DETECTION OF MALARIA ON AN ERYTHROCYTE USING DEEP RESIDUAL NETWORK

Prepared By:

SAMIR MOUHAMAD MOKTAR

Student No. 154449

B.Sc. Eng. CSE

Supervised By:

Rafsanjany Kushol

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT) THE ORGANISATION OF THE ISLAMIC COOPERATION (OIC)

BOARD BAZAR, GAZIPUR, DHAKA BANGLADESH

NOVEMBER, 2019

DETECTION OF MALARIA ON AN ERYTHROCYTE USING DEEP RESIDUAL NETWORK

By

Samir Mouhamad Moktar

Student No. 154449

B.Sc. Eng. CSE

Supervised By:

Rafsanjany Kushol

Submitted in partial fulfillment of the requirements for the Degree of Bachelor of Science and Engineering with specialization in Computer Science and Engineering at the Department of Computer Science and Engineering (CSE)

Islamic University of Technology (IUT) Dhaka, Bangladesh

November, 2019

Declaration

This is to certify that the work presented in this thesis is an original work of me, **Samir Mouhamad Moktar,** under the supervision of **Mr. Rafsanjany Kushol,** Department of Computer Science and Engineering (CSE)., Islamic University of Technology (IUT), The Organization of Islamic Cooperation (OIC), Dhaka, Bangladesh. This work has not been submitted to any other institution for any other degree.

……

Samir Mouhamad Moktar
Student No. 154449
B.Sc. Eng.

……

Mr. Rafsanjany Kushol
Supervisor
Lecturer (On Study Leave) CSE Department IUT

Dedication

This project is dedicated to the Muslim community all around the world, for their best wishes and prayers, our beloved parents, for their endless love, absolute care and guidance's.

Moreover, I would like to dedicate this work to Tchatima Alifa, a dear friend of mine and Ngnawen Vermont Samuel who has instilled in me the love of computer science.

Acknowledgment

We would like to express our gratitude to the Almighty, the Most Gracious and the Most Merciful for His blessings bestowed upon us in completing this thesis.

. We are endlessly grateful to member states of the Organization of Islamic Co-operation (OIC) which granted us with scholarships opportunity to obtain a career in the world of technology, and indeed the host country, people republic of Bangladesh.

We would like to heartily thank **Mr. Rafsanjany Kushol** for the support, valuable comments and availability during the completion of this thesis.

Finally, we would like to thank our families and friends for being helpful and supportive throughout our studies in Bachelor's degree at Islamic University of Technology (IUT).

Moreover, I am grateful to Ousman Badjie for his tremendous help in regard to how to conduct a research.

Abstract

In the recent years, many researches have evolved in the field of CAD particularly in malaria diagnosis. Malaria diagnosis encompasses many domains, such as labelling of blood samples, detection of infected cells and furthermore determining the development stage and parasite type. Various techniques were implemented, among them, Deep learning has recently caught the attention of researchers. In this research project, we have implement ResNet-50, a variant of Deep Residual Learning to detect the presence of malaria on an erythrocyte. For this experiment, a dataset of 27,560 Red Blood Cells from NIH has been used. Samples were equally distributed. The Model achieved a training, validation and testing accuracy respectively of 96.50%, 96.78% and 97%. The whole process took 1 hour, and training-validation lasted for 50 epochs.

Table of Contents

Table of Figures

Table of Tables

Chapter 1
Introduction

1. Background of the Study

Malaria is a life threatening disease that is responsible of thousands of victims per year. According to the report of World Health Organization (WHO) in 2017[19], an estimated 219 million casualties were reported worldwide compared to 239 million cases in 2010 and 217 million cases in 2016. Out of the 219 million registered cases in 2017, an estimated of 435 000 cases were reported dead. The aforementioned statistic shows how serious and rampant the disease is. Most of the tropical regions around the world are malaria-prone, and due to their economic burden, it makes hard for them to actually overcome malaria. In order to tackle this deadly disease, physicist have invented some diagnosis tools such as blood smear and Rapid Diagnosis Test(RDT) which are the most frequent ones. Blood smear consists of visually computing the parasitemia of a giemsa-stained blood film through a microscope. According to the WHO counting protocol [20], a clinician may count manually up to 5000 cells, which is extremely tiresome, time-consuming and probably less efficient at a long run. In order to make the process a bit faster, scientist came up with a new tool, known as RDT. While being significantly faster than blood smear, RDTs are less accurate and also not affordable by most of the malaria-prone regions. So the ideal solution will be to combine the accuracy (if possible more) and the speed of RDTs. On the way of designing such a solution, we are trying to develop a system based on CAD technology that will classify (as infected of normal), further identify the plasmodium development stage and label different cells of a blood film. The objective of labeling is to provide additional information to the physician for detecting only RBCs, as there are required for parasitemia.

Thick blood smear is suitable for Malaria Parasite detection while Thin for species and life-stage classification.

Keywords: CAD, parasitemia, plasmodium, RDT, blood film

2. Definitions

2.1 Malaria

Malaria is a mosquito-borne infectious disease that affects humans and other animals. It causes symptoms that typically include fever, tiredness, vomiting, and headaches.

In severe cases it can cause yellow skin, seizures, coma, or death. It is a deadly disease that affects mostly under-developed countries. As stated in the above section, various diagnosis techniques have been invented. Each of them has its own benefits and drawbacks.

