **200 patients**, including **84,961 annotated parasites** and **62,148 annotated white blood cells**. All smear images have a 3024×4032 pixel resolution.

All blood smear images acquired to date feature the parasite species *Plasmodium falciparum*, which is responsible for the majority of malaria deaths globally. However, we have also started collecting images of *Plasmodium vivax* because this is the second most significant species. Five *Plasmodium* species are known to cause human malaria, but *Plasmodium falciparum* and *Plasmodium vivax* are the most common.

Thin Smear Processing

Our system for thin smears consists of three main processing steps: (i) image acquisition of blood smear images using a standard light microscope with an attached smartphone camera, (ii) detection and segmentation of blood cells [26, 27], and (iii) feature computation and cell classification [28, 29]. Figure 2 illustrates the framework of this pipeline.

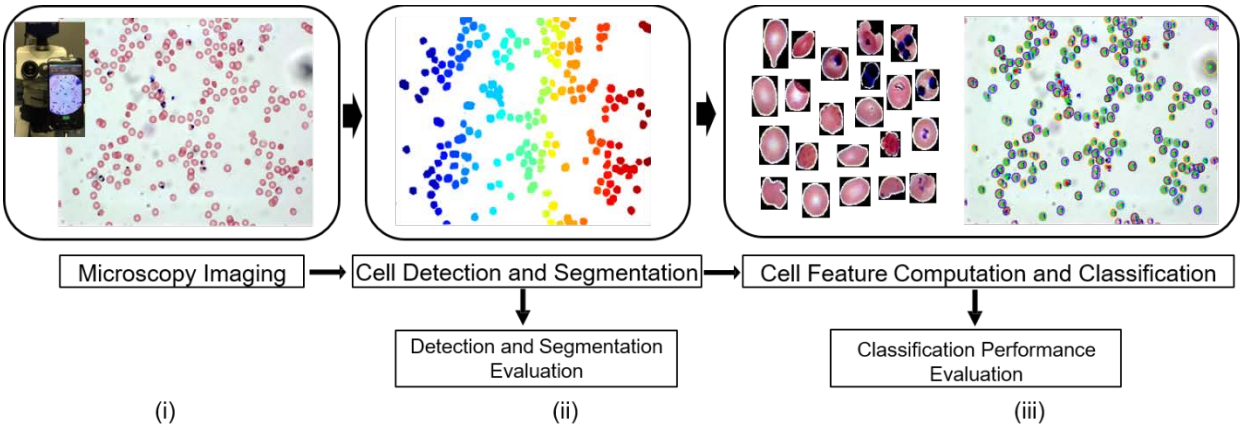


Figure 2. Our image analysis pipeline for counting infected and uninfected red blood cells in thin blood smears. From left to right: (i) Microscopy imaging of blood smears using a standard light microscope with a top-mounted camera or smartphone, (ii) Red blood cells detection and segmentation, (iii) Cell extraction, feature computation, and classification using machine learning.

Cell Detection and Segmentation

Cell detection and segmentation is a crucial step in our processing pipeline. The main challenges are low image contrast, cell staining variations, uneven illumination, cell shape diversity, cell size differences, texture complexities, and particularly touching cells [5]. We investigated methods ranging from Watershed

methods and novel active contour methods to new deep-learning based approaches [25]. Here, we describe our deep learning approach in more detail, which outperformed other methods.

Our framework consists of two deep learning network architectures, Faster RCNN and UNet, applied in a two-step process [23]. First, we train the Faster RCNN network on non-overlapping image tiles extracted from thin blood smear images. Then, in the second step, we apply the trained network to the connected components of cell foreground masks obtained by UNet. Our motivation behind dividing the images into tiles in the training process lies in the weak performance of Faster RCNN on images with small low-resolution objects, which negatively affects selection of region proposals. In addition, training on tiles is much faster than training on full images. During the inference stage, we give the connected cell clumps detected by UNet as input to Faster RCNN, which then detects individual cells. Presenting connected components to the network instead of tiles avoids cutting off cells, which may confuse the network. In Figure 3, we illustrate our deep learning pipeline architecture.

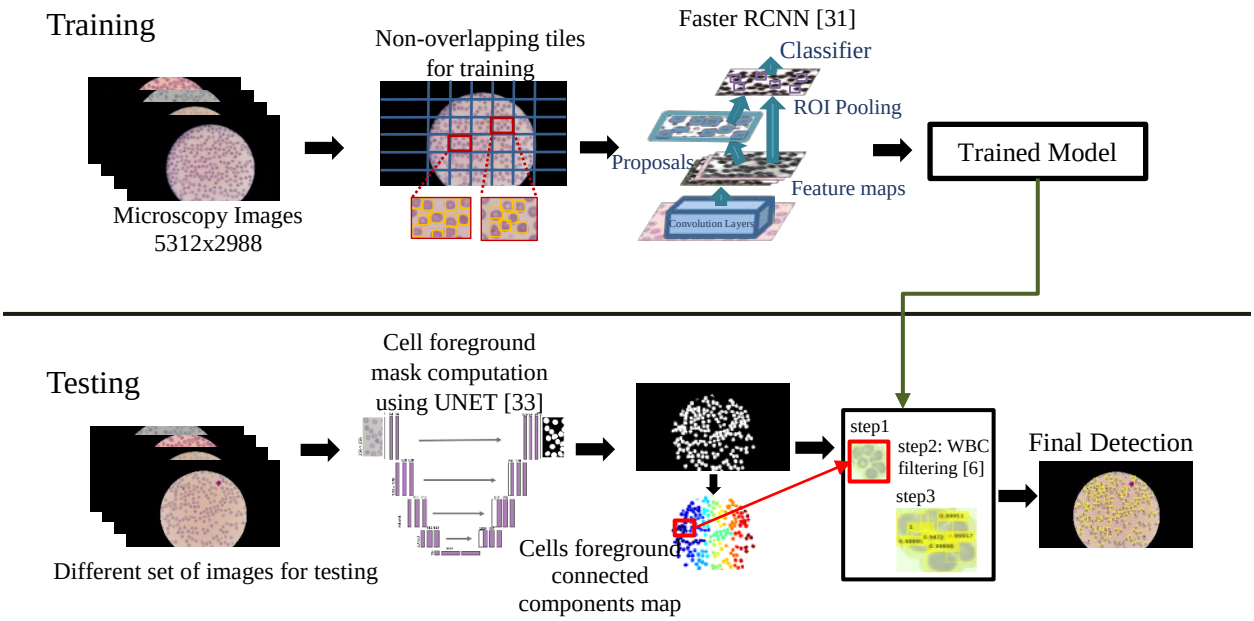


Figure 3. Deep learning pipeline for cell detection and segmentation.

