International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Malaria Cell Image Classification Using Deep Learning

Mohd Zia ur Rehman[1], Areeb Ahmad[2], Firoz Chaudhary[3], Shaikh Mohd Ashfaque[4]

[1,2,3,4]*Department of Computer Science, Rizvi College of Engineering, Mumbai, India*

Abstract - *Malaria is a blood disease caused by the bite of female Anopheles mosquito and by the transmission of Plasmodium parasites . To diagnose and compute parasitemia, Macroscopics is commonly used which uses thick and thin blood smears. The accuracy of the test depends upon accurate classifying and the correct counting of parasitized and uninfected cell. The State of the art image analysis is based on computer aided diagnosis (CADx).It uses machine learning (ML) techniques for microscopic images of the smears using features that demand expertise in analyzing textural, morphological and positional variations of the region of interest (ROI). The Convolutional Neural Networks (CNN)is a sub-class of deep learning (DL) models. It is highly scalable, superior and gives better results with end-to-end feature extraction and classification. It could, therefore, serve as an effective diagnostic aid could be automated malaria screening using DL techniques. In this study, we evaluate the performance of pre-trained CNN based DL models as feature extractors towards classifying parasitized and uninfected cells to aid in improved disease screening. From the underlying data, we determine the optimal model layers for feature extraction Statistical validation of the results demonstrates the use of pre-trained CNNs as a promising tool for feature extraction for this purpose.*

Key Words*:* Malaria, deep learning,

1. INTRODUCTION

Malaria is a disease that has to be detected promptly for on time treatment of the patient and for preventing it from spreading in the community via mosquitoes. Malaria should be treated as a medical emergency and not just a regular illness. Malaria is a disease that has to be detected promptly for on time treatment of the patient and for preventing it from spreading in the community via mosquitoes. Malaria should be treated as a medical emergency and not just a regular illness.

• Diagnosis of malaria can be difficult: Where malaria is not endemic anymore, health-care providers may not be familiar with the disease. Clinicians while examining a malaria patient tend to forget to consider malaria among a potential disease and hence don't suggest the necessary diagnostic tests. Detecting parasites while examining blood smears under the microscope requires experience, which the laboratorians may lack and thus fail to detect.

• In some malaria-endemic areas, a large section of the population is infected by the parasites but is still not ill. Such carriers have the malaria infection but develop just enough immunity to not get the malaria illness. In that case, finding malaria parasites in an ill person doesn't necessarily mean that malaria is the root cause of the person's illness or the parasites are solely responsible for the illness. Thus, automated and scalable feature engineering is made possible with CNNs. Also, tasks like image classification are made possible by plugging in dense layers at the end. For pre-trained model with constrains like less data, Automated malaria detection uses deep learning Models like CNNs coz it's cheap and scalable.

2. PROBLEM STATEMENT

The bite of female Anopheles mosquito is the transmission of the Plasmodium parasites which causes the blood disease Malaria. To diagnose disease and compute parasitemia , Macroscopic is used commonly which uses thick and thin blood smears. However, the efficiency in examining depends on the quality of the smear and expertise in classifying and counting parasitized and uninfected cells.

Convolutional Neural Networks (CNN), a class of deep learning (DL) model with end-to-end feature extraction and classification promises highly scalable and superior results.

That's why automated malaria screening uses DL techniques which can be helpful for effective diagnostic aid.

3. OBJECTIVES

In this study, we performed various experiments to visualize the weights, saliencies, class activation maps, and ROI localization toward understanding the learned behaviour of the customized CNN applied to the challenge of classifying parasitized and uninfected cells to help in malaria screening. An explanation of what made these models arrive at the classification is also provided. Finally, the performance of the models were validated at the cell and patient level with a large-scale clinical dataset and analyzed for the presence/ absence of a statistically significant difference in their performance metrics. Pilot studies are been performed in deploying the customized model into mobile devices and thereby analyzing its performance. The flexibility for dynamic allocation of CPU and GPU for training the models and testing on unseen data is now offered by ML libraries for mobile devices. The trained model can help us minimize delays in environments with limited resources.

