



Guo, Xin (2024) *AI and Blockchain-assisted diagnostics in resource-limited setting*. PhD thesis

<https://theses.gla.ac.uk/84174/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University
of Glasgow

AI and Blockchain-assisted Diagnostics in Resource-Limited Setting

Xin Guo

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow

March 2024

Abstract

Diseases, including communicable and noncommunicable diseases, have been one of the major causes of human morbidity and mortality since the beginning of our history. Although many diseases have become treatable or preventable, thanks to interventions including pharmaceutical and technological advances, many people die each year in developing countries and remote rural areas due to limited (or even no) access to medical facilities and expertise. An accurate, rapid, and reliable diagnostic test is vital to improved disease treatment and prevention. However, running diagnostic tests usually requires complex, expensive instruments, professionally trained operators, and a stable power supply. Unfortunately, these resources are generally limited or unavailable in many low-resource settings.

Although there are countless limitations in running diagnostic tests in low-resource settings, various endeavours have been made to overcome the existing obstacles. One of the most important advances has been the development of point-of-care or point-of-need tests. These diagnostic assays can be delivered in convenient formats and have successfully reduced the cost of running diagnostics, so playing an essential role in disease management and lifesaving in low-income countries.

One key aspect of diagnosis may be the interpretation of the test, which can either be done by an expert in the field or by communicating that data to a remote expert or a “smart” system to interpret the data. Accurately interpreting the test outcome can help the patients receive appropriate treatment timely. However, issues presented in data management during such communication, such as tampered and counterfeited test results and unsecured data sharing between end users (patients) and professionals (doctors, healthcare workers, researchers, etc.). Also, problems like unreliable electricity supply and internet connection were found during the field study conducted by our group previously, and those issues can also delay the diagnosis of the disease.

In this PhD study, an AI-assisted platform for DNA-based malaria diagnostic tests was developed and tested in the field. This platform allows users to run a test with a low-cost

portable heater and record the test information with an Android phone. It can be used to run LAMP-based malaria tests with a portable heater and read the test results automatically with 97.8% accuracy. And it only takes around 20 milliseconds to classify one image on an inexpensive (~£100) Android phone. When the internet connection is available, the test information can be safely kept in a Blockchain network for future use to inform treatment or surveillance activities.

Expertise developed in the deep neural network was also used to train algorithms for the diagnosis of retinopathies, involving developing methods for retina vessel segmentation and classification, which explores the possibility of applying AI to diagnostics in low-resource settings. In such settings, accessing medical expertise can be challenging. It has been found that using only a convolutional neural network is not sufficient in identifying arteries and veins. Models were trained for performing vessel segmentation and classification tasks; for segmenting vessels from the background achieved over 95% accuracy and over 0.8 mean average over the union score (MIoU) on the DRIVE dataset, while for A/V classification tasks, the MIoU decreased to less than 0.7. However, combining it with the traditional approach has the potential to achieve good performance.

In addition, research was conducted on the utilisation of digital technologies to assist other researchers and engage with the public. To assist researchers in determining the minimum required sample size, a web-based calculator was developed during the COVID-19 pandemic. Furthermore, a website was created containing 360-degree images to help individuals comprehend the challenges of diagnostics and healthcare in developing regions and to raise awareness about how infectious diseases spread.

Table of Contents

| | |
|--|----|
| Chapter 1 Introduction | 1 |
| 1.1 Background | 1 |
| 1.2 Malaria | 2 |
| 1.2.1 Clinical diagnosis for malaria | 2 |
| 1.2.1.1 Microscopic diagnosis..... | 2 |
| 1.2.1.2 Quantitative buddy coat (QBC) | 3 |
| 1.2.1.3 Rapid diagnostic tests (RDTs) | 4 |
| 1.2.1.1 Molecular diagnostic..... | 6 |
| 1.2.1.1.1 PCR | 6 |
| 1.2.1.1.2 Isothermal amplification (LAMP)..... | 6 |
| 1.3 Challenges of running diagnostic tests in low-resource settings | 7 |
| 1.4 Digital technologies in diagnostics | 7 |
| 1.4.1 Mobile network | 8 |
| 1.4.2 AI supported diagnostics..... | 9 |
| 1.4.2.1 Deep learning for retina vessel segmentation/classification and diabetic retinopathy detection..... | 10 |
| 1.4.3 Blockchain..... | 12 |
| 1.4.3.1 Types of Blockchain networks..... | 14 |
| 1.4.3.2 Smart contract | 15 |
| 1.4.3.3 Challenges and limitations of using Blockchain..... | 16 |
| 1.4.4 IoT and edge computing..... | 16 |
| 1.5 Aims and objectives | 17 |
| 1.6 Thesis structure | 19 |
| Chapter 2 Mobile phone-controlled heater for LAMP test..... | 20 |
| 2.1 Introduction | 20 |
| 2.2 Multiplex LAMP system..... | 21 |
| 2.3 Material and methods..... | 22 |
| 2.3.1 Mobile heater | 22 |