2.2 Erythrocyte

Commonly known as Red Blood Cell (RBC), it is a red biconcave cell that carries oxygen in the body. The malaria parasite attacks this type of cell in order to weaken the body.

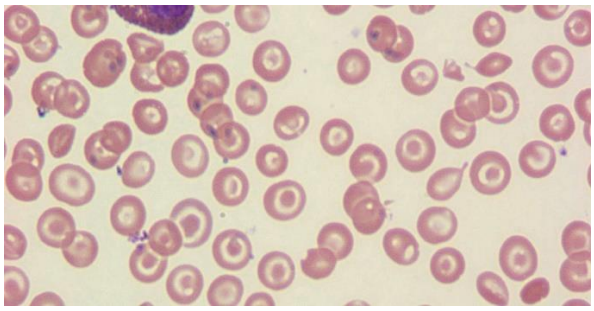

Figure 1 set of erythrocytes on a blood film

2.3 Parasitemia

It is a process of demonstrating the presence of malaria in a body. There are various techniques for doing so.

2.4 Parasite Type

2.4.1 Falciparum

Plasmodium falciparum is a unicellular protozoan parasite of humans, and the deadliest species of Plasmodium that causes malaria in humans

2.4.2 Vivax

Is the most frequent and widely distributed cause of recurring malaria. Although it is less virulent than Falciparum, the deadliest among the malaria parasites, *vivax* malaria infections can lead to severe disease and death.

2.4.3 Malariae

It is found worldwide and causes a so-called "benign malaria", not nearly as dangerous as that produced by P. falciparum or P. vivax. The signs include fevers that recur at approximately three-day intervals.

2.4.4 Oval

It is also responsible for malaria but rare compared to Falciparum and Vivax. Moreover, it is less dangerous than Falciparum.

2.5 Parasite development Stage

2.5.1 Trophozoite

It is the growing stage of the malaria parasite, when it absorbs nutrients from the patient.

2.5.2 Schizont

It is the young stage of the malaria parasite when it is undergoing a nuclear division. This stage is characterized by the infected cell being ring-shaped

2.5.3 Gametocyte

It is the stage at which the malaria parasite divided into gametes are formed in the human body. Though the parasite at this stage is not responsible for clinical symptoms, it ensures the transmission of malaria to another host.

2.6 Artificial Neural Networks (ANNs)

ANNs are nonlinear model motivated by the physiological architecture of the nervous system. They involve a cascade of simple nonlinear computation, when aggregated can implement robust and complex nonlinear function. In fact, depending on how they are constructed, ANNs can approximate any nonlinear function, making them a quite powerful class of models. Multiple-layered neural network are efficient at approximating arbitrary function than other methods.

In the recent years, ANNs that use multiple stages of nonlinear computations (also known as Deep Learning) have been able to obtain outstanding performance on an array of complex tasks, ranging from visual recognition to natural language processing.

2.6.1 Machine Learning

It is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without being explicitly instructed, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task. Machine learning model improves their performance through experience.

2.6.2 Deep Learning

It is a class of machine learning algorithms that uses multiple layers to progressively extract higher level features from the raw input. In contrast to machine learning, where features are extracted before classification, deep learning combine both feature engineering and classification. The deeper the network is, the better its performance will be.

2.6.2.1 Convolutional Neural Networks (CNNs)

It is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN). They have applications in image and video recognition, image classification, medical image analysis and natural language processing.

2.7 Feature Extraction/Selection

Features are essential in any prediction task, so the more relevant they are, the higher the accuracy will be. Irrelevant or less-contributing features used during training a ML model have huge impact on the model performance, in order to attain high performance, one needs to select features carefully. The process of identifying relevant features that provide higher accuracy is known as feature selection. Basically it is a process during which you select features that contribute much to the prediction variable you are interested in. feature selection has some benefits such as reducing over-fitting, improves accuracy and reduce training time.

We may consider a set of feature {a, b, c, d, e, etc.} used to classify a cancer as benign or malignant. Let say {a, d, e} represents {height, skin color and income}, definitely those attributes are irrelevant to the above scenario, hence reducing the accuracy. Thus by applying feature selection, one can discard {a, d, e} from the feature set. However relying too much on relevant variable, there is a risk of over-fitting, so there must be a balance.

Chapter 2
Literature Review

Driven by the motivation to develop affordable, efficient malaria diagnosis many authors worked in this perspective as illustrated in the following a brief recapitulative of the research contributions between 2016 and 2017.

