RESEARCH ARTICLE

Expert-level automated malaria diagnosis on routine blood films with deep neural networks

Petru Manescu[1] | Michael J. Shaw[1] | Muna Elmi[1] | Lydia Neary-Zajiczek[1] | Remy Claveau[1] | Vijay Pawar[1] | Iasonas Kokkinos[1] | Gbeminiyi Oyinloye[2,3] | Christopher Bendkowski[1] | Olajide A. Oladejo[4,5] | Bolanle F. Oladejo[4,5] | Tristan Clark[1] | Denis Timm[1] | John Shawe-Taylor[1] | Mandayam A. Srinivasan[1] | Ikeoluwa Lagunju[2,3,5] | Olugbemiro Sodeinde[1,2,3] | Biobele J. Brown[2,3,5] | Delmiro Fernandez-Reyes[1,2,3,5] 🆔

[1]Department of Computer Science, Faculty of Engineering Sciences, University College London, London, UK

[2]Department of Paediatrics, College of Medicine University of Ibadan, University College Hospital, Ibadan, Nigeria

[3]Childhood Malaria Research Group, College of Medicine University of Ibadan, University College Hospital, Ibadan, Nigeria

[4]Department of Computer Science, University of Ibadan, Ibadan, Nigeria

[5]African Computational Sciences Centre for Health and Development, University of Ibadan, Ibadan, Nigeria

Correspondence
Delmiro Fernandez-Reyes, University College London, Gower Street, London, WC1E 6BT, UK.
Email: delmiro.fernandez-reyes@ucl.ac.uk

Abstract

Over 200 million malaria cases globally lead to half a million deaths annually. Accurate malaria diagnosis remains a challenge. Automated imaging processing approaches to analyze Thick Blood Films (TBF) could provide scalable solutions, for urban healthcare providers in the holoendemic malaria sub-Saharan region. Although several approaches have been attempted to identify malaria parasites in TBF, none have achieved negative and positive predictive performance suitable for clinical use in the west sub-Saharan region. While malaria parasite object detection remains an intermediary step in achieving automatic patient diagnosis, training state-of-the-art deep-learning object detectors requires the human-expert labor-intensive process of labeling a large dataset of digitized TBF. To overcome these challenges and to achieve a clinically usable system, we show a novel approach. It leverages routine clinical-microscopy labels from our quality-controlled malaria clinics, to train a Deep Malaria Convolutional Neural Network classifier (DeepMCNN) for automated malaria diagnosis. Our system also provides total Malaria Parasite (MP) and White Blood Cell (WBC) counts allowing parasitemia estimation in MP/μL, as recommended by the WHO. Prospective validation of the DeepMCNN achieves sensitivity/specificity of 0.92/0.90 against expert-level malaria diagnosis. Our approach PPV/NPV performance is of 0.92/0.90, which is clinically usable in our holoendemic settings in the densely populated metropolis of Ibadan. It is located within the most populous African country (Nigeria) and with one of the largest burdens of *Plasmodium falciparum* malaria. Our openly available method is of importance for strategies aimed to scale malaria diagnosis in urban regions where daily assessment of thousands of specimens is required.

1 | INTRODUCTION

Plasmodium falciparum malaria remains one of the greatest global health burdens with over 219 million cases globally in 2017.[1] It is a widely prevalent disease, especially ubiquitous in parts of sub-Saharan Africa. In 2017 there were approximately 435 000 deaths due to malaria worldwide, with the African region accounting for 93% of these deaths, mostly among children.[1] Early diagnosis is important for reducing the mortality rate due to malaria. Although there are a range of techniques that have been developed for the diagnosis of malaria,[2,3] conventional light microscopy on Giemsa-stained thick and thin blood films remains the gold standard.[1] Techniques such as polymerase chain reaction, flow cytometric assay[4] and fluorescence-dye based[5] approaches lack a universally standardized methodology, present high costs, and require quality control improvement.[2] While some of these approaches have shown promising results independently, they require infrastructure (eg, cold chain logistics for preservation of reagents) which makes them poorly suited for the resource-constrained sub-Saharan African region. Other methods based on lateral flow assays, known as malaria rapid diagnostic tests, are not ubiquitous in all settings, do not provide estimates of parasitemia and have not been able to outperform the well-established TBF clinical microscopy for malaria diagnosis.[6]

Thick Blood Film (TBF) microscopy remains the internationally recognized gold standard.[1] Thick blood film clinical microscopy requires a trained human microscopist to visually inspect Giemsa stained blood films under a light microscope, to identify and count the *P. falciparum* parasites. Unfortunately, visual inspection of thick blood films strongly relies on the availability of trained personnel, and it is time-consuming and subject to human error caused by fatigue and cognitive overload in busy clinical-microscopy services. As with other visual based diagnostic techniques, accuracy depends on individual technician performance, which makes standardization difficult and reliability poor.[7] A wrong diagnosis of malaria can have negative consequences for patients and for anti-malarial therapy resources. Additionally, shortcomings in the availability of trained personnel in certain regions of the world can lead to over-treatment, which subsequently leads to parasite resistance.

The World Health Organization (WHO) has persistently encouraged the development of rapid and efficient diagnostic testing that will allow proper treatment to be given on time. Here we address the problem of automated diagnosis in color brightfield digitized images of Giemsa-stained thick blood films captured with a 100x/1.4 N.A. oil-immersion objective lens. The thick blood film, a concentration technique, is desirable for analysis compared to thin blood smears or red-cell monolayers, because a larger volume of blood is examined, and thus potentially higher parasite density per image field providing greater sensitivity.

