MALARIA DIAGNOSIS BASED ON A MACHINE LEARNING SYSTEM

Submitted by

MADUAKO CHIDINMA CHIDIEBUBE

260756845

To the

INSTITUTE OF PARASITOLOGY,

McGILL UNIVERSITY, CANADA

In partial fulfillment of the requirement of the award of the

[Master of Science in Parasitology]

JULY 2019

DECLARATION

I *Maduako Chidinma Chidiebube* hereby declare that this project titled *Malaria diagnosis based on a machine learning system* was carried out by me under the supervision of *Professor Timothy Geary*. I also declare that it has not been submitted either in part or in full for any other examination.

CHIDINMA MADUAKO

2019-07-18

DEDICATION

This thesis is dedicated in the loving memory of my Cousin Emeka Ngadi, who was always available to help me with computer related issues during the course of my study before he passed away and to my family and friends who have helped me in one way or the other to see that this research thesis was a success.

CERTIFICATION

I certify that the research project titled *Malaria diagnosis based on a machine learning system* was carried out by *Maduako Chidinma Chidiebube* under my supervision.

Timothy Geary (Prof.)

SUPERVISOR

Petra Rohrbach (Prof)

CO-SUPERVISOR

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my Supervisor Prof. Timothy Geary and my Co – supervisor Prof. Petra Rohrbach for their motivation, immense knowledge and continuous support throughout my MSc study. I would like to thank Prof. Michael Ngadi who is a member of my Committee for his comments and encouragement which helped me widen my research study. I would not forget to express my gratitude to Dr. Laura Liu who helped me in the formulation of the algorithm used in this study. My appreciation also goes to my fellow lab. mates for their support and team spirit.

Finally, I must express my profound gratitude to my family: my parents, brothers, cousins and relatives for supporting me spiritually throughout this thesis and life generally.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

WHO - World Health Organization

RBC – Red Blood Cells

ECM – Effective Case Management

FOV – Field of View

RITM – Research Institute of Tropical Medicine

MATLAB – Matrix Laboratory

RGB – Red Green Blue

SVM – Support Vector Machine

CDC – Center for Disease Control

REB – Research Ethics Board

RDT – Rapid Diagnostic Test

HRP-11 – Histidine-rich protein 11

LDH – Lactate dehydrogenase

IFA – Immunofluorescence antibody testing

PCR – Polymerase Chain Reaction

LAMP – Loop mediated Isothermal

RNA – Ribonucleic acid

WBCs – White Blood Cells

RITM – Research Institute of Tropical Medicine

SMRU – Shokla Malaria Research Unit

PPV – Predictive Positive Value

EDTA – Ethylenediamine tetra acetic acid

SUMMARY

Introduction: The latest World Malaria Report released in November 2017 estimated that 219 million cases of malaria occurred and deaths due to malaria reached 435,000 in 2017(1). The WHO considers microscopy to be the gold standard for clinical diagnosis of malaria due to its ready availability. However, microscopy has many shortcomings, including inter-user variability and inconsistency, due to the fact that many microscopy technicians do not assess the standard number of high-power fields, are not adequately trained on recognizing all forms of malaria and the high disparity associated with the quality of manual Giemsa slide production (4) To remedy the mis-use of empiric (symptom-guided) treatment, malaria testing is required by many governmental health organizations before commencing antimalarial drug therapy, thereby resulting in increased demand for up to 500 million malaria tests in 2012 (10). Understanding the diagnostic expertise necessary and representing it by specifically tailored image processing, analysis and pattern recognition algorithms can help in designing an automated diagnosis system. Although it is not yet a widespread research topic, automated diagnosis of malaria directly addresses several current gaps (11). My research aims to develop a machine learning system that can identify the stage and number of *Plasmodium falciparum* in cultured erythrocytes based on their morphology using thin film slides, the second objective is to develop a machine learning system that can identify the number of ring-stage parasites in samples of cultured erythrocytes diluted with fresh whole blood.

Methods: Giemsa stained thin blood smears were made from synchronized cultures of the 3D7 strain of *plasmodium falciparum* stored in an incubator with shaking at 37ºC, 5% CO_2, 3 % O_2, 92 % N_2. Thin blood smears were viewed with an EVOS microscope and digital images were acquired, saved as tiff format and stored in a memory stick. The images were transferred as files to a computer, then the images further pre-processed, segmented, and the parasites and the stages of the life cycle detected. The algorithm formulated with the MATLAB programme was trained using 109 images. For the first objective, 397 images were used and for the second objective 163 images were used.

The Otsu algorithm was used for this study, gray level images were reduced to binary images. The algorithm assumes that the images contain foreground and background pixels.

Results: This study showed a relatively strong, positive linear association/ correlation between automated count and manual counts. The correlation between the manual count and the automated count was 0.85. The Pearson correlation between the automated and manual count was 0.7. The diagnostic tool showed a sensitivity of 94.6% for rings, 96.5% for trophozoites and 98.2% for schizont. Moreso, it showed a specificity of 96.5% for rings, 88.9% for trophozoites and 81.8% for schizonts. The R and G channels of the RGB color scheme had clear features which were used to identify objects containing chromatin in Giemsa-stained blood films. The input images transformed to grayscale highlighted parasites containing chromatin.

Conclusion: This study developed an automated system that could enhance the diagnosis and therefore treatment of malaria. The automated method detected more trophozoites and schizonts than the ring stage parasites as seen with a correlation value of 0.83, 0.86 and 0.94 for the ring, trophozoite and ring stages respectively.

CHAPTER ONE

INTRODUCTION

According to the latest World Malaria Report (November 2017), there were 216 million cases of malaria in 2016,an increase from 211 million cases in 2015.In 2016,the estimated number of malaria deaths was 445,000, a slight decrease from the previous year (446,000) (1).

The WHO African Region consistently carries a disproportionately high proportion of the global malaria burden. In 2016, the region had up to 90% of malaria cases and 91% of malaria deaths. In sub-Saharan Africa, about 15 countries accounted for 80% of the global malaria burden. In 2016, the total funding for malaria control and elimination reached an estimated US$ 2.7 billion. 31% of this funding was contributed by governments of endemic countries (2).

In 2016, about half of the world's population was at risk of malaria. As noted, most malaria cases and deaths occur in sub-Saharan Africa. However, the WHO regions of South-East Asia, Eastern Mediterranean, Western Pacific, and the Americas are also at risk. In 2016, malaria was transmitted in 91 countries.

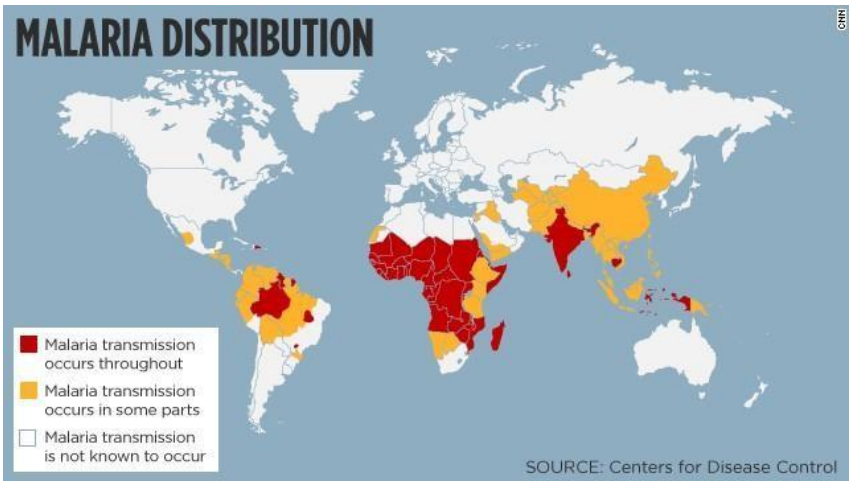


Figure 1: The world distribution of malaria. Source: Center for Disease Control. Malaria 2017 (Available from: https://www.cdc.gov/dpdx/malaria/about/distribution.html.Accessed 19th June 2019[3]

Malaria is caused by protozoan parasites in the genus *Plasmodium*, which are spread to people through the bites of infected female *Anopheles* mosquitoes, called "malaria vectors." Malaria in humans is caused by 5 parasite species – *P. falciparum, P. vivax, P. malariae, P. ovale, and P. knowlesi*. Two of these species – *P.falciparum* and *P.vivax* – pose the greatest threat (2).

In sub-Saharan Africa, *P. falciparum* is the most prevalent malaria parasite, accounting for 99% of malaria cases in 2016. In the WHO Region of the Americas, *P. vivax* is the major malaria parasite, representing 64% of cases. This species also causes more than 30% of malaria cases in the WHO Southeast Asia and 40% in the Eastern Mediterranean regions. This is due to the fact that the environmental conditions are suitable for vector mosquitoes, in addition to the poor socio-economic conditions which make access to health care and disease prevention resources difficult (2).

Plasmodium falciparum has the ability to undergo almost unlimited replication in the human host, unlike other human malaria species. A very high parasitaemia (>50% of erythrocytes infected) is possible. A useful indicator of the intensity of infection is the assessment of the parasite burden, especially in non-immune patients. The level of parasitaemia corresponds generally with clinical features and prognosis. Parasitaemia >4% -5% (>100 000 parasites/µl of blood) is commonly regarded as an indicator of risk of severe malaria in a low-transmission setting, although there is no uniformly agreed definition of hyperparasitaemia (4).

The WHO considers microscopy to be the gold standard for clinical diagnosis of malaria due to its ready availability. However, microscopy has many shortcomings, including inter-user variability and inconsistency, due to the fact that many microscopy technicians do not assess the standard number of high-power fields, are not adequately trained on recognizing all forms of malaria and the high disparity associated with the quality of manual Giemsa-stained slide production. Moreover, there may be no trained microscopy experts available in many laboratories and health care settings to conduct and implement quality assurance (5).

NATURE OF THE PROBLEM

Malaria is a major cause of death in many countries, in part due to the fact that delay in initiating treatment can result from delay in diagnosis (6). The recommended and most basic method and

gold standard for testing blood samples for malaria detection used around the world is the examination of Giemsa-stained thick and thin blood smears through brightfield microscopy. This method can detect as few as 5–10 parasites in 1 μl of blood. However, it is time consuming, needs skilled and well-trained laboratory scientists and is an unautomated protocol (7). Compared to thin smears, thick smears harbor greater parasite numbers, examine lysed red blood cells (RBCs), and employ larger volumes. To evaluate parasite morphology and identify the species present in a thin smear, white light microscopic imaging with a higher magnification and resolution is utilized (8). Furthermore, there can be significant variations in diagnostic accuracy due to subjective measures, based on the training and equipment utilized by the microscopy technician (9).

RATIONALE AND JUSTIFICATION FOR THE STUDY

Malaria infection can be assessed through many types of diagnostic tests (10), but a test that is easy to use, has improved accuracy, and speedily produces results would be of immense benefit to both patients and clinicians. Deficiencies in the quality of test parameters (sensitivity and specificity) cause misdiagnosis and under-treatment of malaria. Parasite drug resistance could arise at least in part as a result of limited availability of convenient and accurate testing methods, leading to increased prevalence of the disease and over-treatment. To reduce the mis-use of empiric (symptom-guided) treatment, malaria testing is required by many governmental health organizations before commencing antimalarial drug therapy, thereby resulting in the demand for up to 500 million malaria tests in 2012. (10)

Understanding the diagnostic expertise necessary and representing it by specifically tailored image processing, analysis and pattern recognition algorithms can help in designing an automated diagnosis system for malaria. Although it is not yet a widespread research topic, automated diagnosis of malaria directly addresses a number of current gaps in practice (11).

There are several reasons why the estimate of parasite burden in *P. falciparum* infections is important: as an indicator of risk of severe and complicated disease, as a measure of response to treatment especially in non-immune patients; as an aid to clinical decision-making in highly endemic areas about the likely cause of febrile illness; and as an end-point in clinical trials of antimalarial drugs or vaccines, at a pre-determined parasite density threshold (12).

To reduce malaria morbidity and mortality in sub-Saharan Africa, Effective Case Management (ECM) remains an important tool. ECM depends on early identification of signs and symptoms, which are interpreted as a malaria episode based on the clinical skills of a peripheral health care worker, as there are often inadequate resources for laboratory diagnosis in many malaria-endemic areas. In many endemic areas, perceived fever is the sign most health workers use to diagnose clinical malaria. However, studies in areas of intense transmission have discovered that reported fever or a history of fever is an unreliable indicator of clinical malaria (13). Nevertheless, in areas of low or unstable transmission, clinical signs and symptoms might be more useful in diagnosing malaria in populations with low immunity. In Asia, studies conducted in areas of low endemicity also found that none of the reported signs or symptoms was a good predictor of malaria (13); how symptoms are perceived is culturally determined and, therefore, the validity of potential diagnostic symptoms must be tested in different socio-cultural and epidemiological settings. Reliance on presumptive clinical diagnosis, in the absence of laboratory diagnosis, results in diagnostic inaccuracy, and over-diagnosis of malaria is common. In areas of lower malaria transmission, it has been observed that there are higher rates of over diagnosis (13). Improper treatment, wastage of drugs and resistance to the few available drugs have contributed to over diagnosis of malaria (14).

Many current research efforts in automating malaria diagnosis use either pixel-based parasite detection based on morphological processing of segmented (schizont-stage) parasites, or extraction of image features from segmented cells. These tests can be improved by machine learning processes.

OBJECTIVES

1. Develop a machine learning system that can identify the stage and number of *Plasmodium falciparum* in cultured erythrocytes based on their morphology using thin film slides.

2. Develop a machine learning system that can identify the number of ring-stage parasites in samples of cultured erythrocytes diluted with fresh whole blood.

HYPOTHESIS

Erythrocytes infected with *P. falciparum* can be accurately identified and quantified using a machine learning system

RESEARCH QUESTIONS

1. How can the life cycle stage of erythrocytic *Plasmodium falciparum* be identified based on parasite morphology using thin film slides analyzed by a machine learning system?

2. Is it possible to develop a machine learning system that can quantify ring-stage *P. falciparum* parasites in thin films made from whole blood?