We compare the results of our deep-learning method with our Watershed and active contour methods. Our deep-learning architecture provides a more accurate cell detection than the other two approaches because a foreground mask guides the prediction, which leads to a notably higher true positive rate. Table 1 shows the comparison, where numbers in bold represent the best values.

Table 1: Cell segmentation evaluation

Method	F1-score	Precision	Recall
Unet + Faster RCNN	**97.94**	**97.54**	**98.39**
Active Contours [9, 24]	95.66	94.98	96.43
Watershed [30]	94.30	95.55	93.51

Our evaluation considers the F1-score, precision and recall, for which we achieve 97.94%, 97.54%, and 98.39%, respectively, on our test set containing 800 images from 160 patients, with five smear images per

patient and 162,443 red blood cells in total. Figure 4 shows an example of our deep-learning segmentation for a thin blood smear image.

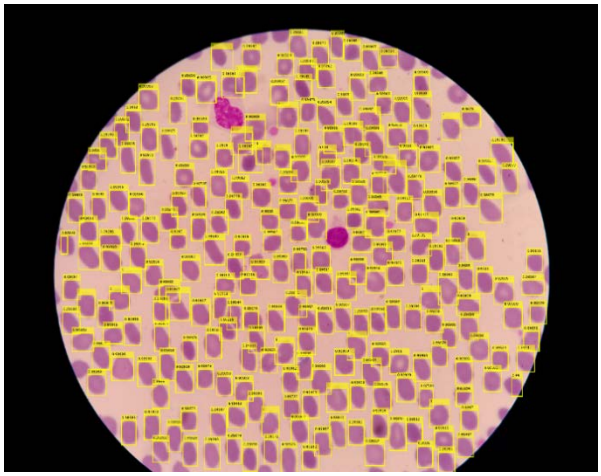


Figure 4. Cell segmentation with deep learning.

Cell Classification

For classification of segmented cells (normal cells vs. parasitic cells), we evaluated the performance of different machine learning methods including Support Vector Machine (SVM), Artificial Neural Network (ANN), and more recently Convolutional Neural Network (CNN). In particular, we evaluated the performance of pre-trained CNNs as feature extractors toward classifying parasitized and uninfected cells segmented by our cell segmentation methods. We experimentally determine the optimal model layers for feature extraction from the underlying data [3]. Figure 5 shows our customized model for malaria cell classification. It achieves 98.9% classification accuracy with lower model complexity and computation time than other models. Moreover, it considerably outperforms the state-of the-art including other pre-trained DL models. This is shown by the performance measures in Table 2, where AUC is the area under the ROC curve and MCC is the Matthews Correlation Coefficient [39]. The latter is a balanced measure that can be used even if the classes are of very different sizes. The classification performance on patient level is lower than on cell level because on patient level all cells of a patient are either in the training set or in the test set, giving the classifier no a priori knowledge about the patient.

Table 2: Comparison of our deep learning cell classification with the state-of-the-art

Method	Accuracy	Sensitivity	Specificity	AUC	F1-score	MCC
Proposed model (cell level) [2]	**98.9**	**99.2**	**98.8**	**99.9**	**99.0**	**97.9**
Proposed model (patient level) [2]	95.1	94.6	95.7	98.5	95.2	90.1
Das et al. [34]	84.0	98.1	68.9	-	-	-
Ross et al. [35]	73.0	85.0	-	-	-	-
Dong et al. 36]	98.1	-	-	-	-	-
Liang et al. [8]	97.3	96.9	97.7	-	-	-
Bibin et al. [37]	96.3	97.6	95.9	-	-	-
Gopakumar et al. [38]	97.7	97.1	98.5	-	-	73.1

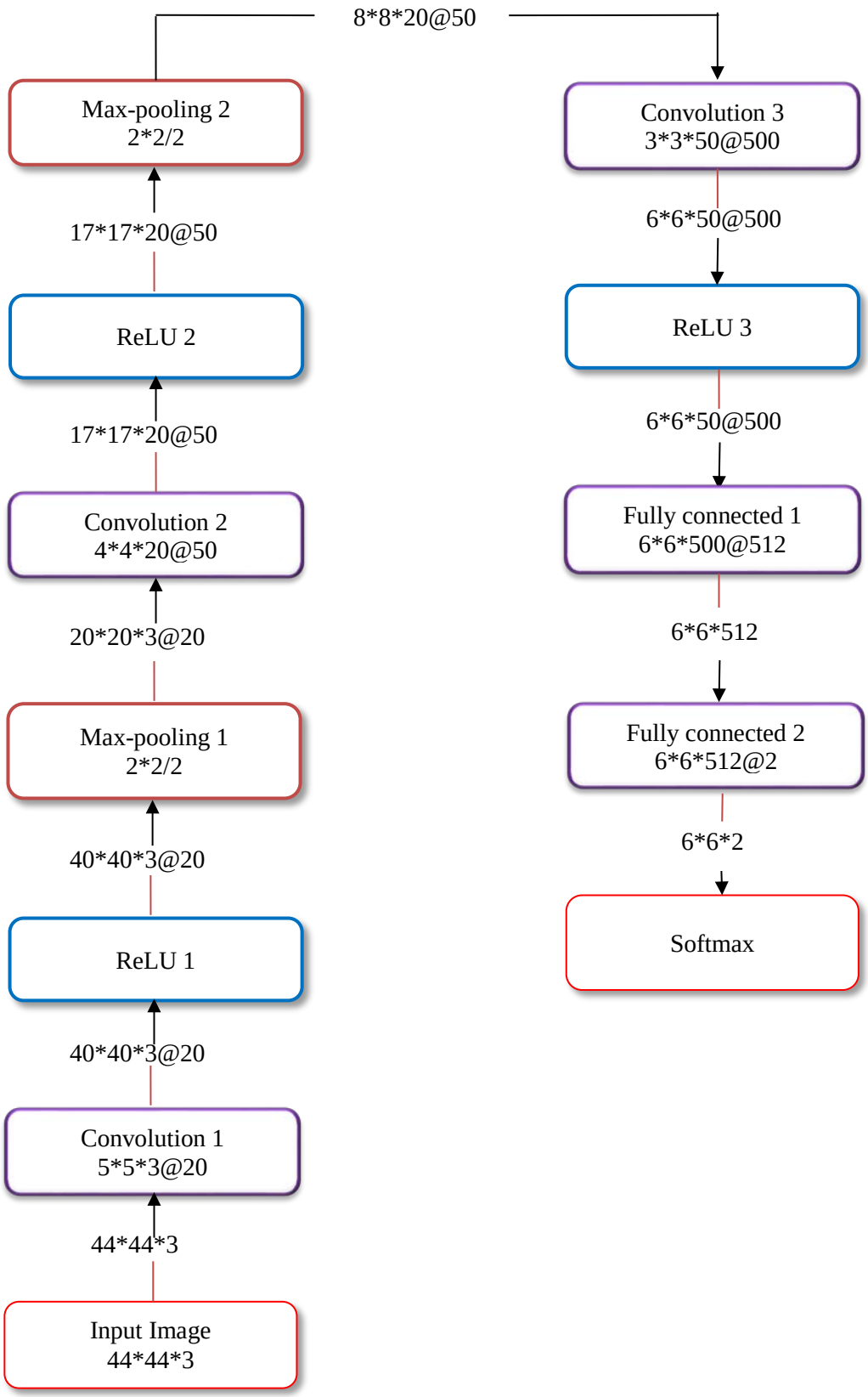


Figure 5. Customized convolutional neural network for cell classification.