| | | |
|-----------|--|----|
| 2.3.1.1 | Material | 23 |
| 2.3.1.2 | Method | 24 |
| 2.3.1.2.1 | Power supply unit..... | 26 |
| 2.3.1.2.2 | Heating unit..... | 27 |
| 2.3.1.2.3 | Temperature sensor unit..... | 28 |
| 2.3.1.2.4 | Microcontroller Unit | 29 |
| 2.3.2 | Android app..... | 30 |
| 2.4 | Results | 32 |
| 2.4.1 | Laboratory testing | 32 |
| 2.4.2 | Field testing..... | 32 |
| 2.5 | Conclusion..... | 33 |
| Chapter 3 | Blockchain network for diagnostic data management..... | 35 |
| 3.1 | Introduction | 35 |
| 3.1.1 | Central concepts in Hyperledger Composer | 36 |
| 3.2 | Methods..... | 36 |
| 3.2.1 | Blockchain network | 36 |
| 3.2.1.1 | Blockchain model..... | 37 |
| 3.2.1.2 | Transaction..... | 38 |
| 3.2.1.3 | Access control and authentication..... | 39 |
| 3.2.1.4 | Authentication | 39 |
| 3.2.2 | Android app..... | 39 |
| 3.3 | Results | 41 |
| 3.3.1 | Blockchain performance test..... | 41 |
| 3.4 | Conclusion and Discussion | 42 |
| Chapter 4 | Convolutional neural network (CNN) for automated malaria test result interpretation | 45 |
| 4.1 | Introduction | 45 |
| 4.1.1 | Deep learning | 45 |
| 4.1.1.1 | Convolutional layers | 45 |
| 4.1.1.2 | Dense layers | 46 |

| | |
|---|----|
| 4.1.2 Activation function..... | 47 |
| 4.1.2.1 ReLU | 47 |
| 4.1.2.2 Sigmoid | 49 |
| 4.1.2.3 SoftMax..... | 49 |
| 4.1.3 Optimizer..... | 50 |
| 4.1.4 Deep CNNs | 51 |
| 4.1.4.1 AlexNet | 51 |
| 4.1.4.2 ResNet..... | 51 |
| 4.2 Methods..... | 52 |
| 4.2.1 Data preparation | 52 |
| 4.2.2 Model training..... | 56 |
| 4.2.2.1 Classification network..... | 56 |
| 4.2.2.1.1 Architecture of CNN | 56 |
| 4.2.2.2 Object detection network | 58 |
| 4.2.3 App implementation..... | 59 |
| 4.3 Results and discussion..... | 60 |
| 4.4 Conclusion..... | 63 |
| Chapter 5 Retina vessel identification and segmentation using deep learning | 65 |
| 5.1 Introduction | 65 |
| 5.1.1 Challenges in retina vessel segmentation and classification..... | 68 |
| 5.1.2 Diabetic retinopathy | 69 |
| 5.2 Methods..... | 69 |
| 5.2.1 Data preparation for vessel classification | 69 |
| 5.2.1.1 DRIVE dataset | 69 |
| 5.2.1.2 UWF retina images | 70 |
| 5.2.1.3 Image labelling..... | 71 |
| 5.2.1.4 Data augmentation | 73 |
| 5.2.2 DNN Models for Semantic Segmentation..... | 74 |
| 5.2.2.1 UNet | 74 |
| 5.2.2.2 PsPNet | 77 |
| 5.2.2.3 Attention UNet..... | 78 |

| | |
|---|-----|
| 5.2.3 Model training..... | 79 |
| 5.2.4 Vessel tracking..... | 80 |
| 5.2.5 Profile extraction using semantic segmentation..... | 81 |
| 5.3 Results and discussion..... | 87 |
| 5.3.1 Vessel classification and segmentation..... | 87 |
| 5.3.1.1 Vessel segmentation..... | 88 |
| 5.3.1.2 Artery/Vein (A/V) classification..... | 89 |
| 5.3.2 Vessel tracking and profiling..... | 90 |
| 5.4 Conclusion..... | 92 |
| Chapter 6 Sample size calculator for diagnostic study | 94 |
| 6.1 Introduction..... | 94 |
| 6.2 Methods..... | 95 |
| 6.2.1 Waterfall methodology..... | 95 |
| 6.2.2 Webpage development..... | 96 |
| 6.2.3 User interface design..... | 96 |
| 6.2.4 Implementing website functionality..... | 99 |
| 6.2.5 Validation..... | 102 |
| 6.3 Results..... | 102 |
| Chapter 7 Discussion..... | 104 |
| Chapter 8 Conclusion and future works..... | 109 |
| 8.1 Thesis conclusion..... | 109 |
| 8.2 Future works..... | 110 |
| Appendix..... | 112 |
| Appendix figures..... | 112 |
| Appendix tables..... | 116 |
| Other appendices..... | 122 |
| References..... | 147 |

List of Figures

| | |
|---|----|
| Figure 1-1 Malaria microscopic diagnosis..... | 3 |
| Figure 1-2 Malaria rapid diagnostic test. | 4 |
| Figure 1-3 Number of RDTs sold by manufacturers and distributed..... | 5 |
| Figure 1-4 Mobile subscriptions and penetration..... | 8 |
| Figure 1-5 An example image from the DRIVE dataset and DNN prediction. | 12 |
| Figure 2-1 Heater design[96]. | 23 |
| Figure 2-2 Circuit schematic diagram [96] | 25 |
| Figure 2-3 Voltage converters..... | 26 |
| Figure 2-4 The basic principle of the buck converter. | 27 |
| Figure 2-5 The TEG control circuit. | 28 |
| Figure 2-6 Temperature calibration device. | 29 |
| Figure 2-7 The PID control mechanism..... | 30 |
| Figure 2-8 The user interface of the mobile App..... | 31 |
| Figure 2-9 The mobile heater was tested at different target temperatures [96] | 32 |
| Figure 2-10 Field test result. | 33 |
| Figure 3-1 The Blockchain architecture..... | 35 |
| Figure 3-2 The system architecture of the diagnostic Blockchain. [96] | 37 |
| Figure 3-3 The Blockchain diagnostics app. [96] | 40 |
| Figure 3-4 Transactions tested with manufacture network card in composer playground. | 41 |
| Figure 4-1 The convolutional layer..... | 46 |
| Figure 4-2 Dense layer. | 47 |
| Figure 4-3 The ReLU function..... | 48 |
| Figure 4-4 Leaky ReLU (left) and ELU (right) activation function. | 48 |
| Figure 4-5 The Sigmoid function..... | 49 |
| Figure 4-6 A residual block in Resnet..... | 52 |
| Figure 4-7 Examples of images from different categories..... | 54 |
| Figure 4-8 The Plastic cartridge [96] | 55 |
| Figure 4-9 An example of image labelling using LabelMe [121]..... | 55 |
| Figure 4-10 The architecture of CNN. | 57 |
| Figure 4-11 Performance of the object detection model..... | 59 |
| Figure 4-12 The performance of the CNN..... | 60 |

