IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

Malaria Parasite using CNN

Ramesh Bhajantri, Harin Gowda, Mohammed Tanzeel, Sidramesh Hiremath, Shruthi Shetty J
Students, BE (Appearing), Department of Computer Science and Engineering[1,2,3,4]
Assistant Professor, Department of Computer Science and Engineering[5]
Alva's Institute of Engineering and Technology, Mijar, Moodbidire, India

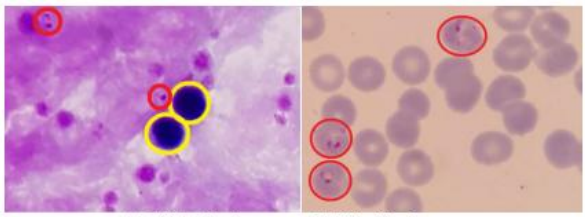
Abstract: *This study looks into the viability of using CNN to automatically detect malaria parasites in thick blood smears. Techniques: We have created the first deep learning technique that works on cellphones and can identify malaria parasites in thick blood smear photos. Our process entails two steps of processing. First, we use a quick screening method called an intensity-based Iterative Global Minimum Screening (IGMS).using a thick smear picture to identify potential parasites. Then, each candidate is categorised as either parasite or background using a customised Convolutional Neural Network (CNN). With this study, we create a dataset that is 1819 thick.150 patients' smear photos made available to the public scientific community. This dataset served as our training and evaluate the deep learning approach we used in this paper. Results: A patient-level five-fold cross-evaluation shows the customised CNN model's effectiveness in differentiating between positive (parasitic) and negative image patches in terms of the performance indicators of accuracy (93.46 percent 0.32 percent), AUC (98.39 percent 0.18 percent), sensitivity (92.59 percent 1.27 percent), specificity (94.33 percent 1.25 percent), precision (94.25 percent 1.13 percent), and negative predictability. High correlation coefficients (>0.98) between ground truth and automatically discovered parasites, at both the patient and picture level, show the effectiveness of our technology. Conclusion: Positive outcomes deep learning techniques, results are produced for parasite detection in thick blood smears. Meaning: Automated parasite detection operating on smartphones offers a promising substitute for manually counting parasites to diagnose malaria.*

Keywords: Deep learning, convolutional neural networks, computer-aided diagnosis, malaria.

I. INTRODUCTION

Malaria is a fatal illness that can affect anyone. The World Health Organization (WHO) malaria report for 2018 states that 219 million cases of malaria were discovered.causes about 435,000 deaths globally in 2017. study of coloured thick and thin blood smears under a microscope is the most reliable way to diagnose malaria [2, 3]. Microscopy Although inexpensive and generally accessible, examination takes time. Additionally, the proficiency of the parasitologists affects how well the microscope diagnostic works [4]. Being very It is typical for parasitologists to work in resource-constrained settings without a strict framework in place to assure the upkeep of their abilities or/and diagnostic capability This results in inaccurate diagnostic findings and hence ineffective treatment [4]. False negative diagnoses, on the other hand, result in the use of antibiotics, a second consultation, and the potential progression of more severe malaria [5]. As an illustration, false positive diagnostic results result in the unnecessary use of anti-malaria drugs and the experience of their side effects, such as nausea and abdominal pain. Therefore, the creation of an automated system for diagnosing malaria is a desirable research objective for enhancing personalised patient care and management. Automated parasite identification offers two major benefits: 1) it can offer a more accurate diagnosis, particularly in locations with limited resources; and 2) it lowers diagnostic expenses. To diagnose malaria and determine the severity of the condition, parasite counts are necessary. This causes inaccurate diagnostic findings, which therefore result in ineffective treatment [4]. For instance, false positive diagnostic results result in the unnecessary use of anti-malaria medications and the resulting side effects, such as nausea and abdominal pain, whereas false negative results result in the unnecessary use of antibiotics, a second consultation, and the progression of potentially more severe malaria [5]. Therefore, one interesting research objective for enhancing customised patient therapy and management is the creation of an automated system for diagnosing malaria. Automated parasite identification has two significant advantages: 1) it can offer a more accurate diagnosis, particularly in resource-limited locations; and 2) it lowers diagnostic costs. Counting parasites is crucial for determining the severity of the disease and for diagnosing malaria.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252