Although a series of classical computational vision and machine learning approaches have been used to identify various types of malaria parasites in digitized thin Giemsa-stained blood smears[8,9] or fluorescence-dye based red-cell monolayer,[5] only a few have attempted parasite detection in digitized thick blood films.[10,11] More recently, some studies have attempted to use deep learning classifiers[12-14] to distinguish malaria parasites from staining artifacts, which remains a challenge.[15]

Advances in deep learning methods for object detection in natural images[16] offer great potential for malaria parasite detection in blood films.[17] However, training of such object detectors involves an extremely laborious process. Human experts label ring malaria parasites in large numbers of field-of-views from digitized thick blood films. Moreover, parasite object detection remains an intermediary step in achieving automatic patient diagnosis, which requires the analysis of multiple fields-of-view (FoV) of the TBF. The inherent parasite false positives detected by the computer vision approaches need to be taken into consideration when establishing such a final diagnosis. To the best of our knowledge, only one group has attempted to do this on samples from patients admitted in clinics for malaria testing.[12,13] Briefly, their approach consisted in classifying a positive malaria sample, if the number of parasites detected by a deep learning model in 300 FoV surpassed a certain empirically determined threshold. Such an approach is likely to misclassify samples with low parasite counts.

To overcome these challenges and achieve a clinically usable Positive and Negative Predictive (PPV/NPV) performance, here we show a novel way to leverage routine clinical-microscopy diagnostic labels. They are from our quality-controlled malaria clinics, and will be used to train a Deep Malaria Convolutional Neural Network classifier (DeepMCNN) suitable for automated malaria diagnosis. We prospectively validate the DeepMCNN against expert-level diagnosis and assess its performance across the all-year-round malaria context of our clinical healthcare settings. This is in the densely populated metropolis of Ibadan, located within the most populous country of Africa (Nigeria), with one of the largest burdens of *P. falciparum* malaria.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

The internationally recognized ethics committee at the Institute for Advanced Medical Research and Training (IAMRAT) of the College of Medicine, University of Ibadan (COMUI) approved this research. It is on the platform of the Childhood Malaria Research Group (CMRG) within the academic Department of Pediatrics, University of Ibadan. It is also at school and Primary Care centers throughout the city of Ibadan, with permit numbers: UI/EC/10/0130, UI/EC/19/0110. Parents and/or guardians of study participants gave informed written consent in accordance with the World Medical Association ethical principles for research involving human subjects.

2.2 | Study site

All study participants were recruited under the auspices of the Childhood Malaria Research Group (CMRG) at the 850-bed tertiary

hospital, University College Hospital (UCH) in the city of Ibadan, Nigeria, in west sub-Saharan Africa. Ibadan is a densely populated urban metropolis in Nigeria with about 5-million inhabitants. Malaria transmission and severe disease occur throughout the year. Although severe malaria syndromes are predominant in children under 5 years, there is still a large burden of severe disease in children up to 16.[18-20]

2.3 | Malaria screening

Malaria parasites (MPs) were detected and counted using human-expert operated microscopy following Giemsa staining of thick and thin blood films. The criterion for declaring a participant to be malaria parasite-free was no detectable parasites in 100 high-power (100x) fields in thick films. We validated the diagnosis outcome by randomly selecting one in ten thick blood films for independent review, by local external experienced senior malaria-microscopy technologists. Parasite density (PD), malaria parasites per microliter (MPs/μL), are calculated by dividing the number of observed MPs by the number of counted white blood cells (WBC), and then multiplied by 8×10^3.[15]

2.4 | Data acquisition and pre-processing

We captured images using an upright brightfield microscope (Olympus BX63) fitted with a 100X/1.4 NA objective lens (MPLAPON100XO), a motorized x-y sample positioning stage (Prior Scientific) and a color camera (Edge 5.5c, PCO) to capture images of Giemsa-stained, thick blood smears prepared in our clinics. For each sample we captured 100 non-overlapping FoV, each covering an area of 166 μm × 142 μm. Such large numerical aperture object lenses have limited depth of field. To capture the entire thickness of the blood film (typically ~5 μm) a z-stack of 14 focal planes with a separation of 0.5 μm was captured for each field. With a camera exposure time of 5 milliseconds the total acquisition time per sample was approximately 5 minutes (Figure S1 in Appendix S1). To reduce the data volume and render images into a form more amenable to annotation, after white balancing, z-stacks were projected onto a single plane using a wavelet-based Extended Depth of Field (EDoF) algorithm.[21] Briefly, each focal plane was decomposed using a 12 level "sym8" wavelet, and for each level and sub-band the coefficients with the maximum values were chosen among the 14 decomposed focal planes. Following a spatial and a sub-band consistency check, the inverse wavelet transform was applied to the selected coefficients.

2.5 | Parasite and white blood cell detection

We tested the use of deep learning-based object detection methods to identify both *P. falciparum* parasites and white-blood-cell (WBC) nuclei in the digitized EDoF thick blood films images. Current state of the art deep learning object detectors usually follow two stages: first a sparse set of region proposals that should contain all foreground objects are generated while excluding most of the background locations.[22] Next, these proposals are fed to a CNN providing each region with a class label probability and a refined bounding box.[23] Various extensions to this approach have been proposed.[24-26] In contrast, simpler and faster one-stage detectors[27] are applied over a regular, dense sampling of possible foreground object locations. Among these detectors, RetinaNet[28] exceeded the performance of previous two-stage approaches thanks to a focal loss function aimed to give more attention to difficult examples. We trained and tested three of these state-of-the-art object detectors: Faster R-CNN[25]; R-FCN[29]; and RetinaNet.[28] Our expert microscopists annotated a total number of 239 EDoF FoV containing 2986 MP and 1272 WBC nuclei (Figure S2 in Appendix S1). Two thirds of the annotated FoV were used to train the object detector models while the rest was used for evaluation (Figure S2 in Appendix S1). These image fields were obtained from 13 unique blood films. Geometrical transformations were applied "on the fly" during training to the image fields to augment the training dataset. At each iteration the image fields were rotated by a uniformly random angle between 0° and 270°. Additionally, the resulting rotated image would be randomly flipped vertically, horizontally or not at all. An example of MP and WBC RetinaNet detections in a full FoV is shown in Figure S5 in Appendix S1. The dataset is available at https://doi.org/10.5522/04/12173568 under open licence CC BY-NC-SA 4.0.