Although CNNs have become the architecture of choice for visual recognition tasks, we recognize that they have often been perceived as black boxes because of a general lack of understanding of the learned behavior. This lack of transparency has been a serious drawback, particularly in applications involving medical screening and diagnosis because poorly understood model behavior could adversely affect the process of decision-making. To address this, we are working on methods to visualize the features and activations in deep learning models toward understanding and interpreting their predictions. In our present work, we visualize the weights, saliencies, and class activation maps (CAM) and localize the region of interest (ROI) to explain model predictions. This provides an explanation for a model's classification decision. Figure 6 shows an example of a gradient-weighted CAM visualization of a red blood cell classified by our deep learning network. The parasite is a salient feature detected by the network, as can be seen in the computed heatmap [2, 7].

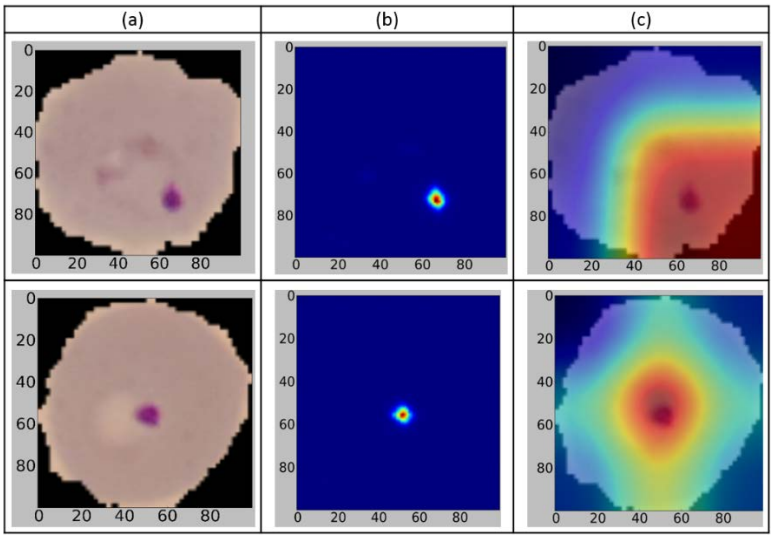


Figure 6. Gradient-weighted CAM visualization: (a) input parasitized cell images, (b) guided saliency maps, and (c) grad-CAM outputs.

Thick Smear Processing

As mentioned earlier, thick blood smears are also important for diagnosing or screening for malaria. Thick blood smears contain more blood and are used to detect the presence of malaria parasites, whereas thin blood smears are used to quantify the degree of parasitemia or to differentiate parasite species. To detect parasites in digital images of thick blood smears, we propose a framework that consists of two stages: screening and prediction. First, we use an intensity-based greedy method to preselect the parasite candidate regions of interest. Second, we train a CNN model to classify the preselected candidates as either parasite or background. The training set for our CNN consists of 2D image patches containing parasites (positive patches) and the same number of 2D patches without parasites (negative patches). Figure 7 shows the flowchart of our proposed framework.

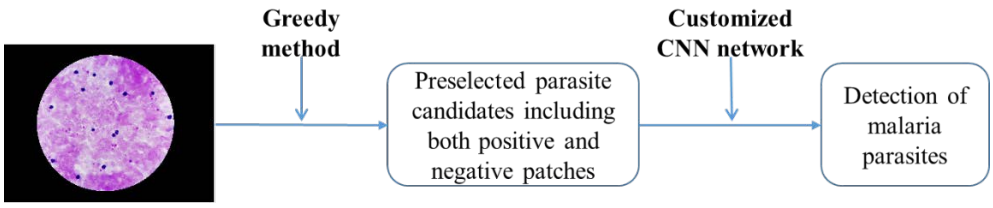


Figure 7. Flowchart of thick smear processing for automated malaria diagnosis.

Parasite candidates preselection via intensity-based greedy method

The screening step reduces the size of the initial search space and preselects a subset of the most suspect parasite candidates. Our intensity-based greedy method selects parasite candidates according to the lowest grayscale intensities, assuming that the nucleus of parasites have darker intensities than the background. To count white blood cells and eliminate their influence on the parasite screening process, we remove them using an Otsu's thresholding method that succeeds due to the relatively higher intensity and contrast of these cells and their much larger size. In our experiments, we extract 500 parasite candidates for each image to ensure that all true parasites become candidates, and to avoid false negatives.

Parasite candidate classification via CNN model

Our proposed customized CNN model consists of seven convolutional layers, three max-pooling layers, three fully connected layers, and a softmax layer as shown in Figure 8, in which the numbers above the cuboids indicate the dimensions of the feature maps, and the numbers below the green dotted line represent the convolutional kernel sizes and the sizes of the max-pooling regions. A batch normalization layer is used after every convolution layer to allow a much higher learning rate and robustness against different initializations, followed by a rectified linear unit (ReLU) as the activation function [21]. We introduce Max-pooling layers after every two successive convolutional layers to select feature subsets. The last convolutional feature map is connected to three fully-connected layers with 512, 50 and 2 hidden units, respectively. Between the three fully connected layers, two dropout layers with a dropout ratio of 0.5 are used to reduce model overfitting [22]. The output of the CNN model is a score vector, with the probabilities of the input being a parasite or a normal patch. We can apply a threshold to these probabilities to obtain a larger or smaller number of predicted parasites.