| | |
|---|-----|
| Figure 4-13 The precision-recall curves | 62 |
| Figure 5-1 An example of images from the DRIVE dataset..... | 67 |
| Figure 5-2 The noise in the UWF fundus image..... | 68 |
| Figure 5-3 Sample images from the DRIVE dataset..... | 70 |
| Figure 5-4 Example of the retinal image used for training. | 71 |
| Figure 5-5 The user interface of LabelMe [161] software..... | 72 |
| Figure 5-6 The labelled retina image. | 72 |
| Figure 5-7 An example of the cropped image..... | 74 |
| Figure 5-8 UNet architecture | 75 |
| Figure 5-9 The architecture of the optimised UNet | 76 |
| Figure 5-10 Overview of the PsPNet used for retina vessel segmentation and classification..... | 78 |
| Figure 5-11 The architecture of the Attention UNet..... | 78 |
| Figure 5-12 Attention Gate [167]..... | 79 |
| Figure 5-13 Intersection over union(IoU)..... | 80 |
| Figure 5-14 Vessel profiling and OD calculation. | 81 |
| Figure 5-15 The vessel label(a) and the vessel skeleton(b). | 82 |
| Figure 5-16 An example of the binarized skeletonised vessel segment..... | 83 |
| Figure 5-17 Vessel tracking process. | 84 |
| Figure 5-18 The tracking function at pixel (i, j)..... | 85 |
| Figure 5-19 Profile calculation..... | 87 |
| Figure 5-20 The outcome of vessel tracking and profile algorithm..... | 91 |
| Figure 5-21 The software GUI..... | 92 |
| Figure 6-1 The waterfall methodology. | 95 |
| Figure 6-2 The user interface of the sample size calculator V1.0..... | 97 |
| Figure 6-3 The UI with responsive design..... | 98 |
| Figure 6-4 Testing the accuracy of a single test..... | 101 |

List of Tables

| | |
|--|----|
| Table 1-1 Comparisons among public Blockchain, consortium Blockchain and private Blockchain [89] | 15 |
| Table 3-1 Performance evaluation results | 42 |
| Table 4-1 Classification criteria | 53 |
| Table 4-2 Model performance comparison | 61 |
| Table 5-1 Performance comparison of retina vessel segmentation and A/V classification | 66 |
| Table 5-2 The Vessel segmentation accuracy of UNet, Att-UNet, and PsPNet | 88 |
| Table 5-3 The MIoU of UNet, Att-UNet, and PsPNet in vessel segmentation | 88 |
| Table 5-4 Sensitivity and specificity of different models on the DRIVE dataset | 89 |
| Table 5-5 The Vessel classification accuracy of UNet, Att-UNet, and PsPNet | 89 |
| Table 5-6 The MIoU of UNet, Att-UNet, and PsPNet | 89 |

List of Appendix Figures

| | |
|---|-----|
| Appendix Figure 1 Screenshots of Android App (latest version) | 112 |
| Appendix Figure 2 Workflow of the Blockchain network..... | 113 |
| Appendix Figure 3 Workflow of the Android App..... | 114 |
| Appendix Figure 4 ETDRS standard photography 2A, 6A and 8A.[157], [159] | 115 |

List of Appendix Tables

| | |
|--|-----|
| Appendix Table 1 The estimated cost of a mobile heater | 116 |
| Appendix Table 2 System resource usage during the Blockchain Performance Test | 117 |
| Appendix Table 3 Performance evaluation results | 118 |
| Appendix Table 4 Species of Plasmodium that cause malaria [223] | 119 |
| Appendix Table 5 $Z_{\alpha 2}$ value and Significance level corresponding to confidence level | 120 |
| Appendix Table 6 $Z_{1 - \beta}$ value under different power | 120 |
| Appendix Table 7 Normal distribution – inverse cumulative distribution function | 121 |

