Detection of Malaria in Blood Cells Using Convolution Neural Network

N. Nalini, Anurag Nepal, Avishek Rijal, Baibhav Dhakal, and Sabin Kandel

Abstract There are a variety of automated diagnostic techniques and models using numerous supervised learning models, but most of these models cater especially to the diseases that are seen in the Western countries, and they rarely see the outbreak of diseases such as malaria, dengue, etc. Early detection of these diseases can control the mortality rate and help save lives. Malaria while being a curable disease still has no vaccine available for it, so early detection of malaria can help determine the risk and can prove to be lifesaving. And with the time taken to collect, analyse and diagnose malaria in the blood is valuable time that can be the difference between a patient's life or death. And less developed countries do not have the proper resources for fast response against the disease. In this project, we are hoping to develop an effective and efficient automated diagnostic model using machine learning models. For this, we have implemented a model based on CNN architecture to detect malarial parasites in blood cells and then use advance image processing techniques to contour and isolate the parasite to track the progression of the disease.

Keywords Malaria · Supervised learning · CNN · Mobile application

N. Nalini · A. Nepal (✉) · A. Rijal · B. Dhakal · S. Kandel
Computer Science Engineering, Nitte Meenakshi Institute of Technology (Visvesvaraya Technological University), Bengaluru, India
e-mail: nepalanurag72@gmail.com

N. Nalini
e-mail: nalini.n@nmit.ac.in

A. Rijal
e-mail: avishekrijal01@gmail.com

B. Dhakal
e-mail: dhakalbaibhav07@gmail.com

S. Kandel
e-mail: mesabmk@gmail.com

N. R. Shetty et al. (eds.), *Advances in Computing and Information*, Lecture Notes in Electrical Engineering 1104, https://doi.org/10.1007/978-981-99-7622-5_19

1 Introduction

Malaria is caused when an infected female Anopheles mosquito bites a human, during which protozoan parasites of the genus Plasmodium are transmitted and the red blood cells are infected with the parasite, and is still considered as a major health problem in the world. Though efforts are being made to regulate and exterminate malaria, it still is a life-threatening disease, especially in developing African and Asian countries where there is less accessibility to proper and quality health services are the contributing factors. Poor living conditions and the increase in freak weather conditions certainly help in spreading the disease and making it life-threatening.

They are of five types, namely *Falciparum, Vivax, Malariae, Ovalae and Knowlesi.* All of which have different identifiers and give off different characteristics in their lifecycle, the summary of which can be found in Fig. 1. Despite the reduction in deaths and cases of malaria in the developing, still an estimated 214 million of cases are seen every year with it claiming more than 400 thousand lives. As in Fig. 2, the death rate is more in countries with poorer reach to health services and sanitation practices as mosquitoes thrive in polluted environments. The main problem these countries have is that they cannot sufficiently fill the gaps in their healthcare systems which starts with early diagnosis of the disease. While the economic burden of detecting and caring for malaria is in billions, there is no improvement in the lives of people in these countries.

There are several methods that can be used to diagnose malaria. Polymerase chain reaction or PCR can be used to show a higher sensitivity to detect the parasite, but it is a complex and high cost per test technology. Similarly, fluorescent microscopy can be used to detect the parasite under fluorescent light, but it is also very costly. That is why manual microscopy is widely used for the detection. Manual microscopic detection of malaria is done by inspection of parasites after Giemsa staining, which is a procedure used to highlight the parasite where the stained blood smears into thick and thin blood smears [1] (Fig. 3).

The thick blood smears, which are prepared with thick layer of blood on the slide, are mainly used for the detection of malaria parasites [1]. But with thick blood smears clear visibility and differentiating the RBCs from other components of blood like WBC's and platelets prove to be major problems [1]. That can be overcome by the thin blood smears which can be used for detailed examination such as the parasite type and the stage of the disease [1]. But the lack of expertise and feasibility to afford microscope can hamper the early diagnosis in rural and endemic regions. But for automated diagnosis of malaria, we must get the image of the blood smears from a microscope. Different kinds of smears can be used to properly develop and train the model for malaria detection. After we acquire and process the images in the required form, it can be run through the model to detect if there is malaria in the blood or not.

