Label-free imaging and classification of live *P. falciparum*

Paul M. Lebel[1*], **Rebekah L. Dial**[2], **Venkata N. P. Vemuri**[1], **Valentina Garcia**[2], **Joseph DeRisi**[1,2], **Rafael Gómez-Sjöberg**[1]

[1]Chan Zuckerberg Biohub; [2]University of California, San Francisco

***For correspondence:**
paul.lebel@czbiohub.org (
)

Abstract Manual microscopic inspection of fixed and stained blood smears has remained the gold standard for *Plasmodium* parasitemia analysis for over a century. Unfortunately, smear preparation consumes time and reagents, while manual microscopy is skill-dependent and labor-intensive. Here, we demonstrate that label-free microscopy combined with deep learning enables both life stage classification and accurate parasitemia quantification. Using a custom-built microscope, we find that deep-ultraviolet light enhances image contrast and resolution, achieving four-category classification of *Plasmodium falciparum* blood stages at an overall accuracy greater than 99%. To increase accessibility, we extended our method to a commercial brightfield microscope using near-ultraviolet and visible light. Both systems were tested extrinsically by parasitemia titration, revealing superior performance over manually-scored Giemsa-stained smears, and a limit of detection below 0.1%. Our results suggest that label-free microscopy combined with deep learning could eliminate the need for conventional blood smear analysis.

Introduction

Historically, malaria is among the deadliest infectious diseases in human history. Of the five *Plasmodium* species that cause illness in humans, *P. falciparum* accounts for 99% of cases in Africa, and 94% of all malaria deaths (***WHO*** (*2019*)). Not only is it distributed across many regions of the world, but its eradication is intimately linked to the complexity of its mosquito vector and the human population it infects. If left untreated, malaria causes severe febrile illness or death, especially in children and infants. The World Health Organization reported an estimated 228 million cases in 2018, and an estimated 405,000 deaths globally (***WHO*** (*2019*)), the large majority of which occur in low-resource areas with little access to healthcare, representing some of the most vulnerable communities in the world.

Although anti-malarial drugs have been employed with success for centuries, chemoresistance remains an ongoing threat as strains evolve in response to widespread treatment regiments (***Packard*** (*2014*)). Despite the effectiveness of chloroquine (CQ), the mass administration of sub-curative doses of the drug in Cambodia in 1955 has been credited with the initial evolution of a CQ-resistant strain (***Packard*** (*2014*)), with resistance also appearing independently in South America in the 1960s (***Gama et al.*** (*2011*)). Artemisinin-based Combination Therapies (ACT) are highly effective, and thought to be less susceptible to resistance due to the combination of drugs with distinct mechanisms of action (***Haldar et al.*** (*2018*)). Unfortunately, front line artemisinin resistance has now been documented in both Southeast Asia and Latin America (***Haldar et al.*** (*2018*), ***Mathieu et al.*** (*2020*)). Further, *P. falciparum* also possesses a well-evolved ability to evade host immunity in both the Anopheles vector and the human host through various allelic adaptations and antigenic variation (***Escalante et al.*** (*2009*)). These factors put considerable pressure on the therapeutic development process to continuously identify and target orthogonal mechanisms of action.

Ongoing investigations into topics such as drug-resistance and host immune evasion hinges on the ability to cultivate strains of *P. falciparum* in a laboratory setting. An essential part of *ex vivo* culturing includes daily assessments of the parasitemia and quantification of the life stages. While methods exist for flow-cytometry assisted quantification of parasitemia, manual counting remains the gold standard for evaluating these factors (***Poostchi et al.*** (***2018***)). For each culture, a smear must be prepared which typically involves: blood smearing, drying, fixation, drying again, staining, and rinsing. On average, the staining process alone consumes 45 minutes (***Poostchi et al.*** (***2018***)). Subsequently, several hundred or even thousands of cells must be inspected manually under the microscope in order to quantify parasite stages with sufficient sampling power, consuming 15-30 minutes for a trained microscopist (***Poostchi et al.*** (***2018***); ***Manser et al.*** (***2013***)). Despite publication of standardized training methods, testing, and cross-validation of microscopist competence (***WHO*** (***2016***)), errors in manual microscopy are likely (***Manser et al.*** (***2013***)), since counting depends on stain quality and technician skill level, and can be hampered by fatigue and reagent quality. The entire end-to-end manual procedure is estimated to be carried out hundreds of thousands of times annually (***Poostchi et al.*** (***2018***)).

The fields of label-free imaging and deep learning are both rapidly advancing (***Rivenson et al.*** (***2019***); ***Guo et al.*** (***2020***); ***Tschandl et al.*** (***2020***)), with a promising convergence in applications related to automated detection of blood pathogens. Label-free methods such as Differential Interference Contrast (***Grüring et al.*** (***2011***)) and Optical Diffractive Tomography (***Kim et al.*** (***2014***)) have succeeded in producing detailed maps of RBCs with sufficient resolution for parasite detection and feature extraction. While they represent important advances, neither of these methods have yielded classification results sufficient for routine laboratory practice or diagnostic settings. On the other hand, quantitative phase microscopy (QPM) has been used to achieve high accuracy differentiation between late parasite stages and healthy cells, but lacks the ability to distinguish parasite stages from each other, and further, the method was not tested on ring stages – the most prevalent stage found in peripheral blood (***Park et al.*** (***2016***)). Meanwhile, there are a growing number of efforts to process conventional Giemsa-stained blood smear images using deep convolutional neural networks (***Rajaraman et al.*** (***2019***); ***Poostchi et al.*** (***2018***); ***Dong et al.*** (***2017***); ***Gopakumar et al.*** (***2018***); ***Yang et al.*** (***2019***); ***Pan et al.*** (***2018***)). These recent efforts have achieved high levels of performance by re-training existing deep networks to classify stained images. These also represent important contributions to the field, because although fixation, staining, and manual microscopy must still be performed, technician time spent inspecting images could potentially be eliminated. Automated, low-cost slide scanning has been recently achieved and combined with embedded computing hardware for image analysis (***Li et al.*** (***2019***)). This advance automates the labor-intensive microscopy portion, but still requires time-consuming fixation and staining prior to imaging, and sophisticated scanning stages. Other non-imaging label-free methods have exploited hemozoin crystallization by the parasite (***Peng et al.*** (***2014***)), or by antibody-mediated electrochemical detection (***Kumar et al.*** (***2016***)). Others have exploited the mechanical hardening of infected RBCs using microfluidic strain sensors (***Kang et al.*** (***2016***); ***Yang et al.*** (***2017***)), yielding up to 92% overall accuracy on binary infection calling, although it lacks clear lifecycle stage breakdown and comes at the cost of complex microfabrication techniques.

The absence of label-free parasite classification of ordinary brightfield microscopy images is likely the result of a century-long preconception that the interaction between visible light and biological matter is too weak (insufficient contrast), especially at sub-micron morphological features (insufficient resolution). Additionally, because the prevalence of parasites (parasitemia) is low compared to healthy cells, the performance requirement for such a technique to be useful is high, imposing stringent requirements on the maximum FPR, precision, and recall. To explore this topic we surveyed the wavelength dependence, from deep-ultraviolet (deep UV) to visible light, of image classification performance scoring parasite infection stages (healthy, ring, trophozoite, or schizont) in live, cultured red blood cells. In doing so, we demonstrated that automated classification is possible over a broad spectrum of wavelengths, and that the combination of higher resolution and contrast at shorter wavelengths yields more clearly-resolved parasite physiology as compared with visible light. Additionally, we employed post-processing techniques that further improve upon raw classifier results, and by extensive parasitemia titration show that our calibrated machine classifiers exceed the performance of manually-counted Giemsa blood smears.

Deep UV imaging methods have previously been developed for use in biological imaging. It has been shown, for example, that cellular dry mass can be recorded from fixed cells, or over time in live cells (***Zeskind et al.*** (***2007***)). Additionally, Ultraviolet Hyperspectral Imaging (UHI) has been used to produce comprehensive molecular imaging signatures (***Ojaghi et al.*** (***2018***)), demonstrating the ability to distinguish different cellular components based on their independently unique absorption and dispersion properties. Deep UV excitation is also commonly used by x-ray crystallographers to excite endogenous protein fluorescence, although in that case the imaged light is Stokes' shifted fluorescence emission that is collected at low magnification (***R.a et al.*** (***2005***)). Here we report the use of a custom high-resolution deep UV microscope used in a transmitted light configuration for the purpose of enhanced imaging of live *P. falciparum* parasites, and demonstrate clear advantages in performance as compared to visible light. We additionally survey the effectiveness of a standard inverted commercial microscope equipped with near-UV wavelengths as well as a standard visible light trans-illumination lamp.

Results

Cultured red blood cells infected with *P. falciparum* were injected into quartz flow cells (see Figure 1 —figure supplement 1) and imaged at multiple wavelengths on a custom-built deep UV microscope ($100\times$/0.85 glycerol immersion) as well as an inverted commercial microscope using a simple brightfield modality (Leica DMi8, $40\times$/1.3 oil immersion). Flow cells were imaged for a maximum of approximately 2-3 hours to avoid parasite health decline outside of incubation conditions. Additionally, deep ultraviolet light exposure was minimized by using a hardware synchronization module that only illuminates the sample for the duration of the camera exposure (see Figure 1 — figure supplement 2). Imaging was performed on freshly-prepared flow cells on multiple distinct dates at various parasitemia levels, and the results were later merged computationally for aggregate analysis.

Our image analysis pipeline used custom data classes in Matlab to store, organize, and process RBC images at multiple wavelengths and focal slices. First, individual datasets were imported for pre-processing. Since the quartz UV objective exhibited substantial aberrations such as chromatic focal shift and lateral distortion between color channels, focal stacks for each color were re-aligned vertically as well as laterally via affine transformation to co-register the three channels. Images from the commercial microscope also required focal offset corrections but the effect was less pronounced and was more easily calibrated prior to image acquisition. Next, semantic segmentation was performed by a ResNet-50 network (***He et al.*** (***2015***)) which we trained by manually segmenting RBCs with diverse appearances and parasite lifecycle stages. Binary masks generated by segmentation were post-processed by an instancing algorithm that separated adjacent cells and also rejected those falling outside a certain size and roundness range, and those intersecting with the image boundary. The primary rationale for morphological filtering was to exclude cells with edge-on orientation, those with high degree of crenation (echinocyte formation – see discussion for more details), and spatially overlapping cells.

The main image classification task — labeling the parasite stage for each RBC (or none if healthy) — was performed using a retrained GoogLeNet instance (***Szegedy et al.*** (***2015***)). Initial retraining was achieved by manually sorting a ~5,000 count subset of all individual RBC instance thumbnails into specified directories, using the native operating system file explorer (Windows 10 Professional or Mac OS). A machine classifier was trained on this initial subset, which was then able to accelerate the sorting process. Subsequently, larger annotated datasets for training and validation were

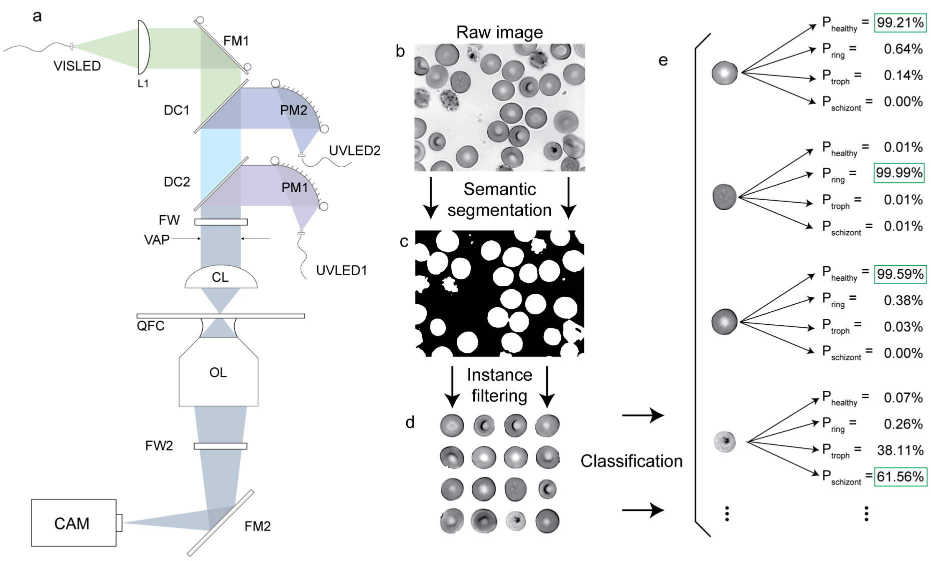


Figure 1. The deep UV microscope (a) was built in a simple transmitted light configuration using a finite conjugate objective lens (OL) to form an image on a camera (CAM) via a rigid fold mirror (FM2) without the need for a specialized tube lens. A custom condenser was built to combine three collimated LEDs (UVLED1, UVLED2, VISLED) using an adjustable fold mirror (FM1) and two dichroic mirrors (DC1, DC2). UV LEDs were collimated using parabolic mirrors (PM1, PM2). Transmitted light numerical aperture was limited with a variable aperture (VAP), then focused onto the sample using a UV fused silica condenser lens (CL). Filter wheels (FW1,2) were added for fluorescence applications (not used in this study). Samples were mounted in Quartz Flow Cells (QFC) for compatibility with deep UV imaging. b) Raw image of parasitized RBCs. c) Binary mask produced by semantic segmentation with a trained ResNet-50 network. d) RBC instances were masked from the raw images and filtered by size and shape parameters to reduce the number of edge-on, misshapen, and/or clipped cells. e) Filtered RBC instances were classified by a retrained GoogLeNet architecture, whose output assigns a probability for each category. Example probabilities are shown for a subset of cells in the raw image. Full classifier results on our dataset are provided as supplementary data.

Figure 1–Figure supplement 1. Flow cells were constructed as follows: 1 mm circular holes (shown in red below) were laser-cut into quartz slides (Ted Pella 26011, outline in blue. Laser-cutter model ULS-P150D). Gaskets (green) were laser-cut from Parafilm (Sigma Aldrich P7793). After cutting, gaskets were aligned by hand, then pressed down gently around the exterior. Quartz coverslips (Ted Pella 26014) were aligned by hand to the gaskets, and then pressed gently. The assembly was sealed by compression of the flow cell (protected inside a single layer aluminum foil envelope) with a mass of 300 g for five minutes, on a hot plate set to 75°C.

Figure 1–Figure supplement 2. Design of the hardware sync module: a) Schematic diagram of the simple circuit used to synchronize LED emission with camera exposure. The circuit uses an analog switch (Maxim DG419) to connect either a programmable analog voltage signal (AIN) or ground (GND) to the output, gated by the camera exposure's logic signal. In series with the analog switch is a manual toggle switch to optional bypass the analog switch, disabling the sync module. b) PCB layout diagram for the hardware sync module. Note that coaxial connector footprints were replaced with standard 0.1" headers, and connected via wires to panel-mount BNC connectors. The circuit was fabricated as a PCB using an LPKF Protomat S103 circuit mill, then hand soldered and assembled. c) Solid model screenshot of the 3D-printed enclosure for the hardware sync module.

Figure 1–Figure supplement 3. Four-channel relay multiplexer. The LED multiplexer uses a four-channel relay (Denkovi DAE-RB/Ro4-JQC-5V) to route the current generated by a Thorlabs LED current driver (LEDD1B) to one of four LEDs. The relay board was mounted inside a box to house the wired connections between relays. The box received a three-pin signal (plus GND). The first relay was used as an overall ON/OFF switch, while the remaining three were arranged in a binary tree to route current from the driver to any of the four LEDs.

Figure 1–Figure supplement 4. The UV microscope control software uses a set of custom classes to configure and execute multi-dimensional image acquisitions. Multiplexed LED illumination was driven by a data class which maps maximum allowable LED currents to control voltages, sending a digital address to the relay multiplexer for LED selection, and an analog voltage representing percent power. The MDEngine class implements the conversion from acquisition parameters to active control of the UVScope, including real-time focus tracking across large sample areas, in order to keep focal stacks centered on the sample. All microscope control software can be found here: https://github.com/czbiohub/UVScope-control.

Figure 1–Figure supplement 5. Overall wiring diagram of the UV Scope. The microscope hardware was controlled by a centralized PC running Windows 10. All the various hardware devices were controlled by a combination of standard and custom Matlab classes, as described in Figure 4. Other specific hardware configurations are possible with minor changes to the software.

Figure 1–Figure supplement 6. Single-wavelength image processing pipeline.

achieved by exporting a fraction of automatically-classified cells with low confidence scores for human correction ("human-in-loop"), which were used to overwrite the imperfect machine labels. In this way, we were able to generate large, fully-annotated datasets consisting of 14,219 cells from the UV scope and over 60,000 cells from the commercial microscope (see Table 1 for dataset statistics). Once our datasets were fully human-annotated, we re-trained new classifiers on a random 90% partition of the pooled N best focus images from all the datasets (N = 5 for UV microscope, N = 1 for commercial microscope). Using additional focus slices served as a natural augmentation of the training dataset size, while simultaneously including examples of slightly de-focused images in the training, in order to reduce the system's dependence on achieving an exact focus.

Using this method, our unmodified four-category classifier was able to achieve an overall classification accuracy of 98.1% for custom UV scope images taken at 285 nm. Full confusion matrices are presented in Figure 2, displaying the precision, recall, false positive rates (FPR), false negative rates (FNR), and mis-classification rates for each category. It is important to note that even for the highest expected parasite densities (laboratory or clinical), samples are always composed predominantly of healthy cells, leading to inherently unbalanced data. Such an imbalance places stringent requirements on the FPR of an image classifier, as even low rates of error will result in a large number of healthy cells labeled as parasitic. We also note that unless the culture is artificially synchronized to the late stages, ring-stage parasites typically predominate over the more mature trophozoites, and even more so over the short-lived schizont stage. The highly unbalanced nature of the sample composition biases the cross-entropy loss function and obscures contributions from minority classes during classifier training. In order to compensate for the imbalance during classifier training, we used the following weighted cross-entropy forward loss function:

$$L = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K}\frac{b_i}{n_i}T_{ni}\log(Y_{ni}) \tag{1}$$

Where N is the total number of training images, n_i are the fractional representation of each class, b_i are empirically-determined training biases, K is the number of classes, Y_{ni} are the predictions, and T_{ni} are the targets (human annotated labels). Here we introduced the term $\frac{b_i}{n_i}$ in order to re-normalize the training weights to account for class imbalance (***Buda et al.*** (***2018***)). As a result, training bias towards dominant classes can be eliminated, such that the resulting classifier's FPR and FNR will, on average, be balanced. For our specific training and validation datasets, it was determined that the optimal values for b_i are [4, 2, 1, 1], corresponding to the classes [healthy, ring, trophozoite, schizont], to re-balance confusion matrices that resulted from processing real samples. The relative balance between false positive and false negative rates is further discussed in the context of confidence thresholding and extrinsic validation.