Serial No	Year of Publication	Description	Characteristics	Performance	Comments
1	2016	- Proposed a multilayered-ANN based on radial basis function feed-forward network trained with back-propagation rule. - This model will classify an RBC as infected or healthy.	-Data set consist of 48 RBCs. - Feature set is made of six statistical/physical parameters (area, circularity, standard deviation, mean, perimeter and covariance) extracted from preprocessed holographic images of RBC. - Experiment was carried out on matlab. - 80% of samples used for training and 20% for testing.	- Achieved a 90% accuracy during testing.	-data set is too small. - in this regards, the paper suggested that the performance may increase when used on bigger dataset with more parameters
2	2016	- They developed 16-layer CNN for detecting malaria. -during the study, they compared their model with another method that uses pre-trained (based on transfer learning) model as feature extractor and SVM as classifier.	- Data set has 27,578 samples of curated RBCs. - The CNN was developed using matconvnet toolbox. - implemented 10-fold cross-validation technique on the whole data set for training and validation. - 90% of the samples were used for training and 10% for testing.	- The baseline had an accuracy of 91.99% while the 16-layer CNN achieved 97.37%	- The 16-layer CNN is believed to be efficient as it outperforms the bassline model.
3	2016	- Design a model for detecting malaria on a thick blood film by identifying trophozoite-infected RBCs. - The approach consist of extracting candidate cell (trophozoite's cytoplasm) using adaptive thresholding technique. The extracted cell is further classifier using a two-class SVM classifier (with RBF kernel).	- Dataset consist of 194 images extracted using a smartphone mounted on a 1000X magnification microscope. - 314 features were combined from geometric color and texture parameters. - 10-fold cross-validation technique was used.	- The model achieved an accuracy of 9138%	- Dataset is too small - As the segmentation is based on chromatin dots, there is a chance to confound a non-infected cell with an infected cell as their chromatin may look alike a earlier stage of malaria.
4	2016	- Proposed a model based on CNN (with a stack of focus) to detect plasmodium falciparum (type of parasite) infected RBCs. -In the meantime they compared their model with a SVM classifier (with linear kernel) and another CNN (with single focus).	- The dataset is made of 765 FoV containing 59,957 cells of which 1,191 are infected. Those images were collected from 41 slide of blood film. - Each focus is 32x32 patches.	- The sensitivity of those models are: 93.12% (SVM), 98.74% (CNN-single focus) and 99.08% (CNN-stack focus).	- It turns out that CNN-stack focus performs better than the other two models.
5	2016	- They proposed an experimental study on 3 diagnosis test: tuberculosis, malaria and intestinal parasite using 4-layer CNN. - The network is trained on single patch. The presence of pathogen on the whole image depends on the patches with high activation score.	- The dataset is made of 7,245 ROIs extracted from 1,185 images (stained thick blood film) using a smartphone mounted on a microscope. Those images were annotated by experts. - 50/50 hold-out validation method was used with 500 epochs. - Test set contains 261,345 patches.	- The model showed better performance according to ROC-AUC curve.	- As the dataset was biased towards negatives samples, flipping/rotation technique was used to augment positive samples.

6	2016	- Design a two-stage ML model to detect from unstained blood cells. - The first stage consists of segmenting candidate cells from the blood smear. They selected COD for this purpose as it performs better than HT. - The second phase is about detecting infected cells. A combination of COD, HT, IT, CT was used for detecting the infected cells. The best performing combination was selected.	- The dataset has 562 templates of blood cells(300 for training and 262 for testing) obtained from 43 images of blood droplets; among them, they were 62 templates of infected cell, 300 templates of uninfected cells and 200 templates of non-cell parts. - ECOC was used as a selector for best feature descriptor. HOG features were found to be the best descriptors compared to PCA and GIST. - Used a 10-fold cross-validation technique.	- The best performing combination infected cells is COD+HT+IT and it achieved a specificity of 86.07%.	- The dataset was small and biased towards uninfected cells and non-cell parts. - An increased dataset is necessary to observe (if possible) better performance.
7	2016	- Conduct an experimental study on classifying RBCs (as healthy of infected) among KNN, SVM, ANN, and Naïve Bayes. - The classification is based on histogram feature set. - The optimal feature set for every classifier was selected as the combination of all feature yielding the highest accuracy.	-The dataset involves in this study had 871 images of an erythrocyte. - Each histogram feature set is represented by 9 statistical features categorized as f1 (Green Channel Histogram), f2 (Saturation Level Histogram), f3 (Chrominance Channel Histogram), f4 (|R-G| Histogram). - A 4-fold cross-validation method was used during training.	- The models achieved following accuracies: ANN (96.32%), SVM (96.09%), KNN (95.29%), and Naïve Bayes (95.86%) - Other performance metrics were used as well.	- The resulting feature set for each classifier are: ANN (f1+f2+f3+f4), SVM (f2+f3), KNN (f3+f4), Naïve Bayes (f2+f3+f4). - It turns out that ANN performs better (with a slight variance) than the others. - The data set is not that much large
8	2017	-Developed two-DL model for classifying RBCs (normal or infected) and detecting the malaria parasite development stage (trophozoite, schizont, ring, gametocyte) or other (unidentified) and leukocyte. - The first stage is labeling cells as RBC or non-RBC from blood smear using Faster R-CNN- the following phase uses ALexNet which classifies the RBC from stage 1 - they also compared this model with a baseline based on traditional segmentation approach and use of RF classifier	- The data is made of 1,300 FoV consisting of 100,000 cells. - The baseline was trained on 300 measurements of intensity, shape and texture features.	- 2 human experts also provided their diagnosis. - The model (Faster R-CNN+ALexNet) showed an accuracy of 98% while the baseline and human experts achieve respectively 50% and 72%.	- The model (Faster R-CNN+ALexNet) showed better performance compared to baseline and human experts
9	2017	-Developed a CNN model to classify infected cell into 3 development stages (ring, schizont, and gametocyte). - This model was also compared to SVM and K-NN.	- The Dataset is diverse and consist of samples of three type of parasites (falciparum, vivax and malariae). - The testing set has 191,519 (with 5,918 infected samples and 185,601 uninfected) cells while training set got 15499 negative samples and 39,374 positive samples. - None of malariae-infected cell was include in the dataset.	- The three models achieved following accuracies per parasite type: falciparum [CNN (98%), SVM (88%), and K-NN (90%)] and vivax [CNN (96%), SVM (86%), and K-NN (90%)]. - CNN performs better with respect to both of the parasite types.	- Though the dataset was quite large, it was biased towards uninfected samples [not equally distributed]. Moreover, testing set was thrice the training set.
10	2017	- Proposed a method for detecting infected RBC along with their estimated count, and classification of parasite into one of the four species (falciparum, vivax, malariae and oval). - Infected RBCs are detected by an algorithm based on morphological features (shape, size) by estimating the roundness of each cell. - Furthermore, The NCC function associated with a threshold of 0.7 was used to classify parasites into corresponding class.	- 160 cell images from different development stages were used for the experiment.	- Achieved an accuracy of 94.97%.	-. raised some limitations, such as, in practice the images need to be normalized to same intensity range in order to avoid uneven background, - NCC is neither rotation or scale invariant. Moreover, the system does not work perfectly unless the template used with NCC have to be of

