# MULTISPECTRAL IMAGING OF HUMAN BLOOD MEDIA

# APPLIED TO MALARIA DIAGNOSTICS

By

Linani Dickson Omucheni, B.Sc (Hons)

I56/72351/2008

A thesis submitted in partial fulfillment of the requirement of Master of Science (M.Sc) in Physics, at the University of Nairobi.

March, 2012.

## Declaration

This work is an original thesis submitted for the Degree of Master of Science (M.Sc) in the Department of Physics, University of Nairobi, and has not been submitted before for examination at any other institution/University.

Linani Dickson Omucheni

I56/72351/08

Department of Physics, University of Nairobi.

Signature......................... Date.....5-3-2012.......

This thesis is submitted with our approval as University of Nairobi Supervisors.

Dr. Kaduki K. A.

Department of Physics, University of Nairobi.

Signature...................... Date......5/3/2012......

Dr. Angeyo H. K.

Department of Physics, University of Nairobi.

Signature..................:.... Date....5-3-2012......

Dr. Bulimo W. D.

Department of Biochemistry, University of Nairobi &

Kenya Medical Research Institute (KEMRI).

Signature...................... Date......5/3/2012......

i

**Abstract**

Malaria is a serious disease whose confirmation in laboratories, usually by optical microscopy, remains a challenge. The process involved in preparation of blood films and their examination in the conventional microscopes is time consuming, and the results may vary depending on the expertise of the examining technician, putting the lives of patients at risk. Other laboratory techniques for confirming malarial infection are expensive and may involve procedures that require extra training of personnel.

In this thesis, a method for rapid detection of *Plasmodia* (malaria parasites) in unstained thin blood smears has been developed. The method is based on microscopically imaging red blood cells using different wavelengths of light in the UV, visible and NIR region for illumination (an emerging field known as multispectral imaging microscopy). Imaging was accomplished by use of an optical microscope modified by replacing its tungsten light source with a set of light emitting diodes (LEDs) whose emission spectra were centered at 375 nm, 400 nm, 435 nm, 470 nm, 525 nm, 590 nm, 625 nm, 660 nm, 700 nm, 750 nm, 810 nm and 940 nm. These LEDs were made to illuminate the samples from different orientations to obtain transmittance, reflectance, and dark-field images to reveal different optical properties of the imaged sample. The microscope was fitted with a monochrome CMOS camera that recorded the intensity of light emanating from the specimen as gray-level images. Using *in vitro* cultures of red blood cells infected with *Plasmodium falciparum* prepared as thin smears and without any stain, congruent intensity images were recorded at different wavelengths and used to compose a multispectral image. Pixels of the multispectral image were used to generate spectral

signatures (with the assistance of multivariate chemometric techniques) to identify infected and non-infected red-blood cells.

Results obtained using Principal Components Analysis (PCA) and Hierarchical Cluster Analysis (HCA) show that detection and identification of malaria parasites in multispectral images is possible in the 375-940 nm range and relies on presence of hemozoin (a pigment generated by the parasites when they digest haemoglobin). Artificial Neural Network (ANN) model shows promising results in accurately discriminating  infected red blood cells from non-infected ones and in determining parasite burden (parasitemia) in unstained human blood.

The developed method is rapid, with the whole diagnostic process taking approximately 10 minutes as opposed to the conventional microscopy which takes about 30 minutes.

## Dedication


To my family

## Acknowledgements

**Table of Contents**

## List of Symbols

| | |
|---|---|
| $S_k(\lambda)$ | Spectral sensitivity of a camera in the k-th band |
| $L(\lambda)$ | Spectral intensity of illumination |
| $z(\lambda)$ | Spectral reflectivity or transmission of an object |
| $x_k$ | Pixel value in the k-th band |
| $I_o$ | Incident intensity of light |
| $I_t$ | Transmitted intensity of light |
| $A$ | Absorbance |
| $\lambda$ | Wavelength of light |
| $I(\lambda)_s$ | Multispectral image of a sample |
| $I(\lambda)_r$ | Multispectral image of a standard reference |
| $I(\lambda)_d$ | Multispectral dark image |
| $I(\lambda)_{spec}$ | Corrected multispectral image |
| $I_{max}$ | Maximum intensity value |
| $I_{min}$ | Minimum intensity value |
| $I_{norm}$ | Normalized intensity value |
| $I_{act}$ | Actual (original) intensity value |

## List of Abbreviations & Acronyms

| | |
|---|---|
| ANN | Artificial Neural Network |
| CMOS | Complementary Metal Oxide Semiconductor |
| DAQ | Data Acquisition Card |
| FTIR | Fourier Transform Infra-Red |
| FWHM | Full Width at Half Maximum |
| HCA | Hierarchical Cluster Analysis |
| HRP-II | Histidine-Rich Protein-II |
| LED | Light Emitting Diode |
| MIA | Multivariate Image Analysis |
| PCA | Principal Component Analysis |
| PCR | Polymerase Chain Reaction |
| RDT | Rapid Diagnostic Test |
| RGB | Red-Green-Blue |
| SIMCA | Soft Independent Modeling of Class Analogies |
| THG | Third Harmonic Generation |
| WHO | World Health Organization |

## List of Figures

## List of Tables

## Chapter 1

## INTRODUCTION

### 1.1 Preamble

Malaria is a serious vector-bone disease that causes significant morbidity and mortality in human beings worldwide. According to the 2009 World Health Organization (WHO) statistics, close to one million of the 243 million people infected by the disease succumbed to this menace in the year 2008 [1]. The disease is caused by any of the following four species of *Plasmodium*: *falciparum, vivax, ovale* and *malariae*, which are introduced into human blood by the saliva of a feeding female anopheles mosquito.

Many control approaches to malaria exist, but one of the major challenges faced in its treatment lies in diagnosis for a suspected infection. Clinical diagnosis (identification of the disease by observation of symptoms exhibited by a patient) is usually based on history or presence of fever and is thus very unreliable because malaria symptoms are not very specific; some patients may not have fever at all and may present other symptoms such as headache, backache, joint pains, and dizziness among others. Presumptive treatment based on clinical diagnosis is associated with over-administration of drugs and erroneous treatment of fever caused by other infections which can lead to development of resistance to drugs by malaria parasites [2]. WHO recommends prompt parasitological confirmation of malaria before start of any treatment, with specific drug administration based on the species of *Plasmodium* present [1].

Examination of stained thick and thin blood smear in a conventional light microscope is currently the "gold standard" for parasitological confirmation of malarial infection [1]. This is done by a trained microscopist or laboratory technician who examines the blood smear after careful preparation [3]. In countries where malaria is uncommon, such as in those outside the tropics, malaria parasites may not be easily recognized by the examining technicians who have insufficient experience in interpreting blood smears. In such a case, the non-immune victim, typically a recently returned traveler is at a great risk of death especially if infected by *Plasmodium falciparum*. In malaria endemic areas, on the other hand, the magnitude of the task and lack of enough skilled microscopists contributes to either misdiagnosis [2] or late diagnosis of the disease.

Conventional microscopy has several advantages, including the ability to speciate and quantify the parasites. However, examination of blood smears is not only labour-intensive, but also a subjective process. For instance, an inexperienced technician may confuse artefacts in a blood smear (such as dye precipitates) or other stained objects (such as blood platelets) with malaria parasites, hence produce false results. The method, therefore, requires the services of experienced personnel.

The limitations of conventional light microscopy in malaria diagnostics has stimulated development of alternative methods with a constraint that they should be rapid, sensitive, easy to use and affordable in order to be adopted in both endemic and non- endemic areas. These methods include: immunochromatographic tests (commonly known as Rapid Diagnostic Tests (RDTs)) which detect specific antigens derived from malaria

parasites in lysed blood [4, 5]; Polymerase Chain Reaction (PCR)-based techniques that detect specific nucleic acid sequences [6-8]; and spectroscopic techniques such as, among others, Third Harmonic Generation (THG) imaging [9] and Raman micro-spectroscopy [10-11], which are based on detection of hemozoin (a digestion residue of the malaria parasite) as a biomarker for identifying malarial infection.

Rapid Diagnostic Tests (RDTs) have already gained acceptance by WHO as an alternative parasitological confirmation of malarial infection [1]. However, RDTs do not allow the user to quantify parasite burden (parasitemia), which is necessary for establishing the severity of the disease. Furthermore, the available dipsticks are mostly specific to *Plasmodium falciparum* and their costs are still prohibitive for widespread application. Similarly, the main hindrance to adoption of PCR and spectroscopic methods lies in the costs involved.

Image processing and analysis is an emerging novel method for applications in machine vision. Systems based on RGB (Red-Green-Blue, i.e. three visible light spectral channels) colour image processing have been proposed for use in automated malaria parasite detection and quantification in microscopy [12-14]. Imaging in three spectral sensitivity bands (such as that in RGB images) is, however, limited in spectral information about the absorption properties of the sample which is valuable for segmentation and classification of objects in microscopic images.

Multispectral imaging involves recording both spatial and spectral data from a sample in order to capture more information from the objects under study [15-22]. Analyses of the large set of variables encountered in multispectral images, which are sometimes correlated, require the use of multivariate statistical tools to extract useful information. The work presented in this thesis combines multispectral imaging with multivariate chemometric techniques to detect, discriminate and quantify malaria parasites in human blood media for the purpose of performing rapid malaria diagnosis. This is accomplished by imaging unstained thin blood smears using a simple multispectral imaging microscope built by replacing the conventional broadband light source (commonly used in optical microscopes) with a set of 13 quasi monochromatic Light Emitting Diode (LED) sources in conjunction with a monochrome camera. The camera captures a gray-level image for each LED illumination, giving a set of 13 gray-level congruent images that compose the multispectral image. Analysis is then carried out by employing Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and Artificial Neural Network to detect, identify and discriminate infected red blood cells from non-infected ones.

## 1.2 Statement of the Problem

Conventional microscopy, which is currently the 'gold standard' for laboratory confirmation of malarial infection, is time consuming and the results are sometimes not reliable because they depend on the expertise of the microscopist examining the specimens. Advancement of Information Technology (IT) exploits the ever-increasing computation power of computers to incorporate image processing in automated disease diagnosis. However, the existing imaging methods designed for automation have not

4

been successful in malaria diagnostics due to their exclusive reliance on the morphology of objects in the blood smear. On the other hand, spectroscopic techniques, which use spectral signatures to detect hemozoin (a malaria pigment that is generated during degradation of haemoglobin by the malaria parasites), use specialized equipment such as, Raman spectrometers, FTIR micro-spectrometers, among others which are rarely available in places where the disease is endemic, hence making them unpopular. Multispectral imaging microscopy of unstained thin blood smear is a potential superior concept to develop by exploiting Multivariate Image Analysis (MIA) techniques to detect, identify, discriminate and quantify malaria parasites in human blood.

## 1.3 Objectives of the Study

The goal of this research was to develop a microscopic system for comprehensive malaria diagnostics in human blood media based on spectral signatures of *Plasmodia* extracted from multispectral images. The specific objectives were:

(i) To identify and characterize the spectral fingerprints of the genus *Plasmodium* in human blood media in the spectral range 375 nm to 940 nm.

(ii) To identify the important spectral signature responsible for discrimination of malaria parasites in blood by exploiting an exploratory chemometric technique (Principal Component Analysis).

(iii) To examine the distribution of parasite components in an infected red blood cell by employing a clustering technique (Hierarchical Cluster Analysis).

5

(iv) To utilize the spectral information from the multispectral microscopy images to develop a supervised pattern recognition model for fast screening of human blood for presence of malaria parasite infection.

(v) To compare the performance of the developed model with results obtained based on conventional light microscopy.

## 1.4 Justification and Significance of the Study

Malaria remains a serious worldwide health concern. For this reason, there is need for quick and accurate diagnosis of malaria. Development of a *Plasmodium* detection method that has the capability to result in automation of malaria diagnostics is of great significance to malaria research. Multispectral imaging microscopy of malaria parasites has the potential of replacing a trained human eye with a trained vision instrument. This will result in improved speed and accuracy in malaria diagnostics translating to effective management of the disease.

# LITERATURE REVIEW

## 2.1 Introduction

This chapter reviews the current methodology and approaches to diagnosis of malaria. These range from the labour-intensive routine laboratory examination of stained blood smears (films) in light microscopes to sophisticated spectroscopic analysis of blood components or application of advanced molecular biology techniques such as Polymerase Chain Reaction (PCR) procedures. Application of digital image processing to the malaria diagnostics problem and the utility of multispectral imaging to myriad applications involving parasite detection and disease diagnostics are also reviewed.

## 2.2 Observation of Malaria Parasites in Conventional Microscopy

Examination of stained blood film under a conventional light microscope is currently the "gold standard" for laboratory confirmation of malaria. Peripheral blood from a finger (or the heel in young infants) is smeared on a glass slide, stained and fixed (for thin blood smear) to highlight the malaria parasites. Giemsa staining is the most commonly used method and allows identification of all the four species of *Plasmodia*, stage and parasitemia (relative quantity of parasites in blood) [1]. The staining process can take up to 30 minutes for thin film preparation while examination of 100 fields (a field is area that is visible in a microscope) in a thick blood smear (the film preparation method that is

a bit faster), apparently, may miss infection of up to 20 % [23] depending on the experience of the personnel involved.

Conventional optical microscopy is labour-intensive as the microscopist or the examining technician has to examine as many red blood cells as possible on a glass slide with his/her own eye. A correct diagnosis is done only after attentive and careful examination of at least 100 microscopic fields in thick blood smear and a number of morphological characteristics have been drawn to identify the species in thin blood smear. The time spent in the staining process and manual examination of many fields is the cause of delays in releasing of laboratory results in clinics and hospitals. In addition, the results are subjective and may not be reproducible depending on the expertise of the examining personnel and the time spent on examination of the specimen.

The aforementioned limitations experienced in conventional light microscopy have stimulated research in development of alternative diagnostic methods that could be rapid, sensitive and affordable. Optical enhancement techniques such as dark field microscopy can be employed to significantly reduce the time spent on parasite staining in conventional transmission light microscopy. Dark-field microscopy is a contrast enhancement technique which is based on abstraction (exclusion) of undiffracted beams of light in image formation in the microscope. Jamjoom [30] used dark-field microscopy to detect presence of malaria parasites in unstained blood films, and observed that intra-erythrocytic stages of *Plasmodium* (rings, trophozoites and schizonts) and gametocytes were visible as bright objects enclosed in red blood cell outlines against a dark

background. This observation may be understood in terms of Huygen's principle that "If a wave-front strikes an interface between materials with different refractive indices at angle different from $90^{\circ}$, every point of the interface becomes a source of spherical waves". However, identification of the various stages of the parasites was difficult and, like in conventional light microscopy, the method is tedious due to manual examination.

## 2.3 Fluorescence-Based Malaria Diagnostic Techniques

Fluorescent dyes such as acridine orange and benzothiocarboxypurine are nucleic acid selective fluorescent cationic dyes that can bind on DNA and RNA of a biological cell in a process known as electrophoresis [24]. The compound formed fluoresces in the visible region when excited by UV light of appropriate wavelength, hence nucleic acid-containing cells such as malaria parasites are highlighted in a blood smear. A microhaematocrit tube containing acridine orange stain and an anticoagulant is filled with a suspected malaria infected blood obtained by the conventional finger pricking method. It is then centrifuged and immediately observed under a microscope fitted with ultraviolet light source. Blood cellular components such as erythrocytes are separated based on density. The parasites are seen as fluorescent bodies standing at different levels of the sedimentation column depending on their stages. Alternatively, acridine orange is used to stain thin blood films which are then observed under a fluorescent microscope fitted with optical filters that allow only the excitation wavelength of the dye to pass to the stained film [25].

9

Although fluorescent-based techniques are fairly quick and easy to use, they have limited performance when other species of *Plasmodium* other than *falciparum* are involved and in parasitemia determination (quantification of relative content of parasites in the blood). These limitations pertain to the fact that the morphology of the parasites is not apparent when fluorescent dyes are used to highlight the parasites [26].

## 2.4 Immunochromatographic Tests for Malaria Antigens

Immunochromatography relies on the migration of liquid across the surface of a nitrocellulose membrane. Immunochromatographic test (commercially presented as Rapid Diagnostic Test kits) are based on the capture of parasite antigen from human blood using monoclonal antibodies prepared against a malaria antigen target and conjugated to a visually detectable marker such as selenium dye or colloidal gold in a mobile phase. A second or third capture monoclonal antibody applied to a strip of nitrocellulose acts as the immobile phase. Blood specimen obtained from a finger-prick is mixed with a buffer solution that contains a hemolyzing compound and applied to the strip. If the target antigen is present in the blood, a labeled antigen/antibody complex is formed and it migrates up the test strip to be captured by the monoclonal antibody of the immobile phase, thus producing a visible coloured line [4, 5].

Malaria antigens currently targeted by Rapid Diagnostic Test kits are Histidine-Rich Protein-2 (HRP-II), Parasite lactate dehydrogenase (pLDH) and *Plasmodium* aldolase. The kits are easy to use and do not require expertise and well established laboratory.

10

However, most RDT kits are based on HRP-II immunochromatography; hence they are only useful for diagnosis of *Plasmodium falciparum*, the species that produce HRP-II. It has also been reported that in some cases the antibody used in the kits reacts with serum rheumatoid factor giving a false positive result, or some individuals may have a gene deletion for the production of HRP-II and so never give positive results. Aldolase and pLDH have been proposed for detection of non *falciparum* infection with positive results reported in *Plasmodium vivax* [4, 5]. However, the cost of RDT kits is still high with price ranging from US$ 1.2-13.5 depending on geographical region [1].

## 2.5 Detection of Malaria Parasites' Nucleic Acid Sequences

Polymerase Chain Reaction (PCR) is a technique in molecular biology that is used to target and amplify any specific nucleic acid from a complex biological sample. In disease diagnostics, the procedure is used to determine whether a clinical sample contains nucleic acid sequence that is known to occur only in a specific pathogen. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (short DNA fragments) containing sequences complementary to the target region along with DNA polymerase (an enzyme that helps to catalyze in the polymerization of deoxyribonucleotides into DNA strand) are the components that enable selective and repeated amplifications. As PCR progresses, the DNA generated is itself used for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. A number of PCR assays have been developed for the detection

of malaria. The major advantages of using PCR-based technique are the ability to detect malaria parasites in patients with low levels of parasitemia and to identify them to the species level. Infections with five parasites or less per micro-litre of blood can be detected with 100 % sensitivity and equal specificity [6-8]. However, PCR-based techniques remain unutilized because they are time-consuming, involve complicated procedures and require a well established laboratory.

## 2.6 Spectroscopic Investigation of Malarial Infection

Spectroscopic techniques such as Third Harmonic Generation (THG) imaging spectroscopy [9], resonance Raman micro-spectroscopy [10-11], and Electron Paramagnetic Resonance (EPR) spectroscopy [27] among others have been used in the investigation of malaria pigment (hemozoin) in malaria drug-target interaction research and in malaria diagnostics studies.

Ong *et al.* [10] employed resonance Raman microspectroscopy to study normal and *Plasmodium berghei*-infected mouse erythrocytes. They found that the Raman spectra for infected and non-infected red blood cells could be differentiated by the bands at 754 $cm^{-1}$ and 747 $cm^{-1}$ respectively. They attributed the slight decrease in frequency between 754 $cm^{-1}$ and 747 $cm^{-1}$ as a result of infection to the fact that, in hemozoin (a polymer formed as a result of haemoglobin degradation by malaria parasites), the porphyrin structures are more packed, resulting in stronger vibrational mode than in haemoglobin. They also observed that the averaged Raman spectra of infected erythrocytes were

12

identical to Raman spectra of β-hematin (a synthetic analog of hemozoin), implying that hemozoin was indeed the spectroscopic biomarker of *Plasmodium* infection.

Belislie *et al.* [9] developed an optical-based malaria detection method using Third Harmonic Generation (THG) imaging of hemozoin using ultra-fast (femto-second), infra-red pulsed laser excitation. THG is a material–dependent non-linear optical effect that involves conversion of fundamental laser frequency to a tripled frequency due to third-order dielectric susceptibility of the material. The results showed that hemozoin yields up to three orders of magnitude stronger THG signal than any other component found in blood. This is because hemozoin contains a condensed phase of iron (III) porphyrin rings which have an extensive polarizable π-electron system as opposed to iron (II) porphyrin rings found in haemoglobin.

Yulia *et al.* [28] used UV-visible spectroscopy to study intra-erythrocytic stages of *Plasmodium falciparum*. Using the first derivatives of the absorption spectra, they enhanced the differences in the absorption spectral features of infected red blood cells. They observed an absorption peak centered at 650 nm whose size increased as the parasites progressed from ring to schizont stage. This peak was due to hemozoin. They also observed that the characteristic oxy-haemoglobin doublet at 540 nm and 575 nm in the haemoglobin spectrum was more pronounced in trophozoite stages but reduced in the schizont stage. The reduction in this doublet is the result of degradation of haemoglobin into the polymeric form of heme (hemozoin).

Webster *et al.* [29] discriminated intra-erythrocytic stages of *Plasmodium falciparum* by employing synchrotron Fourier Transform Infra-Red (FT-IR) micro-spectroscopy in conjunction with principal component analysis. In FT-IR difference spectra between uninfected and infected erythrocytes, they observed bands assigned to hemozoin moiety - 1712 $cm^{-1}$, 1664 $cm^{-1}$ and 1209 $cm^{-1}$. By employing principal component analysis to investigate spectra of parasites at different stages, they found that these bands were important contributors in separating between spectra of infected and non-infected erythrocytes.

The spectroscopic investigations of malaria infection have two things in common: (i) they check for the presence of malaria pigment which has a different spectral response from that of haemoglobin, the dominant protein in the red blood cells; and (ii) they require specialized equipment such as spectrometers, and laser systems. Such a requirement is untenable for application in poor countries where malaria is endemic because they are very expensive. In addition, malaria diagnostics based on exclusive hemozoin detection may miss ring stage of *Plasmodium* since hemozoin is more pronounced in the mature trophozoites and schizont stages of *Plasmodium*.