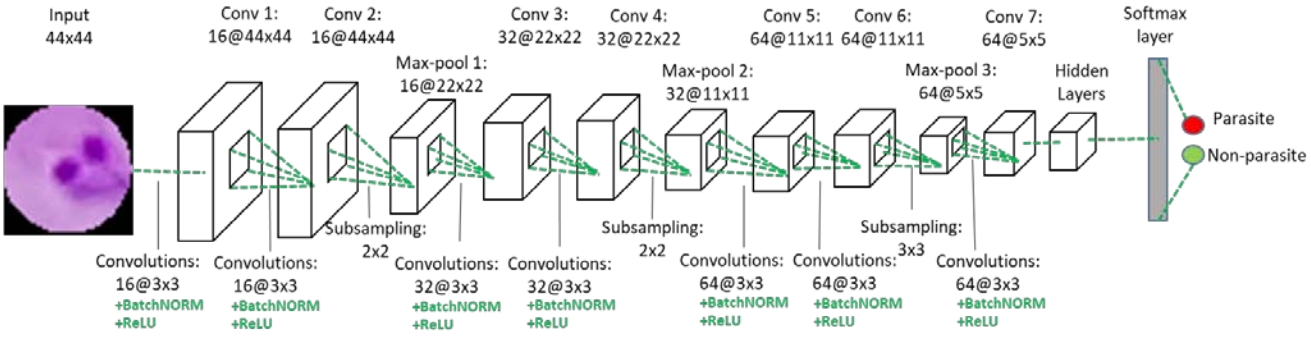


Figure 8. Architecture of our customized CNN model for classification of parasite candidates.

Results

For training and testing our CNN model, we use a subset of our thick smear images, which includes 1444 images and 72,184 parasites from 120 patients. Table 3 shows the results we obtained based on a five-fold cross-evaluation, where we averaged the results over all folds, in terms of accuracy, F1-score, specificity, sensitivity, precision, and area under the ROC curve (AUC).

Table 3: Average parasite classification performance on five folds.

Method	Accuracy	F1-score	Specificity	Sensitivity	Precision	AUC
Mean	**93.46**	**93.40**	**94.33**	**92.59**	**94.25**	**98.39**
Std Deviation	0.32	0.33	1.25	1.27	1.13	0.18

To test our proposed parasite screening method on an independent set, we use the remaining images of our thick smear image repository, including 375 images and 12,777 parasites from 30 patients, for a linear regression between our automatically computed parasite counts and the manual expert counts. Figure 9 shows a strong correlation between the predicted parasite counts and the expert counts, with a correlation coefficient above 0.98, on both image level and patient level.

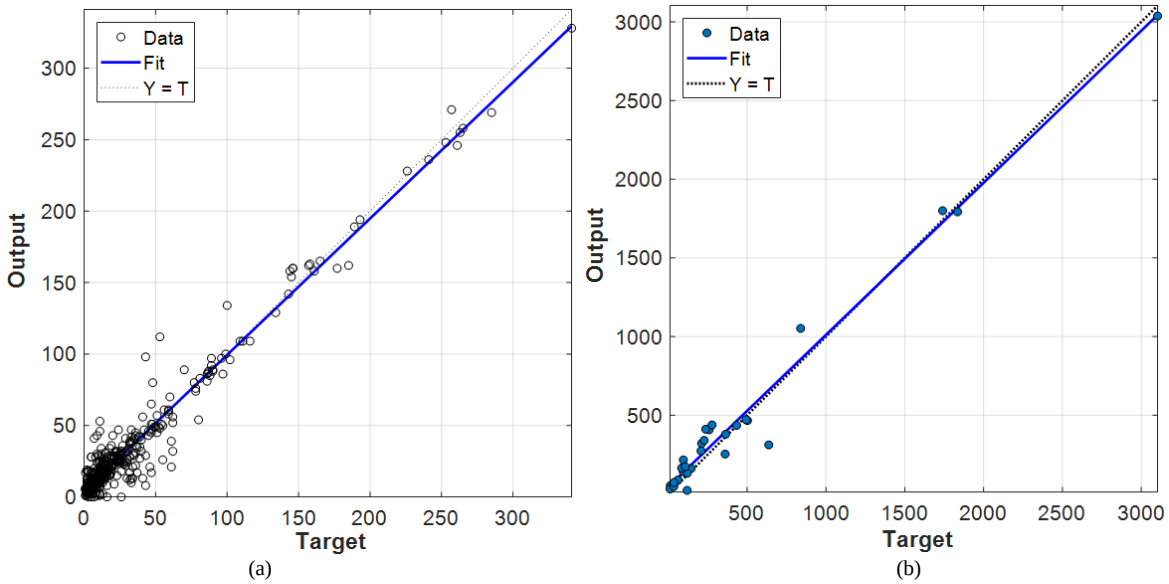
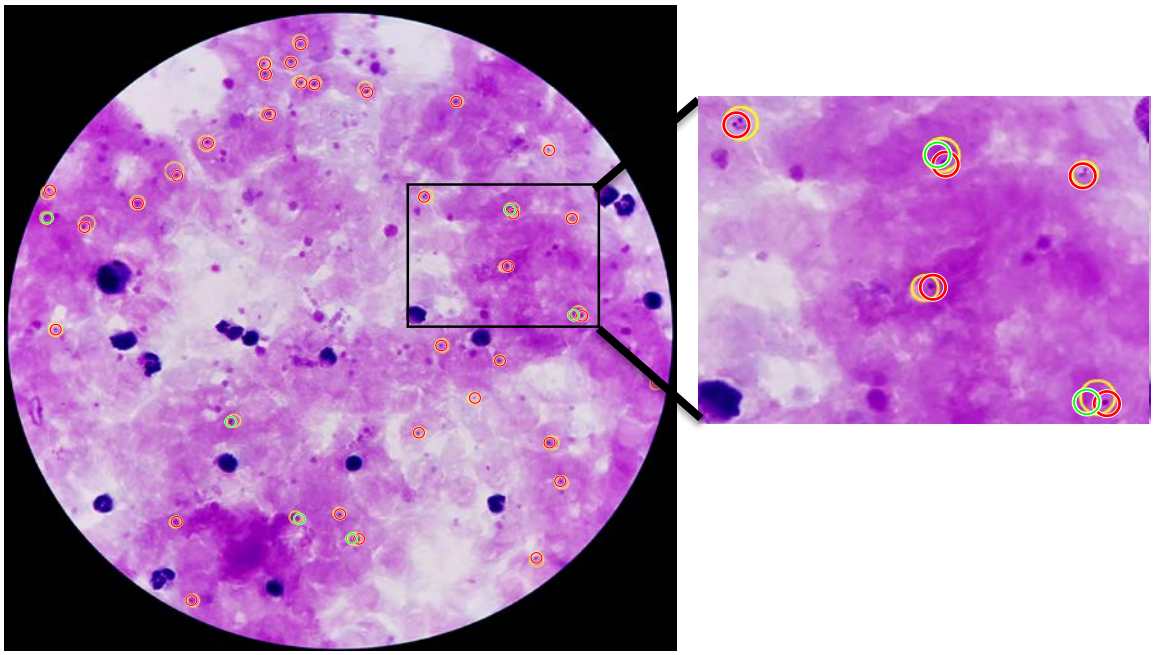