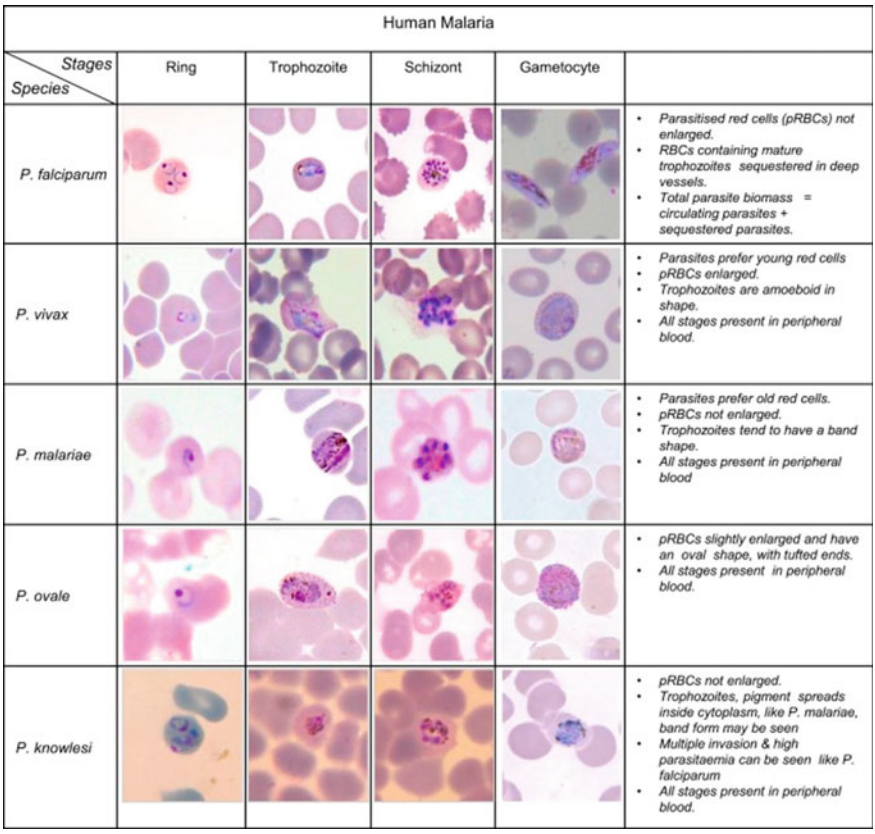

Fig. 1 Summary of types of malaria

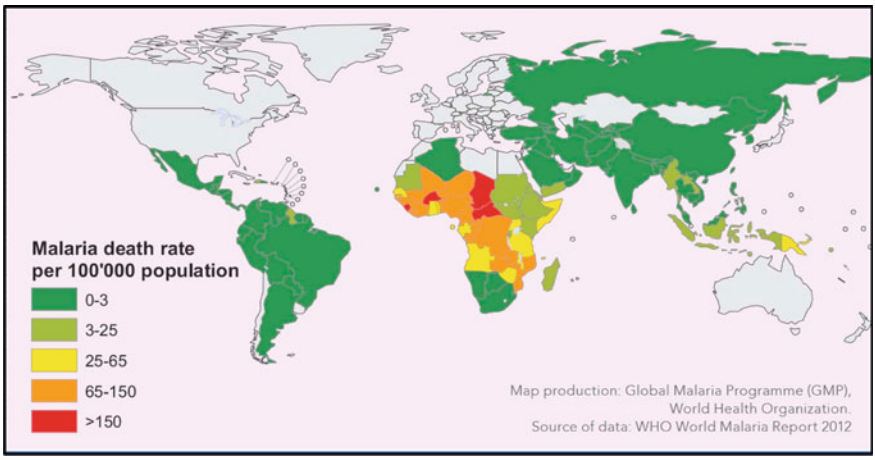

Fig. 2 Malaria cases and death rates

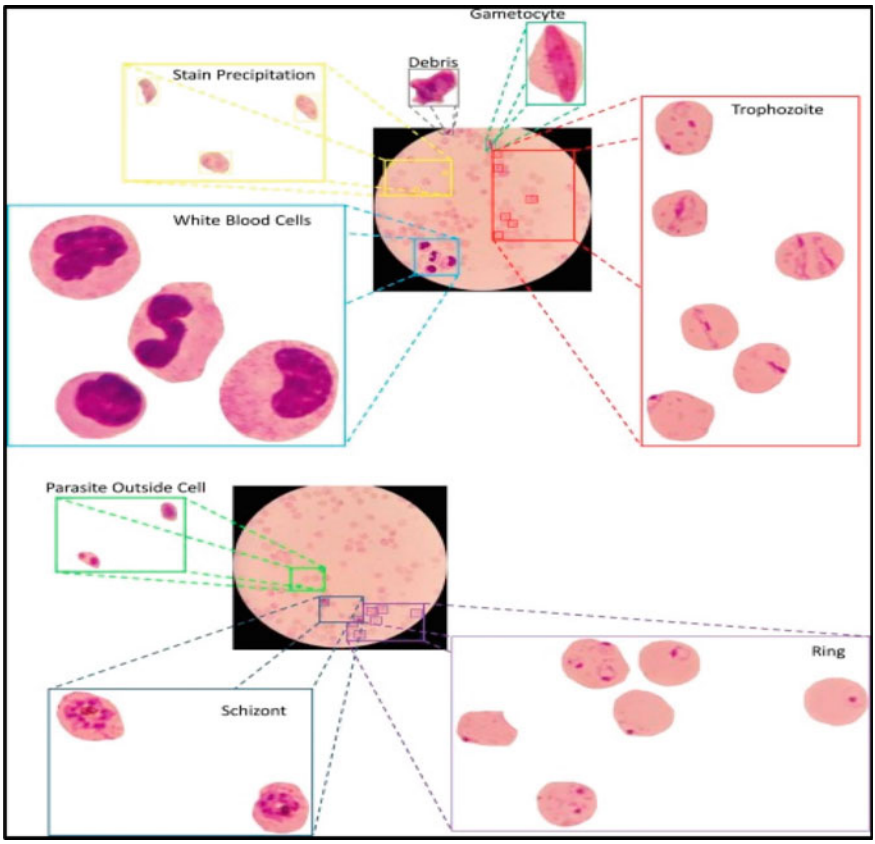

Fig. 3 Progression of the disease

2 Related Works

Pirnstill and Cote Pirnstill and Cote (2015) presented malaria detection using a polarized microscope [2]. They used an inexpensive cell phone based on the idea of a polarized microscope to take samples, in which they were able to see the hemozoin. Hemozoin is a dirty product that is formed after staining the blood. Their cell phone, based on a polarized microscope, was able to take a picture and detect malaria.

Breslauer et al. (2009) introduced a cell phone-based microscope for global health applications, including malaria. Use of microscope attachments helped in advanced adjustments for the mobile camera. A microscope and a mobile camera can capture bright light and the fluorescent image. They used a bright digital image and a blood smear image and were able to identify the malaria parasite. Their system was able to detect malaria parasites thick blood smears and differentiate them with red blood cells.

Rosado et al. (2016) introduced the automatic detection of malaria parasites in blood smears using mobile camera. They used a machine-readable study method, the SVM (vector machine) class, to test the presence of P. falciparum trophozoites and the WBC in Giemsa stains by testing thick blood smears for blood tests. While WBC detection achieved 98.2% sensitivity and 72.1% specificity, P. Falciparum trophozoite detection achieved 80.5% sensitivity and 93.8% specificity [3].

Quinn (2014) presented an automated blood smear analysis for cellular malaria. The design of a 3-D printable adapter i.e., adapter, attaches a cell phone to any type of microscope. With the help of cell phone, they were able to take pictures of blood smears. They also provided a workflow for automatic blood smear testing that includes morphologic and temporal calculation and a combination class of training trained in these tasks to distinguish themselves from the strange fragments containing parasites and common episodes.

Skandarajah et al. (2014) designed and developed a multimedia-based microscope. Their microscope design is compatible with all major cell phones from any manufacturers. They introduced their microscope cell phone and were able to create a multimedia microscope with a light microscope like resolution. Phones with 5 or more-megapixel cameras can produce unlimited diffraction adjustments that compatible to take a single cell image.

Dallet et al. show an android mobile app that can detect and detect malaria in the given blood cultures. This app is based on the annular ring ratio method. In this way, the app detects various blood components such as red blood cells and white blood cells and detects whether malaria parasites are present in red blood cells. It can also classify different categories of parasites and calculate the spread of infection.