## 2.7 Morphological Digital Image Processing in Malaria Diagnostics

Semi-automated classification and quantification of erythrocytes infected with malaria parasites is the beginning of use of Information Technology (IT) in malaria diagnostics and it is based on image processing techniques [12-14]. Development of automated systems for malaria diagnosis can greatly enhance reproducibility of diagnostic results

14

and reduce the time spent on examining specimens. Ross *et al.* [13] developed an image processing algorithm to automate diagnosis of malaria on thin blood smears based on morphological and threshold selection techniques. They used features based on RGB colour space, texture and geometry of the cellular components of blood and the parasites to classify infections in a two-stage tree classifier employing back propagation feed forward neural networks.

Minh *et al.* [12] and Gloria *et al.* [14] developed semi-automatic methods for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. For Minh *et al.* [12], the images were first analyzed for nucleated components based on the differences between intensity values of different colour channels of RGB images, then decomposed into solid and non-solid matters in which the latter was considered the background. Parasitemia was estimated by differentiating white blood cells, red blood cells and gametocytes based on estimated average size. Gloria *et al.* [14] on the other hand, used normalized RGB colour space to classify pixels as either erythrocyte or background followed by a two step classification process based on the histograms of the images to identify infected from non- infected red blood cells for parasitemia estimation. They allowed user intervention whenever the approach failed; for example in the case when the classifier was not able to allocate an erythrocyte to a unique class, it was left to the user to classify the red blood cell by visual inspection.

Systems based on morphological, size and geometrical mathematical operators for detection of parasites [12-14] are necessary but not sufficient for accurate diagnosis of

malaria because they rely on RGB colour images which have limited spectral information that can aid in object identification. RGB colour imaging mimics human eye and hence samples have to be stained in order for the algorithms to work. The process of staining a thin blood smear (which is ideal for RGB image processing) is the major cause of delay in obtaining diagnostic results because the staining process alone can take up to 20 minutes [23].

## 2.8 Multispectral and Hyperspectral Imaging Applications

Multispectral imaging is a novel imaging technology that integrates spatial and spectral properties of an object under observation in order to record more information. Its use has been reported in a number of applications including, detection of parasites in cod fillets [16], cell differentiation in thyroid cytology [17], white blood cell segmentation [18,22], detection and classification of latent defects and diseases on raw French fries [20], mapping of lipid oxidation in chicken meat [21], identification of faecal contamination on apples [31], measurement of ripeness in tomatoes [32], defect detection in cherries [33] and detection of malignant melanomas on skin lesions [34], among others. The utility of multispectral imaging owes its success to the fact that in addition to recording the morphological aspects of an object (e.g. a cell), its interaction with electromagnetic radiation at different wavelengths can be used to reveal its spectral signature. Such information is valuable for segmentation of objects in the images.

A multispectral image consists of a set of congruent gray-level images, acquired at different wavelengths of light. The gray-levels (the numerical value in each pixel) represent a measured intensity of the radiation used in image acquisition. The number of gray level (intensity) images generally varies from four to a few tens depending on the spectral resolution required or equipment limitations.

Advances in computer technology have led to the development of hyperspectral imaging systems. Unlike multispectral imaging, hyperspectral imaging systems capture image stacks made of a series of tens or hundreds of narrow, adjacent spectral band congruent gray-level images. Applications have been reported in food quality control [35], such as bruise damage detection on white mushrooms [36] and pickling cucumbers [37]. However, due to extensive time required in hyperspectral image acquisition and equipment requirement it is not suited for automated systems but rather for identification of important spectral bands to be used in multispectral imaging systems [37]. Multispectral imaging is an equally powerful basis for segmentation and classification as it can be used to visualize the chemical composition of materials with fewer spectral bands. This thesis incorporates the advantages of multispectral imaging, optical contrast enhancement and Multivariate Image Analysis (MIA) to detect, discriminate and quantify *Plasmodia* in human blood media for the purpose of malaria diagnostics.

## 2.9 Multivariate Chemometric Techniques in Image Analysis

Chemometrics has its origin in chemistry. Its use arose from the realization that traditional univariate statistics was not sufficient in describing and modeling chemical

17

experiments performed with instruments giving multivariate responses for each sample analyzed. Chemometrics overcomes the limitations of univariate statistics in experimental design, multivariate classification and multivariate calibration by exploiting methods derived from mathematics, statistics and computer science. The whole idea of chemometrics is that matrices obtained from a multivariate response instrument contain redundant information and can be reduced substantially to ease interpretation [38].

Multispectral and hyperspectral images are huge three-way arrays, usually of highly correlated data. Multivariate chemometric techniques can therefore be used to generate a smaller number of statistically independent images, each revealing a unique contrast phenomenon. The most commonly used technique for achieving this is Principal Component Analysis (PCA) which has been extensively utilized in food quality analysis such as detection and classification of latent defects and diseases on raw French fries [20], mapping of lipid oxidation in chicken meat [21] and identification of faecal contamination on apples [31], among others. Another multivariate chemometric technique useful in image analysis is Soft Independent Component of Class Analogies (SIMCA) which has been used in detection of parasites in cod fillets [16]. In fact, there is no limit on applicability of chemometric techniques in multivariate image analysis, hence techniques such as Hierarchical Cluster Analysis (HCA), Artificial Neural Network (ANN) and many others can be used.

**Chapter 3**

**THEORETICAL BACKGROUND**

### 3.1 Introduction

In this chapter, the life cycle and diagnostic stages of *Plasmodium* are presented followed by an overview of the biophysical changes introduced by the parasite in blood. An overview of image formation in an optical microscopy, which is currently the "gold standard" for malaria diagnosis, is given together with a general perspective of UV-visible-NIR spectroscopy. The concept of multispectral imaging and Multivariate Image Analysis (MIA), which are the cornerstones of this thesis, are also are discussed.

### 3.2 The Life Cycle of *Plasmodium*

Malaria parasite exhibits a complex lifecycle that involves a human being (and some primates) and an insect vector (female anopheles mosquito). Malaria infection is initiated when sporozoites in mosquito saliva are injected into the human body by a feeding female *Anopheles* mosquito. The sporozoites are carried to the liver by the human circulatory system where they undergo asexual reproduction to form an invasive stage of the parasite known as merozoites. Merozoites are released into the blood stream to invade erythrocytes (red blood cells) and undergo a trophic period in which the parasites enlarge. The early trophozoite stage is known as 'the ring stage' due to its morphology. Trophozoite enlargement is accompanied by active metabolism which includes ingestion

19

of host cytoplasm and proteolysis of haemoglobin into amino acids. The end of the trophic period culminates into schizonts, an intra-erythrocytic parasite stage involving repetitive nuclear division. Merozoites bud from mature schizonts and are released following rapture of the infected erythrocytes which reinitiates the cycle by attacking other red blood cells.

As an alternative to the trophic replicative cycle, the merozoites can differentiate into male and female sexual gametes known as macro and microgametocytes respectively. Ingestion of gametocytes by a feeding mosquito facilitates the completion of the vector (sexual) stages of the *Plasmodium* life cycle [39]. The complete life cycle of malaria parasites is illustrated in Figure 3.1.
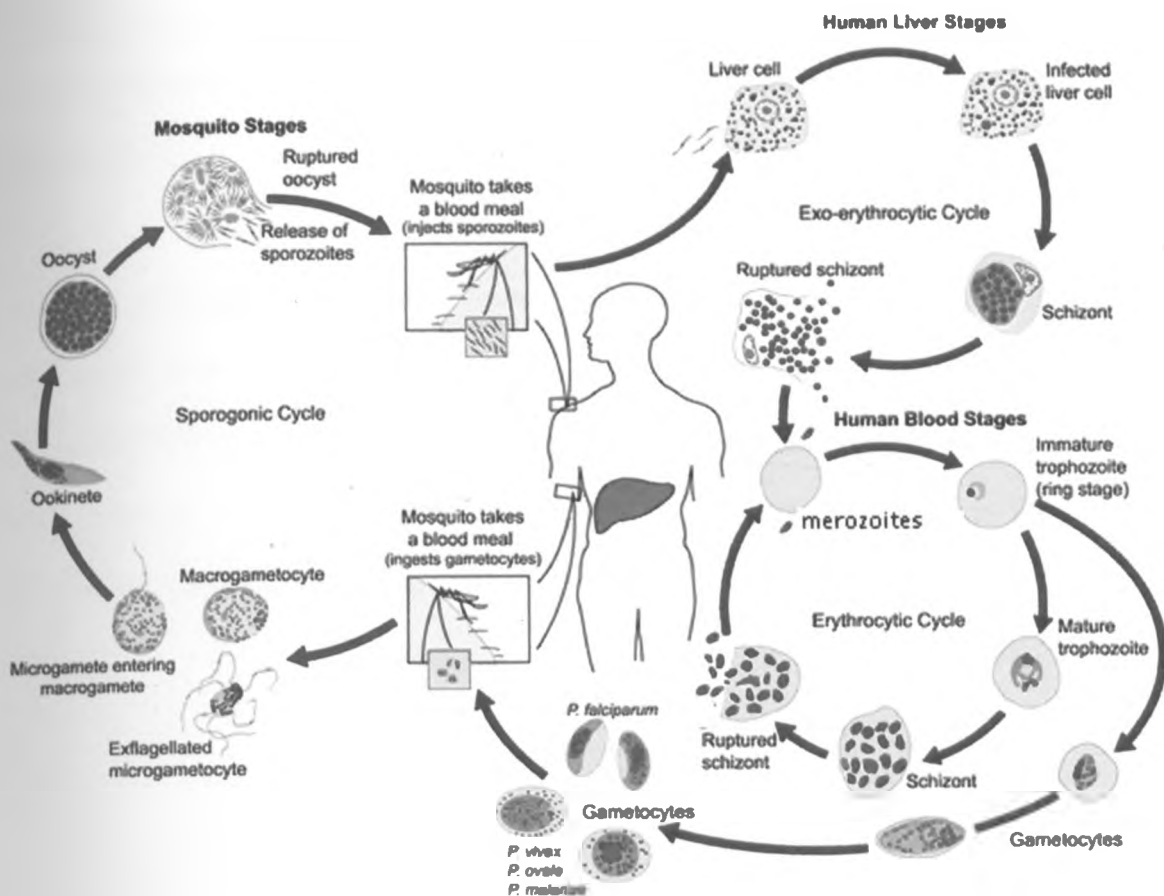
**Figure 3.1:** Life cycle of *Plasmodia* that cause malaria in humans |Source:Centre for Disease Control website: http://www.dpd.cdc.gov/dpdx/HTML/ImageLibrary/M-R/Malaria/ b ody_Malaria_il1.htm (accessed on September 24th, 2010)].
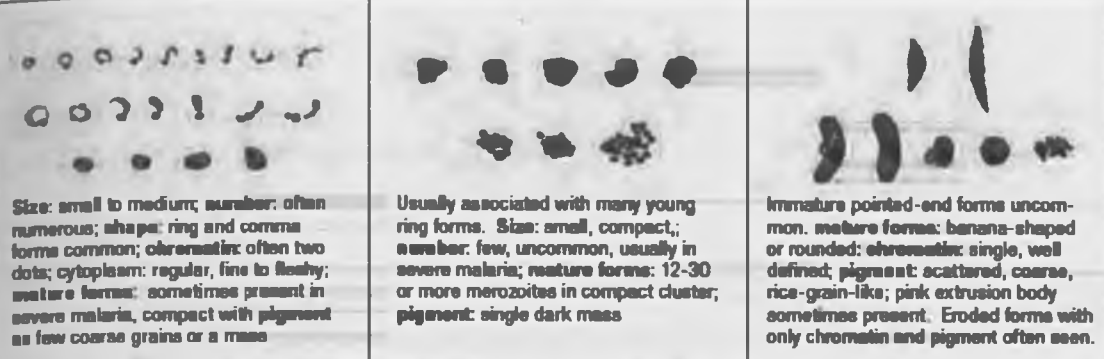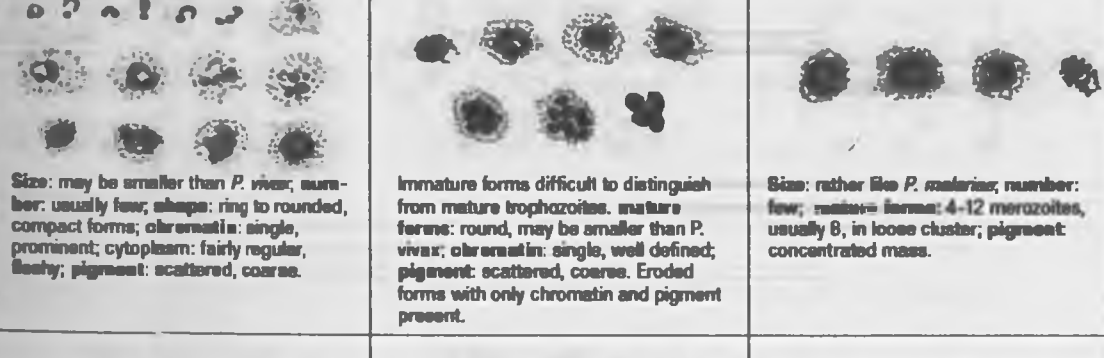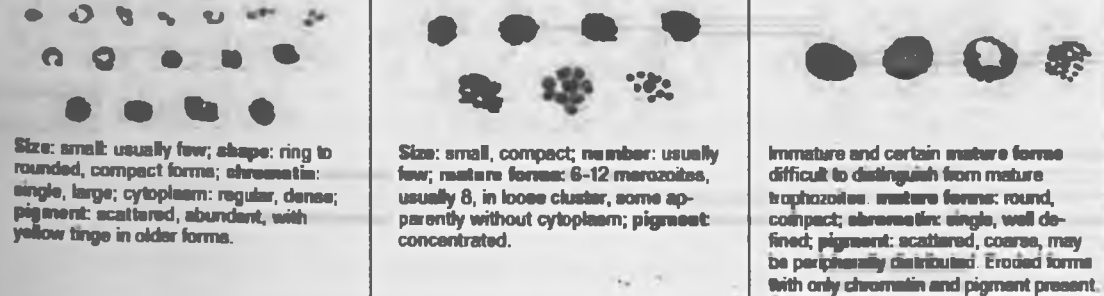
## 3.3 Malaria Diagnostics

When malaria infection is suspected on clinical ground, it is important to obtain laboratory confirmation for the presence of *Plasmodia*, stages and the degree of infection (parasitemia) [1]. Confirmation can be achieved by direct or indirect methods. Direct diagnosis is pursued by looking for the parasites' whole cells or nucleic acid or parasites'

products in blood. On the other hand, indirect diagnostics (also known immunodiagnostics) is conducted by testing a patient's immune response to the infection in the blood.

Many morphological and spectral features of *Plasmodium* can be revealed during examination of a peripheral blood smear. In conventional microscopy for instance, trophozoites, schizonts and gametocytes are identified by their shapes, sizes and colour in stain. Table 3.1 gives a summary of the main diagnostic features used to identify different species of *Plasmodium* in thin blood smears.

**Table 3.1:** Features used in malaria parasites identification in conventional microscopy (Source: [3]).

| Species | | Trophozoite | Schizont | Gametocyte |
|---|---|---|---|---|
| *Plasmodium falciparum* | Young, growing trophozoite and/or mature gametocytes usually seen |  Size: small to medium; number: often numerous; shape: ring and comma forms common; chromatin: often two dots; cytoplasm: regular, fine to fleshy; mature forms: sometimes present in severe malaria, compact with pigment as few coarse grains or a mass |  Usually associated with many young ring forms. Size: small, compact,; number: few, uncommon, usually in severe malaria; mature forms: 12-30 or more merozoites in compact cluster; pigment: single dark mass |  Immature pointed-end forms uncommon. mature forms: banana-shaped or rounded: chromatin: single, well defined; pigment: scattered, coarse, rice-grain-like; pink extrusion body sometimes present. Eroded forms with only chromatin and pigment often seen. |
| *Plasmodium vivax* | All stages seen; Schüffner stipping in 'gowl' of host red cells, especially at film edge |  Size: small to large; number: few to moderate; shape: broken ring to irregular forms common; chromatin: single, occasionally two; cytoplasm: irregular or fragmented; mature forms: compact, dense; pigment: scattered, fine |  Size: large; number: few to moderate, mature forms: 12-24 merozoites, usually 16, in irregular cluster; pigment: loose mass |  Immature forms difficult to distinguish from mature trophozoites. mature forms: round, large; chromatin: single, well defined; pigment: scattered, fine. Eroded forms with scanty or no cytoplasm and only chromatin and pigment present. |
| *Plasmodium ovale* | All stages seen; prominent Schüffner stipping in 'gowl' of host red cells, especially at film edge |  Size: may be smaller than *P. vivax*; number: usually few; shape: ring to rounded, compact forms; chromatin: single, prominent; cytoplasm: fairly regular, fleshy; pigment: scattered, coarse. |  Immature forms difficult to distinguish from mature trophozoites. mature forms: round, may be smaller than P. vivax; chromatin: single, well defined; pigment: scattered, coarse. Eroded forms with only chromatin and pigment present. |  Size: rather like *P. malariae*; number: few; mature forms: 4-12 merozoites, usually 8, in loose cluster; pigment: concentrated mass. |
| *Plasmodium malariae* | All stages seen |  Size: small; usually few; shape: ring to rounded, compact forms; chromatin: single, large; cytoplasm: regular, dense; pigment: scattered, abundant, with yellow tinge in older forms. |  Size: small, compact; number: usually few; mature forms: 6-12 merozoites, usually 8, in loose cluster, some apparently without cytoplasm; pigment: concentrated. |  Immature and certain mature forms difficult to distinguish from mature trophozoites. mature forms: round, compact; chromatin: single, well defined; pigment: scattered, coarse, may be peripherally distributed. Eroded forms with only chromatin and pigment present. |

This thesis is focused on direct diagnostics of malaria by employing unique microscopy techniques based on spectral properties of blood and the chemical changes induced by *Plasmodium* in the red blood cells during an infection.

## 3.4 Effect of *Plasmodium* Parasitization on Erythrocytes

Red blood cells make up 99 % of cellular component of human blood and about 43 % of whole blood [37]. They are highly specialized cells for facilitation of gas transfer in the body. Some of their adaptations to this function include: lack of cell organelles and nuclei, high concentration of haemoglobin and a biconcave disk shape for high surface-to-volume ratio and ease of movement within small blood vessels [40]. In its intra-erythrocytic stage, the malaria parasite alters biophysical properties of the red blood cells. For instance, *Plasmodium falciparum*, the most lethal of the four species, digests a considerable amount of haemoglobin for its nutrition and produces a huge amount of waste products known as hemozoin [41-43], and creates nano-scale protrusions on the red blood cell membrane leading to reduced deformability of the host cell and increased cytoadherence to vascular endothelium and to other red blood cells [43-45]. The result is sequestration of the red blood cells in microvasculature in later stages of parasite development which is linked to vital organ dysfunction in severe malaria [46].

The various components within infected red blood cells (the parasite whole cell with its myriad organelles, haemoglobin and hemozoin) interact differently with electromagnetic radiations of different energies. This is important for spectroscopic analysis of malaria.

24

In the optical region of the electromagnetic spectrum, for instance, haemoglobin (the dominant protein in red blood cells) is known to have a Soret absorption band centred at around 416 nm and doublet absorption bands centred at 540 nm and 575 nm [27, 28]. Intra-erythrocytic parasite growth is associated with a corresponding decrease in haemoglobin but an increase in hemozoin that accumulate in the parasite's digestive vacuole [39, 41]. Hemozoin is known to have Raman spectra that are identical to commercial β-hematin and can be differentiated from haemoglobin using 754 cm$^{-1}$ band [10]. UV-visible absorption spectra show that hemozoin has a broad absorption band centered at around 650 nm [28, 47- 48] in addition to the Soret band close to that of haemoglobin. Both haemoglobin and hemozoin have a porphyrin containing heme prosthetic group but differ slightly in their spectra due to polymerization in hemozoin which makes it more stable than haemoglobin.

The infected red blood cells exhibit varying index of refraction due to inhomogeneity brought about by the invading parasites as opposed to healthy ones which have homogeneous distribution of haemoglobin and hence uniform refractive index. The refractive indices of these stages have been found to be 1.399, 1.395, 1.383 and 1.373 for healthy red blood cells, ring, trophozoite and schizont stage respectively [49].

