Master of Science Thesis

Automated malaria diagnosis using convolutional neural networks in an on-field setting

The analysis of low quality smartphone based microscope images

R. Sorgedrager

January 9, 2018

DCSC

TU Delft

Delft Center for Systems and Control

Delft University of Technology

Automated malaria diagnosis using convolutional neural networks in an on-field setting

The analysis of low quality smartphone based microscope images

Master of Science Thesis

For obtaining the degree of Master of Science in System Control at Delft University of Technology

R. Sorgedrager

January 9, 2018

Faculty of Mechanical, Maritime and Materials Engineering · Delft University of Technology

Delft University of Technology

Delft University Of Technology
Delft Center for Systems and Control

The undersigned hereby certify that they have read and recommend to the Faculty of Mechanical, Maritime and Materials Engineering for acceptance a thesis entitled **"Automated malaria diagnosis using convolutional neural networks in an on-field setting"** by **R. Sorgedrager** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: January 9, 2018

Supervisor:

Prof.dr.ir. M. Verhaegen

Readers:

Prof. Dr. G. Vdovin

Dr.ir. J. Kober

Dr. ir. S. Baldi

Abstract

This study focuses on automated malaria diagnosis in low quality blood smear images, captured by a low-cost smartphone based microscope system. The aim is to localize and classify the healthy and infected erythrocytes (red blood cells) in order to evaluate the parasitaemia in an infected blood smear. Due to the lower quality of the smartphone microscope system compared to traditional high-end light microscopes, conventional algorithms fail to process these images. We propose a framework using a convolutional neural network as a pixel classifier to localize the erythrocytes. Afterwards we classify them accordingly, using a convolutional neural network as an object classifier. Such a system can offer in-the-field malaria diagnosis without human intervention or can act as an aid for human experts to lower workload and increase diagnosis accuracy. The algorithm successfully localizes the erythrocytes with an average sensitivity of 97.31% and precision of 92.21%. Classification performed inadequate, in terms of low agreement with two human experts. This can be due to the low image quality or the small amount of training data available at the time.

Table of Contents

List of Figures

Acronyms

DUT	Delft University of Technology
OSMD	Optical Smart Malaria Diagnostic
RDT	Rapid Diagnostic Test
SVM	Support Vector Machines
MLP	Multilayer Perceptron Network
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
ReLU	Rectified Linear Unit

Chapter 1

Introduction

Nearly half of the world population is at risk of malaria [WHO, 2016]. Malaria is a serious disease caused by a peripheral blood parasite of the genus Plasmodium, transmitted to humans via mosquitoes [Sherman, 1979]. In 2015 alone, the global tally of malaria reached 212 million cases of which 429 000 fatal. Most of these deaths occurred in the African region (92%) and the vast majority is due to P. falciparum malaria (99%) [WHO, 2016]. Early and accurate malaria diagnosis and prompt treatment can cure a patient, preventing severe malaria cases and possible fatal disease states [WHO, 2016] [Blumberg and Frean, 2007]. All around the world there are still millions of people lacking access to malaria prevention and treatment [WHO, 2015].

Recent developments made by the Optical Smart Malaria Diagnostic (OSMD) research team of the Technicical University Delft shows promising potentials in malaria diagnostics. The smartphone based imaging system developed by the team [Agbana et al., 2017], can mean drastic improvements to the cost, availability and readiness of malaria diagnosis in low-resource settings.

However, the problem of optical malaria diagnosis is a twofold. Not only the costs of purchasing and maintaining microscopes are high [Carpenter et al., 1991], trained personal is scarce and also expensive since the technique is labour intensive and time-consuming [Uzochukwu et al., 2009]. The accuracy of optical malaria diagnosis is ultimately determined by the expertise and reader technique of the microscopist and quality of the blood smear [O'Meara et al., 2006]. While theoretically very high sensitivity and specificity can be reached using microscopy analysis, practice shows big discrepancies between different technicians and on-field versus laboratory diagnosis results [O'Meara et al., 2006][McKenzie et al., 2003].

Automated image analysis software could remove the most serious limitation of the new smartphone based microscope system and microscopy in general, dependency on human expert performance for the diagnostic accuracy of the results. This software is not new, different image processing techniques have been applied in the past to tackle this problem. Most of these techniques are dedicated to the quantification of parasites with respect

to the total amount of erythrocytes (red blood cells). Conventional techniques like morphology [Di Ruberto et al., 2002][Ross et al., 2006], edge detection [Sio et al., 2007], region growing [Theerapattanakul et al., 2004] and more, have reported positive results with respect to their data. However all these researches are based on image data acquired using digital cameras coupled to high-end light microscopes, examining Giemsa (or similar) stained thin blood smears under oil immersion. This data acquisition results in high resolution, highly detailed, uniform illuminated, altogether high quality images, and even in this type of images inconsistent light intensities are causing problems [Ross et al., 2006]. A low cost device using a smartphone camera to capture images will inherently produce inferior images, and these conventional techniques will not be sophisticated enough to analyze these images.

1-1 Aims and objectives of this thesis

In this thesis a novel approach to the automated malaria diagnosis problem is presented. We show a two-phase procedure to localize erythrocytes and classify them as healthy or infected in relative low quality thin blood smear images. Localization of the erythrocytes is done using a Convolutional Neural Network (CNN) as a pixel classifier [Cireşan et al., 2013]. The classification of individual erythrocytes is done using a CNN as an object classifier. The system must be able to provide a quantification and exact locations of healthy and infected erythrocytes shown in different fields of view captured using the smartphone camera. Such a system would either enable a fully automated diagnosis or act as an aid to improve accuracy and lower workload of human experts.

To show if the algorithm is an applicable diagnostic tool, the performance will be evaluated against two human experts. From these results, recommendations for the future will be made.

1-2 Thesis Structure

The thesis is organized as follows. Chapter 2 gives background information on malaria diagnosis, neural networks and the acquired images from the smartphone based microscope. We will also show the performance of conventional algorithms on this data, which serves as a motivation for this study. Chapter 3 proposes the method, including the two-phase approach using convolutional neural network to localize erythrocytes and elaborates on the classification problem. At the end of this chapter, information on algorithm efficiency along with a proposed improvement is given. In chapter 4 the experimental results can be found, both the localization and classification phases are tested against two human experts. Chapter 5 consists of the discussion and conclusion of this work.

It is important to note that the results of this thesis work has been summarized for publication in a peer review journal. Before publication the content and results of this paper will be subject to change. To provide the reader with a more compact alternative to thesis we refer to appendix A, containing the version of the paper as this thesis is written. The paper is advised for readers that have most background information readily available. This thesis will however provide an outlined procedure and detailed implementation steps for readers with little or no theoretical background in deep learning algorithms.

Chapter 2

Background and Thesis Motivation

2-1 Malaria Diagnosis by Human Experts

Rapid and precise diagnosis is essential to address the global burden of malaria. Low-cost malaria diagnosis can roughly be divided into two classes, microscopy diagnosis and antigen detection. The performance of diagnostic tools is normally measured in two parameters, sensitivity and specificity. Sensitivity expresses share of true positives correctly identified as such. Specificity is the share of true negatives correctly identified as such. Using these two proportions the diagnosis ability of a test can be evaluated [Altman and Bland, 1994]. To be an applicable diagnostic for malaria, a test must achieve greater than 95% sensitivity and a specificity of at least 90% [WHO, 2000]. In this section both diagnostic methods are summarized with their respective advantages and disadvantages

2-1-1 Microscopy Malaria diagnosis

Microscopy examination of stained thick and thin blood smears is seen as the most suitable diagnostic instrument for malaria. It is regarded as the gold standard for malaria diagnosis [Moonasar et al., 2007]. Blood is collected from a patient, stained with a Romanovsky stain (most often Giemsa), which is a dye to make DNA visible, and spread over glass as either a thick or thin blood smear. This smear is examined under a light microscope with a 100x oil immersion objective [WHO, 2010]. Thick blood smears contain more blood compared to thin ones, and are used to determine the presence of malaria parasites. If malaria parasites are found, thin blood smears (example shown in figure 2-1) are used to differentiate between malaria species and to estimate the parasitemia, the amount of infected erythrocytes to healthy erythrocytes. Parasitemia is used as an indication for the severeness of the disease.

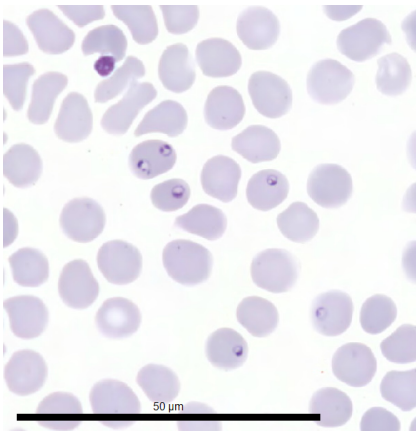

Figure 2-1: Example of a thin blood smear, the dark purple rings inside the erythrocytes are malaria infections. Image taken with a digital camera coupled to a light microscope, 100x magnification under oil imersion. Taken from the Mamic database, see Linder [Linder et al., 2014] for source

Microscopy has numerous advantages. It yields the lowest cost per test [Shillcutt et al., 2006], supports direct parasite count and discrimination between different species. Results can be achieved with very high sensitivity and specificity, both close to 100%.

The main disadvantage is dependency on the human expert. The technique requires skill and experience that is scarce even in developed countries [Blumberg and Frean, 2007][Wongsrichanalai et al., 2007]. The accuracy is ultimately determined by the expertise and reader technique of the microscopist and quality of the blood smear [O'Meara et al., 2006]. A recent study shows big differences in parasitemia estimation between different technicians. The differences are even bigger between on-field versus laboratory diagnosis results [O'Meara et al., 2006][McKenzie et al., 2003].

2-1-2 Antigen Malaria Diagnosis

An antigen is a protein expressed by bacteria or virus to which the immune system can respond by producing specific antibodies. Malaria has its own specific antigens, these can be detected using Rapid Diagnostic Tests (RDTs). These 'dipstick' tests, also known as lateral flow immunochromatographic assays, are simple tests that can be used by persons without much diagnostic expertise [Wongsrichanalai et al., 2007].

The advantage of RDT-based diagnosis is its relative ease to support at peripheral health facilities compared to appropriate microscopes including trained personnel. This means it may facilitate access to diagnosis in areas where this was not possible before [Shillcutt et al., 2006]. These days it can also achieve high sensitivity ($> 95\%$) and relative high specificity ($> 90\%$) [WHO, 2010].

Unfortunately, their results depend heavily on the storage, quality assurance and end user training [Moonasar et al., 2007]. So despite being able to be used by people without knowledge of malaria or diagnosis in general, the usage is not trivial. Also, they cost more (around two times the cost per test) compared to microscopy [Shillcutt et al., 2006].

Despite some advantages, current RDTs are not intended to replace microscopy.

2-2 Previous Work on Automated Malaria Detection

There have been numerous studies aimed at automating the malaria diagnosis process. All these studies are an extension to microscopy diagnosis, they work towards automatic quantification of thin smeared blood films. Most (earlier) studies show results on an image level (i.e. microscope field-of-view) [Tek et al., 2009], while more recent studies also attempt to diagnose malaria on a patient level [Linder et al., 2014].