2.6 | Automated diagnosis with negative adjustment

In this previously suggested approach,[12,13,30] only the negative samples from the training set (Table 1) were used. The trained RetinaNet parasite detection model described in the previous section was applied to these samples. Next, the average number of false positives per 100 image fields $mean_{fp}$ and its SD std_{fp} were computed. Further on, a threshold θ computed using the $mean_{fp}$ and std_{fp} values was applied to the test samples for diagnosis.

$$D = \begin{cases} 1 & if \ N_{mpd} > \theta \\ 0 & otherwise \end{cases} \qquad (1)$$

TABLE 1 DeepMCNN vs NA expert-level diagnostic performance on validation set

Model	Sensitivity	Specificity	Accuracy	PPV	NPV
NA	0.66 [0.54-0.77]	0.95 [0.86-0.98]	0.79 [0.71-0.86]	0.94 [0.83-0.97]	0.70 [0.62-0.76]
DeepMCNN	0.92 [0.83–0.97]	0.90 [0.80-0.96]	0.91 [0.84-0.95]	0.92 [0.83-0.96]	0.90 [0.80-0.95]

Abbreviations: DeepMCNN, Deep Malaria Convolutional Neural Network; NA, negative adjustment method; NPV, negative predictive value; [95% CI]; PPV, positive predictive value.

$$\text{where} \begin{cases} N_{mpd} = \text{number of potential MP detected} \\ \theta = mean_{fp} + \alpha \cdot std_{fp} \end{cases}$$

with D the automated diagnostic (malaria positive or negative) and $\alpha \in$ [0,2] a sensitivity parameter. That is, a sample was classified positive if the number of parasites detected in 100 image fields was larger than the threshold θ (Figure S4 in Appendix S1).

2.7 | The deep malaria CNN classifier (DeepMCNN)

Here we propose a novel approach to leverage routine clinical-microscopy labels from our malaria diagnosis clinics. We trained the DeepMCNN (Figure 1 and Figure S3 in Appendix S1) classifier as follows. First, the RetinaNet was applied to each of the 100 FoVs obtained from each sample from the training set (Table S1 in Appendix S1). Second, stacks of detected MP regions extracted from each image field are then used together, with the human-expert clinical-microscopy diagnostic (malaria-positive or malaria-negative) label, to train the DeepMCNN classifier (Figure 1 and Figure S3 in Appendix S1).

In more detail, we obtained a variable number of potential malaria parasites (Np) from each sample 100 FoVs (Figure 1). These were then cropped from the FoV using a 64 × 64 pixel window corresponding to 4.2 × 4.2 μm which is large enough to encompass a malaria ring parasite (3 × 3 μm). A VGG-19 model[31] was trained to classify these stacks of potential parasite images as positive or negative. The weights of the convolutional layers were initialized with weights from a VGG-19 model pre-trained on the ImageNet dataset.[32] For each stack of variable images, the Np features vectors corresponding to the input of the fully connected layers are averaged into one single feature vector (Figure S3 in Appendix S1). This allows a variable number of potential parasite images as an input for the MCNN classifier.

Mathematically, the classification problem can be re-formulated as described in the next paragraph. In general, the outcome of a CNN classifier can be written as

$$D = argmax \begin{pmatrix} a \\ softmax(\mathcal{M}(I)) \end{pmatrix}$$

where D represents the output label and I is the image to be classified. A batch size of 1 is assumed for simplicity, that is, one input image
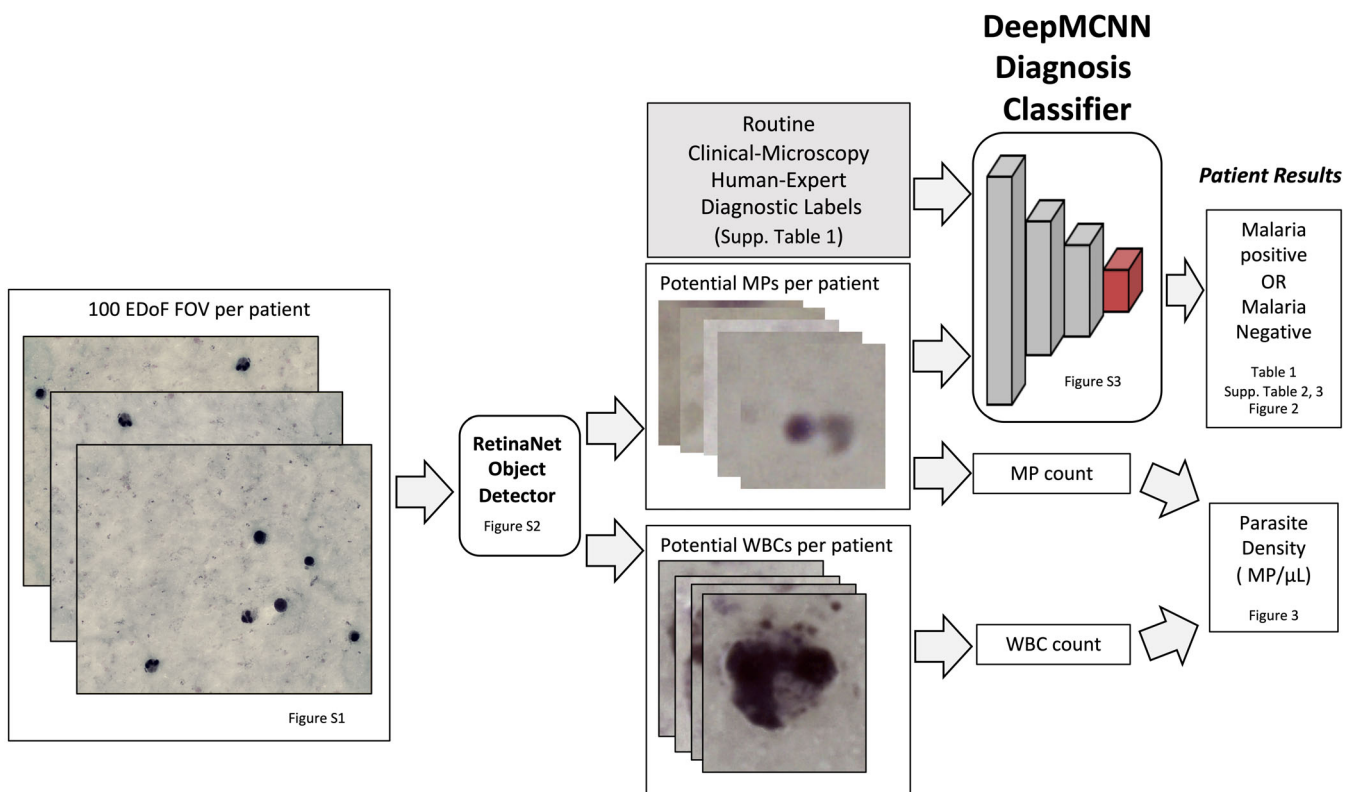