Cesario et al. describe in detail cell-based solutions and support for animal diseases in areas where health care services are limited. They focus is mainly on image analysis and classification, in which remote diagnoses of the rarest diseases are performed [4]. This reduces the burden on health workers and reduces the risk of misdiagnosis of rare diseases. Their work in developing and implementing image analysis and classification in the medical field has received constructive feedback from health professionals.

Shah et al. present an in-depth study of how you can identify malaria in blood cells. They use the convolution neural network (CNN) to differentiate between infected and healthy blood samples. Their model works with minimal resources and has a low calculation time that provides instant results. The CNN separator used has several flexibility layers with a few filters that show good accuracy even under low-level resources. Their CNN algorithm is 95% accurate and operates with limited resources.

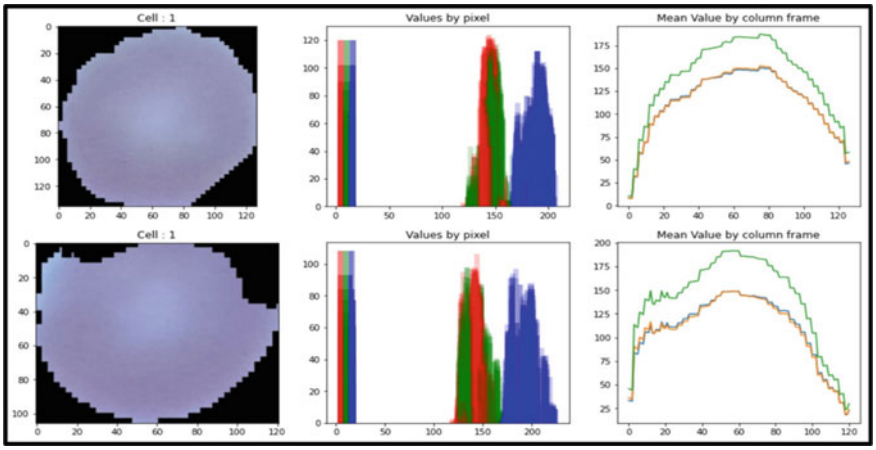

Fig. 4 Uninfected blood sample

3 Proposed Work

Data collection and data preprocessing were to be done before starting to form the model to identify the malarial parasites in the blood samples. With what we are trying to do with supervised learning the dataset and the data is the main thing that controls the outcome's quality. That is why while collecting and processing the data utmost diligence is required.

3.1 Dataset

The dataset taken contains 27,558 images of the cells which are divided evenly into two categories i.e., "Parasitized" and "Uninfected". Looking at the mean pixel values of the images, we can differentiate between the samples and take some idea about how the computer will go about classifying the classes. As in Fig. 4, the uninfected sample will have no discernible difference in the pixel values while in Fig. 5, the infected samples will have difference in mean pixel values.

3.2 Tracking the Parasite Within the Blood Sample

The images must be manually corrected before feeding them to the model. As the images contain false labelling and noise in them, they must be investigated. The images have been resized into 32 × 32 size before feeding them. The images are not uniform, and the resized images help the model run more efficiently than feeding

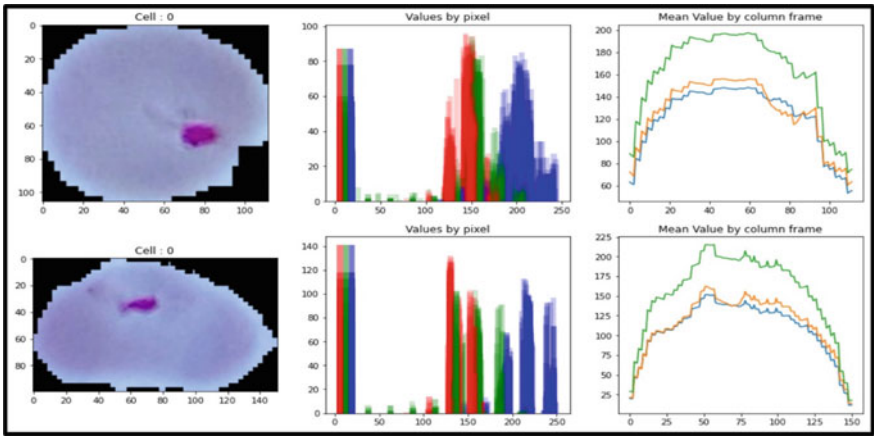

Fig. 5 Infected blood sample

original sized images. Also, for getting the progression of the parasite, we are tracking the ratio of parasite to the blood samples. For this, the contrast of the image must be changed to better suit the algorithm as well as the noise inside the image must be removed. This helps in isolating the parasite inside the blood sample and to contour and highlight it in the image. This can help in tracking the progression of the parasite (Fig. 6).