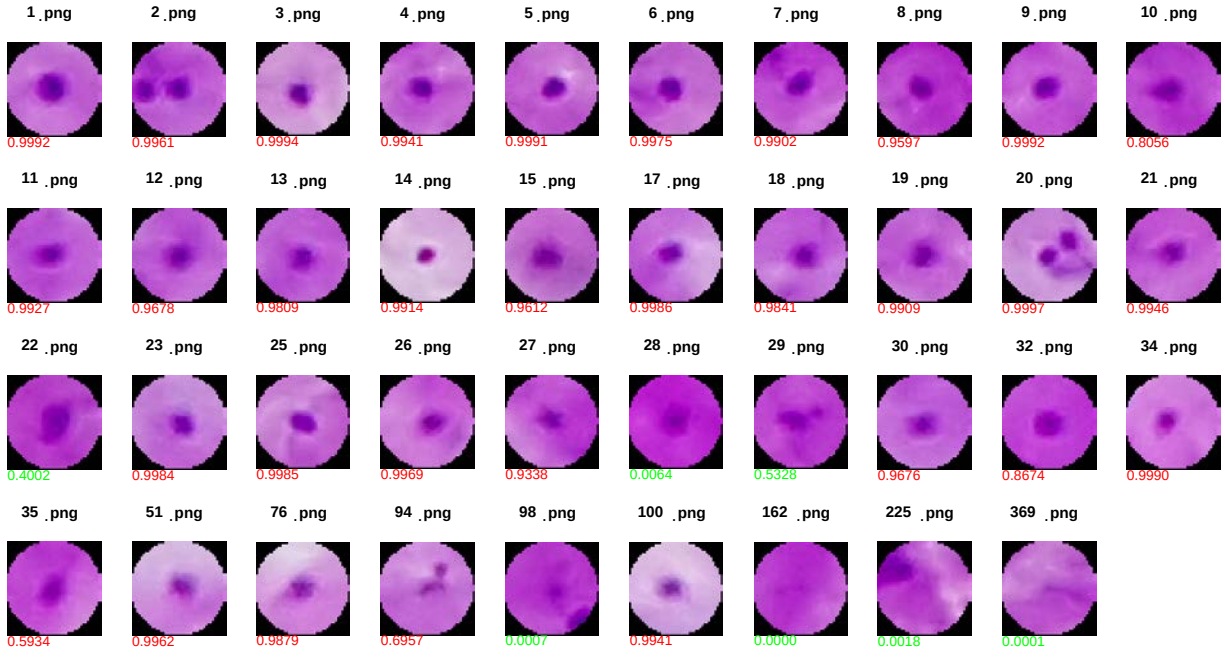
Figure 9. Linear regression between the automatically computed parasite counts (output) and the expert counts (target) on image level (a) and patient level (b).

Figure 10(a) gives an example of our parasite detection using the proposed framework. Parasites annotated by the expert are marked by yellow circles. Red and green circles indicate the preselected parasite candidates (using our greedy method) that overlap more than 50% with the ground truth (regions marked by the expert), with red circles indicating the patches predicted as parasites by the CNN model, and green circles indicating those predicted as non-parasites by the CNN model. The rectangular region displayed in Figure 10(a) is shown enlarged on the right-hand side for a better illustration. In this area, there are five annotated parasites (yellow circles). Our greedy method has identified seven parasite candidates in this area (red and green circles), and five of them are predicted as parasites by our customized CNN model (red circles). Figure 10(b) shows the probability of each patch being a parasite, as predicted by the CNN, along

with the patch image. For this example, we have 36 parasites listed in our ground truth data. Our greedy method extracted 39 parasite candidates, with 33 of them predicted as parasites by the CNN model.



(a)



(b)

Figure 10. Parasite prediction on a thick smear image using our proposed framework. (a) Parasites annotated in the ground truth (yellow circles) and preselected parasite candidates (red and green circles). Red circles indicates the preselected candidates that are predicted as parasites, and green circles indicates those that are predicted as non-parasites by our CNN model. (b) The preselected parasite candidates and their probabilities. The number above each patch is the identifier (filename), and the numbers under the patches are the probabilities of each patch being predicted as a parasite by our customized CNN model. Red and green numbers indicate a probability larger than 0.6 or smaller than 0.6, respectively, which was the optimal threshold in our experiments.

Smartphone Tool: NLM MalariaScreener

In this section, we discuss NLM MalariaScreener, an Android mobile app that we developed to test and validate our research in automated blood smear analysis for malaria diagnosis in the field [4]. When using this app, the camera of the smartphone is attached to the eyepiece of the microscope. The user adjusts the microscope to find the target field in the blood smear and takes pictures with the app. The algorithm in the app will then process these images locally. Since in malaria diagnosis both thin and thick blood smears are examined, we designed the app to process both types. The app records the automatic parasite counts along with patient and smear metadata, and saves them in a local database on the smartphone, where they can be used to monitor disease severity, drug effectiveness, and other critical parameters.

We implemented an embedded camera function to preview and capture the image seen through the microscope. Unlike most other camera apps, NLM MalariaScreener avoids digital zooming because we do not want to risk losing image resolution. A user will operate with the optical zoom of the microscope to bring the image into focus and enlarge the image. However, the app does provide the option to adjust white balance. The color of the image is very important since we use color features to describe and process cells; therefore, we give our users the option to adjust the color of the image among different lighting conditions. The app presents the captured image to the user for review. Once the user accepts the image, the app processes the image and counts the infected cells or parasites.

As mentioned above, the app processes both thin and thick blood smears. The processing targets for these two types of smears are different. For thin smears, we are computing the number of red blood cells, including cells infected by malaria parasites and uninfected normal cells. For thick smears, we detect and count parasites directly because there are no red blood cells visible in thick smears. Therefore, we designed the user interface of our app in a way that suits both types of smears and that lets users easily switch between the two types.

For thin smears, we segment all the cells and identify whether they are infected or uninfected. We currently use blob detection in combination with a Watershed algorithm for image segmentation and a support vector machine for cell classification. To implement such algorithms, we use OpenCV4Android SDK, which supports most of the image processing functions needed. After an image is processed, it will be displayed on a result page where all the cells are annotated with colored labels indicating whether they are infected or uninfected. Then, the user moves the slide to a different field in the blood smear to capture and process another image, and repeats this process until enough cells are collected to meet the standard protocol.

For thick smears, we are looking for parasites rather than red blood cells. The app detects, counts, and records parasite numbers and displays the results in the user interface. Similar to thin smears, users will take several images until they have collected enough data to meet the requirements of their local protocols. The app will aggregate the parasite counts across all images. The algorithm we use to process thick smear images is also implemented using the OpenCV4Android SDK library.