## 3.5 Image Formation in an Optical Microscope

An optical microscope is an instrument that uses visible light to produce a magnified image of an object (or specimen) that is projected onto the retina of the eye or onto an imaging device. In its simplest form, image formation in a light microscope may be

25

represented as shown in Figure 3.2. The objective lens collects light diffracted by the specimen and forms a magnified real image at the real intermediate image plane near the eyepiece. The eye or the imaging device detects a magnified virtual image of the intermediate real image. In practice, the objective and the ocular (or the eyepiece) contains multiple lens elements and the position of the specimen is dictated by the overall focal length of the objective [49].
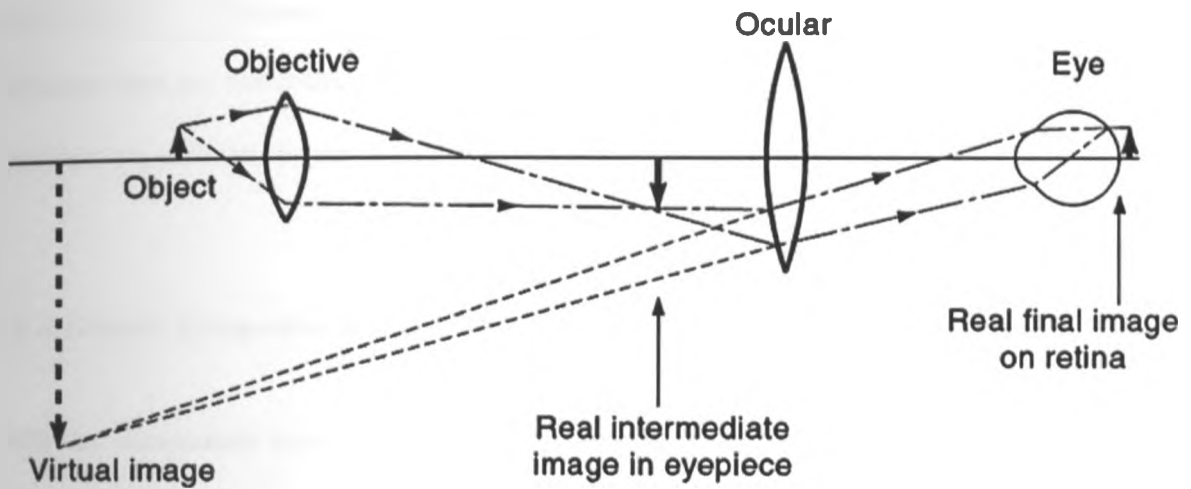


**Figure 3.2**: Perception of a magnified virtual image of a specimen in a transmission microscope (Source: [50]).

In conventional microscopy, the specimen is illuminated by transmission using a broad band source of light that covers part of or the entire visible region. This is called bright field. Other modes of illumination are also possible. For instance, by including a reflective objective, images can be generated by use of reflected light. The microscope's objective collects both diffracted and non-diffracted rays from the specimen. According to Abbe's theory of image formation in a microscope, interference between $0^{th}$ and higher order diffracted rays in the image plane generates contrast [50]. For unstained transparent

26

specimens, however, the component of non-diffracted background light is very large, resulting in bright, low contrast images in which details are poorly visible. Such objects can be viewed in dark-field microscopy in which the non-diffracted rays are removed altogether so that the image is composed solely of diffracted wave components. This technique is sensitive because images based on small amounts of diffracted light from minute objects are seen against a black or very dark background. Dark-field conditions are obtained by illuminating the specimen at an oblique angle such that direct, non-diffracted rays are not collected by the objective. As such only diffracted rays from the specimen are collected by the objective to be used in image formation [50, 51].

## 3.6 A General Perspective of UV- Visible-NIR Spectroscopy

Different compounds have different spectral response to excitation by electromagnetic radiation of different wavelengths. The wavelength of light that a compound absorbs is characteristic of its chemical structure. Specific regions of the electromagnetic spectrum are absorbed when they excite specific types of molecular and atomic motion of matching energy to higher energy levels. Absorption of visible and ultraviolet (UV) radiation is associated with excitation of lower electronic states to higher energy states in both atoms and molecules, while visible and near infrared (NIR) is associated with vibrational molecular processes. Absorption spectra are therefore characteristic of molecular structure and may be used to qualitatively identify atomic and molecular species. Since they uniquely identify compounds, the characteristic spectra can act as fingerprints (or

27

spectral signatures) of the compounds. The absorbance of a transparent solvent is governed by Beer Lambert Law:

$$A = -\log_{10}\left(\frac{I_t}{I_o}\right)$$

(3.1)

where $I_o$ is the incident intensity, $I_t$ the transmitted intensity and A the absorbance. Hence by measuring the intensity of a monochromatic beam before and after its interaction with a sample, absorbance of such a sample can be calculated.

## 3.7 Multispectral Imaging

When a specimen is imaged in an optical microscope, contrast is created by the colour of the imaged scene, which makes features in the specimen to become visible. RGB imagers mimic the human eye by detecting three spectral bands: red, green and blue which combine in various proportions to give a colour image. However, contrast is not an inherent property of the specimen; rather, it results from the interaction of the specimen with light and the sensitivity of the detector. Light intensity recorded by a camera after interacting with the specimen results from such phenomena as absorption, reflection and diffraction of light. Contrast therefore arises from the differences in these optical phenomena for the different parts of the specimen as a function of wavelength. In a transmission set up, spatial gradients in the index of refraction also contribute to contrast in the specimen [52].

Spectral imaging involves measurement of the wavelength spectrum (transmitted, reflected, emitted, etc) at every pixel of an image. Multispectral images are composed of

28

multiple bands (see Figure 3.3) in which each pixel represents the spectral characteristics at a specific position in the image. The resulting spectrum acts like a fingerprint which can be used to characterize the composition of that particular pixel [15-22] as explained in section 3.6.



(a)                                                              (b)
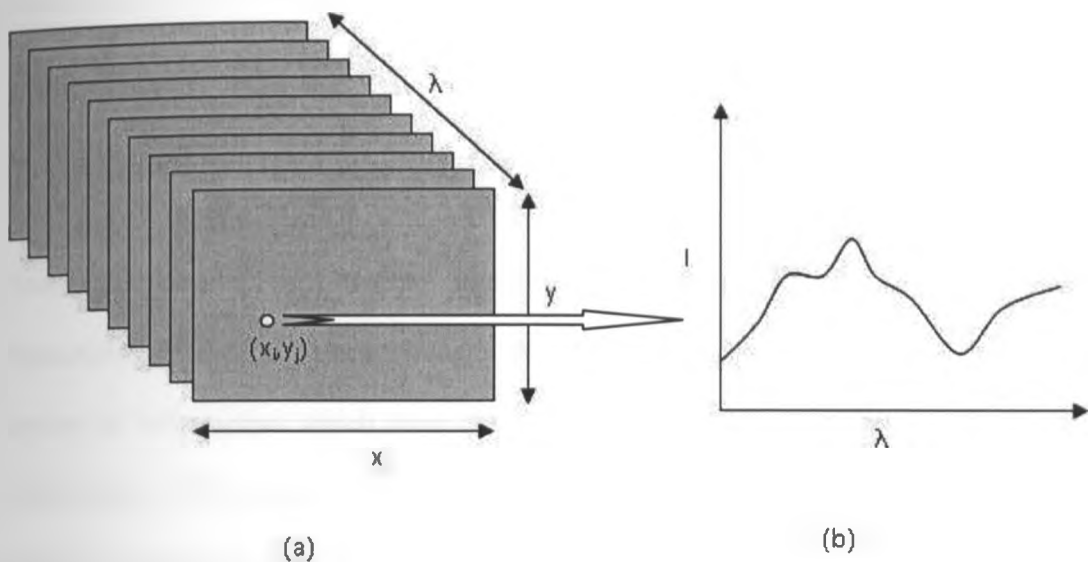
Figure 3.3: (a) Representation of a multispectral image and (b) an arbitrary spectrum of a pixel $(x_i, y_j)$ from a multispectral image.

When the illumination conditions and the spectral characteristics of a camera are known, physical variables such as the reflectivity or transmittance of an object being observed in specified illumination conditions can be determined [52]. Suppose that the spectral sensitivity of a camera in the k-th band is $S_k(\lambda)$, the spectral intensity of the illumination is $L(\lambda)$, and the spectral reflectivity or transmission of an object is $z(\lambda)$, then the pixel intensity value $x_k$ in the k-th band is represented as

$$x_k = \int S_k(\lambda)L(\lambda)z(\lambda)d\lambda \qquad\qquad (3.2)$$

where $\lambda$ is the wavelength [15]. In a three band colour representation such as RGB, the sensitivity bands are broad and overlapping and emphasizes the colour composition ratio rather than the physical variables [15]. In multispectral imaging, sensitivity bands are significantly narrow; hence more information about the physical variable is obtained in addition to pseudo-colour composition from the various spectral bands.

## 3.8 Multivariate Image Analysis (MIA)

Pixel classification can enable identification of regions with similar spectral characteristics in a multispectral image. Within a multispectral image, there are different sources of information which have characteristic spectra that can be used to map the distribution of different compounds in the image. These characteristic spectra are mixed together, meaning that useful information can only be extracted by image analysis such as by the use of multivariate statistical analytical tools. In order to apply two-way multivariate techniques such as Principal Component Analysis (PCA), the three-way multivariate images (consisting of one spectral and two spatial dimensions) are unfolded to two-way object-variable data matrix in which the columns represent the variables and rows the objects (pixels). Analysis is done and the results interpreted in the usual way. In addition, visualization as images is done by folding the transformed data matrix back into the original dimension of the multivariate image [53].

Multivariate data analysis may be divided into pattern recognition and multivariate regression techniques [54]. Image analysis is basically a pattern recognition problem.

30

Pattern recognition techniques are further divided into supervised and unsupervised learning procedures [54]. In supervised pattern recognition, a priori knowledge about classes contained within a sample is required. In contrast, unsupervised pattern recognition does not require a priori knowledge about classes in the training set samples. As such, unsupervised pattern recognition techniques are exploratory methods for data analysis, which seek inherent similarities of data, then group the data in a "natural" way. Supervised pattern recognition, on the other hand, groups data into predefined classes thereby allowing a more precise classification within the class boundaries. In this thesis, Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) have been used to explore the spectral response of pixels from different regions of multispectral images whereas Artificial Neural Network (ANN) has been used for classification. PCA and HCA are unsupervised techniques while ANN is a supervised pattern recognition technique.

### 3.8.1 Principal Component Analysis (PCA)

PCA evaluates total variances within a data set and reduces its dimensionality via eigen value analysis. It searches for correlating features and converts them to a few uncorrelated linear combinations of the original variables that capture most of the information in the original variables of the dataset. To find the latent (hidden) structure in a data matrix $X$ (made up of n objects and q variables), PCA decomposes $X$ into an outer product of score vector $t_a$ and loading vectors $p_a^T$ plus a matrix of residuals, $E$ (the noise part) as expressed in equation 3.3.

$$X = \sum_{a=1}^{A} t_a \, p_a^T + E = TP^T + E \qquad (3.3)$$

where A is equal to or less than the rank of X. The rank of a matrix is the maximum number of independent rows or columns; therefore A must be the smallest value for which equation 3.3 still holds [38]. T and P are known as score and loading matrices respectively. $P^T$ is the transpose matrix of P.

For multispectral images, equation 3.3 is applicable after unfolding has been done as shown in Figure 3.4. The three-way Object-Object-Variable (OOV) data image $\underline{X}$ (pixels as objects and wavelengths as variables) is unfolded to a two-way Object-Variable (OV) data matrix (X) which is decomposed into score (T), loading (P) and E (residual) matrices [55]. The score vectors have same dimension as the unfolded original gray-level images of the multispectral image and thus can be reshaped (folded back to the original dimension of the multispectral image) to make gray-level images. To do this, the score vectors are rescaled to specified gray-level resolution (for example 0-255 for 8-bit images) to avoid having negative values and positive values that are beyond the required resolution of the images. These values may arise in PCA decomposition because all the calculations are based on floating-point numbers. Images have to be represented as short integers for example in the range 0 to 255 for 8-bit images in order to be displayed on a computer screen. The first few score images usually account for most variation in the multispectral image. In addition, score and loading vectors can be plotted against each

32

other as scatter plots in order to visualize important analytical properties such as outliers, gradients and clusters [51].



$$\underline{X} \rightarrow X = T * P^T + E$$
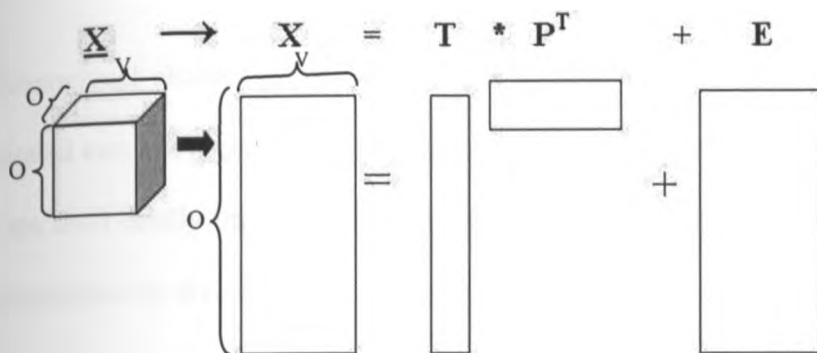
**Figure 3.4:** Decomposition of a 3-way data matrix in PCA.

PCA is thus a powerful data-reduction technique that can condense original data with many variables to a dataset with only a few variables reflecting the most relevant analytical information. For spectral data, it assists in resolving overlapping spectral features by evaluating variance across the selected spectral range. In this work, PCA was used to separate pixel clusters for malaria parasite identification and to single out the most important variables (wavelengths) for malaria diagnostics in the 375-940 nm wavelength range.

### 3.8.2 Hierarchical Cluster Analysis (HCA)

Cluster analysis, in general, uses statistical methods to identify groups that behave similarly or show similar characteristics in a dataset [56]. There are many clustering algorithms such as K-means cluster analysis, Hierarchical Cluster Analysis (HCA) and C-means clustering techniques. In this work, Hierarchical Cluster Analysis (HCA) has been

33

used because of its popularity in production of dendrograms which can provide a two-dimensional pictorial representation of the clustering process [57].

In hierarchical clustering, the original data composed of $m$ objects and $n$ variables are separated into a few general classes, each of which is further subdivided into still smaller groups until finally the individual objects themselves remain. Such algorithm may be agglomerative or divisive [56].

In agglomerative clustering, small groups, starting with individual samples, are fused to produce larger groups. In contrast, divisive clustering starts with a single cluster, containing all objects which are successively divided into smaller partitions. Agglomerative hierarchical clustering is the commonly implemented algorithm in most application software. It begins with the computation of similarity or distance matrix between the objects, finds the smallest elements in the distance matrix and joins the corresponding objects into a single cluster. This process is repeated until all the objects are fused into one large set. The most popular measure of dissimilarity between objects $x_i$ and $x_j$ is the Euclidean distance, which is a special case of the Minkowski distance with $p=2$, generally expressed as

$$d_{minkowski}(x_i, x_j) = \left( \sum_{l=1}^{m} |x_{il} - x_{jl}|^p \right)^{\frac{1}{p}} \qquad (3.4)$$

where $x_i=\{x_{il}, ...., x_{id}\}$ and $x_j=\{x_{jl}, ...., x_{jd}\}$. The Minkowski distance with $p=1$ is called Manhattan distance. Other distances can also be defined, such as Chebyshev, Mahalanobis and Bahattacharyya distances [57]. The algorithm employed in this work

uses Euclidean distance and has been used to show spatial distribution of heme species in an infected red blood cell.

### 3.8.3 Artificial Neural Network (ANN)

An ANN consists of a large number of processing units known as neurons connected together, that attempt to simulate the biological structure of the brain and nervous system. A biological neuron has components that are of particular interest in understanding an artificial neuron: the dendrites which receive electrical signals from other neurons into the cell body of the neuron, the soma (or the cell body) which sums and thresholds the incoming signals and fires depending on the threshold value, and the axon which transmits the resulting signal from soma to other neurons [58] (see Figure 3.5). A point of contact between an axon of one neuron and a dendrite of another neuron is known as a synapse. The strengths of the synapses change in an organism's life. Strengthening or weakening of synaptic junctions is what is known as 'a learning process' of the organism [59].
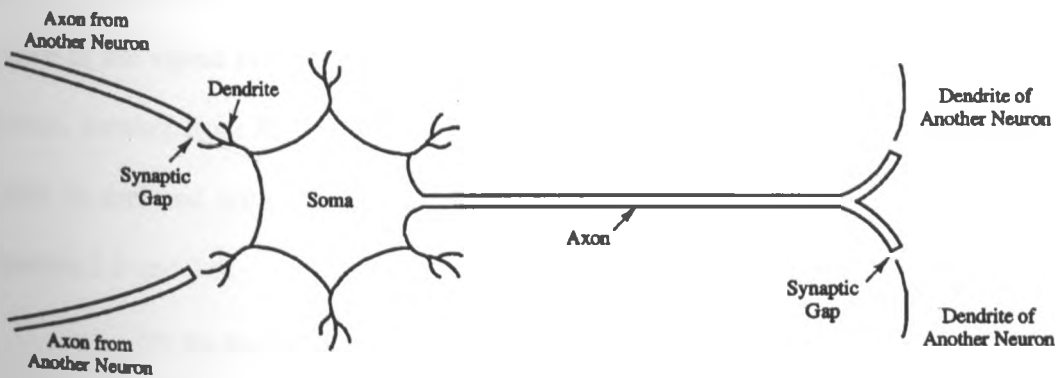


Figure 3.5: A schematic representation of a biological neuron (Source: [58]).

Components of an artificial neuron model relate to those of a biological neuron as follows: the dendrites, synapses, soma and axon correspond to inputs, weights, transfer function and output of the neuron respectively. An illustration of an artificial neuron is shown in Figure 3.6.
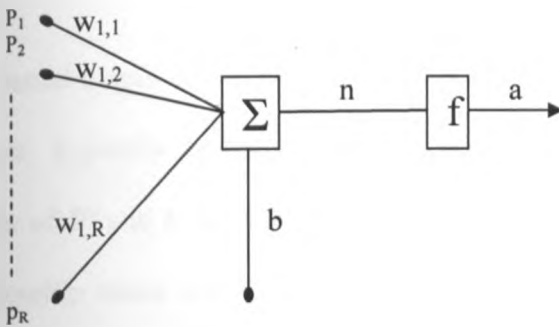


**Figure 3.6:** A schematic representation of an artificial neuron.

The individual inputs $p_1$, $p_2$, ..., $p_R$ form the input vector, and are each weighted by the corresponding elements $w_{1,1}$, $w_{1,2}$, ...,$w_{1,R}$ of a weight matrix **W**. The first index on the elements of the weight matrix indicate the particular neuron destination for that weight (in the case where there are more than one neuron), and the second index indicates the source of the signal fed to the neuron, for example, $w_{1,2}$ means a connection to the first neuron, coming from the second element of the input vector. The neuron has a bias $b$ which is summed with the weighted inputs to form the net input $n$, as expressed in equation 3.5 and 3.6.

$$n = w_{1,1}p_1 + w_{1,2}p_2 + ... + w_{1,R}p_R + b \tag{3.5}$$

or

$$n = Wp + b \tag{3.6}$$

where $W$ is the weight matrix, $p$ is the input vector and $b$ the bias. The summer output $n$ (referred to as net input) is evaluated in a transfer function, $f$ (also known as the activation function) to produce a scalar output $a$ of the neuron as expressed in equation 3.7

$$a=f(Wp+b) \tag{3.7}$$

The actual output of the neuron depends on the particular transfer function used in the neuron. Typically the transfer function is chosen by the neural network designer and the values of $W$ and $b$ are adjusted by some learning rule so that the neuron input/output relationship meets some specific goal. The neural network "learns" by modifying the weights of the neurons in response to the errors between the actual output values and the target output values [60].

A layer of neurons, such as the one just described, is known as a perceptron. A perceptron is limited by the fact that it can only be used to solve a classification problem in which there exist a linear boundary. In this work, though the main problem is to classify a red blood cell as either infected or uninfected, there were no identifiable linear boundaries between the spectra of the two. As such a multilayer perceptron neural network with a Levenberg-Marquardt back propagation learning algorithm which can be trained to solve non-linear problems [61] was chosen. Since output of the network is expected to classify red blood cells in two states, positive or negative, a symmetrical function in the hidden layer and a linear function with a threshold value in the output

layer were selected. These are: hyperbolic tangent function defined by equation 3.8 whose outputs varies between -1 and +1 and a linear transfer function whose input is equal to the output for the output layer with a threshold value set to zero.

$$a = \frac{e^n - e^{-n}}{e^n + e^{-n}} \qquad\qquad (3.8)$$

where $a$ is function output and $n$ is the summer output.

Training of such a network is carried out by repeatedly presenting the training patterns (updating the weights) until the average error over all the training patterns is minimal and within tolerance specified for the problem. This is carried out through gradient descent approach in which changes in weights are proportional to the negative of the derivative of the sum of the squares of the errors for all the training patterns. One pass through the set of training pattern, together with associated updating of the weights, is called a cycle or an epoch. After training, the neural network should be able to reproduce the target output values for the training data with minimal errors. The associated trained weights of the neurons are then stored in the neural network memory and used in prediction of target output values when different sets of data are fed to the ANN [59].

The performance of the neural network model can be assessed by several criteria; these include the coefficient of correlation R, root mean squared error and mean absolute error among others. A well trained model should result in an R value close to one and small values of error terms [62].

# Chapter 4

## MATERIALS AND METHODS

### 4.1 The Multispectral Imaging Microscope

The imaging system used in this study was a multimodal, multispectral imaging microscope developed by Mikkel *et al.* [52] and advanced by Aboma [63]. The equipment consists of a commercial optical microscope (Brunel compound microscope) modified by replacing the conventional illumination light (white light) with monochromatic LED lighting system. Thirteen LEDs emissions centred at 375 nm, 400 nm, 435 nm, 470 nm, 525 nm, 590 nm, 625 nm, 660 nm, 700 nm, 750 nm, 810 nm, 850 nm and 940 nm are used. This wavelength range covers UV, visible and NIR optical region. Figure 4.1 shows the emission spectra of the LEDs as measured using USB2000 Ocean Optics® spectrometer.
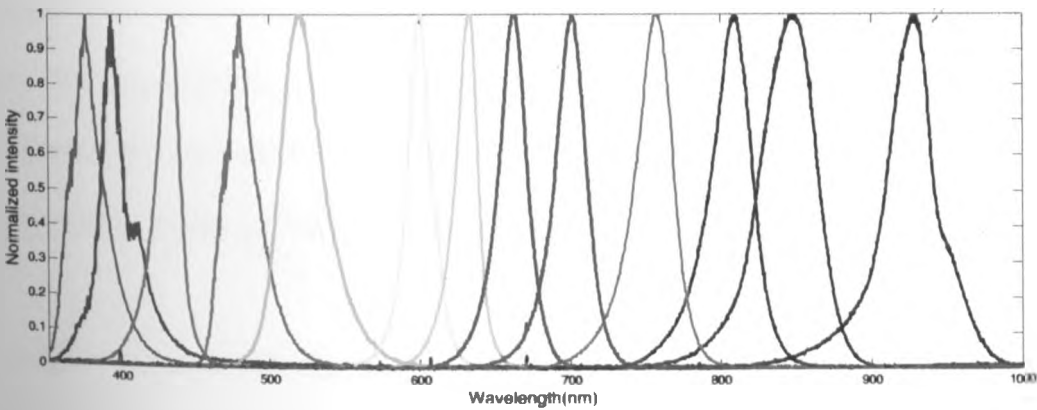


**Figure 4.1:** Emission spectra of LEDs used for illumination in the microscope.

It may be noted that all the LEDs show an appreciable chromaticity though there is little overlap of some of the spectra of neighbouring LEDs. The LEDs are mounted in a quasi hemispherical container made of teflon and illuminate same spot on a 5 mm Opal light diffuser (Edmund Optics) which in turn provides a Lambertian-like illumination to the sample. A Lambertian source of light is one which has same apparent radiance when viewed at any angle, hence provides uniform illumination to the sample.