List of Abbreviations

A

acrylonitrile butadiene styrene

(ABS) · 20

artery vein

(A/V) · 19, 67

artificial intelligence

(AI) · 8

attention gate

(AG) · 80

B

Bluetooth low energy

(BLE) · 31

C

cardiovascular disease

(CVD) · 10

case-fatality rate

(CFR) · 1

computer vision

(CV) · 10

Conditional Random Fields

(CRF) · 11

Convolutional Neural Network

(CNN) · 11, 64, 77

D

Decision Tree

(DT) · 17

deoxyribonucleic acid (DNA) · 3

Digital Retinal Images for Vessel Extraction

(DRIVE) · 11

E

electronic health records
(EHRs) · 8

G

ground
(GND) · 26

I

internet of things
(IoT) · 8

K

K-Nearest Neighbour
(KNN) · 17

L

light-emitting diodes
(LEDs) · 24

long-lasting insecticidal nets
(LLINs) · 1

loop-mediated isothermal amplification
(LAMP) · 2

Loop-mediated isothermal amplification
(LAMP) · 7

low and middle-income countries
(LMICs) · 1

low- and middle-income countries
(LMICs) · 64

low-income countries
(LICs) · 1

M

Machine learning

(ML) · 10
metal-oxide-semiconductor field-effect transistor
(MOSFET) · 23
microcontroller unit
(MCU) · 22
Ministry of Health
(MoH) · 21
multi-channel module
(MM) · 91
multi-feature fusion
(MF) · 91
multi-scale transformation
(MT) · 91

N

natural language processing
(NLP) · 45
neural processing units
(NPU) · 63

O

On-To-Go
(OTG) · 20
optical density
(OD) · 13

P

Plasmodium falciparum

(*Pf*) · 54

Plasmodium pan

(*Ppan*) · 54

polymerase chain reaction

(PCR) · 2, 6

Pulse Width Modulation

(PWM) · 28

Pyramid Scene Parsing Network

(PsPNet) · 78

Q

Quantitative buddy coat
(QBC) · 3

R

Rapid diagnostic tests
(RDTs) · 4
retinopathy of prematurity
(RoP) · 89

S

seasonal malaria chemoprevention
(SMC) · 1
sub-Saharan Africa
(SSA) · 1
support vector machine
(SVM) · 11

T

thermoelectric generator
(TEG) · 24

V

Vector Control Division
(VCD) · 21
vertical flow assay
(VFA) · 10

List of Accompanying Material

Publication

Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and Blockchain technology for security. *Nat Electron* **4**, 615–624 (2021).
<https://doi.org/10.1038/s41928-021-00612-x>

Acknowledgements

I would like to express my sincere gratitude to those who have generously supported me throughout my PhD studies.

First and foremost, I am deeply grateful to my supervisors, Professor Jon Cooper and Professor Julien Reboud, for their constant guidance and support throughout my research and daily life. Their valuable insights, constructive feedback, and unwavering encouragement have been instrumental in helping me overcome various challenges and find my direction.

I also extend my heartfelt appreciation to Dr Muhammad Arslan Khalid for his valuable assistance and collaborative efforts during the initial stages of my PhD journey. His guidance, advice, and suggestions have played a pivotal role in shaping the direction of my research and have been of great value to me.

Moreover, I would like to express my gratitude to Dr Xiaoxiang Yan and Dr Shantimoy Kar for their invaluable assistance with the wet lab experiments. Their guidance and support greatly enhanced my understanding of the experiments.

I am grateful for the support of the many colleagues in the Vector Control Division of the Uganda Ministry of Health, including administrators, technicians, nurses, and drivers, who supported the field testing (with their contributions also recognised through co-authorship on the primary publication arising from this work).

Furthermore, I am grateful to Ivo Domingos for his support in my Android development. His programming skills and experience have accelerated the development process.

I am also thankful for Dr Victor Ochoa's assistance in providing retinal images and data for my study, from whom I have learned a lot from our collaboration, and to Professor Andy Harvey, whose team collected the source data.

My family has been a constant source of support, encouragement, and understanding. I thank them from the bottom of my heart for their unwavering belief in my abilities and unconditional love.

I am also deeply grateful for the support and encouragement of my friends and colleagues, who have been a constant source of inspiration and motivation through insightful discussions and sharing of experiences.

Lastly, I would like to express my gratitude to the China Scholarship Council (CSC) for providing me the opportunity to pursue my PhD and for their financial assistance. This work was supported by China Scholarship Council (CSC No. 201806070160)

Author declaration

I, hereby, declare that this submission is entirely my own work, in my own words, except where explicit reference is made to the contribution of others. It has not been submitted, in whole or in part, by another person or me for the purpose of obtaining any other degree in the University of Glasgow or any other Institution.

Chapter 1 Introduction

1.1 Background

Infectious diseases remain one of the most crucial challenges globally. As a recent example, during the SARS-CoV-2 (COVID-19) outbreak, more than 6 million people have died since 2019 [1], and as of September 2022, more than 500 million cases have been confirmed worldwide [1]. During the pandemic, medical resources were inadequate to deal with the quick spread of the virus [2]. According to the WHO COVID-19 dashboard, the United States of America reported the highest number of confirmed cases, over 90 million, with more than 1 million deaths as of 2022 [1].

If a patient has multiple medical conditions, known as comorbidities, they may be at a greater risk of experiencing severe illness or even death [3]. Some non-communicable diseases, such as diabetes, cancer and hypertension, can also significantly increase the case-fatality rate (CFR) from infectious diseases [3], [4]. Studies from China, the US, and Italy revealed that comorbidity is one of the major causes of critical illness [5], [6], [7], [8]. The mortality of critically ill patients with at least one comorbidity was extremely high (56%-97%) [9].

Although many reports show that COVID-19 mortality in sub-Saharan regions was surprisingly lower than elsewhere [3], the situation still caused concern. Many countries imposed lockdown policies, which successfully suppressed coronavirus transmission but led to interruptions in the provision of treatment for other diseases. In sub-Saharan Africa (SSA), malaria cases and death rates rose during the lockdowns [10], [11], because the distribution of long-lasting insecticidal nets (LLINs) [12] and seasonal malaria chemoprevention (SMC) was suspended. According to WHO, the mortality linked to malaria has increased by approximately 12% from 2019 to 2020 [13].