					same size as the reference image and also similarly oriented. - Threshold might be tuned appropriately to attain better accuracy.
11	2017	- Developed a model based on ANN to classify an erythrocyte as either infected or normal.	-The dataset used throughout the experiment consisted of 1120 RBCs cropped from 77 full blood smear images, with 120 infected samples and 1000 samples as normal. 740 randomly chosen samples were used for training, and the rest plus 244 samples from different images for testing. - A set of two intensity-based features (variance and skewness) is selected as salient features for classification.	- The model achieved a 99. 68% accuracy after comparing the output to the results of microscopist.	-The dataset is a bit small for the model to generalize. - A set of other salient features might be investigated.
12	2017	- Performed a comparative study on 3 CNN architectures (AlexNet, LeNet and GoogleNet) and SVM.	- The data set involved in the experiment consisted of 2565 RBCs (1034 as infected and 1531 as normal), approximately equally divided into training-validation (25% randomly selected for validation) set and testing set. - Those cell images were cropped from a whole slide image of thin blood smear, subjected to wright geimsa staining, later labeled by four pathologist [a cell is classified based on unanimous decision of those pathologist, otherwise excluded]. - SVM was trained on seven handcrafted features which are Hu's moment (7, 5, 2, 3 and 6), MinIntensity and shanon's entropy.	- The models' performance are: AlexNet (95.79%), LeNet (96.18%), GoogleNet (98.13%) and SVM (91.66%).	- Overall GoogleNet outperforms other models probably owning it to its depth that may have extracted higher order knowledge. - Dataset was not equally distributed.
13	2017	-Developed a model to identify specific malaria parasite (P. Vivax and P. falciparum) along with corresponding life cycle (gametocyte and trophozoite) from infected RBCs using a transfer learning on inception V3 CNN.	- The data set consisted of 363 cell images (142 P. vivax and 221 P. falciparum). - Since the data set is small for learning the classifier, those cells were rotated by 15 degree increment to produce additional instances.	- As a result, they obtained training accuracy of 94% as well as validation accuracy of 84.6% for species identification. In top of that, the model yields an accuracy of 92.4% on test set with sensitivity of 95.2% and specificity of 84.7% for parasite detection and 87.9% for species classification.	- though samples were rotated to increase the dataset, it may still be not sufficient for better performance.
14	2017	- Conduct an experiment which consisted of classifying stained cells as either infected or normal using Deep Belief Network.	- The dataset has 4100 cells images (3431 non-parasite and 669 parasite cells). Those images were extracted from blood smears using level set segmentation. - 2122(336 infected cells and 1786 normal cells) samples were used for training, while 1978(333 infected cells and 1645 normal cells) for testing.	- The model yields an accuracy of 96.21%.	- In both training and testing set, there were more positive samples than negative ones.

			- As part of training, they used the concatenation of feature color (histogram-based features and color coherence vector) and texture (GLCM-based textural features, LBP feature and GLRLM-based textural features) as image descriptors due to the fact that pathologist also consider color and texture variation for identifying infected and normal objects in stained a blood smear		
15	2017	- Carried out a study to build a mobile based diagnosis system for malaria that will detect presence of malaria parasite at a ring development stage from a blood sample. - Viola Jones algorithm was used for detecting candidate cells and Adaptive boosting for classification.	- The dataset is made of RGB images (555 positive slides and 777 negative slides). - Blood slide were preprocessed using Gaussian filter (for smoothing) and morphological operations for improving geometric structure of cells. - OTSU segmentation method was used to reduce the search space for detection algorithm by removing background (objects other than RBCs). - Those RGB images were converted to grayscale. - 10-fold cross-validation method was used during learning.	- Observe a performance with 96% of specificity and 91% of accuracy with a 20 stage cascade classifier.	-Adaptive Boosting is sensitive to noise and outliers.

Table 1 resume of some research papers [2016-2017]

The task of malaria diagnosis can be divided into three main subtasks which are: detection of candidate cells, classification of the cell as either infected or normal and determining the development stage of the parasite/type of the parasite. Some of the research papers combine some of these subtasks.