To achieve multispectral imaging, the LEDs are activated sequentially through a computer controlled data acquisition card (DAQ from National Instruments Inc.). The sample is illuminated in three different modes: bright field (transmission), dark field (scattering) and reflection. For transmission measurement, the sample is illuminated by the set of LEDs located directly below the sample. In reflection mode, light reaches the sample from above via a system of optics, which includes a cassegrain objective. The cassegrain objective (X15 Reflx$^{TM}$® objective from Edmund Optics) is a reflective objective that is essential for reduction of chromatic aberration in all the three modes of imaging. Dark field imaging is achieved through the use of a ring coupled on an optical fibre. The ring provides an oblique illumination that cannot be accepted by objective's aperture, hence only light which is scattered by the specimen can enter the objective. An image formed in this way has bright objects superimposed upon a dark background.

A computer controlled 12-bit (resolution 640X480 - 1600X1200) monochrome CMOS camera (Guppy GF503B, Allied Vision Technologies) is fitted on the microscope ocular to capture an image of the sample at each LED illumination. Figures 4.2 and 4.3 show

40
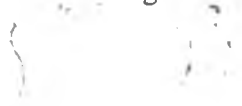
the spectral sensitivity of the camera and the complete multispectral imaging set-up respectively. It can be noted that the spectral sensitivity of the camera matches the emission range of the LEDs shown in Figure 4.1.
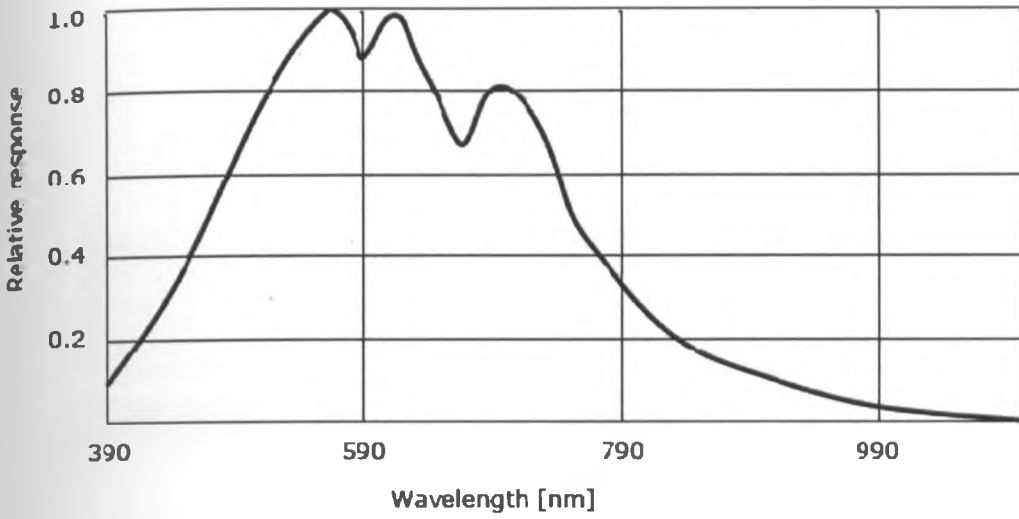


**Figure 4.2:** Spectral response of Guppy GF503B camera (Source: [64]).
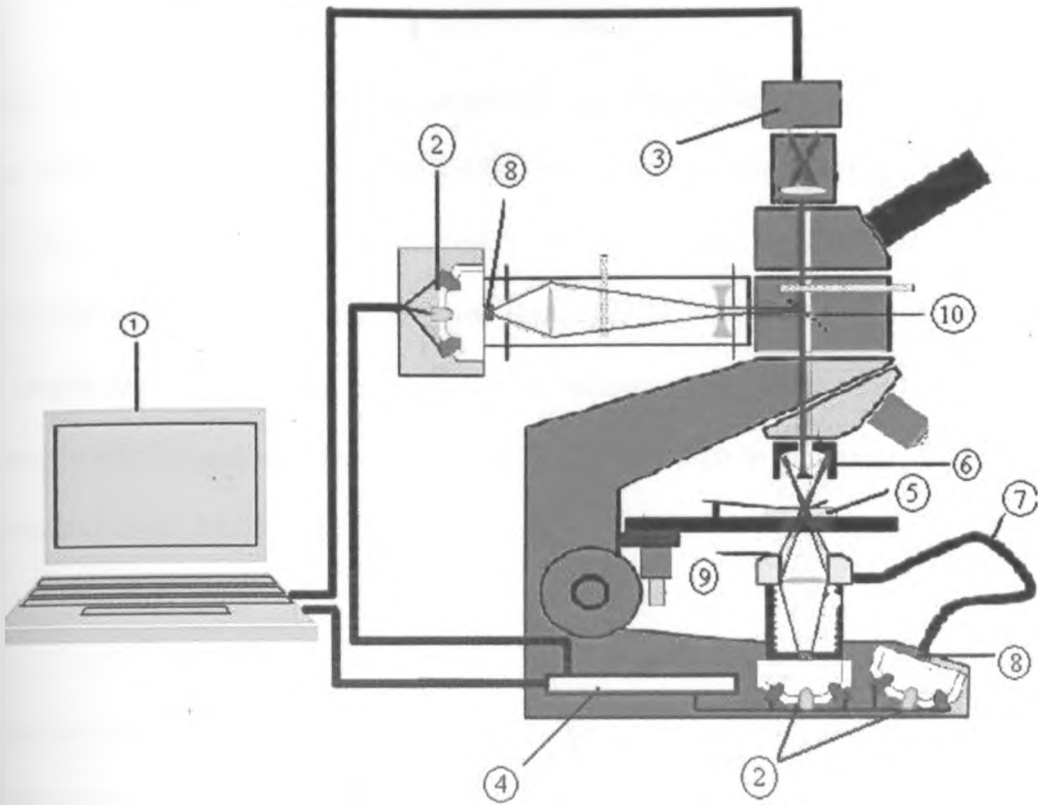
**Figure 4.3:** The Multispectral Imaging Microscope (Adapted from [63]).

<u>Key</u>

1. Computer
2. LED set
3. CMOS camera
4. DAQ card and circuitry
5. Sample

6. Cassegrain objective
7. Optical fibre
8. Diffuser
9. Ring coupled on optical fibre
10. Beam splitter

## 4.2 Measurement Procedure

*In vitro* cultures of malaria-infected unstained thin blood smears were obtained from Kenya Medical Research Institute (KEMRI) and Institut National d'Hygiène Publique (Côte d'Ivoire). The samples had already been positively identified by an expert microscopist as positive and containing all the stages of *Plasmodium falciparum*. For each sample, three measurements (transmission, reflectance, and scattering) were made. The sample was focused such that the camera 'sees' the objects in the sample from all the three modes at same image plane.

The spectral sensitivity of the CMOS camera (as show shown in Figure 4.2) varies with the wavelength of the light to be detected [64]. For instance, light from a 590 nm LED will easily saturate the camera as opposed to one from a 940 nm LED when both LEDs are driven by equal proportion of current as per their current ratings. As such, a protocol (camera and current settings for each LED) was created, at every measurement, in the software (developed in LabView-National Instruments Inc.) by subjectively adjusting the driving currents of the LEDs, the exposure time and the gain of the camera to obtain a good dynamic range without saturating the camera. A single measurement consisted of taking three images; sample image, $I(\lambda)_s$, dark image, $I(\lambda)_d$, and reference image, $I(\lambda)_r$ using same protocol. The three images were then used in calibration (see section 4.3).

The process of acquiring reference image, $I(\lambda)_r$ and dark image, $I(\lambda)_d$ varied slightly depending on the mode of imaging in use. For transmission measurements, $I(\lambda)_r$ was an image of an empty glass slide with identical optical properties as those of the slide holding the sample. In reflection and scattering modes, $I(\lambda)_r$ was the image of reflecting and scattering sides respectively of an Opal diffuser (PO:006208-00 SM-Edmund Optics) known to have an average reflectance value of 50 %. The dark current image $I(\lambda)_d$ in transmission mode was taken by using the same transmission protocol as that of the sample but with the power supply to the LEDs switched off. In reflection, $I(\lambda)_d$ was obtained using the same protocol as that of the reflection sample image with nothing on the light path (sample and slide removed) while in dark field mode $I(\lambda)_d$ was obtained with an empty glass slide using scattering sample image protocol.

## 4.3 Calibration

To account for the differences in the spectral response of the camera to the different LED light sources, the 'dark' camera response and different noise sources, a corrected spectral image, $I(\lambda)_{spec}$ was obtained from the images captured in measurements using equation 4.1.

$$I(\lambda)_{spec} = \frac{I(\lambda)_s - I(\lambda)_d}{I(\lambda)_r - I(\lambda)_d} \qquad (4.1)$$

where $I(\lambda)_s$ is sample image, $I(\lambda)_r$ is reference image, and $I(\lambda)_d$ is dark image. The subtraction is done on a pixel-by-pixel basis for each wavelength ($\lambda$). Image correction and the subsequent processing and analysis were performed using MATLAB® (MathWorks Inc.) software. A code was written to import the raw images stored in

unsigned 16-bit representation, convert them to double class representation and apply the correction using equation 4.1 in readiness for analysis. A number of selected Matlab codes used in the processing are presented in the Appendix.

For images taken in transmission mode, transmittance values were obtained directly from equation 4.1 since the reference is a plain transparent glass slide. The ratio between the intensity of light transmitted through sample slide to the intensity of light transmitted though an empty slide is the transmittance value. However, the values of reflectance were obtained by multiplying images taken in reflectance mode by a factor Q given by equation (4.2), because the reference used in this case (Opal diffuser) is 50 % reflecting.

$$Q = \frac{1-I(\lambda)_d}{\frac{1}{2}-I(\lambda)_d} \tag{4.2}$$

where $I(\lambda)_d$ is the dark image.

In all imaging modes, the size of the imaged scene was estimated by imaging homogeneous polymer microspheres of standard diameter 10 µm each. From the images, a microsphere with diameter of 10 µm occupies 40 X 40 pixel space. Thus for a 480 X 640 pixel image, the imaged scene is approximately 120 X 160 µm with a resolution of 16 pixels per square micro-metre.

## 4.4 Image Preprocessing

The corrected images resulting from equation 4.1 have both missing and infinite values arising from points where there was division by zeros. To remove these "dead" pixels, all

pixel values in the images were scaled to a maximum value of one, all infinite values set to one and all missing values set to zero. A median filter was then applied on each gray-level image using a kernel of 3 X 3 to remove "salt and pepper" noise.

## 4.5 Spectral Capability Test for the Microscope

To test the capability of the microscope to generate spectral signatures of imaged components, an absorption spectrum of thin blood smear was taken using SolidSpec® 3700 spectrophotometer in the range 375-940 nm. The same sample was also imaged using the LED multispectral imaging microscope working in transmission mode and absorption spectra obtained by using Beer-Lambert's law after normalization. The spectra from the two systems were then compared in order to visualize the main spectral features of blood.

## 4.6 Multivariate Image Analysis

Image analysis consisted of two parts. First, the images were explored for presence of parasites in the red blood cells. To do this, an image from the dark-field mode was used to identify possibly infected red blood cells. In dark-field, an infected red blood cell would contain a bright spot inside an enclosed outline of the red blood cell against a dark background, whereas uninfected one does not contain such spot. The bright spot in an infected cell is the result of diffraction (scattering) effect of the parasite. However, since the morphological features of the parasite (usually known and used by microscopists) were not clear to conclusively identify it as a parasite, exploratory analysis based on the

46

spectral properties of the imaged red blood cells was done. Principal components analysis was done (after mean-centering to minimize undue influence from individual wavelengths) to investigate the spectra of both suspected infected red blood cells and uninfected ones. The reason for this was that in the transformed latent space of PCA, scores and loadings could be used to visualize similarities and differences in the spectra and also highlight the variables (wavelengths) causing the highest variances.

PCA was applied by first inspecting a single red blood cell suspected to be infected by *Plasmodium* using analytical method of image analysis developed by Geladi *et al.* [53, 65-66]. Individual parasitized red blood cells, which are visible in the images, were cropped (using a function developed in Matlab to crop a 13-band image) and then subjected to PCA to explore variance in the spectra. To minimize undue influence from individual variables (wavelengths at which images have been acquired), all spectra were scaled to 0-1 dynamic range and then mean-centered before applying the PCA algorithm in Matlab. The transformed data was then interpreted by selecting a cluster of pixels in the score plot and highlighting them in the image space (a process known as masking) to identify the origin of spectra. Secondly, PCA was applied by labeling each red blood cell as either positive (P) or negative (N) based on the results of dark-field image. The red blood cell was then cropped and its spectrum found by averaging intensity values in each gray-level image for all the 13 bands. The averaged spectra were then subjected to PCA analysis and the scores studied. Correctly classified red blood cells were those which appeared in the expected cluster. Red blood cells that appeared in the unexpected cluster or those which appeared to be outliers were excluded from the second stage of analysis. HCA algorithm

was then applied to the unfolded image spectra using Matlab Statistical Tool box to show the distribution of similar spectral features and hence the distribution of important pigments in an infected red blood cell.

Part two of analysis was development of Neural Network model based on the correctly identified spectra of red blood cells from PCA results. The neural network was developed using Matlab Neural Network tool box (MathWorks Inc.) software. The inputs to the network were raw spectra obtained by averaging pixel values from individual red blood cells. The network was trained to solve a two-class classification problem to identify each red blood cell as either infected (output coded +1 or uninfected (output coded -1).

To avoid overfitting, which makes a neural network to memorize training patterns in such a way that it cannot generalize to new data, spectra were split into three subsets: a training set (60%), a validation set (20%) and a test set (20%). The training set is usually used to update the network weights as the error is monitored whereas the validation set is used to ensure that, when the network generalizes, training stops just before overfitting occurs. A one hidden layer ANN was chosen since it can approximate any continuous function [62]. Network structure was optimized by testing varying number of neurons in the hidden layer for the lowest training and validation dataset errors. The prediction accuracy of the network was then determined from classification performed on the test set. The performance of the network was calculated by dividing the number of correctly classified spectra by the total number test spectra and expressing it as a percentage.

## Chapter 5

## RESULTS AND DISCUSSION

### 5.1 Microscope's Capability in Reproducing Red Blood Cells Spectra

Before any analysis was carried out on the spectral dimension of the images, the microscope was tested on its capability to reproduce measured spectra. Testing was done using a thin blood smear on a standard laboratory slide. It is known that haemoglobin is the major constituent of red blood cells and has a spectral signature in the visible region. Figure 5.1 shows normalized absorption spectrum of blood smear obtained from Shimadzu® DUV 3700 spectrophotometer for comparison to spectra from the microscope.
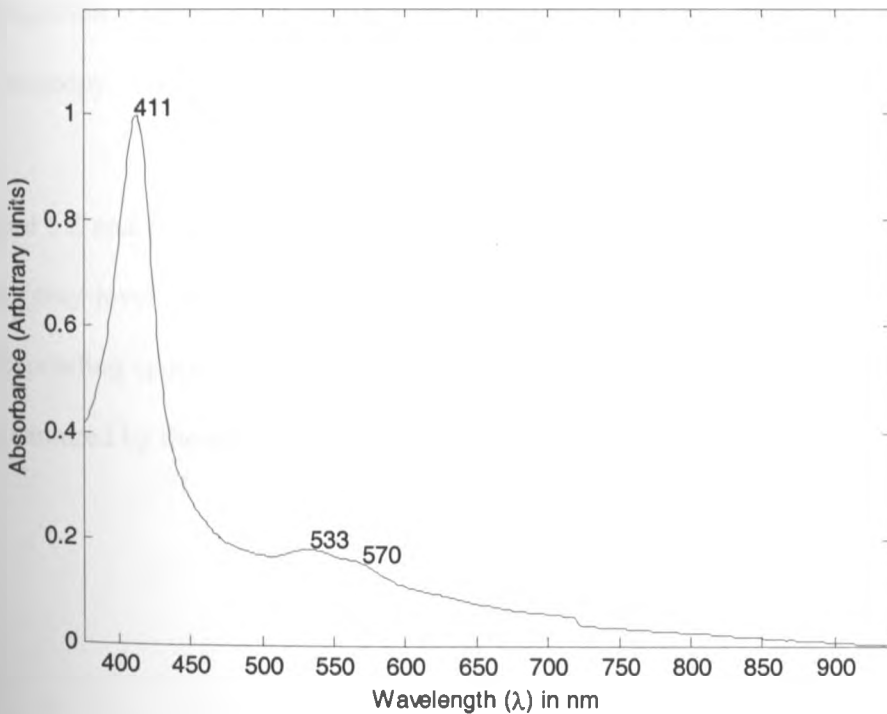


**Figure 5.1:** Absorption spectrum of a thin blood smear taken in the spectral range 375-940 nm using DUV 3700 spectrophotometer.

The spectrum was taken from a macroscopic view without consideration of the contribution of individual red blood cells and normalized using equation 5.1 in order to scale all values to range between zero and one.

$$I_{norm} = \frac{I_{act} - I_{min}}{I_{max} - I_{min}} \qquad (5.1)$$

where $I_{norm}$ is the normalized intensity value, $I_{act}$ is the actual (original) intensity value, $I_{max}$ is the maximum intensity value in the dataset and $I_{min}$ is the minimum intensity value in the dataset.

From Figure 5.1, it can be seen that haemoglobin (the dominant constituent of blood) has a strong absorption at 411 nm (known as Soret band). It also has relatively weaker absorptions at 533 nm and 570 nm (a doublet) which are only observable for oxy-haemoglobin. These bands are important for identifying oxy-haemoglobin in absorption spectroscopy.

Figure 5.2 and Figure 5.3 show, respectively, a representative spatial image of red blood cells (gray-level image captured at 590 nm for the purpose of visualization) and the corresponding spectrum of the whole multispectral image (imaged scene size 120 X 160 µm) captured by the microscope.
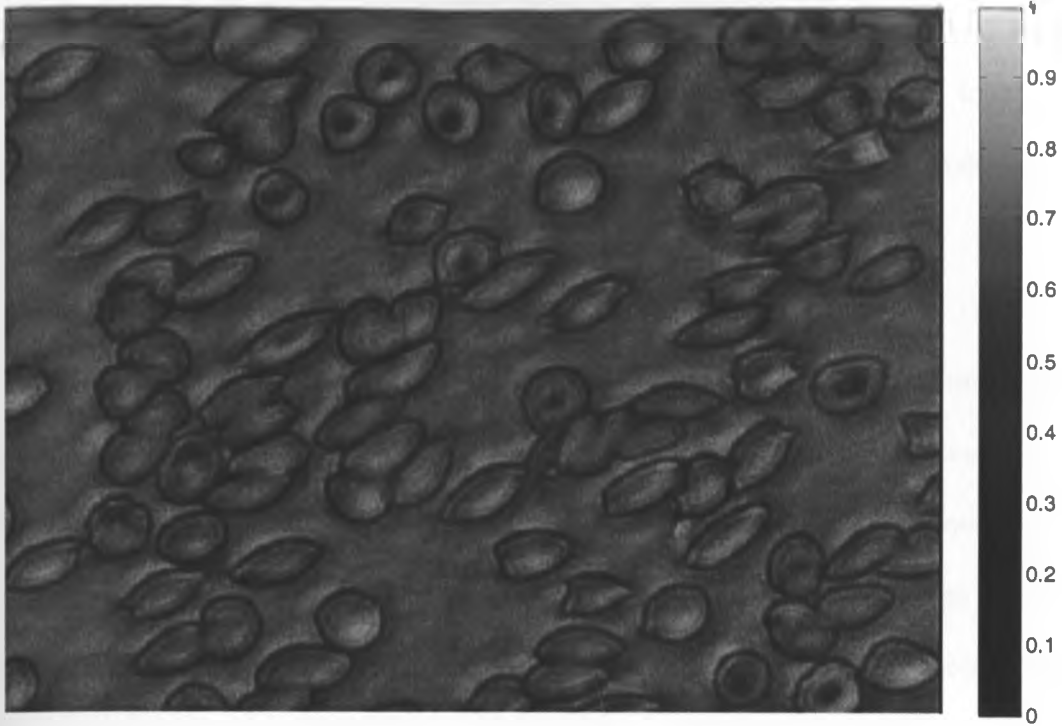
**Figure 5.2:** A microscopic image of thin blood smear taken in transmission mode.
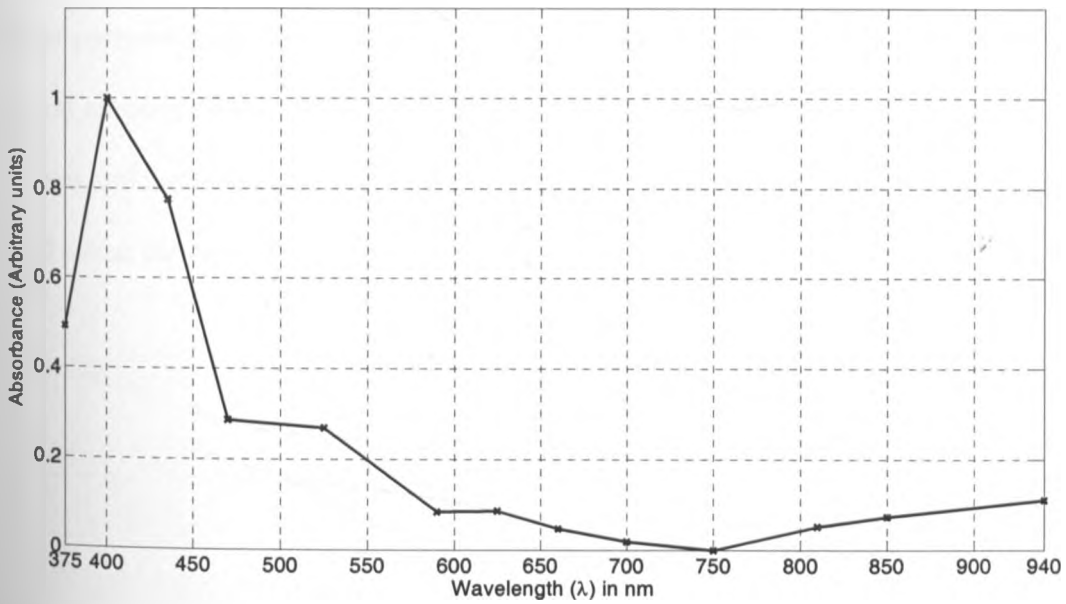


**Figure 5.3:** Absorption spectrum of thin blood smear derived from a multispectral image.

The normalized spectrum from the image was obtained by taking an average pixel value in each transmittance gray-level image for all the channels (corresponding to all the wavelengths of measurement) to obtain transmittance spectrum and then transformed to absorbance spectrum using Beer-Lambert's law.