(a) Thick blood smear (b) Thin blood smear

Examples of thick and thin blood smears. Red circles are parasites and yellow circles are white blood cells. species and stages of parasite growth. As seen in Fig. 1, thick and thin blood smears require various processing techniques for parasite detection. White blood cells (WBCs) and red blood cells (RBCs) can both be seen clearly in thin blood smears. RBCs are typically segmented and then classified as infected or uninfected as part of the automatic parasite detection process in thin smears [5]–[7]. In thick blood smears, however, only WBCs and the nuclei of RBCs are visible (see Fig. 1(a)).Therefore, parasites need to be detected directly, and a typical step is to first preselect parasite candidates and then classify the candidates as either actual parasites or background noise.

II. RELATED WORK

In recent years, a number of methods for image processing and analysis on both thin and thick blood smears have been put forth with the goal of automating the detection of parasites. You can find reviews of the published literature in [5], [8], and [9]. We give a quick summary of the methods for detecting malaria in thick blood smears in the paragraph that follows. Conventional parasite detection methods are frequently used.

Based on thresholding and morphological processes with segmentation [10]–[13]. Kaewkamnerd and colleagues [10] suggest a approach employing the V-value histogram and an adaptive threshold the HSV picture to identify potential parasites and white separating white blood cells (WBCs) from the background WBC parasites are classified according on size. Analysis of 20 photos demonstrate the proposed technique's accuracy.sixty percent. Utilizing an intensity-stretching technique is Hanif et al.255 thick blood smears were enhanced for contrast, and then To divide up malaria parasites, utilise an empirical threshold. The authors display qualitative findings on various photos, whereTo achieve satisfactory segmentation results, various empirical criteria are utilised. To find parasites in thick blood smears, Chakrabortya et al. [12] combine morphological segmentation with colour information. Patch level evaluation of 75 photos used in the experiments reveals a successful detection rate of 95% and a false positive ratio of 10%. When working with denoised images, Dave et al. [13] apply histogram-based adaptive thresholding and morphological operations to find RBCs harbouring malaria parasites in both thin and thick blood smears.On 87 photos, patch level analysis reveals that the technique detects 533 parasites as opposed to the 484 parasites marked as ground truth. Traditional methods for parasite detection are quick and easy to use, but they are challenging to scale to huge datasets. . This is because parameters are frequently established empirically and traditional approaches are particularly sensitive to image fluctuations. When evaluating on large datasets, on the picture level, or at the patient level, performance evaluation on patch level on small datasets (from 20 to 300 images) can alter significantly.

Feature-based approaches use machine learning-based feature extraction and classification. [14]–[18]. Elter et almethod .'s [14] of extracting 174 features from pre-detected plasmodia candidates and using an SVM classifier to identify the parasite. According to the authors, the sensitivity for 256 images at the patch level is 97 percent. Purnama and others Identify features in the RGB channel, H channel in HSV space, and H channel in HIS space histograms,then determine the sort of parasite through genetic programming.and setting. On 180 patches, their classification model achieves a parasite detection accuracy of 95.58 percent on average, and 95.49 percent were able to identify no parasites. From the pre-segmented image, Yunda et al. [16] extract colour characteristics, co-occurrence texture features, and wavelet-based texture features.minimise superfluous data using Principal Component Analysis (PCA)features, followed by a classification model based on a neural network. 110 photos evaluated indicate that the sensitivityfor finding parasites is 76.45 percent. According to Quinn et al. [17], each picture should first be divided into 475 randomly overlapping patches using downsampling and sliding window screening. The patches should then be used to extract linked component and moment features before being classified using a randomised tree classifier.The technique yields a precision of 90%

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

at a recall of 20% on patch level when tested on 2903 pictures from 133 patients.Rosado et al[18] .'s adaptive thresholding method is used to detect parasites, and for the identification of WBC and parasites, they combine geometry, colour, and texture characteristics with an SVM classifier that is based on an RBF kernel. analysis of 94 pictures from 6 patients demonstrates their automatic parasite prediction has attained a 91.8 percent patch level accuracy and a 80.5 percent sensitivity and 93.5 percent specificity, while their WBC 72.1 percent specificity and 98.2 percent sensitivity are achieved during detection. The feature-based methods gauge their effectiveness on patches level. That is, the input sample is a single patch image and the Evaluation typically involves the accuracy of a classification patch. However, the ultimate objective for diagnosing malaria patients is to find and categorise all patches, including parasitic and false-positive ones, for a patient. Having successful categorization performance at the patch level not always guarantee strong performance on an aesthetic level or tolerant level Due to its greater performance on massive data, deep learning is the newest trend in machine learning. It already improved performance in a number of non-medical fields. lately, deep Learning is becoming more and more popular in computer-assisted diagnostic apparatus. There were two key causes of this development: 1) Deep learning requires neither segmentation nor feature-based approaches, in contrast to conventional techniques.