Annular and dendritic ring stage parasites are both observed

Traditional fixation and Giemsa staining procedures modify the sample in a way that improves contrast for transmitted visible light, but may also modify it in other ways ***Bessis*** (***1974***). For instance, ring-stage parasites are traditionally identified by their intense staining density around a prominent annular region, thought to be caused by nuclear staining combined with an elevated peripheral density of ribosomes (***Bannister et al.*** (***2004***)). However, in addition to the canonical annular form, ring stage parasites may also assume a dendritic (or amoeboid) appearance, as observed in both live-cell imaging (***Grüring et al.*** (***2011***)) and electron microscopy (***Bannister et al.*** (***2004***)), and even the dynamic interconversion between the two forms (***Grüring et al.*** (***2011***)). In the present work we observe both annular and dendritic ring forms with nearly equal regularity in our static images (see Figure 2, Supplement 5), as well as some limited support for dynamic interconversion (Figure 5, supplement 2). We note that although the two forms are visually distinct, in this study our classifiers were trained to recognize a spectrum of formations from annular to amoeboid, all as belonging to a singular ring category.

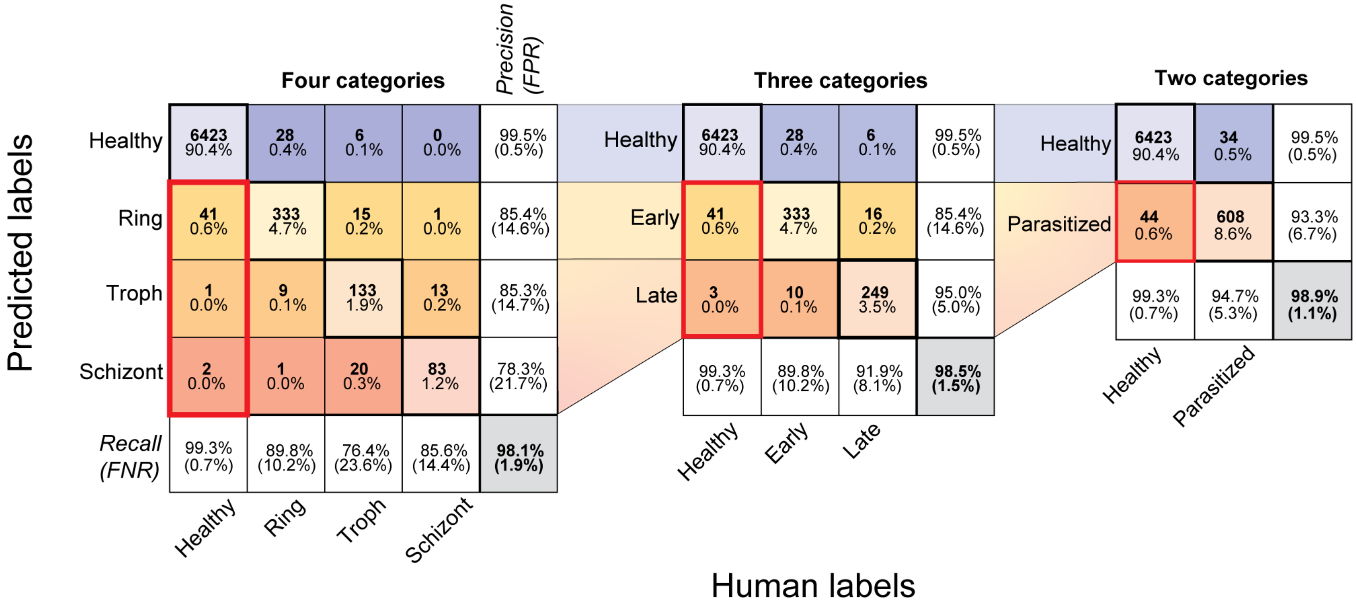


Figure 2. Classification performance of deep UV microscopy at 285 nm: Classification was evaluated with either two, three, or four categories. For each level of granularity, confusion matrices are used to represent detailed classifier performance. Diagonal entries correspond to correctly classified instances and off-diagonal entries correspond to incorrectly classified instances. 'Predicted labels' represents the classifier predicted categories and 'Human labels' represents the human-annotated ground truth. Overall classification accuracy is shown in the bottom right cell of each matrix and is defined as the percent of all instances that were classified correctly. The precision for each category is shown along the far right column (false positive rates underneath in parentheses), while the recall is shown along the bottom row (false negative rates underneath in parentheses). Within each cell in the central colored matrix, the total number of counts is shown (bold, top), as well as the corresponding percentage of the total population (bottom). Thick red boxes indicate the rates of healthy cells classified as parasitic. The classifier was trained on a random 90% partition of the pooled dataset containing all five best focus slices, comprising 58,203 healthy, 3,344 ring, 1,566 trophozoite, and 873 schizont images. The validation was performed on the remaining 10% partition consisting of 6,467 healthy, 371 ring, 174 trophozoite, and 97 schizont images.

Figure 2–Figure supplement 1. Classifier confidence statistics indicate mis-classified cells are more likely to be low-confidence. In all three plots, log-scaled histograms of classifier confidence scores are displayed in the pattern of a confusion matrix. Statistics are derived by evaluating the validation dataset from the 285 nm classifier. Total instance counts for each matrix element as well as the medians of the distributions are shown as text insets.

Figure 2–Figure supplement 2. Confusion matrices from the custom UV scope at 285 nm show improved statistics after rejection of RBCs based on meeting the minimum classifier confidence criteria. These confusion matrices show the result of applying the empirically-optimal threshold values of 61%, 76%, and 96%, to the four, three, and two-category classifier confusion matrices.

Figure 2–Figure supplement 3. A healthy control dataset was processed using our four-category classifier operating at 285 nm. Top: Sample composition estimates are shown as a function of confidence threshold, demonstrating that the approximately 0.5% raw FPR can be reduced by rejecting cells with low classifier confidence. Bottom: The number of cells kept (blue) and rejected (red) by the thresholding process are plotted as a function of threshold value.

Figure 2–Figure supplement 4. Statistics from training the 285 nm classifier. The upper plot shows the overall accuracy as a function of iteration and epoch number. The blue line plot (no markers) shows training dataset accuracy while the black line plot (dashed with circle markers) shows validation accuracy. The lower plot shows the loss function (red with no markers: training loss; black with circle markers: validation loss

Figure 2–Figure supplement 5. Example RBC instances from each category in the confusion matrix for the UV Scope at 285 nm wavelength. The arrangement of the categories is the same as Figure 2 from the main text. For this montage, a maximum array size of 10 x 10 was used for display. Note that for many of the categories (especially off-diagonal entries) there were not enough examples to populate the array, in which case the space was left blank. All data from this figure was taken from the validation dataset, as opposed to the training dataset.

Table 1. Summary of training and validation datasets.

Microscope	Objective	Wavelengths (nm)	Healthy	Rings	Trophs	Schizonts	Total
UV	100×/0.85	285	12,938	734	359	188	14,219
UV	100×/0.85	285,365,565	10,575	261	124	27	10,987
Leica DMi8	40×/1.3	405	56,898	5,213	2,272	583	64,966
Leica DMi8	40×/1.3	365,405,Broadband	61,811	6,502	1,682	4,393	74,388

* UV microscope data additionally included five best focus slices, acting as hardware-based data augmentation system.

** Single- and multi-wavelength datasets consisted of distinct but partially-overlapping merged sets of raw data.

Post-processing improvements to raw classifier output

In addition to reduction of labor-intensive steps, there are other advantages of machine classifiers over human scoring. The potential for longitudinal reproducibility, elimination of inter- and intra-observer variation, the potential to create a centralized repository with consensus expert annotations, and finally, the ability to analyze numerical classifier confidence scores enables further improvements in performance through statistical analysis. We describe multiple ways in which leveraging these attributes can lead to improved performance over raw classifier output.

Category merging

In Figure 2, we summarize the rates of correct and incorrect classification of all the instances in the validation dataset (a 10% random partition of all the annotated data). The confusion matrices compare the results for all four categories from the raw classifier. However, we noted that in some instances lifecycle stage appears to be transitional, share morphological features common to more than one stage, or simply be difficult to distinguish for other reasons. In particular, rings transitioning to the trophozoite stage began accumulating hemozoin (visible as dark highly-absorbing puncta) with highly variable morphologies, and conversely, some early trophozoites had not yet grown in size but exhibited hemozoin accumulations (see Figure 2 – figure supplement 5 for examples). Likewise, many mature trophozoites had grown large in size and accumulated substantial hemozoin, while some early schizonts had only begun displaying increased cytoplasmic texture indicative of nascent merozoite formation. We therefore found it useful to study the statistics for merged classifiers with three- and two- category schemes, which resulted in higher accuracy by not attempting to distinguish borderline transitional instances, at the cost of decreased granularity. The three-category classifier was created by using a single 'late' category; the summation of the trophozoite and schizont probabilities resulted in higher confidence and higher accuracy. The three-category output better reflected partial information in the case of high total confidence spread across two or more categories.

Our four-category classifier achieves a raw overall accuracy of 98.1%, and combined misclassification of all types occurs at a rate of 0.7%. We note that by merging the trophozoite and schizont categories, the precision and recall of the resulting 'late' category is substantially improved (Figure 2); as is the overall accuracy (98.5%), reflecting a reduction in the overall number of misclassified cells. Further reduction of the model to two categories (Healthy and Parasitized) results in an overall accuracy of 98.9%, a parasitic recall rate of 94.7%, and a false-positive rate of 0.7%.

Confidence thresholding

We also noted that although the large majority of the population is classified with very high confidence (healthy cell median confidence is 99.9% – see Figure 2, supplement 1), the confidence distribution has a long tail extending into the low-confidence regime (<90%). This could be explained by many factors including variation of individual parasite anatomy, life cycle stage, relative image focus, non-standard RBC morphology, obstruction by other RBCs, or external stochastic factors such as mechanical vibrations or stage drift. We posited that this subset of cells might exhibit a higher than average error rate, and that excluding it from analysis would be net beneficial to the results. However, it should be noted that removal of data poses an inherent risk of bias and should be applied judiciously. Further, knowledge of the statistics underlying the distribution of confidence scores and their typical correlation with predictive power is essential in applying an appropriate threshold value on classifier confidence. Selection of optimal threshold value is primarily a trade-off between error reduction and introduction of bias. As the threshold value is increased, the rate of rejection of misclassified cells should be higher than the incremental rejection of correctly-classified cells, and the estimate of overall sample composition should improve. However, certain categories may be inherently more difficult to score than others, implying lower confidence values on average. In this case, as the threshold value is increased the more difficult categories will be erroneously rejected at higher rates than easier categories, introducing bias error.

In Figure 2 — supplement 1, we studied the effect of confidence thresholding on our data by plotting the distributions of confidence scores stratified by predicted and human labels, arranged in the form of a confusion matrix. Indeed, the median confidence scores for off-diagonal matrix elements (incorrectly scored cells) were found to be substantially lower than those for on-diagonal positions (correctly scored cells), suggesting that without prior knowledge of ground truth the results could be improved by confidence thresholding. As a performance metric, we considered the estimated overall sample composition accuracy, as opposed to optimization for any one particular element of the confusion matrix. Indeed, the pragmatic output from the machine classifier is the estimated overall sample composition as opposed to the correctness of any one individual cell. Correspondingly, we found the utility of confidence thresholding to be limited in cases where the confusion matrix was already balanced across the diagonal, but more effective in cases where there were more false-positives than false-negatives (or vice-versa). In fact, balanced classifier results were usually worsened with increasing threshold. We further note that the four-category classifier often has difficulty distinguishing late trophozoites from early schizonts, despite a high combined confidence. Confidence thresholding in that case therefore tends to erroneously reject late stage parasites. Merging into a combined 'late' stage category resolves this issue and results in a greater improvement with threshold application.

Confidence-based post-processing strategies

The optimal confidence threshold value cannot be determined without prior knowledge of the underlying sample composition. We therefore discuss methods of improving classifier performance in the absence of prior knowledge, and without the need for large-scale annotation efforts.

False-positive rate parameterization: As shown in Figure 2 supplement 2, the median confidence value for the healthy population is typically very high (99.9%), presumably due to the excess of available healthy cell examples in the training dataset as compared to parasitized examples. Increasing the confidence threshold therefore rejects parasitized cells at a higher relative rate compared to healthy cells, leading to a decrease in the observed parasitemia as a function of threshold value. There is thus a trade-off between elimination of false positives and the recall performance of the classifier. One method of resolving this trade-off is to choose a maximum acceptable false positive rate as an independent parameter, and setting the minimum confidence threshold for which the desired rate is achieved. In Figure 2 supplement 3), we show the result of analyzing a negative control (healthy) sample with our classifier at 285 nm. As confidence threshold is increased, the FPR is reduced to zero. Recall will be also be modified as a result, but can be compensated for in the resulting measurement of sample composition. For stringent applications at low parasitemia levels, the FPR must be held as low as possible. While recall compensation can offset this known effect, rejection of cells by confidence thesholding correspondingly increases the total number of cell images required to ensure adequate sampling.

Population sub-sampling: Human annotation is used to generate labels for a small random sub-sample of the entire dataset. Optimal confidence threshold can then be assessed by minimizing experimental error as a function of threshold for the human-annotated sub-sample, and applying the same value to the remainder of the dataset.

Human collaboration: A human annotates a non-random, low-confidence subset of the data, which is most likely to be incorrect. The new human labels are incorporated into the dataset by overwriting the machine labels, and the classification results are updated using the merged data.

Focus-dependence of classification performance

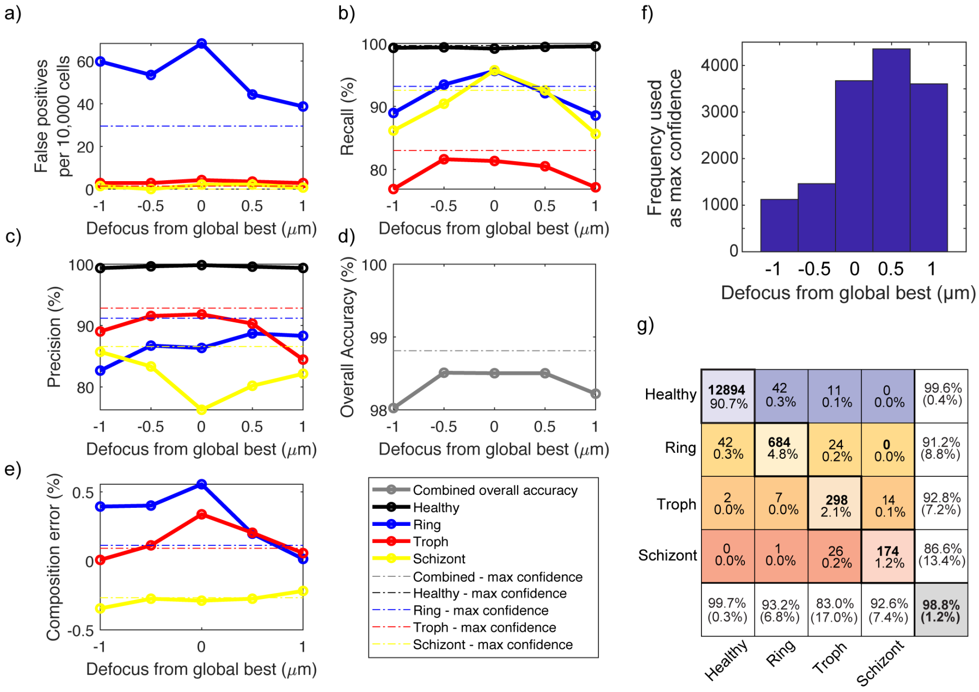
In order to ensure that the overall best focus images were captured, we acquired complete focus stacks for all datasets on the custom UV microscope. Since the optimal focus for any one parasite in general does not coincide with the global best focus of the image field of view, any one focal plane likely does not capture optimal views of each cell. Therefore, focal stacks provide an opportunity to improve classifier robustness by capturing a range of potential positions over which parasite features might be used for classification. We studied how various metrics of performance were affected by focus relative to global optimum, as well as take advantage of the additional information to boost performance beyond raw classifier output.

Because relative parasite location and orientation with respect to the global best focus plane was variable and not known *a priori*, we reasoned that robustness could be improved by consideration of all the focal slices available for each RBC instance. Among many possible approaches making use of focal information, our most effective strategy was based on selecting the focus slice with the highest classification confidence for each cell. Each focal slice of each RBC was independently processed by the classifier, but only the slice with the highest classification confidence was used to produce the final label for each given cell. In this manner, different RBCs in the same field of view can be classified independently at their optimal focal plane. We found this method to be more robust across nearly all performance metrics, resulting in a more favorable result than any one focal plane in isolation. When classifying four categories, overall accuracy improved from 98.1% to 98.8%, while the FPR for rings (healthy incorrectly identified as rings) decreased dramatically from 0.7% to 0.21%. This reduction in FPR is especially helpful for estimation of overall sample composition, as the balance against the FNR was nearly equalized. In Figure 3, we quantified several key performance metrics as a function of focal offset for all four categories.