4. LITERATURE REVIEW

[1] The development of malaria diagnostic techniques: a review of the approaches with focus on dielectrophoretic and magnetophoretic methods

Author: Surasak Kasetsirikul, Jirayut Buranapong

Approach: Manual diagnosis of blood smears requires expertise in classifying and counting parasitized and uninfected cells, also it's an intensive manual process and hence takes time. This process may not do well, especially in regions where the right expertise is difficult to find. To extract hand engineered features and build machine learning based classification models, some advancements have been made in leveraging state of the art image processing and analysis techniques. However, the models with more data being available for training are not scalable and given the fact that hand-engineered features take a lot of time..

[2] Image Analysis and Machine Learning for detecting Malaria

Author: Mahdieh Poostchi, Kamolrat Silamut, Richard J Maude, Stefan Jaeger, George Thoma

Approach: Deep learning is serving as the latest trend in machine learning. It has already boosted the performance in many nonmedical areas. Deep learning is an extension of the well-known multilayer neural network classifiers trained with back propagation, except that several more layers are used. There are different layers that are employed in typical successions. Deep learning typically requires large training sets. This is often the reason why medical applications are among the last applications to adopt deep learning, as annotated training images are significantly harder to get due to expert knowledge requirements and privacy concerns. The primary article to use deep learning to malaria diagnosis is by Liang et al., who use a convolutional neural network to discriminate between infected and uninfected cells in thin blood smears, after applying a classic cell segmentation approach. This is often a perfect application for deep learning because images of segmented red blood cells are a natural input for a convolutional neural network. Deep learning doesn't require the design of handcrafted features, which is one in every of its biggest advantages.

[3] Image Analysis and Machine Learning for detecting Malaria

Author: Sivaramakrishnan Rajaraman, Kamolrat Silamut, Md. A. Hossain, I. Ersoy, Richard J. Maude Stefan Jaeger, George R. Thoma, Sameer K. Antani

Approach: Deep learning is the latest trend in machine learning, which has already boosted the performance in many nonmedical areas. Deep learning is an extension of the well-known multilayer neural network classifiers trained with back propagation, except that many more layers are used. There are also different kind of layers that are used in typical successions. Deep learning typically requires large training sets. This is the reason why medical applications have been among the last applications to adopt deep learning, as annotated training images are significantly harder to obtain because of expert knowledge requirements and privacy concerns. The first article to apply deep learning to malaria diagnosis is by Liang et al., who use a convolutional neural network to discriminate between infected and uninfected cells in thin blood smears.. This is an ideal application for deep learning because images of segmented red blood cells are a natural input for a convolutional neural network. Deep learning does not require the design of handcrafted features, which is one of its biggest advantages.

[4] Detection of malaria parasites in thick blood smear: A review

Author: Faza Maulaf Azif, Hanung Adi Nugroho, Sunu Wibirama

Approach: The clinical diagnosis of malaria is one of traditional malaria detection methods. The diagnosis is clinically done by observing the visible physical symptoms and characteristics of patient's body. Medics will try to identify the disease though early symptoms happened are still common, such as fever, headache, diarrhoea, cold, chills, myalgia, abdominal pain, and vomiting. The number of symptoms that experts need to consider in clinical diagnosis is leading to the difficulty for accurate identification of the disease, because the symptoms of the disease are likely still too general to determine whether that the patient has malaria or not. The diagnosis of malaria using PBS(Peripheral Blood Smear) is based on blood smear. The presence of malaria parasites in the blood smear was observed using microscope. The blood smear can be thick and thin. The thick blood smear is used to detect the presence of plasmodium parasites, whereas the thin blood preparation is used to detect the type and phase of plasmodium. The results of sensitivity and specificity depend on the technique used, the reagent, and the expertise of medical personnel. The required time to do the diagnosis is about 30-60 minutes.