Taken together, infectious, or communicable diseases are one of the leading causes of death in low-income countries (LICs) and low and middle-income countries (LMICs) [14]; diseases like malaria, TB, HIV/AIDS etc., have threatened people's lives for centuries. According to WHO, there were 241 million malaria cases and 627,000 deaths caused by malaria in 2020 [15], and over 65% of confirmed HIV cases are in Africa [16].

1.2 Malaria

Malaria is an acute infectious disease caused by Plasmodium parasites, spreading through the bites of infected female Anopheles mosquitoes. There are five different types of malaria parasites: *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. Four types of those malaria species can infect human beings, *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae*, with *P. falciparum* and *P. vivax* showing the greatest threat to human life [13], [17].

Although malaria has been eliminated in many countries in the past 150 years [18], it remains a heavy burden for many LMICs. Early diagnosis is crucial to stop the spread of malaria and provide adequate treatment to patients, especially for children [19]. There are various methods for malaria diagnosis, including microscopy, polymerase chain reaction (PCR), loop-mediated isothermal amplification (LAMP), and Rapid Diagnostic Tests (RDTs) using immunodiagnostics to measure antigens/antibodies. However, each approach has its advantages and disadvantages, not least how easily they can be adopted into the field, at the point-of-care.

1.2.1 Clinical diagnosis for malaria

Traditional clinical malaria diagnosis was based on the patient's symptoms, such as fever, dizziness, and diarrhoea. Diagnosis by such observation of symptoms is the cheapest and is still used widely. However, the symptoms are nonspecific and are easily confused with other common "febrile" diseases. That can often lead to, misdiagnosis, over-diagnosis [20] and over-treatment, all of which can result, in the long term, in the emergence of drug resistance. The use of symptomatic diagnosis ultimately may lead to increases in the costs of treatment and mortality rates.

1.2.1.1 Microscopic diagnosis

Light microscopy remains the gold standard of malaria diagnosis through examination of the red blood cells in thick or thin blood smears/blood films using Giemsa, Wright's, or Field's stains [20]. Detection and identification of Plasmodium under the microscope is widely accepted because it is simple, has a relatively low-cost and can identify the infectious species. However, the stain and interpretation require a trained technician and manual scanning of samples under the microscope. This process is tedious, time-

consuming, error-prone and relies on human expertise, which is infeasible in remote rural areas (although a well-trained microscopist can differentiate between species of plasmodium). In addition, the sensitivity of microscopy is relatively low at low parasite levels [21], with only 39.3% in sensitivity when using quantitative PCR (qPCR) as the reference [22].

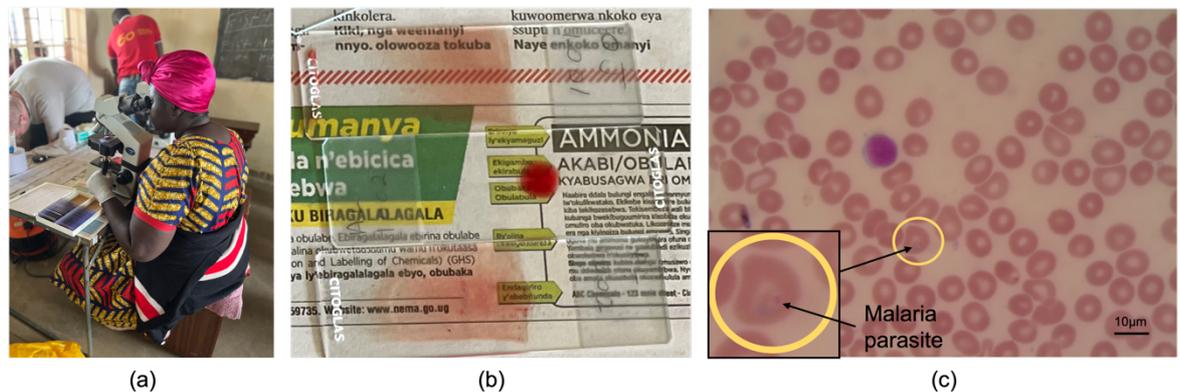


Figure 1-1 Malaria microscopic diagnosis. (a) the microscopic test carried out by the local technician using a brightfield microscope, and (b) the patients' thin and thick blood smears. (c) The parasite lives in a human red blood cell (the cell in the yellow circle); even after 100X magnification, it still can be hard to visualise.

1.2.1.2 Quantitative buddy coat (QBC)

The QBC technique improved and simplified the microscopic diagnosis. In this approach, the parasite's deoxyribonucleic acid (DNA) is stained in a micro-haematocrit tube with fluorescent dyes, e.g., acridine orange, and detected by epi-fluorescent microscopy, a more expensive format to implement than upright (in-line) microscopy [23]. The QBC has been proven as a sensitive and rapid diagnostic test [24], [25]. Tests based on QBC can achieve sensitivity of over 90% and specificity of 95% [26]. Because of its increased sensitivity, it is preferred in epidemiologic studies in the asymptomatic population in endemic areas [27]. Furthermore, portable devices for QBC are available in the market. However, even if it is a reliable and straightforward approach for malaria diagnosis, the cost is much higher than the microscopic diagnostics [28]. QBC malaria test costs about £0.5-£2 per test, while microscopy is £0.01-0.02 per test [28], [29]. The instrument for QBC is, as stated, a more expensive format of equipment, costing ~£7,000 [30].