On a general level the automated techniques work in the same way. The first step is to segment the erythrocytes from the background. Afterwards infected erythrocytes are identified, sometimes along with their respective life stage. At last healthy erythrocytes and infected erythrocytes are counted and the parasitemia can be estimated. Problems arise when erythrocytes are clustered or overlap [Ross et al., 2006]. A lot of these techniques make assumptions on uniform object shapes or sizes and use area fitting techniques to identify erythrocytes [Di Ruberto et al., 2002][Ross et al., 2006][Linder et al., 2014]. Naturally, if these assumptions are violated (by overlap or clustering) these approaches fail. The algorithms vary in complexity, below some examples are given of the most common approaches.

One of the most basic, and frequently used techniques to segment erythrocytes from the background is based on morphological operators. Di Ruberto [Di Ruberto et al., 2002] implemented an automatic thresholding method based on Otsu's threshold method [Otsu, 1979], to separate the erythrocytes from the background. Using different morphological steps clustered and overlapped erythrocytes are split. At last two different methods to classify the erythrocytes are proposed, one based on morphological operators and one based on colour histogram similarity. They report good results on their data, however also state that the result is heavily dependent on the image quality. This is not a problem in an scientific laboratory setting, where high-end microscopes produces high quality images, but is a different story in the field. Morphological methods have the great downside that the operators affect the whole image in the same way. When all erythrocytes have nearly equal shape, size and colour this is not a problem, but when differences appear, they will still be processed in the same way. This processing can lead to one erythrocyte being segmented properly but a different erythrocyte in the same image not being recognized as one and being discarded. Dealing with these differences using morphological operators is possible, but require human intervention during the process to tweak settings in an trial-and-error approach, this is not desired in an automated diagnosis system.

A slightly more advanced segmentation technique is used by Sio [Sio et al., 2007]. Using edge detection, the edges of all erythrocytes are detected and linked together at their terminal points to form closed boundaries around all erythrocytes. These closed boundaries may compromise clumps of overlapping erythrocytes, and therefore a clump splitting routine that separates these regions is implemented. Finally, the infected erythrocytes are differentiated from healthy erythrocytes using the edge correlation coefficient and binary morphology. Again this technique relies on high quality images, small differences in contrast and luminance variations can be handled using the edge detection algorithm, but differences in shape or size still poses problems. Enclosed regions that differ from those of average erythrocytes are removed during the segmentation process. The algorithm was able to handle slight overlap, but in situations with greater overlap or multiple erythrocytes touching each other, the technique failed.

A last example is given by the work of Díaz [Díaz et al., 2009]. Their main effort is aimed at the classification of infected erythrocytes in their respected life cycle. Their segmentation process is based on a pixel classification method, that labels every pixel in the image as either background or foreground (belonging to an erythrocyte) based on its colour features. Afterwards an Inclusion-Tree-Structure based on [Monasse and Guichard, 2000] removes irrelevant objects from the image and keeps only the erythrocytes for the classification stage. Splitting of erythrocytes is based on a template matching approach. Which comes down to finding a 'best match' when two or more template shapes are superimposed in different ways to match the clumped shape. The classification is done using predetermined features of individual erythrocytes, such as; the color histogram, saturation level histogram, Tamura texture histogram and Sobel histogram. Two different classification algorithms are evaluated using these features to classify erythrocytes accordingly, a multilayer perceptron neural network (MLP) and a support vector machine (SVM). The advantage of their segmentation method over morphological operators is the usage of expert knowledge through a trained pixel classifier, this should result in the ability to process lower quality images. The classification stage is also more extensive using a good deal of information available in the different histograms and classifying using state of the art classifiers. The blood smears are used from a unrelated study from one year ago, which resulted in noisy smears with stain differences and artifacts in them. However, the images are still captured using high end microscopes coupled to a digital camera. So despite the lower quality blood smears, the images are still of high quality.

2-3 Smartphone based Microscope Blood Smear Images and the Performance of Conventional Techniques

The major difference between this thesis and previous researches on automated malaria diagnosis is the input data that is being used. In all researches, including this one, the input data consist of RGB digital images of blood smears. Differences between these images can be found due to the capturing device (different cameras), magnification at which images are captured and differences in microscopes. However, all other researches have very similar images with respect to each other. High quality images captured using digital cameras coupled to scientific light microscopes in a laboratory setting. We are dealing with a very different quality type of images. The smartphone based microscope is aimed to be as low cost as possible with sufficient resolution to detect an early ring malaria parasite. The images it produces are inherently worse than the digital camera coupled to light microscopes images, which brings a new obstacle to the automated diagnosis problem.

The smartphone based microscope has specifications listed in table 2-1, as a comparison a standard light microscope used for malaria diagnosis is also listed. The smartphone based imaging platform consist of a Motorola Camera XT1572 with 20.7 megapixels camarea sensor with a pixel size of 1.12 microns. Directly on the camera a 0.5 mm ball lens is placed, as explained in [Agbana et al., 2017] . Images are captured in the JPEG format in the maximum resolution possible by the camera, 4080x5344 pixels. Since the imaging system has been optimized to minimize system aberrations, the usable field-of-view for the 0.5 mm ball lens is therefore limited to 100 microns (equivalent to approximately 1300x1300 pixels).

	Mag	NA	Dl Res	FOV
Standard Light Microscope	50x	1.25	0.22 μm	180 μm
Smartphone (0.5mm Ball Lens)	8.5x	0.2	1.8 μm	100 μm

Table 2-1: Specifications of smartphone based microscope system versus and a standard laboratory light microscope

The differences in quality can be explained by table 2-1. For the optimal quality the following applies; magnification, higher is better, numerical aperture, higher is better, diffraction limit resolution, lower is better, field of view, higher is better, but can be nullified by examining more field of views per sample. The smartphone microscope is outperformed by the light microscope on all fronts by at least a factor 4, hence produces low contrast and low resolution images.

Two example images can be found in figure 2-2, a comparison is shown between an example image captured using the smartphone based microscope and a digital camera coupled to a light microscope, taken from the Mamic Database used by Linder [Linder et al., 2014]. Both images show only the green colour channel in which the purple colour from stained parasites is most visible [Ross et al., 2006]. Big differences in illumination consistency, contrast and detail (not visible at print-scale) can be found.

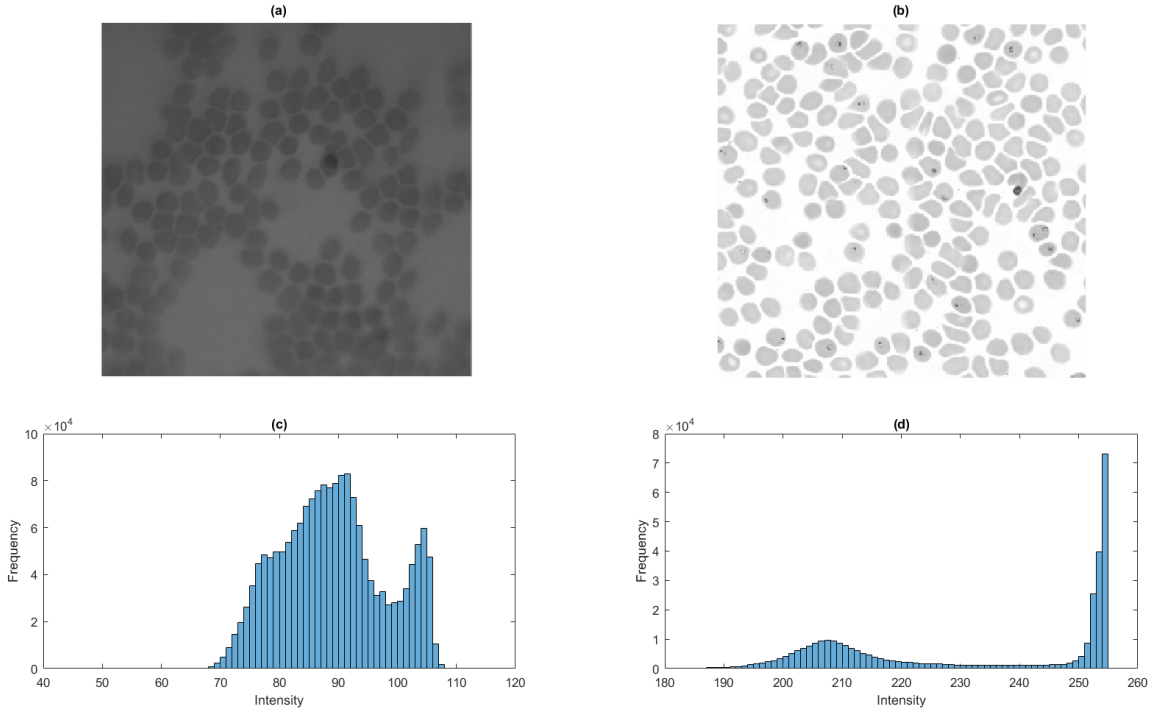

Figure 2-2: Example of smartphone based microscope image (a), with corresponding image histogram (c). Example of light microscope image, taken from Mamic Database (b) , with corresponding image histogram (d).

The differences in figure 2-2 also becomes clear looking at the distribution of the pixel

value histograms. Histogram (d), corresponding to the light microscope, shows a typical bimodal histogram, the principal mode corresponds to the greyscale intensities of the background colour, and the second mode to those that make up the erythrocytes [Ross et al., 2006] [Linder et al., 2014]. Histogram (c), that of the smartphone image, shows a less distinct bimodal distribution. The researchers discussed in section 2-2, for example [Ross et al., 2006][Linder et al., 2014][Di Ruberto et al., 2002] use this distribution to their advantage. An easy way of segmenting the erythrocytes from the background is by selecting a threshold level that maximizes the separability of the resultant classes, this can be done automatically using Otsu's method [Otsu, 1979].

The result of Otsu's method and the more advanced edge detection method of Sio [Sio et al., 2007] on both examples from image 2-2 is shown in figure 2-3. Note that the code for Otsu's method is standard in MATLAB 2017, the code for Sio's method is published along with their paper, see [Sio et al., 2007].