After the image acquisition and processing stage, the app will go through a series of input masks for the user to fill in the information associated with the current patient and smear. This information is saved in the local database of the app, which we built with the SQLite API provided by Android. The app offers a user

interface to the database where the user can view the data and images of previous smears, allowing hospital staff to monitor the condition of patients.

Since malaria is a disease that is widespread in different areas around the world, the app does support several languages to accommodate users of different countries. With English being the default language, the app also supports Thai and Chinese. We are working on supporting more languages.

OMERO

As a service to our collaborators and the general community, we have installed the microscopy environment OMERO that allows remote users to apply our algorithms to their blood smear images to quantify the degree of malaria. OMERO is an open source client/server system written in Java for visualizing, managing, and annotating microscope images and metadata. It is a joint project between European and U.S. universities, with University of Dundee playing a leading role. OMERO allows us to establish a client-server platform where users can run our programs from anywhere via internet. We have developed a prototype program that allows users to upload a thin blood smear image to a cloud storage, segment the cells in the image by our segmentation method, and visualize the segmentation results and cell counts.

5. Project Status

As noted, we have implemented processing methods for thin and thick smears, which we trained and tested on our images from Bangladesh. The performance is generally very promising. Deep learning has outperformed traditional methods in cell detection, cell segmentation, cell classification, and parasite classification. The results we published in the literature have been received well by the research community. So far, we have run tests of our methods at the following universities and hospitals:

- Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand
- Chittagong Medical College & Hospital, Chittagong, Bangladesh
- Lyantonde Hospital, Lyantonde, Uganda
- Holy Innocents Children Hospital, Mbarara, Uganda.
- University of Veterinary and Animal Sciences, Lahore, Pakistan

We have already made some of our training data of this project publicly available alongside our publications. In the near future, we plan to make the entire repository available to the research community, either within the framework of a benchmark competition or simply for download. We also provide access to our smartphone application for partners interested in collaborating with us. In the future, we will release the software to a public open-source repository.

We are currently working on adding deep learning models to our app as classifiers for both thin and thick smears. The deep learning classifier models that were developed by our researchers have been re-trained based on the DL4J (Deep learning for Java) framework so that they can be added to our Android app. We will also experiment with TensorFlow to retrain our models in order to find the best deep learning framework for our app.

NLM MalariaScreener is an innovative mobile app that provides a fast and inexpensive solution for malaria diagnosis in the field. The average processing speed per image is 5.4 seconds for the SVM classifier (on a Samsung Galaxy S6 phone), which is several times faster than a human would be able to count. The accuracy of cell classification on image level is around 96%. The workflow of the diagnosis and the user interface is very intuitive. The app also provides a detailed user manual. It takes a first-time user very little time to learn how to operate the app with no training.

We have also started research into automated discrimination between parasite species, in particular between *P. falciparum* and *P. vivax*, and between different parasite development stages [26, 29]. These problems are important for both diagnosing malaria in the field and for laboratory research. Extending our methods to automate these classification problems would then cover the full spectrum of microscopic malaria diagnosis. Therefore, in 2019, in addition to thin smear images for *P. vivax*, we plan to acquire and annotate images of thick blood smears for *P. vivax*.

6. Evaluation Plan

We have evaluated our system performance by comparing our automated counts with ground-truth counts produced by expert microscopists, as shown in Figure 9. Evaluations were computed on cell-level, image-level, and patient-level depending on the annotations available. For the future, we plan extensive field-testing at various international sites, in particular with our partner in Thailand. However, to test the stability of our methods across a wider range of laboratories and hospitals, with potentially different smear preparation protocols and local malaria parameters, we are also seeking collaboration with other partners. For example, we are about to finish a field study in Uganda, where we are running our software in parallel to the daily screening routine, comparing manual counts with our automatic counts for thin and thick smears. In addition, sites in India, Vietnam, United States, and other countries have expressed their interest to collaborate with us on researching and testing automated malaria diagnosis. To continuously improve our algorithms, we add all images collected via phone in the field to our image collection for training and testing. We will explore running benchmark competitions with our data as another potential avenue to evaluating our methods.

7. Project Schedule and Resources

The project started in 2014 when we saw researchers manually counting infected cells for malaria diagnosis at the National Institute of Allergy and Infectious Diseases (NIAID). The idea was born to find a more efficient workflow for researchers in the laboratory, as well as workers in the field, by automating cell counting. Since then, the project has made rapid progress as shown in Table 4.

We use the following tools and computation resources for this project:

Tools. For annotating training images, we use the Firefly web-based annotation tool. For developing our deep learning models, we use Matlab, Python, and Keras with Theano/Tensorflow backend. We use Singularity and Docker containers for packaging scientific workflows, software, libraries, and data. For statistical validation, we use IBM SPSS version 25.0. To develop our smartphone app, we use Android, OpenCV4Android SDK library, and DL4J.

Computation Resources. We use resources at both NIH as well as in-house at Lister Hill Center: Biowulf (95,000+ core/30 PB Linux cluster) at NIH high performance computing facility (HPC); LHC's NVIDIA DGX-1 workstation with Ubuntu Linux host OS and V100 GPUs for accelerated DL applications, and Windows and Linux desktop computers with multiple NVIDIA GEFORCE GTX 1080 GPUs.

Table 4: Project Timeline.

2014	- Initial project idea after NIAID visit - Image acquisition and proof of concept in MATLAB for mouse malaria
2015	- HHS Ventures Fund Award - Official project start
2016	- Image acquisition and annotation of **thin blood smears for *P. falciparum*** - First ever application of deep learning to malaria diagnosis - First implementation of cell detection, segmentation, and classification in NLM MalariaScreener to help testing and validating research results
2017	- Image acquisition and annotation of **thick blood smears for *P. falciparum*** - First methods for parasite detection in thick smears - First deep learning method in NLM MalariaScreener
2018	- Image acquisition and annotation of **thin blood smears for *P. vivax*** - First deep learning pipeline for thick smear processing of *P. falciparum* - Porting of thick smear methods to NLM MalariaScreener - Starting automatic detection of parasite types and development stages - Starting first systematic field test in Uganda

8. Summary and Future Plan

Malaria is a serious global health problem, claiming more than 400,000 lives per year. It is caused by parasites transmitted through mosquito bites, which infect the red blood cells and lead to symptoms such as seizures and coma in severe cases. We are investigating deep learning and image analysis for a computer-aided system for malaria screening because a fast and reliable diagnosis of malaria is one of the most promising ways of fighting the disease. In fact, we were the first to use deep learning methods for malaria screening to the best of our knowledge. The common method for malaria diagnosis is microscopy, in which an expert visually inspects blood smears for parasites. With millions of smears inspected every year all over the globe, this is an extremely laborious, costly, and unreliable process. However, accurate parasite counts are essential to diagnosing malaria correctly, testing for drug-resistance, measuring drug-effectiveness, and classifying disease severity. Another problem is that microscopic diagnostics is not standardized; it depends heavily on the experience and the skill of the microscopist, which may lead to incorrect diagnostic decisions in the field. We have shown that deep learning and image analysis techniques can play key roles in fighting a major disease like malaria and improving diagnosis. Our software can count parasites and parasite-infected as well as uninfected red blood cells in digital images of blood smears. By computing the quantitative content of parasites in blood smear images, it relieves microscopists and field workers of this tedious task. Our software offers several advantages: It provides a reliable and standardized interpretation of blood smears and it reduces diagnostic costs by reducing the workload through automation.