1.2.1.3 Rapid diagnostic tests (RDTs)

RDT and microscopy are generally considered the most suitable tests for use in low-resource settings. Unlike laboratory diagnosis, immunodiagnostic RDTs do not require a complicated operation, specific knowledge, expensive equipment or even a power supply. The RDTs for malaria are based on detecting malaria antigens in a finger-prick of blood applied to a membrane containing specific anti-malaria antibodies [31].

In Figure 1-2 (b), the process of diagnosis using an RDT is depicted. The first step is mixing the patient's blood with a lysing agent on a test strip or a well to release more parasite protein. Next, the blood and buffer solution can be placed on the strip or in the well and mixed with a dye-labelled antibody, which is specific for the target antigen. The dye-labelled antibody is present on the lower end of the nitrocellulose strip or in a plastic well provided with the strip. Then, due to capillary action, the mixture moves towards the test and control lines. If the target antigen is present in the sample, the antibody-antigen complex will form and accumulate on the test line, while excess-labelled antibodies will accumulate on the control line [32].

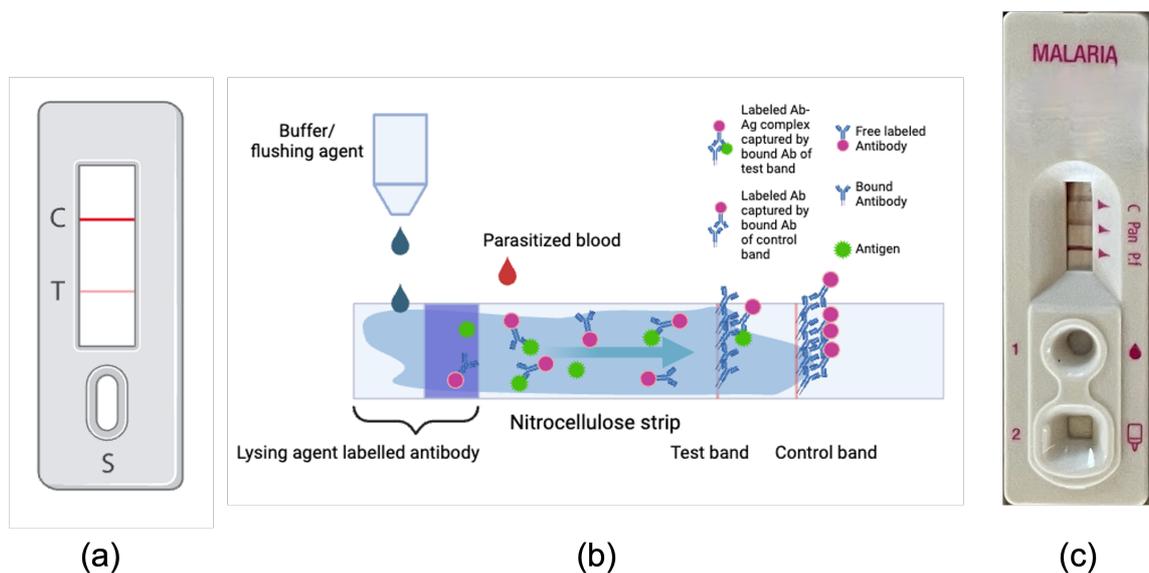


Figure 1-2 Malaria rapid diagnostic test. (a) the schematic drawing of the malaria RDT kit. (b) the illustration of how the malaria RDT works. (c) the Pf-Pan RDT kit used in the field.

Although the RDTs reduce the requirements of running malaria diagnostic tests, there is still concern about their accuracy when the parasite density is low, especially in sensitivity

[33]. According to WHO, the RDTs can achieve a sensitivity over 90% when the *P. falciparum* density exceeds 100 parasites per μl blood [34]. S. Opoku Afriyie *et al.*'s study showed that RDTs can have higher accuracy than microscopic tests. However, when the parasite density is low, its accuracy can drop greatly when using qPCR as a reference [22]. Therefore, sometimes, when the parasite level is low, it needs to be combined with other diagnostic methods, such as microscopy and PCR, to confirm the results. As it is easy to use and cheap, many RDTs have been sold in endemic areas since 2010, and most of them were distributed to sub-Saharan Africa (Figure 1-3).

Most malaria RDTs rely on detecting the histidine-rich protein 2/3 (HRP2/3) of *P. falciparum*. In recent years, the parasite has evolved with a variant that has a deletion of the HRP protein (as a method to avoid detection and treatment). Infections caused by parasites carrying the gene *pfhrp2/3* deletion will result in false negative tests [35]. Therefore, there is a need to create diagnostic tests that have a higher level of sensitivity and specificity as a strategy for controlling the spread of malaria, through diagnosis and treatment.