As a summary, some of the related works (5, 7, 9, 11, and 12) have used conventional image processing techniques for detecting candidate cells annotated by an expert from a stained blood smear. The problem that may arise with this approach is that there is a likelihood of identifying non-RBC object as candidate cell, for instance platelets, when stained look like erythrocyte. By candidate cell, we mean RBCs needed for parasitemia. However, [6], [1] used Viola Jones detection algorithm for labeling candidate cells. Viola Jones has been discovered more than a decade (since 2003), thus is efficiency may be questioned. Hence [8] proposed a cell labeling approach using a novel object detection algorithm called Faster R-CNN.

Pertaining to classification, some of the contributions (1, 2, 7, 12, 10, and 14) performed mainly binary classification, i.e. a cell is classified as either infected or normal. This approach is not enough for malaria diagnosis, as neither the development stage (**trophozoite, schizont, and gametocyte**) of the parasite nor the type of parasite (**falciparum, vivax, malariae and oval**) is determined. Life stage and nature of parasite are really necessary for medicines prescriptions. To alleviate to the aforementioned limitation, (3, 4, 8, 9 and 13) provided further classification by identifying parasite development stage. Moreover some works (1, 3, and 7) mainly provided classification based of handcrafted features. However, (2, 4, 8, 9, 12, 13, 14) implemented CNN architectures that eliminated the need for features extraction beforehand. Moreover, [12] evaluated the performance of three Deep learning models and concluded that all of them achieved an accuracy above 95% compared to SVM which had 92%. This study proved the better performance of Deep learning over Machine Learning approaches. However, [16] argues that some deep learnings models suffers from a degradation problems, thus need for residual learning approach.

Chapter 3
Proposed Approach

1. Proposed method

Our objective is to build a fully automated system for malaria diagnosis, in the sense that a physician does not need any more to identify the erythrocytes from a stained blood smear that will be later classified as infected or not. Basically with the object detection model, cells on a stained blood film will be labeled automatically, and therefore the user have to cropped RBCs only and later fed them into the classifier. With this approach, there is a chance of reducing errors that may be committed by microscopist while cropping actual RBC instead of other cells as proposed by some authors in the literature review (…). In addition to that, infected cells will be further classified by providing their development stage as well as type of plasmodium, which plays an important role in prescribing medicines.

The prototype will consist of 2 phases: the first phase will be about **cell detection from a blood film** and the second one concerns the **Parasite detection and identification of development stage**

1.1 Stained blood film labeling (Cell detection)

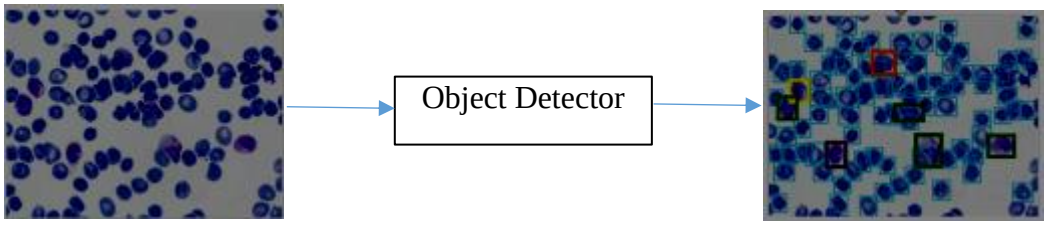

Figure 2 unlabeled *Figure 3 labeled*

In this regard we will investigate cell labeling using other object detection algorithm such as **YOLO**, **SSD** (Single Shot MultiBox Detector) and **R-FCN** (region fully connected network) with the aim of achieving higher performance.

1.2 Parasite detection and development stage identification

Our method will classify a given blood erythrocyte as infected or nor along with type of plasmodium and development stage. Those erythrocytes are detected by the object detector then cropped by the user, which are later fed into the classifier for further investigation. As most the

research paper have worked with some CNN architecture, like **DBN**, **AlexNet**, and **customized CNN**, little consideration was given on **ResNet**, **Inception V4 and DenseNet**. Therefore, we will carry out the experiment on some of those CNN architectures, and select the one achieving better performance.

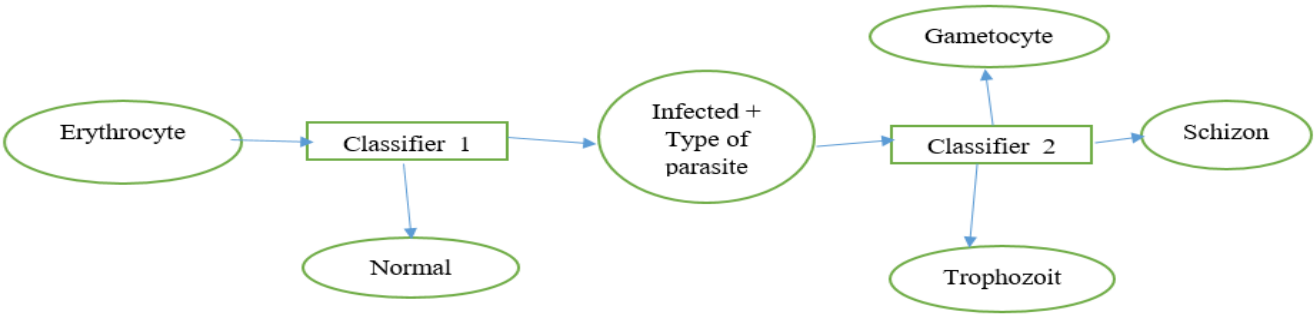

Figure 4 initial model

Due to some limitations such as:

- Unavailability of data set

  We could not find a dataset for blood samples ready for the experiment on object detection. Though some research papers have conducted cell detection [8]. Their source was not mentioned. Moreover we did find prepared dataset but only for infected/non-infected RBC. Data set for type parasites and development stages mentioned in [10] and [13] were not accessible.