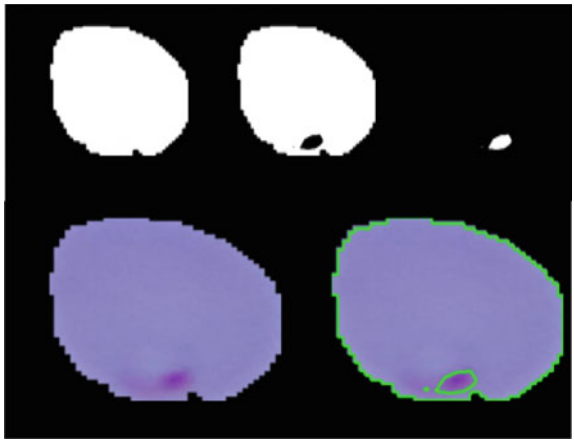

Fig. 6 Preprocessing to remove noise and highlight the parasite

3.3 Training Model

We are trying to detect malaria parasite in each blood smear using CNN, also known as CNN. Now the obtained microscopic image of the blood sample is of very high quality and dimension. We will apply some image processing techniques to obtain an absolute dimension without losing the essential features of the image. We hope to do real-time analysis and detection of malaria, so we need to normalize and shrink the image for faster computation and processing. We are using CNN to find out whether a blood sample contains malaria parasite or not. In addition, we plan to use image segmentation to find out where parasites are present in each blood sample.

ConvNet/CNN is a deep learning algorithm which takes an input image and assigns importance to different aspects/objects of the image and be able to differentiate one from the other could. ConvNet can successfully capture the spatial and temporal dependencies in an image through the application of contextual filters. The architecture allows for better fitting to image datasets due to the reduction in the number of parameters involved and the reusability of the weights. In other words, the network can be trained to better understand the sophistication of the image. Using a convolution model with a sigmoid activation function, we give the resized array as the input to the network; we achieved 94.68% accuracy.

4 Results and Performance Analysis

4.1 Results

4.1.1 First Attempt

For the first attempt, we made a model with 17 layers of CNN with convolution, batch normalization, max pooling, flattening and dense. In this attempt, we took the images directly form the folder using Keras Flow from directory into the model, and we get the output as the 0 and 1. The flow from directory is an iterator that leverages the GPU to pull the images directly from the specified directory and the classes are identified based on the subfolders of the directory, i.e., in our case "Parasitized" and "Uninfected" as the classes. With this we do not have to preprocess the images as the Keras flow from directory iterator has an image data generator function which takes in the images and then resizes the images to the specified size which can be directly fed into the neural network (Fig. 7).

We were getting spikes in the 11th epoch of the validation testing. The spike happened in epoch no 11 which we can refer to as the problematic epoch. Since this is a binary classification model detecting between parasitized cell and uninfected one, the confidence or accuracy threshold is between 49 and 51%. The output in the form of Dense (1), i.e., there is only one output value. This happened as in our first model we took the results as a dictionary of image and value (0 for infected and 1