CHAPTER TWO

LITERATURE REVIEW

Life Cycle of *Plasmodium falciparum*

Early detection of malaria infection leads to prevention and cure by providing more effective treatment and case management. Red blood cells harbor the pathogenic stage of malaria parasites. There are five main life stages of malaria parasites in human blood: rings, trophozoites, schizonts which contain the merozoites and gametocytes. These stages are visible under the light microscope in cultures of *P. falciparum*, but in peripheral blood smears made from infected patients, ring stages (sometimes referred to as early trophozoite stages) are the only stage present (15). During clinical diagnosis of *P. falciparum*, gametocytes may be present, but mature trophozoites and schizonts are rarely seen because they remain sequestered in capillaries and bone marrow (16).

The life cycle of malaria involves two hosts. A malaria-infected female *Anopheles* mosquito injects sporozoites into the human host during a blood meal. These sporozoites infect liver cells, where they mature into exo-erythrocytic schizonts which release merozoites as they rupture. In *P. vivax* and *P. ovale* infections, a dormant stage [hypnozoites] can persist in the liver and cause relapses by invading the bloodstream weeks, or even years after the initial infection. In the liver, after the parasite undergoes initial replication (exo-erythrocytic schizogony), merozoites that enter the blood stream undergo asexual reproduction in erythrocytes (erythrocytic schizogony). Ring stages mature into schizonts which rupture, releasing merozoites that infect new red blood cells and produce new schizonts. Some ring stages differentiate into the sexual erythrocytic stages known as gametocytes. During a blood meal, male (microgametocytes) and female gametocytes (macrogametocytes) are ingested by an *Anopheles* mosquito. The parasites mate and replicate in the mosquito in the sporogonic cycle. Microgametes penetrate macrogametes to generate zygotes in the mosquito stomach. Zygotes in turn become motile and elongated (ookinetes) and then enter the midgut wall of the mosquito where they become oocysts. Oocysts increase in size and eventually rupture, releasing sporozoites which move to the salivary glands. The life cycle is perpetuated when the sporozoites are injected into a new human host (17).

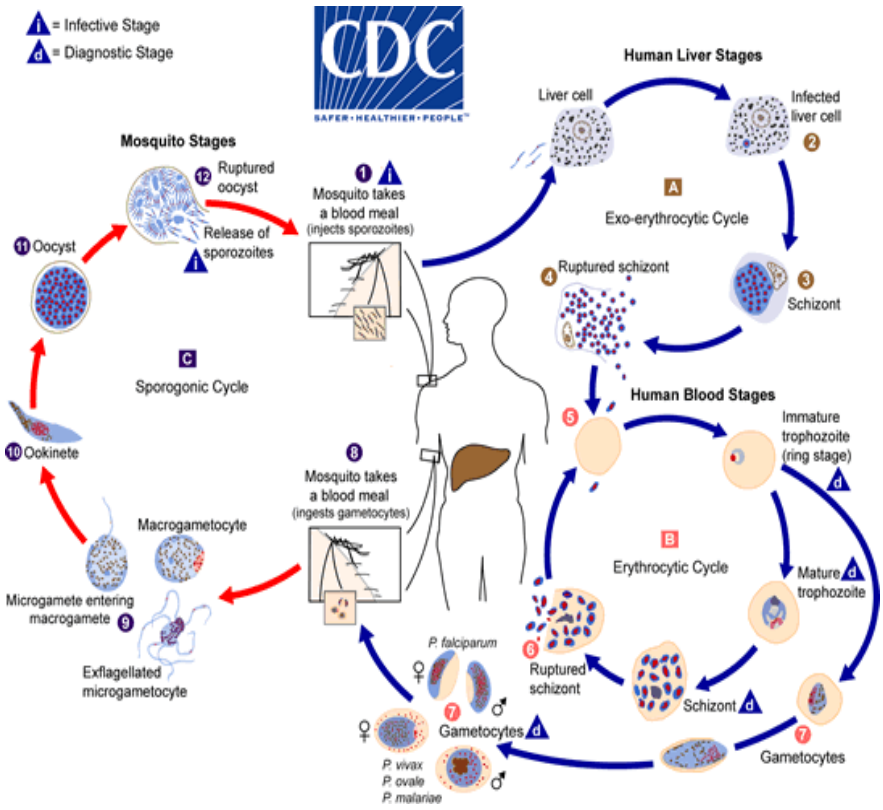


Figure 2: The life cycle of malaria parasites. Source: Downloaded from the Centers for Disease Control https://www.cdc.gov/dpdx/malaria/index.html. Accessed 19th June 2019[17]

MALARIA DIAGNOSIS

A brief review of malaria diagnostic methods

One of the leading causes of death in malaria endemic countries is delay in diagnosis and treatment of malaria. Malaria diagnosis involves identifying malaria parasites or malaria antigens or nucleic acids in patients' blood. Importantly, the combination of clinical and parasite based findings can greatly enhance the accuracy of malaria diagnosis (18).

Laveran originally discovered the malaria parasite and improvement of staining techniques for analyzing these parasites in blood was done by Romanowsky in the late 1800s (19). More than a century later, microscopic detection and identification of *Plasmodium* species in Giemsa-stained thin blood films remains the gold standard for laboratory diagnosis (19).

It is difficult to maintain required levels of malaria diagnostic expertise especially in countries where the disease is not endemic. Microscopy is time - and labor demanding and is challenging for species determination at low parasite density (20).

The general acceptance of this method by most laboratories in the field all over the world is due to its simplicity, low cost to patients, ability to identify the presence and species of parasites and assess parasite density, and species (21).

Community transmission of malaria infection can be decreased by rapid and effective deployment of malaria diagnostic procedures, which also alleviates suffering and prevents serious illness. Over-treatment of malaria or non- treatment of other diseases in malaria- endemic areas and misdiagnosis in non-endemic areas could be due to the nonspecific nature of the clinical signs and symptoms (22).

Effective management of malaria can be achieved through prompt and accurate diagnosis. The global impact of malaria has generated increased interest in developing effective diagnostic strategies not only for developing countries with limited resources where malaria is a substantial burden on the society, but also in developed countries where practical expertise in malaria diagnosis is often lacking (23).

Rapid diagnostic test (RDT)

The urgent need for new, simple, quick, accurate and cost-effective diagnostic tests has been recognized by the World Health Organization (WHO). Numerous new techniques have been developed to overcome the deficiencies of light microscopy. Most notable is the increased use of Rapid Diagnostic Tests RDTs for malaria diagnosis which are fast and easy to perform and do not require electricity. RDTs detect malaria antigens in blood flowing along a membrane containing specific anti- malarial antibodies in a simple design that does not need sophisticated laboratory equipment. Most RDTs target a *P. falciparum*-specific protein, e.g. histidine-rich protein II (HRP-II) or lactate dehydrogenase (LDH). Some RDTs detect *P. falciparum* specific and pan-specific antigens and can distinguish non-*P. falciparum* infections from mixed malaria infections; some claim to effectively and rapidly diagnose *P. vivax* malaria (24).

RDT performance for diagnosis of malaria has been reported to be excellent. RDTs create an opportunity for the benefits of parasite based diagnosis of malaria to be extended beyond the confines of light microscopy with significant advantages in the management of febrile illness in remote malaria endemic areas (25).

For healthcare workers, RDTs appear to be highly valuable, but must be used in conjunction with other methods to confirm the results, characterize infection and monitor treatment. Improvements are needed for ease of use in malaria- endemic areas where light microscopy equipment may be available and where sensitivity for non-falciparum malaria infection, stability, and affordability are also required. Guidelines are being developed by the WHO to ensure lot-to-lot quality control for these tests, which is necessary to build community confidence in the diagnostic tool (26).

Immunofluorescence Antibody Testing (IFA)

Recently, immunofluorescence antibody testing (IFA) has been advanced as a reliable serological test for malaria based on the detection of antibodies against asexual blood stage malaria parasites (27).

IFA is beneficial in many ways including for use in; epidemiological surveys, screening prospective blood donors to avoid transfusion- transmitted malaria and occasionally for providing evidence of recent infection in non-immune individuals. This is due to the fact that it is simple, highly sensitive and specific. Its shortcomings include being time consuming and subjective, as well as requiring fluorescence microscopy and trained technicians. Hence, it cannot be automated, and this limits the number of sera that can be studied daily. It also requires sophisticated equipment, reliable power and training of laboratory personnel; readings for serum samples with low antibody titers, can be influenced by the level of training of the technician (28).

Polymerase Chain Reaction (PCR) based technique

For the molecular diagnosis of malaria, PCR-based techniques are a recent development and have been shown to be one of the most specific and sensitive diagnostic methods, especially for malaria cases with low parasitemia or mixed species infections. Malaria diagnosis by PCR has

been confirmed extensively, and the technique has been used to follow-up therapeutic response, and identify drug resistant strains (29).

PCR has higher sensitivity and specificity than conventional microscopy as it can detect as few as 1-5 parasites/µl of blood (≤ 0.0001% of infected red blood cells) compared with around 50-100 parasites/µl of blood by microscopy or RDT. Moreover, PCR can be automated to process large numbers of samples (30).

PCR utilization is limited due to its complex methodologies, high cost and the need for sophisticated equipment, consistent power supply and specially trained technicians. It is not routinely implemented in developing countries due to the lack of resources to carry out these tests adequately and routinely. This technique may not be suitable for routine malaria diagnosis in clinical settings and in remote rural areas because quality control and equipment maintenance are necessary for reliable use (31).

LAMP Technique

The Loop – mediated Isothermal Amplification (LAMP) technique amplifies and detects the conserved 18S ribosomal RNA gene of *P. falciparum* in a single tube, providing a simple and inexpensive molecular malaria diagnostic test.(32) It is highly sensitive and specific, suitable for *P. falciparum*, as well as *P. vivax*, *P. ovale* and *P. malariae* (33).

LAMP can be used in regions where malaria is endemic. It generates results quicker and is lower in cost than PCR. However, the reagents need cold storage and further clinical trials must be done to validate the feasibility and clinical utility of LAMP in the field (34, 35).

Microarrays

Microarrays may play an important role in the diagnosis of infectious diseases in the future (36). Microarrays work on the same principle as traditional Southern hybridization. Hybridization of labeled targets derived from nucleic acids in the test sample to identical sequence probes on the array enables the probing of multiple pathogen gene in a single experiment. *P. falciparum* has been accurately identified in clinical specimens using a pan-microbial oligonucleotide

microarray developed for infectious disease diagnosis (37), although the technical platforms needed for routine use in the field remain to be fully developed and validated.

Flow Cytometry

Flow cytometry has been used for malaria diagnosis. This technique is based on the principle of the detection of hemozoin, which is produced when intra-erythrocytic malaria parasites digest host hemoglobin and crystallize the released toxic heme into hemozoin in the acidic food vacuole.

Depolarization of laser light can be used to detect hemozoin ingested by phagocytes as cells pass through a flow-cytometer channel. This method has been reported to provide sensitivity of 49-98%, and specificity of 82-97%, for malarial diagnosis (38).

However, the need for expensive equipment and advanced training makes it highly unlikely that flow cytometry will find a place for routine or field-based malaria diagnosis.

Mass Spectrometry

Mass spectrometry identifies a specific chemical biomarker in clinical samples. In malaria, heme from hemozoin is a parasite-specific biomarker of interest. This technique can be rapid, analyzing a sample in < 1 min, and can be automated. Remote rural areas without electricity are not favorable for existing mass spectrometers. This method could be made more practicable with improvements in equipment and techniques (39).

Light Microscopy

The gold standard and traditional method for malaria diagnosis is light microscopy which is time consuming and can be challenging to perform in some field situations. It requires the expertise of a skilled and trained laboratory technician. In particular, poor people living in rural endemic areas who lack access to health care are at greater risk for the disease (40).

In areas where malaria is not endemic, microscopic diagnosis can be difficult for healthcare providers who are unfamiliar with the disease. Clinicians may not remember to consider malaria

among the differential diagnoses. Technicians may also fail to detect the parasite when examining a blood smear under the microscope, because they lack experience with the disease. In endemic areas, malaria transmission can be so intense that a large proportion of the population may be infected but remain asymptomatic. In endemic countries, a major barrier to reliable and timely diagnosis is the lack of resources. Health personnel are too often not adequately trained, equipped or paid. They are too often overwhelmed by the work load and have to divide their attention between malaria and other severe infectious diseases such as tuberculosis or HIV/AIDS (41).

Identifying the presence of parasites is the key to malaria diagnosis and recognising the abundance of infected erythrocytes is important for predicting the severity of the disease. Determining the level of parasitaemia also allows monitoring patients by measuring drug efficacy and potential drug resistance (42).

The most common method for malaria diagnosis in health facilities with laboratories is the use of simple light microscopy examination of Giemsa-stained blood films. The advantages of microscopy include differentiation between species, quantification of parasite density and the ability to distinguish clinically important asexual parasite stages from gametocytes-the sexual stage of the parasite which could be asymptomatic (43). Standard laboratory microscopy has a sensitivity of approximately 90%, a figure which drops dramatically in the field. The accuracy of microscopy performed by laboratory scientists was low and only 51% of parasite quantification results were considered acceptable in a recent study (44).

Thus, routine microscopy may be of low quality as a result of poor training, inadequate skill of laboratory staff, poor infrastructure, inadequate reagents and equipment, and unreliable electricity in many regions of Sub-Saharan Africa (45). These challenges may result in over diagnosis of malaria by laboratory technicians, perhaps as a result of previous WHO recommendations based on the assumption that instead of missing a true case of malaria, it is safer to treat non-malarial patients with antimalarial drugs than to not treat a true case of malaria. Although in places with limited resources and funding, such as Sub-Saharan Africa, microscopy can give accurate and exact malaria diagnosis, its use can be linked with over-diagnosis of malaria (46).

Physical features or morphology as well as the appearance of red blood cells that have been invaded are used to recognise the presence of the malaria parasite [7]. The ability to detect the parasite visually requires the use of Giemsa stain. This process slightly decolourises the RBCs, but highlights the parasites, white blood cells (WBCs), platelets and debris [7].

Therefore, the diagnostic process requires an ability to distinguish between malaria parasites in infected RBCs from non-parasite stained components (normal RBCs, WBCs, platelets and artefacts) using visual information. However, visual recognition is time consuming and can be challenging, especially in cases where large number of samples require reliable analysis (47).