It is clear from the image (Figure 5.2) that there is spatial variation in the transmittance values (pixel values) between regions occupied by red blood cells and where there are no red blood cells. In fact, there exist variances even within the individual red bloods. If all pixels from a single red blood cell (approximately 900 pixels since the average size of a red blood cell in the image is about 30 X 30 pixels) are plotted as absorption spectra, intra-cell variances are seen (see Figure 5.4). The global spectrum from the multispectral image is an average value of the all pixels in each channel represented as a single intensity value at that particular wavelength and hence cannot reveal intra-cell variances but reveals the general spectral features like those captured by the spectrophotometer. Since each pixel contains unique information in the image, spectral signatures of different components in blood smear can be found by examining the average value of localized pixel spectra.
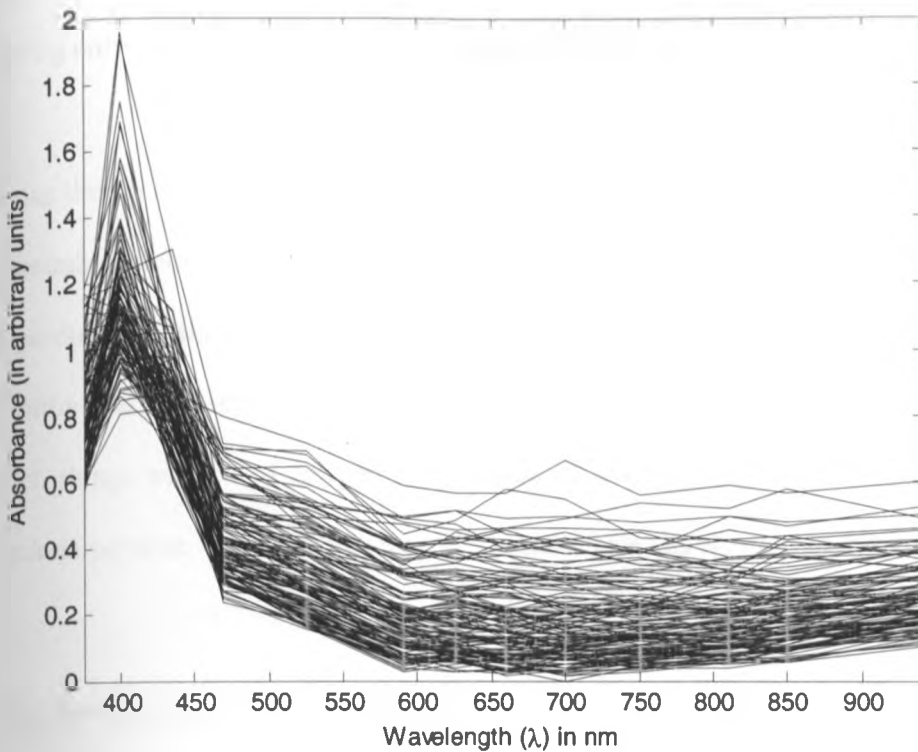
**Figure 5.4:** Pixels of a red blood cell presented as absorption spectra.

When the absorption spectrum obtained by Shimadzu® 3700 DUV spectrophotometer (Figure 5.1) is compared to the global spectrum measured by the multispectral imaging microscope (Figure 5.3), basic features of haemoglobin are observed. The spectrum from the spectrophotometer reveals a Soret absorption band centered at 411 nm and the Q bands (doublet) centered at 533 nm and 570 nm respectively. This is due to the fact that the spectrophotometer has high spectral resolution (0.1 nm full width at half maximum (FWHM)) which enables acquisition of more than 5000 data points in the range 375-940 nm. These absorption bands are a signature that identifies oxy-haemoglobin which exists when red blood cells are exposed to oxygen. On the other hand, the global spectrum measured by the multispectral microscope reveals the Soret band clearly but the Q bands

53

are inseparable due to the fact that images were taken at discrete (thirteen) spectral bands giving only thirteen data points in the range 375-940 nm.

Since the spectrum from the whole multispectral image (Figure 5.3) basically captures the spectral signature of haemoglobin, each red blood cell in the image is expected to reproduce a similar spectrum. However, this can only be true if all the red blood cells are healthy and exposed to oxygen. Haemoglobin in the red blood cells can get altered by such things as lack of oxygen or presence of disease causing micro-organism such as malaria parasite. This spectroscopically is manifested as changes in the spectral signature.

## 5.2 Identification of Infected Red Blood Cells in Dark-Field

Figure 5.5 shows dark field microscopic image of red blood cells infected with *Plasmodium falciparum*. The infected red blood cells are indicated using white arrows. Dark-field mode of imaging was used as a starting point for identifying red blood cells as either infected or not infected because parasites inside the red blood cells cause diffraction (scattering) which make them to be highlighted as tiny bright spots inside the red blood cells. The bright (white) outline of the red blood cells is also the result of scattering of light by the edges of the red blood cells. Since uninfected red blood cells are homogenous only their edges are seen to have scattered light. The whole scenario is observed against a dark background representing regions of little or no scattering.
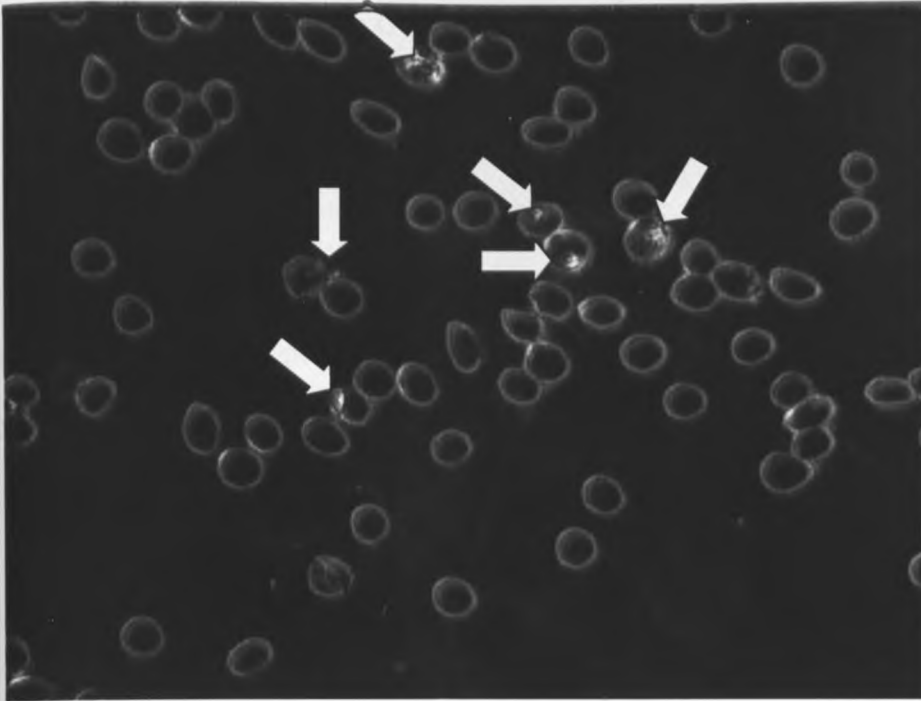
**Figure 5.5:** Suspected *Plasmodium falciparum* in erythrocytes as seen in dark-field microscopic image of *in vitro culture* thin blood smear.

However, identifying a red blood cell based on dark-field microscopic examination was not conclusive owing to the fact that morphological features well known to clinicians (and any other microscopists) could not be resolved in the image. As such it was difficult to separate a true positive from a false positive (such as an artifact e.g. a dust particle) in the thin blood smear. Thus, a red blood cell identified as a positive in dark-field had to be subjected to further scrutiny in order to arrive at a conclusion that it is a true positive. Due to the limitation in the magnifying power of the microscope (X15 objective), it was necessary to study the spectral dimension of images of both infected and non-infected red blood cells as identified in dark-field.

## 5.3 Spectral Signature of the Genus *Plasmodium* in Blood

Figure 5.6 shows normalized absorption spectra of *Plasmodium falciparum*-infected red blood cell and that of uninfected one plotted together. These plots represent a trend after examining spectra of 200 non-infected red blood cells and 50 infected ones. The spectra were obtained by cropping transmittance images of individual red blood cells (using a custom-made Matlab® function that crops a 13-band image) which gave transmittance spectra to be transformed to absorbance spectra as explained in section 5.1.
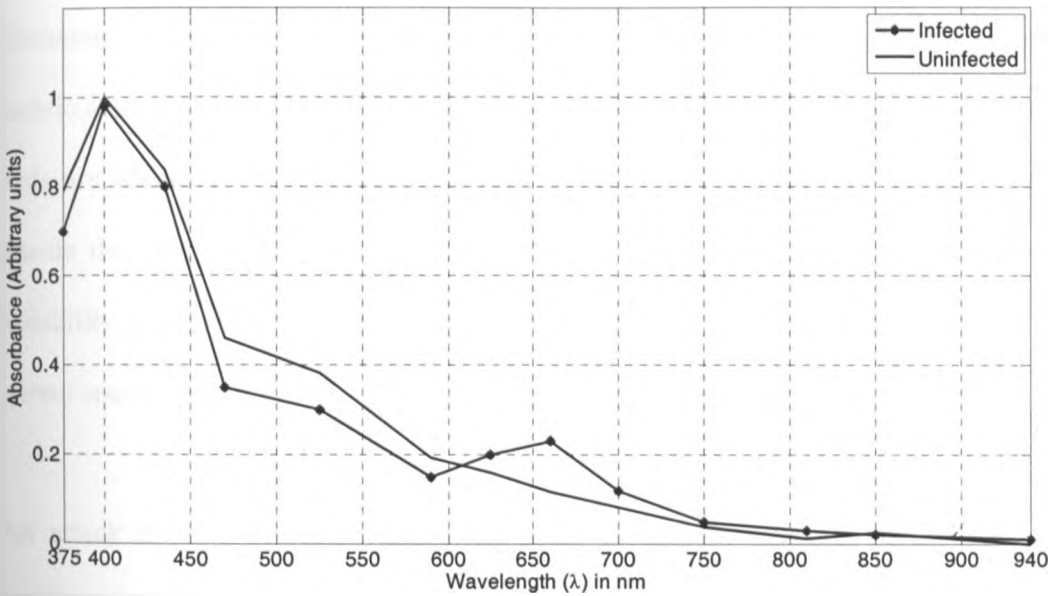


**Figure 5.6:** Absorption spectra of infected and non-infected single red blood cells.

It is important to note that these spectra are not absolute but only illustrate the main spectral features. The absorption spectra of infected and non-infected red blood cells have a common intense absorption band centered near 400 nm which is due to electronic transitions of the porphyrin ring found in heme. However, there are slight differences in the spectra: a healthy red blood cell has slightly higher absorbance in the 470-590 nm

wavelength range whereas an infected one has slightly higher absorbance in the 625-700 nm spectral range.

*Plasmodium*, the malaria parasite, is a complete unicellular organism with cell membrane, nucleus and myriad organelles. On the other hand, a mature red blood cell consists mainly of haemoglobin protein without nucleus. The spectral signature of haemoglobin in the 375-940 nm range was shown in Figure 5.1. The invasion of the red blood cell by *Plasmodium* with its cellular complexity presents some spectral response that requires a biochemical understanding of the precise origin of the spectra. Malaria parasites have nucleic acid material which is generally transparent in the visible region. In fact, nucleic acids are identified by their characteristic intense absorption at 260 nm [67] which is outside the range of the microscope used in this work. This, therefore, rules out any possibility of detecting the parasites' DNA using the multispectral microscope in its current spectral range.

An attack by blood feeding organisms such as malaria parasites is associated with production of a large amount of waste product known as hemozoin (also known as malaria pigment). Hemozoin is a heme crystal produced by *Plasmodia* in order to detoxify free heme released when haemoglobin is digested in the digestive vacuoles of the parasites. Free heme is a potent cytotoxic agent to the parasites because it promotes lysis of the cellular organisms and generates Reactive Oxygen Species (ROS) which damage many biomolecules [68]. Both haemoglobin and hemozoin contain tightly bound metal-ion-containing molecules known as prosthetic groups. The spectral signatures of prosthetic

groups are known to be found in the visible region of the electromagnetic spectrum [67]. Hemozoin is, therefore, of central importance in identifying infected red blood cells spectroscopically using the LED multispectral imaging microscope working in 375-940 nm range. The slightly higher absorbance of infected red blood cell in the spectral range 625- 700 nm but lower absorbance in the 470-590 can therefore be attributed to reduction of haemoglobin but increase in hemozoin in an infected red blood cell in the trophic stage of the parasite.

Reflectance curves for both parasite-infected and non-infected red blood cells were also generated from reflectance images as was done in section 5.2 for transmittance images and are shown in Figure 5.7. However the differences between infected and uninfected red blood cell spectra were difficult to observe in this imaging mode. One of the reasons for the failure of reflection mode to highlight key spectral features is that the microscope was not optimized to separate specular and diffuse reflections. The spectra obtained include both and hence some information may have been lost due to specular reflection. A recording from transmission mode involves light that has adequately interacted with the sample hence generating full information about the sample.
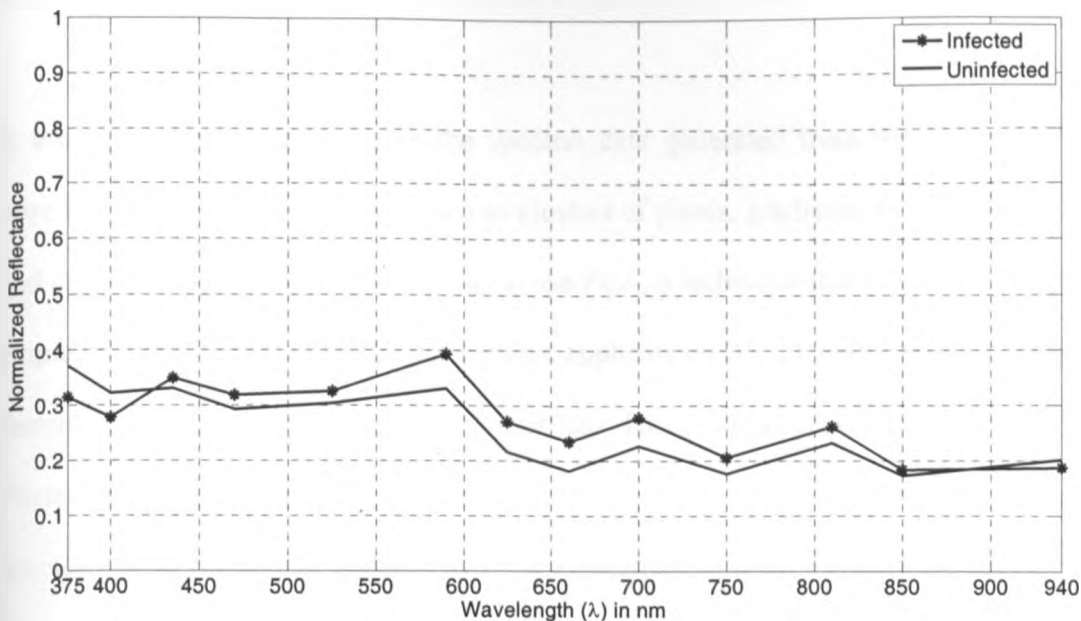
**Figure 5.7:** Reflectance spectra of infected and non-infected red blood cells.

From the spectra, both absorbance and reflectance, there is very little spectral differences between infected and uninfected red blood cells. This is due to the fact that in either case, the spectra are averaged values from the contribution of each pixel in the image. A typical red blood cell image is resolved by approximately 30X30 pixels. This means that the averaged spectrum was found from approximately 900 pixels (that is 900 spectra) whose plots are shown in Figure 5.4. Such complicated and highly correlated spectra which cannot be easily separated or interpreted by simple observation with the eye can be interpreted by employing multivariate chemometric techniques such as Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and Artificial Neural Network among others.

## 5.4 Image Analysis Based on Multivariate Chemometric Techniques

It was necessary to first explore the spectral data generated from the images for any observable relational phenomena such as clusters of pixels, gradients between the clusters and outliers. A good starting point was to use PCA, a technique that has been extensively employed successfully in multivariate image applications [17, 21, 36-37, 69-70] to segment features based on their spectral characteristics. HCA was expected to give clusters of *Plasmodium* species as demonstrated by Royston *et al.* [71] in fingerprinting bacterial species. However, due to limitations in the availability of other *Plasmodium* species, low spectral resolution of the microscope and the unfolding of the fact that, spectroscopically, in the 375-940 nm spectral range hemozoin was central in determining if a cell was infected or not, HCA has only been used to show the spatial distribution of heme-containing pigments in an infected red blood cell. The ultimate goal of this work was to develop a blood screening model based on a supervised pattern recognition technique. ANN was a good choice, having successfully been used to discriminate intra-erythrocytic stages of malaria parasites based on spectra obtained in synchrotron Fourier Transform-Infra-Red (FTIR) micro-spectroscopy [29].

## 5.4.1 Principal Component Analysis of Thin Blood Smear Images

PCA gives two important values: scores and loadings. The role of score values in PCA image analysis is two-fold: they can be observed as both gray-level decorrelated images and/or scatter plots. The first three score images (each of the score images is a gray-level

image) which contain most of information can be assigned to red, green and blue channels to form a false RGB image that has greatest contrast since the score vectors are orthogonal to each other. In addition, the scores can be visualized as score plots and segmentation done by studying the density of pixel clusters. Figure 5.8 shows a false colour composite of transmittance image generated by assigning the first, second and third principal components (the first 3 PCs contain 99 % variance) to red, green and blue channels respectively.



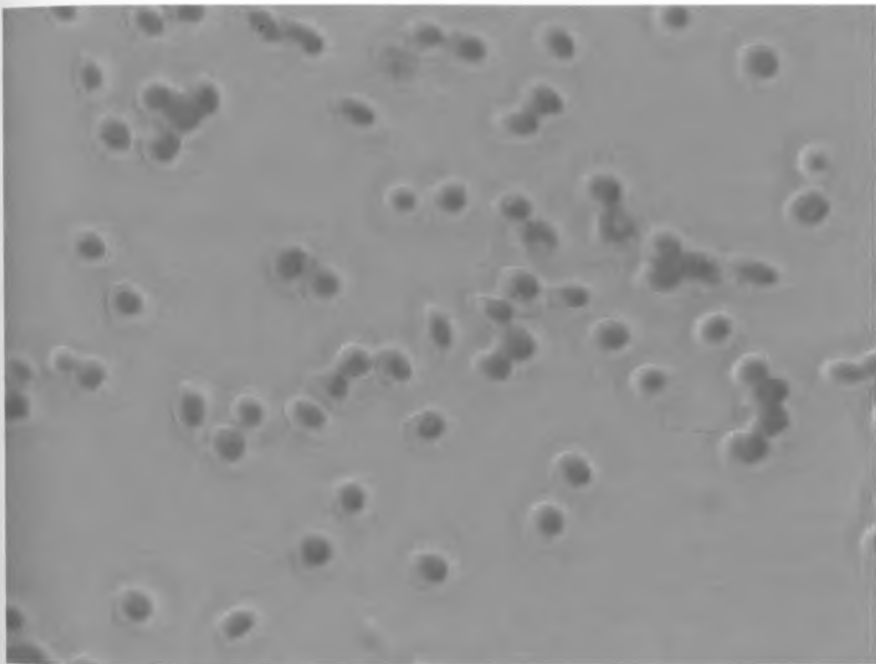**Figure 5.8:** A false colour composite of transmittance image generated by assigning the first three PCs to red, green and blue channels.