Wavelength-dependence of classification

The majority of malaria researchers likely do not have the time or resources to build a dedicated deep UV microscope, nor is it practical or economical to fabricate custom quartz flow cells (required for deep-UV imaging) for routine parasite analysis. We therefore explored the possibility of label-free parasite classification at longer wavelengths both on our custom microscope as well as using a commercial light microscope. By producing spatially-registered images at multiple wavelengths, we performed direct comparisons of classifiers across different wavelengths and focal slices, using the same set of human-generated ground truth labels. Refinements to spatial registration were made in software by digital refocussing and affine transformations – see Figure 4 supplemental figure 2. Manual annotation was performed only on the highest resolution and global best-focus data, with human-generated labels transferred correspondingly to all other color channels and focus planes. Note that for the wavelength-dependence study, training and validation datasets consisted of distinct, but partially-overlapping sets of cell images, as compared with single wavelength classifier training and validation.

Observation of RBC appearance over a range of wavelengths, focal planes, and parasite stages permitted qualitative characterization of image features, and quantitative characterization of machine classifier performance. In Figures 4b and c, the same RBC is shown across an array of five focal positions and three wavelengths. It is apparent that the parasites' contrast changes as a function of both variables. This effect can be attributed to the parasites' real and imaginary refractive index components being lower in value as compared to the RBC (***Kim et al.*** (***2014***)), manifesting as



	Healthy	Ring	Troph	Schizont	
Healthy	**12894** 90.7%	42 0.3%	11 0.1%	0 0.0%	99.6% (0.4%)
Ring	42 0.3%	**684** 4.8%	24 0.2%	**0** 0.0%	91.2% (8.8%)
Troph	2 0.0%	7 0.0%	**298** 2.1%	14 0.1%	92.8% (7.2%)
Schizont	0 0.0%	1 0.0%	26 0.2%	**174** 1.2%	86.6% (13.4%)
	99.7% (0.3%)	93.2% (6.8%)	83.0% (17.0%)	92.6% (7.4%)	**98.8%** **(1.2%)**

Figure 3. Analysis of classification statistics vs. focal offset in deep UV images acquired at 285 nm. In all panels, solid lines refer to statistics from processing all cells at the same global focus offset, while dashed lines correspond to using the optimal focus slice on a cell-by-cell basis, determined via maximum confidence strategy. a) False positive rate vs. defocus from global optimum for all three parasite stages, b) Recall vs. defocus for all categories, c) Precision vs. defocus, d) Overall accuracy vs. defocus, and e) Sample composition estimation error vs. defocus. f) Histogram showing frequency of slice usage in the max confidence method. Slice number 3 is the global best focus slice, with slices evenly spaced in 0.5 μ increments. g) Resulting improved confusion matrix after applying the max confidence method. To increase the number of datapoints, all focus-dependent dataset statistics in this figure and its supplementary figures were derived from analyzing pooled classifier training and validation datasets as a function of focus slice.

Figure 3–Figure supplement 1. Slice-dependent four-category classification performance arranged as a confusion matrix. All panels spatially correspond to confusion matrix elements from Figure 2, and display relative changes in matrix values as a function of defocus. Diagonal matrix entries (true positives) corresponding to all three parasite life cycles stages tend to improve wtih global best focus, while many off-diagonal entries (false negatives / false positives) see reductions at the global best focus. However, exceptions include ring false positives (second row, first column) and troph/schizont cross-identification.

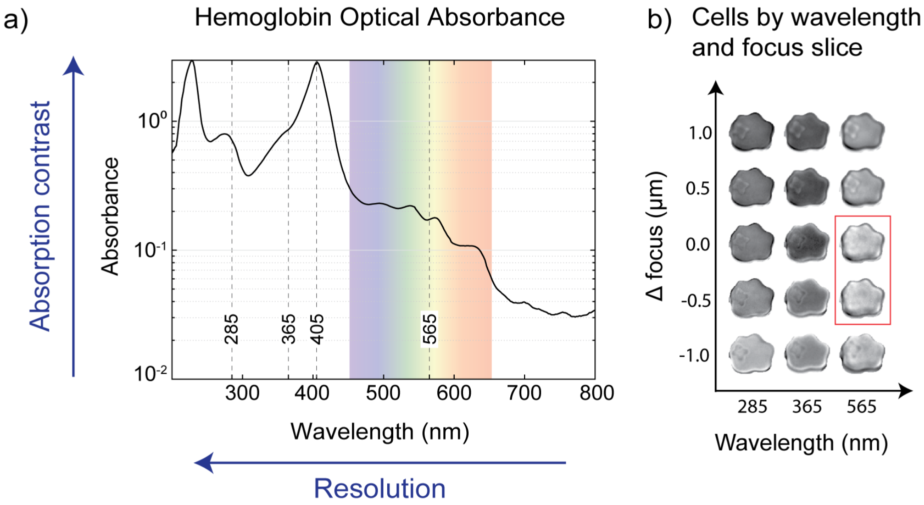


Figure 4. Wavelength and focus affect parasite contrast and classification efficiency. a) Optical absorption spectrum of hemoglobin is plotted, with the visible portion of the spectrum highlighted in color. Vertical dashed lines indicate the wavelengths used for imaging on the UV scope (285, 365, and 565 nm), and on the commercial microscope (365, 405, and conventional visible spectrum lamp). The hemoglobin (Hb) spectrum is suggestive that parasite identification should be influenced by absorption contrast (highest at peaks), while the shortest wavelengths enable the highest resolution (highest at short wavelengths). b) An array of images of the same ring-stage infected RBC generated by multi-dimensional image registration, enabling assessment of classifier performance as a function of wavelength and focal offset. The parasite becomes nearly impossible to observe in some of the focus slices at 565 nm (red box), illustrating the benefit of increased contrast and resolution.

Figure 4–Figure supplement 1. Color-focus arrays for many example RBCs at various stages of infection. For all panels, the color/focus layout is identical to Figure 4. Red boxes highlight cells that are not easily distinguished with visible light due to loss of contrast at certain focal planes. Examples of all four categorized life cycle stages are shown.

Figure 4–Figure supplement 2. Our multi-wavelength image analysis pipeline had the following steps: 1. Load data structures (UV scope) or raw images (commercial scope). 2. Apply re-focusing. 3. Register wavelength channels using affine transformation. 4. Export pre-processed images. 5. Do semantic segmentation on the master wavelength/slice using ResNet-50. 6. Apply instancing algorithm to all wavelengths/slices, exporting RBC instances to disk. 7. Classify all RBC instances for only the master wavelength and slice. 8. Apply the resulting labels to all the corresponding instances in the other slices/channels. 9. Export the lowest-confidence subset of the machine-classified images for human annotation. 10. Import the results of human annotation and propagate labels to all the other channels/slices. 11. Train new classifiers with the high confidence dataset.

Figure 4–Figure supplement 3. Affine transformations were performed for image registration between different wavelengths. Left: overlay of two raw images from the UV scope, acquired at 285 nm (green) and 565 nm (pink). Right: overlay of the same images after image registration was completed. The sample in the images consists of fixed *E.Coli* cells (rods) mixed with 1 μm beads (circles).

a combination of optical phase and absorption mechanisms, respectively. In particular, parasites can appear as either brighter than the RBC cytoplasm, darker than the cytoplasm, or exactly the same, depending on the focal plane. The phenomenon of through-focus contrast inversion is expected for purely phase objects (no imaginary component) imaged in brightfield (***Hernández Candia and Gutiérrez-Medina*** (***2014***)); an effect that is also explained by the Transport Of Intensity equation (***Petruccelli et al.*** (***2013***)). Early stage parasites without visible hemozoin accumulations in particular exhibit this effect. On the other hand, the imaginary component is directly related to the hemoglobin (Hb) molecular absorbance and contributes to a signal that does not invert sign through focus. The interplay of these two factors can explain the relative prominence of the parasite across the focus-wavelength array. Notably, in Figure 4b, a dendritic ring-form is seen to vanish entirely at certain focal planes when imaged with visible light due to the phase component exactly cancelling the absorptive component of the image on the host RBC backdrop. We remark that ultraviolet wavelengths, possessing both higher resolution and higher molecular absorbance by Hb, are more suitable for robust label-free imaging of *P. falciparum* because a) the parasite membrane is sharply resolved (higher resolution) and b) there are no focal planes for which the cytoplasm fully vanishes by contrast cancellation (absorption contrast exceeds the max phase contrast). We also note that although in many visible light images ring stage parasites were clearly identifiable by phase effect, that there were also cases where the definition of the parasite membrane was crucial for human parasite identification, for which the longer wavelengths were not always sufficient to resolve.

Near-UV modification of a commercial microscope improves classification

In order to increase the accessibility of our method, we conducted imaging and classification experiments on an inverted commercial microscope using the built-in trans-illumination lamp source as well as a custom light source configurable with two near-UV wavelengths (405 nm, and 365 nm – see Figure 5). Hypothesizing that Hb optical absorbance played a role in parasite detection, we selected these wavelengths according to the spectrum in Figure 4a. In the case of 405 nm, RBC absorption contrast was maximized with respect to both background and parasite cytoplasm, while resolution was hypothetically maximized at 365 nm (the shortest wavelength transmitted by most commercial microscopes), and finally the broadband lamp representing the unmodified system. We observed that with the high molecular absorption coefficient at 405 nm, early-stage parasite contrast was apparently dominated by the spatial occlusion of Hb (Figure 5, supplement 2), causing the parasite images to manifest as bright bodies on dark RBC backgrounds. This follows directly from that fact that early ring forms have not yet consumed a large Hb fraction, leaving the RBC nearly opaque. Later stages displayed internal enrichment of hemozoin crystal (***Ribaut et al.*** (***2008***)), which is also characterized by a spatial re-distribution of pigment throughout the cell. As a result, RBCs infected with later stages of parasite maturation can additionally be distinguished by depletion of pigment from the cytoplasm (Figure 5). Incorporation of these two near-UV wavelengths to the commercial microscope permitted exploration of optimal configurations of our method that are accessible to many groups around the world. We opted to acquire data with lower magnification than the 100× magnification on the custom microscope, using a 40×/1.30 oil immersion objective in order to acquire datasets with a larger numbers of cells. At 40×, we could acquire approximately 500-1,000 cells per field of view without without significant overlapping of cells, at the cost of reduced resolution. However, the larger cell count provided increased statistical resolving power that is important for analysis of low-parasitemia samples.

We assessed the performance of the modified commercial microscope by the same metrics as with deep UV, using confusion matrices to compare classifier predictions with human annotations (results from a classifier optimized on a single wavelength dataset at 405 nm is shown in Figure 5, and a direct comparison across wavelengths is shown in supplement 1). On the basis of overall classification accuracy, the results were inferior when compared to the higher magnification and shorter wavelength used in the UV microscope. However, of the three wavelengths, 405 nm dis-

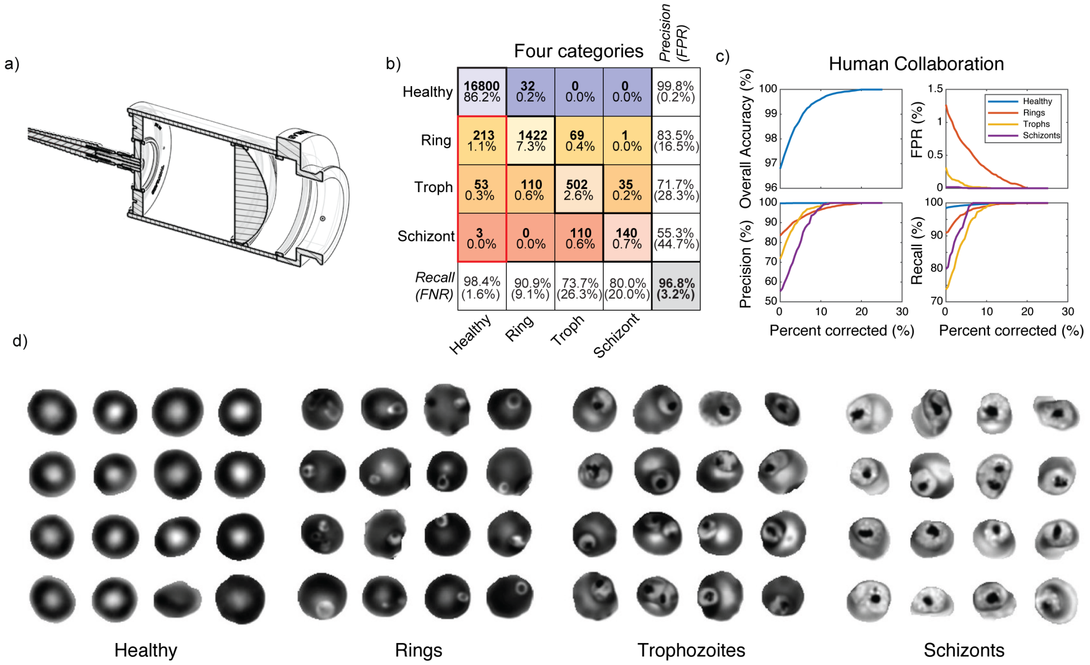


Figure 5. Equipping a commercial widefield microscope with 405 nm excitation generates high-contrast label-free images of infected RBCs. a) Cross-sectional view of our fiber-coupled, collimated light source used for near-UV excitation through the microscope transillumination port. b) Confusion matrix from the single-wavelength validation dataset of the classifier trained at 405 nm on the commercial microscope. c) Results from human collaboration, ie. human correction of labels for only the cells with the lowest classifier confidence. All four graphs depict performance metrics as a function of the total percentage of cells corrected. Upper left: Overall accuracy, Upper right: False-Positive Rate, Lower left: Precision, and Lower right: recall. The human plus machine classifier achieves perfect performance after correction of 20% of the dataset. d) Example cell images of all four categories, acquired at 405 nm. From left to right: Healthy, Rings, Trophozoites, and Schizonts.

Figure 5–Figure supplement 1. Confusion matrices for classifiers trained at each wavelength on our commercial microscope. Left: 365 nm, Middle: 405 nm, Right: Visible light lamp. All classifiers were trained on the identical set of RBC instances, using images acquired each at their respective wavelength (multiwavelength dataset, see table 1).

Figure 5–Figure supplement 2. Example RBC images acquired on our commercial microscope for all three wavelengths (left columns: 365 nm, middle columns: 405 nm, right columns: standard visible lamp). Four example cells are displayed for each category, where each row within a category are images of the same physical cell. Ring stage parasites are seen to be highly mobile, as evidenced by their differing positions and even morphology between acquisitions at different wavelengths, which were separated in time on the order of a few minutes.

Figure 5–Figure supplement 3. Parameterization of the false-positive rate from a healthy control sample imaged on our commercial microscope at 405 nm.The upper plot displays the sample composition error for all four stages as a function of confidence threshold. Data processed were from a healthy donor sample, uninfected by *Plasmodium*. In particular, the FPR for rings at 50% confidence threshold is substantially lower than the results reported in Figure 5. We explain this by noting this particular healthy control sample exhibited very few echinocytes, to which we attribute the low basal FPRs as compared to Figure 5.

Custom UV Microscope: 100x/0.85 glycerol immersion

Focus condition	Overall Accuracy (%)			FPR-Rings (%)			Recall - Rings (%)		
	285 nm	365 nm	565 nm	285 nm	365 nm	565 nm	285 nm	365 nm	565 nm
Max Confidence	98.7	98	97.7	0.5	0.6	0.9	93.8	87.4	88.3
1.0 µm	98	97	96.9	0.7	1.2	1.4	88.7	83	80.5
0.5 µm	98	97.3	97	1	1.3	1.5	90.7	83.2	85
0 µm	97.9	97.3	96.8	1.3	1.2	1.7	92.9	85	85
-0.5 µm	98.2	97.6	96.9	1.1	1.1	1.6	93.1	88.3	88
-1.0 µm	98	97.6	97	1	1.1	1.5	89.4	86.1	86

Commercial brightfield microscope: 40x/1.3 oil immersion

	Overall Accuracy (%)			FPR-Rings (%)			Recall - Rings (%)		
	365 nm	405 nm	Lamp	365 nm	405 nm	Lamp	365 nm	405 nm	Lamp
0 µm	93.3	96.1	94.9	2.5	1.2	0.9	87	96.6	84

Figure 6. Top: Wavelength- and slice-dependent classification statistics on the custom UV microscope. Overall accuracy (left), FPR for rings (center), and recall for rings (right) were selected for detailed breakdown as a function of wavelength (columns), and focal offset (rows). The top row in each table shows the result of processing the data with our maximum confidence strategy for focal plane selection on a cell-by-cell basis. Bottom: The same statistics were plotted for a single focal plane on the commercial light microscope at 40× magnification. For each grouping by microscope adn statistic, cells are color-coded on a scale from blue (lowest performance) to yellow (highest performance). All statistics were derived from multi-wavelength datasets, which consisted of distinct but partially-overlapping sets of data with single-wavelength datasets (see Table 1). In both cases, each wavelength-specific classifier was trained and validated uniquely on images at its particular wavelength. To ensure a fair comparison between the wavelengths, all three wavelengths for each microscope shared the exact same random partitions between training and validation, ie. the same distinct sets of RBCs and corresponding ground truth labels were used across wavelengths.

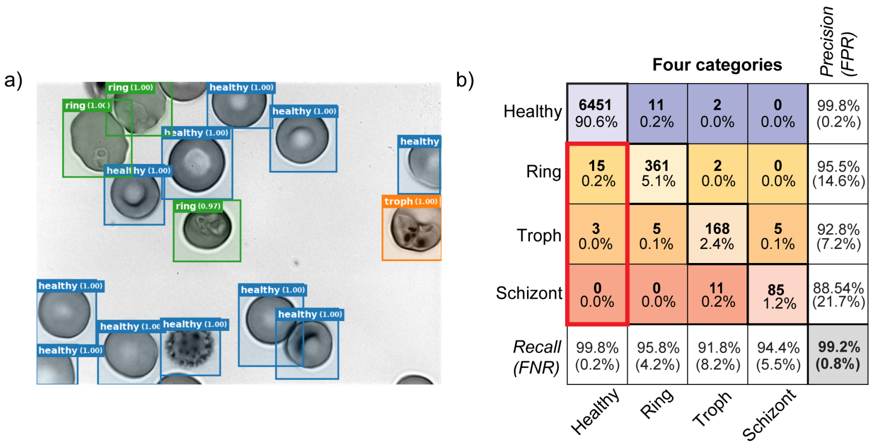
played the most robust contrast for ring stage parasites, which is corroborated by an exceptionally high recall of 96.6%. Interestingly, images acquired at 365 nm, while expected to exhibit the highest nominal resolution on the commercial microscope, did not perform better than visible light. This is potentially explained by optical aberrations in the microscope objective in the near-UV regime, given that this trend was not observed using our custom UV microscope.