PARAITAEMIA ESTIMATION

Plasmodium falciparum, unlike the other human malaria species, has the capacity for nearly unlimited multiplication in the human host, and very high parasitaemia (>50% of erythrocytes infected) is possible in *falciparum* infections. Assessing the parasite burden gives useful information on the severity of infection, especially in non-immune patients, and the level of parasitaemia (parasite burden or load) corresponds generally with clinical features and prognosis. Therefore, 4 or 5% (or higher) parasitaemia, or > 100 000 parasites/µl, is generally considered an indicator of risk of severe malaria in a low-transmission setting (48). Parasite load estimation is also an objective measure of response to treatment and an aid to clinical decision-making about the likely cause of febrile illness in highly endemic areas (49).

MORPHOLOGICAL FEATURES

Ring Stage

- Rings are the young trophozoite stages found inside RBCs. The name ring is derived from the physical appearance of the stage, which has a ring-like structure in blood films.
- It consists of a central vacuole and nucleus present at the center of the cytoplasm. Often two or more ring forms of the parasite are found inside a single RBC. This is not the case in patient's blood; mainly found in cultured samples.

- In stained smears, ring shaped cytoplasm surrounds a central blue colored vacuole with red colored nucleus (50).

Trophozoite Stage

- Trophozoites are vacuolated and more or less irregular (amoeboid) shaped with one nucleus.
- They are small, delicate and measure 1.25 - 1.5 µm in size
- In stained preparations, they show a thin ring of blue cytoplasm and darkish stained nucleus.
- A single large mass of yellow- to black-colored pigment crystals called hemozoin is present.

Schizont Stage

- They are small, immobile, asexual and dividing forms of the parasite.
- They measure 4.5-5 µm in diameter and occupy about 2/3rd of the infected RBC.
- Each schizont contains 8-24 merozoites and an aggregate of dark stained pigment crystals.
- They have a circular shape with more than two cytoplasm dots and an even number of chromatin dots (51).

Gametocyte Stage

- Gametocytes are sexual erythrocytic stages and are infectious to mosquitoes.
- They are typically crescent (banana) shaped with round or pointed ends.
- Mature gametocytes are 1-1.5 times larger than RBCs.
- There are two types of gametocytes:
    - Microgamete: male form
    - Macrogamete: female form

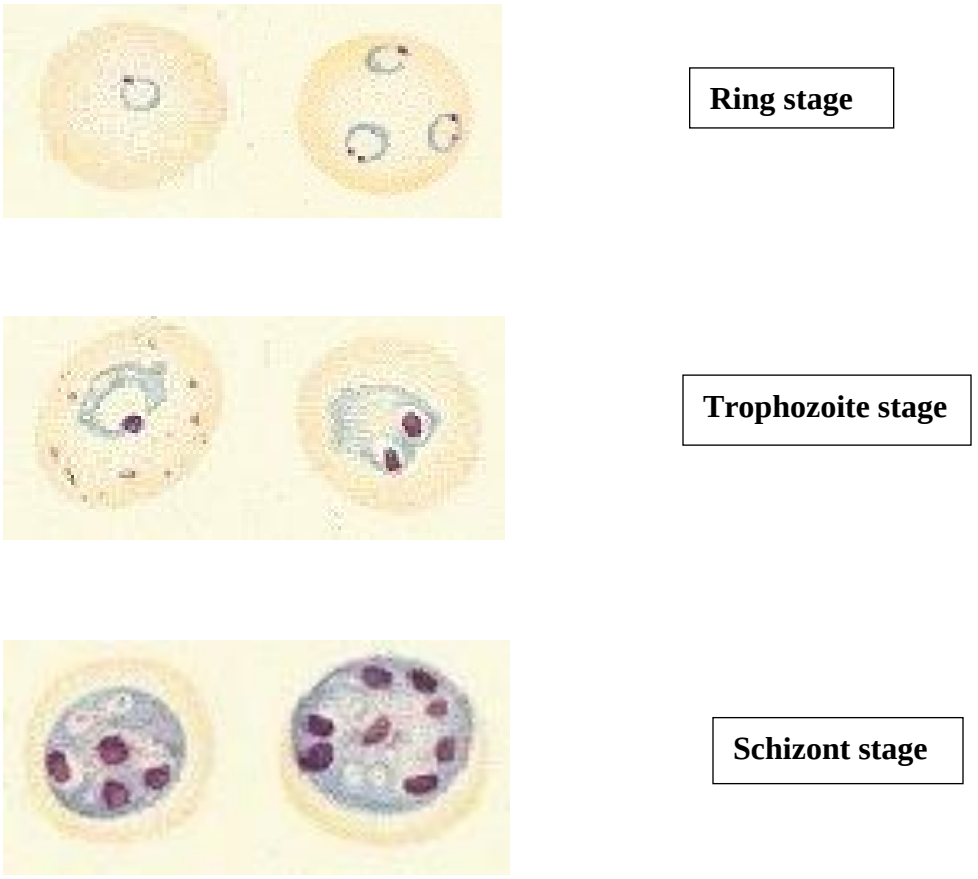


Figure 3: Image showing the different morphological features associated with the stages of *plasmodium falciparum.* Source: (Available from : https://www.onlinebiologynotes.com/plasmodium-falciparum-morphology-life-cycle-pathogenesis-and-clinical-disease/.Accessed 19th June 2019[50]

MISDIAGNOSIS OF MALARIA

Laboratory Misdiagnosis

Rapid and effective malaria diagnosis alleviates suffering, but also decreases community transmission. It may be difficult to retain personnel with the laboratory expertise required for microscopic diagnosis, taking into consideration the large number of laboratory staff providing emergency services. In countries with imported malaria, each institution may see only a few cases each year. Two factors can cause misdiagnosis: unfamiliarity with reading malaria smears, especially for on-call staff who might not be able to readily consult more experienced staff out of routine hours. The second, which occurs as a result of inexperience, is difficulty in diagnosing cases with low-level parasitemia, which may represent a substantial proportion. However, as soon as the alert clinician has considered the possibility of malaria, good medical training has taught him or her that a single negative smear does not exclude the possibility of malaria, and it is good practice to ask the patient to return for repeat smears if symptoms persist and suspicion is high (52).

Clinical Misdiagnosis

Clinical diagnosis of malaria is least expensive and practised generally among medical doctors. This is based on the patient's signs and symptoms, as well as physical findings at examination. The initial symptoms of malaria are nonspecific and variable. Moreover, due to the non-specific nature of the signs and symptoms of malaria, which overlap with other common, as well as life threatening diseases, this makes a clinical diagnosis challenging. This can promote the indiscriminate use of antimalaria drugs and compromise the quality of care for patients with non-malaria fever (53).

An on-call physician who is inexperienced in travel medicine or is not clinically alert may see a febrile patient. The absence of clinical suspicion may have more serious consequences. Although the reason for this may remain elusive, looking at the cases in retrospect, the patients have frequently provided the necessary details that should point to the possibility of malaria, warranting a blood film examination (52).

MACHINE LEARNING AND AUTOMATED DIAGNOSIS OF MALARIA

Machine learning, a branch of Artificial Intelligence (AI) is the science and technology that has the main objective of developing computers that can think, see, perceive, hear, talk and feel. It involves the development of a machine (computer system), with its functions related to human intelligence, including reasoning, inference, hearing and problem solving. In machine learning, information processors perform tasks of sorting, assembling, assimilating, and classifying information. In supervised learning, the machine infers a function from a set of training examples. In unsupervised learning, the machine tries to find hidden structure in unlabeled data (54).

There are many known and potential applications for machine learning, the most significant of which is data mining. Machine learning can often be successfully applied to these problems, improving the efficiency of systems and the designs of machines. Every instance in any dataset used by machine learning algorithms is represented using the same set of features. If instances are given with known labels, then the learning is called supervised, in contrast to unsupervised learning in which instances are unlabeled. Applying these unsupervised (clustering) algorithms could lead to useful classification of items (55).

There are 3 broad types of machine learning algorithms

1.Supervised Learning

This algorithm consists of dependent or outcome variable which is predicted from a given set of independent variables. The algorithm is trained using input variables, when fully trained, the algorithm will be able to identify new data and predict a new label for it. Examples include: Linear Regression, Decision tree, Naïve Bayes, Nearest neighbor, Support Vector Machines (SVM), Neural Networks, Random forest, KNN, Logistic regression, and Gradient Boosted Trees.

2.Unsupervised Learning

In this algorithm, there are no target or outcome variables to predict or estimate. Clustering population are segmented in different groups based on their similarities. Examples include: Aporiori algorithm, K-means clustering, Association Rule learning algorithm, t-SNE (t-Distributed Stochastic Neighbor Embedding), and PCA (Principal Component Analysis).

3.Reinforcement Learning

The algorithm trains itself to make specific decisions, based on continuous use of trial and error. The algorithm learns from experience and tries to capture the best possible knowledge to make accurate decisions. Examples are Markov Decision Process, Q-Learning, Temporal Difference (TD), Monte-Carlo Tree Search (MCTS), Asynchronous Actor-Critic Agents (A3C) and Deep Adversarial Network.

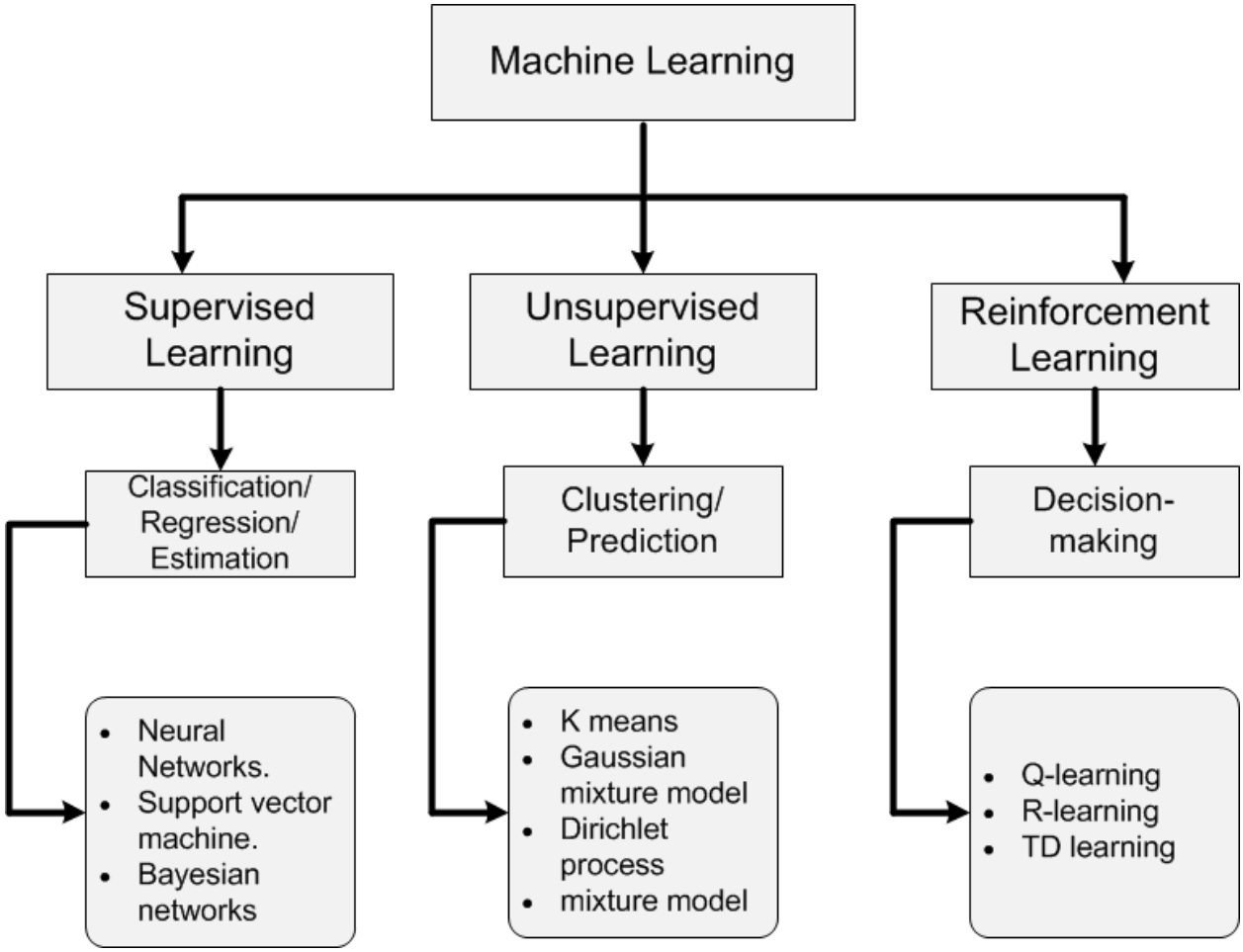


Figure 4: Types of machine learning. SOURCE: (Available from: https://www.researchgate.net/publication/325928183big_data_perspective_and_challenges_in_next_generation_networks)[56]

MATLAB

Matrix Laboratory is a high-performance programming language and interactive environment for technical and numerical computation, visualization and programming (57). It allows for development and implementation of algorithm, data analysis, exploration and visualization. The MATLAB application is designed around the MATLAB scripting language. A common usage of the MATLAB application involves using the Command window as an interactive mathematical shell or executing text files containing MATLAB code. The MATLAB system consists of 5 main parts which are;

The MATLAB language

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs.

The MATLAB working environment

This is the set of tools and facilities that you work with as the MATLAB user or programmer. It includes facilities for managing the variables in your workspace and importing and exporting data. It also includes tools for developing, managing, debugging, and profiling M-files (multi-file), MATLAB's applications.

Handle Graphics

This is the MATLAB graphics system. It includes high-level commands for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level commands that allow you to fully customize the appearance of graphics as well as to build complete Graphical User Interfaces on your MATLAB applications.

The MATLAB mathematical function library

This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.

The MATLAB Application Program Interface (API)

This is a library that allows you to write C and Fortran programs that interact with MATLAB. It includes facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.