In the RGB image (Figure 5.8), any reddish, greenish or bluish colours observed signify the first, second and third principal components respectively. It can be seen that all the red blood cells in the image appear reddish, implying high values of the red channel (first

principal component) in the red blood cells. This means that the first principal component represents haemoglobin, a dominant pigment in the red blood cells. The image background appears grayish-green, a colour that arises from slightly higher values of the green channel (second principal component) than the other channels in this region of the image. The second principal component therefore represents the image background, which is majorly blood plasma. The little bluish colour on the red blood cells in the image is due to intra-cell variance resulting from curvature of the red blood cells, an implication that the third principal component has little significance in specific features in the whole image.

Visualizing score images as colour images is a rather subjective exercise because there may be differences in screen settings and people usually perceive colours differently. A more objective way is to study the PCs as score plots. The score plots generate clusters based on the fact that similar spectral features in the original image yield almost identical score combinations. Therefore, pixels from a specific feature in the image will overlap to create a dense cluster. These clusters are of great use in segmenting features in the image by delineating pixel clusters in the score plot which have similar spectral fingerprints. Pixel delineation is achieved by toggling between the score space and the image space. A cluster in the score space is selected and their original spatial location in the image space highlighted based on a threshold value in a gray level image calculated using equation 5.1

$$h = \frac{PC1 - PC2}{PC1 + PC2} \qquad\qquad (5.1)$$

where $h$ is a gray-level image upon which thresholding is done, PC1 and PC2 are the scatter-plotted principal components. This procedure of calculating $h$ is synonymous to that of calculating Normalized Difference Vegetation Index (NDVI) in satellite images [72]. Pixels falling below the threshold can be isolated and projected back in the image. By varying the threshold value of $h$, specific features in the image can be highlighted including the parasite. Figure 5.9 shows clusters in the score space and their corresponding origin in the image space.



(a)                                              (b)

**Figure 5.9:** Pixels of red blood cells for a transmission image (colour coded-blue) (a) delineated in the score space and (b) highlighted in the image space.

It is clear that the pixels in the blue cluster originated from the red blood cells. However, highlighting malaria parasites inside the red blood cells at this level may be a daunting task as the pixels from the parasites form a very small cluster overlapping with those of the red blood cells. As such, individual single red blood cells were cropped and the

process of delineating pixels in the score space and highlighting them in the image space repeated. The results of this process are shown in Figures 5.10 and 5.11.
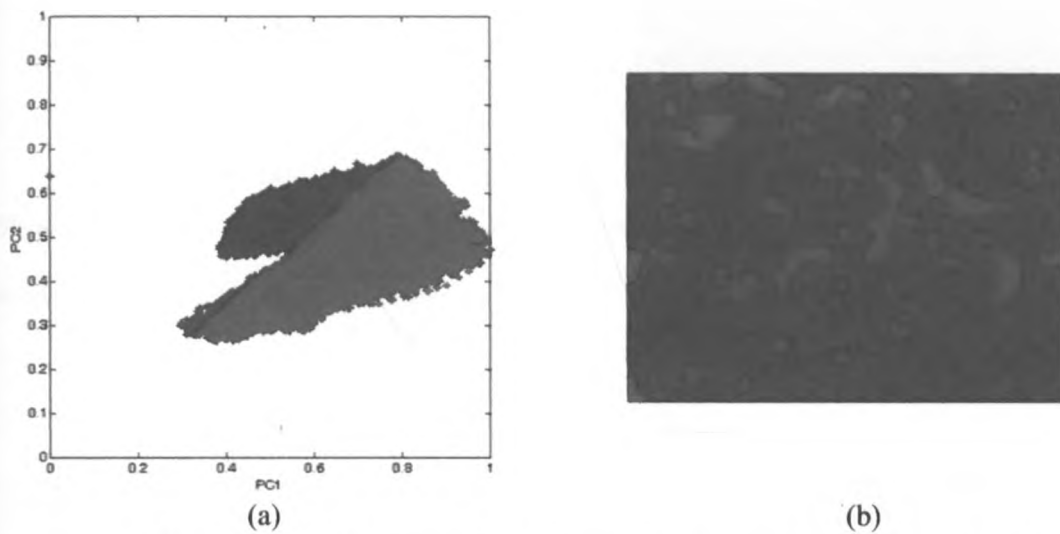


(a) (b)

**Figure 5.10:** Pixels of a single red blood cell (colour coded-blue) (a) delineated in the score space and (b) highlighted in the image space.



(a) (b)

**Figure 5.11:** Pixels of a malaria parasite (colour coded-blue) (a) delineated in the score space and (b) highlighted in the image space.

From these figures, it can be seen that the pixels from the parasite form a small cluster within the pixel cluster of the red blood cell. Overlapping of these clusters may have arisen from the fact that pixels in an infected red blood cell image contain mixed spectra of haemoglobin and hemozoin. In addition, the spectral features of hemozoin and haemoglobin are highly correlated. Both haemoglobin and hemozoin contain heme (a prosthetic group) which exists as a large heterocyclic organic ring called porphyrin with an iron ion at the centre. In hemozoin, this ring exists as ferriprotoporphyrin IX [73] whereas the non-protein part of haemoglobin exists as ferroprotoporphyrin IX. These two compounds exhibit $\pi$-$\pi^*$ electronic transitions of the porphyrin ring visualized as the strong absorption band (Soret band ) centered near 400 nm as seen in Figures 5.1, 5.3, 5.4 and 5.6. Their differences, however, arise from different locations of their vibronic transitions which occur at 540 nm and 575 nm for oxy-haemoglobin and 650 nm for hemozoin.

The procedure was repeated for images captured in reflection and dark-field modes of imaging with the microscope and the results are shown in Figures 5.12 and 5.13. From these clusters, it can be seen that reflection and dark-field mode images do not bring out clusters that can be easily delineated. The clusters from these modes are rather continuous with few outlier pixels far away from the main clusters. This makes it difficult to delineate specific features in the images.
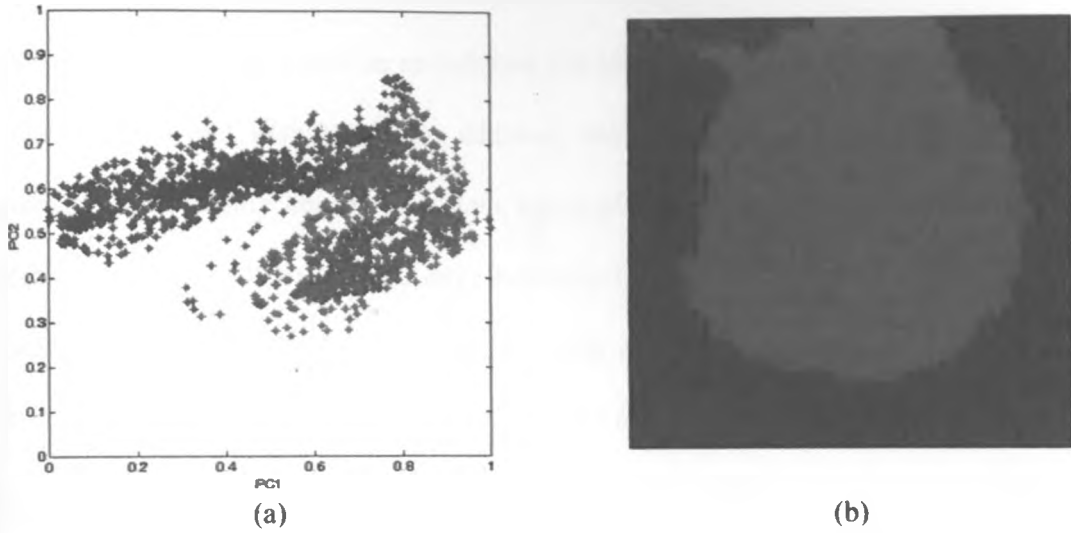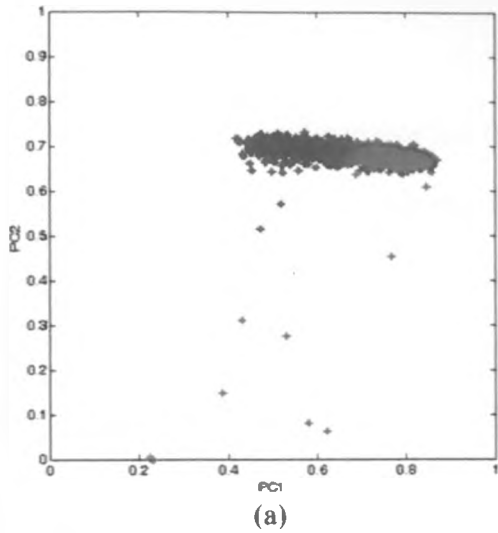
**Figure 5.12:** Pixels of a red blood cell (colour coded-blue) (a) delineated in the score space and (b) highlighted in the image space for an image captured in reflection mode.



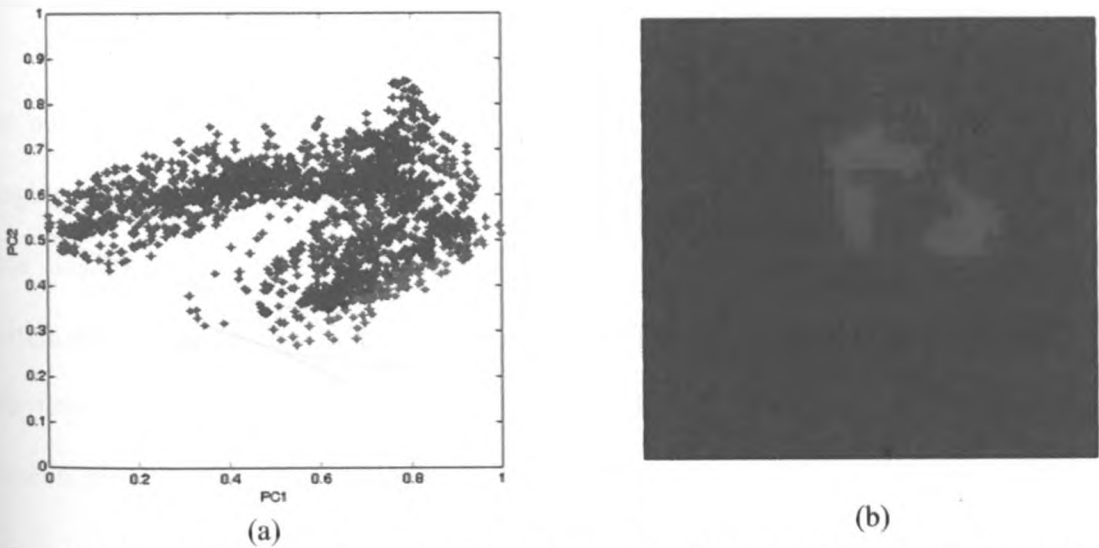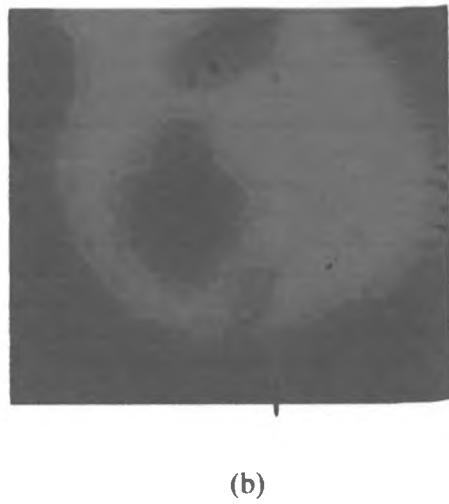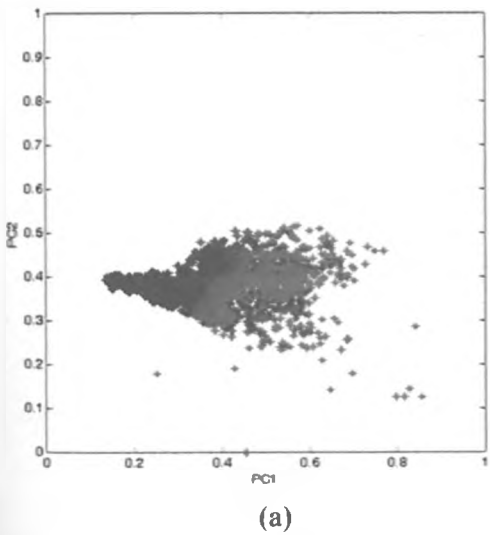**Figure 5.13:** Pixels of a red blood cell (colour coded-blue) (a) delineated in the score space and (b) highlighted in the image space for an image captured in dark-field mode.

## 5.4.1.1 Important Spectral Band Selection

The influence a measured variable has in each score is given by its loading. Hence, the loading plots are helpful in understanding the meaning of groups (in this case pixel clusters). Variables responsible for higher variance in the datasets generally give higher coefficients (positive or negative) on a certain principal component. In this regard, the contribution from different spectral bands to the PC eigen vectors can be observed in the loading plots. Figure 5.14 shows the loading plot of the first two PCs of images captured in the transmission mode.
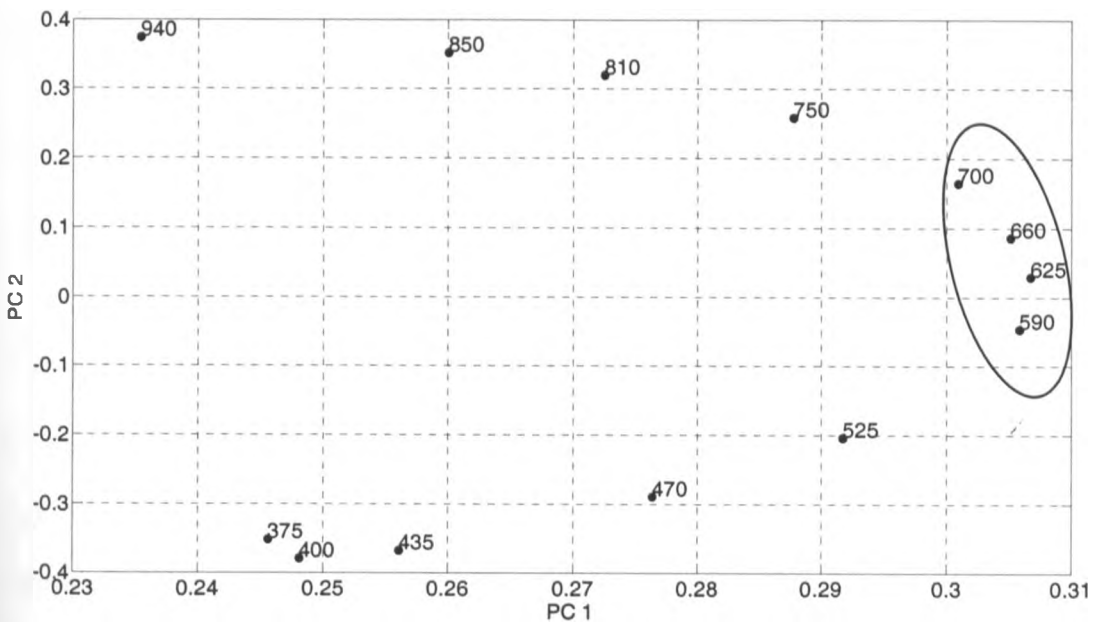


**Figure 5.14:** Loading plot of PC2 against PC1 from transmission image.

It was observed that the loading plots of the first and the second principal components of transmittance images were quite systematic for different red blood cells. From Figure 5.14, it can be observed that the variables loading heavily on the first principal component

are 590 nm, 625 nm, 660 nm and 700 nm. These are bands associated with the vibronic absorption bands of haemoglobin and hemozoin. They are the important bands for differentiating between infected and uninfected red blood cells. Along the second principal component, infrared bands (750-940 nm) have high positive loading values whereas ultra-violet (i.e. 375 nm) and blue region of visible light (400-525 nm) have high negative values. It is a clear indication of contrasting absorption behaviour of haemoglobin in the long wavelength (IR) radiations and short wavelength (UV and blue part of visible light) of the optical spectrum. In the former, haemoglobin exhibits very little (or no absorption) whereas in the latter, it absorbs strongly. This is also the case with hemozoin. Since haemoglobin and hemozoin have identical optical characteristics in these two spectral regions, it implies that detection of malaria parasites can be done by employing fewer LEDs emitting in the 590-700 nm spectral range in the multispectral microscope.

The loading plots of various PC combinations for reflection and scattering images were not reproducible, giving completely different results on different red blood cells; hence they have not been presented. Failure of the reflection and scattering results may be attributed to the random nature of light scattering on the surface of the red blood cells. Spectra from reflection mode, in particular, may contain both diffuse and specular reflections depending on the particular location from which light has been reflected on the (spherical) surface of a red blood cell.

## 5.4.1.2 Image Generation by Selective Illumination of Red Blood Cells

After PCA results had shown that the variables (wavelengths) of importance in malaria diagnostics for the microscope were 590 nm -700 nm, raw gray-level transmittance images captured at these important variables (wavelengths) were transformed to absorbance [74, 75] (by use of Beer-Lambert's law as explained in section 3.6) and compared to images captured at wavelengths contributing little variance such as those captured at 375 nm, 400 nm 850 nm and 940 nm. Figures 5.15, 5.16 and 5.17 show gray-level absorbance images captured at 400 nm, 625 nm and 940 nm respectively. In the image captured at 625 nm (Figure 5.16), the parasite region appears brighter due to its higher absorption in the image in comparison with other parts of red blood cells (see the gray-level scale). However, in the image captured at 400 nm (Figure 5.15) and that captured at 940 nm (Figure 5.17), there is little intensity difference between the parasites and the red blood cells. The reason for this is that at 400 nm and at 940 nm, the malaria pigment has almost identical absorption properties to those of haemoglobin. At 625 nm (and at the other three important spectral bands viz 590 nm, 660 nm and 700 nm), however, there is slight differences in absorption of haemoglobin and hemozoin, making the parasite more visible.
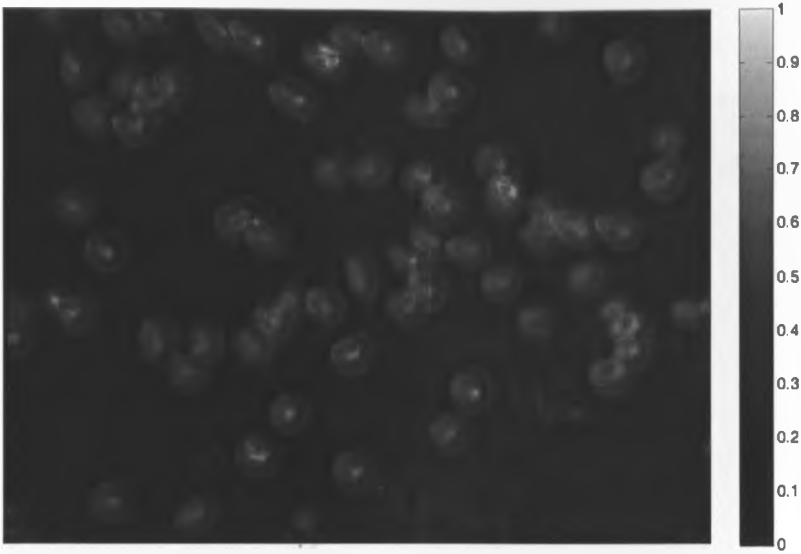
**Figure 5.15:** Image captured at a common absorption band (400 nm) for hemozoin and haemoglobin.
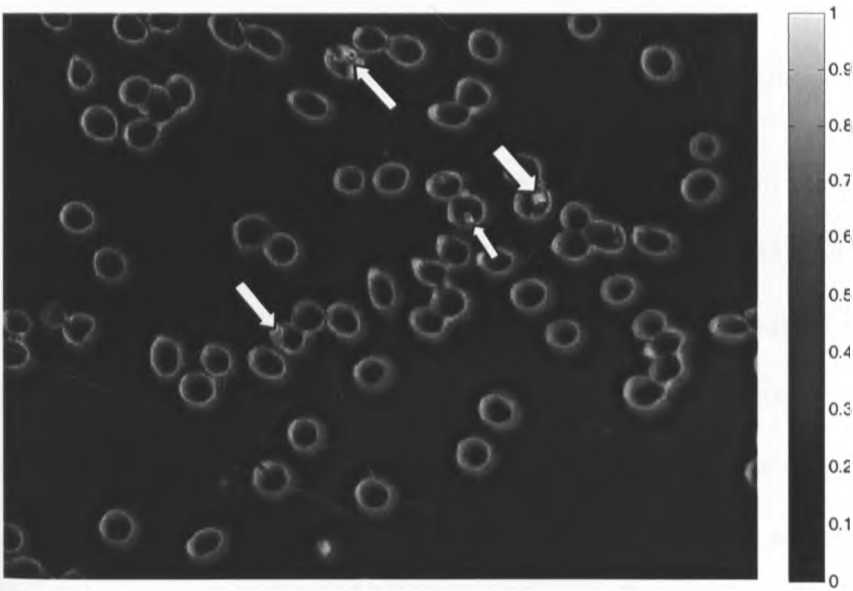


**Figure 5.16:** Gray-level image captured at 625 nm showing high absorption values for parasite region (shown with white arrows).