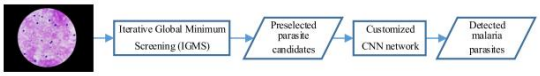


Fig. 2. Pipeline of the proposed system for automated parasite detection.

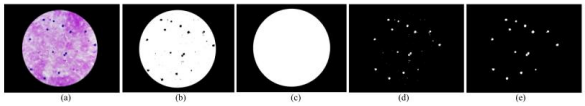


Fig. 3. Example of WBC detection. (a) A sample slide image of a thick blood smear acquired with a smartphone. (b) Detected objects after using Otsu thresholding. (c) Detected field of view ROI mask. (d) Detected WBCs including small areas of noise. (e) Detected WBCs after filtering noise in (d).

2.1 Contributions

We add the following to the processing of thick blood smears in comparison to previous work: First, using our proposed intensity-based Iterative Global Minimum Screening (IGMS) method for quick automatic preselection of parasite candidates and our specialised CNN model [19]-[21], [26]-[28] for classification of parasite candidates as either parasites or background, we create a smartphone system for automated parasite detection in thick blood smears. . To the best of our knowledge, this is the first deep learning-based effort created for parasite detection in cellphones for thick blood smears. Furthermore, our system is quick. On a typical Android smartphone, it takes roughly 10 seconds to find parasites in a 3024 x 4032 image. Third, we evaluate our methodology using a considerably larger dataset of photos from 150 patients, which includes 1819 thick smear images and 84,961 tagged parasites. This dataset is made available to the public along with this research. The remainder of the paper is organised as follows: The specifics of our suggested method for automatically detecting parasites are presented in Section II. The dataset, the experimental setting, and the findings are introduced in Section III. We present our findings in Section IV, followed by a conclusion in Section V.

III. METHODS

Our problem can be processed more quickly by being divided into a screening and classification step since we only need to make predictions on a limited number of pixel patches, which lowers the overall processing cost. In Fig. 2, we depict the method's pipeline.

3.1 Parasite Candidate Screening

A subset of the most likely parasite candidates is preselected during the screening stage, which also minimises the size of the initial search area. Utilizing the fact that the nuclei of parasites and WBCs have darker intensities than the background, parasite candidates are chosen based on the lowest grayscale intensities using a histogram analysis (Fig. 3(a)). We filter out WBCs before doing the parasite candidate screening to remove WBC distraction. Therefore, WBC

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

detection and parasite candidate generation comprise our intensity-based screening method for parasite candidate preselection. . All WBCs present in the image are first filtered by the WBC detection. The parasite candidate generation then creates areas of interest by focusing on the areas with the lowest intensities within a dense blood smear image.

1) Detection of WBC: In Fig. 3, an example smear image is displayed (a). The RGB image is first transformed into a grayscale image. Then, we apply Otsu's technique [29] to transform the grayscale image into a binary mask M1. The broad ROI area corresponding to the field of view is shown as the foreground (white) in this binary mask M1, whereas WBCs are shown as the background (black); see Fig. 3. (b). The big field of view ROI area's holes are filled to create the field of view mask M2, as shown in Fig. 3. (c).The binary mask M1 and ROI mask M2 can then be subtracted to separate the WBCs (see Fig. 3(d)).The final step is to filter out little noisy areas to obtain clean WBCs.