To validate and test our software, we have designed a smartphone application called NLM MalariaScreener, which runs on a phone attached to a microscope where it can process blood smear images captured by the smartphone camera. LHC has trained NLM MalariaScreener on hundreds of thousands of manually annotated blood cells to learn the typical visual appearance of infected and uninfected cells and discriminate between both. This training set of images and annotations acquired by LHC at a hospital in Bangladesh is one of the largest training sets for malaria parasites in the world (Big Data). NLM MalariaScreener uses novel deep learning techniques and advanced imaging methods for cell detection, segmentation, and classification. It runs on a highly portable and inexpensive smartphone platform for field use in resource poor settings. NLM MalariaScreener is an independent system that does not need an internet connection. Nevertheless, future research could target remote decision making and interfacing with an electronic health record system **(**EHR**)**. The app takes the human expertise out of the equation of malaria diagnosis, thus giving it the potential to expand malaria diagnosis to more regions and populations around the world than we can reach now. We expect it to save a large amount of manual labor on the part of field workers and hospital medical staff, making hospitals function more efficient, consistent, and accurate when it comes to malaria diagnosis. MalariaScreener is currently being used and tested in several malaria-prone regions around the world and has drawn interest from many researchers. The software is freely available for interested partner sites and will eventually become publicly available.

To Malaria and Beyond with Deep Learning

In future work, we plan to add deep learning methods to NLM MalariaScreener that can discriminate between different parasite species and development stages, which is important for proper patient medication and research [32]. Moreover, automatic blood analysis could help in the diagnosis of a wide spectrum of diseases encountered in the blood, parasitic and non-parasitic, including but not limited to other mosquito-borne tropical diseases. For example, we could extend our software to help in the diagnosis of parasitic diseases such as the following, among others:

- **Babesiosis** is a malaria-like parasitic disease caused by the *Babesia* species that infect red blood cells. Most human cases of *Babesia* infection in the United States are caused by the parasite *Babesia microti*. Tick-borne transmission is most common in particular regions and seasons. In symptomatic people, *babesiosis* usually is diagnosed by examining blood under a microscope and seeing *Babesia* parasites inside red blood cells.
- **Elephantiasis** is a parasitic disease caused by microscopic, thread-like worms. It affects over 120 million people in 73 countries throughout the tropics and sub-tropics of Asia, Africa, the Western Pacific, and parts of the Caribbean and South America. The disease spreads from person to person by mosquito bites. It affects the lymph system that results in fluid collection and swelling, affecting mostly the legs. The standard method for diagnosing active infection is the identification of parasites by microscopic examination.
- **Chagas disease** is a parasitic tropical disease that is spread to humans by insects. The parasites cause symptoms like fever, enlarged lymph nodes, and headache. Approximately 6.6 million people are infected with Chagas, with about 8,000 deaths, as of 2015. Diagnosis of early disease is by finding the parasite in the blood using a microscope.

We can train deep learning methods to detect all kinds of foreign organisms and abnormalities in the blood, such as parasites. Therefore, we plan to extend NLM MalariaScreener to a general computer-aided blood screening and monitoring tool that we can use to detect and analyze a broad spectrum of abnormalities in

the blood. To do so, we will acquire training data so that we can train deep learning models to discriminate between healthy and unhealthy states. Similarly, we will adapt our image segmentation methods to the new objects of interest.

9. Related publications and other references

Related Publications

Journal Articles

[1] Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma GR. Image analysis and machine learning for detecting malaria. J. of Laboratory and Clinical Medicine. In-Depth Review: Diagnostic Medical Imaging (Translational Research), v.194, pp. 36:55, 2018.

[2] Rajaraman S, Silamut K, Hossain MA, Ersoy I, Maude RJ, Jaeger S, Thoma GR, Antani SK. Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images. J. Med. Imag. 5(3), 034501, 2018.

[3] Rajaraman S, Antani SK, Poostchi M, Silamut K, Hossain MA, Maude RJ, Jaeger S, Thoma GR. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ, 2018.

Conference Papers

[4] Yang F, Yu H, Poostchi M, Maude RJ, Jaeger S. Smartphone-Supported Automated Malaria Parasite Detection, SIIM conference on Machine Intelligence in Medical Imaging, 2018.

[5] Moallem G, Sari-Sarraf H, Poostchi M, Maude RJ, Silamut K, Antani S, Jaeger S, Thoma GR. Detecting and segmenting overlapping red blood cells in microscopic images of thin blood smears. Proc. SPIE Medical Imaging: Digital Pathology, vol.10581, pp. 105811F, 2018.

[6] Moallem G, Poostchi M, Yu H, Silamut K, Palaniappan N, Antani SK, Hossain Md Amir, Maude RJ, Jaeger S, Thoma GR. Detecting and segmenting white blood cells in microscopy images of thin blood smears. IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2017.

[7] Rajaraman S, Antani SK, Jaeger S. Visualizing deep learning activations for improved malaria cell classification. Proceedings of the First Workshop in Medical Informatics and Healthcare, v. 69, p. 40-47, 2017.

[8] Liang Z, Powell A, Ersoy I, Poostchi M, Silamut K, Palaniappan K, Guo P, Hossain M, Antani SK, Maude R, Huang J, Jaeger S, Thoma GR. CNN-based image analysis for malaria diagnosis. IEEE International Conference on Bioinformatics & Biomedicine (BIBM), pp. 493-496, 2016.

[9] Poostchi M, Ersoy I, Gordon E, Bansal A, Palaniappan K, Pierce S, Antani S, Thoma GR, Jaeger S. Image analysis of blood slides for automatic malaria diagnosis. NIH-IEEE Strategic Conference on Healthcare Innovations and Point-of-Care Technologies for Precision Medicine (HI-POCT), 2015.