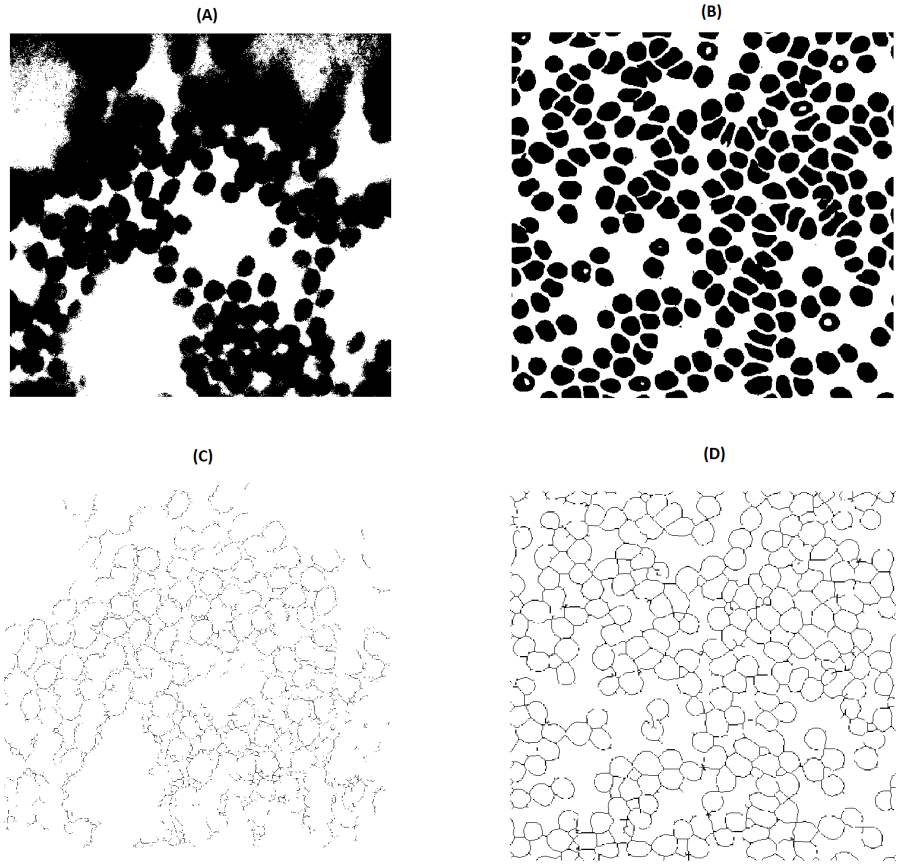

Figure 2-3: (A) Smartphone based microscope example image with Otsu's thresholding method. (B) Example image from a light micrsoscope taken from the Mamic Database with Otsu's thresholding method. (C) Smartphone based microscope example image processed with Sio's edge detection algorithm. (D)Example image from a light micrsoscope taken from the Mamic Database processed with Sio's edge detection algorithm.

As can be seen in the upper part of figure 2-3 (A-B) the threshold method is not capable of segmenting all the erythrocytes from the background in the smartphone image (A). The

image from the Mamic Database (B) with its distinct bimodal histogram shown in figure 2-2, is segmented almost perfectly. The difference can be explained due to a lack of uniform lightning in the smartphone based microscope. Pixels that belong to erythrocytes in the top part of the image share the same intensity as pixels in the background in the lower part of the image, whereas the Mamic image has a very even background with distinct pixel values for the erythrocytes.

The lower part, image C and D show the result of the first step of Sio's algorithm, the edge detection. Again the algorithm performs much worse on the smartphone image (C) compared to the image from the Mamic database (D). Due to less contrast and detail the edges of a erythrocyte in the smartphone image are more blurred and a smooth transition instead of a sharp step at the edges of erythrocytes in the Mamic image. This results in edges not to be found or scattered as small pieces around a erythrocyte. The next step in the algorithm would be to link together the edges to form closed boundaries around the erythrocytes, however the algorithm is unable to perform this step in the smartphone image, the preliminary result from the Mamic image (D) poses no problems for this next step.

2-3-1 Lightning Differences between the smartphone based microscope and light microscopes

One of the important differences in image quality is found due to the different lightning techniques in the smartphone based microscope and a light microscope. Modern scientific light microscopy makes use of Köhler illumination [Society, 1894]. This illumination method makes use of additional optical elements in the illumination element which results in extremely even illuminations, however, this increases the cost of the system. The smartphone based microscope is aimed to be very low cost and used a LED covered with a white paper, this can be improved a bit but will never reach the even distribution of Köhler illumination.

2-4 Convolutional Neural Networks Background

Convolutional Neural Networks (CNNs) is a class of deep, feed-forward Artificial Neural Networks (ANNs), specialized at analyzing image data [Krizhevsky et al., 2012].

An ANN processes information in roughly the same way that a biological nervous network, like a brain, processes information. The structure of ANNs especially stands out, it consist of a large number of neurons that are highly interconnected and work in unison to solve problems. A very important aspect of ANNs is the ability to adapt to certain problems by 'training' the network. Just like humans have the ability to learn by example, an ANN can do the same. This also means that the same network could solve several different problems by training the network differently.

Neural Networks have a remarkable ability to derive meaning from large amounts of complicated (imprecise) data. Patterns and trends can be discovered that are not visible for a human expert. All these characteristics make ANN very suitable for processing images and especially pattern recognition in images. The networks that excels in this area are called Convolutional Neural Networks.

In this section background information on (Convolutional) Neural Networks will be given. First general information on Neurons, Network Architectures and Training characteristics will be given, which hold for all Artifical Neural Networks, afterwards the information will narrow down to Convolutional Neural Networks.

2-5 Artifical Neurons

An ANN consist of many artificial neurons, a single neuron is a mathematical function that can be seen as a model for a biological neuron. This model is quite simple and explained below.

For a given neuron, let there be n inputs, with signals x_1 to x_n. Each input will be weighted with an according weight, w_1 to w_n. Next to this there will also be a bias value, b. The output of the neuron will be given as

$$y = f(\sum_{i=1}^{n} x_i w_i + b) \tag{2-1}$$

Note that equation Eq. (2-1) not only contains a dot product of its inputs and the weights plus a bias but also a f which stands for the activation function. This activation function takes the output of the neuron and applies a certain fixed non-linearity function to it. A visualization of this single neuron can be seen in Figure 2-4

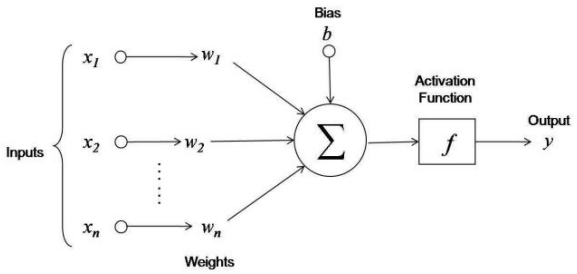

Figure 2-4: Visual representation of the model of a single Artificial Neuron

The activation function, f, is an important property of a neuron and determines its functioning. In its simplest form the activation function could be step function, resulting in a binary (0 or 1) output of the neuron. Using other non-linear functions however, such as the *tanh* or *ReLU* and many more non-linear functions, a small network with few nodes (neurons) can solve nontrivial problems.

2-5-1 Rectified Linear Unit (ReLU)

A neuron using a rectifier as the activation function is also know as a Rectified Linear Unit (ReLU) [Nair and Hinton, 2010]. ReLU's have drastically increased in popularity over the last years. The activation function is expressed as; $f(x) = max(0, x)$, i.e. the output of the

neuron is thresholded at zero, so only positive arguments are propagated as outputs, a visual representation of this function is given in figure 2-5

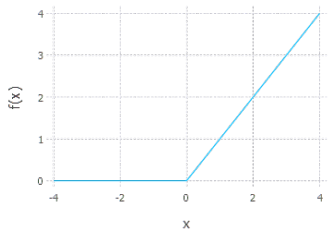

Figure 2-5: Rectified Linear Unit (ReLU) activation function , which is zero when $x < 0$ and then linear with slope 1 when $x > 0$

One of the reasons for its popularity is the fact that this activation is a simpler computation compared to a traditional *tanh*, which requires computing an exponent. This results in decreased training and evaluation times[Krizhevsky et al., 2012]. Next to this, networks using ReLU activation usually result in sparse networks.Sparsity in a network is a measure for the amount of 'dead' neurons (neurons that never activate, thus always output 0) in a network. Sparsity in neural networks has various advantages, these are explained neatly by Glorot [Glorot et al., 2011] in section 2.2. Nevertheless, too much sparsity in a network can have negative effects on the performance, various adaptions to the ReLU have been made to solve this problem, for example leaky rectified linear unit (Leaky ReLU), parametric rectified linear unit (PReLU) and a randomized leaky rectified linear units (RReLU) [Xu et al., 2015].

2-6 Artificial Neural Network Architectures

The neurons in an ANN can be connected in various ways, here we will only discuss feed-forward neural networks and disregard recurrent neural networks. This means all neurons are connected in a acrylic graph, meaning the output of neurons can be the input to the next neurons. Also there can be no cycles present in the connection, thus a signal can only propagate forward trough the network. Most of the time ANNs are organized in a layer-wise architecture. Examples of this architecture are shown in Figure 2-6 and Figure 2-7

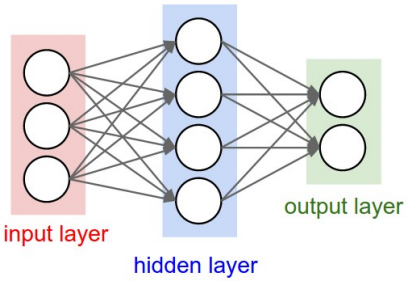

Figure 2-6: Example of a 2-layer feed-forward Neural Network (Input layer of 3 neurons, hidden layer with 4 neurons and an output layer with 2 neurons)

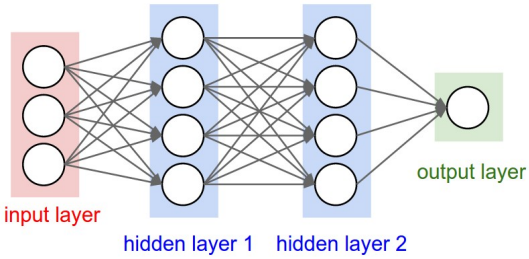

Figure 2-7: Example of a 3-layer feed-forward Neural Network (Input layer of 3 neurons, 2 hidden layers with 4 neurons each and an output layer with 2 neurons

As we can see in Figure 2-6 and Figure 2-7 both networks, and all networks for that matter, consist of at least a input and output layer. In between these layers there can be none, one or multiple hidden layers. The most common type of hidden layer is the fully-connected layer, which are shown blue in Figure 2-6, where all neurons in the layer are fully connected to each adjacent layer but share no connections with each other inside the layer.

A network can have multiple hidden layers. This means we can speak of a depth in the network, a N-layer network counts all layers except for the input layer. For example, a single layer-network is a input layer directly connected to a output layer. A network with one hidden layer is called a shallow network, more than one hidden layer and the network can be considered a deep neural network, however the exact definition of a deep neural network is not unambiguous and often also take into account what happens in the layers and not just the count.

Deep Neural Networks are the current state-of-the-art type of Neural Networks. They have proven to outperform previous state-of-the-art algorithms on various problems and contests [Krizhevsky et al., 2012]. A good overview of the history and recent successes is given by [Schmidhuber, 2015].

The main thing that defines Deep Neural Networks and also where the name comes from is the large amount of layers in the network. Due to the amount of layers the original input is transformed more times than shallow networks. This way the network is able to 'learn' harder tasks than shallow networks, for example more complicated features can be extracted in image recognition. There are also downsides to Deep Neural Networks. A large network needs a big amount of training data and is extremely computational expensive to train. For this reason a lot of times GPU's are necessary to train Deep Neural Networks. Also they do not have a strong theoretical background, thus choosing the right network with its parameters, training method, topology, etc. is considered more an art than science. Getting insight in what is happening is not easy to comprehend in a lot of cases, which results into trial-and-error design processes.

2-7 Training the Neural Network

In order to make an ANN solve problems it has to be trained. But before training can take place a network has to be initialized, more specifically the weights and biases have to be initialized. Biases can generally be initialized to zero, weights are normally chosen at random from a distribution specified by the expert. A suitable distribution is dependent on the network and type of activation functions used.