Given the trade-off between raw classification performance and throughput, the most important output from a machine classifier in this context is an accurate estimate of the sample composition, which depends not only on the per-cell accuracy, but also on the balance between FPR and FNR, and the statistical resolving power. We later assess the sum contribution of these factors by titration of parasitemia, comparing also to the century-old gold standard method of manual inspection of Giemsa-stained smears.

Direct detection of parasitized RBCs via Faster R-CNN

The above results were used to train and validate a python-based Faster R-CNN model (***Ren et al.*** (***2015***)) to recognize both healthy and parasitized RBCs directly on raw images, as opposed to performing a segmentation step prior to classification. The method uses a Region Proposal Network (RPN) to first identify image regions likely to contain cells, which are then processed by a unified network to generate object detection confidence scores. Training such a network to recognize malaria presented logistical challenges but had some advantages. First, the direct detection on raw images is simpler to implement, with fewer intermediate steps. Second, the python-based framework is more easily open-sourced and disseminated. Finally, this method has a more direct path for implementation on a low-cost, embedded system.

The main logistical challenge to implementing this method was assembling enough human-annotated instances of all the parasite life stages for training. In contrast with pre-segmented RBCs which can be hand-annotated very quickly (see Methods section), generation of labels for R-CNN requires carefully drawing and labelling bounding boxes around a large number of cells of



	Healthy	Ring	Troph	Schizont	*Precision (FPR)*
Healthy	**6451** 90.6%	**11** 0.2%	**2** 0.0%	**0** 0.0%	99.8% (0.2%)
Ring	**15** 0.2%	**361** 5.1%	**2** 0.0%	**0** 0.0%	95.5% (14.6%)
Troph	**3** 0.0%	**5** 0.1%	**168** 2.4%	**5** 0.1%	92.8% (7.2%)
Schizont	**0** 0.0%	**0** 0.0%	**11** 0.2%	**85** 1.2%	88.54% (21.7%)
Recall (FNR)	99.8% (0.2%)	95.8% (4.2%)	91.8% (8.2%)	94.4% (5.5%)	**99.2%** **(0.8%)**

Figure 7. a) The Faster-RCNN method detects objects directly on images without using an intermediate semantic segmentation step (***Ren et al.*** (***2015***)), generating bounding boxes and confidence scores for each detected object. The image shown was acquired at 285 nm, and includes examples of healthy cells, ring stage parasites, trophozoites, as well as an echinocyte (spiky RBC, lower middle-left). b) A four-category confusion matrix is shown for the Faster R-CNN method, using the same format as Figure 2. Faster R-CNN exhibited a reduced FPR and higher overall accuracy than the two-step method.

Figure 7–Figure supplement 1. Total loss function (training + validation) during Faster-RCNN training.

each category. Further, the bounding boxes might also contain edges of other nearby cells, adding deleterious background features to the training data. Additionally, since the raw images themselves contain largely healthy cells and very low fractions of trophozoites and schizonts, human annotation performed directly on the raw images would be excessively time-consuming in order to obtain enough annotated examples of rare classes. To resolve this issue, we used our previous image segmentation and classification results to generate class-balanced synthetic raw images for training. These balanced synthetic training images for the Faster R-CNN consisted of a random training partition of pre-annotated instances of the various RBC classes randomly distributed over a field of view. Subsequently, the trained network was evaluated on the same series of raw images from which the original training data were derived, but with the analysis statistics drawn only from cells within the validation partition.

The Faster R-CNN method performed well on the 285 nm UV images – better than our original two-step method used for training data generation. Overall accuracy was 99.2%, better than even using slice consensus – see Figure 3. Notably, for rings (the most common stage in peripheral blood) the FPR was only 0.2%, with approximately 95% precision and recall. For trophozoites and schizonts the FPR was negligible, and precision and recall near 90%.

Quantitative extrinsic validation

Thus far we have performed intrinsic statistical analyses using human-annotated images as ground truth, without comparison either to existing methods or with respect to external control parameters. In order to validate our method beyond self-consistency, we conducted a rigorous extrinsic test of our method, by comparing it to hand-counted Giemsa-stained blood smears over parasitemia levels from 18.6% down to less than 0.1% . First, parasites were cultured to high parasitemia, then serially-diluted into fresh, healthy blood at 2% hematocrit, forming a ten-point series. For each point on the curve, the sample was removed from the incubator immediately prior to imaging in order to make a blood smear and load flow cells. Samples were imaged on both microscopes in order to facilitate simultaneous image acquisition, minimizing time delay between the methods. Giemsa-stained blood smears were also prepared at each concentration at the time of imaging, and subsequently counted manually by three experienced technicians. Manual counting was intentionally limited to approximately 300 cells per titration point in order to represent a realistic laboratory counting practice. However, in order to generate a more precise reference point, manual counting was performed on over 2,000 cells at only the highest concentration, by a combination of all three annotators. All nominal parasitemia values were computed based on dilution factors from this reference point.

Results of the comparison are shown in Figure 8. Fundamentally, Poisson counting statistics impose a minimum uncertainty with a variance equal to the mean number of counted parasites. Correspondingly, we expect variance in manual counting of 300 cells to diverge at low parasitemia. To convey this quantitatively, we overlaid grayscale error bands on each plot, each corresponding to one standard deviation of the underlying Poisson distribution for several values of total counted cells, demonstrating how resolving power improves as a larger number of cells is counted. All bands are relatively narrow at high parasitemia values, and begin expanding asymptotically as they approach an expectation value of one parasite per sample – below which the technician (or algorithm) is not likely to encounter any parasites in the sample.

Our results in Figure 8a demonstrate that with deep-UV microscopy, our raw classifier exhibits a linear response from the high point of 18.6% parasitemia down to approximately 0.5%, where it is further limited by the detection of false positives. By capturing many thousands of cells per titration point (at 100× with z-stacks), the dataset exhibits superior counting statistics than manual counting of 300 cells – albeit with reduced recall– as defined by comparison with manual inspection of Giemsa-stained smears. The FPR can be modestly reduced by applying the slice max confidence technique (blue curve in Figure 8a), and substantially improved by human collaboration (light blue in Figure 8a). In the latter case, we exported a pool of all 1,595 putative infected cell images as well as the 5,000 lowest confidence putative healthy cells for manual inspection. Image export and label correction was performed on the pool, without knowledge of the underlying titration point for any given RBC. In total, a net 124 cells labeled as parasitic by the classifier were re-labeled as healthy during collaboration, out of a total of 75,343 cells in the dataset. The procedure was performed on a pool of ten total imaging experiments and consumed 90 minutes of hands-on time. It can therefore stand to reason that a single experiment could be human-corrected in approximately ten minutes. Aside from consistently lower parasitemia estimates, our human-collaboration results exhibit superior ratiometric error as compared to both manual scoring of Giemsa slides and raw classifier output. Faster R-CNN analysis of the same data results in very similar performance as the raw classifier, but with marginally higher recall and FPR.

Although human collaboration increases classifier performance for the 285 nm data at 100×, we note that to the extent that the error statistics are stationary, both recall and FPR can be compensated for by introducing multiplicative and additive correction factors to the sample composition estimates. In Figure 8b, corrections were made to the raw data such that each compensated point $C_{comp} = (C_{raw} - FPR)/recall$, and plotted for a range of confidence threshold values. Compensation values were co-optimized by weighted least squares fitting to the ten-point dilution series, whose fit values are shown as a function of confidence threshold in Figure 8c. These data suggest that the classifiers' inherent false-positive rate as measured by best fit to an extrinsic control variable (parasite dilution factor) is significantly lower than the values reported in Figure 2. However, it is important to note that all reported statistics depend to some extent on characteristics of the sample. Specifically, deviations from the discocyte cell morphology tend to degrade classification performance, as evidenced by a high fraction of low-confidence and misclassified cells consisting of echinocytes and spatially-overlapping cells. Notably, the fresh RBCs used in the titration displayed a particularly low number of echinocytes. We hypothesize this to underlie the lower observed FPR in the titration as compared to intrinsic validation. To further

Titration data collected on the commercial microscope using 40× magnification at 405 nm facilitated acquisition of a far greater number of cells per condition, by virtue of a larger field of view and single focal plane acquisition. We captured images of 20,000-130,000 cells per condition – an order of magnitude more data than with deep UV, leading to a further improvement in counting statistics. Despite inferior intrinsic validation performance, we observed that imaging of a large number of

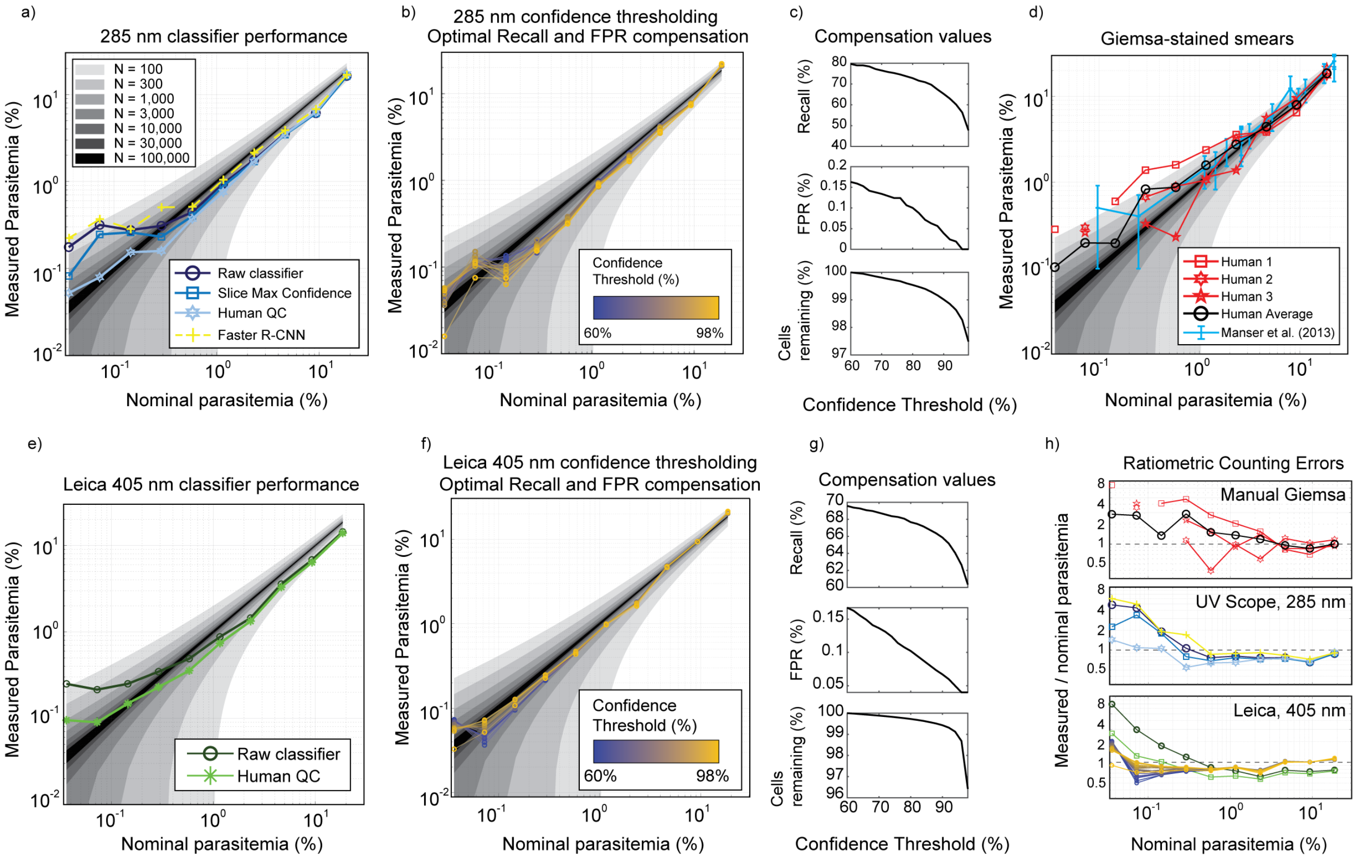


Figure 8. Parasitemia titration: Logscale plots a-b,d-f contain shaded regions corresponding to nominal mean parasitemia plus or minus one standard deviation of the underlying Poisson distribution when counting N total cells, where N is varied from 100 up to 100,000 in roughly 3× increments. The nominal parasitemia in panels a-b,d-f is derived from a manual count of the titration high point using 2,214 total cells, with all remaining nominal diluted values computed by theoretical dilution factor. a) Sample parasitemia extracted from the raw 285 nm classifier results is plotted in dark blue. Post-processing using the maximum slice confidence technique is shown in blue, resulting in a reduction of the false-positive rate evidenced by a further drop in the lowest points of the curve. The light blue data show the result of human collaboration on the putative infected and lowest-confidence healthy cells. The Faster R-CNN method was used to detect cells directly on the raw images (yellow). b) The data from a) were re-plotted after compensation for recall and FPR, for an array of confidence threshold values applied to the data. c) Recall and FPR compensation values are plotted as a function of confidence threshold, which both decrease with threshold value, as healthy cells' median confidence is on the order of 99.9%. d) Manual counting of Giemsa-stained smears by three experienced technicians are plotted with solid red curves (human 1: squares, human 2: hexagrams, human 3: pentagrams), and their arithmetic mean plotted in solid black (circles). Missing markers correspond to points where no parasites were located in the 300 cell sample. Blue solid lines with error bars are an overlay of the data from Table 2 in ***Manser et al.*** (***2013***), showing the result of extensive validation across hundreds of trained microscopists. Error bars indicate standard deviation across the participants' parasitemia estimations. The "Reference mean parasitemia" data from the cited table were used as x-values during overlay. e) The titration results were plotted for the classifier trained on commercial microscope at 405 nm. Dark green circles: raw classifier results; light green '*': results from human collaboration, as described in a). f) Commercial microscope classifier results after compensation, for various confidence threshold values. Large cell count numbers permit low relative counting variance, reducing deviation from the underlying distribution. g) Compensation values for recall (top), FPR (middle), and remaining cells after thresholding (bottom) from the data in f). h) Ratiometric error of the data in panels a,b,d,e, and f) vs. nominal parasitemia. Top: Manual counting of Giemsa-stained smears (d), middle: UV Scope classifier at 285 nm (a,b), and bottom: Commercial microscope images at 405 nm (e,f). All data in h) are referenced to the legends in their corresponding raw data panels. All titration data was purely 'test data', acquired after classifier training and validation had been completed.

Figure 8–Figure supplement 1. Confusion matrices for four, three, and two-category classifiers at 285 nm tested on a random sub-partition of pooled titration data, which as a whole contained a low echinocyte fraction. As compared with Figure 2, the FPR for rings is reduced presumably due to fewer cells containing anomalous morphology, and is in closer agreement to the result extracted by least squares fit to the titration data in Figure 8. The random sub-partition was configured to contain overall parasite prevalence similar to Figure 2.

cells led to superior results. The raw classifier curve in Figure 8e is more amenable to compensation, with the majority of datapoints lying within one standard deviation for the N=30,000 band and many points lying within the N=100,000 band down to 0.07% parasitemia. We also observe from Figure 8f that although confidence thresholding aids in reducing FPR and therefore is beneficial at low parasitemia, there is only minor additional benefit beyond data compensation. Figures 8c,g exhibit similar trends, showing that FPR reduction occurs with increasing threshold, at the cost of lost recall.

Discussion

This work demonstrates that visible and ultraviolet brightfield microscopy images of live *P. falciparum*-infected red blood cells can be classified using deep-learning. We validated this method intrinsically by statistical analysis of confusion matrices, and extrinsically by titration of the parasitemia alongside manual scoring of Giemsa-stained smears. Our data suggest that the method exceeds the performance requirements for routine laboratory analysis of parasite cultures from below 1% up to at least 18% parasitemia (***Schuster*** (***2002***)), and when compensated for known recall and FPR rates, likely exhibits a limit of detection lower than 0.1%. By virtue of improved sampling power, automated imaging and classification offers superior performance over manual analysis for low parasitemia samples, while eliminating the time-consuming and variable steps of fixation, staining, and manual inspection.

We employed a custom-built deep UV microscope equipped with wavelengths as short as 285 nm, as well as a modified commercial light microscope operating from near-UV to visible light, in order to explore performance across a range of imaging conditions. Our results suggest that image classification performance depends on both image resolution and contrast: on our custom microscope, the best intrinsic classification was observed at 285 nm / 100× where resolution was maximal. Of the wavelengths accessible to commercial microscopes, 405 nm performed best. While resolution monotonically improves with lower wavelength, contrast is the result of more nuanced light-matter interactions involving both the real and imaginary refractive indices of the parasite and RBC cytoplasm (***Petruccelli et al.*** (***2013***), ***Hernández Candia and Gutiérrez-Medina*** (***2014***)). The phase effect (real component) is sensitive to and inverts in sign through image focus, but the absorption effect (imaginary component) results in positive parasite contrast independently of focus. Over the wavelengths we tested, Hb absorbance varied approximately a factor of 15 (see Figure 4a), and in the case of 405 nm appears to be large enough such that ring parasites are visible in stark positive contrast with respect to the RBC cytoplasm, as compared with other wavelengths where it is possible in certain focal planes for ring stage parasites to entirely vanish (Figure 4b). We hypothesize that this factor explains the high recall rate for rings at this particular wavelength. Independent of wavelength, the use of information from multiple focal planes consistently increases the likelihood of identifying parasites (or ruling them out).