3.2 Parasite Preselection Using Iterative Global Minimum Screening (IGMS)

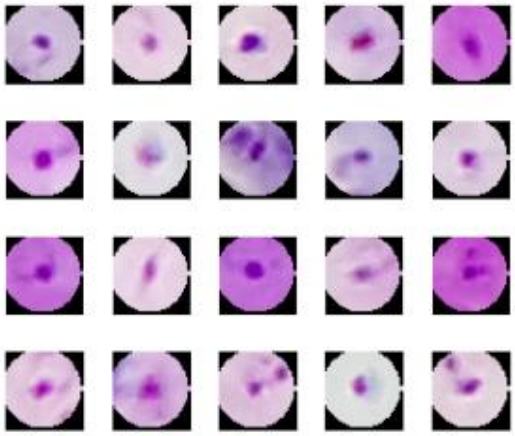
IGMS generates RGB parasite candidates by localizing the minimum intensity values in a grayscale image. If only one pixel is localized, a circular region centered at this pixel location with a pre-defined radius of 22 pixels (average parasite radius) is cropped from the original RGB image and is selected as a parasite candidate (Fig. 5(a)). If more than one pixel is localized, a new parasite candidate centered at the ith pixel is added when all the distances between the ith pixel and previously selected pixels are larger than 22. Once a parasite candidate is selected, the intensity values inside this region of the grayscale image will be replace by zeros to guarantee the convergence of the IGMS method. The screening stage stops when the number of parasite candidates reaches a given number. In our experiments, we select 500 parasite candidates for each image to cover the true parasites as much as possible. Experiments on our dataset of 150 patients show that we can achieve a sensitivity above 97% on patch level, image level, and patient level when using this number. Each parasite candidate is a $44 \times 44 \times 3$ RGB patch image, with pixels having a distance greater than 22 to the center set to zero. Fig. 4 shows the processing flowchart for IGMS and Fig. 5 shows examples of positive and negative patches extracted by IGMS.

3.3 Parasite Classification

Once the parasite candidates are extracted, we use a CNN model to classify them either as true parasites or background. In this work, we customize a CNN model consisting of seven convolutional layers, three max-pooling layers, three fully connected layers, and a softmax layer as shown in Fig. 6. A batch normalization layer is used after every convolution layer to allow a higher learning rate and to be less sensitive to the initialization parameters [29], followed by a rectified linear unit (ReLU) as the activation function [19]. Max-pooling layers are introduced after every two successive convolutional layers to select feature subsets. The last convolutional feature map is connected to three fully connected layers with 512, 50, and 2 hidden units, respectively. Between the three fully connected layers, two dropout layers [30] with a dropout ratio of 0.5 are applied to reduce model overfitting. The network is derived from VGG19 [27] by selecting the first six convolutional layers and three corresponding max-pooling layers from the VGG19 architecture to stop the feature maps at $64@5 \times 5$, followed directly by the fully connected and dropout layers. This shorter network structure provides similar performance while being faster and requiring less memory, which is important for smartphone applications. The output of the CNN model is a score vector, which gives the probabilities of the input image patch being either a parasite or background. We can obtain a larger or smaller number of predicted parasites by applying an adaptive probability threshold to the score vector. Compared with pre-trained networks such as VGG [27], GoogLeNet [28], ResNet-50 [26], our customized CNN model has several advantages: 1) runtime is reduced by using a smaller set of customizable parameters, with the input size of the model being determined by the average parasite size in thick smear images ($44 \times 44 \times 3$), which is much smaller than the input size used by the other networks ($224 \times 224 \times 3$); 2) our smaller network structure with fewer layers is more suitable for smartphones. Since the input size is smaller, our network should in fact be less deep to avoid feature maps that are too small. A smaller network structure with less parameters also avoids over-training on the smaller input space. Compared to the pretrained networks mentioned above, our customized CNN model achieves a better accuracy, despite having less network layers, and a shorter runtime. For an input image of $4032 \times 3024 \times 3$ pixels, our system can complete the parasite detection within ten seconds (about eight seconds for candidate screening and two seconds for classification) on a standard Android smartphone. Both the smaller set of

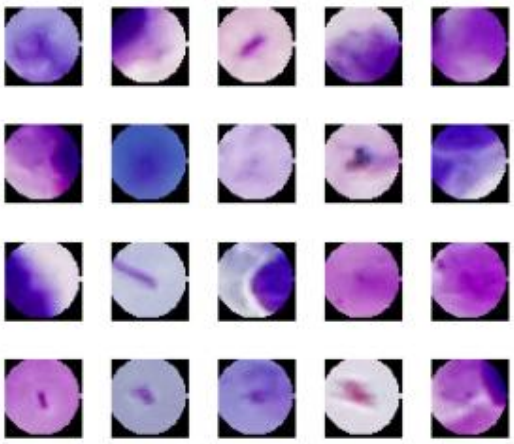
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