RELATED WORK

Several efforts have been made to detect malaria infection using machine learning algorithms by automated analysis of microscopic images of stained red blood cells (58). For example, Delahunt et al. (30) reported a study on automated microscopy and machine learning for expert-level malaria field diagnosis, using an algorithm trained on 27 *P. falciparum*-positive samples, examining 50 fields of view (FOV) with several thousand parasites, and 36 negative samples (50 FOVs each). 90% sample level specificity was ensured by setting the algorithm's diagnosis parameters. The algorithm was then applied to a separate test set, which consisted of the *P. falciparum* component of a WHO55 slide set as well as 3 low-parasitemia samples. This test set included 20 negative samples, 10 RITM (Research Institute of Tropical Medicine) *P. falciparum* samples (parasitemia 100 – 200 p/µL), and 14 RITM and SMRU (Shoklo Malaria Research Unit, Thailand) *P. falciparum* samples (200 – 1200 p/µL). Autoscope diagnosis results showed sensitivity and specificity exceeding the diagnosis requirements for a WHO Level 1 (expert) microscopists (59).

Dhanya Bibin (60) carried out a study on malaria parasite detection from peripheral blood smear images using deep belief networks. A Canon A60 camera attached to a Brunel microscope with a 100X objective using an optical coupler was used to capture images as RGB color images with a 1600×1200-pixel resolution. The experiment was conducted in MATLAB (R2016a) using image processing, neural network and the statistical pattern recognition toolbox. 630 RGB images consisting of non-overlapping fields of 9 peripheral blood smear slides were utilized for the experiment. Hematopathologists with expertise in malaria diagnosis provided the ground truth for the collection of images. 669 and 3431 parasites and non- parasite objects, respectively, were used in this research. The result obtained showed an F-score (the measure of a test accuracy, it is the mean between the precision and recall) of 89.66% with a sensitivity of 97.60% and specificity of 95.92%. (60)

Savkare and Narote (61) used an SVM multi-classifier to demonstrate that methods of feature extraction can produce a very rich group of parameters. The area occupied by early and late trophozoite stage parasite in erythrocytes was 20-50%, increasing in schizont stages to 40-80%, whereas gametocytes can occupy the entire RBC. The SVM binary classifier showed 96.26% sensitivity and 99.09% specificity and identified life stages of malaria parasite through color and

geometrical features which were extracted from infected erythrocytes. SVM kernels such as linear, polynomial and RBF (Radial basis function kernel) were applied to 71 images, SVM multiclass classifier (i.e. RBF kernel) gave 96.42% accuracy for correct identification of life stage of parasite. Accuracy is the function of predictions a model is able to get right The above algorithms were implemented using MATLAB (61).

Purwar et al. (62) developed a method based on digital image processing of Giemsa-stained thin smear images to hasten the diagnostic process. The diagnostic procedure was divided into two parts; enumeration and identification. The image-based method was designed to automate the processes of enumeration and identification, with the main advantage being the ability to carry out the diagnosis in an unsupervised manner and yet have high sensitivity thereby reducing cases of false negatives. The image-based method was tested on over 500 images from two independent laboratories. The aim was to distinguish between positive and negative cases of malaria using thin blood smear slides. This method requires minimal human intervention due to its unsupervised nature, hastening the process of diagnosis. The overall sensitivity to capture cases of malaria was 100% and specificity ranged from 50-88% for all species of malaria parasites (62).

Linder et al. (63) proposed a method using computer vision detection and visualization of only the diagnostically most relevant sample regions in digitized blood smears. Approximately 50,000 erythrocytes per sample were captured using Giemsa-stained thin blood films containing *P. falciparum* ring-stages (n=27) and uninfected controls (n=20), which were digitally scanned with an oil immersion objective (0.1 mm/pixel). Identification of parasite candidate regions was based on color and object size, followed by extraction of image features (local binary patterns, local contrast and scale-invariant feature transform descriptors) used as input to a support vector machine classifier.

The classifier was trained on digital slides from 10 patients and validated on 6 samples. 31 samples (19 infected and 12 controls) were used to test diagnostic accuracy. 128 most probable parasite candidate regions were generated from each digitized area of a blood smear. Two expert microscopists visually inspected the panel on a tablet computer to determine if the patient was infected with *P. falciparum*. The method achieved a diagnostic sensitivity and specificity of 95% and 100% as well as 90% and 100% for the two readers, respectively. The automated system

calculated parasitemia and the correlation coefficient between manual and automated parasitemia counts was 0.97 (63).

Nicholas et al. (64) developed an image processing algorithm to automate the diagnosis of malaria using thin blood smears. A charge-coupled device camera connected to a light microscope was used to acquire the image. To automate the diagnosis of malaria, an image processing algorithm was developed. The image classification system is designed to positively identify malaria parasites present in thin blood smears, and differentiate the species of malaria, with a sensitivity of 92% and a PPV of 40%, potential parasites were segmented. The classifier was able to positively identify malaria parasites with a sensitivity of 85% and a Positive predictive Value (PPV) of 81%. Relating these results to examination of a full sample reveals that high sensitivity could be achieved by examining a large number of slides. A classifier accuracy of 73% was determined for 11 out of 15 samples. To positively identify malaria parasites, present in thin blood smears and differentiate the species of malaria, an image classification system was designed.(64)

In conclusion, factors that determine the choice of malaria diagnostic method include; the urgency of diagnosis, the experience of the physician, effectiveness of the healthcare workers and the level of malaria endemicity. Previous work suggests that an automated diagnostic system could significantly improve the efficiency of the microscopist and reduce the need for dedicated microscopists in rural villages.(65)

CHAPTER THREE

MATERIALS AND METHODS

SAMPLE PREPARATION

Red blood cell samples were mixed with samples of cultured *P. falciparum* parasites (3D7 strain);(66) in the laboratory, and was maintained in continuous culture in human erythrocytes in complete media (RPMI 1640 supplemented with 0.5% AlbuMAX® as serum substitute,0.1mM hypoxanthine,25ug/mL gentamicin) at 37 °C under controlled atmospheric conditions of 5 % CO_2, 3 % O_2, and 92 % N_2 in an incubator.(66) *P. falciparum* was cultured in 10 cm petri dishes and parasitemia was maintained at 5%. Parasite cultures were diluted every 48 hr to a parasitemia of 0.4% using prewarmed complete media and uninfected erythrocytes. To make thin blood smears,5 µl of the suspension was transferred to a microscope slide. A spreader slide held at a 30 - 45° angle was used to push the suspension forward. The dried smear was dipped in 100 % methanol to fix the cells to the slide. The slide was then incubated in Giemsa solution for 15 min, after which it was rinsed and air dried.(67) The slide was analyzed under a microscope at 1000 x magnification in oil immersion. The culture was synchronized when it had more ring stages and high parasitemia. An interval of 6 hr was observed between two consecutive synchronization steps: 10 ml culture was transferred to a 15 ml tube which was centrifuged at 2000 rpmg, acceleration 9, deceleration 7 (Thermo scientific Sorvall ST 16R centrifuge), after which the supernatant was removed. The pellet was resuspended in 10 ml 5% D-sorbitol in a 15 ml Falcon tube, which was incubated for 5 min at 37 °C in a water bath and then centrifuged for 2 min at 800 xg. The supernatant was removed, and the cells washed in 20 ml RPMI incomplete media. RPMI incomplete media is not supplemented with hypoxanthine, albumax and antibiotic (Gentamycin mixture). Cells were resuspended in 20 ml RPMI complete medium and put back into the incubator with shaking. The shaker ensures that the parasites developed synchronously.(68)

For the second objective, a whole blood sample was collected by venipuncture into an EDTA vacutainer from a healthy, non-pregnant donor with informed written consent (REB File # 202 – 0918).

The fresh blood sample (5 ml) was mixed with cultured red blood cells infected with ring stages of *P. falciparum* strain 3D7, obtained from on-going synchronous cultures of malaria parasites in the laboratory. Thin blood smears were prepared as described for the first objective. (68)

MALARIA DIAGNOSIS

For clinical diagnosis of malaria, a drop of the patient's blood is applied to a glass slide, which is then immersed in a staining solution to make parasites more easily visible under a conventional light microscope, usually with a 100× oil objective. For this study, thin blood smears were made with cultured 3D7 *P. falciparum* ring stages, and from whole fresh blood samples mixed with the cultured 3D7 *P.* ring stages for the first and second objectives, respectively.

GIEMSA STAINING METHOD

The staining process colorizes the RBCs and highlight the parasites, WBCs and platelets. To detect malaria parasites by AI (Artificial Intelligence) requires the detection of the stained objects. Moreover, the stained objects must be analyzed further to determine if they are parasites or not. This is to prevent false diagnosis. As discussed above, many methods can be used to detect malaria infection, including image segmentation and feature extraction. (69)

Thin film slides were dipped in methanol as noted and then air dried for about 2 min, then placed in a 10% Giemsa working solution for about 10 min. Slides were removed from the Giemsa stain and rinsed under running buffered water at about 7.2 pH to float off the iridescent scum on the surface of the stain. The slides were placed in a rack to dry before analysis. The images were viewed and evaluated by three experienced microscopists and the ground rule was set by them. (70)

In this study, supervised learning method was used. OTSU method was used to generate data for training of deep learning image segmentation model. The algorithm was trained using 109 images. The complexity rate of OTSU thresholding algorithm is very high and processing rate is very slow. The segmentation result of OTSU segmentation algorithm is good, involves simple calculation and stability. (71)

DIGITAL IMAGE ANALYSIS

Image analysis is the process of extracting meaningful information from images, such as finding shapes, counting objects, identifying colors, or measuring object properties (72).

Computer imaging is important since our main sense is our visual sense – one picture is worth a thousand words. The computer representation of an image equals many thousands of words without a corresponding amount of information, the medium would be prohibitively inefficient. The main reason for the development of many subareas in the field of computer imaging such as image segmentation and compression are the huge amount of data required for image analysis.

To develop an image processing algorithm or system, many images must be examined and tested. The system has to be exposed to a lot of images and trained to distinguish between the classes (classes are formed according to their similarities), you can usually teach it new classes in the same domain.(73)

ILLUSTRATION OF THE MALARIA IMAGE ANALYSIS ALGORITHM

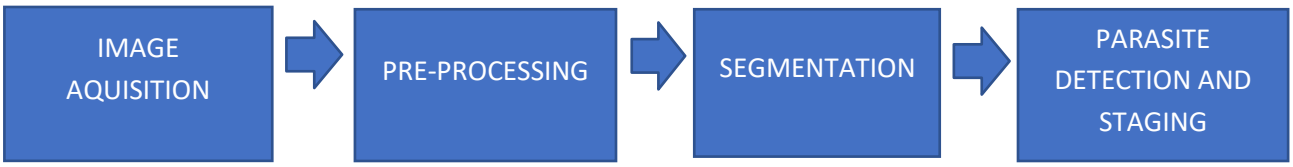


Figure 5: The steps involved in malaria image analysis

IMAGE ACQUISITION

Digital images were the main data sources for this research. The images used in this research were acquired with the use of an EVOS digital microscope, saved in the tiff format. Digital images were stored on a memory stick attached to the microscope and transferred as files to a computer and used for automated detection of the stage and time point of malaria parasite in culture

PRE-PROCESSING

Preprocessing is mainly applied to improve the quality of the image and to reduce variations in the images that would unnecessarily complicate the subsequent processing steps. Three key objectives can be identified: noise removal, contrast improvement, illumination and staining correction. To reduce complexity without compromising the importance of parasite details, the color image was converted to grayscale image. (74)

The primary goal of this step is to make the acquired images more suitable for subsequent processing, mainly image segmentation and feature extraction. The second objective of image pre-processing is to reduce or eliminate noise from the acquired image. The third objective is to enhance image contrast for visual evaluation. Considerable color variability between images can exist. The pre-processing module transforms the scanned color images to generate a consistent appearance. The pre-processing stage also removes unwanted effects such as noise from the image and transforms and adjusts the image as necessary for further processing. Captured images were converted from RGB to gray scale. Grayscale images carry only intensity information, color is too much information and can be very distracting, grayscale improves luminosity and brightness, reduces the image to tone, light and shadow (75).

(A) (B)

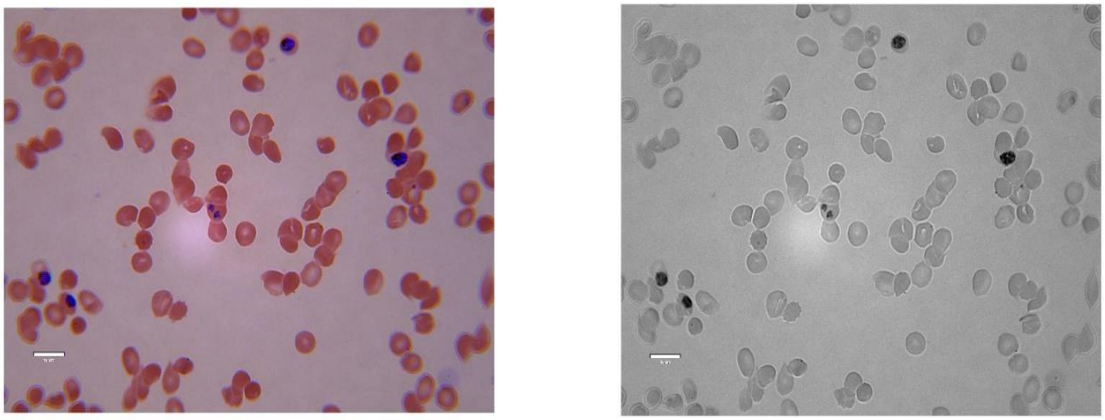


Figure 6: Original image and grayscale images of *P. falciparum.* (A) Original image captured with a light microscope. (B) Grayscale image. Scale bar = 10 µm

SEGMENTATION

Segmentation is the process of dividing or separating an image into regions with respect to objects. Identifying common properties can be used to segment regions. The simplest property that pixels in a region can share is intensity. Therefore, the best method is to segment such regions through thresholding, which is the separation of light and dark regions. Binary images are created from grey – level images through thresholding by turning all pixels below some threshold to zero and all pixels above that to one. The basic aim of this method is grouping pixels into clusters having similar properties. For instance, this algorithm was designed by identifying and dividing the image into regions or segments, that is the background and foreground. For example, an object may be segmented from the background by defining a threshold value, T, such that all pixels having intensity values less than T belong to the background and those pixels values greater than T are foreground. An image histogram is a distribution of gray levels of an intensity image. Often in an image histogram, there are two or more modes, corresponding to different regions of the image. A good threshold value is located near the "valley" bordering two adjacent modes (76).