The first objective during training is to perform the task the network is supposed to do as good as possible. This measure is done via a cost function. The cost function represent the difference between the prediction and the ground-truth. A common way to express this is via a quadratic loss function, shown in equation Eq. (2-2) in its most basic form. Where y_j is the output of output neuron j and E_j is the expected outcome (ground-truth) of that neuron.

$$C_{MSE} = 0.5 \sum_j (y_j - E_j)^2 \tag{2-2}$$

This cost value is also known as the mean squared error. It is one of the ways to express the quality of the estimator. Minimizing the cost function, which is minimal at zero in this case, is the objective during the training phase.

Optimizing an ANN during training is done with backpropagation [Hecht-Nielsen, 1988]. This is the workhorse of learning in ANNs. Backpropagation is based on calculating derivatives of the errors starting in the last layer, with those error derivatives, one can compute error derivatives in previous layers, thus propagating backwards trough the network, from output to input. During backpropagation the weights are updated using a gradient descent method each cycle until a desired quality of the estimator is achieved, that means once the network output matches the desired output good enough, and thus the cost function is small enough. For a full explanation and proof we refer to the work of Hecht-Nielsen [Hecht-Nielsen, 1988].

During the training phase the Neural Network is prone to overfitting. Overfitting occurs when a model is too complex to describe an underlying relationship and instead describes random error or noise. To prevent overfitting of the network, regularization methods can be used. There are different methods that show good performance in different applications, thus there is not one golden regularization method. L2 regularization is a common regularization method, while dropout [LeCun et al., 2015] shows promising results. Regularization can be seen as a way of penalizing the complexity of the model and is the second objective during training.

2-8 Convolutional Neural Networks

Convolutional Neural Networks, called ConvNets for convenience, are a special type of Artificial Neural Networks. All information given in the last section about ANNs hold but since the explicit assumption is made that the inputs to these networks are images some properties can be encoded in the network. These make the networks more efficient to implement and vastly reduce the amount of parameters to tune.

Most recent well performing ConvNets are mainly build from three types of (hidden) layers, convolution layers, pooling layers and fully connected layers. The fully connected layers are

exactly the same as described in the previous section. The convolutional and pooling layers are described below.

2-8-1 Convolutional Layers

Convolutional layers are responsible for feature extraction in images. Convolutional layers can be seen as a sort of filters. These filters are in fact a window of small dimensions that slide over the input image so that the centre has been in every position. At every spatial location the filter does a convolution with the input image and produces a activation map that gives a response of the filter at all locations. An example of this procedure is shown in Figure 2-8 where 3 filters, also known as kernels perform a convolution with the input and produce 3 separate feature maps.

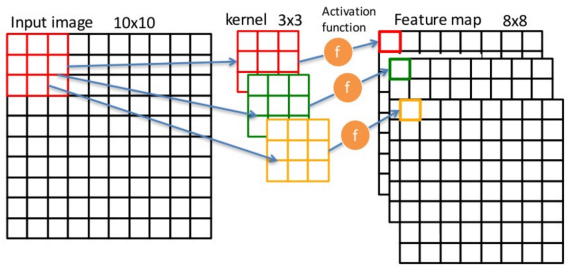

Figure 2-8: Convolution of an input image (10x10 pixels) with 3 kernels (3x3 pixels) shown in red, green and yellow, resulting in 3 feature maps (8x8 pixels) that contain the output activation's for a given kernel

The convolutions that are performed are very alike to 1-D convolutions known from signal processing that take the form of time series. The main difference is of course the change in dimensionality and that convolution in ANNs are actually a slightly modified version of convolutions that is called cross correlation in signal processing.

The formula to perform the convolution is given in equation Eq. (2-3)

$$y_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} x_{(i+a)(j+b)} \tag{2-3}$$

Where the input is a $N \times N$ square neuron input layer called x, the filter is called w and has size $M \times M$. The output of the convolution is y at place (i, j). This convolution can easily be visualized in Figure 2-9.

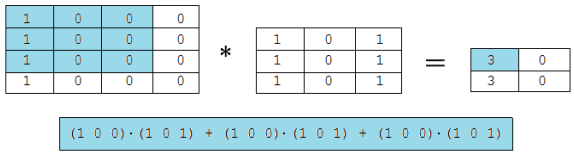

Figure 2-9: Convolution equation example, the input image (4x4 pixels) is shown in the left top corner, the kernel (3x3 pixels) in the middle and the output (2x2) pixels in the right top corner, the blue box at the bottom is one the 4 convolutions done, represented in blue in the top images

In Figure 2-9 we can see one convolution done, the blue window will slide across the whole input image and cover each possible position, resulting in 4 convolutions, shown in the most right feature matrix.

Normally there are multiple convolutional layers, each with different filters. The filters are trained in such a way that they all will 'detect' separate features in an image, from low-level features such as edges and corners up to high-level patterns like honeycomb or grid patterns. A lot of these filters have fixed weights for the whole image, and do not have different values for the left part of the image than the right part. This is called parameter sharing and is very effective to bring down the total amount of parameters to train. After the convolution of the input image with all these convolutional layers it gives an output of multiple activation maps that shows all the locations where certain features are present in the image.

2-8-2 Pooling Layers

Pooling layers are used to reduce the dimensionality of the feature maps and to prevent overfitting [Hinton et al., 2012]. There are multiple pooling techniques like L2-pooling, average pooling and MAX pooling, where MAX pooling is the most popular and will be explained here.

A MAX pooling layer produces a map that shows the most dominant feature in divided parts of the feature map. MAX pooling layers leads to improved generalization [Rawat and Wang, 2017] It is often enough information to know what is the most dominant feature in an approximate area and not the exact location. Just like the convolutional layer it is again a filter with a small dimensionality, most of the times 2x2, that is moved with stride 2 over the activation maps and takes the max value of each 2x2 window. An example of this pooling layer is shown in Figure 2-10

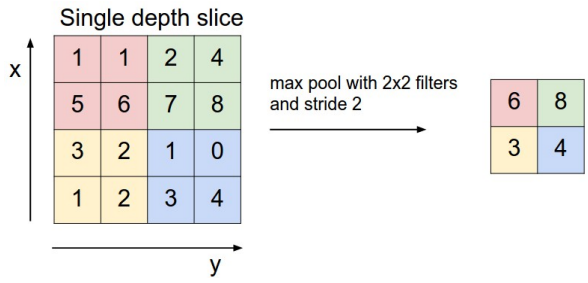

Figure 2-10: Max Pooling Layer on a 4x4 input (left) with a 2x2 filter with stride 2, resulting in an output image (right) that is down-sampled by a factor 2 and holds only the maximum values of the coloured areas in the input image

Using a pooling layer as depicted in Figure 2-10 means the activation maps are down sampled by a factor 2 over the width and height, thus leaving less inputs for the next layer and improves computation times. Since this inevitably also reduces the amount of information, the improvement in computation times sometimes comes at the cost of accuracy.

Chapter 3

Method

Our proposed automated malaria diagnosis procedure consists of two stages: 1) Localization of all erythrocytes present in the image, 2) Classification of each erythrocyte as either infected or healthy. Both stages use Convolutional Neural Networks, first as a pixel classifier, secondly as an object classifier. In this chapter the choice for this approach and both stages are discussed in detail. The overview of the process is shown in figure 3-1.

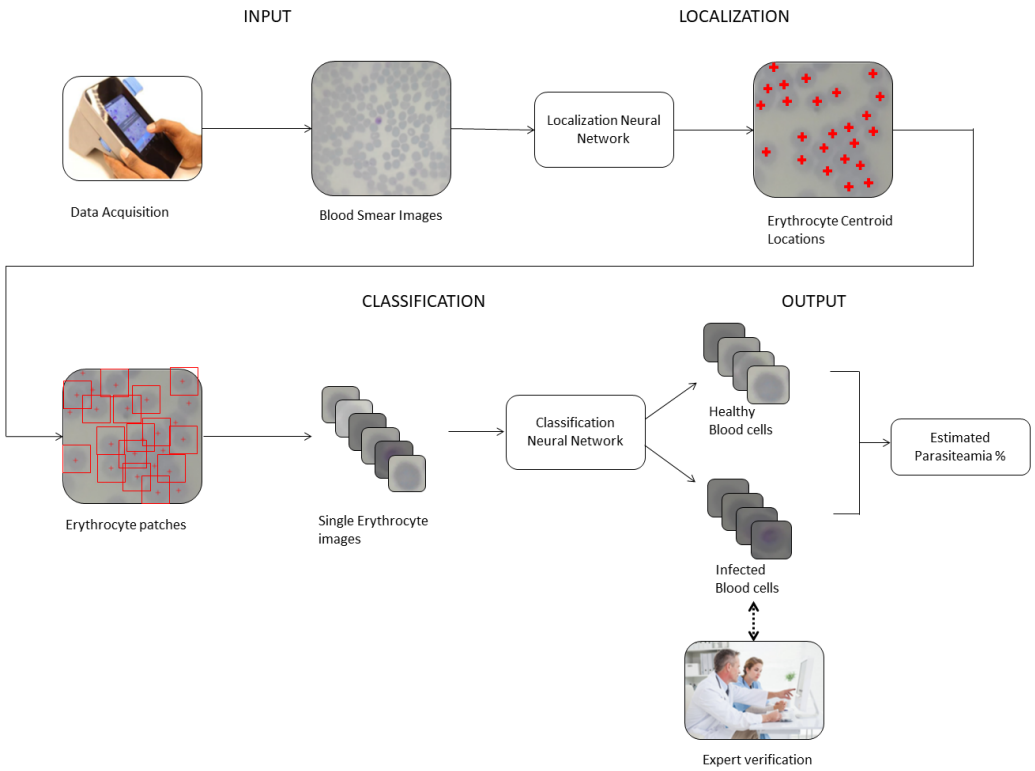

Figure 3-1: Overview of the proposed automated diagnostic method for the smartphone based microscope

As can be seen in figure 3-1 the total process consist of various steps to achieve our goal. The first step is data acquisition, this is done with the smartphone based microscope discussed in 2-3. The second step is to localize the centre coordinates of all erythrocytes in the input image. With these coordinates cropped images can be made containing the individual erythrocytes in a image. In the third step these images are classified as being a healthy or infected erythrocyte. In the last step, results from the classification stage that hold confidence scores for each classified erythrocyte can be shown to human expert(s) for validation, afterwards parasitemia can be estimated.

3-0-1 Stage 1: Erythrocyte Localization

Given a RGB input image I the goal is to find the set $C = \{c_1, c_2, c_3, ..., c_n\}$ with elements c living in $\mathbb{R}^2$ and n is unknown beforehand. This set contains the centroid coordinates of all n erythrocytes present in I. To achieve this goal a CNN is trained to classify cropped images of I in two classes: *erythrocyte* or *non-erythrocyte*.

Using this CNN every pixel p in image I is classified, this process is shown in figure 3-2. This process includes classifying small x by x windows of I that have pixel p as a centre and moves over every pixel p in I. At every pixel p the CNN computes the class-score for that specific window, stores this in a matrix and moves to the next pixel. After this process, every pixel p has a class designated and confidence matrix M can be constructed. This matrix depicts the confidence scores at every pixel for a specific class. For the first class, *erythrocyte*, we expect to see high scores (close to 1) around the erythrocyte centroids and low scores (close to 0) everywhere else. Using the post-processing steps explained in subsequent section 3-0-4 we can compute the set C from this matrix M.