Our results using a modified commercial microscope extend the impact of this work beyond the limited set of researchers with the time and resources to build a custom deep UV microscope. The simple addition of a near-UV light source enables a boost in performance, effectively increasing the robustness of information used by the machine classifier. We have made all of our software open-sourced, including resources for rapidly generating large annotated datasets for re-training on other microscopes. We anticipate that adopters of this technique can quickly replicate this work on existing microscopes by first using our published, pre-trained classifiers to sort RBCs imperfectly, then provide corrections to the labels by careful inspection of a subset of low-confidence cells, and then using the resulting dataset to train an improved classifier. In our experience, this procedure requires less hands-on time than generating training labels from scratch, especially because as the classifier improves, the fraction of cells requiring re-labelling diminishes and median machine confidence increases. Since our software keeps a record of all the existing annotated files across iterations, repeated annotations of the same cells does not need to be performed across iterations.

Titration of parasitemia over a wide range suggests automated counting exhibits, by most metrics, better performance than manual counting of Giemsa stains, which is the gold standard in the field. With the compensated commercial microscope datapoints in Figure 8f lying close to the expected single standard deviation band for 100,000 counted cells, it is possible that the limit of detection had not yet been reached even at the lowest point on the curve (0.036% nominal parasitemia). Similarly, ratiometric errors shown in Figure 8h indicate little divergence from unity for human collaboration (middle, 285 nm) as well as compensated data (bottom, 405 nm). On the other hand, manual scoring of blood smears exhibits larger variation over most of the experimental range due to the effects of limited sample size, noting that averaging results from all three annotators substantially reduced error. Finally, it should be noted that although all nominal parasitemia data was generated from a "deep" manual count, the three annotators nonetheless separately assessed substantially different results of 17.4% (N=1,332 cells), 19.1% (N=460 cells), and 21.6% (N=422 cells), which presumably was due both to Poisson sampling error and variation in annotation accuracy between humans. From this we conclude that there is substantial uncertainty in the standard itself to which we reference, highlighting the very need for more consistent and robust counting methods. In fact, according to World Health Organization guidelines (***WHO*** (***2016***)), malaria microscopists are graded on a competence scale of 1-4, where the highest competence level requires a parasite counting accuracy within 25% of the correct answer only 50% of the time – our three annotators' scores lie well within this window from their aggregate mean.

A future direction enabled by this work will be the development of a low-cost module to perform label-free clinical diagnostics in the context of low-resource settings. This direction holds promise, but would likely encounter several challenges. First, clinical parasitemia levels can vary widely (***Adu-Gyasi et al.*** (***2012***)), and must be distinguished from healthy with a high degree of confidence. As a result, clinical diagnostic procedures typically include both thick and thin blood smears, with a recommended minimum 2,000 counted cells per thin smear (***WHO*** (***2016***)). While counting a large number of cells is challenging for human observers, automated systems have clear advantages in this respect. Second, individuals may be infected with more than one *Plasmodium* species (we did not perform speciation in this work), complicating the classification task. Third, screening whole blood will present additional challenges such as the presence of lymphocytes, platelets, and variable patient RBC count. It is also worth noting that in many of our datasets, we observe a relatively large fraction of echinocytes, or RBCs exhibiting spiked morphologies (***Bessis*** (***1974***)). Previous work has shown that echinocytes and other exotic RBC morphologies can be influenced by the confinement between glass surfaces, and can also depend sensitively on the details of sample preparation (***Eriksson*** (***1990***)). Our flow cells fall within the regimes reported in the literature as enhancing this effect, which we observed with regularity in our experiments. Fortunately, our classifiers could in general distinguish the echinocytes' sharp spicular morphologies from parasites, albeit at higher error rates. Depending on collection methods and imaging consumable design, clinical diagnostic imaging in whole blood may or may not suffer from this effect.

Implementation of label-free imaging and classification of live *P. falciparum* will enable major reductions in technician time, training, reagent cost, and observer variability with respect to the analysis of parasitized blood samples around the world, in both laboratories and ultimately in the field. Our results strongly suggest that advances in label-free imaging should enable rapid, automated, and highly accurate clinical diagnosis of fresh, whole blood – replacing error-prone and labor-intensive manual slide preparation and counting, a process that has changed little in over a century.

Methods and Materials

Hardware

Deep UV Microscope

Deep UV microscopy was performed on a custom-built microscope (Figure 1) using a Zeiss 100×/0.85 Ultrafluar quartz objective lens to form images onto a UV-sensitive camera (PCO.Ultraviolet). A multi-wavelength condenser was built for transmitted light excitation using commercially-available UV LEDs to illuminate the sample with 285 nm (Thorlabs M285L5), 365 nm (Thorlabs M365FP1), or 565 nm (Thorlabs M565L3) light. The condenser used one dichroic mirror to combine the two UV wavelengths, and a second dichroic mirror to merge visible light from the third LED, or alternatively, any standard microscopy lamp. Each UV LED was collimated with an off-axis parabolic UV-enhanced aluminum mirror (Thorlabs MPD129-F01) before being combined. An adjustable iris was used to control the numerical aperture of the illumination.

Commercial microscope

All commercial microscope experiments were performed on a Leica DMi8 equipped with a Plan Apochromatic 40×/1.30 oil immersion objective (Leica #11506358), motorized XY and focus stages, and custom automated image acquisition software. To facilitate near-UV wavelengths, we constructed a fiber-coupled equivalent of a commercial pre-built microscopy light source (Thorlabs M405L3-C2) and attached it to the TL-port of the microscope. Our custom source used a multimode fiber which could be moved between multiple LED sources. The standard TL lamp was used for visible light experiments as a comparison.

Software

Instrument control

The deep UV microscope was controlled using custom classes, functions, and scripts written in MATLAB. All software is freely available here: https://github.com/czbiohub/UVScope-control. Overall architecture is detailed in Figure 1, supplement 4. The microscope operating system is implemented by a custom “UVScope” data class, orchestrating all the high level processes such as state of the system and interaction with each of the hardware device classes. Other data classes act as hardware interface drivers, image data processing, metadata generation, and storage.

Image Processing

Two different image processing pipelines were used in this work. The main results (Figures 1 - 6) all used a two-step method employing semantic image segmentation of all the RBCs, which were subsequently classified by a network trained to distinguish all four categories of cell. We also employed a single-step method (Faster R-CNN).

Two-step method (segment and classify)

All two-step method training and classification was performed using the Matlab Deep Learning Toolbox and custom data classes to store and organize the image data, and deep network training was performed on a Windows 10 desktop PC with 128 GB of RAM, and an Nvidia GeForce GTX1060 6144MB GPU card. Image datasets containing multiple fields of view, wavelengths and focus slices were imported, digitally re-focused, registered, segmented, annotated, trained, and validated (see Figure 1 — supplementary figure 2). Since multiple distinct datasets were required in order to increase the size of the training dataset, they were merged prior to network training. After dataset merging, the various other dimensions of the data (fields of view, wavelengths, focus slices, dataset IDs) could then either be merged, kept separate, or both, for purposes of classifier training and validation.

Manual human annotation was performed to provide ground truth training data. In order to generate enough annotations, cropped RBC image instances were written to disk and sorted into labeled directories. Using this method, cells could be manually labeled at a rate up to several thousand cells per hour, which was significantly faster than other methods we explored. A human-in-loop strategy was implemented to further increase the quantity of training data, whereby a network was initially trained on a small dataset (5,000 cells), then used to classify a larger dataset. The 5,000 cells with the lowest confidence scores were exported for human annotation and re-imported for iterative training. In all cases, data augmentation was used during network training and included rotation, scaling, translation, and reflection. All software used for the two-step method can be found here: https://github.com/czbiohub/Label-Free-Malaria.

Single-step method (Faster R-CNN)

For Faster R-CNN analysis we modified the Luminoth package made available by Tryolabs (***Tryolabs*** (***2018***)). For training, comma-separated-value (.csv) files (produced by our two-step method) containing the locations and labels for each cell were imported. The initial dataset was randomly split into separate training (90% of data) and validation (10% of data) datasets. The datasets were then converted to tensorflow records for training. Hyperparameter tuning was used to optimize validation loss function after iterative network training. The hyperparameters tuned include: the loss function for RPN and R-CNN layers in Faster R-CNN, weights of the RPN and R-CNN loss functions, data augmentation techniques (image scaling, reflection, and rotation), and learning rate. For training, simulated raw images were constructed to enrich the density of low-frequency parasite stages, and were constructed by assigning random locations and orientations for each pre-segmented RBC from the two-step pipeline. All of our codebase is publicly available at: https://github.com/czbiohub/luminotl uv-imaging/.

Processing commercial microscope images

Images from the commercial microscope did not require re-focussing. However, alignment between color channels was performed using 2D cross-correlation (rigid body translation). Subsequently, all steps were identical to UV scope image processing.

Sample Preparation

Cell culture and manual blood smear counting

Plasmodium falciparum strains W2 and 3D7 were cultured in 50 mL flasks containing RPMI (Thermo Fisher #31800089, supplemented with 0.5% Albumax II (GIBCO 290 Life Technologies), 2 g/L sodium bicarbonate, 0.1 mM hypoxanthine, 25 mM HEPES (pH 7.4), and 50 μg/L gentamicin). Cultures were maintained at 2% hematocrit in a temperature and gas-controlled environment set to 37°C, 5% oxygen, and 5% CO2. Cultures are periodically split in order to maintain parasitemia levels of 1-5% in order to prevent overgrowth, with daily media changes. They were checked daily by briefly centrifuging 500 microliters of the culture in an Eppendorf tube, aspirating the supernatant, and adding 10 microliters of the infected red blood cells to a clear glass slide. To visualize P. falciparum, standard Giemsa staining technique was used and the slide was subsequently viewed using light microscopy with a Zeiss light microscope at 100× magnification.

Preparation for microscopy

Prior to imaging, cultures are diluted two-fold into culture media in order to reduce the likelihood of RBCs overlapping during imaging. 60 μL of diluted cell culture is loaded directly into a custom quartz flow cell (see Figure 1, supplement 1). Flow cell ports are subsequently sealed with clear nail polish, and all outer surfaces cleaned with isopropanol. Quartz flow cells were fastened into a custom 3D-printed adapter with exterior dimensions of a standard SBS well plate. The adapter is directly mounted on the microscope stage.

Enrichment of schizont stage

Classifier training requires numerous examples of each category. Since schizonts were the shortest-lived and least frequently-observed life cycle stage, we synchronized the life cycle stages four days prior to imaging by adding a 5% sorbitol solution to ring-stage parasites for 10 minutes, ensuring that all trophozoite or schizont-infected RBCs were lysed. Synchronization was confirmed in the following days by regular smearing of the parasite culture, and imaging took place when schizonts were abundant.

Titration of parasitized red blood cells

Malaria was grown to high parasitemia by splitting a ring-dominant culture to 5% parasitemia into freshly-drawn, healthy RBCs two days prior to the experiment. Frequent media changes were made subsequently, but no further splits were done prior to the experiment. A ten-point serial dilution into healthy RBCs was performed in 2× increments. The entire series was prepared into flasks in the morning, and stored in the incubator at 37°C prior to imaging. Each sample was taken out of the incubator and loaded into a flow cell on demand, starting with the lowest concentration. Dilutions were made using RPMI complete medium at a hematocrit of 2%, with the healthy RBCs stored in the incubator at 2% hematocrit after the blood draw. As with all other conditions, the sample was further diluted twofold in PBS in order to prevent the overlap of RBCs while imaging. Each concentration point was imaged immediately after loading the flow cell. Giemsa-stained blood smears were performed at each concentration and counted manually as described above.

Image annotation and ground truth generation

Ground truth labels were generated by manually sorting exported RBC images into labelled directories, by using a standard Windows 10 explorer window to display the images as large icons. Incorrectly labelled cells were moved by a drag and drop to their new directory. In the case of UV microscope, images from the 285 nm channel were exported for sorting and labels applied to both other wavelengths, whereas in the case of the commercial microscope, 405 nm was used for sorting. In the latter case the image contrast and dynamic range were too high for viewing on the monitor as a result of the very high Hb absorption coefficient. To circumvent this issue we computed the logarithm of the image pixels using FIJI (Fiji Is Just ImageJ) after the original export operation. Once the images were sorted, their filenames were used to apply the new labels to the original (unprocessed) images.

Specific human annotation criteria

Manual sorting was used to generate ground truth labels to the following standards (for example labelled parasite images see Figure 2, — figure supplement 5). The following paragraphs outline class- and microscope-specific annotation criteria. In general, cells without any of the following features were labelled as 'healthy', regardless of other morphological abnormalities, debris, or otherwise indistinguishable features.

Deep UV — ring stage

Ring stage parasites were primarily identified by a membrane boundary contour defining the parasite morphology, which varied greatly but could be identified either by dendritic projections, canonical annular morphologies, or more rarely, circular blob forms. Importantly, the parasite contrast with respect to the background Hb could be either positive or negative depending on the relative position of the parasite with respect to the focal plane. This effect can be seen in the z-stacks shown in Figure 4.

Deep UV — trophozoite stage

Trophozoite stages were distinguished from rings by their moderate size, increased shape solidity (lack of projections or annular voids), and varying stages of hemozoin crystal accumulation. Trophozoites were distinguished from schizonts by their lack of nascent merozoites and their less centralized hemozoin accumulation. RBCs infected by mature trophozoites exhibited brighter cytoplasmic pixel backround values due to the parasite's Hb sequestration effectively lowering the cytoplasmic absorption.

Deep UV — schizont stage

Schizont stage parasites were identified by their dominant size (usually greater than half the cell area), prominent and centralized accumulation of hemozoin, and their unique characteristic texture due to merozoite formation. However, since early stage schizonts shared many features with mature trophozoites, the distinction was sometimes subtle or non-existent.

Near UV — ring stage

At 405 nm, the Hb absorption coefficient is nearly threefold higher than at 285 nm (see Figure 4 for a full spectrum) and the resolution was lower. The effect of high absorption dominated the image contrast, enabling ring-stage parasites to be identified by the parasites' occlusion of Hb. We could identify ring stage parasites in 405 nm images based entirely on this effect, combined with the characteristic range of shapes they are known to assume, based partially on our experience with the higher resolution deep UV images acquired at 285 nm.

Near UV — trophozoite stage

Since the Hb absorption was so high, we could easily identify puncta of accumulated hemozoin within the parasite, which, along with increased size and solidity, was the primary means of distinguishing trophozoites from rings. Additionally, the cytoplasmic depletion of Hb becomes more pronounced with mature trophozoites.

Near UV — Schizont stage

At near UV our resolution was not sufficient to resolve merozoite texture to the degree that was possible with deep UV, which hindered our ability to distinguish schizonts from mature trophozoites. Overall size, centralized hemozoin accumulation, and RBC Hb depletion were the main image features used to score schizonts at near UV.

Hemoglobin absorption measurement

Lyophilized hemoglobin was purchased from Sigma-Aldrich (#H7379-1G) and dissolved into Dulbecco's Phosphate Buffered Saline (DPBS) at 10 mg/mL. Absorption measurements were performed on a SpectraMax M3 plate and cuvette reader using a 1 cm path length. UV-transparent cuvettes were used (Thermo #13-878-123) to provide adequate transparency over the range of the measurement.

Acknowledgments

The authors would like to acknowledge Shalin Mehta, Loic Royer, Bin Yang, and Kevin Yamauchi for useful discussions related to construction of the microscope and its software design, and Robert Puccinelli for building the relay-based LED multiplexer. We would also like to thank Madhura Raghavan from the DeRisi lab for assistance with manual parasite counting, and Joshua Batson for discussions related to statistics and deep learning. We are grateful to the late Prof. James Pawley for providing the quartz microscope objective and for insightful written communications on ultraviolet imaging. This work was supported by the Chan Zuckerberg Biohub.

Competing Interests

The authors declare provisional patent application 63/072,037 filed on 08/28/2020.

References

Adu-Gyasi D, Adams M, Amoako S, Mahama E, Nsoh M, Amenga-Etego S, Baiden F, Asante KP, Newton S, Owusu-Agyei S. Estimating malaria parasite density: assumed white blood cell count of 10,000/uL of blood is appropriate measure in Central Ghana. Malaria Journal. 2012 Jul; 11:238. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3411500/, doi: 10.1186/1475-2875-11-238.

Bannister LH, Hopkins JM, Margos G, Dluzewski AR, Mitchell GH. Three-Dimensional Ultrastructure of the Ring Stage of Plasmodium falciparum: Evidence for Export Pathways. Microscopy and Microanalysis. 2004 Oct; 10(5):551–562. http://www.cambridge.org/core/journals/microscopy-and-microanalysis/article/threedimensional-ultrastructure-of-the-ring-stage-of-plasmodium-falciparum-evidence-for-export-pathways/3C7F86320F26D1459342E6406BC79830, doi: 10.1017/S1431927604040917.

Bessis M. Echinocytes. In: Bessis M, editor. *Corpuscles: Atlas of Red Blood Cell Shapes* Berlin, Heidelberg: Springer; 1974.p. 13–20. https://doi.org/10.1007/978-3-642-65657-6_3, doi: 10.1007/978-3-642-65657-6_3.

Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks. 2018 Oct; 106:249–259. http://www.sciencedirect.com/science/article/pii/S0893608018302107, doi: 10.1016/j.neunet.2018.07.011.

Dong Y, Jiang Z, Shen H, Pan WD, Williams LA, Reddy VVB, Benjamin WH, Bryan AW. Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In: *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*; 2017. p. 101–104. doi: 10.1109/BHI.2017.7897215.

Eriksson LEG. On the shape of human red blood cells interacting with flat artificial surfaces — the 'glass effect'. Biochimica et Biophysica Acta (BBA) - General Subjects. 1990 Dec; 1036(3):193–201. http://www.sciencedirect.com/science/article/pii/030441659090034T, doi: 10.1016/0304-4165(90)90034-T.