parameters and the smaller network structure contribute to the reduced runtime.considerable risk of error and data loss in the current system. Because humans undertake the majority of the job, the workload is higher. Students' safety does not meet the standards of today's modern environment. System is more time consuming both for checking and updating. In the current Leave Management System once a data entered incorrectly by human can end in overall confusion in the data. Maintaining data in the present Leave Management system necessitates physical space. In the present Leave Management System, the same data is written down and replicated multiple times at various levels of the system. The present Leave Management System's data response time is excessive. . In the existing Leave Management System, wardens are not accessible to respond 24 hours a day, seven days a week since they may have other responsibilities, such as lunch, or they may only work part-time. It may take time for a new warden to adjust to the current Leave Management System, and training may be required in some circumstances. During a natural disaster, the existing Leave Management System may fail, causing data to be lost.



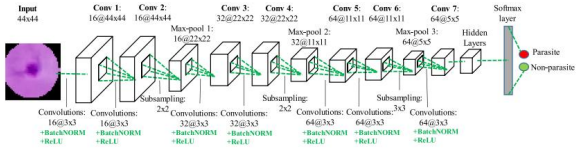
(a)

Parasite candidates generated by the IGMS method. (a) 20 positive patches.



(b)

Parasite candidates generated by the IGMS method. (a) 20 negative patches.



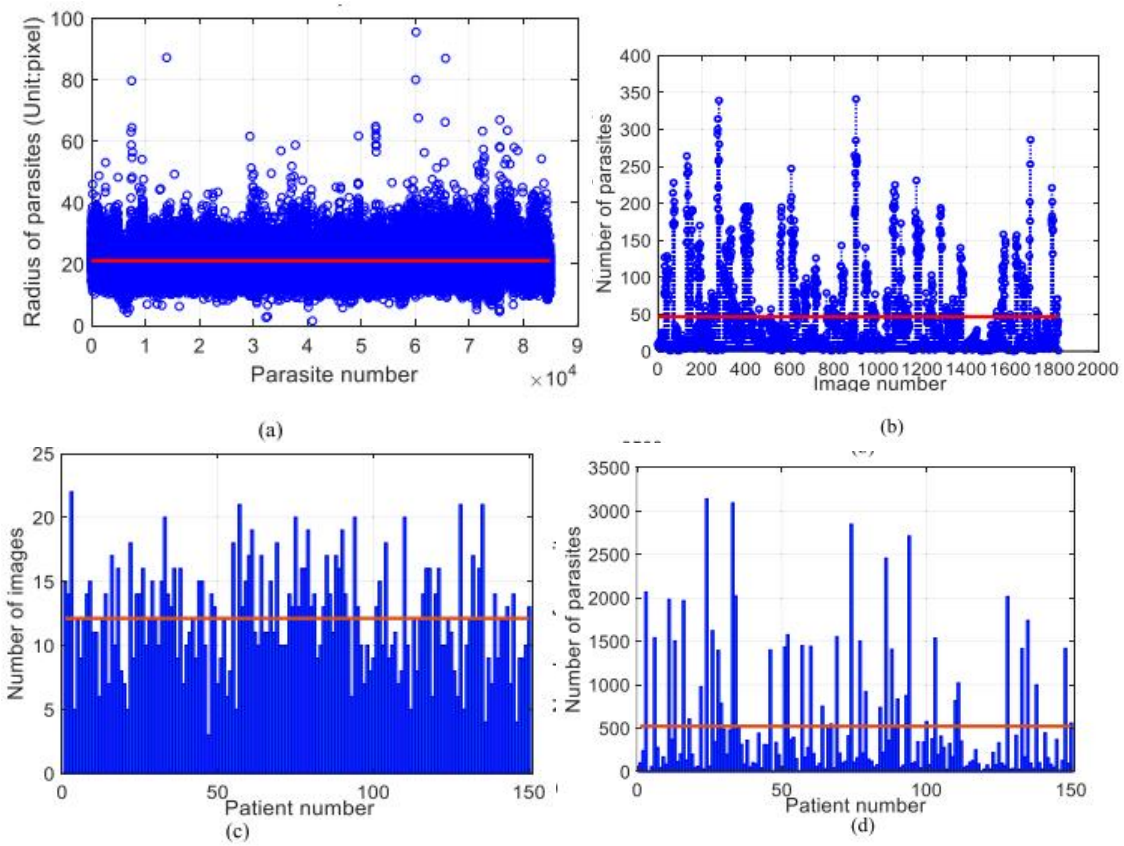
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