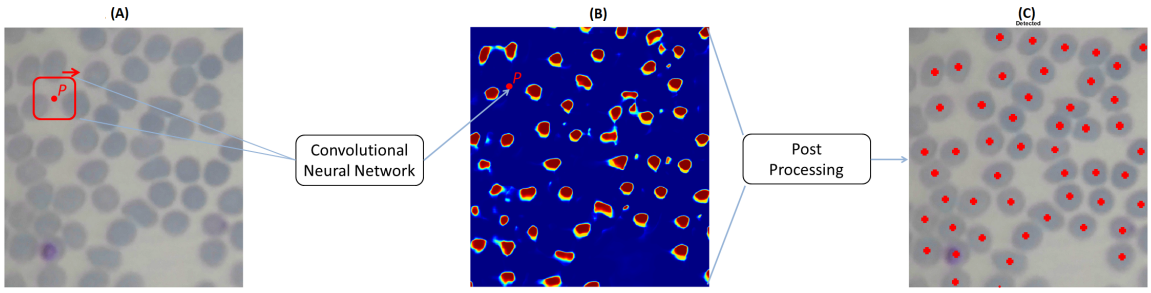

Figure 3-2: Localization process, (A) is a part of the input image, where p depicts the centre of the n by n window to be classified as erythrocyte or non-erythrocyte, (B) shows the confidence matrix M that holds output scores for one output class, (C) shows the centre locations found after post-processing the matrix M

3-0-2 Erythrocyte Localization network architecture

We utilize a relatively narrow and shallow CNN for the localization phase, this is possible because of the simple shape of the erythrocyte that has to be detected. This network has to analyze very large amounts of data, a separate window for every pixel in I, so a small

network is necessary to get acceptable computation times. The malaria infections, which are more detailed features inside the erythrocyte are only of importance in the second stage.

The network architecture is build up (from input to output) by 2 convolutional layers followed each by a max-pooling layer, hereafter 2 fully connected layers with a final double class soft-max layer. The convolutional, pooling and fully connected layers are explained in section 2-4 on Convolutional neural networks. The soft-max layer at the output ensures that the outputs of the network all add up to 1. This layer is used because there are two mutually exclusive classes, *erythroctye* or *non-erythrocyte*. By forcing the outputs to represent a probability distribution across discrete alternatives this information is included in the network. The cost function used during training is the cross-entropy cost function, which is commonly used with a softmax output layer, this function is given as;

$$C = \sum_j t_j \; log \; y_j \tag{3-1}$$

with t_j being the target value, which is 1 for correct answer and 0 for the wrong answers. This cost function has the nice property that it has a very big gradient when the target value is 1 and the output is almost zero, this gradient is used in the back propagation algorithm to train the CNN.

Furthermore we use neurons with a ReLu activation function, the advantages of these are explained in section 2-5-1 on rectified linear units, in short we refer to [Glorot et al., 2011] for an overview of advantages of this activation function.

3-0-3 Erythrocyte Localization Training data

An expert manually annotated a total of 1473 erythrocytes centroids in 10 blood smear training images. To create a large enough training data set, an extra augmentation step is included. 9 windows are cropped of each erythrocyte, the first window having its centre coincide with the annotation, and 8 others all shifted $+/-$ 1 pixel in either x or y direction. This results in a total of 13257 training images for the *erythrocyte* class.

For the *non-erythrocyte* class 13000 patches are randomly cropped from the same 10 blood smear images annotated by the expert. The only rule we impose on the these patches is that the centre of a patch q can not be closer to a annotated centroid c by a euclidean distance of 20 pixels, i.e. $d(q_i, c_j) > 20 \; \forall \; i, j \in \mathbb{Z}$. This way we ensure that all non-erythrocyte training patches contain either off-centre erythrocytes or no erythrocytes at all.

Examples of training images for both classes are given in figure 3-3. Note that these images are enhanced in both contrast and brightness to make them easier visible for this report, for the network training and processing of new images the raw RGB images are used.

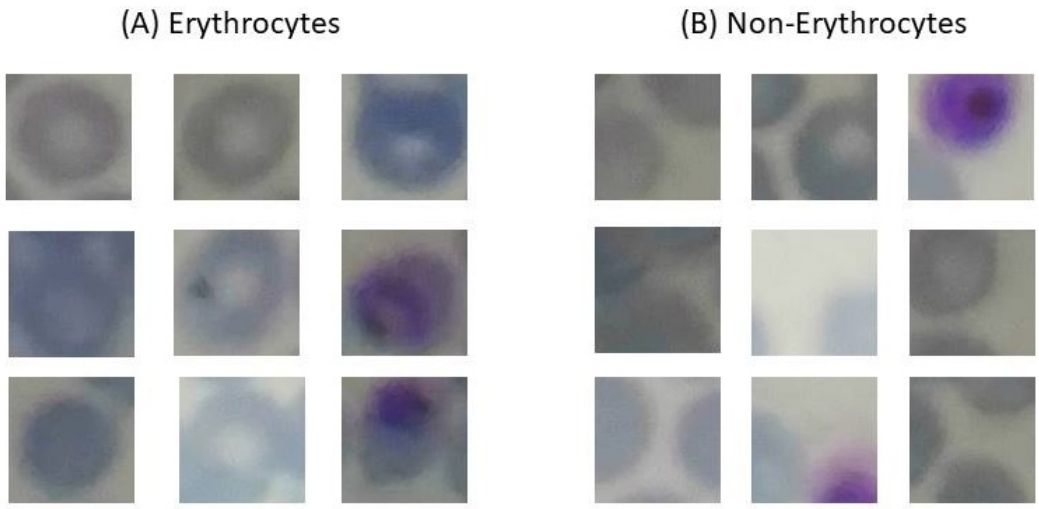

Figure 3-3: (A) shows 9 examples of images belonging to the erythrocyte class for training, (B) shows 9 examples of images belonging to the non-erythrocyte class used for training. Note that image brightness and contrast is enhanced for print visibility

Erythrocyte detection is rotational invariant. During each network training step we have a 50 % chance to augment the input image by rotating it 90 degrees, this should ensure a rotational invariance after training.

3-0-4 Localize erythrocytes in a new unseen image

Localizing the erythrocytes in an unseen image I is straightforward at this point. Since the network learned the features of the erythrocytes on raw RGB values we do not have to do any preprocessing on a new image. Using the sliding window technique discussed before we get confidence matrix M, which yield the classification score for the erythrocyte class for every pixel p in image I.

To obtain set C we have to post-process the matrix M. First we threshold matrix M with a threshold level of 0.9, this leaves us with a new binary matrix that has value 1 for all areas where the network has the confidence of a erythrocyte centroid being located at that position of at least 90 % and a 0 every else. Secondly, we segment this binary image and split clumped areas with a watershed algorithm [Meyer, 1994]. Thirdly we calculate the centre of mass of all connected regions found by the watershed algorithm, these points make up the set C. At last we use our expert knowledge of the erythrocytes, due to their round shape we know the centres of two erythrocytes can not lie too close to each other. To ensure this, a constraint is added to set C, all points must have a minimum distance d_{min} from all other points. If two points lie too close to each other, both points are excluded from the set and a new point that is the average of both is added, this is done until the constraint is met. This yield final set C with all centroid coordinates.

Before the next phase we create a set of cropped images from I with the centre coordinates equal to those of the elements in C, if the algorithm was successful this set of images contain all erythrocytes present in I in separate images.

3-0-5 Stage 2: Erythrocyte Classification

In the previous step, the erythrocyte centroid locations are determined and individual crops are taken showing the erythrocytes in separate images. Given this set of erythrocyte images, the goal is to classify these as *infected* or *healthy*.

Since the two classes are mutually exclusive, a softmax output layer is used again. This gives a nice result for eventual expert validation. Every image will have a confidence score attached and can be interpret in a intuitive way, for instance, the network can be 95% certain the erythrocyte is infected and 5% certain it is healthy, the same way a human could give his results.

In order to calculate the parasitemia a binary output per erythrocyte image is needed, healthy or infected. A threshold value is selected on the confidence score to count as a infected erythrocyte. For now the value is set to 0.9 for the infected class, so the network has to have at least 90 % confidence in order to count a erythrocyte as infected. If the system would be employed in the field and used a diagnostic aid, it would be wise to also show an expert all erythrocytes that have a score of greater then 50% on the infected class, to be more confident that no infections are missed. Incorrect classified instances can be added to the training database to improve the network performance over time.

With the erythrocytes divided into the two classes, i for the infected set and h for the healthy set, perhaps with adjustments from the expert validation input, the parasitemia is calculated as follows;

$$Parasitemia = \frac{n_i}{n_i + n_h} \tag{3-2}$$

with n_i the amount of infected erythrocytes and n_h the amount of healthy erythrocytes.

3-0-6 Erythrocyte Classification network architecture

The network utilized for classification of the erythrocytes is fairly similar to the localization network. From input to output the network is build as follows: 2 convolutional layers each followed by a max-pooling layer, hereafter 2 fully connected layers with a final double class soft-max layer. Note that this network is similar to the localization network in the order of layers but differs in the amount of nodes per layer, the filter size in the convolutional layers and of course is trained on very different data from the localization network.

We would have liked to try different, especially deeper neural networks for this classification problem, because of their outstanding performance lately in different competitions, for example the traffic sign classification [Cireşan et al., 2012a]. Unfortunately, as becomes clear in the next section, we suffered from a lack of training data. Using a low amount of training data for a deep neural network easily converts the network to a memory cell. Such a network can memorize every training image perfectly and reaches very high classification performance during training phases, but is unable to generalize the underlying relation that is needed to classify unseen objects.

The current architecture of the classification- and localization network, with multiple convolutional layers and pooling layers in between, achieve very high performance on image

classification these days [Krizhevsky et al., 2012]. Most hyper parameters, such as: number of nodes in each layer, filter sizes of convolutional layers or pooling layers, number of convolutional layers, number of fully connected layers, learning rate, and more, are initialized to values either found in the literature or logically reasoned. Afterwards, these values are adjusted through a trial-and-error process. We know we have probably not found the optimal values, but doing so is a problem on its own. This optimization problem suffers heavily from the curse of dimensionality since there exist a large amount of hyper parameters to be optimized at the same time, therefore we have chosen that a trial-and-error process is sufficient at this time in the research.

3-0-7 Erythrocyte Classification Training Data

The erythrocytes annotated by the expert before are divided in the *healthy* and *infected* class, examples images are shown in figure 3-4. Our training dataset for infected erythrocytes is significantly insufficient due to minimal availability of blood smear training images, combined with relative low parasitemia in those that are available. To circumvent this condition a little bit, we augmented all sample windows, flipping them vertically to increase the dataset by a factor of 2. Although 100 training examples of the infected erythrocytes were generated, it is still considered very small as compared to the data requirement for training an efficient and well performing neural network. While there is no hard rule for a minimum dataset size, most neural networks start to perform well with a database of a few thousand samples per class [Cireşan et al., 2012b].