FIGURE 1 The Deep Malaria Convolutional Neural Network (DeepMCNN) diagnostic classifier approach architecture. DeepMCNN leverages routine clinical microscopy human-expert diagnostic labels to provide expert level malaria diagnosis of a thick blood film specimen. EDoF, extended depth of field; FOV, field of view; MP, malaria parasite; Parasitemia MP/μL, parasitemia in malaria parasites per microliter; WBC, white blood cell.

and one output label. So, $\mathcal{M}$ is the CNN transformation of the input image until the *softmax* layer and can be written as:

$$\mathcal{M}(I) = relu(\mathfrak{C}(I) \cdot W_1 + b_1) \cdot W_2 + b_2$$

With $\mathfrak{C}(I)$ the flattened output of the convolutional layers (CNN feature vector) and W_i and b_i the weights corresponding to the fully connected layers. To accommodate a variable number (N_p) of input images corresponding to one single label, $\mathcal{M}$ was modified as follows:

$$\mathcal{M}'(I_1, I_2, ..., I_{N_p}) = relu\left(\frac{1}{N_p} \sum_{i=1}^{N_p} \mathfrak{C}(I_i) \cdot W_1 + b_1\right) \cdot W_2 + b_2$$

This is equivalent to an average pooling of the feature vectors. Equation (2) becomes:

$$D = argmax(softmax(\mathcal{M}'(I_1, I_2, ..., I_{N_p}))) \quad (2)$$

Where D is the patient diagnostic (malaria positive or negative) after inspecting N_p potential parasites from 100 FoV. A gradient descent optimizer with a fixed learning rate of 0.0003 and a cross entropy loss function were chosen to optimize the CNN weights.

2.8 | Parasitemia estimation

Once a patient sample has been classified as positive, the patient parasitemia was estimated in the following manner: Let re_{mp} and pr_{mp}, re_{wbc} and pr_{wbc} be the recall (or sensitivity) and precision (or positive predictive value) of the object detector for the MP and the WBC respectively on the test image fields. The patient parasitemia pp (MP/μL) computation according to the WHO recommendation[15] was adjusted:

$$pp = 8000 \cdot \frac{mp_{detected} \cdot pr_{mp}}{re_{mp}} \cdot \frac{re_{wbc}}{wbc_{detected} \cdot pr_{wbc.}}$$

where $mp_{detected}$ and $wbc_{detected}$ represent the number of MP and WBC, respectively, detected by the object detector. This formula assumes on average 8000 WBC per μL. The predicted parasitemia was compared to the parasitemia computed with the human MP/WBC count.

3 | RESULTS

3.1 | Study participants, datasets and annotations

Training and validation data used for DeepMCNN automated patient diagnosis is described in Table S1 in Appendix S1. Each Thick Blood Film (TBF) corresponds to an individual with a total of 169 in the training set and 130 in the validation set. The training set is comprised of 84 malaria-positive and 85 malaria-negative TBFs, each with 100 EDoF fields of view (Table S1 in Appendix S1). The validation set contains 60 malaria-positive and 70 malaria-negative TBFs each with

100 EdoF fields of view (Table S1 in Appendix S1). Malaria-positive thick blood films have a range of parasitemia from 60 to 10^5 MP/μL. All the specimens have been collected and prepared at our quality-controlled malaria clinics and assessed by our expert microscopists and clinicians.

3.2 | Automated malaria diagnosis with the deepMCNN classifier

Malaria parasite detection in an individual FoV from a TBF only represents an intermediate step in achieving patient final malaria diagnosis. To achieve patient level diagnosis, we proposed and trained (see methods section) a novel Deep Malaria Convolutional Neural Network (DeepMCNN). It leverages routine clinical microscopy labels from our malaria diagnosis clinics, to achieve an automated final diagnosis by assessing 100 FoVs (Figure 1 and Figure S3 in Appendix S1).