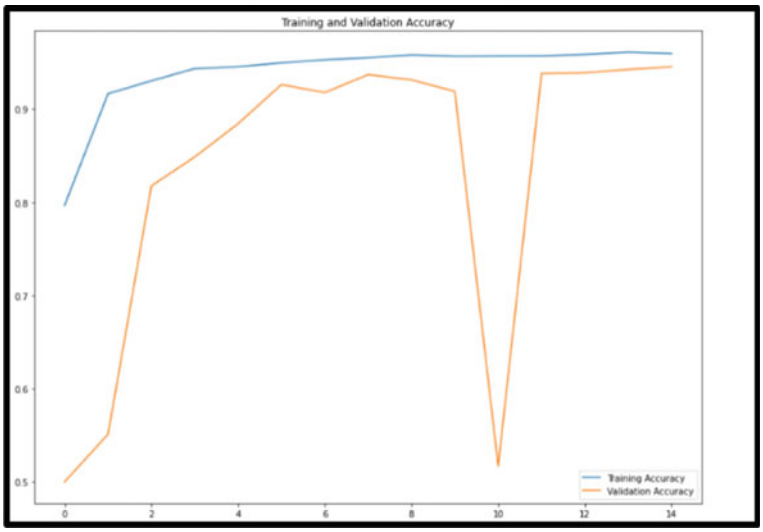

Fig. 7 Training and validation accuracy for first model

for uninfected) pairs and after some time the model defaulted back to the threshold value as it got confused during classification and can only return one value back as the output. The reason that we see the drop in accuracy can be attributed to batch training and bad labelled dataset. This process is more computationally expensive making the model inefficient in handling our case as the network kept on defaulting to the threshold value after it encountered unknown test instances.

4.1.2 Second Attempt

For the second attempt, we chose to simplify the model by reducing the layers as well as taking the values of the images as an array. The time to load the images onto the model was high due to the inefficiency of going to directory each time when an image is needed. So, on order to solve this problem we store every image onto a multidimensional NumPy array containing the images and its respective class. As a result of the images being in an array, resizing them to feed to the neural network is easier and it takes up less storage and time. The output layer in this attempt was changed to a Dense (2) layer, i.e., the output is in the form 2 output values which are probability of belonging to class 0 and probability of belonging to class 1. So there is less chance of the model getting confused whenever there are unknown test instances. Even if the network gets confused, the network will return the value with the higher probability so there is no erratic behaviour during training and validation (Fig. 8).

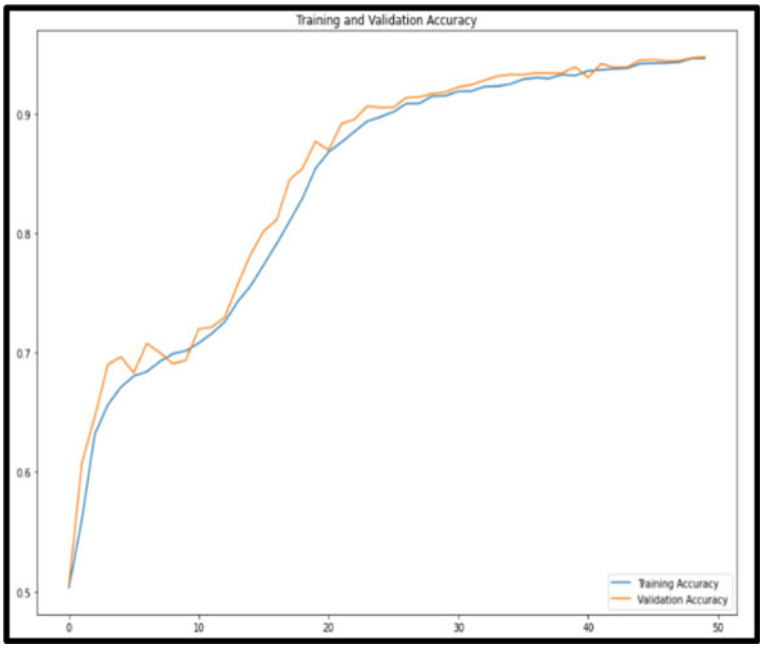

Fig. 8 Training and validation accuracy for second model

4.2 Performance Analysis

For analysis of our model, we have used various performance metrics and the summary of the same can be found in Table 1. With a total testing set of 5512 images, the uninfected image classification has a precision of 96% with a F1-score of 0.95. Similarly, infected image classification has a precision of 93% with a similar F1-score of 0.95. The lower precision in the infected images is due to the model getting the white blood cells with the parasites. Also, the confusion matrix shows the accuracy across the validation sets (Fig. 9).