Image segmentation identifies and segments potential parasites and erythrocytes from the image background. It is necessary to first identify infected erythrocytes from the combination of parasites and erythrocytes in the image and then segment them from the background. For this algorithm, a method to find thresholds based on the image histogram was developed. From the histogram, two threshold levels need to be determined: one for erythrocytes and the other for the parasites (77).

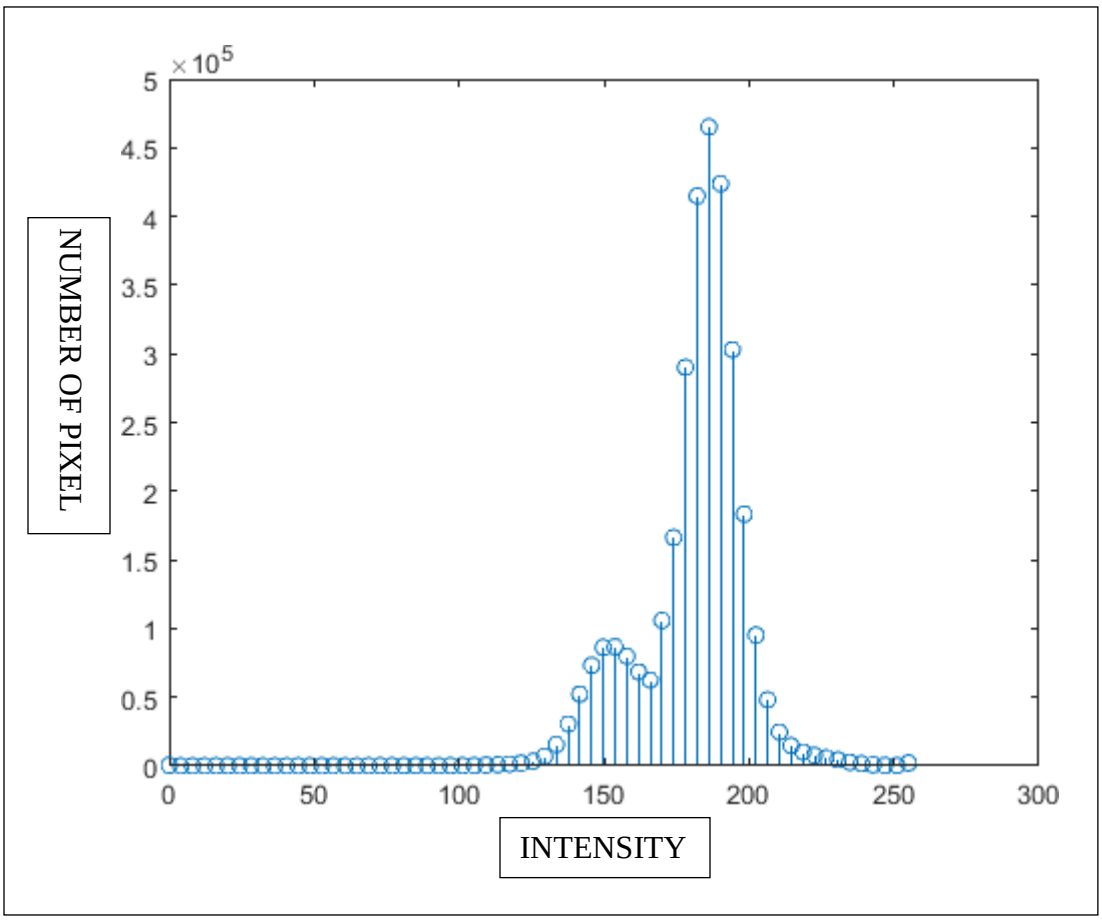


Figure 7: Histogram image of *P. falciparum* in cultured erythrocytes.

The eventual success or failure of an image analysis procedure is determined by the accuracy of segmentation. Due to this fact, additional care is taken to improve the likelihood of accurate segmentation. A considerable portion of the segmentation algorithm is based on two basic properties of intensity values: discontinuity and similarity (76).

The identification and classification of *P. falciparum* depends on the presence of chromatin dots or hemozoin and variations in the area and shape of the cytoplasm. These variations are translated to changes in the distribution of color, intensity, texture and edges in the cell image, and these features were selected for the extraction process. Erythrocyte feature is a random variable described by a histogram. Grayscale histogram represents the intensities distribution on the image, a 256-bit histogram was extracted from gray scale images (78).

FEATURE EXTRACTION/PARASITE IDENTIFICATION AND LABELLING

Simple morphological features such as color, shape and size can be used to identify malaria parasites. The ring stage parasite has a ring shape and one or two chromatin dots; size is determined by finding the width and the length of the parasite. The trophozoite stage has an irregular shape and the size is measured by finding the perimeter that is the total distance between consecutive points of the border. Schizonts are circular shaped, and the size is measured by averaging the length of the radial line. Schizonts contain 8-24 merozoites upon maturation which are arranged like grapes in a cluster. Schizonts are rarely seen in peripheral blood smears obtained from patients infected with *P. falciparum*, and the presence of schizonts in peripheral blood indicates the presence of a severe infection. (79)

The ground truth which is the accuracy of the training set's classification for supervised learning techniques, was leveraged to train the machine learning algorithms by establishing a correspondence between the detected objects and the annotated parasites. A detected object that corresponds with an annotated parasite was used as a positive sample to train the machine learning algorithm. Annotated parasites that are not detected are also used as positive training samples. Detected objects that do not correspond to any annotated parasite are used as negative training samples (distractors). The output of the correspondence algorithm is a set of objects that are labeled as either parasites or distractors.

Area = Number of pixels on the interior of the cell

Pixel value = Scale bar value / Perimeter

Perimeter =The total distance between consecutive points of the border of the parasite

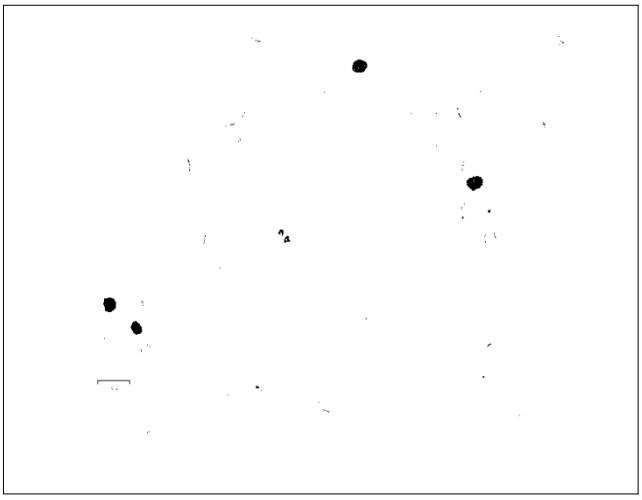


Figure 8: Detected malaria parasites

DATA PROCESSING AND ANALYSIS

The algorithm was developed using the MATLAB R2018a (REF). Data are presented in tables, scatter grams, bar charts and Receiver Operator Curves (ROC).

ETHICAL CLEARANCE

This study was conducted with the approval of the Research Ethics Board (REB), Faculty of Agriculture and Environmental Sciences. It was conducted in accordance with the policies and procedures governing the ethical conduct of research involving participants at McGill University. Participants were students of the Institute of Parasitology, McGill University who voluntarily donated their blood samples for this research. They were non-pregnant, healthy adults, free from malaria infection with no known blood disorder and the participants provided written informed consent to participate in this study. REB File #: 202 – 0918

CHAPTER FOUR

RESULTS

OBJECTIVE 1

Develop a machine learning system that allows identification of the stage and number of *Plasmodium falciparum* parasites in culture based on morphology using thin film slides made from infected red blood cells in culture

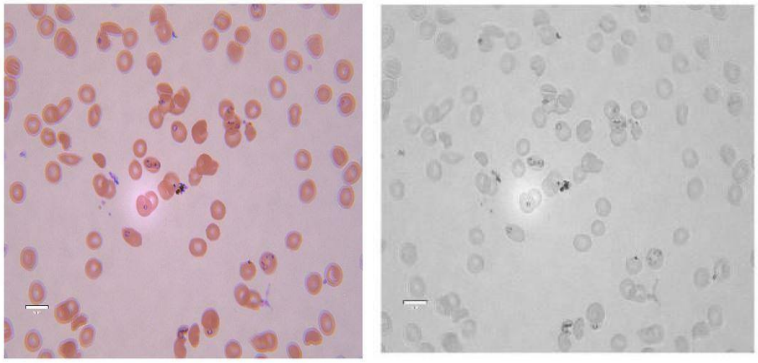
In the course of this research, 5 slides, and 5 fields of views (FOVs) taken randomly at each time point from each slide formed the set of images for analysis.

For the first objective, 203 images were used for sample 1 and 194 images for sample 2. Sample 1 represents 3D7 samples cultured in petri dish 1 and sample 2 represents 3D7 samples cultured in petri dish 2.

Parasite cultures were synchronized as described and representative images of the three life cycle stages (ring, trophozoite and schizont) are shown in Figures 8 – 10.

RING STAGE

(A) **(B)**



(C) **(D)**

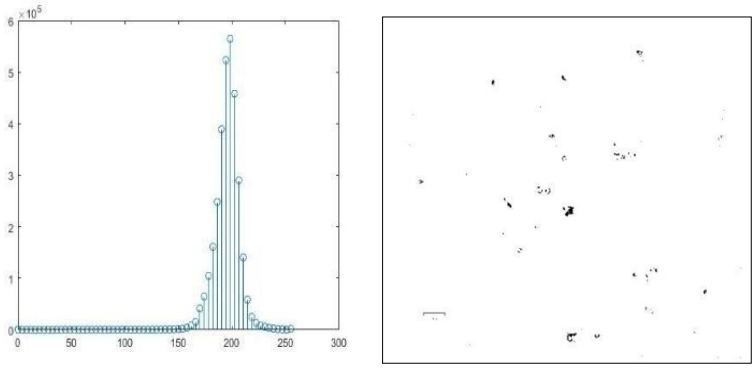
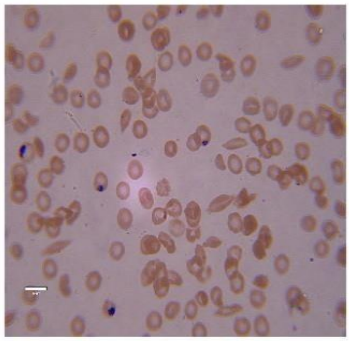
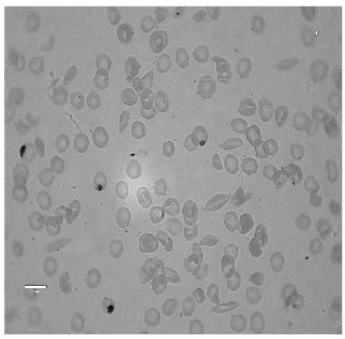


Figure 9: Images of ring stage parasites analyzed with MATLAB. (A) Original image showing the parasite in the red blood cell. (B) Grayscale image showing the parasite in the red blood cell. (C) Histogram with the x-axis showing intensity and y axis showing the number of pixels. (D) Automated image showing the detected malaria parasites

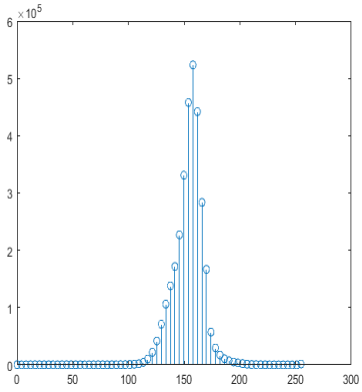
TROPHOZOITE STAGE

(A) **(B)**





(C) **(D)**



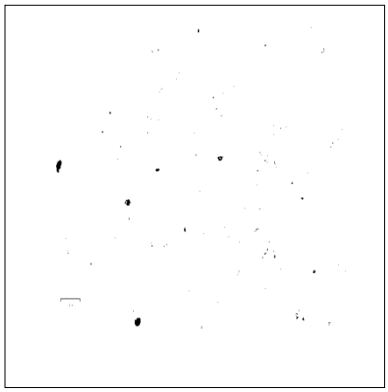
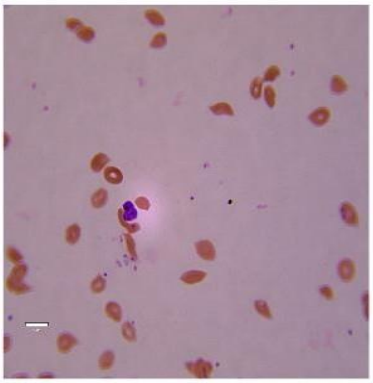
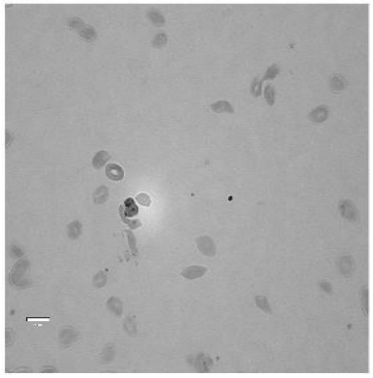


Figure 10: Images of trophozoite stage parasite analyzed with MATLAB (A) Original image showing the parasite in the red blood cell. (B) Grayscale image showing the parasite in the red blood cell. (C) Histogram with the x-axis showing intensity and y axis showing the number of pixels. (D) Automated image showing the detected malaria parasites

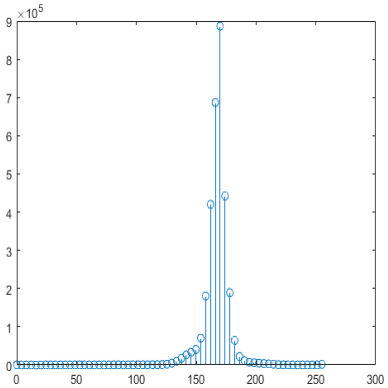
SCHIZONT STAGE

(A) **(B)**





(C) **(D)**



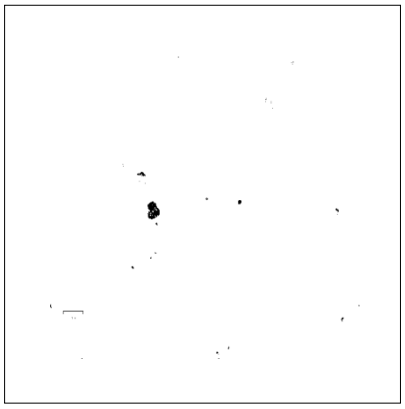


Figure 11: Images of schizont stage parasites analyzed with MATLAB. (A) Original image showing the parasite in the red blood cell. (B) Grayscale image showing the parasite in the red blood cell. (C) Histogram with the x-axis showing intensity and y axis showing the number of pixels. (D) Automated image showing the detected malaria parasites

Sensitivity and specificity are fundamental to understanding the utility of a clinical test. They are independent of the population of interest subjected to the test. The validity of a diagnostic test is measured by sensitivity and specificity.