The diagnosis performance of the DeepMCNN on the validation set is shown in Table 1 and Figure 2. We benchmarked our DeepMCNN automated diagnostic method to a previously proposed method referred to as Negative Adjustment[30] (NA), as described in the methods section and Figure S4 in Appendix S1. The NA detection threshold θ in Equation (1) was estimated at 177 MPs per 100 FoVs for a specificity on the train set (Table S1 in Appendix S1) of 0.9. Our DeepMCNN achieves a sensitivity of 0.92; a specificity of 0.90 and an accuracy of 0.91 on the validation set with PPV/NPV of 0.92/0.90, outperforming the NA approach (Table 1). The trained DeepMCNN outputs a higher sensitivity (0.92) than the NA approach (0.66) for a specificity, equal or higher than 0.9 (Table 1).

To explore the clinical utility of DeepMCNN we calculated PPV and NPV for malaria prevalence values ranging from zero to one, and compared to that of NA (Figure 2A). DeepMCNN NPV clearly outperforms the NA approach (Figure 2A red-line). Moreover, DeepMCNN PPV and NPV performance across these prevalence ranges makes it usable in a wide range of clinical settings.

To evaluate PPV/NPV performance in relation to our Ibadan holoendemic (all-year-round) setting, we calculated PPV and NPV using the actual mean monthly prevalence obtained from our large clinical settings, serving five million inhabitants of the city of Ibadan in the sub-Sahara (Figure 2B). The mean monthly prevalence data (Figure 2B dotted line) is calculated from our large database over a five-year period from 2014 to 2019, and therefore represents an accurate and current snapshot of the burden of malaria in our clinical settings. DeepMCNN clearly shows NPV of over 0.9 across all months which is clinically usable in our settings (Figure 2B red-line). On the contrary, the NA approach falls below 0.9 during the long Ibadan rainy season (Figure 2B red-line) which hinders its utility in sub-Saharan settings.

Looking closer at the classification of the positive samples (Table S2 in Appendix S1), the NA method[30] completely misses all the low parasite count samples (less than 160 MP/μL). However, our DeepMCNN classifies 0.75 of these as positives, for a diagnostic specificity ≥0.90 (Table 3). In medium (160 to 1600 MP/μL) and high
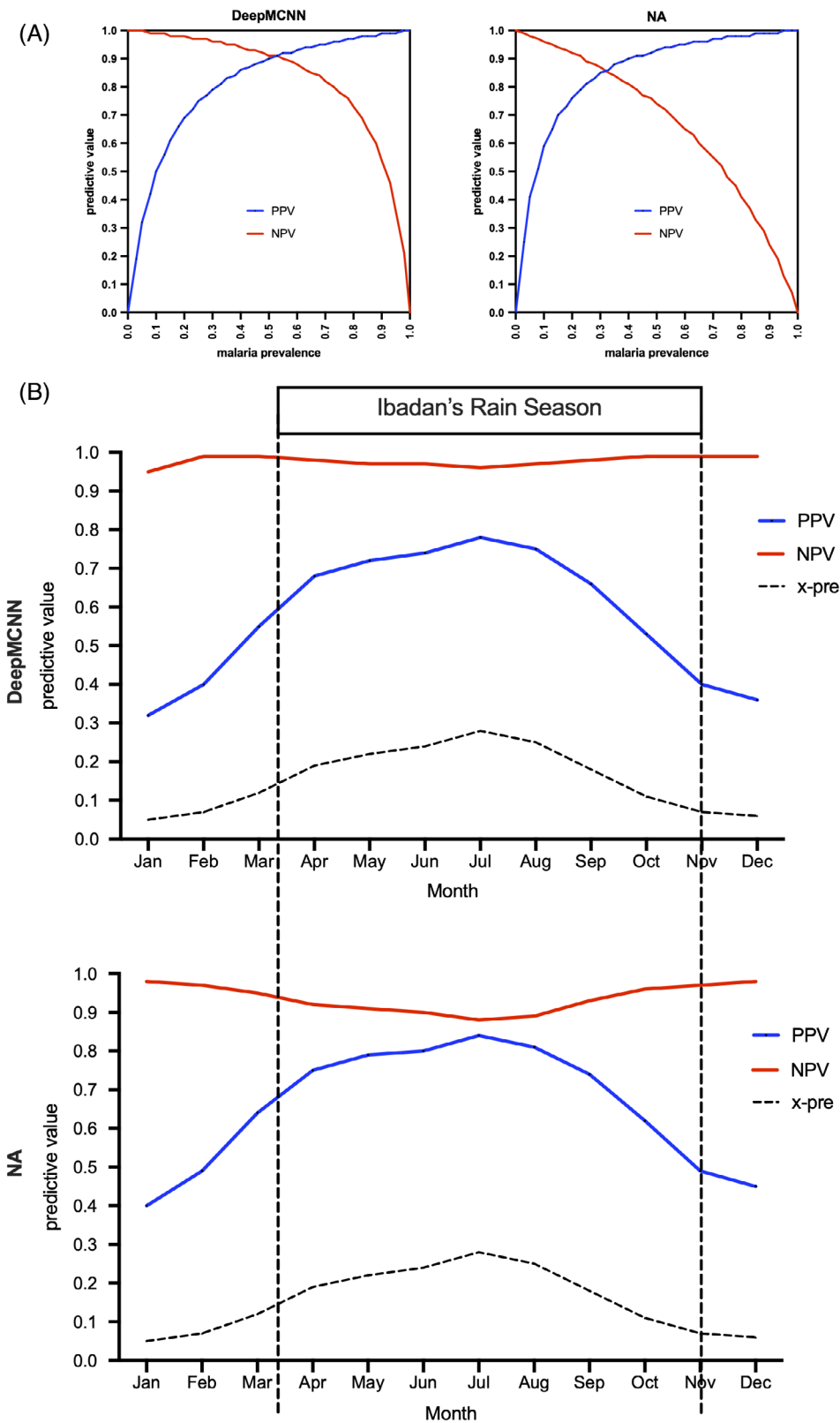
(A)