- Time constraint

  As the project is a bit large, the allocated time will not be sufficient to achieve it.

We therefore decided to implement a portion of this project. In this report, we will focus on the classification of erythrocytes. Hence the refined method is as follow:

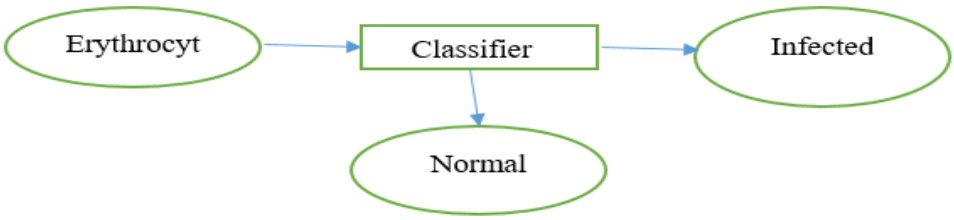

Figure 5 refined model

2. Definitions

2.1 Deep Residual Learning

Since the inception of Deep learning, the performance of a model is related to the depth of its network. The deeper the network is, the greater its accuracy. However, [16] argue that there is a degradation problem associated with the depth of a network. The degradation problem is defined as when a deeper network start to converge, its accuracy gets saturated then eventually degrades rapidly [16]. In order to resolve the issue, [16] proposed a learning paradigm called residual learning. Residual Learning consist of adding a residual function after a block of layers. This residual function is expressed as the sum of the input X and the output F(X) of the block of convolutional layers of that input. The resulting output of the residual function will act as an input to the next block of convolutional layers. A Deep Residual Learning can be defined as applying the residual learning concept in a deep network. [16] Designed new networks that implement the Deep Residual Learning. Among them we have ResNet (residual network) with 50 layers, 101 layers, 152 layers, etc.… Due the Deep Residual Learning framework, [16] won 1st place in ILSVRC 2015 classification competition, ImageNet Detection and Localization. [16] Proved that deep residual network perform better than simply deep network.

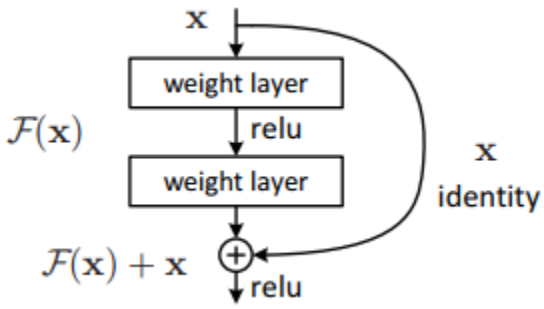

Figure 6 Residual Learning: a building block

2.1.1 ResNet-50

ResNet-50 is a 50-layer deep residual network where each block is 3-layer deep. It has the ability to classify images up to 1000 classes.

The architecture of ResNet-50 is as follow:

Layer Name	Output Size	
Conv1	112x112	7x7, 64, stride 2
		3x3 max pool, stride 2
Conv2_x	56x56	[1x1, 64] [3x3, 64] x 3 [1x1, 256]
Conv3_x	28x28	[1x1, 128] [3x3, 128] x 4 [1x1, 512]
Conv4_x	14x14	[1x1, 256] [3x3, 256] x 6 [1x1, 1024]
Conv5_x	7x7	[1x1, 512] [3x3, 512] x 6 [1x1, 2048]
	1x1	Average pool, softmax

Table 2 architecture of ResNet-50

2.2 Training, Testing and Validation

2.2.1 Training

Training (or Learning) is a process by which a model (or hypothesis) adapts changes by adjusting its mutable parameters (or weights). Those parameters change in such a way that there is an association between a set of inputs and desired outputs. In case of supervised learning (eg: logistic and linear Regression), we can use Gradient Descent (or Linear algebra to find θ_s with respect to X and) to tune parameters θ_s so that hypothesis $h(\theta_s) = X^T\theta$ can related X to Y (desired output).

For example, considering the scenario above, training will consist of using a large (usually 80-90%) portion of the labeled dataset. During the process, unknown weights $\theta_s$are associated to every features $x_i \in X$ and then later equalized to Y (where X represents set of features while Y stands for set of labels) per sample. The ideal values of those weights will be determined based on all the samples used so that they can fit the prediction model.

2.2.2 Testing

Testing in relation to ML is a process of evaluating the performance of the model in terms of some metrics like accuracy, precision, etc. During the process, a portion of the dataset that was kept aside earlier before training is feed into the model; for every predicted output, one can observe

either it is a **TP** (True Positive), **TN** (True Negative), **FP** (False Positive), and **FN** (True Negative). At the end, we collect the values of TP, TN, FP and FN to compute the performance metrics and later judge if the model is well trained enough or there is any need for improvement.
For instance, 10% of the dataset will be used for testing the new model.