Sensitivity is the possibility of a positive result given true the presence of the parasite. Sensitivity is the most important issue because false-negative results can lead to missing a malaria case for treatment and this can lead to severe disease or death, with a sensitivity of close to 100%, parasite densities ≥ 100 asexual parasites per µl blood should be detected reliably (80).

It is essential that a diagnostic measure has a high sensitivity since all localized parasites will be the basis for classification of stages, missed infected erythrocytes would mean missed infections. In addition, high precision is also important to estimate the parasitemia correctly. Precision refers to the closeness of two or more measurements to each other and not necessarily how accurate the result is. Sometimes, precision can be separated into repeatability and reproducibility.

- Repeatability – This is the variations that arise despite all efforts made to keep conditions constant by using the same instrument and operator and repeating the measurements during a short time period.
- Reproducibility – The variation arising using the same measurement process among different instruments and operators, and over longer time periods.

$$\textbf{SENSITIVITY} = \frac{TP}{TP + FN} \times 100$$

TP=True positive, FN = False negative

Specificity is the possibility of a negative result or absence of a symptom given the absence of the disease.

$$\textbf{SPECIFICITY} = \frac{TN}{TN + FP} \times 100$$

TN=True negative, FP=False positive

Accuracy is how close a measured value is to the actual (true) value.

$$\textbf{ACCURACY} = \frac{TP + TN}{TP+FP+TN+FN} \times 100$$

TP=True positive, FN = False negative, TN=True negative, FP=False positive

Source: Public Library of Science ONE[62]

We then compared each FOV by manual counting and the automated procedure to analyze the performance of the machine-learning algorithm.

Table 1. Sensitivity, specificity and accuracy of automated images in their different stages

STAGE	SENSITIVITY	SPECIFICITY	ACCURACY
RING	94.6%	96.5%	95.3%
TROPHOZOITE	96.5%	88.9%	95.4%
SCHIZONT	98.2%	81.8%	96.7%

The number of parasites detected with the light microscope was labelled as the manual count while those that were automatically detected using the algorithm were labelled as automated count. This was analyzed using a scrambler software. Figure 11 shows the correlation for all the stages of the parasite. Figure 12,13and 14 show the scatterplot for the ring, trophozoite and schizont stage parasites respectively.

These indices were used because they are the best indices for a diagnostic test. Other indices can be used to access the performance of Machine Learning, but sensitivity, specificity and accuracy are best fit for a diagnostic test

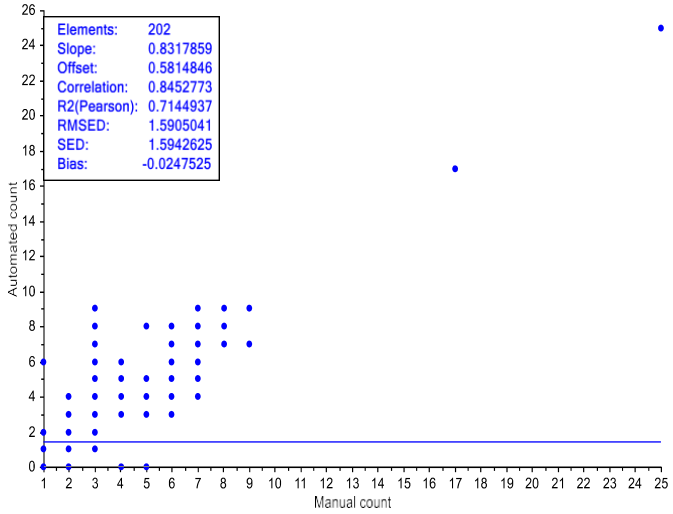


Figure 12: Scatterplot for all the stages comparing detection by visual inspection vs the machine learning algorithm. The Scatterplot shows a relatively strong, positive linear association / correlation between the automated count and the manual count with an increase in trend. The correlation between the manual count and the automated count is 0.85.

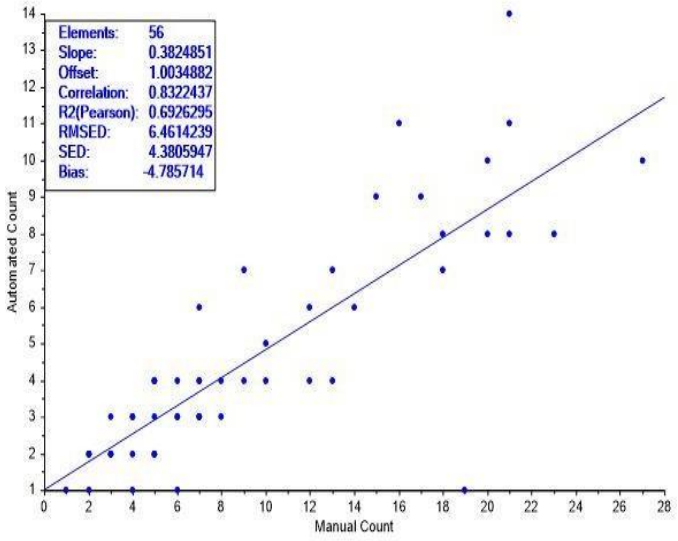


Figure 13: Scatterplot for ring stage parasites. The scatterplot shows a relatively strong, positive linear association / correlation between the automated count and the manual count with an increase in trend. The correlation between the manual count and the automated count is 0.83.

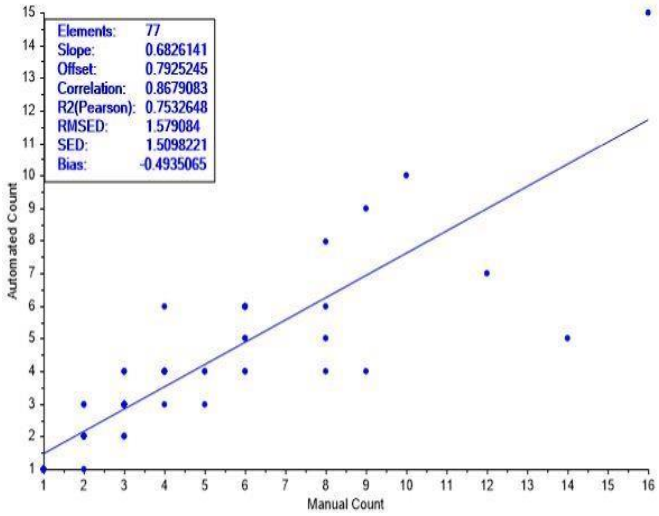


Figure 14: Scatterplot for trophozoite stage parasites. The scatterplot shows a relatively strong, positive linear association / correlation between the automated count and the manual count with an increase in trend. The correlation between the manual count and the automated count is 0.87.

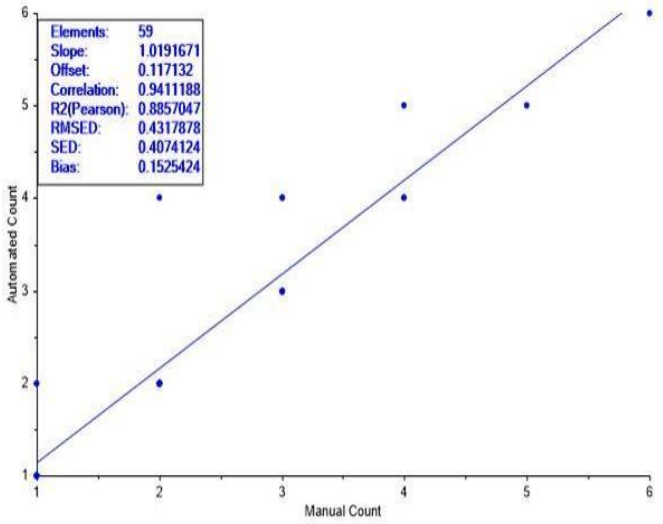


Figure 15: Scatterplot for schizont stage parasites. The scatterplot shows a relatively strong, positive linear association / correlation between the automated count and the manual count with an increase in trend. The correlation between the manual count and the automated count is 0.94.

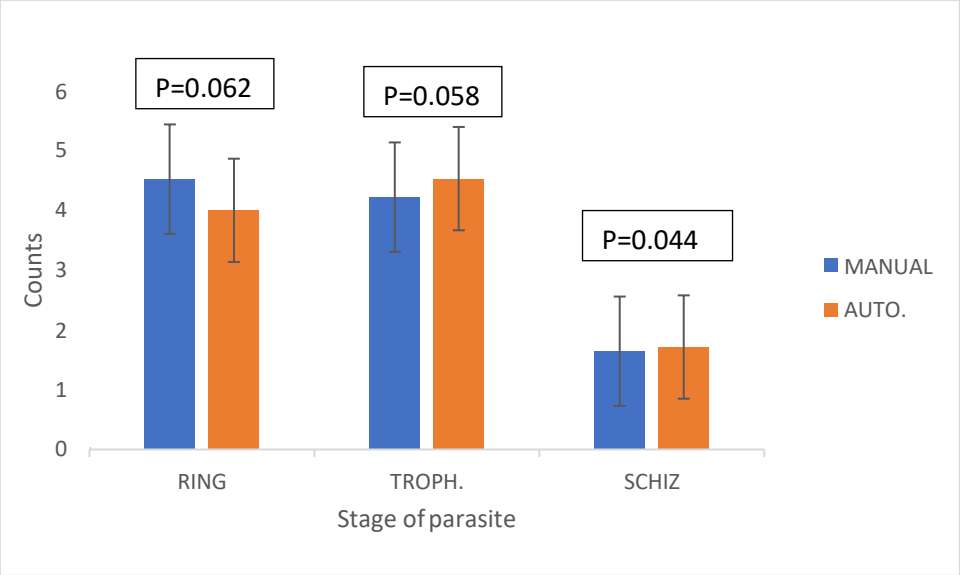


Figure 16: Difference between manual and automated counts at various stages

The two counting methods produce similar results for all stages of the parasites and are not statistically different for schizonts. The analysis shows that, overall, the manual and machine-learning approaches are about equal in performance. P value of ring, trophozoites and schizonts is 0.062, 0.058 and 0.044 respectively

The Receiving Operating Characteristics (ROC) curve, is a graphical plot that shows the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve involves plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The TPR is also known as sensitivity, recall or probability of detection in machine learning, while the FPR is also known as the fall-out or probability of false alarm. The ROC curve is the sensitivity as a function of fall out.

The detection accuracy of the classifier was evaluated with the area under the ROC curve (AUC_{ROC}). The above-mentioned metrics are based on the outcomes of a binary test (true positive (TP), false positive (FP), true negative (TN) and false negative (FN)

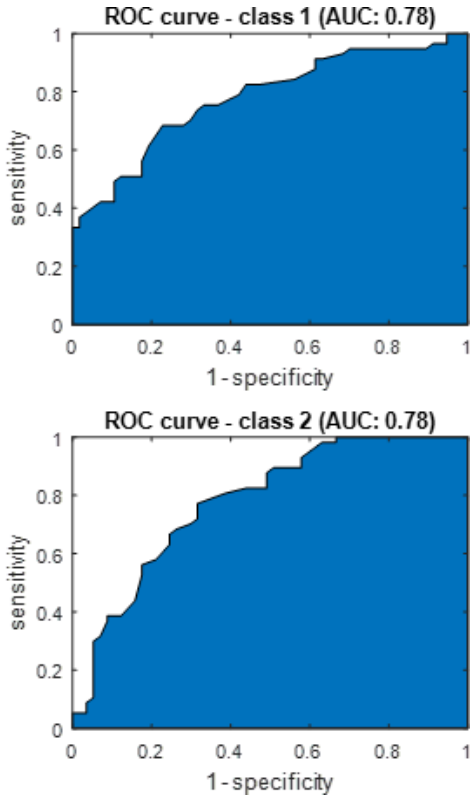


Figure 17: ROC curves; top figure = manual count; bottom figure = automated count

The ROC curves depict the relationship between the sensitivity and specificity of the manual count and automated counting methods. The AUC (Area under the curve) which measures the quality of the model's prediction and how well predictions are ranked is 0.78 for the manual count and for the automated count. This analysis again leads to the conclusion that the two methods perform about equally well.