(B)

FIGURE 2 Deep Malaria Convolutional Neural Network validation performance. A, PPV and NPV performance for DeepMCNN vs NA method across a range of malaria prevalence values. B, PPV and NPV evaluation of DeepMCNN and NA methods in Ibadan's holoendemic malaria urban clinical settings. DeepMCNN, Deep Malaria Convolutional Neural Network; NA, negative adjustment method; NPV, negative predictive value (red line); PPV, positive predictive value (blue line); x-pre, actual Ibadan mean monthly prevalence (2014-2019)

(>1600 MP/μL) parasite densities DeepMCCN has sensitivity greater than 0.9 also clearly outperforming the NA method (Table S2 in Appendix S1).

Figure 3A shows that DeepMCCN automated patient diagnosis is achieved by assessing a median of well above 1000 WBC per 100 FoV, for both malaria positive and malaria negative specimens. For the vast majority of specimens more than 500 WBCs were assessed to achieve patient diagnosis and parasite counts (Figure 3A and Figure S3 in Appendix S1). This is twice the required WHO sampling protocol for the human expert microscopist. We
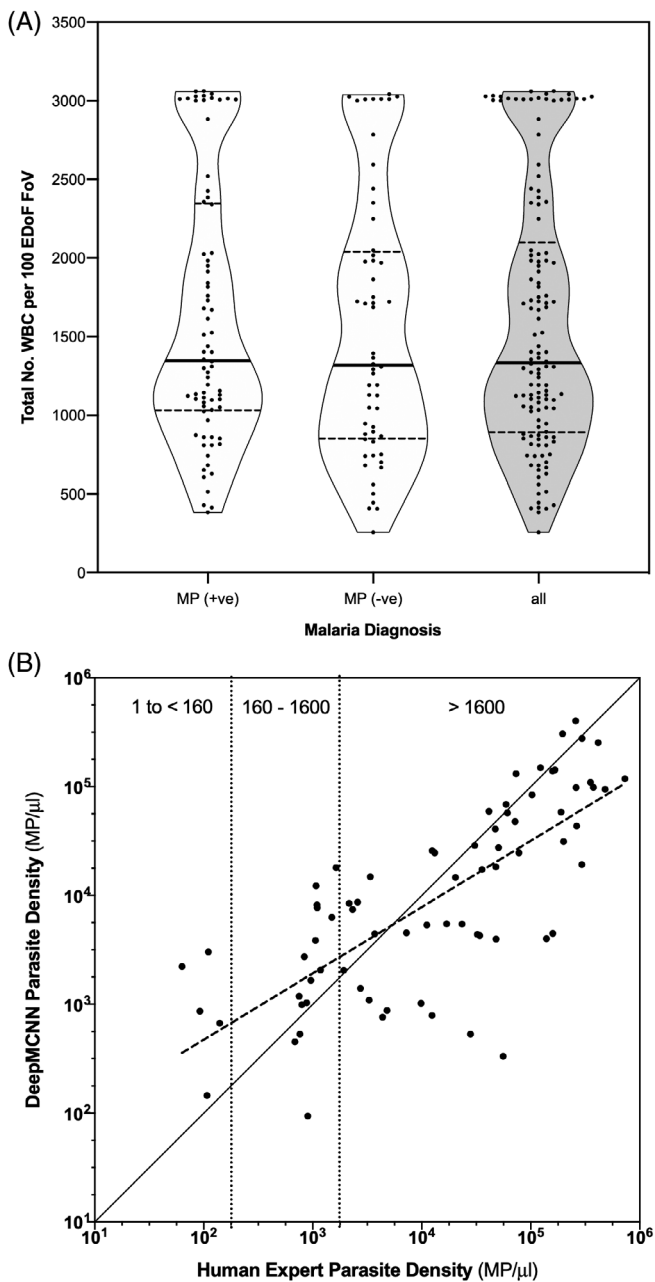
FIGURE 3 Deep MCNN WBC counts and malaria parasite density estimation. A,DeepMCNN total number of WBC assessed per sample 100 field of views. WBC, white blood cell; EDoF, extended depth of field; FoV, field of view; MP, malaria parasite; MP (+ve), malaria parasite positive; MP (−ve), malaria parasite negative. Violin-plot horizontal line = median; violin-plot horizontal dotted-lines = inter-quartile range. B, Scatterplot of estimates of parasite densities by human-expert vs DeepMCNN estimates. X-Y axes parasite densities in MP/μL, parasitemia in malaria parasites per microliter. $R^2 = 0.55$; 95% CI = [0.48-0.73]

then compared the estimated DeepMCNN parasitemia (see methods) against the computed parasitemia, using the manual count reported by the human-expert microscopist, across a range of low, mid and high parasitemia (Figure 3B). In patients with high parasitemia, our approach agrees closely with the human-expert

estimates. In those patients with low and mid parasitemia our approach overestimates the parasite densities (Figure 3B).

4 | DISCUSSION

Prompt, reliable and accurate malaria diagnosis is a challenge for healthcare providers servicing large urban metropoles in holoendemic malaria settings, such as the one presented in this work. Leveraging both the well-established malaria diagnosis gold-standard, and deep-learning image processing approaches, could provide automated scalable solutions amenable to be deployed in these clinical settings. However, the bottleneck of every deep learning-based approach is the lack of sufficient annotations. Obtaining a large number of accurate object-level human-expert annotations of malaria parasites is extremely time consuming and immensely laborious. To overcome these challenges and to create a deployable clinically usable automated diagnosis system, here we show that routine clinical-microscopy human-expert diagnostic labels could be leveraged to train a Deep Convolutional Neural Network. It achieves NPV and PPV performance suitable for clinical services within Ibadan, a densely populated metropolis located in west sub-Sahara Nigeria, where malaria is prevalent all year.