Table 1 Performance metrics

	Precision	Recall	F1-score	Support
Uninfected	0.96	0.93	0.95	2727
Parasitized	0.93	0.97	0.95	2785
Accuracy			0.95	5512
Macro avg	0.95	0.95	0.95	5512
Weighted avg	0.95	0.95	0.95	5512

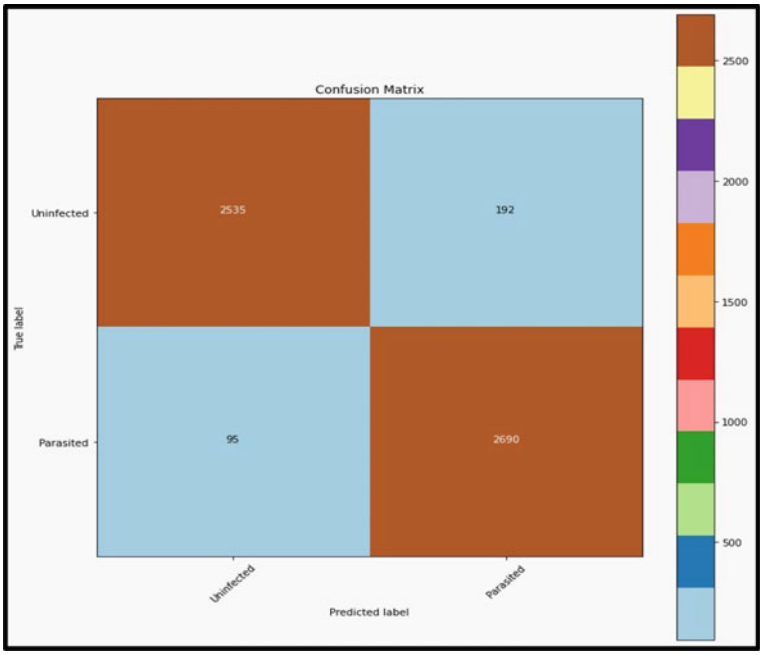

Fig. 9 Confusion matrix

5 Conclusion

Malaria is a very important issue to be tackled in the developing countries which has been left unnoticed and is claiming lives to this day. Unavailability of vaccine is the main contributing factor for it still being a problem for the developing countries, and we hope to help just a little by developing a cost-effective way to diagnose the parasite. So, using CNN to segment and analyse a photo from cheaply available methods can help in detecting the parasite in the blood smears. This would help in subsetting the cost of using big and bulky microscopes and help the developing countries detect the parasites in the patients' blood as early as possible, so that they can help reduce the casualties. We hope that our application will be a modular one with more modules to detect more vector-based parasites in the future.

References

1. Rehman A, Abbas N, Saba T, Mehmood Z, Mahmood T, Ahmed KT (2018) Microscopic malaria parasitaemia diagnosis and grading on benchmark datasets. Microsc Res Tech 81(9):1042–1058. https://doi.org/10.1002/jemt.23071
2. Pirnstill CW, Coté GL (2015) Malaria diagnosis using a mobile phone polarized microscope. Sci Rep 5:13368. https://doi.org/10.1038/srep13368.PMID:26303238;PMCID:PMC4548194
3. Rosado L, da Costa C, Manuel J, Elias D, Cardoso J (2016) Automated detection of malaria parasites on thick blood smears via mobile devices. Procedia Comput Sci 90:138–144. https://doi.org/10.1016/j.procs.2016.07.024
4. Canty MJ (2014) Image analysis, classification and change detection in remote sensing: with algorithms for ENVI/IDL and Python, (3rd edn). CRC Press
5. Breslauer DN, Maamari RN, Switz NA, Lam WA, Fletcher DA (2009) Mobile phone based clinical microscopy for global health applications
6. Skandarajah A, Reber CD, Switz NA, Fletcher DA (2014) Quantitative imaging with a mobile phone microscope. PLoS ONE 9
7. Quinn JA, Andama A, Munabi I, Kiwanuka FN (2014) Automated blood smear analysis for mobile malaria diagnosis. Mobile Point Care Monit Diagno Dev Design 31:115
8. Dallet C, Kareem S, Kale I (2004) Real time blood image processing application for malaria diagnosis using mobile phones international conference on circuits and systems. IEEE 2405–2408
9. Herrera S, Vallejo AF, Quintero JP, Arévalo-Herrera M, Cancino M, Ferro S (2014) Field evaluation of an automated RDT reader and data management device for Plasmodium falciparum/Plasmodium vivax malaria in endemic areas of Colombia. Malar J 13:87