2.2.3 Validation

A model performs well on the training set (faces problem of over-fitting and under-fitting) as it was trained on that set, but the prediction accuracy is measured based on the performance on an independent set (validation set). Basically a model is validated base on its performance on a new data set completely different from the training set. So there is a need for a validation method that judges how well a model is performing.

Cross-validation is a validation technique that assess the result of a prediction model on an independent set. It is a rule of thumb that both training and validation set shall be drawn from the same distribution. Its goals is to define a training and testing or validation set in order to avoid aforementioned problems. There are several types of cross-validations techniques, among them we have: **k-fold**, **holdout** (or train/test split) and **stratification**.

For example, if we consider task of classifying an image of an animal as either cat of dog with 2-fold cross validation. Assuming we have a data set of 3000 samples, the k-fold cross-validation means that we will have k rounds of training/testing (no matter the size of the training or testing set). During each round the model will be trained on k-1 of the data set and tested on 1 set. After every round the training/testing set are renewed by reshuffling the data set. With this approach each sample is used once for training and get to be used exactly once for testing. As a result, under-fitting is reduced as we are using most of the data point for training and over-fitting is reduced as well as most of the data point are used for validation.
Similarly for k-fold, each samples is used k-1 times for fitting and exactly once for validation.

This method is good when we do not have enough data for training. 10-fold cross-validation is a commonly used.

2.3 Performance Analysis

There are various performance analysis metrics in the literature such as ROC and confusion matrix (of different dimension).

2.3.1 Confusion Matrix (2-by-2)

Given a classifier and a set of instances from the test set, a 2-by-2 confusion matrix (also known as contingency matrix) can be constructed, representing the disposition of the set of instances.

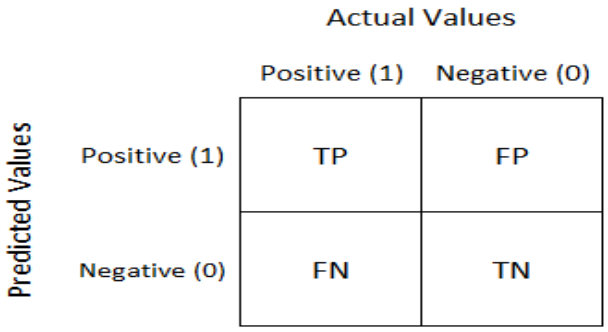

Figure 7 2-by-2 confusion matrix

Numbers along the diagonal (TP, TN) represent the correct decision made, and the numbers off the diagonal represents the errors. Following are some common metrics derive from the confusion matrix.

$$\boldsymbol{Sensitivity/Recall/\text{True Positive rate}} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\textbf{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

$$\textbf{False Positive rate} = \frac{\text{FP}}{\text{TP+FN}}$$

$$\textbf{True Positive rate} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\textbf{Specificity} = \frac{\text{TN}}{\text{FP+TN}}$$

$\textbf{Accuracy} = \frac{\text{TP+TN}}{\text{P+N}}$, where P denotes total positive and N, total negative.

In other words……

3. Setting up the Environment

Computer	HP PROBOOK
Operating System	Window 10 Pro, 64 bits
Hardware	CPU: Intel core i5, 8th Gen, @ 1.60~1.80 GHz
	GPU: NVIDIA GeForce 930MX
	RAM: 8 GB
Deep Learning Framework	Front-end: Keras

	Back-end: Tensor Flow
Numerical Processing and ML Packages	NumPy & Scikit-learn
Plotting tool	Matplotlib
Image Processing and Deep Learning Convenience	Imutils

Table 3 characteristics of the environment

Chapter 4
Experiment

1. Experiment

The experiment consist of Determining whether a given RBC is infected or not.

1.2 Dataset

The dataset for this task was collected from NIH[1] [17]. It consists of 27,560 images belonging to two separate classes:

1. **Parasitized:** Implying that the region contains malaria.
2. **Uninfected:** Meaning there is no evidence of malaria in the region.

The number of images per class is equally distributed with 13,780 images per each respective class.

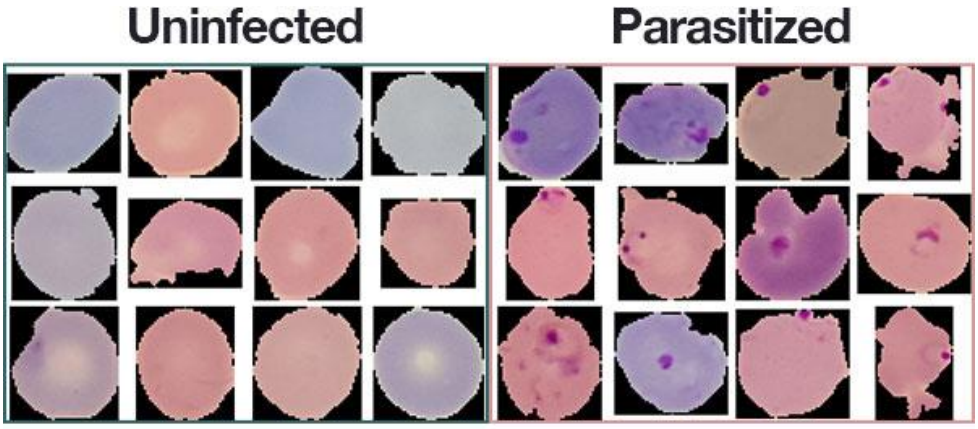

Figure 8 RBC from the data set

80% of the samples will be used for training and the remaining 20% for testing. Moreover, 10% out the training set will be reserved for validation purpose.