Escalante AA, Smith DL, Kim Y. The dynamics of mutations associated with anti-malarial drug resistance in Plasmodium falciparum. Trends in Parasitology. 2009 Dec; 25(12):557–563. https://www.cell.com/trends/parasitology/abstract/S1471-4922(09)00216-5, doi: 10.1016/j.pt.2009.09.008.

Gama BE, Lacerda MVG, Daniel-Ribeiro CT, Ferreira-da Cruz MdF. Chemoresistance of Plasmodium falciparum and Plasmodium vivax parasites in Brazil: consequences on disease morbidity and control. Memórias do Instituto Oswaldo Cruz. 2011 Aug; 106:159–166. http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0074-02762011000900020&lng=en&nrm=iso&tlng=en, doi: 10.1590/S0074-02762011000900020.

Gopakumar GP, Swetha M, Siva GS, Subrahmanyam GRKS. Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. Journal of Biophotonics. 2018; 11(3):e201700003. https://onlinelibrary.wiley.com/doi/abs/10.1002/jbio.201700003, doi: 10.1002/jbio.201700003.

Grüring C, Heiber A, Kruse F, Ungefehr J, Gilberger TW, Spielmann T. Development and host cell modifications of Plasmodium falciparum blood stages in four dimensions. Nature Communications. 2011 Jan; 2(1):1–11. http://www.nature.com/articles/ncomms1169, doi: 10.1038/ncomms1169.

Guo SM, Yeh LH, Folkesson J, Ivanov IE, Krishnan AP, Keefe MG, Hashemi E, Shin D, Chhun BB, Cho NH, Leonetti MD, Han MH, Nowakowski TJ, Mehta SB. Revealing architectural order with quantitative label-free imaging and deep learning. eLife. 2020 Jul; 9:e55502. https://doi.org/10.7554/eLife.55502, doi: 10.7554/eLife.55502, publisher: eLife Sciences Publications, Ltd.

Haldar K, Bhattacharjee S, Safeukui I. Drug resistance in Plasmodium. Nature Reviews Microbiology. 2018 Mar; 16(3):156–170. http://www.nature.com/articles/nrmicro.2017.161, doi: 10.1038/nrmicro.2017.161.

He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv:151203385 [cs]. 2015 Dec; http://arxiv.org/abs/1512.03385, arXiv: 1512.03385.

Hernández Candia CN, Gutiérrez-Medina B. Direct Imaging of Phase Objects Enables Conventional Deconvolution in Bright Field Light Microscopy. PLoS ONE. 2014 Feb; 9(2). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3928359/, doi: 10.1371/journal.pone.0089106.

Kang YJ, Ha YR, Lee SJ. High-Throughput and Label-Free Blood-on-a-Chip for Malaria Diagnosis. Analytical Chemistry. 2016 Mar; 88(5):2912–2922. https://doi.org/10.1021/acs.analchem.5b04874, doi: 10.1021/acs.analchem.5b04874.

Kim K, Yoon H, Diez-Silva M, Dao M, Dasari RR, Park Y. High-resolution three-dimensional imaging of red blood cells parasitized by Plasmodium falciparum and in situ hemozoin crystals using optical diffraction tomography. Journal of Biomedical Optics. 2014 Jan; 19(1):011005. doi: 10.1117/1.JBO.19.1.011005.

Kumar B, Bhalla V, Bhadoriya RPS, Suri CR, Varshney GC. Label-free electrochemical detection of malaria-infected red blood cells. RSC Advances. 2016 Aug; 6(79):75862–75869. http://pubs.rsc.org/en/content/articlelanding/2016/ra/c6ra07665c, doi: 10.1039/C6RA07665C.

Li H, Soto-Montoya H, Voisin M, Valenzuela LF, Prakash M. Octopi: Open configurable high-throughput imaging platform for infectious disease diagnosis in the field. bioRxiv. 2019 Jun; p. 684423. https://www.biorxiv.org/content/10.1101/684423v1, doi: 10.1101/684423.

Manser M, Olufsen C, Andrews N, Chiodini PL. Estimating the parasitaemia of Plasmodium falciparum: experience from a national EQA scheme. Malaria Journal. 2013 Nov; 12(1):428. https://doi.org/10.1186/1475-2875-12-428, doi: 10.1186/1475-2875-12-428.

Mathieu LC, Cox H, Early AM, Mok S, Lazrek Y, Paquet JC, Ade MP, Lucchi NW, Grant Q, Udhayakumar V, Alexandre JS, Demar M, Ringwald P, Neafsey DE, Fidock DA, Musset L. Local emergence in Amazonia of Plasmodium falciparum k13 C580Y mutants associated with in vitro artemisinin resistance. eLife. 2020 May; 9:e51015. https://elifesciences.org/articles/51015, doi: 10.7554/eLife.51015.

Ojaghi A, Fay ME, Lam WA, Robles FE. Ultraviolet Hyperspectral Interferometric Microscopy. Scientific Reports. 2018 Jul; 8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6028425/, doi: 10.1038/s41598-018-28208-0.

Packard RM. The Origins of Antimalarial-Drug Resistance. New England Journal of Medicine. 2014 Jul; 371(5):397–399. https://doi.org/10.1056/NEJMp1403340, doi: 10.1056/NEJMp1403340.

Pan WD, Dong Y, Wu D. Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks. Machine Learning - Advanced Techniques and Emerging Applications. 2018 Sep; https://www.intechopen.com/books/machine-learning-advanced-techniques-and-emerging-applications/classification-of-malaria-infected-cells-using-deep-convolutional-neural-networks, doi: 10.5772/intechopen.72426.

Park HS, Rinehart MT, Walzer KA, Chi JTA, Wax A. Automated Detection of P. falciparum Using Machine Learning Algorithms with Quantitative Phase Images of Unstained Cells. PLOS ONE. 2016 Sep; 11(9):e0163045. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0163045, doi: 10.1371/journal.pone.0163045.

Peng WK, Kong TF, Ng CS, Chen L, Huang Y, Bhagat AAS, Nguyen NT, Preiser PR, Han J. Micromagnetic resonance relaxometry for rapid label-free malaria diagnosis. Nature Medicine. 2014 Sep; 20(9):1069–1073. http://www.nature.com/articles/nm.3622, doi: 10.1038/nm.3622.

Petruccelli JC, Tian L, Barbastathis G. The transport of intensity equation for optical path length recovery using partially coherent illumination. Optics Express. 2013 Jun; 21(12):14430. https://www.osapublishing.org/oe/abstract.cfm?uri=oe-21-12-14430, doi: 10.1364/OE.21.014430.

Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G. Image analysis and machine learning for detecting malaria. Translational Research. 2018 Apr; 194:36–55. https://linkinghub.elsevier.com/retrieve/pii/S193152441730333X, doi: 10.1016/j.trsl.2017.12.004.

R a J, K S, C G. An ultraviolet fluorescence-based method for identifying and distinguishing protein crystals. Acta Crystallographica Section D. 2005; 61(1):60–66. //scripts.iucr.org/cgi-bin/paper?wd5024, doi: 10.1107/S0907444904026538.

Rajaraman S, Jaeger S, Antani SK. Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. PeerJ. 2019 May; 7:e6977. https://peerj.com/articles/6977, doi: 10.7717/peerj.6977.

Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. *Advances in Neural Information Processing Systems 28* Curran Associates, Inc.; 2015.p. 91–99. http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf.

Ribaut C, Berry A, Chevalley S, Reybier K, Morlais I, Parzy D, Nepveu F, Benoit-Vical F, Valentin A. Concentration and purification by magnetic separation of the erythrocytic stages of all human Plasmodium species. Malaria Journal. 2008; 7(1):45.

Rivenson Y, Wang H, Wei Z, de Haan K, Zhang Y, Wu Y, Günaydın H, Zuckerman JE, Chong T, Sisk AE, Westbrook LM, Wallace WD, Ozcan A. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. Nature Biomedical Engineering. 2019 Jun; 3(6):466–477. http://www.nature.com/articles/s41551-019-0362-y, doi: 10.1038/s41551-019-0362-y.

Schuster FL. Cultivation of Plasmodium spp. Clinical Microbiology Reviews. 2002 Jul; 15(3):355–364. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC118084/, doi: 10.1128/CMR.15.3.355-364.2002.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2015. p. 1–9. doi: 10.1109/CVPR.2015.7298594, iSSN: 1063-6919.

Tryolabs, Luminoth: Deep Learning toolkit for Computer Vision. Tryolabs; 2018. https://github.com/tryolabs/luminoth.

Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Janda M, Lallas A, Longo C, Malvehy J, Paoli J, Puig S, Rosendahl C, Soyer HP, Zalaudek I, Kittler H. Human–computer collaboration for skin cancer recognition. Nature Medicine. 2020 Jun; p. 1–6. https://www.nature.com/articles/s41591-020-0942-0, doi: 10.1038/s41591-020-0942-0, publisher: Nature Publishing Group.

WHO. Malaria microscopy quality assurance manual. Version 2 ed. Geneva: World Health Organization; 2016. OCLC: ocn961249779.

WHO. World Malaria Report 2019. World Health Organization; 2019. https://www.who.int/publications-detail/world-malaria-report-2019.

Yang F, Poostchi M, Yu H, Zhou Z, Silamut K, Yu J, Maude RJ, Jaeger S, Antani S. Deep Learning for Smartphone-based Malaria Parasite Detection in Thick Blood Smears. IEEE Journal of Biomedical and Health Informatics. 2019; p. 1–1. doi: 10.1109/JBHI.2019.2939121.

Yang X, Chen Z, Miao J, Cui L, Guan W. High-throughput and label-free parasitemia quantification and stage differentiation for malaria-infected red blood cells. Biosensors and Bioelectronics. 2017 Dec; 98:408–414. http://www.sciencedirect.com/science/article/pii/S0956566317304670, doi: 10.1016/j.bios.2017.07.019.

Zeskind BJ, Jordan CD, Timp W, Trapani L, Waller G, Horodincu V, Ehrlich DJ, Matsudaira P. Nucleic acid and protein mass mapping by live-cell deep-ultraviolet microscopy. Nature Methods. 2007 Jul; 4(7):567–569. http://www.nature.com/articles/nmeth1053, doi: 10.1038/nmeth1053.

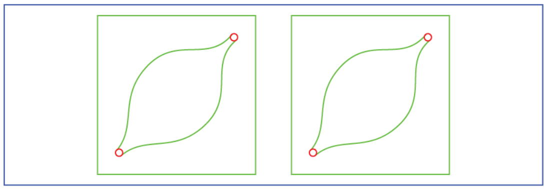


Figure 1–Figure supplement 1. Flow cells were constructed as follows: 1 mm circular holes (shown in red below) were laser-cut into quartz slides (Ted Pella 26011, outline in blue. Laser-cutter model ULS-P150D). Gaskets (green) were laser-cut from Parafilm (Sigma Aldrich P7793). After cutting, gaskets were aligned by hand, then pressed down gently around the exterior. Quartz coverslips (Ted Pella 26014) were aligned by hand to the gaskets, and then pressed gently. The assembly was sealed by compression of the flow cell (protected inside a single layer aluminum foil envelope) with a mass of 300 g for five minutes, on a hot plate set to 75°C.

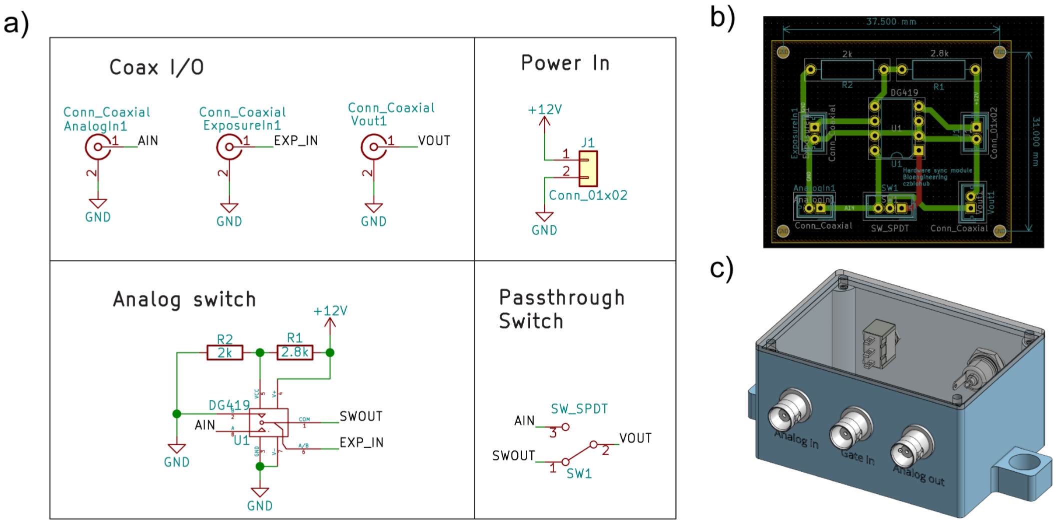


Figure 1–Figure supplement 2. Design of the hardware sync module: a) Schematic diagram of the simple circuit used to synchronize LED emission with camera exposure. The circuit uses an analog switch (Maxim DG419) to connect either a programmable analog voltage signal (AIN) or ground (GND) to the output, gated by the camera exposure's logic signal. In series with the analog switch is a manual toggle switch to optional bypass the analog switch, disabling the sync module. b) PCB layout diagram for the hardware sync module. Note that coaxial connector footprints were replaced with standard 0.1" headers, and connected via wires to panel-mount BNC connectors. The circuit was fabricated as a PCB using an LPKF Protomat S103 circuit mill, then hand soldered and assembled. c) Solid model screenshot of the 3D-printed enclosure for the hardware sync module.

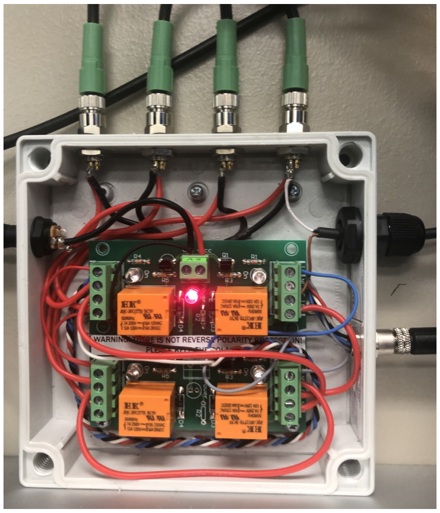


Figure 1–Figure supplement 3. Four-channel relay multiplexer. The LED multiplexer uses a four-channel relay (Denkovi DAE-RB/Ro4-JQC-5V) to route the current generated by a Thorlabs LED current driver (LEDD1B) to one of four LEDs. The relay board was mounted inside a box to house the wired connections between relays. The box received a three-pin signal (plus GND). The first relay was used as an overall ON/OFF switch, while the remaining three were arranged in a binary tree to route current from the driver to any of the four LEDs.

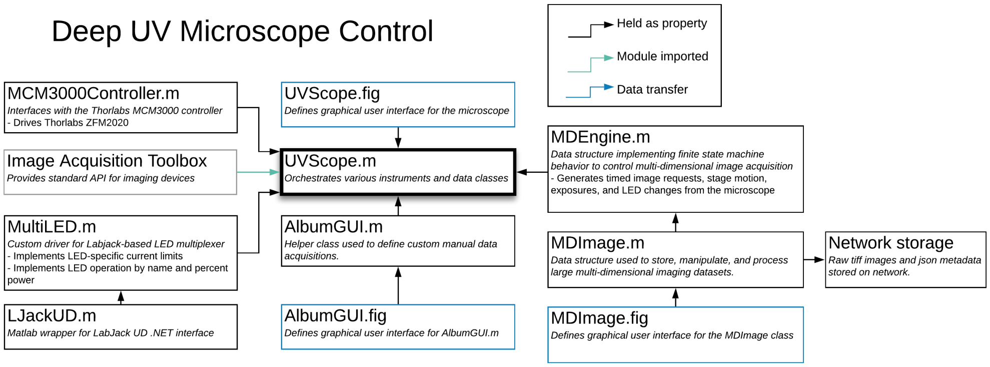

Deep UV Microscope Control
Held as property
Module imported
Data transfer
MCM3000Controller.m
Interfaces with the Thorlabs MCM3000 controller
- Drives Thorlabs ZFM2020
UVScope.fig
Defines graphical user interface for the microscope
Image Acquisition Toolbox
Provides standard API for imaging devices
UVScope.m
Orchestrates various instruments and data classes
MDEngine.m
Data structure implementing finite state machine behavior to control multi-dimensional image acquisition
- Generates timed image requests, stage motion, exposures, and LED changes from the microscope
MultiLED.m
Custom driver for Labjack-based LED multiplexer
- Implements LED-specific current limits
- Implements LED operation by name and percent power
AlbumGUI.m
Helper class used to define custom manual data acquisitions.
MDImage.m
Data structure used to store, manipulate, and process large multi-dimensional imaging datasets.
Network storage
Raw tiff images and json metadata stored on network.
LJackUD.m
Matlab wrapper for LabJack UD .NET interface
AlbumGUI.fig
Defines graphical user interface for AlbumGUI.m
MDImage.fig
Defines graphical user interface for the MDImage class

Figure 1–Figure supplement 4. The UV microscope control software uses a set of custom classes to configure and execute multi-dimensional image acquisitions. Multiplexed LED illumination was driven by a data class which maps maximum allowable LED currents to control voltages, sending a digital address to the relay multiplexer for LED selection, and an analog voltage representing percent power. The MDEngine class implements the conversion from acquisition parameters to active control of the UVScope, including real-time focus tracking across large sample areas, in order to keep focal stacks centered on the sample. All microscope control software can be found here: https://github.com/czbiohub/UVScope-control.