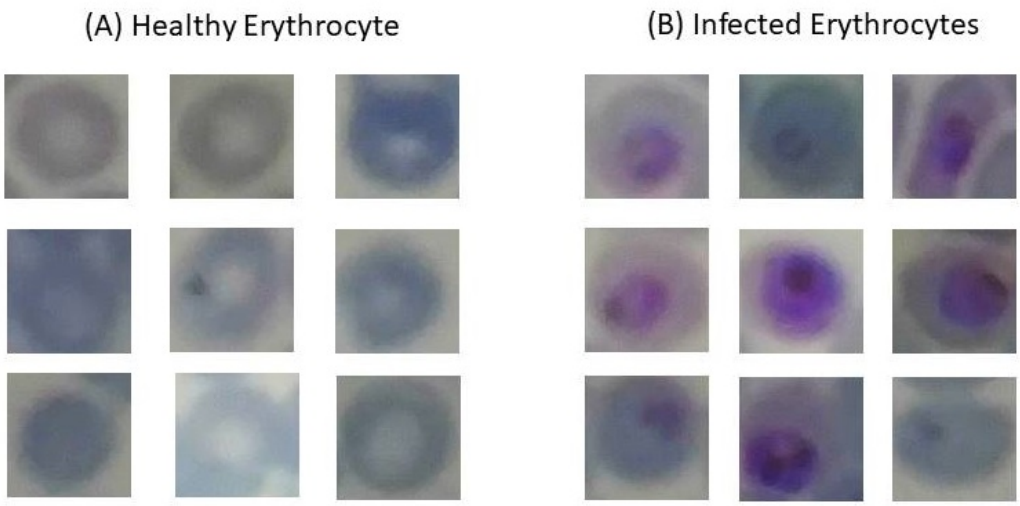

Figure 3-4: (A) shows 9 examples of images belonging to the Healthy erythrocyte class for training, (B) shows 9 examples of images belonging to the Infected erythrocyte class used for training. Note that image brightness and contrast is enhanced for print visibility

3-1 Two Stage Approach to Automated Malaria Detection explained

The two stage approach to automated malaria detection has been well studied and reported in literature. Linder [Linder et al., 2014] for instance first uses morphological operators to find candidate regions (possible infected erythrocytes) and classifies these in a second stage on malaria infections. However, an intuitive reader who has read the first part of erythrocyte localization may wonder why identification of *infected* erythrocytes is not done in one-shot. The paragraph below presents our reasons for the separation.

First of all by using two networks, both of them can be optimized for their specific task. The first (localization) network has to process large amounts of data, a small cropped image for every pixel in the input image. The input image easily scale up to 1300x1300 pixels, which adds up to almost 1.7 million (small) images to be analyzed. To do this in a manageable time, the network has to be small. The task to classify between a erythrocyte and a non-erythrocyte is fairly easy due to its distinguishable round shape, so the network can be kept quite small. For the second task, the classification between an infected and healthy erythrocyte, a lot less data is needed to be processed, a window per found erythrocyte location, usually around 70 to 150 windows for a 1300x1300 pixel image. However, more detailed information is needed to differentiate between a healthy and infected erythrocyte. In this stage, larger and/or deeper networks, multiple networks in parallel, or any other solution can be used that would have taken too much computation power if used the first stage.

Secondly, this solution lends itself perfectly to be extended upon in the future. Where this thesis focuses on one of the five parasite species present in humans and does not take into consideration the life cycle of the found parasites, these are things that should be implemented before being able to function in the field as an automated tool. To implement these features, the first part (localization of erythrocytes) can be left intact, and only the classifier should be updated or extended, which requires little work compared to a overhaul of the whole algorithm. And again, extending the first phase to also classify the species and life cycles, would have led to a big increase in computation time when all 1.7 million images would have had to be classified instead of the 70-150 erythrocytes.

At last, to calculate parasitemia, the total amount of erythrocytes should be known, either healthy or infected. Since we have shown in section 2-3 the segmentation of erythrocytes from the background is not possible with conventional techniques, due to image quality. We would have to implement some other solution to get the total erythrocyte count after finding the infected erythrocytes. Since this solution will probably be looking similar to the localization algorithm as is, the current solution seem to make more sense.

3-2 Localization Algorithm Efficiency

In this thesis the efficiency of the algorithm is not of high priority, since we are keen on demonstrating the proof of concept at this stage. However, the ultimate goal is to have a all-in-one solution to the malaria diagnosis problem, which would mean to run this algorithm on the smartphone that captures the images. A smartphone naturally has limited computation power, so algorithm efficiency becomes more important. Some efficiency is already introduced

by working with pooling layers, but looking carefully at how the algorithm works, a huge improvement can be made, which will be explained in this section.

Unfortunately the Neural Network Toolbox of MATLAB 2017B did not allow certain changes in the code to be made, due to read-only properties. These changes are needed for a implementation of this method. Therefore the following part remains theoretical and is not tested yet.

The problem of erythrocyte localization is quite a unique one from the perspective of object classification. The problem is not so much the unique shape or visual properties to distinguish an erythrocyte from the background, but to classify many instances of the same object in one image. We have already found a solution to this problem by splitting the bigger input image I into small pieces and classify these accordingly, where the class output is allocated to the centre pixel of this window, i.e. the network is working as a pixel classifier. Due to its current architecture, the classification of these smaller images involve many computations that are redundant since they have already been done before, this becomes clear looking at figure 3-5, where the current method (A) is shown in comparison to the proposed method (B).

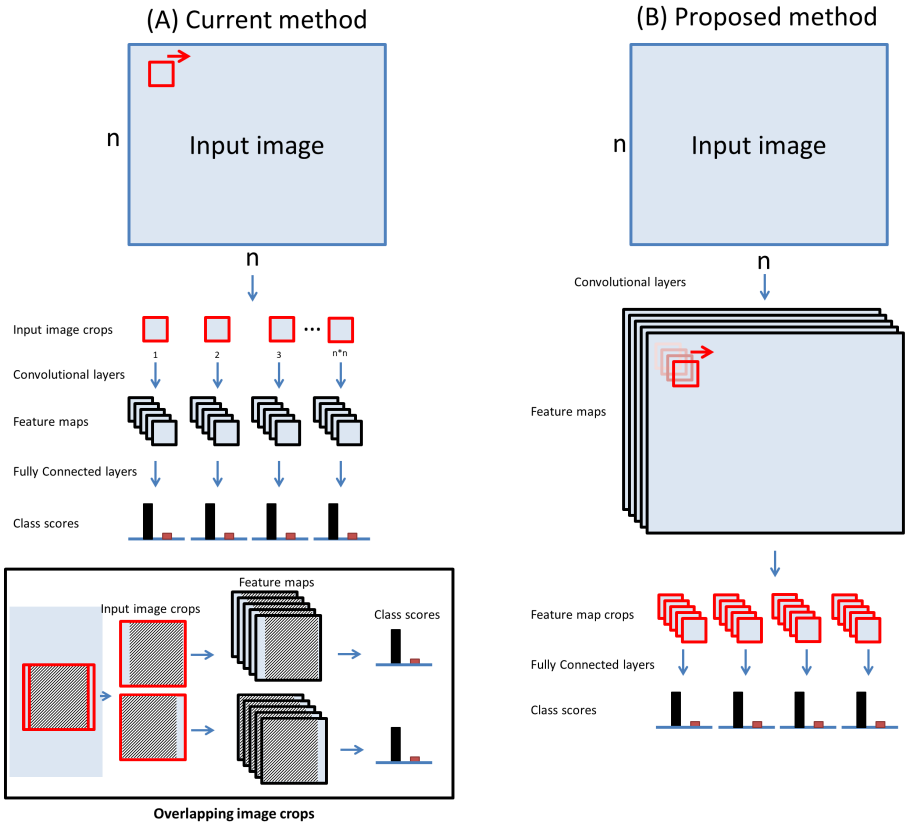

Figure 3-5: (A) The current method implemented, cropping images from the input image and using those as an input for the neural network (B) The proposed method which first computes features maps from the input image and crops parts of these maps as an input for the fully connected layers of the neural networks. The black box bottom left shows the problem with method (A), where two frames next to each other overlap in the original image, and computations are done multiple times with the same result

In image 3-5 (A) the current method is shown. As explained in section 3, crops are taken from the original input image, for every crop, feature maps are computed using the convolutional layers in the neural network, afterwards the information from these maps are combined using fully connected layers and an output is computed. The main issue occurs when two frames overlap in the original image, for example, let us use a 100x100 pixel window for the crops. If two frames lie next to each other, as shown in the black box in figure 3-5, for the first window (top) the feature maps are computed, fed into the fully connected layers and an output is derived. The second window (bottom) is shifted by 1 pixel, so there are only 100 new pixels, 99% of the image is already processed in the previous window. However, for this window again, the feature maps are computed (which contain 99% the same information as the previous window), fed into the fully connected layer and an output is derived. In other words, 99% of our computations in the convolutional layers are redundant for this example, because this information was already computed in the previous step.

For this reason we come up with method (B), where the neural network is split up. Please note that the neural network is trained in exactly the same way as network (A), so the filter values in the convolutional layers and neuron weights in the fully connected layers are identical. With this method, before we take small crops we first compute the feature maps on the input image, because of the same training method these should look identical to those of method (A), but now only computed once for the whole image. Afterwards we take crops of the feature maps and use those as an input to our fully connected layers. The main difference is omitting the redundant calculations done in (A) while computing the feature maps, but this adds up. If one considers the crop window to be of size n x n, the amount of calculations done in the convolutional layers scale down by approximately a factor $n * n/4$. This would result in around 2500 times less calculations for the 100x100 pixel windows. Note that the amount of calculations in the fully connected layers will remain the same, since a class has to be computed for every pixel in I.

We believe this proposed method can mean a huge improvement in computation efficiency and yield the same performance result. Care must be taken when one would implement the pooling layers, which can result in different feature maps using method (A) or (B), while we believe this would not affect the performance much, research should prove this.

Chapter 4

Experimental Results

To asses the performance of our algorithm we will treat it as if it is a fully automated diagnostic tool, this means no human validation is used after the classification stage.

The algorithm is analyzed in two stages, the ability to localize the erythrocytes and the ability to diagnose an image, which means to estimate the parasitemia (amount of infected erythrocytes to total amount of erythrocytes).

Two human experts are shown 10 blood smear images captured by the smartphone based microscope. These images are enhanced using image editing software, where changes to brightness and contrast are made to make the parasites more visible for the human experts. Note that the original images are used for the algorithm evaluation. Expert 1 (5 years work experience) counted all erythrocytes manually using a tally counter, as if the image was viewed through a microscope. Expert 2 (4 months research experience) used computer software to manually annotate all erythrocytes and retrieved the counts from this result. After counting, the amount of infected erythrocytes per image are determined by both the experts. After this manual step, the same 10 images, now annotated with all erythrocyte centroids having a red dot, is shown to the experts and verified as the ground truth. For the infections we are not able to establish this ground truth due to the poor image quality, uncertain cases will always persist even after careful inspections.

4-0-1 Localization Results

For the first stage, the ability to correctly localize the erythrocytes we will make use of two performance measures, the precision and sensitivity. Precision will depict the amount of erythrocytes among all identified spots by the network. Sensitivity will depict the amount of erythrocytes localized among all erythrocytes annotated in the ground truth image.

$$Precision = \frac{TP}{TP + FP} \qquad Sensitivity = \frac{TP}{TP + FN}$$

True positive (TP) is a localized centroid also annotated in the ground truth. False Positive (FP) is a localized centroid not annotated in the ground truth. False Negative (FN) is a centroid annotated in the ground truth but not localized by the network. For all these binary measures above, the hits and misses of the network, we stick to a small radius around the annotated ground truth, if a localized centroid is close enough to an annotated centroid it 'hits' and count as a TP, if it too far off it 'misses' and counts as a FP.