**Figure 5.17:** Image captured at a common non-absorbing band (940 nm) for hemozoin and haemoglobin.

### 5.4.1.3 Infected and Uninfected Red Blood Cell Discrimination

Having known that pixels emanating from a parasite region in transmission mode images formed clusters different from those of image background and non-parasite region, individual red blood cells were cropped again and spectra obtained (as in section 5.1) and labeled as either positive (infected) or negative (uninfected) based on observations in dark-field images and PCA segmentation. These spectra were then subjected to PCA and their positions in the score plots used to identify true positives and negatives for Artificial Neural Network model development.
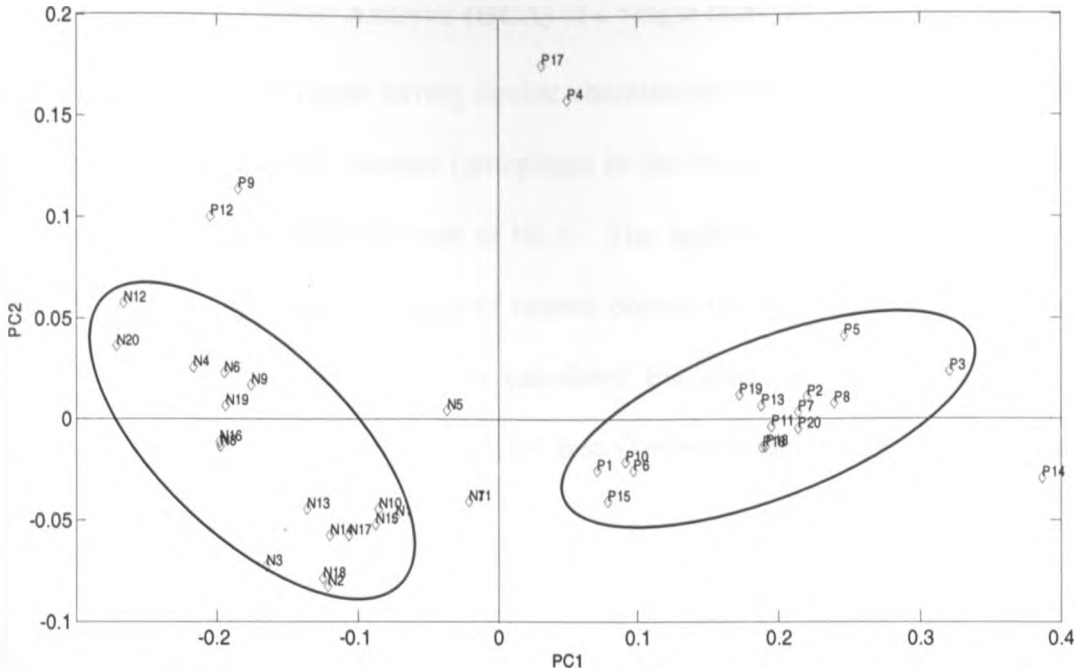
**Figure 5.18:** Score plot of labeled red blood cells; Ns are suspected non-infected and Ps are the suspected infected cells.

Figure 5.18 shows score plot of labeled red blood cells. P1 to P20 had been considered to be positive and N1 to N20 were the non-infected according to dark-field images. From this figure, the two main classes are separated along the first principal component. The clusters arise from the spectral variation of infected and uninfected red blood cells. As such, all P's were expected to form one class and all N's another. However, P9 and P12 are appearing in a wrong cluster while P4, P14 and P17 are outliers. The red blood cells appearing in unexpected cluster and those which were outliers were removed from the data set as they presented confusion on whether they were true positives or not, probably due to differences in the stages of plasmodia.

72

## 5.4.2 Hierarchical Cluster Analysis (HCA) of a Single Infected Red Blood Cell

Automatic grouping of pixels having similar characteristics in a multispectral image can be used to reveal 'natural' clusters (groupings) in the image data. Such clusters can be obtained by clustering methods such as HCA. The agglomerative HCA algorithm used here proceeds by successive mergers of nearest objects (in this case pixels) followed by linkages of nearest clusters based on calculated Euclidean distances between objects (pixels). This yields a hierarchical cluster tree (known as dendrogram) that represents how clusters are joined.

In analysis of a multispectral image, the resulting clusters can also be viewed as a spatial distribution of chemical species present in the image because HCA classifies pixels (spectra) into groups of spectral similarity. Figure 5.19 shows a dendrogram obtained by hierarchical cluster analysis of a single *Plasmodium falciparum*-infected red blood cell cropped from a transmittance image.
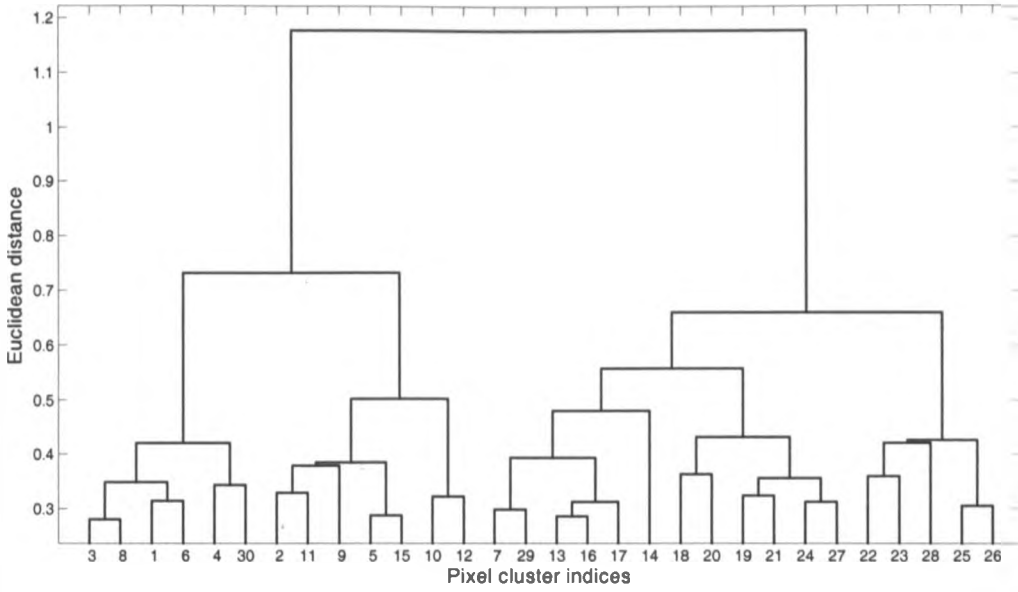
**Figure 5.19:** A dendrogram of a parasitized red blood cell cropped from transmittance image.

It is difficult to interpret pixel similarity indices from the dendrogram. However, by successively limiting the maximum number of clusters generated by the algorithm, a clustered spatial image map can be correlated with branches in the resulting dendrogram. Figure 5.20 shows the results of such a process.
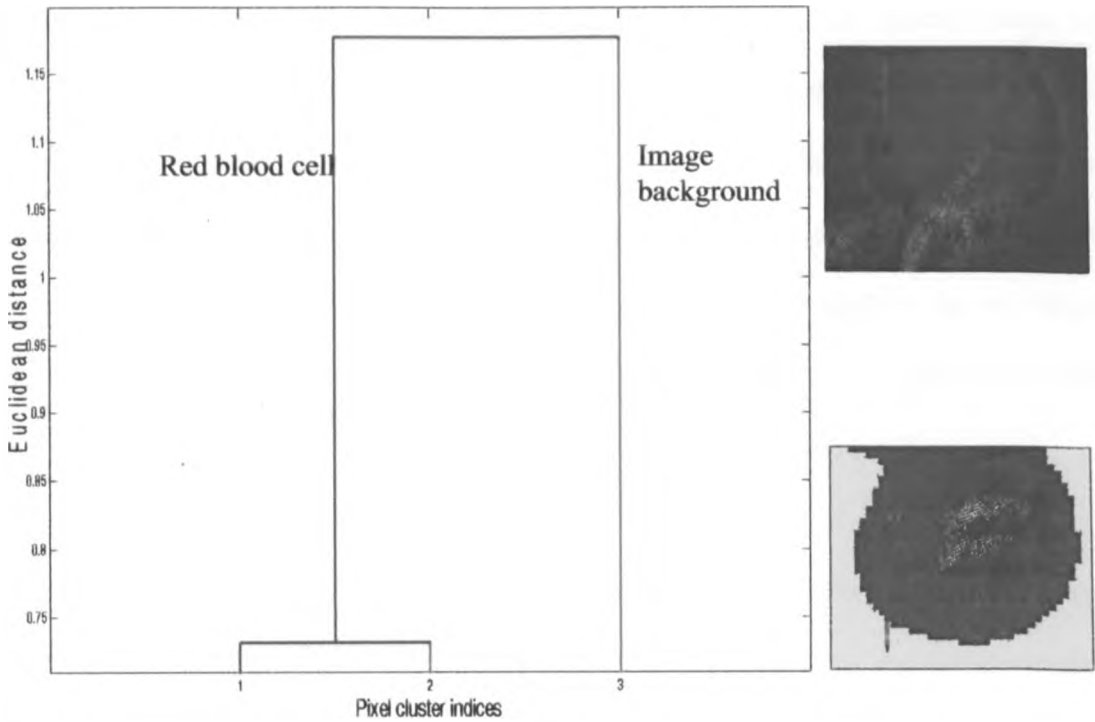
**Figure 5.20:** A dendrogram of an infected red blood cell and the corresponding spatial map generated by limiting the maximum number of clusters to three.

It can be seen that the red blood cell is separated from the background in the image in a two cluster map. Exceeding three clusters for this image resulted in clusters that were difficult to comprehend because each pixel is at least slightly different from the others and hence up to thirty clusters could be obtained (see Figure 5.19). A three cluster map therefore appears more appropriate for displaying the spatial distribution of heme species in an infected red blood cell.

In an unstained parasitized red blood cell, haemoglobin and hemozoin are the most abundant pigments that we can expect to detect in the 375-940 nm spectral range by transmittance spectra, since they are known to absorb light in this spectral range. The red and blue clusters in the three-cluster-red blood cell in Figure 5.20 (lower right) can be attributed to hemozoin and haemoglobin respectively, because haemoglobin is the most abundant pigment in a red blood cell and is usually distributed throughout the red blood cell. On the other hand, hemozoin is usually localized in the parasite's digestive vacuole and is therefore concentrated in one specific region inside the red blood cell.

HCA was also performed on cropped infected red blood cells captured in reflectance and scattering modes. The clustering results in form of dendrograms and the corresponding spatial maps are shown in Figures 5.21, 5.22, 5.23, and 5.24.
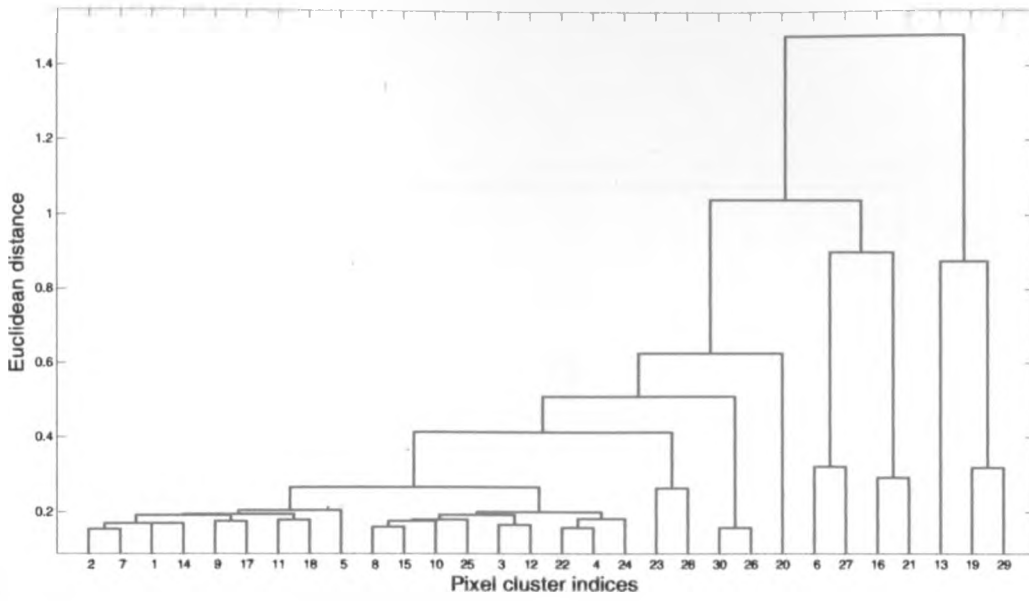
**Figure 5.21:** A dendrogram of a parasitized red blood cell cropped from reflectance image.
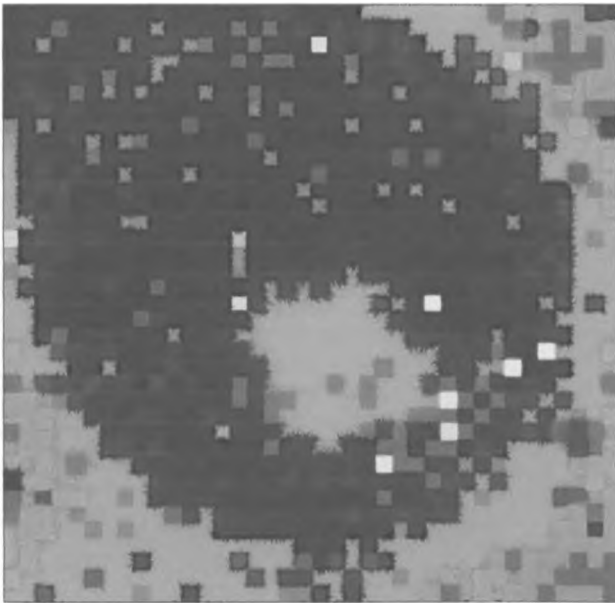


**Figure 5.22:** Spatial map of clusters obtained from reflectance image.
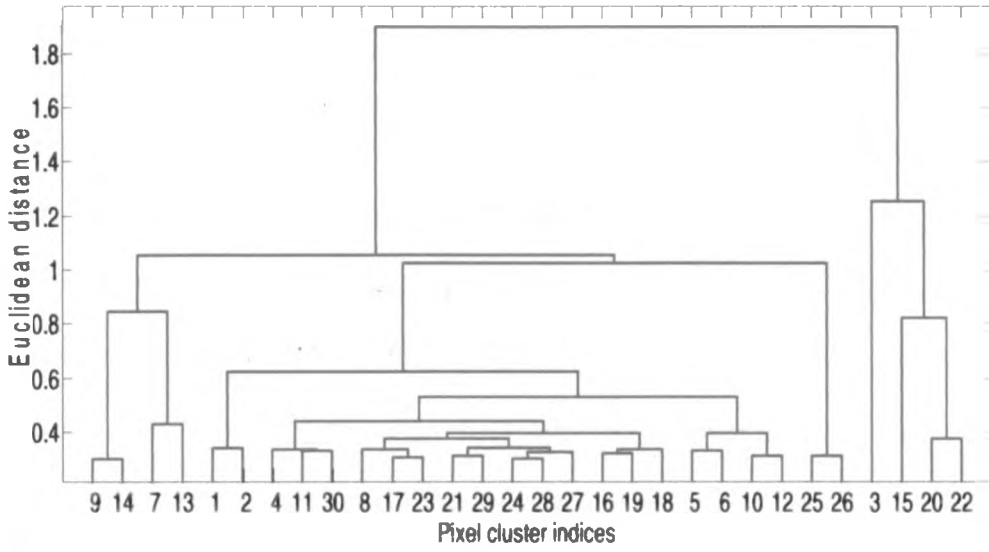
**Figure 5.23:** A dendrogram of a parasitized red blood cell cropped from scattering image.
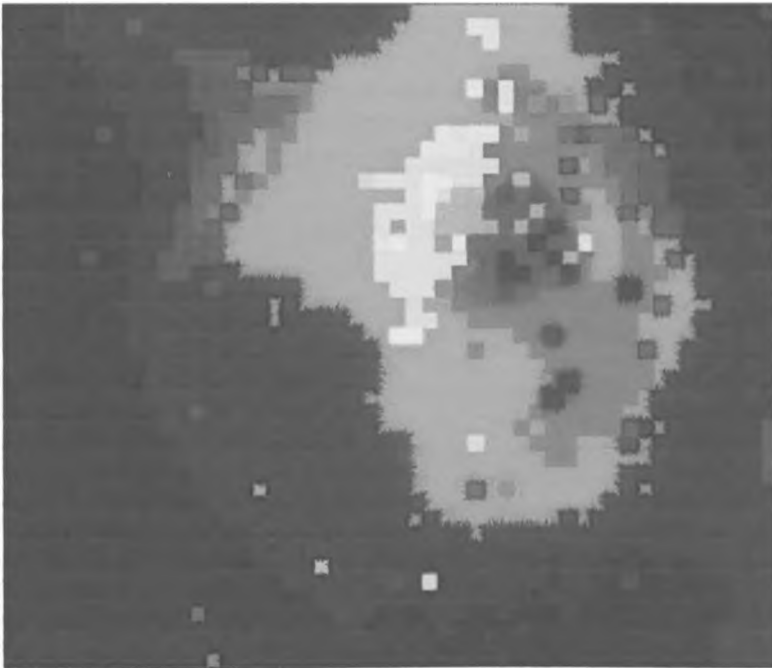


**Figure 5.24:** Spatial map of clusters obtained from scattering image.

In Figure 5.21 the dendrogram shows that there are two main clusters in reflectance image representing the image background and the red blood cell. We would expect the red blood cell to divide into two clusters representing haemoglobin and hemozoin as is the case in transmittance images. However, this is not the case. Instead, by inspecting the corresponding cluster map (Figure 5.22) of the dendrogram, it can be seen that some pixels inside the red blood cell are identical to those found outside the red blood cell and part of the edges of the red blood cell (and indication of similar spectral similarity). This unexpected clustering can be attributed to the shape of the red blood cell. Since the red blood cell is not flat, different parts reflect light into different directions at different points on the red blood cell in form of diffuse and specular reflections.

In Figure 5.23, the dendrogram of the scattering image also has two main branches. These two branches represent the high and low scattering parts of the infected red blood cell. The high scattering part is colour-coded 'light blue' in Figure 5.24 and is the parasite region. The 'dark blue' colour in the spatial map (Figure 5.24) is the low scattering part (the red blood cell and its background). However, it is difficult to tell what the other clusters represent. Thus, the best clustering results for malaria diagnostics were those of images taken in transmission mode.

## 5.4.3 Development of Artificial Neural Network (ANN) Model

Unsupervised pattern recognition techniques are only good for data exploration. To develop a blood screening model for malaria detection, a supervised technique is more

useful. Artificial Neural Network is a powerful tool for classification whose performance largely depends on the chosen parameters such as the transfer function employed, the number of layers and neurons used and the linearity or non-linearity of the data.

Since infected and non-infected red blood cells had shown spectral overlapping features, a multi-layer perceptron neural network with a Levenberg-Marquardt back propagation learning algorithm with one hidden layer was chosen because of its ability to solve non-linear problems [61]. The optimum number of neurons in the hidden layer was determined by trial starting from one neuron and increasing the number by one in the subsequent trials as the models were trained and tested.

As discussed earlier (section 3.8.2), the performance of a trained neural network can be measured by the errors in training, validation and test sets. However, it is also useful to investigate the network response by performing a regression analysis between the network response and the corresponding targets. A correlation coefficient (known as R-value) can then be used to explain the correlation between the targets and the outputs of the network. R-value is a measure of how well the variation in the output is explained by the targets. If its numerical value is close to 1, then there is good correlation between targets and the outputs. Table 5.1 shows the results of training the neural network for various numbers of neurons in the hidden layer. It can be noted that most of the models are efficient in predicting the state of infection as their R value is close to unity. The best model was that with high R value, with low mean-squared error and with fewer neurons in the hidden layer. A high number of neurons in the hidden layer would cause over-fitting while too

80

few would not be able to classify accurately. As such, the model with 5 neurons was adopted because it exhibited lower validation mean-squared error with a consistent R-value that is close to one.

After training, the neural network was challenged with new red blood cell images. Due to the limitation of the magnification of the microscope, training was based on identifying each red blood cell as either infected (-1) on not infected (+1) without identification of stage. The developed network was able to classify all the red blood cells correctly. Since classification is done at cell level, parasitemia was obtained by taking the number of infected cells as a percentage of all examined cells. The time taken to produce diagnostic results with this model was approximately ten minutes, which was way below that of conventional microscopy which takes about 30 minutes.

**Table 5.1:** Results of training neural network for varying number of neurons in the hidden layer.

| Number of neurons in the hidden layer | Best validation performance mean squared error | No. of epochs | Correlation coefficients | | | |
|---|---|---|---|---|---|---|
| | | | Training R-value | Validation R-value | Test R-value | Overall R-value |
| 1 | $5.000 \times 10^{-1}$ | 23 | 1 | 0.7746 | 0.7746 | 0.90453 |
| 2 | $3.361 \times 10^{-3}$ | 15 | 0.99998 | 0.9992 | 1 | 0.99719 |
| 3 | $6.798 \times 10^{-2}$ | 4 | 1 | 0.89002 | 0.9946 | 0.97912 |
| 4 | $2.6798 \times 10^{-1}$ | 2 | 0.99852 | 0.79571 | 0.8004 | 0.92553 |
| 5 | $8.8317 \times 10^{-4}$ | 38 | 1 | 0.99993 | 0.95462 | 0.9928 |
| 6 | $8.0126 \times 10^{-5}$ | 437 | 1 | 0.99999 | 0.99393 | 0.99884 |
| 7 | $5.7077 \times 10^{-2}$ | 9 | 1 | 0.98028 | 0.98491 | 0.9915 |
| 8 | $9.1095 \times 10^{-11}$ | 37 | 1 | 1 | 0.78055 | 0.95169 |
| 9 | $6.9042 \times 10^{-2}$ | 5 | 1 | 0.96782 | 0.99714 | 0.99196 |
| 10 | $7.9496 \times 10^{-2}$ | 9 | 1 | 0.93701 | 0.97479 | 0.9793 |
| 11 | $7.0875 \times 10^{-3}$ | 6 | 1 | 0.99788 | 0.620.06 | 0.91034 |
| 12 | $1.2602 \times 10^{-3}$ | 8 | 1 | 0.99357 | 0 | 0.9812 |
| 13 | $7.4448 \times 10^{-2}$ | 6 | 1 | 0.95243 | 0.97782 | 0.98837 |
| 14 | $8.6322 \times 10^{-2}$ | 5 | 1 | 0.95896 | 0.99985 | 0.99125 |
| 15 | $3.7646 \times 10^{-1}$ | 2 | 1 | 0.69333 | 0.96159 | 0.90589 |
| 16 | $1.6261 \times 10^{-2}$ | 3 | 1 | 0.99093 | 0.9953 | 0.99709 |
| 17 | $1.0209 \times 10^{-2}$ | 6 | 1 | 0.97754 | 0.98191 | 0.98962 |
| 18 | $3.5364 \times 10^{-2}$ | 4 | 1 | 0.98698 | 0.94035 | 0.97977 |
| 19 | $5.6368 \times 10^{-2}$ | 3 | 1 | 0.98006 | 0.98453 | 0.99104 |
| 20 | $4.0484 \times 10^{-2}$ | 3 | 1 | 0.99132 | 0.99459 | 0.99495 |

## CONCLUSION AND FUTURE PROSPECTS

### 6.1 Conclusion

A rapid malaria diagnostic method has been developed in this thesis. The method is based on detection of *Plasmodium* spectral signature derived from pixels of a multispectral microscopic image of unstained thin blood smear. The signatures identify red blood cells as either positive or negative depending on the presence or absence of hemozoin. The fingerprints of haemoglobin and hemozoin (two pigmented compounds that would be found in an infected red blood cell) have been found to lie between 590 nm and 700 nm (a small section of the visible region of the electromagnetic spectrum).