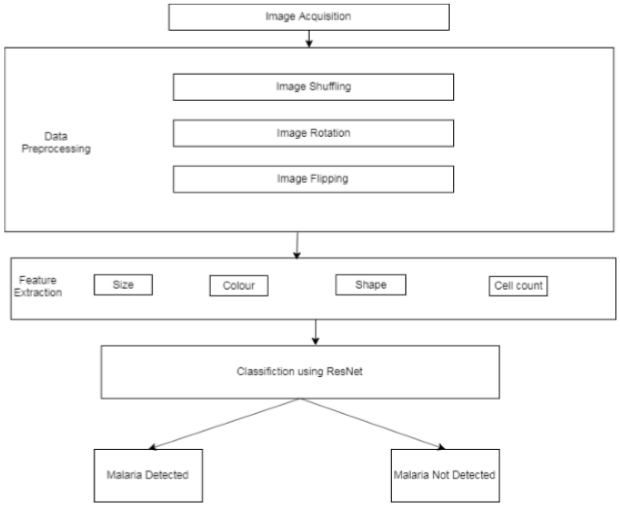
[5] Transfer Learning with ResNet-50 for Malaria Cell-Image Classification

Author: A. Sairaj Bharadwaj Reddy & D. Sujitha Juliet

Approach: Deep learning models were inspired by the functioning of human nervous system. Deep learning has been a proven method which increases the performance in any field. Convolutional Neural Networks (CNN) perform well in image recognition problems, whereas Long Short-Term Memory (LSTM) perform well in speech recognition. Deep learning models have also shown promise in fields like financial fraud detection, Natural Language Processing (NLP), Bioinformatics, and many more. Deep learning models such as Artificial Neural Networks (ANN) and CNN are feed forward networks with one input layer, one output layer and many hidden layers. These networks can be trained by back propagation. The building block of a neural network is called a neuron. The functionality of it is like that of a human neuron, i.e. to give an output to the given input. In mathematical terms neuron of a neural network gives an output by applying a mathematical function to the given

inputs. The output from a neuron is input to the many neurons in the next layer. The mathematical functions used to by a neuron in a neural network as generally termed as Activation functions. Some, of these activation functions are Uni-Polar Sigmoid Function, Bi-Polar Sigmoid Function, Hyperbolic Tangent Function and many more. Choosing an activation function for the last layer is often tied to the type of the problem (Regression or Classification). Whereas choosing activation function for hidden layers can be interesting, i.e. choosing sigmoid for very deep neural networks can cause vanishing gradient problem and sometimes using Relu as activation function can cause "Dead Neuron's" in the hidden layers. Choosing activation function to make the network perform at its best is one of the core problems.

5. METHODOLOGY



Deep learning models, or more specifically Convolutional neural networks (CNN's), have proven very effective in a wide variety of computer vision tasks. Shortly, the key layers in a CNN model include convolution and pooling layers.

CNNs help with automated and scalable feature engineering. Moreover, tasks like image classification are performed by plugging in dense layers at the end of the model. Automated malaria detection models implementing deep learning models like CNNs could be cheap, scalable and very effective especially with the advent of transfer learning and pre-trained models that work quite well, even with constraints like less data.

The paper leverages six pre-trained models on a dataset to obtain an impressive accuracy of 95.9% in detecting malaria vs non-infected samples. Our focus is to try some simple CNN models from scratch and a couple of pre-trained models using transfer learning to see the results we can get on the same dataset. We will use open source tools and frameworks, including Python, JSON and YAML, to build our models.

STEPS:

I. *The first step is usually the acquisition of digital images of blood smears, which largely depends on the equipment and materials being use. The Image acquisition section breaks down the different approaches for the different types of microscopy, blood slides (thin or thick), and staining.*

II. *Following image acquisition, to remove noise and to normalize lighting and color variations inherent in the image acquisition and staining process most systems perform one or several preprocessing methods. The Preprocessing section sorts the publications according to the pre-processing methods implemented.*

III. *The next step usually involves the detection and segmentation (outlining) of individual blood cells and maybe other objects that can be visible in a blood slide image, such as parasites or platelets.*