While some approaches attempt to validate automated detection of malaria using the polymerase chain reaction (PCR) as reference,[33] we evaluated our approach against human-expert Thick Blood Film microscopy, since it remains the internationally accepted and realizable gold standard in sub-Saharan regions. Similarly, our diagnosis and parasitemia estimation follows well accepted WHO protocols in the region. In large urban holoendemic settings such as ours, healthcare providers often lack the capacity to carry out TBF microscopy every six hours, once malaria treatment has commenced, which is required in severe malaria clinical pathways. Our automated approach could further facilitate the healthcare provider to process follow-up TFB to support these clinical pathways.

Overall, our DeepMCNN approach provides better accuracy in terms of diagnosing samples as malaria positive or negative compared to the NA approach.[30] The NA approach classifies a sample as malaria positive, if the number of detected parasites exceeds a specific threshold determined empirically on a hold-out set of negative samples, so that the specificity on that set exceeds 0.9. Our experiments show that the NA method misdiagnoses samples with low to mid parasite densities where the number of overall detections in 100 image fields is below the decision threshold. In contrast, our DeepMCNN approach does not have this limitation as it does not rely on a decision threshold.

From the clinical point of view, it is generally accepted that in any child with fever, malaria diagnosis is so important that a false-positive is better than a false-negative. Our DeepMCNN achieves NPV consistently greater than 0.9 across all months in the Ibadan settings rendering the system well-suited to provide pediatric clinical pathway support. This is reinforced by the DeepMCNN PPV performance observed during Ibadan's lengthy rainy season. Furthermore, its

performance at low parasitemia levels of less than 160 MP/μL is well suited to handle the adult population in high-transmission West sub-Saharan regions, whicht are more likely to have low to asymptomatic parasitemia.

Our DeepMCNN system provides a WBC count which, together with the MP count estimation, is used to determine a patient's diagnosis and parasitemia according to the WHO recommendations. With a median of more than 1000 WBCs observed, our approach is well above the recommended 500 WBCs required in low-parasitemia specimens.

Parasite densities estimates produced by both, human-expert and deep-learning system have their own drawbacks. The human expert is subject to cognitive load when counting objects over a large number of FoV, while the automated approach is limited by the ability of the object detector to discard staining artifacts. Taking this into account, our method overestimates parasite density, when compared to the human-expert, in low and mid parasitemia specimens. However, the human expert is prone to fatigue and as a consequence their counting accuracy might fluctuate over time.[7] In contrast, our method consistently uses the parasite detection accuracy to adjust the parasite density estimate. We are of the opinion that this leads to a more robust estimation of the patient's parasite density.

Patient level human-expert diagnostic labels routinely produced by our malaria clinical microscopy services are exponentially easier to obtain than object level labels from digitized blood films. Our study shows that our strategy does deliver a deep-learning system, that is capable of handling the burden of malaria disease observed in a large *Plasmodium falciparum* holoendemic setting.

Our open data and easily deployable DeepMCNN provide a clinically relevant platform, where other healthcare providers could harness their readily available patient level diagnostic labels, to tailor and further improve the accuracy of the DeepMCNN classifier for their clinical pathway settings. In turn, this should increase their quality, allowing them to process large number of blood films as required in large urban holoendemic malaria sub-Saharan settings. Further investment in research and development is needed to make advances in deep learning assisted diagnosis accessible in peri-urban and rural settings.

CONFLICT OF INTEREST
The authors have declared that no competing financial interests exist.

AUTHOR CONTRIBUTIONS
D.F.-R., B.J.B., O.S., J.S.-T., M.A.S., M.J.S., V.P., I.K. designed the study. D.F.-R., B.J.B., O.S., I.L., G.O., P.M., M.J.S., V.P. carried out the study. B.J.B., I.L., D.F.-R., O.S. carried patient recruitment at the CMRG, Ibadan, Nigeria. D.F.-R., B.J.B., G.O., M.E. collected and processed clinical specimens. D.F.-R., L.N.-Z., M.J.S., M.E., P.M., R.C., C.B., A.O., B.F.O., T.C., D.T. microscope data capture, storage and processing of image datasets, clinical data curation, labeling and curation of image datasets. P.M., I.K., D.F.-R. designed deep-learning diagnostic approach and carried out computational work. P.M., D.F.-R. analyzed the data. P.M., D.F.-R. wrote the manuscript. D.F.-R. is project lead and senior corresponding author.

ORCID
Delmiro Fernandez-Reyes ![ORCID] https://orcid.org/0000-0001-5070-9198

REFERENCES
1. World Health Organization. *World Malaria Report 2018*. Geneva: WHO; 2018.
2. Siahaan L. Laboratory diagnostics of malaria. *IOP Conf Ser Earth Environ Sci*. 2018. https://doi.org/10.1088/1755-1315/125/1/012090
3. Amir A, Cheong FW, De Silva JR, Lau YL. Diagnostic tools in childhood malaria. *Parasit Vectors*. 2018;11:53. https://doi.org/10.1186/s13071-018-2617-y
4. Pillay E, Khodaiji S, Bezuidenhout BC, Litshie M, Coetzer TL. Evaluation of automated malaria diagnosis using the Sysmex XN-30 analyser in a clinical setting. *Malar J*. 2019;18:15. https://doi.org/10.1186/s12936-019-2655-8
5. Eshel Y, Houri-Yafin A, Benkuzari H, et al. Evaluation of the parasight platform for malaria diagnosis. *J Clin Microbiol*. 2017;55:768-775. https://doi.org/10.1128/JCM.02155-16
6. Falade CO, Ajayi IO, Nsungwa-Sabiiti J, et al. Malaria rapid diagnostic tests and malaria microscopy for guiding malaria treatment of uncomplicated fevers in Nigeria and prereferral cases in 3 African countries. *Clin Infect Dis*. 2016;63:S290-S297. https://doi.org/10.1093/cid/ciw628
7. Bowers KM, Bell D, Chiodini PL, et al. Inter-rater reliability of malaria parasite counts and comparison of methods. *Malar J*. 2009;8:267. https://doi.org/10.1186/1475-2875-8-267
8. Das DK, Mukherjee R, Chakraborty C. Computational microscopic imaging for malaria parasite detection: A systematic review. *J Microsc*. 2015;260:1-19. https://doi.org/10.1111/jmi.12270

9. Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G. Image analysis and machine learning for detecting malaria. *Transl Res*. 2018;194:36-55. https://doi.org/10.1016/j.trsl.2017.12.004

10. Arco JE, Górriz JM, Ramírez J, Álvarez I, Puntonet CG. Digital image analysis for automatic enumeration of malaria parasites using morphological operations. *Expert Syst Appl*. 2015;42:3041-3047. https://doi.org/10.1016/j.eswa.2014.11.037

11. Rosado L, Da Costa JMC, Elias D, Cardoso JS. Automated detection of malaria parasites on thick blood smears via mobile devices. *Procedia Comput Sci*. 2016;90:138-144. https://doi.org/10.1016/j.procs.2016.07.024

12. Mehanian C, Jaiswal M, Delahunt C, et al. Computer-automated malaria diagnosis and quantitation using convolutional neural networks. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*; Venice, Italy; 2017:116-125. https://doi.org/10.1109/ICCVW.2017.22

13. Torres K, Bachman CM, Delahunt CB, et al. Automated microscopy for routine malaria diagnosis: a field comparison on Giemsa-stained blood films in Peru. *Malar J*. 2018;17:339. https://doi.org/10.1186/s12936-018-2493-0

14. Yang F, Poostchi M, Yu H, et al. Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE J Biomed Health Inform*. 2019;1-1. https://doi.org/10.1109/jbhi.2019.2939121

15. WHO. *Basic Malaria Microscopy: Learner's Guide*. 2010.

16. Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. *IEEE Trans Neural Netw Learn Syst*. 2019;30(11):3212-3232. https://doi.org/10.1109/TNNLS.2018.2876865.

17. Hung J, Ravel D, Lopes SCP et al. Applying Faster R-CNN for object detection on malaria images. 2018. http://arxiv.org/abs/1804.09548. Accessed April 25, 2020.

18. Burté F, Brown BJJ, Orimadegun AEE, et al. Severe childhood malaria syndromes defined by plasma proteome profiles. *PLoS One*. 2012;7(12):e49778. https://doi.org/10.1371/journal.pone.0049778

19. Burte F, Brown BJ, Orimadegun AE, et al. Circulatory hepcidin is associated with the anti-inflammatory response but not with iron or anemic status in childhood malaria. *Blood*. 2013;121(15):3016-3022. https://doi.org/10.1182/blood-2012-10-461418

20. Abah SE, Burté F, Marquet S, et al. Low plasma haptoglobin is a risk factor for life-threatening childhood severe malarial anemia and not an exclusive consequence of hemolysis. *Sci Rep*. 2018;8:17527. https://doi.org/10.1038/s41598-018-35944-w

21. Forster B, Van De Ville D, Berent J, Sage D, Unser M. Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images. *Microsc Res Tech*. 2004;65:33-42. https://doi.org/10.1002/jemt.20092

22. Uijlings JRR, Van De Sande KEA, Gevers T, Smeulders AWM. Selective search for object recognition. *Int J Comput Vis*. 2013;104(2):154-171. https://doi.org/10.1007/s11263-013-0620-5

23. Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH*; 2014:580-587. https://doi.org/10.1109/CVPR.2014.81

24. Girshick R. Fast R-CNN. *Proc IEEE Int Conf Comput Vis*. 2015;2015 Inter:1440-1448. https://doi.org/10.1109/ICCV.2015.169

25. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(6):1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

26. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell*. 2018;42:386-397. https://doi.org/10.1109/TPAMI.2018.2844175

27. Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. *Lect Notes Comput Sci*. 2016;9905 LNCS:21-37. https://doi.org/10.1007/978-3-319-46448-0_2

28. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. *2017 Proceedings of the IEEE International Conference on Computer Vision*. Institute of Electrical and Electronics Engineers Inc.; 2017:2999-3007. https://doi.org/10.1109/ICCV.2017.324

29. Dai J, Li Y, He K, Sun J. R-FCN: object detection via region-based fully convolutional networks. *Advances in Neural Information Processing Systems*. Neural Information Processing Systems Foundation; 2016:379-387.

30. Delahunt CB, Mehanian C, Hu L, et al. Automated microscopy and machine learning for expert-level malaria field diagnosis. In: Proceedings of the fifth IEEE Global Humanitarian Technology Conference, GHTC 2015:393-399. https://doi.org/10.1109/GHTC.2015.7344002

31. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Third International Conference on Learning Representations ICLR 2015 – Conference Track Proceedings; 2015:1-14.

32. Deng J, Dong W, Socher R, Li L-J, Kai L, Li F-F. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2010:248-255. https://doi.org/10.1109/cvpr.2009.5206848

33. Dave IR. Image analysis for malaria parasite detection from microscopic images of thick blood smear. In: Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2017:1303-1307. https://doi.org/10.1109/WiSPNET.2017.8299974

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.