882

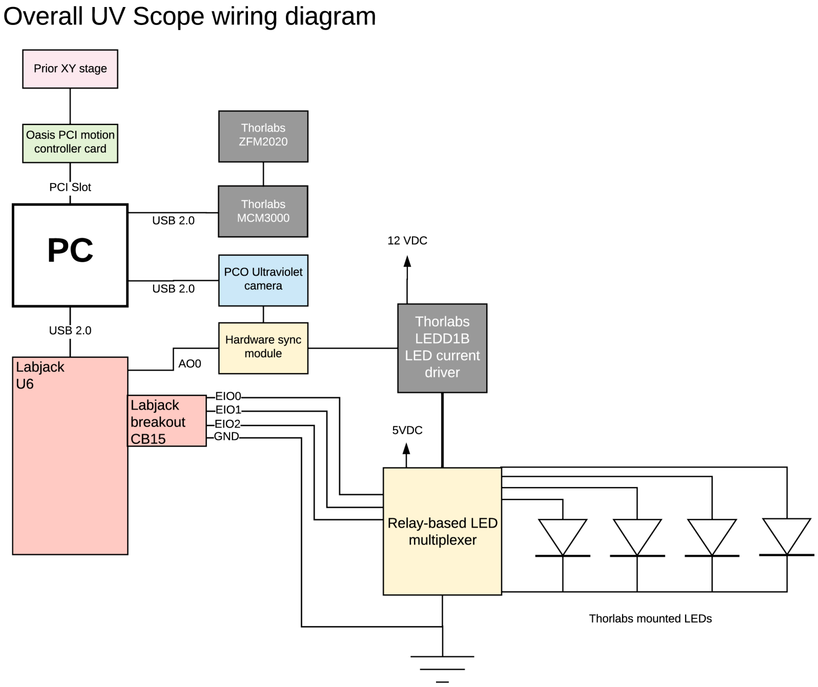


Figure 1–Figure supplement 5. Overall wiring diagram of the UV Scope. The microscope hardware was controlled by a centralized PC running Windows 10. All the various hardware devices were controlled by a combination of standard and custom Matlab classes, as described in Figure 4. Other specific hardware configurations are possible with minor changes to the software.

883

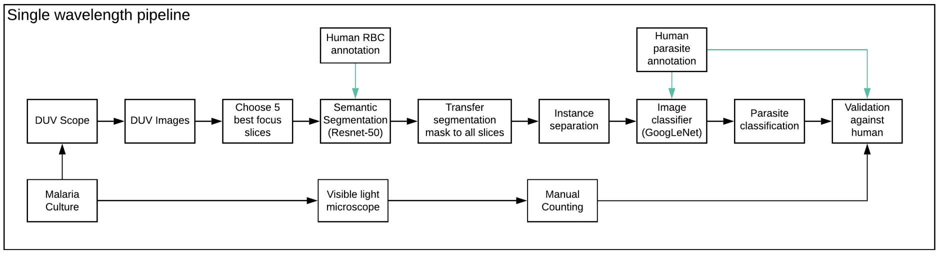

Single wavelength pipeline
Human RBC annotation
Human parasite annotation
DUV Scope
DUV Images
Choose 5 best focus slices
Semantic Segmentation (Resnet-50)
Transfer segmentation mask to all slices
Instance separation
Image classifier (GoogLeNet)
Parasite classification
Validation against human
Malaria Culture
Visible light microscope
Manual Counting

Figure 1–Figure supplement 6. Single-wavelength image processing pipeline.

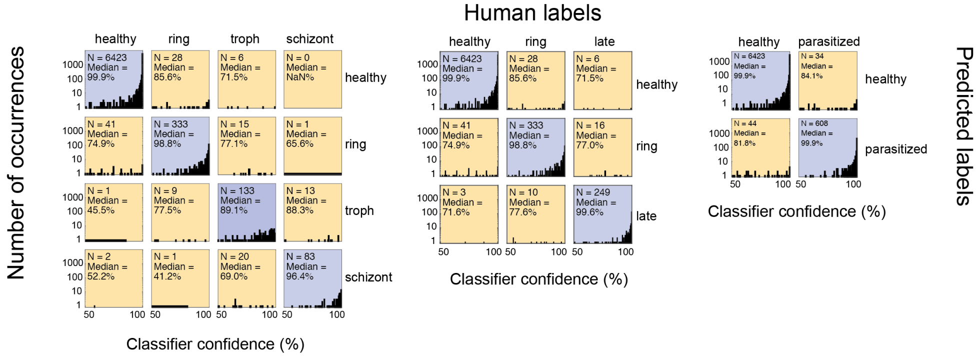


Figure 2–Figure supplement 1. Classifier confidence statistics indicate mis-classified cells are more likely to be low-confidence. In all three plots, log-scaled histograms of classifier confidence scores are displayed in the pattern of a confusion matrix. Statistics are derived by evaluating the validation dataset from the 285 nm classifier. Total instance counts for each matrix element as well as the medians of the distributions are shown as text insets.

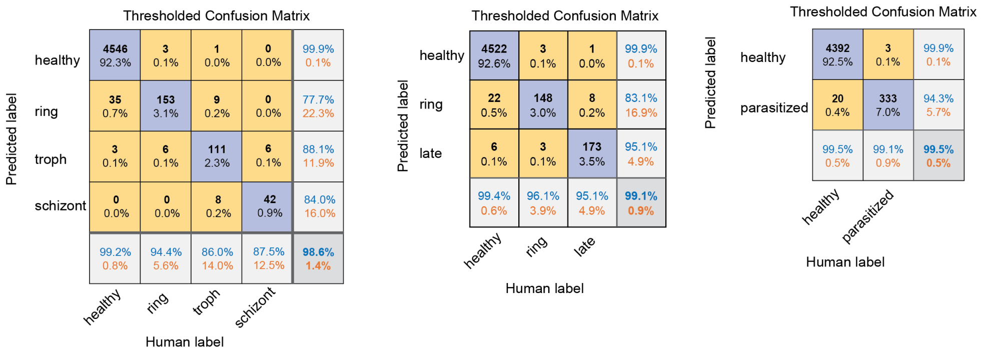


Figure 2–Figure supplement 2. Confusion matrices from the custom UV scope at 285 nm show improved statistics after rejection of RBCs based on meeting the minimum classifier confidence criteria. These confusion matrices show the result of applying the empirically-optimal threshold values of 61%, 76%, and 96%, to the four, three, and two-category classifier confusion matrices.

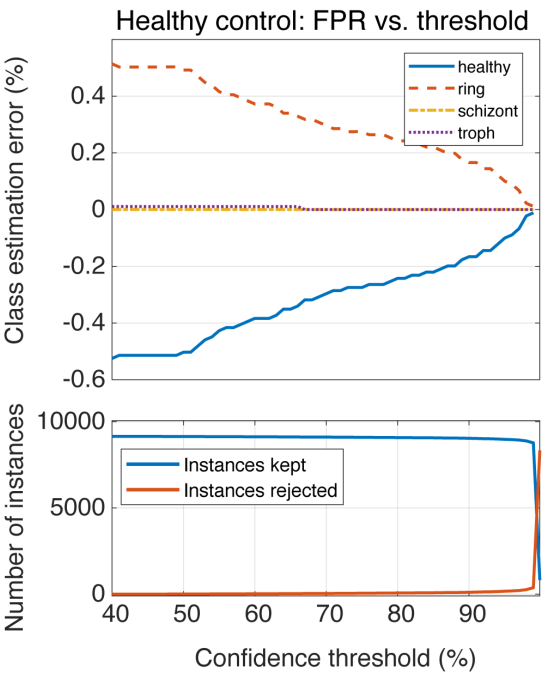
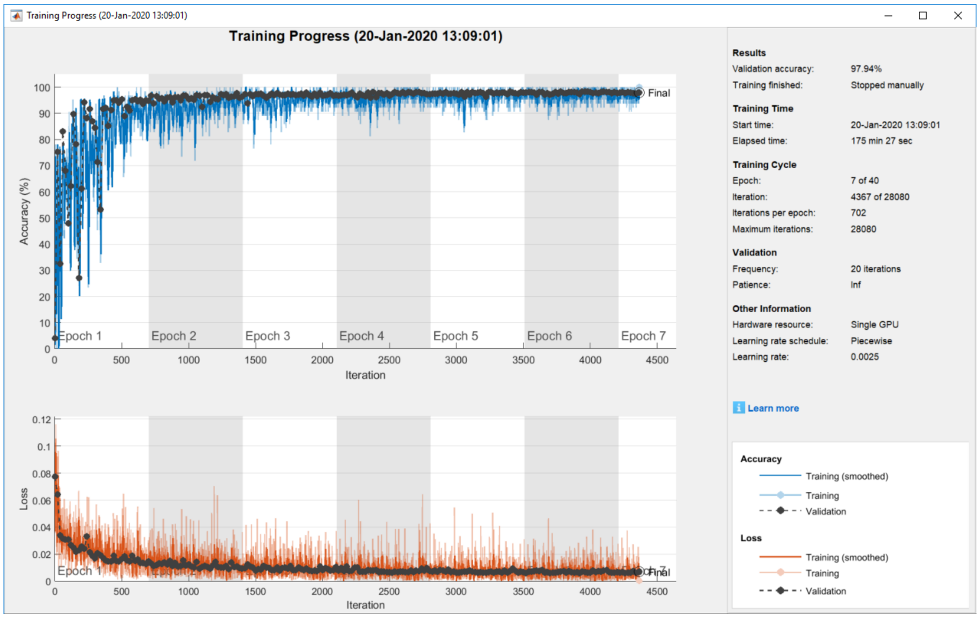


Figure 2–Figure supplement 3. A healthy control dataset was processed using our four-category classifier operating at 285 nm. Top: Sample composition estimates are shown as a function of confidence threshold, demonstrating that the approximately 0.5% raw FPR can be reduced by rejecting cells with low classifier confidence. Bottom: The number of cells kept (blue) and rejected (red) by the thresholding process are plotted as a function of threshold value.


Training Progress (20-Jan-2020 13:09:01)

Results
Validation accuracy: 97.94%
Training finished: Stopped manually

Training Time
Start time: 20-Jan-2020 13:09:01
Elapsed time: 175 min 27 sec

Training Cycle
Epoch: 7 of 40
Iteration: 4367 of 28080
Iterations per epoch: 702
Maximum iterations: 28080

Validation
Frequency: 20 iterations
Patience: Inf

Other Information
Hardware resource: Single GPU
Learning rate schedule: Piecewise
Learning rate: 0.0025

Figure 2–Figure supplement 4. Statistics from training the 285 nm classifier. The upper plot shows the overall accuracy as a function of iteration and epoch number. The blue line plot (no markers) shows training dataset accuracy while the black line plot (dashed with circle markers) shows validation accuracy. The lower plot shows the loss function (red with no markers: training loss; black with circle markers: validation loss

888

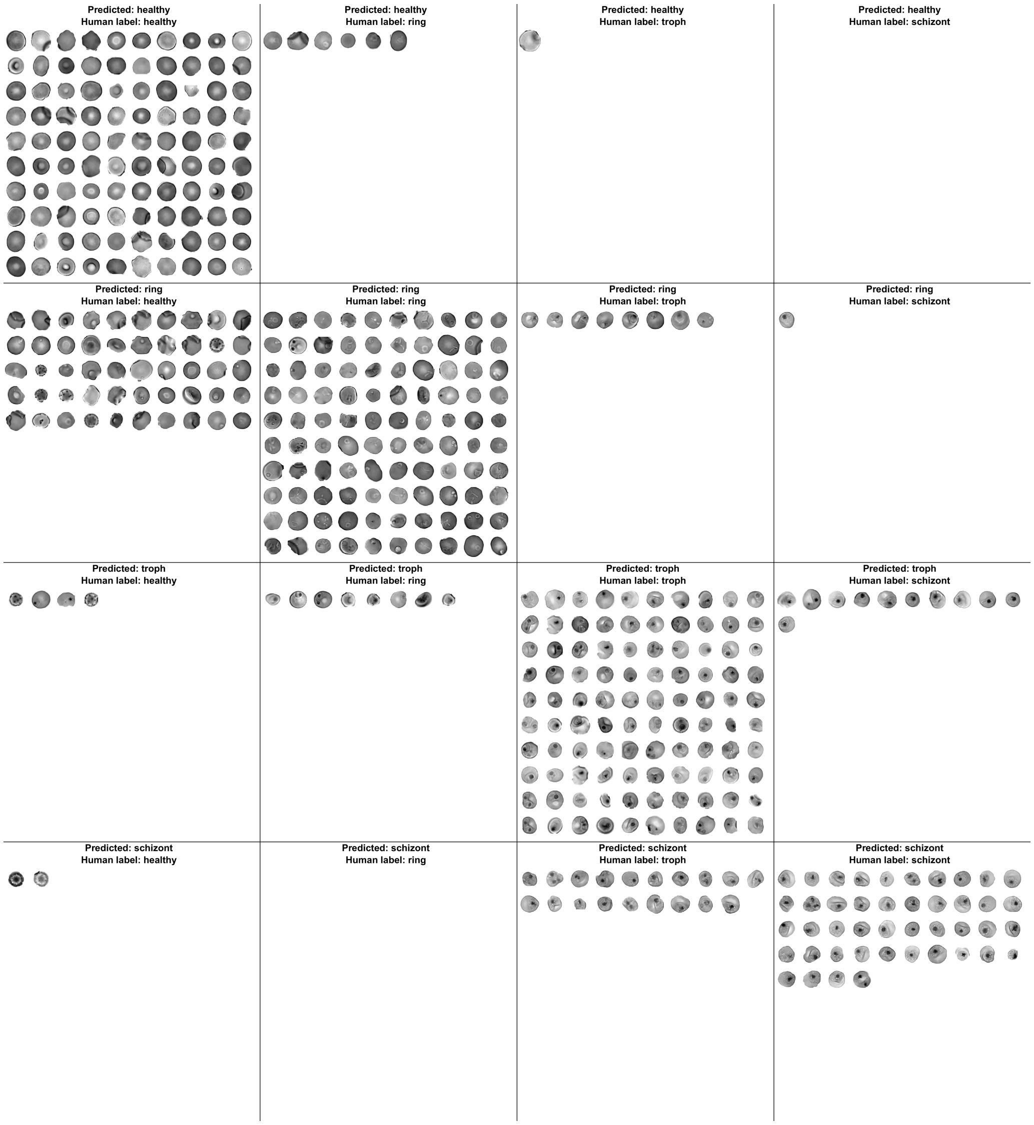


Figure 2–Figure supplement 5. Example RBC instances from each category in the confusion matrix for the UV Scope at 285 nm wavelength. The arrangement of the categories is the same as Figure 2 from the main text. For this montage, a maximum array size of 10 x 10 was used for display. Note that for many of the categories (especially off-diagonal entries) there were not enough examples to populate the array, in which case the space was left blank. All data from this figure was taken from the validation dataset, as opposed to the training dataset.

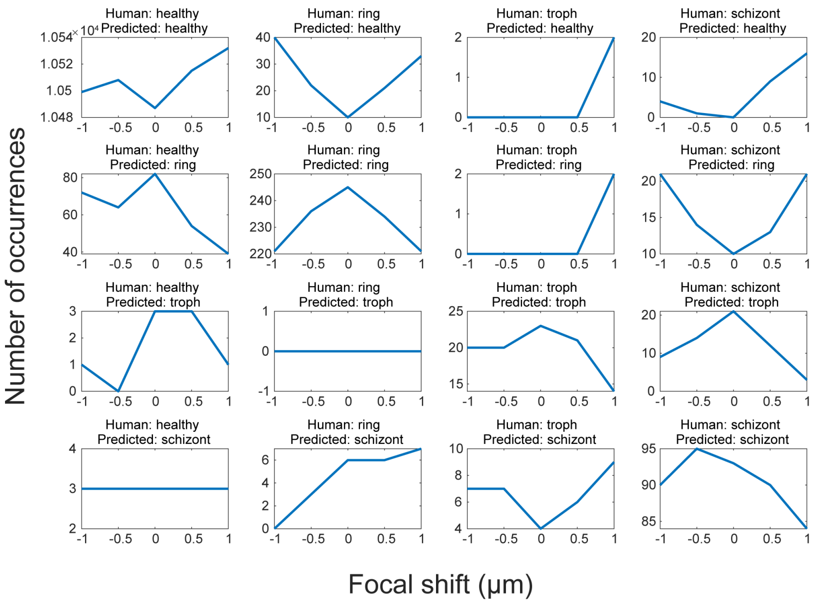


Figure 3–Figure supplement 1. Slice-dependent four-category classification performance arranged as a confusion matrix. All panels spatially correspond to confusion matrix elements from Figure 2, and display relative changes in matrix values as a function of defocus. Diagonal matrix entries (true positives) corresponding to all three parasite life cycles stages tend to improve wtih global best focus, while many off-diagonal entries (false negatives / false positives) see reductions at the global best focus. However, exceptions include ring false positives (second row, first column) and troph/schizont cross-identification.