[1] National Institute of Health, USA

2. Result Analysis

After implementing the network, we obtain following statistics:

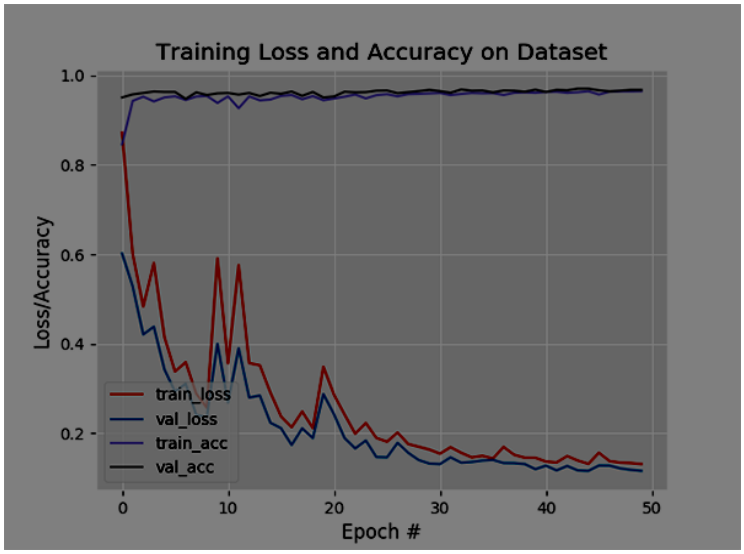

Figure 9 training-validation loss/accuracy curves

Metrics	Training	Validation	Testing
Accuracy	**96.50%**	**96.78%**	**97%**
Training duration	1 hour, 50 epochs		

Figure 10 results

Chapter 5
Conclusion and Future Work

1. Conclusion

As per the literature review, many research has been done for malaria diagnosis in general. Some have focus on classification of candidate cells as either normal or infected, others have worked on detection of erythrocyte from a blood smear and furthermore, determining the type of parasite and development stage of the malaria. Those research contributions have investigated the task of malaria diagnosis using numerous techniques, ranging from image processing, Machine Learning to Deep learning. Most of these approaches have their drawbacks as well as benefits as stated in the literature review.

Though Deep learning approach has been the backbone of CAD in the recent years, [16] has raised some limitations associated with it. In this regards, we decided to carry out the experiment on detecting malaria using deep residual Learning. During this research, we used a dataset of curated Red Blood Cells from NIH. As a result, we obtained a testing accuracy of 97% which is better compared to most of the methods proposed in the literature review.

2. Future Work

As stated earlier, malaria diagnosis encompasses many subtasks. Among them, we have only work on the detection of malaria.

Our motivation was to design a fully automated malaria diagnosis system, meaning from cell detection from stained blood film, detecting the presence of malaria and determining the type of parasite involve and eventually the development stage of the parasite. As we have mentioned in chapter 3, we have encountered some limitations.

As for future work, we would like to solve the challenges (in chapter 3) in order to achieve our goal. Moreover, as performance is concerned, it will be wise to carry out the experiment with other networks and/or other learning approach to increase the accuracy if possible.

References

[1] Artificial Neural Networks for Detection of Malaria in RBCs [2016]

[2]CNN-Based Image Analysis for Malaria Diagnosis [2016]

[3]Automated detection of malaria parasites on thick blood smears via mobile devices [2016]

[4]Automatic Detection of Malaria Infected RBCs from a Focus Stack of Bright Field Microscope Slide Images [2016]

[5]Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics [2016]

[6]Image Classification of Unlabeled Malaria Parasites in Red Blood Cells [2016]

[7]Malaria Infected Erythrocyte Classification Based on the Histogram Features using Microscopic Images of Thin Blood Smear. [2016]

[8]Applying Faster R-CNN for Object Detection on Malaria Images [2017]

[9]Automated plasmodia recognition in microscopic images for diagnosis of malaria using CNN [2017]

[10]Detection and Classification of Malaria in Thin Blood Slide Images [2017]

[11]Detection of Malaria Parasites Using Digital Image Processing [2017]

[12]Evaluations of Deep CNN for Automatic Identification of Malaria Infected Cells [2017]

[13]Malaria Parasite Detection and Species Identification on Thin Blood Smears using a CNN [2017]

[14]Malaria Parasite Detection from Peripheral Blood Smear Images using Deep Belief Networks [2017]

[15]The Malaria System MicroApp: A New, Mobile Device-Based Tool for Malaria Diagnosis [2017]

[16] Deep Residual Learning for Image Recognition

[17] https://lhncbc.nlm.nih.gov/publication/pub9932

[18] https://www.pyimagesearch.com/2018/12/03/deep-learning-and-medical-image-analysis-with-keras/

[19] World Malaria Report 2017, WHO

[20] Malaria Counting Protocol, WHO

ACRONYMES

NIH: National Institute of Health

RBC: Red Blood Cell

ML: Machine Learning

DL: Deep Learning

ECOC: Error Correction Output Code

COD: Cascade Object detection

HT: Hough Transform

IT: Intensity Thresholding

CT: Color Thresholding

WHO: World Health Organization

ResNet: Residual Network