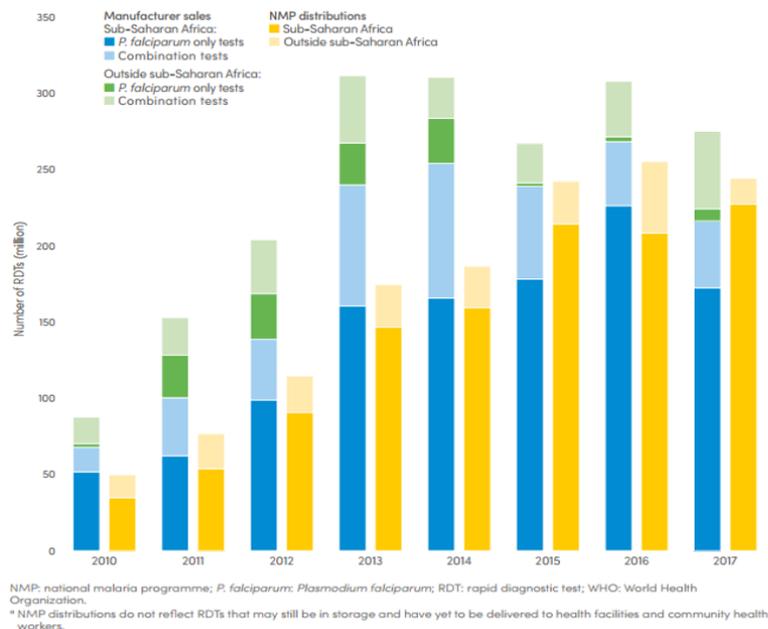


Figure 1-3 Number of RDTs sold by manufacturers and distributed by NMPs for testing suspected malaria cases, 2010-2018 [19].

1.2.1.1 Molecular diagnostic

The developments of molecular biological technologies provide more possibilities for diagnosing malaria. Nucleic acid-amplification-based diagnosis methods have been considered the future of diagnostic techniques. Advantages include using sequence complementarity to improve the specificity, target gene amplification to enable theoretical single copy detection and much more rapid turnaround time [20]. Polymerase chain reaction (PCR) and isothermal amplification are two strategies of nucleic acid amplification methods and can be used in diagnostics.

1.2.1.1.1 PCR

PCR-based malaria diagnosis has been proven to be one of the most sensitive and specific diagnostic methods, especially for low-level parasites, mixed infections, and drug-resistant parasites [23], [36]. It was found to be more sensitive than QBC and some RDTs [23], [31], [37], [38]. Compared with the microscopy diagnosis method, PCR shows higher sensitivity and specificity. PCR can detect as few as 1-5 parasites/ μ l of blood compared with approximately 50-100 parasites/ μ l by microscopy or RDT [31]. As a result, PCR has been considered the best approach to malaria diagnosis. Unfortunately, due to its high cost and complexity, this technology may not be well-suited for point-of-care testing in remote areas where laboratory infrastructure is lacking.

1.2.1.1.2 Isothermal amplification (LAMP)

Loop-mediated isothermal amplification (LAMP) [39] is one of many isothermal nucleic acid amplification approaches that have been developed as an alternative to PCR.

Compared with PCR, the advantage of LAMP is that it is much cheaper and more straightforward, as these methods require no thermal cycling [37]. The amplification of gene sequences in the LAMP reaction is processed at a constant temperature, and it can amplify a few copies of DNA (or RNA) to a large amount in under an hour [40].

Moreover, the LAMP does not require complex equipment to perform the assay (primarily as a consequence of the simplicity in the thermal management). These advantages suggest that LAMP might become an excellent solution for malaria diagnosis in low-resource settings. However, LAMP primer design is complicated and requires a number of primer sequences (making storage in a point-of-care format more difficult). LAMP amplicons

contain a mixture of stem-loop DNA molecules of different sizes, which are not suitable for gene cloning or identifying specific targets based on size differences [39], [41].

1.3 Challenges of running diagnostic tests in low-resource settings

Diagnostic testing has played an essential role in the containment of outbreaks and in preventing unnecessary deaths during pandemics and epidemics [42]. Poor infrastructure and lack of funds are some of the main barriers to introducing diagnostic strategies to LMICs and other areas with limited resources. Products used for disease diagnostics must strictly comply with medical clinical regulations, which could significantly increase the cost. The expensive equipment may be unaffordable and maladapted for use in low-resource settings.

However, some techniques used for high-resource settings could promote the development of tests in low-resource environments. For instance, the cultural isolation of bacteria or pathogens is still the gold standard for detecting most infectious diseases in modern hospitals. Isolation of bacteria is necessary for complete microbe identification, epidemiology, drug-resistance testing, patient triage, and nosocomial monitoring [20], which requires expert knowledge and training; hence, not well suited in low-resource settings. However, new automated systems are being developed to detect the organisms automatically. These systems reduce the cost and personnel training requirements and thus could be adaptable in some low-resource settings.

Including lack of access to expensive diagnostic instruments, there may also be a shortage of well-trained operators, and minimal access to modern medical services in remote, rural/under-resourced settings [43], [20]. Many deaths result by delayed treatment, and many of these cases, death could have been prevented if the disease had been diagnosed at an early stage [43].

1.4 Digital technologies in diagnostics

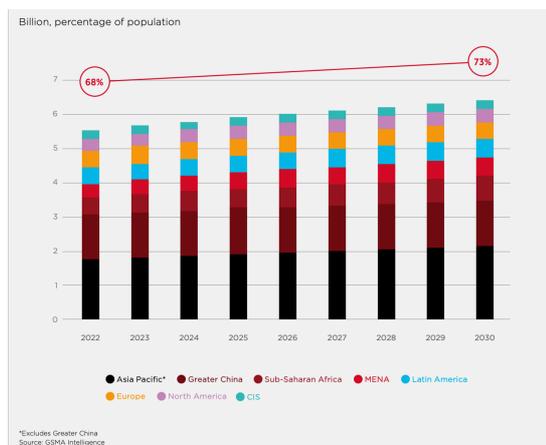
New technologies, including Blockchain, artificial intelligence (AI) and the Internet of Things (IoT), have recently changed how people work and live their lives. Numerous research projects have investigated the application of these technologies to diagnostics and the healthcare industry, such as AI-supported diagnostic systems for medical imaging,

Blockchain-based electronic health records (EHRs), and trillions of IoT devices that have been used for monitoring patients' health status and improving the patients' experience [44], [45]. However, these technologies have not been widely applied across the world, especially in the LMICs. Generally, Blockchain and AI as methods are amenable to use within digital or electronic health systems and can be implemented on mobile phones to provide secure communication of patient data and records as well as diagnostic decision support, in remote areas.