OBJECTIVE 2

Develop a machine learning system that allows identification of the number of ring-stage parasites in samples diluted with fresh whole blood

Table 2. Sensitivity, specificity and accuracy for ring stage parasite in infected blood whole blood samples

SENSITIVITY	SPECIFICITY	ACCURACY
96.2%	60.6%	71.2%

(A) **(B)**

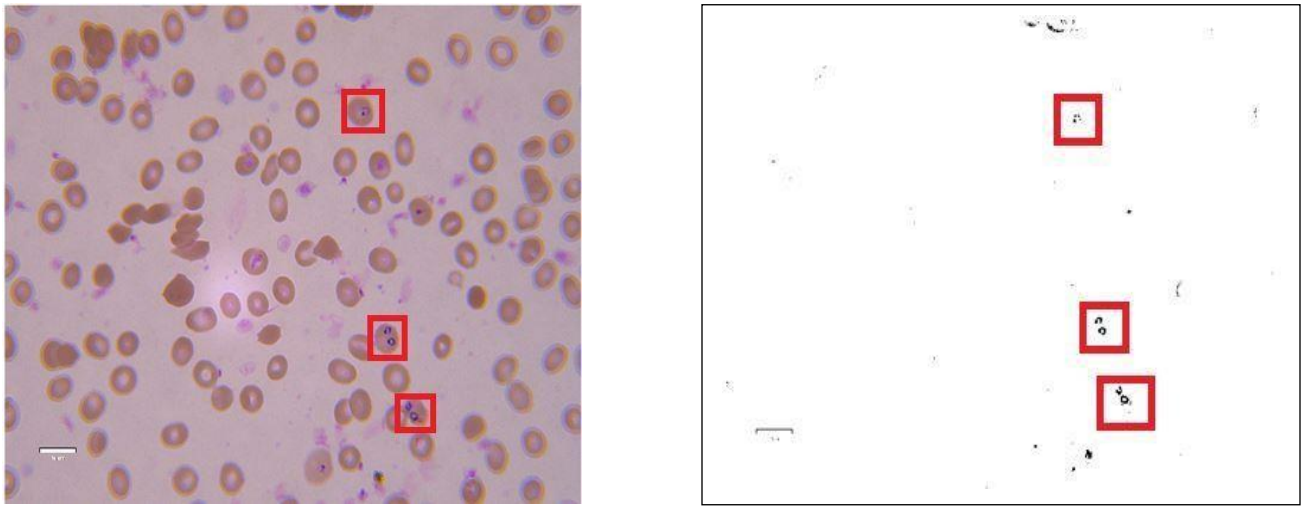


Figure 18: (A) Ring stage parasites detected with the light microscope. (B)Ring stage parasites automatically detected

Thin blood smears were made from synchronized culture of whole blood sample infected with 3D7 strain of *p. falciparum* stained with Giemsa stain and viewed at 100x magnification with the

EVOS microscope, digital images which were saved in tiff format were acquired. The red boxes in the images above show red blood cells infected with *plasmodium falciparum* parasites.

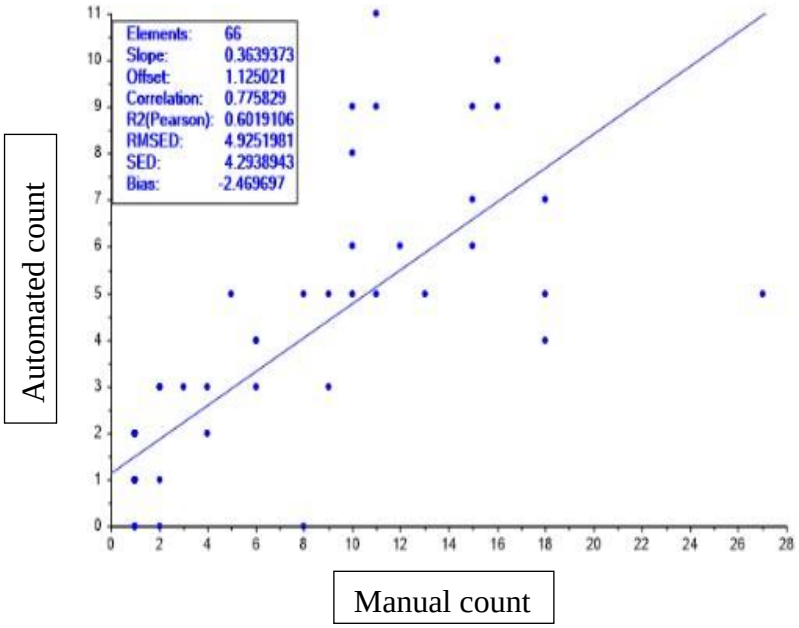


Figure 19: Scatterplot comparing the two counting methods for ring stage parasites in objective 2. The scatterplot shows a moderate, positive linear association / correlation between the automated count and the manual count with an increase in trend. The correlation between the manual count and the automated count is 0.78 and the Pearson (R2) correlation between the automated and manual count is 0.60.

CHAPTER FIVE

DISCUSSION

This study describes automating malaria diagnosis based on a machine learning system. This system is designed to assist Laboratory scientists/Microscopists in the diagnosis of malaria infection. We were able to develop a fully automated image analysis algorithm for malaria diagnosis. The aim of the algorithm is to correctly develop a machine learning system that allows the identification of the stage and number of *P. falciparum* parasites in cultured erythrocytes based on their morphology using thin film slides, as well as the number of ring- stage parasites in samples of cultured erythrocytes mixed with whole blood based on their morphology using thin film slides. The algorithm identified infected red blood cells, using the morphology of the parasites to train the algorithm using features like color, shape and size to detect and classify the parasite in different stages – ring, trophozoite and schizonts. The developed algorithm accurately quantified each parasite stage in culture erythrocytes. We provide a detailed account of the methodology for automating malaria diagnosis from thin film slides using acquired digital images that are subjected to pre-processing, segmentation and feature extraction. The accuracy of this technique was very similar to what we achieved with manual microscopic analysis, making it potentially applicable for malaria diagnosis. This study used the OTSU algorithm (named after Nobuyuki Otsu) to identify infected red blood cells and segment the parasites from the blood smear images. Gray level images were reduced to binary images as the algorithm assumes that the image contains foreground and background pixels. To confirm the stage of parasites using an automatic system, infected erythrocyte segmentation and parasite segmentation are both important processes.

For the first objective, the diagnostic tool showed a sensitivity of 94.6% for rings, 96.5% for trophozoites and 98.2% for schizonts. The schizont stage has a higher sensitivity basically due to its bigger size compared to the other stages. The larger the parasite the more easily the algorithm can detect it. Therefore, more schizont stage parasites could be detected compared to ring stages. Moreso, it showed a specificity of 96.5% for rings, 88.9% for trophozoites and 81.8% for schizonts. Sensitivity is the ability of the diagnostic test to correctly detect true positive rates (those with the disease) while specificity is its ability to correctly identify those without the disease (true negative rate).

For the second objective, detecting ring stages in whole blood, the algorithm showed a sensitivity of 96.2% and specificity of 60.6%.

To assess parasitemia, there is need for accurate segmentation of erythrocytes. Defining a subset of well-defined erythrocytes requires the application of automated erythrocyte segmentation.

For identifying parasites, the R and G channels of the RGB color scheme had clear features which were used to identify objects containing chromatin in Giemsa stained blood films. The input images transformed to grayscale highlighted parasites containing chromatin. This is because the grayscale image subjects each image to contrast stretching and gives intensity information.

Nicholas et al., (64) developed an image processing algorithm to automate the diagnosis of malaria on thin blood smears with a sensitivity of 92%. Our algorithm has a higher sensitivity of 94.6%,96.5% and 98.2% for ring, trophozoite and schizont stages respectively.

The sensitivity of our algorithm for the trophozoite which read 96.5% and schizont stage parasite which was 98.2% is higher than that from the study of Savkare and Narote (60) who used an SVM multi-classifier to demonstrate that methods of feature extraction can produce a very rich group of parameters. Their SVM binary classifier showed 96.3% sensitivity and 99.1% specificity.

According to a study by Linder et al. (63) using an automated system calculated parasitemia and reported the correlation coefficient between manual and automated parasitemia counts of 0.97, whereas, our study showed that correlation between manual and automated count is 0.85.There was a stronger correlation in their study between manual and automated count than in our study.

A study on malaria parasite detection from peripheral blood smear images using deep belief networks by Dhanya Bibin (60) had a sensitivity of 97.6% which is higher than the sensitivity of our study except for schizont stage parasites. Their study also showed a specificity of 95.9% which is higher than that of our trophozoite and schizont but lower than that for the ring stage.

CONCLUSION

This study developed an automated system that could enhance the diagnosis and treatment of malaria parasite. It can also reduce the stress associated with the manual method. The use of artificial intelligence is an emerging concept in the diagnosis of malaria and other diseases.

- This study reported an automated diagnostic method that had a relatively high AUC reading (0.76) in the ROC curve, meaning that it provides the basis for the development of a useful diagnostic method
- The automated method detected more trophozoites and schizonts than ring stage parasites as seen with correlation values between the manual and automated counts of 0.83,0.86 and 0.94 for the ring, trophozoite and ring stages respectively
- An automated diagnostic system could significantly improve the efficiency of microscopists in the field by reducing variability in results and reducing the labor needed to obtain a diagnosis.

SIGNIFICANCE OF THE STUDY

- It is possible to accurately determine the stages of cultured *P. falciparum* in research laboratories where frequent, large scale, and efficient determination of malaria counts is required
- Clinical detection of malaria can be enhanced by machine learning approaches

CONTRIBUTIONS TO THE BODY OF KNOWLEDGE

This study reports a method that can automate the gold standard for malaria diagnosis, thereby minimizing the need for a Microscopists. In addition, it has high sensitivity and specificity.

The algorithm developed from this research can be further developed to ensure accurate determination of parasitemia and stages of *P. falciparum* in research laboratories where frequent, large-scale, efficient determination of parasitemia is required, in addition to being useful for the clinical detection of parasitemia.

LIMITATIONS OF THE STUDY

- The technique was tried on only *P. falciparum* and not on other malaria species
- The digital images acquired had different color schemes

FURTHER STUDIES

Development of an app or system using the algorithm from this research to automate malaria diagnosis by incorporating cell phone-based microscopy platforms.

The algorithm could be further automated to improve its performance.

REFERENCES

1. World Health Organization. World Malaria Report 2017 .Available from: www.who.int/malaria/publications/world-malaria-report-2017/en/.Accessed 13th Sept.,2018
2. World Health Organization. WHO, "Fact sheet: World Malaria Report 2016," 2016. Available from: http://www.who.int/malaria/media/world-malaria-report-2016/en.Accessed 5th Sept.,2018
3. Center for Disease Control. Malaria 2017. Available from: https://www.cdc.gov/dpdx/malaria/about/distribution.html.Accessed 19th June 2019
4. World Health Organization. Guidelines for the Treatment of Malaria. Geneva: World Health Organization; 2006 . Available from: http://whqlibdoc.who.int/publications/2006/9241546948_eng.pdf.Accessed 20th Oct.2018
5. Ahmed L SL, Ainley C, De la Salle B, Brereton M, Hyde K, et al. Web-based virtual microscopy of digitized blood slides for malaria diagnosis: an effective tool for skills assessment in different countries and environments. J Med Internet Res. 2016;18(8):3-5
6. World Health Organization, Malaria 2015 [Available from: http://www.who.int/mediacentre/factsheets/fs094/en/.Accessed 1st July2019
7. Butykai A et al. Malaria pigment crystals as magnetic micro-rotors: key for high-sensitivity diagnosis. Sci Rep. 2013; 3(1431):1-2
8. Breslauer DN, Maamari R. N., Switz N. A., Lam W. A., Fletcher D. A. . Mobile phone based clinical microscopy for global health applications. Public Library of Science ONE 2009;7(4):1-3.
9. Maguire JDea. Production and validation of durable, high quality standardized malaria microscopy slides for teaching, testing and quality assurance during an era of declining diagnostic proficiency. Malar J 2006;5(92):2-3

10. Kiszewski A JB, Schapira A, Delacollette C, Crowell V, Tan-Torres T, et al. . Estimated global resources needed to attain international malaria control goals. Bull World Health Organ. 2007;85(8):623–30.

11. Sio SW, Sun W, Kumar S, Bin WZ, Tan SS, Ong SH, et al. MalariaCount: an image analysis-based program for the accurate determination of parasitemia. Journal of microbiological methods. 2007;68(1):11-8.

12. Bejon P LJ, Olotu A, et.al. Efficacy of RTS, S/AS01E vaccine against malaria in children 5 to 17 months of age. New England Journal of Medicine. 2008;359:2521-32.

13. Richard Ndyomugyenyi PMSC. Diagnosis and treatment of malaria in peripheral health facilities in Uganda: findings from an area of low transmission in south-western Uganda Malaria J. 2007; 6:39-10.

14. Mosha JF, Conteh L, Tediosi F, Gesase S, Bruce J, Chandramohan D. Cost implications of improving malaria diagnosis: findings from north-eastern Tanzania. J Public library of Science ONE. 2010;5(1):e8707.

15. World Health Organization. Basic Malaria Microscopy: Part I. Learner's Guide 2nd ed. Geneva, Switzerland 2010 p. 51–67.

16. Cuomo MJ, Noel, L.B., White, D.B., . Diagnosing Medical Parasites: A Public Health Officers Guide to Assisting Laboratory and Medical Officers 2012. Available from: http://www.phsource.us/PH/PARA/Diagnosing_Medical_Parasites.pdf.Accessed 22nd December 2018

17. Centers for Disease Control. Malaria 2017 [Available from: https://www.cdc.gov/dpdx/malaria/index.html.Accessed 12th Nov.,2018

18. Kyabayinze DJ TJ, Odong GW, Rwakimari JB, Counihan H. . Operational accuracy and comparative persistent antigenicity of HRP2 rapid diagnostic tests for Plasmodium falciparum malaria in a hyperendemic region of Uganda. Malar J. 2008;7:221

19. Bharti AR PK, Chuquiyauri R, Kosek M, Gilman RH, Llanos-Cuentas A, Vinetz JM. . Polymerase chain reaction detection of Plasmodium vivax and Plasmodium falciparum DNA from stored serum samples: implications for retrospective diagnosis of malaria. Am J Trop Med Hyg 2007;77:444–6

20. Erdman LK KK. Molecular diagnostic and surveillance tools for global malaria control. Travel Med Infect Dis 2008;6(1-2):82-99.

21. Ngasala B MM, Warsame M, Petzold MG, Massele AY, Gustafsson LL, Tomson G, Premji Z, Bjorkman A. Impact of training in clinical and microscopy diagnosis of childhood malaria on antimalarial drug prescription and health outcome at primary health care level in Tanzania: a randomized controlled trial. Malar J. 2008;7:199

22. Bhandari PL RC, Rajeev A, Bhandari PD. Comparative study of peripheral blood smear, quantitative buffy coat and modified centrifuged blood smear in malaria diagnosis. Indian J Pathol Microbiol 2008;51(1):108-12.

23. Reyburn H MH, Mwangi R, Mwerinde O, Olomi R, Drakeley C, Whitty CJ. . Rapid diagnostic tests compared with malaria microscopy for guiding outpatient treatment of febrile illness in Tanzania: randomised trial. British Medical Journal. 2007;334:403.