It is of very high importance to have a high sensitivity since all found locations will be the input for classification stage, missed erythrocytes could mean missed infections. Besides, high precision is also needed to estimate the parasitemia correctly.

Results of the localization stage of all 10 images are shown table 4-1.

Table 4-1: Erythrocyte localization results of the algorithm against the ground truth for 10 images

Image #	Ground Truth Erythrocytes	Algorithm Detections	Precision	Sensitivity
1	151	150	98.00%	97.35%
2	149	155	94.84%	98.66%
3	132	132	96.97%	96.97%
4	185	193	94.82%	98.92%
5	87	96	90.63%	100.00%
6	77	81	90.12%	94.81%
7	93	104	85.58%	95.70%
8	129	141	87.23%	95.35%
9	141	146	96.58%	100.00%
10	128	140	87.14%	95.31%

The average precision has a mean of 92.21 % with a standard deviation of 4.60 %, the average sensitivity reaches a mean of 97.31 % with standard deviation of 1.99 %. As being said the sensitivity is the most important measure and scores on 9 out of 10 samples above 95 %, which is the lower bound to be an applicable diagnostic [WHO, 2000]. The precision is also reasonable. Most mismatches are found at areas where erythrocytes are clustered, which is a well known problem where other researches devote a lot of energy to [Ross et al., 2006]. This approach does not need extra steps to solve this problem, which is a nice advantage over other solutions. The solution can probably be improved by adding more training samples containing overlapped erythrocytes to the training database.

4-0-2 Diagnostic Results

The results of the algorithm at diagnosing an image, which means first localizing the erythrocytes and afterwards classifying them accordingly is compared against the two experts. The results are shown in table 4-2, as an illustrative example the result of one of the images is shown in figure 4-1, please note that the right part of the image, showing the single erythrocyte images along with their respective class scores, only shows a small selection of all healthy erythrocytes.

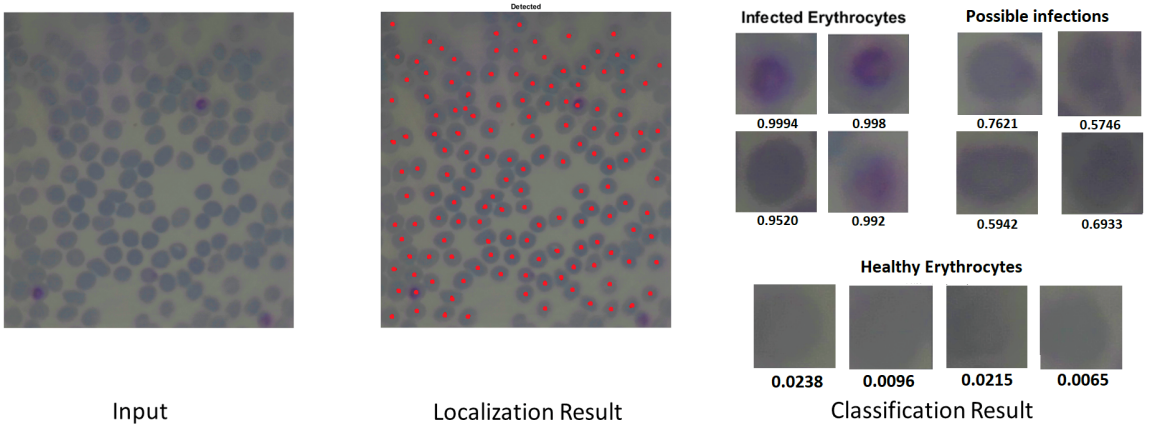

Figure 4-1: Erythrocyte localization and classification result of image 9, from left to right; the input image, the found locations annotated with a red dot and a review of the classification result, showing the 4 infected erythrocytes (confidence scores greater than 0.9), 4 erythrocytes with possible infections (scores above 0.5 but below 0.9) and 4 examples of healthy erythrocytes with low scores, note that there are many more of these healthy erythrocytes but only 4 are shown in this figure

The result of image 9 shown in figure 4-1 yields interesting information. The algorithm has found 4 erythrocytes being infected with high confidence, 4 erythrocytes in between both classes and 138 marked as healthy with high confidence. Expert 1, with most work experience from both experts, annotated 2 erythrocytes as infected. The 2 infected erythrocytes annotated by the expert, match with those erythrocytes that have the highest confidence score given by the algorithm. Even though there is a difference between the number of infected erythrocytes given by the algorithm and the expert, the infected annotations from the expert are at least a subset of the algorithms infected annotations. This result is also observed in other images, where experts and algorithm disagree on numbers, often the annotated infected erythrocytes by the experts are a subset of the algorithms annotations.

Table 4-2: Diagnosis Results of two human experts versus algorithm

	Amount of Erythrocytes			Infected Erythrocytes			Parasitemia		
Image #	Expert 1	Expert 2	Algorithm	Expert 1	Expert 2	Algorithm	Expert 1	Expert 2	Algorithm
1	112	150	150	3	3	1	2.68%	2.00%	0.67%
2	118	152	155	2	4	4	1.69%	2.63%	2.58%
3	116	134	132	1	4	3	0.86%	2.99%	2.27%
4	177	189	193	0	2	4	0.00%	1.06%	2.07%
5	81	88	96	1	2	5	1.23%	2.27%	5.21%
6	65	77	81	4	6	6	6.15%	7.79%	7.41%
7	91	94	104	3	4	5	3.30%	4.26%	4.81%
8	123	132	141	3	6	6	2.44%	4.55%	4.26%
9	125	143	146	2	5	4	1.60%	3.50%	2.74%
10	128	130	140	1	1	5	0.78%	0.77%	3.57%

To compare the results between the individual experts and the algorithm , one observes the relation between the results of expert 1 (A), expert 2 (B) and the algorithm (C). The covariance, correlation coefficient and the Root Mean Squared Error (RMSE) is computed

between A-B, A-C and B-C. Relations are computed on the total number of erythrocytes annotated, total number of infected erythrocytes annotated and the estimated parasitemia in the blood smear. The results are shown in table 4-3.

Table 4-3: Diagnostic result comparison of expert 1 (A), expert 2 (B) and the algorithm (C). Relation between each participant is given by the covariance, correlation coefficient and Root mean Squared Error (RMSE). Relations are computed on the total number of erythrocytes annotated, total number of infected erythrocytes annotated and the estimated parasitemia from table 4-2

		Covariance	Corrrelation	RMSE
# Erythrocytes	A-B	971.18	0.93	19.18
	A-C	953.80	0.96	22.09
	B-C	1100.87	0.99	6.31
# Infections	A-B	1.56	0.73	2.02
	A-C	0.33	0.18	2.85
	B-C	0.77	0.30	1.90
Parasitemia (%)	A-B	0.03	0.90	1.40%
	A-C	0.02	0.65	2.08%
	B-C	0.03	0.73	1.45%

The erythrocyte localization performance was already determined against the ground truth, which gave good results. Compared to the experts it is also positive, from the results shown in table 4-3 we can conclude the correlation is stronger between the algorithm and the individual experts than the experts with each other, which indicates a nice in-between performance.

In the infection classification stage our performance is unfortunately deficient, looking at table 4-3. The scores are not enough in agreement with the human experts to be applicable as an automated diagnosis system. The foremost reason for this result could well be the size of the training data set, while there is no hard rule for a minimum dataset size, most neural networks start to perform well with a database of a few thousand samples per class [Cireşan et al., 2012b]. We only have around 50 samples (not augmented) of infected erythrocytes in our database, acquiring more examples of infected erythrocytes should result in a better trained network with higher accuracy.

From closer inspection of the full results in table 4-2 we can see the algorithm has a excess of erythrocytes marked as infected, in 9 out of 10 cases compared to expert 1 and in 4 out of 10 compared to expert 2. This is essentially not a bad thing, proposed in figure 3-1 there is a possibility for human expert validation after the classification stage. We rather see the algorithm returning too much erythrocytes as infected which can be easily checked and discarded by an expert than the algorithm missing infections.

Chapter 5

Discussion

The recent developments by the OSMD project can mean a big improvement to the availability, readiness and costs of malaria diagnosis The low-cost smartphone based microscope system should enable in-the-field optical malaria diagnosis in areas where they currently lack proper diagnostic tools. But the diagnosis problem is a two-fold, not only proper equipment is lacking but experienced personnel is also scarce [Carpenter et al., 1991] [Uzochukwu et al., 2009] . The accuracy of optical malaria diagnosis is heavily dependent on the expertise and reader technique of the microscopist. [O'Meara et al., 2006]. To overcome one of the biggest obstacles of this system and microscopy diagnosis in general, an automated diagnosis tool is proposed. Such a system is not new, various algorithms have been proposed for the quantification of parasites in stained thin blood smears [Linder et al., 2014][Díaz et al., 2009][Di Ruberto et al., 2002][Ross et al., 2006] However, all these algorithms are based on high quality images captured using digital cameras coupled to high-end laboratory microscopes with extremely even illuminations due to Köhler illumination [Society, 1894]. We have shown that the images from the low-cost smartphone based microscope are of too low quality to be analyzed with these conventional algorithms, and so the need for smarter algorithms arises.

In this thesis we presented an automated malaria diagnosis algorithm that analyzes smartphone acquired blood smear images. The problem is split in two stages, the first stage is to localize all erythrocytes so their locations and total number is known, hereafter in the second stage the erythrocytes are classified as either healthy or infected. This algorithm can serve as an independent diagnostic tool to calculate the parasitemia of a patient, which is an indication for the severeness of the disease. It can also serve as a diagnostic aid for human experts to lower workload and increase accuracy.

By posing the problem in such a way to enable the use of Convolutional Neural Networks for the localization of erythrocytes, the problems that arose due to low quality images with uneven lightning, were solved. Where previous studies used tools like morphological operators or edge detection algorithms, as explained in section 2-3, these are shown to be insufficient for the smartphone images. The visual appearance of different erythrocytes in the same smartphone image differ too much as opposed to those from the high-end light microscope where

all erythrocytes closely resemble each other. A Convolutional Neural Network is trained to recognize these different erythrocytes and to base its judgment on a deeper understanding of the shape, texture and colour of the erythrocyte rather than a difference in pixel intensity as a morphological threshold method would. Via this more advanced method, not only was the network able to localize the erythrocytes in the low quality images with a high average sensitivity of 97.31 % and precision of 92.21 %, but was also able to deal with clustered and overlapping erythrocytes without extra processing steps as other researches have implemented [Linder et al., 2014][Díaz et al., 2009][Di Ruberto et al., 2002][Ross et al., 2006]. This provides a great perspective for the use Convolutional Neural Networks in the erythrocyte localization problem. Not only to employ this algorithm in the low-quality smartphone based microscope system. But also to work in an on-field setting where blood smears are often of sub-optimal quality compared to the laboratory setting most researches aim at.

The way we use the Convolutional Neural Network as a pixel classifier to localize multiple objects is not a common practice, actually is not to be found in the literature in this way. One of the downsides is the huge amount of data to be processed this way, where other researches use high quality, proper smeared blood samples the need for this more complex algorithm is absent. However, there is a lot of improvement to be made concerning efficiency. The proposed theory given in section 3-2 can mean huge improvements to the algorithm efficiency, this combined with pooling layers and the sparse network characteristic from the ReLu activations can mean very fast computing networks. Once these networks efficiency matches the conventional techniques they could provide benefits also for the high quality images. In sub-optimal quality blood smears that show more clustering or overlap of erythrocytes, we believe CNNs could provide the solution, however a detailed comparison using the same input data and different algorithms should prove this first.