[10] Gordon E, Waisberg M, Ersoy I, Pena M, Jaeger S, Pierce S. The eye as a window to investigate the CNS microvasculature during a dynamic malaria infection. Third Annual Seminar on Molecular Imaging of Infectious Diseases, 2013. [Abstract]

Poster Presentations

[11] Poostchi M, Yang F, Yu H, Maude RJ, Silamut K, Jaeger S, Thoma GR. Computational Image Analysis and Deep Learning for Malaria Parasite Detection. National Library of Medicine Informatics Training Conference, 2018.

[12] Poostchi M, Liang Z, Jaeger S, Thoma GR. Malaria Parasite Detection and Quantification Using Deep Learning, CVPR Women in Computer Vision Workshop, 2018.
[13] Moallem G, Poostchi M, Yu H, Palaniappan K, Silamut K, Maude RJ, Hossain MA, Jaeger S, Antani SK, Thoma GR. Automatic white blood cell detection and segmentation in microscopy images of thin blood smears. Annual Meeting of the American Society of Tropical Medicine & Hygiene (ASTMH), 2017.
[14] Jaeger S, Silamut K, Yu H, Poostchi M, Ersoy I, Powell A, Liang Z, Hossain M, Antani SK, Palaniappan K, Maude R, Thoma GR. Reducing the diagnostic burden of malaria using microscopy image analysis and machine learning in the field. Annual Meeting of the American Society of Tropical Medicine & Hygiene (ASTMH), 2016.
[15] Gordon, E, Ersoy, I, Jaeger S, Waisberg, M, Pena, M, Thoma GR, Antani SK, Pierce, S, Palaniappan, K. Retinal microcirculation dynamics during an active malarial infection. Annual Meeting of the American Society of Tropical Medicine and Hygiene (ASTMH), 2014.

References

[16] WHO. World Malaria Report. 2017.
[17] WHO. Microscopy for the detection, identification and quantification of malaria parasites on stained thick and thin blood films in research settings, 2015.
[18] WHO. Guidelines for the treatment of malaria (3rd edition). World Health Organization, 2015.
[19] Makhija KS, Maloney S, Norton R. The utility of serial blood film testing for the diagnosis of malaria. J. of Pathology, vol. 47, no. 1, pp. 68–70, 2015.
[20] WHO. Malaria microscopy quality assurance manual (v.2). World Health Organization, 2016.
[21] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Int. Conf. on Machine Learning (ICML), vol. 37, no. 2, pp. 81–87, 2015.
[22] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J. Mach. Learn. Res., vol. 15, pp. 1929–1958, 2014.
[23] Kassim YM, Palaniappan K, Maude RJ, Thoma G, Jaeger S. Robust detection for malaria cells in thin blood smear microscopy using deep learning, SIIM conference on Machine Intelligence in Medical Imaging, 2018.
[24] Ersoy I, Bunyak F, Higgins J, Palaniappan K. Coupled edge profile active contours for red blood cell flow analysis. IEEE International Symposium on Biomedical Imaging, pp. 748-751, 2012.
[25] Dong Y, Jiang Z, Shen H, Pan WD, Williams LA, Reddy V, Benjamin WH, Bryan AW. Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. International Conference on Biomedical and Health Informatics, pp.101--104, 2017.
[26] Das DK, Maiti AK, Chakraborty C. Automated system for characterization and classification of malaria-infected stages using light microscopic images of thin blood smears. Journal of microscopy, vol. 257(3), pp. 238-252, 2015.
[27] Pirnstill CW, Cote GL. Malaria diagnosis using mobile phone polarized microscope. Scientific Reports, 5:13368, 2015.
[28] Tek FB, Dempster AG, Kale I. Parasite detection and identification for automated thin blood film malaria diagnosis. Computer vision and image understanding, vol. 114(1), pp. 21-32, 2010.
[29] Silamut K, White NJ. Relation of the stage of parasite development in the peripheral blood to prognosis in severe falciparum malaria. Transactions of the Royal Society of Tropical Medicine and Hygiene, vol. 87(4), pp. 436-443, 1993.
[30] Rafael GC, Woods RE. Digital Image Processing, 3rd edition. Prentice Hall, 2007.

[31] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in neural information processing systems, pp.91-99, 2015.
[32] He K, Gkioxari G, Doll P, Girshick R. Mask R-CNN. IEEE Int. Conf. on Computer Vision (ICCV), pp. 2980-2988, 2017.
[33] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Int. Conf. on Med. Image Computing and Computer-assisted Intervention, pp. 234-241, 2015.
[34] Das DK, Ghosh M, Pal M, Maiti AK, Chakraborty C. Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron. 45, pp. 97–106, 2012.
[35] Ross NE, Pritchard CJ, Rubin DM, Duse AG. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. J. of Med. Biol. Eng. Comput. Vol. 44(5), pp.427–436, 2006.
[36] Dong Y, Jiang Z, Shen H, Pan WD, Williams LA, Reddy VB, Benjamin WH, Bryan AW. Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. IEEE Int. Conf. on Biomedical and Health Informatics (BHI), pp. 101–104, 2017.
[37] Bibin D, Nair MS, Punitha P. Malaria Parasite Detection from Peripheral Blood Smear Images Using Deep Belief Networks. IEEE Access. vol. 5, pp. 9099–9108, 2017.
[38] Gopakumar GP, Swetha M, Sai SG, Sai SG. Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. J. of Biophotonics. pp. e201700003, 2018.
[39] Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure. vol. 405(2), pp. 442–451, 1975.

Acknowledgements

We would like to express our most sincere thanks to the following researchers, experts, and students, who have contributed directly or indirectly to this project:

National Institute of Allergy and Infectious Diseases (NIAID):
Laboratory of Malaria and Vector Research:
Dr. Chanaki Amaratunga
Dr. Abhisheka Bansal*
Dr. Rick Fairhurst (Chief, Malaria Pathogenesis and Human Immunity Unit) *
Dr. Louis Miller (Chief, Malaria Cell Biology Section)
Dr. Prakash Srinivasan*

Laboratory of Immunogenetics:
Dr. Susan Pierce (Chief)
Emile Gordon*

Twinbrook Imaging Facility:
Joseph Brzostowski (Chief)

Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand:
Dr. Richard Maude (Head of Epidemiology)
Dr. Kamolrat Silamut (Image annotations) *

University of Missouri:
Dr. Ilker Ersoy (Cell detection and segmentation)
Dr. Kannappan Palaniappan (Firefly annotation tool)

Summer students:
Yasmin Kassim, Dr. Zhaohui Liang, Wilson Lin, Paul LoBuglio, Katie McMenamin, Golnaz Moallem, Nila Palaniappan, Andrew Powell

*as of project start date