24. Lee SW JK, Jeon BR, Park I. Rapid diagnosis of vivax malaria by the SD Bioline Malaria Antigen test when thrombocytopenia is present. J Clin Microbiol 2008 46(3):939-42.

25. Kyabayinze DJ TJ, Odong GW, Rwakimari JB, Counihan H. Operational accuracy and comparative persistent antigenicity of HRP2 rapid diagnostic tests for Plasmodium falciparum malaria in a hyperendemic region of Uganda. Malar J 2008;7(221).

26. World Health Organization. List of known commercially available antigen-detecting malaria RDTs 2008. Available from: http://www.wpro.who.int/sites/rdt.Accessed 16th Aug.,2018

27. She RC RM, Mohl R, Perkins SL, Hill HR, Litwin CM. Comparison of immunofluorescence antibody testing and two enzyme immunoassays in the serologic diagnosis of malaria. J Travel Med. 2007 14(2):105-11

28. Oh JS, Lee CH, Nam DH, Kim SH, Park DW, Lee CK, Lim CS, Park GH. Evaluation of a malaria antibody enzyme immunoassay for use in blood screening. Mem Inst Oswaldo Cruz 2008;103(1):75-8.

29. Rakotonirina H BC, Raherijafy R, Andrianantenaina H, Ratsimbasoa A, Randrianasolo L, Jahevitra M, Andriantsoanirina V, Ménard D. Accuracy and reliability of malaria

diagnostic techniques for guiding febrile outpatient treatment. Am J Trop Med Hyg 2008;78(2):217-21.

30. Swan SL, Muyombwe A, Chavalitshewinkoon-Petmitr P, Krudsood S, Leowattana W, Wilairatana P,et.al. Evaluation of a real-time polymerase chain reaction assay for the diagnosis of malaria in patients from Thailand. Am J Trop Med Hyg 2005;73(5):850-4.
31. Mens PF vAA, Sawa P, Kager PA, Schallig HD. Molecular diagnosis of malaria in the field: development of a novel 1-step nucleic acid lateral flow immunoassay for the detection of all 4 human Plasmodium spp. and its evaluation in Mbita, Kenya. Diagn Microbiol Infect Dis. 2008;61(4):421-7.
32. Poon LL WB, Ma EH, Chan KH, Chow LM, Abeyewickreme W, Tangpukdee N, Yuen KY, Guan Y, Looareesuwan S, Peiris JS. Sensitive and inexpensive molecular test for falciparum malaria: detecting Plasmodium falciparum DNA directly from heat-treated blood by loop-mediated isothermal amplification. Clin Chem 2006;52(2):303-6.
33. Aonuma H SM, Iseki H, Perera N, Nelson B, Igarashi I, Yagi T, Kanuka H, Fukumoto S. Rapid identification of Plasmodium-carrying mosquitoes using loop-mediated isothermal amplification. Biochem Biophys Res Commun. 2008;376(4):671-6.
34. Erdman LK KK. Molecular diagnostic and surveillance tools for global malaria control. Travel Med Infect Dis 2008;6(1-2):82-99.
35. Poon LL, Ma EH, Chan KH, Chow LM, Abeyewickreme W, Tangpukdee N, Yuen KY, Guan Y, Looareesuwan S, Peiris JS. Loop-Mediated Isothermal Amplification Assay for Rapid Diagnosis of Malaria Infections in an Area of Endemicity in Thailand. J Clin Microbiol. 2014;52(5):1471–7.
36. K. P. Role of DNA microarray in infectious diseases. Chula Med J 2008;52:147–53
37. Palacios G QP, Jabado OJ, Conlan S, Hirschberg DL, Liu Y, Zhai J, et.al.,. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. Emerg Infect Dis 2007;13(1):73-81.
38. Izumiyama S OM, Takasaki T, Ohmae H, Asahi H. Plasmodium falciparum: development and validation of a measure of intraerythrocytic growth using SYBR Green I in a flow cytometer. Exp Parasitol 2009;121(2):144-50

39. Scholl PF KD, Demirev PA, Feldman AB, Lin JS, Sullivan DJ Jr, Kumar N. Rapid detection of malaria infection in vivo by laser desorption mass spectrometry. Am J Trop Med Hyg. 2004;71(5):546-51.

40. Center for Disease Control. Frequently asked questions (FAQs) 2016 [Global Health – Division of Parasitic Diseases and Malaria Available from: https://www.cdc.gov/malaria/about/faqs.html.Accessed 5th Sept.,2018

41. Noppadon T, Polrat W, Srivicha K.et.al Malaria Diagnosis: A Brief Review. Korean J Parasitol 2009;47:93-102.

42. Poostchi M, Silamut K, Maude R, Jaeger S, Thoma G. Image analysis and machine learning for detecting malaria National Center for Biotechnology Information. 2018 Volume 194:36–55

43. Kahama-Maro J DAV, Mtasiwa D, Genton B, Lengeler C et.al.,. Low quality of routine microscopy for malaria at different levels of the health system in Dar es Salaam. Malar J. 2011;10(1186/1475-2875-10-332):332-10.

44. Frean J PO, Fensham V, McCarthy K, von Gottberg A,Alamu W,de Gouveia et al. . External quality assessment of national public health laboratories in Africa, 2002–2009.National Center for Biotechnology Information 2012;90:191–9A.

45. Nankabirwa J ZD, Njogu JN, Rwakimari JB, Counihan H, Snow RW, Tibenderana JK et.al. Malaria misdiagnosis in Uganda-implications for policy change. Malar J 2009;8: 66-10.

46. Runsewe-Abiodun IT, Ghebremedhin B, Sotimehin AS, Ajewole J, Akinleye J, König B et.al. Malaria diagnostics: a comparative study of blood microscopy, a rapid diagnostic test and polymerase chain reaction in the diagnosis of malaria. . J Trop Pediatr 2012;58.

47. Purwar Y, Sirish L, Clarke G, Almugairi A, Muehlenbachs A. Automated and unsupervised detection of malarial parasites in microscopic images. Malaria Journal. 2011;10:1-10

48. World Hwealth Organization. Guidelines for theTreatment of Malaria Geneva: WHO; 2006 Available from: http://archives.who.int/publications/2006/9241546948_eng.pdf.Accessed 15th Oct.2018

49. O'Meara WP MF, Magill AJ, Forney JR, Permpanich B, Lucas C, Gasser RA, Wongsrichanalai C et.al. Sources of variability in determining malaria parasite density by microscopy. American Journal of Tropical Medicine and Hygiene. 2005;73:593-8.

50. Elter M,Hasslmeyer E,Zerfass T. Detection of malaria parasites in thick blood films.(Available from :https://www.onlinebiologynotes.com/plasmodium-falciparum-morphology-life-cycle-pathogenesis-and-clinical-disease/

51. Kshipra C. Bairagi V. Automated malaria parasite and their stage detection in microscopic blood images. IEEE Sponsored 9th International Conference on Intelligent Systems and Control (ISCO)2015 ;2-3

52. Hänscheid T. Current strategies to avoid misdiagnosis of malaria. Clinical Microbiology and Infection. 2003;9(6):497-504.

53. McMorrow ML, Abdulla SM, Kahigwa E, Kachur SP. Challenges in routine implementation and quality control of rapid diagnostic tests for malaria. Am J Trop Med Hyg. 2008;79:385-90.

54. Lloyd S, Mohseni M, Rebentrost PJapa. Quantum algorithms for supervised and unsupervised machine learning. arxiv 2013 :1307.0411 .

55. Kotsiantis SB. Supervised machine learning: A review of classification techniques.Informatica 2007;160:3-24.

56. Kashif S, Hazrat A, Zhang Z. Big Data Perspective and Challenges in next generation networks. Research gate. 2018:10(7):56.Available from:http://www.researchgate.net/publication/325928183big_data_perspective_and_challenges_in_next_generation_networks

57. Aravindan S. What is MATLAB? 2019. Available from: https://study.com/academy/leson/what-is-matlab-html.Accessed 14th Nov.2018

58. Das DK GM, Pal M, Maiti AK, Chakraborty C. Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron Oxf Engl 2013;45: 97–106.

59. Charles B., Courosh M, Liming H, Shawn K, Cary R, Mattew P. et.al.,. Automated microscopy and machine learning for expert-level malaria field diagnosis IEEE Global

Humanitarian Technology Conference 2015;978:1-4.

60. Dhanya Bibin, Madhu S, Punitha P. Malaria parasite detection from peripheral blood smear images using deep belief networks. IEEE Access. 2017;5:9105-6.

61. Savkare S, Narote S . Automatic system for classification of erythrocytes infected with malaria and identification of parasite's life stage. Science Direct 2012;6:405-10.

62. Purwar Y, Shah SL, Clarke G, Almugairi A, Muehlenbachs A. Automated and unsupervised detection of malarial parasites in microscopic images.Malaria Journal. 2011;10(1):364.

63. Linder N, Turkki R, Walliander M, Mårtensson A, Diwan V, Rahtu E, et al. A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears.Public Library of Science ONE 2014;9(8):e104855.

64. Nicholas ER, Pritchard CJ, Rubin DM, Adriano GD. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. Med Bio Eng Comput 2006;44(5):427–36

65. Das DK, Madhumala G, Mallika P, Asok K, Chandan C et.al.,. Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron. 2013;45:97-106.

66. Trager W, Jensen J.B. Human malaria parasites in continuous culture. Science. 1976;193:673-5.

67. Wendi B, Williams J, Bain B, Parker-Williams J, Chiodini P. Guideline: the laboratory diagnosis of malaria.British Journal of Haematology 2013;163:573-580.

68. Leidenberger M, Voigtlander C, Simon N, Kappes B. SYBR® Green I-Based Fluorescence Assay to Assess Cell Viability of Malaria Parasites for Routine Use in Compound Screening. National Center for Biotechnology Information. 2017; 1601:97- 110.

69. Ghate D, Chaya J, Usha R. Automatic detection of malaria parasite from blood images. International Journal of Advanced Computer Technology. 2017;4(1):129-32.

70. World Health Organization. Giemsa staining of Malaria blood film. Malaria microscopy standard operating procedure-MM-SOP-07A 2016.p 1-5, Philippines

71. . Kang W, Yang Q, Liang R,"The Comparative Research on Image Segmentation Algorithms", IEEE Conference on ETCS.2009;703-707.

72. Poostchi M, Silamut K, Maude R, Jaeger S, Thomas G. Image analysis and machine learning for detecting malaria. Sci Dir. 2018; 194:36-55.

73. Umbaugh SE. Computer imaging:digital image analysis and processing. 1st ed: CRC Press; 2005. p.3

74. Wu Q. MF, Castleman K. Microscope Image Processing. New Jersey: Academic Press; 2008.

75. Bhargava N, Kumawat A, Bhargava R. Threshold and binarization for document image analysis using otsu ' s Algorithm. Int J Comput Trends Technol. 2014 17(5):272–5.

76. Gonzalez RC, Woods, R.E., Eddins, S.L. Digital Image Processing using MATLAB. 2nd ed. Pearson Prentice Hall, Upper Saddle River: Gatemarks; 2009.

77. Otsu N. A threshold selection method from gray level histograms. IEEE Trans Syst Man Cybern SMC. 1979;9(1):62–6.

78. Díaz G, González FA, Romero E. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. Sci Dir. 2009;42(2):296-307.

79. Ahirwar A. PS, Acharya B. Advancced Image Analysis Based System for Automatic Detection and Classification of Malarial Parasite in Blood Images. International Journal of Information Technology and Knowledge Management. 2012; 5(1):59 - 64.

80. World Health Organization. New perspectives: malaria diagnosis. Report of a joint WHO/USAID informal consultation 2000. Available from: https://www.who.int/tdr/publications/documents/malaria-diagnosis.pdf.Accessed July,2nd 2019.

McGill

Research Ethics Board Office

Faculty of Agricultural and Environmental Sciences

Macdonald Campus

21 111 Lakeshore

Saint-Anne-de Bellevue, QC H9X 3V9

Email: lynda.mcneil@mcgill.ca Tel: (514)398-6831

Website: www.mcgill.ca/macdonald/research/compliance/human/

FAES Research Ethics Board

Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 202-0918

Project Title: Malaria diagnosis based on a machine learning system

Principal Investigator: Chidinma Maduako **Department:** Parasitology

Status: Master's Student **Supervisor:** Professor Timothy Geary

Co-Investigator(s)/Other Researchers: Prof. Petra Rohrbach

Approval Period: October 10, 2018 to October 9, 2019

The FAES REB reviewed and approved this project by delegated review in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Participants and the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

Deanna Collin

Senior Ethics Review Administrator

---

* Approval is granted only for the research and purposes described.

* Modifications to the approved research must be reviewed and approved by the REB before they can be implemented.

* A Request for Renewal form must be submitted before the above expiry date. Research cannot be conducted without a currentethics approval. Submit 2-3 weeks ahead of the expiry date.

* When a project has been completed or terminated, a Study Closure form must be submitted.

* Unanticipated issues that may increase the risk level to participants or that may have other ethical implications must be promptly reported to the REB. Serious adverse events experienced by a participant in conjunction with the research must be reported to the REB withoutdelay.

* The REB must be promptly notified of any new information that may affect the welfare or consent ofparticipants.

* The REB must be notified of any suspension or cancellation imposed by a funding agency or regulatory body that is related to thisstudy.

* The REB must be notified of any findings that may have ethical implications or may affect the decision of the REB.

LINK FOR IMAGES

IMAGES FOR OBJECTIVE 1

RINGS STAGE

https://drive.google.com/drive/folders/1ekmWzjbbUrSLo1MjwLJbZjfkk4HMgyfb?usp=sharing

TROPHOZOITE STAGE

https://drive.google.com/drive/folders/1NYaSdXm-xnZlqFMISijjvWyKTq00DKkn?usp=sharing

SCHIZONT STAGE

https://drive.google.com/drive/folders/1nb8I4iCI5TwFb1ukkG5OCxwTJIbCI2BZ?usp=sharing

IMAGES FOR OBJECTIVE 2

RING STAGE

https://drive.google.com/drive/folders/12kwgSyHQbgIn8_GRAMBoZUPlJ3IEUi2l?usp=sharing