IV. DATA PREPARATION AND EXPERIMENTAL RESULTS

4.1 Dataset

We photographed Giemsa-stained thick blood smear slides from 150 P. falciparum infected patients at Chittagong Medical College Hospital, Bangladesh, using a smartphone camera for the different microscopic field of views. Fig. 7 shows the smartphone-microscope setup and a screenshot of the phone displaying a thick smear image. Images are captured with 100x magnification in RGB color space with a 3024 × 4032 pixel resolution. An expert slide reader manually annotated each image at the Mahidol-Oxford Tropical Medicine Research Unit (MORU), Bangkok, Thailand. We de-identified all images and their annotations, and archived them at the National Library of Medicine (IRB#12972). In this work, we use 1819 thick blood smear images from these 150 patients

4.2 Statistics of the Dataset



Statistical analysis of our dataset (including both Set A and Set B). The distribution of radii size of 84,961 parasites is plotted in (a), and the distribution of the number of parasites in 1819 images is illustrated in (b). The number of images and parasites in each patient is illustrated in (c) and (d) respectively. The red lines in the four subfigures indicate the average parasite radius, the average number of parasites in each image, the average number of images for each patient, and the average number of parasites for each patient, respectively.

4.3 Performance of the Customized CNN model

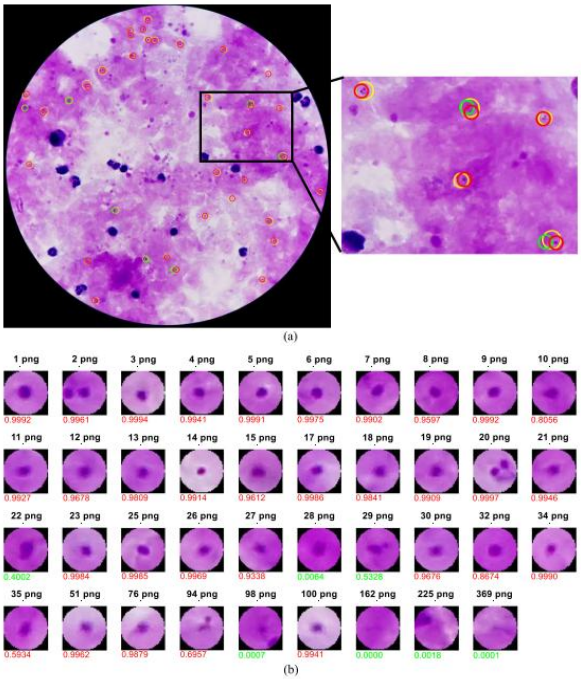
We evaluate the performance of the customized CNN model on Set A using five-fold cross evaluation. Each fold contains 24 patients. Table II and Fig. 11 present the classification performance and receiver operating characteristic (ROC). According to Fig. 11, our customized CNN model achieves an average AUC score of 98.39%, and a standard deviation of 0.18%, showing its robustness and effectiveness. The average accuracy, F-score, specificity, sensitivity, precision, and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

negative predictive values for our customized CNN model are 93.46%, 93.40%, 94.33%, 92.59%, 94.25%, and 92.74%, respectively.

V. DISCUSSION

In this work, we develop a smartphone-supported parasite detection application based on our IGMS method and deep learning. In Section III, we show that our application achieves an accuracy of 97.26% and an AUC of 97.34% on patch level, while obtaining a correlation coefficient above 98% on both image level and patient level. This is mainly due to two factors: First, the preselection of parasite candidates by IGMS covers parasites in the ground truth quite well. Second, our CNN model with the customized input size and network layers can classify the preselected candidates with high accuracy. Our IGMS method generates false positive patches (negative patches) that are very similar to parasites (positive patches) so that our CNN model learns to reduce false positives. We have also performed experiments with negative patches randomly selected from the background. However, the accuracy decreased to less than 75% on Set B. This is because the random selection of negative patches generates too many clean negative patches. Therefore, training the CNN model on such patches leads to many false positives. We have compared the performance of our method using three different input patch sizes: 36 × 36 × 3, 44 × 44 × 3 and 52 × 52 × 3. We observe that with a small patch size of 36 × 36 × 3, too many false positives are detected. Using this patch size, the method does not work because the patch size is too small to include enough information for identifying parasites. When using an increased patch size of 52 × 52 × 3, the AUC value on patch level is 97.30%, which is very close to our reported results for a patch size of 44 × 44 × 3. However, the correlation coefficients decrease to 0.96 and 0.97 on the image and patient levels. This is because more background noise is introduced when the patch size increases.