IV. *In the last step, a mathematical discrimination method that classifies the segmented objects into different classes based on the computed features is implemented. For example, labelling each red blood cell as either infected or uninfected is a key classification task performed in this step, which then allows to compute the parasitemia.*

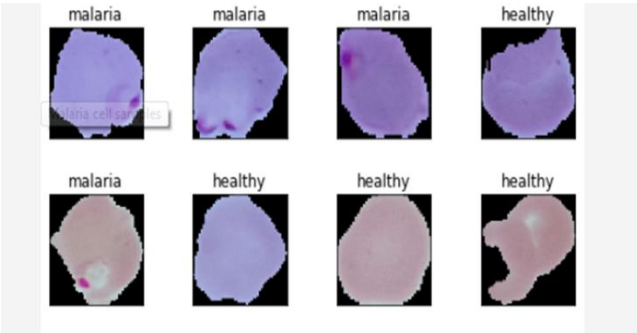


Fig -1: Classification of the images from the dataset as Malaria or Healthy

5.1 Dataset

The dataset we would be using for analysis (training, validation and testing) is taken from The Lister Hill National Center for Biomedical Communications, who have cautiously accumulated and annotated this dataset of fit- healthy and infected blood smear pictures.

The dataset consists of 27,558 segmented cellular images, with the same instances of 13,779 parasite and 13,779 Non-parasite segmented RBC images. In Positive samples

plasmodium is present and in negative samples plasmodium is absent. After applying data augmentation, a pre-processing technique our count of 27,558 input dataset images would increase and become 163314. Thus, our training dataset increase by 6 folds helping us achieve a good accuracy for the testing dataset.

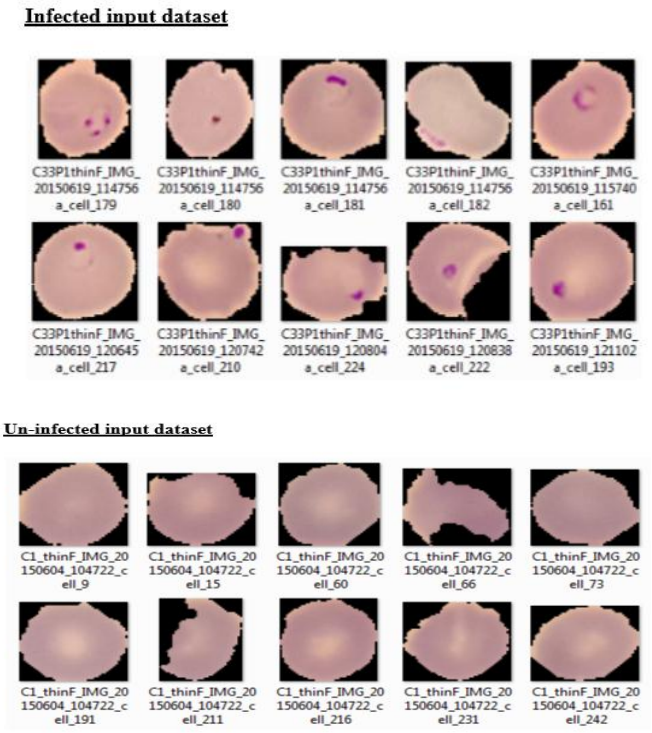


Fig -2: Infected and Uninfected Blood smears

To further check the cross validation and the accuracy of our system, we tested our system with a new dataset other than the previous dataset acquired from the NLM. This new Malaria ParasiteImage Database for testing the accuracy of our system has been taken from an article inside the book "Processing and Analysis of Biomedical Information" . MPIDB which is the open image dataset comprises of 3 species of Malaria parasites: Ovale, Vivax, Falciparum. For all the classes, 4 distinctive stages of life, depicted in the filenames as follows:

• R: implies the occurrence of Ring stage parasites

• T: implies the occurrence of Trophozoite stage parasites

• S: implies the occurrence Schizont stage parasites

• G: implies the occurrence Gametocyte stage parasites

The proportion of execution and exactness of the strategy was assessed by two measurements: sensitivity and specificity. We use binary confusion matrix, to keep a track on the True Positives, True Negatives, False Positives and False Negatives. We printed the classification report to get an idea of accuracy, precision, f1-score and recall for individual class labels. Sensitivity is characterized as the likelihood (rate) that patients with the contamination will have a positive outcome utilizing the test under assessment. Specificity is characterized as the likelihood (rate) that patients without the disease will have a negative outcome utilizing the test under assessment.