Reliance on hemozoin signature to identify the malaria parasites without the associated morphological features of the parasites is inadequate to differentiate species of *Plasmodia*. To differentiate the four species of plasmodia affecting humans, a higher resolution microscope that can reveal the morphology of either the whole parasite cells or differences in hemozoin crystals for the different *Plasmodium* species is required. Thus, this work has been limited to detection of *Plasmodium falciparum* only. Due to correlated nature of spectral signatures of hemozoin and haemoglobin, multivariate chemometric techniques have been employed to enhance detection. PCA and HCA were used in data exploration. PCA was used to find the most important spectral bands for malaria detection in 375-940

nm spectral range in absorption spectroscopy. HCA was used to highlight the distribution of pigments in *Plasmodium*-infected red blood cell. Finally, ANN was used for classification of red blood cells as either infected or not infected.

## 6.2 Future Prospects

PCA results have shown that the discriminating wavelengths for infected and non-infected red blood cells are found in the region 590-700 nm. This is in agreement with the results of Yulia *et al.* [28] and Oliveira *et al.* [47] which showed that it is a fingerprint region for hemozoin. The microscope can therefore be optimized for malaria detection by employing LEDs whose emissions are centered at 650 nm to pick out hemozoin as the highest absorbing objects in the gray-level images acquired. In addition, presence of nucleic material in a red blood cell can also be a biomarker of malarial infection, since red blood cells do not generally have nucleus. As such, detection of nucleic material may be made possible by use of a 260 nm LED in conjunction with UV detectors such as photomultiplier tubes in the microscope.

An artificial neural model has been developed to classify red blood cells as either positive (infected) or negative (non-infected). The limitation of this network is pegged on inability of humans to classify the parasites in unstained smear. Since ANN is a supervised pattern recognition technique, its inputs were based on correctly classified cells from dark-field microscopy and PCA. Further exploitation of this technique is possible by preparation of samples on slides containing pure stages of the parasites unlike those used in this work

which contained all stages on a single slide. For field study, the microscope should be used alongside the established conventional microscopy to determine its sensitivity and specificity in field conditions.

# REFERENCES

[1]     Aregawi, M., Cibulskis, R., Otten, M., Williams, R. and Otten, M. (2009), World Malaria Report 2009, WHO.

[2]     Lema, O. E., Carter, J. Y., Nagelkerke, N., Wangai, M. W., Kitenge, P., Gikunda, S. M., Arube, S. M., Munafu, C. G., Materu, S. F., Adhiambo, C. A. and Mukunza, H. K. (1999), "Comparison of five methods of malaria detection in the outpatient setting", American Journal of Tropical Medicine and Hygiene 60(2): 177-182.

[3]     Storey, J. (2010), Basic Malaria Microscopy: Part 1. Learner's Guide, World Health Organization, Geneva.

[4]     Garcia, M., Kirimoama, S., Marlborough, D., Leafasia, J. and Rieckmann, K. H. (1996), "Immunochromatographic test for malaria diagnosis (Letter)", Lancet 347(9014):1549.

[5]     Shiff, C. J., Minjas, J. N. and Premji, Z. (1994), "The Parasight®-F test: a simple rapid manual dipstick test to detect *Plasmodium falciparum* infection", Parasitology Today 10:494-495.

[6]     Oliveira, D. A., Shi, Y. P. and Oloo, A. J. (1996), "Field evaluation of polymerase-chain reaction based nonisotopic liquid hybridization assay for malaria diagnosis", Journal of Infectious Diseases 173:1284-1287.

[7]     Snounou, G., Viriyakosol S., Jarra, W., Thaithong, S. and Brown, K. N. (1993), "Identification of the four human malaria parasites in field samples by the polymerase-chain-reaction and detection of a high prevalence of mixed infections", Molecular and Biochemical Parasitology 58:283-292.

[8]     Tirasophon, W., Rajkulchai, P., Ponglikitngkol, M., Wilairat, P., Boonsaeng, V.
and Panyim, S. (1994), "A highly sensitive, rapid and simple polymerase-chain-
reaction method to detect human malaria (*Plasmodium falciparum* and
*Plasmodium vivax*) in blood samples", American Journal of Tropical Medicine and
Hygiene 51:308-313.

[9]     Belisle, J. M., Costantino, S., Leimanis, M. L., Bellemare, M. J., Bohle, D. S.,
Georges, E, and Wiseman, P. W. (2007), "Sensitive detection of malaria infection
by third harmonic generation imaging", Biophysical Journal: Biophysical Letters:
L26-L28.

[10]    Ong, C. W., Shen, Z. X., Ang, K. H., Kara, U. A. K. and Tang, S. H. (2002),
"Raman microspectroscopy of normal erythrocytes and *Plasmodium berghei*-
infected erythrocytes", Applied Spectroscopy 56:1126-1131.

[11]    Grant, T. W., Leann, T., Samantha, D., Don, M. and Bayden, R. W. (2008),
"Resonance Raman spectroscopy can detect structural changes in haemozoin
(malaria pigment) following incubation with chloroquine in infected erythrocytes",
FEBS Letters 582:1087-1092.

[12]    Minh, T. L., Timo, R. B., Claudia, K. 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 2008:9-15.

[13]    Ross, N. E., Pritchard, C. J., Rubin, M. D. and Adriano, G. D. (2006), "Automated
image processing method for the diagnosis and classification of malaria on thin
blood smears", Med Biol Eng Compt 44:427-436.

[14] Gloria, D., Fabio, A. G. 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 42:296-307.

[15] Uchiyama, T., Yamaguchi, M., Ohyama, N., Mukawa, N. and Kaneko, H. (2006), "Multispectral image retrieval using a distance based on vector quantization", Electronics and Communications in Japan 3 (89):1546-1555.

[16] Wold, J. P., Westad, F. and Heia, K. (2001), "Detection of parasites in cod fillets by using SIMCA classification in multispectral images in the visible and NIR region", Applied Spectroscopy 55:1025-1034.

[17] Shah, S. and Thigpen, J. (2008), "Multispectral microscopy for cell differentiation in thyroid cytology", Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems TA5-3:267-271.

[18] Wu, Q., Zeng, L. and Guo, N. (2006), "A method based on multispectral imaging technique for white blood cell segmentation", Computers in Biology and Medicine 37:70-76.

[19] Zoueu, T. J., Loum, G. L., Haba, C. T., Mikkel, B. and Menan, H. (2008), "Optical microscope based on multispectral imaging applied to *Plasmodium* diagnosis", Journal of Applied Sciences 8 (15):2711-2717.

[20] Noordam, J. C., Broek, W. H. and Buydens, L. M. C (2005), "Detection and classification of latent defects and diseases on raw French fries with multispectral imaging", J Sci Food Agri 85:2249-2259.

[21]    Kvaal, K. and Wold, J. P. (2000), "Mapping lipid oxidation in chiken meat by multispectral imaging of autofuorescence", Applied Spectroscopy 54:900-909.

[22]    Wu, Q., Zeng, L., Zheng, H. and Guo, N (2009), " Precise segmentation of white blood cells by using multispectral imaging analysis techniques", First International Conference on Intelligent Networks and Intelligent systems:491-494.

[23]    Craig, M. H. and Sharp, B. L. (1997), "Comparative evaluation of four techniques for the diagnosis of *Plasmodium falciparum* infections", Transanctions of the Royal Society of Tropical Medicine and hygiene 91:279-282.

[24]    Makler, M. T., Ries, L. K., Ries, J., Horton, R. J. and Hinrichs, D. J. (1991), "Detection of *Plasmodium* infection with fluorescent dye, benzothiocarborxypurine", American Journal of Tropical Medicine and Hygiene 44:11-16.

[25]    Kawamoto, F. (1991) "Rapid diagnosis of malaria by fluorescence microscopy with light microscope and interference filter", Lancet 337:200-202.

[26]    Htut, Y., Aye, K. H., Han, K. T., Kyaw, M. P., Shimono, K. and Okada, S. (2002), "Feasibility and limitations of acridine orange fluorescence technique using a malaria diagnosis microscope in Mynamar", Acta Med. Okayama 46 (5):219-222.

[27]    Bremard, C., Girerd, J. J, Kowalewski, P., Merlin, J. C. and Moreau, S. (1993) "Spectroscopic investigations of malaria pigment", Applied Spectroscopy 47:1837-1842.

[28]  Yulia, M. S., Patel, J. and Garcia-Rubio, L. H. (2010), "Interpretation of the ultraviolet-visible spectra of malaria parasite *Plasmodium falciparum*", Applied Optics 49 (2):180-188.


[29]  Webster, G. T., Katherine, A.V., Egan, T. J., Tilley, T., Tobin, M. J., Bambery, K. R., McNaughton, D. and Wood, B. R. (2009), "Discriminating the intraerythrocytic lifecycle stages of the malaria parasite using synchrotron FT-IR microspectroscopy and an Artificial Neural Network", Anal Chem 81: 2516-2524.


[30]  Jamjoom, G. A. (1983) "Dark-field microscopy for detection of malaria in unstained blood films", Journal of Clinical Microbiology 17:717-721.


[31]  Kim, M. S., Lefcourt, A. M., Chao, K., Chen, Y. R., Kim, I. and Chan, D. E. (2002) "Multispectral detection of fecal contamination on apples based on hyperspectral imagery. Part I. Application of visible and near-infrared reflectance imaging", Tranact ASAE 45(6):2027-2037.


[32]  Polder, G., Heijden, G. W. A. M. and Young, I. T. (2002), "Spectral image analysis for measuring ripeness of tomatoes", Transact ASAE 45(4):1155-1161.


[33]  Guyer, D. and Yang, X. (2000), "Use of genetic artificial neural networks and spectral imaging for defect detection on cherries", Computers and Electronics in Agriculture 29:179-194.


[34]  Dhawan, A. P., Alessandro, B. D., Sachin, P. and Mullani, N. (2009), "Multispectral optical imaging of skin-lesions for detection of malignant melanomas", Annual International Conference of the IEEE (31):5352-5355.

[35]     Gowen, A. A., Donnel, C. P., Cullen, P. J., Downey, G. and Frias, J. M. (2007), "Hyperspectral imaging – an emerging process analytical tool for food quality and safety control", Trends in Food Science & Technology (18): 590-598.

[36]     Gowen, A. A., Donnell, C. P., Taghizadeh, M., Cullen, P. J., Frias, J. M. and Downey, G. (2007), "Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (*Agaricus bisporus*)", Journal of Chemometrics 22:259-267.

[37]     Ariana, D. P., Lu, R. and Guyer, D. E. (2006), "Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers", Computers and Electronics in Agriculture 53:60-70.

[38]     Geladi, P. (2003), "Chemometrics in Spectroscopy. Part 1. Classical Chemometrics" Spectrochimica Acta Part B 58:767-782.

[39]     Bannister, L. H., Hopkins, J. M., Fowler, R. E., Krishna, S. and Mitchell, G. H. (2000), "A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages", Parasitology Today 6:427-433.

[40]     Valey, V. T. (2002), Handbook of Optical Biomedical Diagnostics, SPIE, Washington D.C.

[41]     Goldberg, D. E., Slater, A. F. G., Cerami, A. and Anderson, G. B. (1990), "Haemoglobin degradation in the malaria parasite *Plasmodium falciparum*: An ordered process in a unique organelle", Proc Natl Acad Scie USA 87:2931-2935.

[42] Sherman, I. W. (1979), "Biochemistry of *Plasmodium* (malarial parasites)", Microbiol Rev 43 (4):453.

[43] Nash, G. B., O'Brlen, E., Smith, G. E. C. and Dormandy, J. A. (1989), "Abnormalities in the mechanical properties of red blood cells caused by *Plasmodium falciparum*", Blood 74 (2):855-861.

[44] Paulitshke, M. and Nash, G. B. (1993), "Membrane rigidity of redblood cells parasitized by different strains of *Plasmodium falciparum*", J. Lab Clin Med 122(5):581-589.

[45] Cranston, H. A., Boylan, C. W., Carroll, G. L., Sutere, S. P., Williamson, J. R., Gluzman, I. Y. and Krogstad, D. J. (1984), "*Plasmodium falciparum* maturation abolishes physiologic red cell deformability", Science 223 (4634):400-403.

[46] Miller, L. H., Baruch, D. I., Marsh, K. and Doumbo, O. K. (2002), "The pathogenic basis of malaria", Nature 415 (6872):673-679.

[47] Oliveira, M. F., Gandara, A. N. P., Braga, C. M. S., Silva, J. R., Mury, F. B., Petretski, M. D., Menezes, D., Marcos, A. V. and Oliveira, P. (2007), "Heme crystalization in the midgut of triatomine insects", Comparative Biochemistry and Physiology, Part C 146:168-174.

[48] Hempelmann, E. and Marques, H. M. (1994), "Analysis of malaria pigment from *Plasmodium falciparum*", Journal of Pharmacological and Toxicological Methods (32):25-30.

[49]   Yong, K. P., Monica, D. S., Popscu, G. Lykotrafitis, G., Choi, W. Feld, M. S. and Suresh, S. (2008), "Refractive index maps and membrane dynamics of human red blood cells parasitized by *Plasmodium falciparum*", PNAS 105(37):13730-13735.

[50]   Murphy, D. M. (2001), Fundamentals of Light Microscopy and Electronic Imaging, Wiley-Liss Inc, New York.

[51]   Cox, G. (2007), Optical Imaging Techniques in Cell Biology, Taylor & Francis, New York.

[52]   Mikkel, B., Zuguang, G. and Sune, S. (2008), "Broad-band multispectral microscope for imaging transmission spectroscopy employing an array of light-emitting diodes", American Journal of Physics 77(2):104-110.

[53]   Geladi, P. and Grahn, H. (1996), Multivariate Image Analysis, John Wiley & Sons Inc., New York.

[54]   Liqun, W. and Boris, M. (2008), "Application of multivariate data- analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy", Anal Bioanal Chem 391:1641-1654.

[55]   Huang, J., Wium, H., Qvist, K. B. and Esbensen, K. H. (2003), "Multi-way methods in image analysis: relationships and applications", Chemometrics and Intelligent Laboratory Systems 66:141-158.

[56]   Kaufman, L. and Rousseeuw, P. J. (2005), Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons Inc., New Jersey.

[57]  Thanh, N. T., Wehrens, R. and Buydens, L. M. C. (2004), "Clustering multispectral images: a tutorial", Chemometrics and Intelligent Laboratory Systems 77:3-17.

[58]  Fausett, L. (1994), Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Prentice-Hall, Upper Saddle River.

[59]  Hagan, M., Demuth, B. H. and Beale, M. (1996), Neural Network Design, PWS, New York.

[60]  Hagan, M., Demuth, B. H. and Beale, M. (2009), Neural Network Tolbox user's guide, Math Works Inc., New York.

[61]  Marini, F., Bucci, R., Magri, A. L. and Magri, A. D. (2007), "Artificial Neural Networks in Chemometrics: History, Examples and Perspectives", Microchemical 88(2):178-185.

[62]  Yusuf, E., Rao, B. H. and Singh, D. N. (2008), "Artificial neural network model for predicting soil thermal resistivity", International Journal of Thermal Sciences 47:1347-1358.

[63]  Aboma, M. (2010), "Multispectral microscopy with applications to malaria detection", M.Sc thesis, Lund University.

[64]  Allied Vision Technologies (2009), AVT Guppy Technical Manual, Germany.

[65]  Geladi, P., Britta, S., Nystrom, J., Lillhonga, T., Torbjorn, L. and Burger, J. (2004), "Chemometrics in Spectroscopy: Part 2. Examples", Spectrochimica Acta Part B (59):1347-1357.

94

[66] Geladi, P., Isakson, H., Lindqvist, L., Wold, S. and Esbensen, K. (1989), "Principal Component Analysis of Multivariate Images", Chemometrics and Intelligent Laboratory Systems 5: 209-220.

[67] Hammes, G. G. (2005), Spectroscopy for the Biological Sciences, Wiley InterScience, New Jersey.

[68] Metzler, D. E. (2003), Biochemistry the chemical reactions of living cells, Vol 2, Elsevier Science, London.

[69] Costa, J. C., Alves, M. M. and Ferreira, E. C. (2009), "Principal component analysis and quantitative image analysis to predict effect of toxics in anaerobic granular sludge", Bioresource Technology 100:1180-1185.

[70] Baronti, S., Casini, A., lotti, F. and Porcinai, S. (1997), "Principal component analysis of visible and near-infrared multispectral images of works of art", Chemometrics and Intelligent Laboratory Systems 39:103-114.

[71] Goodacre, R., Timmins, E. M., Burton, R., Kaderbhai, N., Woodward, A. M., Kell, D. B. and Rooney, P. J. (1998), "Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks", Microbilogy 144:1157-1170.

[72] Ray, D. J. and Huete, A. R. (1991), "Interpreting vegetation indices", Preventive Veterinary Medicine 11:185-200.

[73] Coy, D. F. and Phitsamai, K. (1987), "The state of ferriprotoporphyrin IX in malaria pigment", Biological Chemistry 262:15552-15555.

[74] Sun, D. (2010), Hyperspectral Imaging for Food Quality Analysis and Control, Elsevier Inc., London.

[75] Geladi, P., Burger, J. and Lestander, T. (2004), "Hyperspectral imaging: calibration problems and solutions", Chemometrics and Intelligent Laboratory systems 72:209-217.

## APPENDIX

Matlab codes:

**To import gray level images (should be saved as** `importImages`**)**

```
dn=uigetdir('C:\Documents and Settings\diod\Desktop\from
desktop\MyApp\Images\');

fn=dir([dn '\*.tiff']);

clear M
for n=1:length(fn)
    M(:,:,1+str2num(fn(n).name(1:(end-5))))=imread([dn '\'
fn(n).name]);
end

pn=dir([dn '\*.xls']);
```

**To calculate a corrected multispectral image**

```
importImages
bright=double(M);
clear M
importImages
dark=double(M);
clear M
importImages
sample=double(M);
clear M


image=(sample-dark)./(bright-dark);
```

**To crop a 13-band multispectral image**

```
function I=imstackcrop(x);
%I=IMSTACKCROP(X) crops and m-by-n-by-13 image
%from X which is also m-by-n-by-13.  The function
%displays the first gray scale image in X to allow
%you to zoom and define the point to be cropped.
%Press ENTER after zooming, the cursor will change
%to crosshair.  Define the cropping rectangle by dragging
%the crosshair over the region to be cropped.
```

```matlab
imshow(x(:,:,1),[])
pause


rect=getrect;
for k=1:13
    I(:,:,k)=imcrop(x(:,:,k),rect);
end
```

**To apply a median filter to a 13-band image**

```matlab
function I=imstackfilt(x)


for k=1:13
    I(:,:,k)=medfilt2(x(:,:,k));
end
```

**To perform Principal Component Analysis**

```matlab
function [I1, coeff, var]=imprincomp(X)
%[I1 coeff]=IMPRINCOMP(X) performs Principal Component
Analysis
%on X and returns score image (I1) and the corresponding
loadings (coeff)
%


%convert from unsigned 8-bit integers to double class and
remove missing
%values
X=im2double(X);

%check the size of the image to get the number of variables
and objects
[row,col,n]=size(X);

%unfold the image to variables and objects
Y=reshape(X,row*col,n);

%perform PCA
[coeff,score,var]=princomp(zscore(Y));
```

```
%reshape score matrix to score image
I1=reshape(score,row,col,n);
```

**Delineating pixels in scatter plot and highlighting them in image space**

```
%Take two decorrilated images and get their ratios
s=mat2gray(s);
PC1=s(:,:,1);
PC2=s(:,:,2);
w=(PC1-PC2)./(PC1+PC2);

%Apply threshold by assigning it a value
threshold=0.1;
q=(w>threshold);

%Make scatter plot and map them in the imape space
h=figure;
p=get(h,'Position');
set(h,'Position',[p(1,1:3),p(3)/2])

subplot(1,2,1)
plot(PC1,PC2,'*r'),set(gca,'XTick',[0 .2 .4 .6 .8 1])
hold on
plot(PC1(q(:)),PC2(q(:)),'b*')
set(gca,'XLim',[0 1],'YLim',[0 1])
axis square
xlabel('PC1')
ylabel('PC2')
subplot(1,2,2)
imshow(q)
set(h,'colormap',[1 0 0 ;0 0 1])
```

**To perform Hierarchical Cluster Analysis**

```
function HCAcluster(X)

[row,col,n]=size(X);
X=reshape(X,row*col,n);
Y=pdist(X,'Euclidean');
Z=linkage(Y,'average');
[H T]=dendrogram(Z);
```

## Neural Model Development

```
net=newff(Inputs,Targets,5);
net.layers{1}.transferFcn='tansig'
net=train(net,Inputs,Outputs);

y=sim(net,Inputs)
```