Parasite detection on an example image using our proposed method. (a) Parasites annotated in the ground truth (yellow circles) and screened parasite candidates that overlap more than 50% with the parasites in the ground truth (red and green circles). Red circles indicate candidates that are finally predicted as parasites (true preselected parasites), and green circles indicates those that are predicted as non-parasites (false preselected parasites). (b) Probabilities of parasite candidates that overlap more than 50% with parasites in the ground truth. The number under each patch denotes the output probability of the CNN. Red and green numbers indicate probabilities larger than 0.6 and smaller than 0.6, respectively.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, July 2022
Impact Factor 6.252

VI. CONCLUSION

In this paper, we implement a deep learning application for smartphones to detect malaria parasites in thick smear images. Our processing pipeline for automated parasite detection consists of two stages: parasite screening and classification. An intensity-based Iterative Global Minimum Screening (IGMS) first performs a fast screening of an entire thick smear image to generate parasite candidates. A customized CNN model then classifies each candidate as either parasite or background. Our experimental results demonstrate the practicality of our method for automatic detection of malaria parasites. To the best of our knowledge, our paper is the second paper that has developed a smartphone application for thick blood smear screening [18], and the first paper that has applied deep learning techniques for parasite detection in thick smears on smartphones, with evaluation on patient level. We make our dataset of 1819 images from 150 patients publicly available, as a service to the research community, which will mitigate the problem of lacking training data for automated malaria diagnosis in thick blood smears.

ACKNOWLEDGMENT

Conflicts of Interest: All authors have read the journals policy on disclosure of potential conflicts of interest and have none to declare. All authors have read the journals authorship agreement and the manuscript has been reviewed and approved by all authors.

REFERENCES

[1] "World malaria report," WHO, Geneva, Switzerland, 2018.
[2] WHO, Guidelines for the Treatment of Malaria, 3rd ed. Geneva, Switzerland: World Health Organization, 2015.
[3] K. S. Makhija, S. Maloney, and R. Norton, "The utility of serial blood film testing for the diagnosis of malaria," Pathology, vol. 47, no. 1, pp. 68–70, 2015. [4] WHO, Malaria Microscopy Quality Assurance Manual. Geneva, Switzerland: World Health Organization, 2016.
[5] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, "Image analysis and machine learning for detecting malaria," Transl. Res., vol. 194, pp. 36–55, Apr. 2018.
[6] Z. Liang et al., "CNN-based image analysis for malaria diagnosis," in Proc. IEEE Int. Conf. Bioinf. Biomed., Shenzhen, China, 2017, pp. 493–496.
[7] S. Rajaraman et al., "Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images," J. Med. Imag., vol. 5, no. 3, Jul. 2018, Art. no. 034501.
[8] L. Rosado, J. M. Correia da Costa, D. Elias, and J. S. Cardoso, "A review of automatic malaria parasites detection and segmentation in microscopic images," Anti-Infective Agents, vol. 14, no. 1, pp. 11–22, Mar. 2016.
[9] P. A. Pattanaik and T. Swarnkar, "Comparative analysis of morphological techniques for malaria detection," Int. J. Healthcare Inf. Syst. Inform., vol. 13, no. 4, pp. 49–65, Oct. 2018.
[10] S. Kaewkamnerd, A. Intarapanich, M. Pannarat, S. Chaotheing, C. Uthaipibull, and S. Tongsima, "Detection and classification device for malaria parasites in thick-blood films," in Proc. IEEE Int. Conf. Intell. Data Acquisition Adv. Comput. Syst., Prague, Czech Republic, 2011, pp. 435–438.
[11] N. S. M. M. Hanif, M. Y. Mashor, and Z. Mohamed, "Image enhancement and segmentation using dark stretching technique for Plasmodium Falciparum for thick blood smear," in Proc. Int.