The second phase, the classification of erythrocytes performed inadequate. The comparison with the human experts showed too little agreement, specifically, a correlation coefficient of 0.65 with expert 1 and 0.73 with expert 2 on the estimated parasitemia. Note must be made that the correlation of 0.90 between both experts is also considered quite low, as other researches report correlations between 0.97 and 0.99 [Díaz et al., 2009] [Le et al., 2008]. This lower correlation can be the cause of difference in experience between both the experts but is more likely to be the result of the low image quality, since both experts reported to feel uncertain in classifying particular erythrocytes due to the image quality.

On a positive note, the classification network annotated more erythrocytes as infected compared to the experts in 9 out of 10 cases compared to expert 1 and 4 out of 10 compared to expert 2. Closer inspection showed that the experts annotations were often subsets of the algorithms annotations. This is an encouraging result to use the network as an diagnostic aid, since the excess of infections can be easily checked by experts while still having confidence that no infections are missed by the algorithm.

The best result so far on erythrocyte classification is reported by Diaz [Díaz et al., 2009], which classifies erythrocytes based on infections and also in life stages (ring/trophozoite or schizont infection). Diaz uses machine learning in the form of a SVM (support vector machine) and MLP (multilayer perceptron neural network) to perform the classification. A big difference with Diaz and a very probable cause for our lower performance is the size of our training data set, with only 50 (not augmented) samples of infected erythrocytes, compared to over 600 samples of Diaz, the training set is very small to train a CNN [Cireşan et al., 2012b].

Another big difference is the image quality, with various luminance, smudged texture and edge distributions, and variations in saturation levels per individual erythrocyte from the same smartphone input image, the classification task becomes harder. We do believe convolutional neural networks could provide the outcome for this problem, especially Cirecsan, [Cireşan et al., 2012a] shows promising results on the classification of traffic signs that deals with similar differences in detail and contrast using a multi-column deep neural network. Cirecsan has around 1000 training images per class. Future work should prove the increase in performance when using a larger training database. Also with an increase in size of the training database different possibilities open up for deeper convolutional neural networks, multi-column networks and other state-of-the-art solutions.

Chapter 6

Conclusion

Smartphone microscope technology has great potential to lighten the burden of malaria in remote sub-Saharan Africa. However, the accuracy of optical malaria diagnosis depends on the human expert. We propose an automated diagnostic system that can exclude the human expert from the process or serve as an aid for an expert to lower workload and improve accuracy. The images made by the low-cost smartphone based microscope are of low quality compared to the usual images captured in laboratory settings with digital cameras coupled to high-end light microscopes. We propose an algorithm, based on convolutional neural networks, that can deal with the problems inherent in low quality images. We have shown great performance in localizing the erythrocytes, at least on-par with two human experts. The classification performance is considered to be inadequate because of its too low agreement with the human experts. We must, however, note that the experts were also novices to the diagnosis based on these images and no ground truth could be established due to the image quality. The algorithm efficiency was of low priority up to this point but should be considered to enable the algorithm to operate on a smartphone. We have proposed a method to considerably increase efficiency that should be implemented in future work.

To conclude and summarize our contributions. We have investigated and proposed a solution to the diagnosis of malaria in low quality blood smear images. The quality of these images presented problems that conventional techniques of similar researches on automated malaria diagnosis could not solve. We have split the problem in two parts, first localization of erythrocytes and secondly the classification of these as healthy or infected. By using the framework proposed in this thesis, so that convolutional neural networks can be used for the localization of erythrocytes we believe to have found a suitable solution for the first part. The second part, the classification, is performing inadequate to serve as an automated diagnosis tool. Based on other sources in the literature, we do believe that the convolutional neural network provided here can offer the solution to this problem. Before the peer review journal will be published, new experiments will take place. These should prove the increase in performance with larger training data sets.

The battle against the devastating burden of malaria will continue. Early and accurate diagnosis is one of the keystones in the fight against this disease [WHO, 2016]. This thesis, as a part of the OSMD project at the TU Delft, is a component in the development of a tool that will bring a healthy and malaria-free Africa one step closer.

Bibliography

[Agbana et al., 2017] Agbana, T. E., Diehl, J.-C., Pul, F. v., Khan, S. M., Verhaegen, M., and Vdovin, G. (2017). Imaging identification of malaria parasites using cellphone microscope with a ball lens.

[Altman and Bland, 1994] Altman, D. G. and Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, 308(6943):1552.

[Blumberg and Frean, 2007] Blumberg, L. and Frean, J. (2007). Malaria control in south africa-challenges and successes. *South African Medical Journal*, 97(11):1193–1197.

[Carpenter et al., 1991] Carpenter, C. C., Pearson, G. W., Mitchell, V. S., Oaks Jr, S. C., et al. (1991). *Malaria: obstacles and opportunities*. National Academies Press.

[Cireşan et al., 2012a] Cireşan, D., Meier, U., Masci, J., and Schmidhuber, J. (2012a). Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32:333–338.

[Cireşan et al., 2013] Cireşan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 411–418. Springer.

[Cireşan et al., 2012b] Cireşan, D. C., Meier, U., and Schmidhuber, J. (2012b). Transfer learning for latin and chinese characters with deep neural networks. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–6. IEEE.

[Di Ruberto et al., 2002] Di Ruberto, C., Dempster, A., Khan, S., and Jarra, B. (2002). Analysis of infected blood cell images using morphological operators. *Image and vision computing*, 20(2):133–146.

[Díaz et al., 2009] Díaz, G., González, F. A., and Romero, E. (2009). A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *Journal of Biomedical Informatics*, 42(2):296–307.

[Glorot et al., 2011] Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323.

[Hecht-Nielsen, 1988] Hecht-Nielsen, R. (1988). Theory of the backpropagation neural network. *Neural Networks*, 1(Supplement-1):445–448.

[Hinton et al., 2012] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

[Le et al., 2008] Le, M.-T., Bretschneider, T. R., Kuss, C., and Preiser, P. R. (2008). A novel semi-automatic image processing approach to determine plasmodium falciparum parasitemia in giemsa-stained thin blood smears. *BMC Cell Biology*, 9(1):15.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[Linder et al., 2014] Linder, N., Turkki, R., Walliander, M., Mårtensson, A., Diwan, V., Rahtu, E., Pietikäinen, M., Lundin, M., and Lundin, J. (2014). A malaria diagnostic tool based on computer vision screening and visualization of plasmodium falciparum candidate areas in digitized blood smears. *PLoS One*, 9(8):e104855.

[McKenzie et al., 2003] McKenzie, F. E., Sirichaisinthop, J., Miller, R. S., Gasser Jr, R. A., and Wongsrichanalai, C. (2003). Dependence of malaria detection and species diagnosis by microscopy on parasite density. *The American journal of tropical medicine and hygiene*, 69(4):372–376.

[Meyer, 1994] Meyer, F. (1994). Topographic distance and watershed lines. *Signal processing*, 38(1):113–125.

[Monasse and Guichard, 2000] Monasse, P. and Guichard, F. (2000). Fast computation of a contrast-invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872.

[Moonasar et al., 2007] Moonasar, D., Goga, A. E., Frean, J., Kruger, P., and Chandramohan, D. (2007). An exploratory study of factors that affect the performance and usage of rapid diagnostic tests for malaria in the limpopo province, south africa. *Malaria journal*, 6(1):74.

[Nair and Hinton, 2010] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

[O'Meara et al., 2006] O'Meara, W. P., Barcus, M., Wongsrichanalai, C., Muth, S., Maguire, J. D., Jordan, R. G., Prescott, W. R., and McKenzie, F. E. (2006). Reader technique as a source of variability in determining malaria parasite density by microscopy. *Malaria Journal*, 5(1):118.

[Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

[Rawat and Wang, 2017] Rawat, W. and Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*.

[Ross et al., 2006] Ross, N. E., Pritchard, C. J., Rubin, D. M., and Duse, A. G. (2006). Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Medical and Biological Engineering and Computing*, 44(5):427–436.

[Schmidhuber, 2015] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–117.

[Sherman, 1979] Sherman, I. W. (1979). Biochemistry of plasmodium (malarial parasites). *Microbiological reviews*, 43(4):453.

[Shillcutt et al., 2006] Shillcutt, S., Morel, C., Coleman, P., Mills, A., and Goodman, C. (2006). Cost-effectiveness of malaria diagnosis in sub-saharan africa: the role of rapid diagnostic tests in rural settings with high plasmodium falciparum transmission.

[Sio et al., 2007] Sio, S. W., Sun, W., Kumar, S., Bin, W. Z., Tan, S. S., Ong, S. H., Kikuchi, H., Oshima, Y., and Tan, K. S. (2007). Malariacount: an image analysis-based program for the accurate determination of parasitemia. *Journal of microbiological methods*, 68(1):11–18.

[Society, 1894] Society, R. M. (1894). *Journal of the Royal Microscopical Society*. Royal Microscopical Society.

[Tek et al., 2009] Tek, F. B., Dempster, A. G., and Kale, I. (2009). Computer vision for microscopy diagnosis of malaria. *Malaria journal*, 8(1):153.

[Theerapattanakul et al., 2004] Theerapattanakul, J., Plodpai, J., and Pintavirooj, C. (2004). An efficient method for segmentation step of automated white blood cell classifications. In *TENCON 2004. 2004 IEEE Region 10 Conference*, pages 191–194. IEEE.

[Uzochukwu et al., 2009] Uzochukwu, B. S., Obikeze, E. N., Onwujekwe, O. E., Onoka, C. A., and Griffiths, U. K. (2009). Cost-effectiveness analysis of rapid diagnostic test, microscopy and syndromic approach in the diagnosis of malaria in nigeria: implications for scaling-up deployment of act. *Malaria Journal*, 8(1):265.

[WHO, 2000] WHO (2000). New perspectives: malaria diagnosis. report of a joint who/usaid informal consultation, 25-27 october 1999. *New perspectives: malaria diagnosis. Report of a Joint WHO/USAID Informal Consultation, 25-27 October 1999*.

[WHO, 2010] WHO (2010). *Basic Malaria Microscopy: Tutor's guide*. World Health Organization and Center for Disease Control.

[WHO, 2015] WHO (2015). *Global Technical Strategy for Malaria 20162030*. World Health Organization.

[WHO, 2016] WHO (2016). *World malaria report 2016*. World Health Organization.

[Wongsrichanalai et al., 2007] Wongsrichanalai, C., Barcus, M. J., Muth, S., Sutamihardja, A., and Wernsdorfer, W. H. (2007). A review of malaria diagnostic tools: microscopy and rapid diagnostic test (rdt). *The American journal of tropical medicine and hygiene*, 77(6_Suppl):119–127.

[Xu et al., 2015] Xu, B., Wang, N., Chen, T., and Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

Appendix A

Appendix A: Journal paper: High sensitive malaria diagnosis using convolutional neural networks in an on-field setting

The following journal is written in parallel to this thesis and is aimed to be published in the near future. Please note that before publication, new experiments will be carried out and therefore, the content and results of this journal is subject to change.