6. RESULT AND DISCUSSION

From the very different methods published during the last 10 years, we can see that there has been a lot of experimenting done to reach the current state of the art.
A well-performing system will require the interplay of several factors, such as the characteristics of the microscope, the type of staining, the slide preparation, and the image analysis and machine learning software

Exponential progress has been made as can be seen by the natural development of methods used for image analysis and machine learning. In fact, this development has largely followed the development in other fields and has adopted major techniques and successfully applied them to malaria diagnosis. Many of these methods are general-purpose methods that are independent from the application domain. The values for sensitivity and specificity are expressed in terms of true positives (TP), false positive (FP), false negative (FN) and true negative (TN) a defined below in expressions 1 and 2:

Sensitivity=TP/TP+FN (1) Specificity= TN/TN+FP (2)

By using a function, we plot something known as a binary confusion matrix. A confusion matrix will give us information about the true positives, true negatives, false positives and false negatives. It gives us knowledge not just into the blunders being made by a classifier however more significantly the sort of mistakes that are being made

Shown below is the model evaluation on unseen data

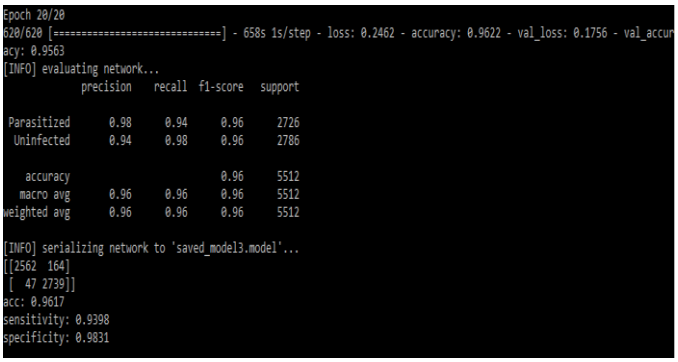


Fig -3: Result analysis of testing data

This means in the Fig 3,

● The accuracy for the testing data is 96.17% and the validation accuracy is 95.63%. ● 164 out of 2726 (6 %) infected cells were classified as clean — False Negative. ● 47 out of 2786(1.6 %) clean cells were classified as infected — False Positive.

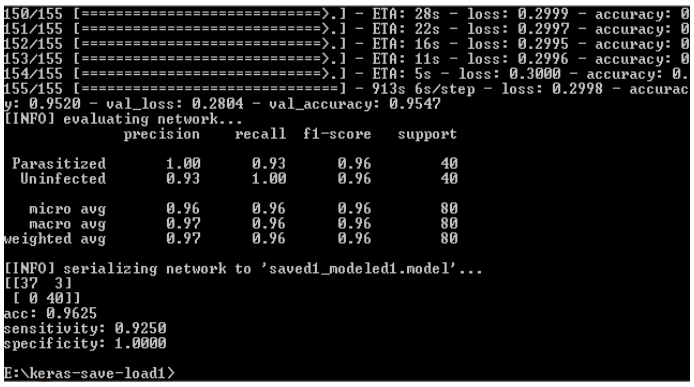


Fig -4: Result analysis of testing data.

This means in Fig. 4

, ● The accuracy for the testing data is 96.25% and the validation accuracy is 95.47%. ● 3 out of 40 (7.5 %) infected cells were classified as clean— False Negative. ● 0 out of 40 clean cells were classified as infected—False Positive.