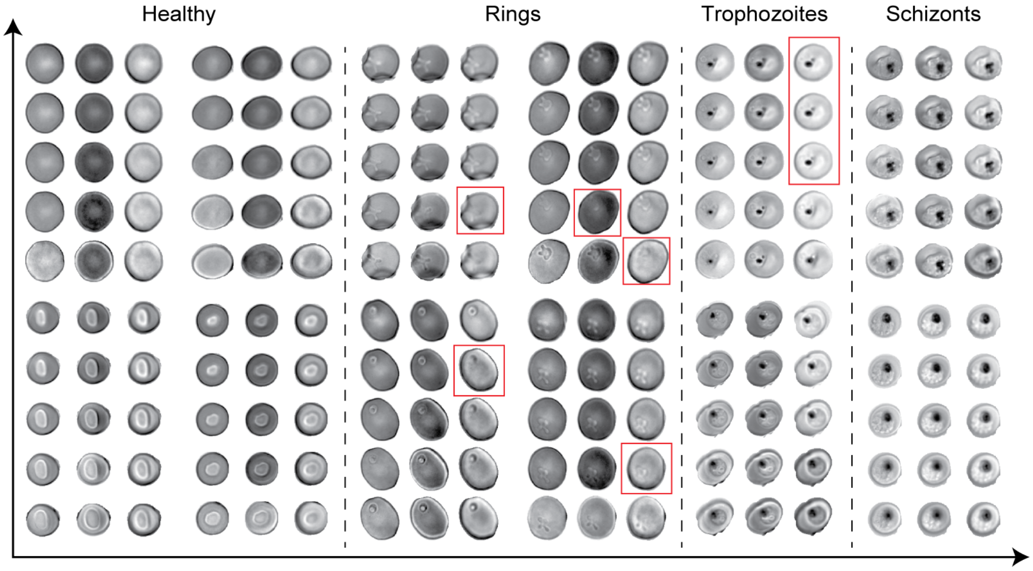
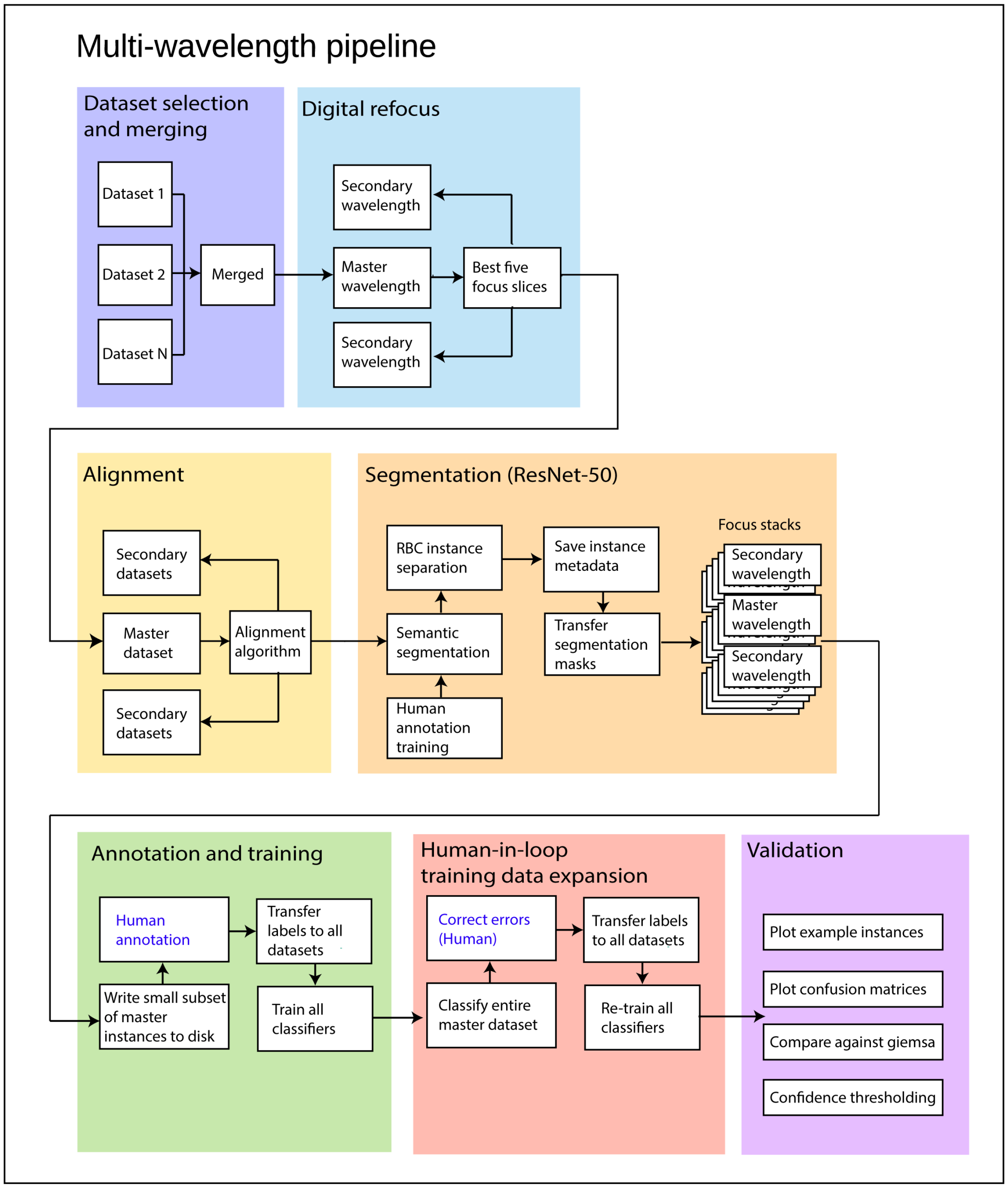


Figure 4–Figure supplement 1. Color-focus arrays for many example RBCs at various stages of infection. For all panels, the color/focus layout is identical to Figure 4. Red boxes highlight cells that are not easily distinguished with visible light due to loss of contrast at certain focal planes. Examples of all four categorized life cycle stages are shown.



891

Figure 4–Figure supplement 2. Our multi-wavelength image analysis pipeline had the following steps: 1. Load data structures (UV scope) or raw images (commercial scope). 2. Apply re-focusing. 3. Register wavelength channels using affine transformation. 4. Export pre-processed images. 5. Do semantic segmentation on the master wavelength/slice using ResNet-50. 6. Apply instancing algorithm to all wavelengths/slices, exporting RBC instances to disk. 7. Classify all RBC instances for only the master wavelength and slice. 8. Apply the resulting labels to all the corresponding instances in the other slices/channels. 9. Export the lowest-confidence subset of the machine-classified images for human annotation. 10. Import the results of human annotation and propagate labels to all the other channels/slices. 11. Train new classifiers with the high confidence dataset.

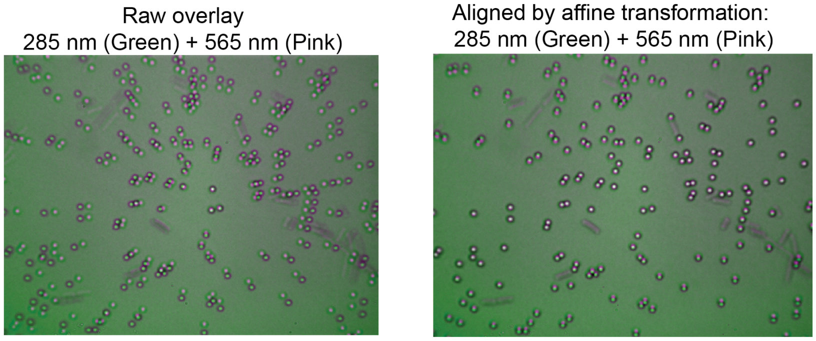
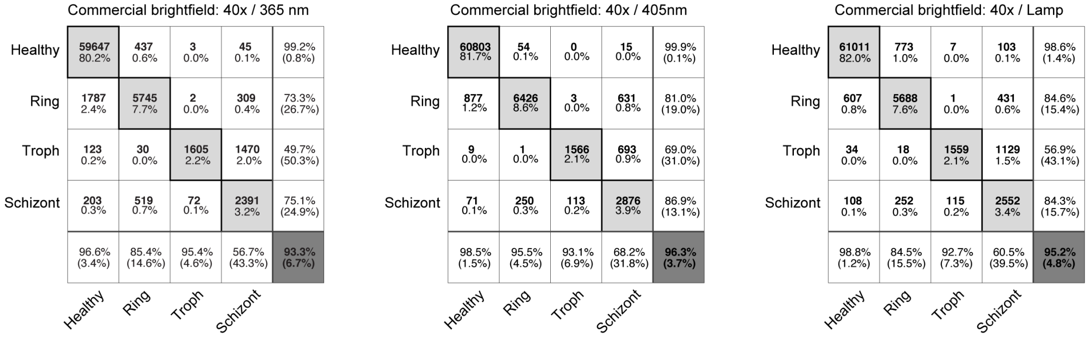

Raw overlay
285 nm (Green) + 565 nm (Pink)
Aligned by affine transformation:
285 nm (Green) + 565 nm (Pink)

Figure 4–Figure supplement 3. Affine transformations were performed for image registration between different wavelengths. Left: overlay of two raw images from the UV scope, acquired at 285 nm (green) and 565 nm (pink). Right: overlay of the same images after image registration was completed. The sample in the images consists of fixed *E.Coli* cells (rods) mixed with 1 μm beads (circles).



Commercial brightfield: 40x / 365 nm

	Healthy	Ring	Troph	Schizont	
Healthy	59647 80.2%	437 0.6%	3 0.0%	45 0.1%	99.2% (0.8%)
Ring	1787 2.4%	5745 7.7%	2 0.0%	309 0.4%	73.3% (26.7%)
Troph	123 0.2%	30 0.0%	1605 2.2%	1470 2.0%	49.7% (50.3%)
Schizont	203 0.3%	519 0.7%	72 0.1%	2391 3.2%	75.1% (24.9%)
	96.6% (3.4%)	85.4% (14.6%)	95.4% (4.6%)	56.7% (43.3%)	93.3% (6.7%)

Commercial brightfield: 40x / 405nm

	Healthy	Ring	Troph	Schizont	
Healthy	60803 81.7%	54 0.1%	0 0.0%	15 0.0%	99.9% (0.1%)
Ring	877 1.2%	6426 8.6%	3 0.0%	631 0.8%	81.0% (19.0%)
Troph	9 0.0%	1 0.0%	1566 2.1%	693 0.9%	69.0% (31.0%)
Schizont	71 0.1%	250 0.3%	113 0.2%	2876 3.9%	86.9% (13.1%)
	98.5% (1.5%)	95.5% (4.5%)	93.1% (6.9%)	68.2% (31.8%)	96.3% (3.7%)

Commercial brightfield: 40x / Lamp

	Healthy	Ring	Troph	Schizont	
Healthy	61011 82.0%	773 1.0%	7 0.0%	103 0.1%	98.6% (1.4%)
Ring	607 0.8%	5688 7.6%	1 0.0%	431 0.6%	84.6% (15.4%)
Troph	34 0.0%	18 0.0%	1559 2.1%	1129 1.5%	56.9% (43.1%)
Schizont	108 0.1%	252 0.3%	115 0.2%	2552 3.4%	84.3% (15.7%)
	98.8% (1.2%)	84.5% (15.5%)	92.7% (7.3%)	60.5% (39.5%)	95.2% (4.8%)

Figure 5–Figure supplement 1. Confusion matrices for classifiers trained at each wavelength on our commercial microscope. Left: 365 nm, Middle: 405 nm, Right: Visible light lamp. All classifiers were trained on the identical set of RBC instances, using images acquired each at their respective wavelength (multiwavelength dataset, see table 1).

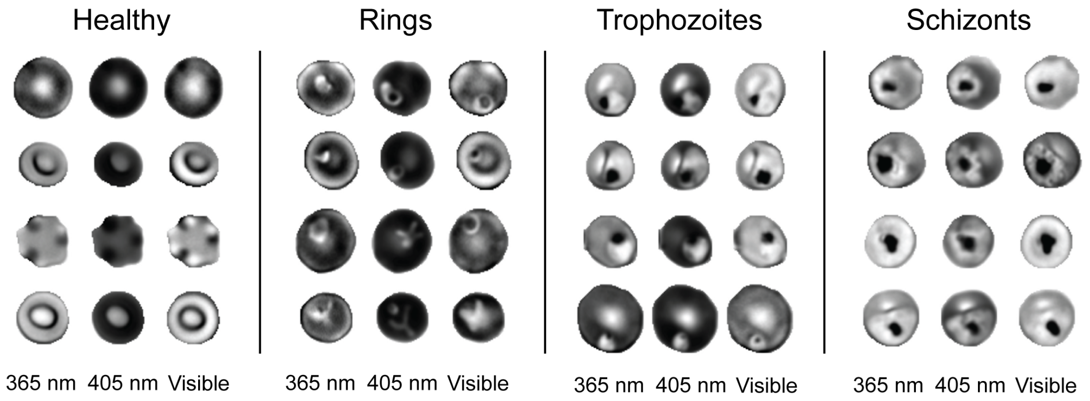


Figure 5–Figure supplement 2. Example RBC images acquired on our commercial microscope for all three wavelengths (left columns: 365 nm, middle columns: 405 nm, right columns: standard visible lamp). Four example cells are displayed for each category, where each row within a category are images of the same physical cell. Ring stage parasites are seen to be highly mobile, as evidenced by their differing positions and even morphology between acquisitions at different wavelengths, which were separated in time on the order of a few minutes.

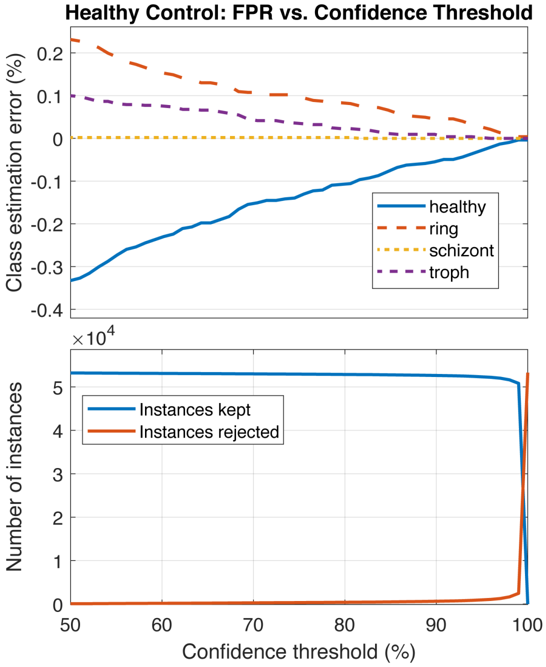
895 

Figure 5–Figure supplement 3. Parameterization of the false-positive rate from a healthy control sample imaged on our commercial microscope at 405 nm.The upper plot displays the sample composition error for all four stages as a function of confidence threshold. Data processed were from a healthy donor sample, uninfected by *Plasmodium*. In particular, the FPR for rings at 50% confidence threshold is substantially lower than the results reported in Figure 5. We explain this by noting this particular healthy control sample exhibited very few echinocytes, to which we attribute the low basal FPRs as compared to Figure 5.

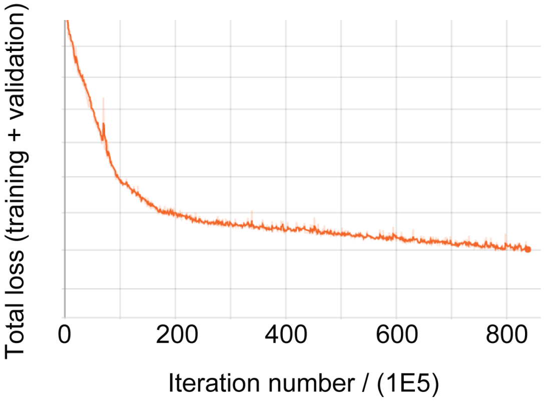
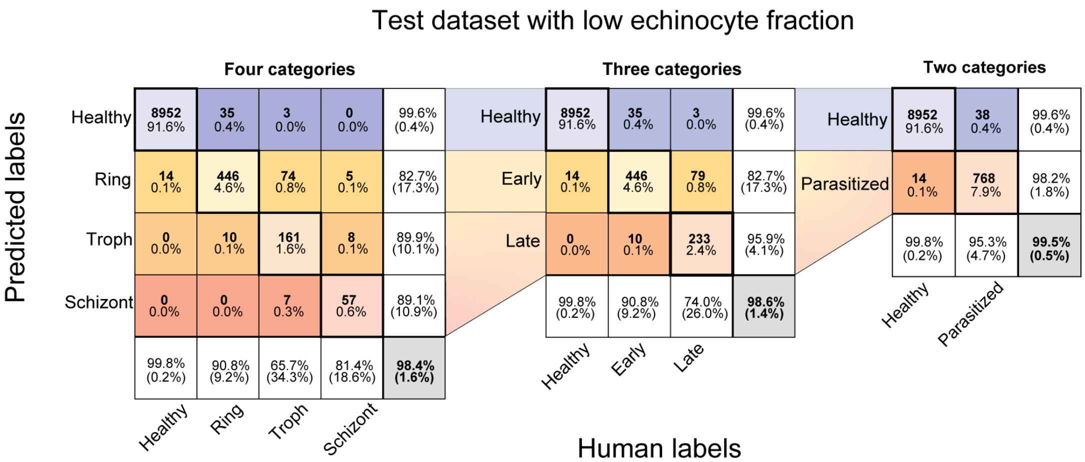
896 

Figure 7–Figure supplement 1. Total loss function (training + validation) during Faster-RCNN training.



Four categories

	Healthy	Ring	Troph	Schizont	
Healthy	**8952** 91.6%	**35** 0.4%	**3** 0.0%	**0** 0.0%	99.6% (0.4%)
Ring	**14** 0.1%	**446** 4.6%	**74** 0.8%	**5** 0.1%	82.7% (17.3%)
Troph	**0** 0.0%	**10** 0.1%	**161** 1.6%	**8** 0.1%	89.9% (10.1%)
Schizont	**0** 0.0%	**0** 0.0%	**7** 0.3%	**57** 0.6%	89.1% (10.9%)
	99.8% (0.2%)	90.8% (9.2%)	65.7% (34.3%)	81.4% (18.6%)	**98.4%** **(1.6%)**

Three categories

	Healthy	Early	Late	
Healthy	**8952** 91.6%	**35** 0.4%	**3** 0.0%	99.6% (0.4%)
Early	**14** 0.1%	**446** 4.6%	**79** 0.8%	82.7% (17.3%)
Late	**0** 0.0%	**10** 0.1%	**233** 2.4%	95.9% (4.1%)
	99.8% (0.2%)	90.8% (9.2%)	74.0% (26.0%)	**98.6%** **(1.4%)**

Two categories

	Healthy	Parasitized	
Healthy	**8952** 91.6%	**38** 0.4%	99.6% (0.4%)
Parasitized	**14** 0.1%	**768** 7.9%	98.2% (1.8%)
	99.8% (0.2%)	95.3% (4.7%)	**99.5%** **(0.5%)**

Figure 8–Figure supplement 1. Confusion matrices for four, three, and two-category classifiers at 285 nm tested on a random sub-partition of pooled titration data, which as a whole contained a low echinocyte fraction. As compared with Figure 2, the FPR for rings is reduced presumably due to fewer cells containing anomalous morphology, and is in closer agreement to the result extracted by least squares fit to the titration data in Figure 8. The random sub-partition was configured to contain overall parasite prevalence similar to Figure 2.