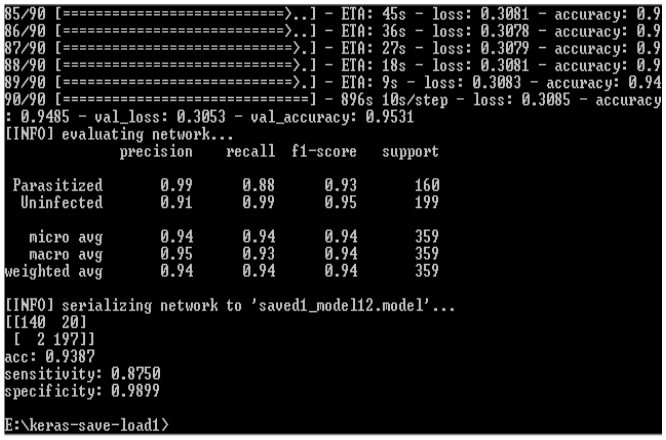


Fig -5: Result analysis of testing data.

This means in Fig. 5,

● The accuracy for the testing data is 93.87% and the validation accuracy is 95.31%. ● 20 out of 160 (12.5 %) infected cells were classified as clean—False Negative. ● 2 out of 199 (1 %) clean cells were classified as infected—False Positive

7. CONCLUSION

There has been a lot of fine-tuning of these methods to make them perform better for blood smear images, and more so for the image analysis methods than for machine learning. There is certainly the potential that some of these methods gain importance outside malaria diagnosis, in particular for preprocessing and for detecting and segmenting red blood cells in other applications.

Yes, there are different models such as AlexNet, VGG- 16 and DenseNet-121 which can also be used to carry out this project. However, they do have certain limitations which can be solved by using ResNet-50. ResNet-50 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 50 layers profound and can categorize images into 1000 object classification, such as keyboard, mouse, pencil, and numerous animals. Subsequently, the system has learned rich feature representations for a wide collection of images. Previous to ResNet, the problem of vanishing gradients was present in training deep neural networks. in addition, ResNet50 has Skip Connections. The center thought of ResNet is presenting an alleged "identity shortcut connection" that skips at least one layer. The proposed method is working well because the use of large amount of thick blood smear image data which suitable for neural network-based method. Using different techniques such as Dropouts and Batch Normalization have helped in solving the problem of over- fitting the model. Utilizing data augmentation has enormously diminished the odds of over-fitting the model just as the quantity of bogus negatives. Utilizing decaying learning rate demonstrated shockingly powerful as the model could arrive at the actual optimal solution. The latency for a solitary image forecast is quite low, which is acceptable considering the fact that this is an image classification difficulty.

Given the wide acceptance of deep learning, the importance of large annotated data image repositories for training is now widely understood, leading to a great support of data acquisition efforts. This will likely lead to larger test suites on patient level, allowing for more standardized evaluations and extensive field testing. Given these developments, automated microscopy is very much in the race toward a cheap, simple, and reliable method for diagnosing malaria

REFERENCES

[1] Surasak Kasetsirikul, Jirayut Buranapong, "The development of malaria diagnostic techniques: a review of the approaches with focus on dielectrophoretic and magnetophoretic methods", Malaria Journal 2016 15:358

[2] Mahdieh Poostchi, Kamolrat Silamut, Richard J Maude, Stefan Jaeger, George Thoma, "Image Analysis and Machine Learning for detecting Malaria", Translational Research Vol. 194, April 2014

[3] Sivaramakrishnan Rajaraman, Kamolrat Silamut, Md. A. Hossain, I. Ersoy, Richard J. Maude Stefan Jaeger, George R. Thoma, Sameer K. Antani, "Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images", Journal of Medical Imaging 5(3) 034501 August 2018

[4] Faza Maulaf Azif, Hanung Adi Nugroho, Sunu Wibirama, "Detection of malaria parasites in thick blood smear: A review", Communications in Science & Technology 3(1) (2018) 27-35

[5] A. Sairaj Bharadwaj Reddy & D. Sujitha Juliet, "Transfer Learning with ResNet-50 for Malaria Cell-Image Classification", International Conference on Communication and Signal Processing