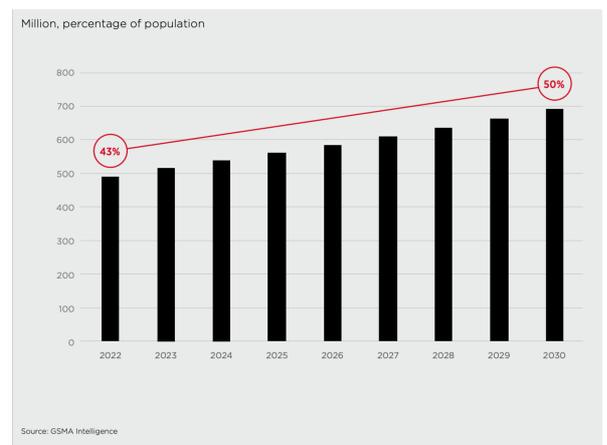
In this study, the smartphone was used to deliver the diagnostic tests in low-resource settings. Combining AI and Blockchain technology, the platform can automatically read out and diagnose the test result and safely communicate or store the test information on the Blockchain network.

1.4.1 Mobile network

Mobile phones are one of the most accessible personal electronic instruments in many developing countries and regions. The number of unique mobile subscribers approached 5.4 billion in 2022 with a 68% penetration rate globally, and this number was predicted to be 6.3 billion by 2030[46]. Although, in many countries the mobile penetration is approaching to saturation, the emerging market still has room for growth. In the 2022 to 2030 period, India and Sub-Saharan Africa will contribute around half of new mobile subscribers globally[46], and by 2030, the mobile penetration rate in Sub-Saharan Africa was predicted to reach 50%[47].



(a)



(b)

Figure 1-4 Mobile subscriptions and penetration (a) worldwide[46] and in (b) sub-Saharan Africa[47] from 2022 to 2030.

The fast growth of mobile subscriptions and smartphone penetration provides an excellent opportunity to deliver diagnostic services to remote villages and monitor the spread of diseases.

The smartphone is not only a tool for communication but also a portable device integrated with multiple sensors and high computational power [48]. The latest smartphone CPU integrated more than 10 billion transistors (the same level as the transistor count of a desktop CPU 10 years ago) [49], [50], which makes it possible to run complex calculations on the mobile device. Combined with multiple sensors, including geolocation (to confirm the site at which a test was performed), as well as cameras (to record patient data or test results), the smartphone can handle as many tasks as most computers [51], providing an excellent platform for mobile-diagnostics (m-diagnostics) or mobile health (m-health).

1.4.2 AI supported diagnostics

AI has been regarded as one of the most promising technologies of the 21st century. The ambition for AI is to design hardware and software systems which are able to mimic the sensing and decision-making capacity of humans. This concept was proposed in the 1950s [52], although because of the limitation of computational power, the extremely high cost of computers, limited available data, and immature algorithms, AI was not developed substantially until the 1980s.

Machine learning (ML) is a branch of AI which uses a large amount of data to teach algorithms to perform specific tasks, such as performing a classification or prediction. The algorithms are trained on a library of data and subsequently are informed by this database in making a decision. An important milestone is the nearest neighbour algorithm that was proposed in 1967 [53]. Other algorithms, such as decision trees, reinforcement learning, and support vector machines, were also developed in the period from the 1980s to the 1990s.

Deep learning, a technology built on machine learning, outperforms other approaches in tasks like image recognition and natural language processing [54]. DL is good at processing high-dimensional data by using deep neural networks (DNN). DNNs consist of

multiple layers, and each layer can extract the features from the input. In a DNN, for example, a deep convolutional neural network, most of the processes are non-linear [54], and with the combination of several layers, the model can better map the input to the output.

In recent years, DNNs have shown great potential in many fields. In medical diagnostics, DNN has been widely used in computer vision (CV) to support doctors in analysing medical images and making decisions [55]. The methods can also help experts work more efficiently and even run some diagnostic tests automatically when combined with hardware.

Zachary S. Ballard *et al.* proposed a deep learning-enabled mobile phone reader to automatically read out the result of vertical flow assay (VFA) for cardiovascular disease (CVD) [56]. Their system benefited from the deep learning approaches in quantification accuracy, quality assurance of assay production, and reduced response error caused by the hood effect [56]. Valérian Turbé *et al.* utilised a convolutional neural network (CNN) for classifying the HIV RDT results; the CNN achieved 97.8% sensitivity and 100% specificity. And compared with the human vision interpretation (by experienced nurses and healthcare workers), CNN was able to reduce the number of false negatives and positives [57]. Liping Huang *et al.* combined deep learning approaches with Raman spectroscopy for liver cancer diagnosis [58]. Thus, it has already been proved that deep neural networks are fully capable of handling medical data, and compared with the traditional computational approaches, deep learning usually shows better performance in both sensitivity and specificity [56], [57], [59].

1.4.2.1 Deep learning for retina vessel segmentation/classification and diabetic retinopathy detection

There are various studies about using machine learning approaches for automated retina vessel segmentation. In 2007, Elisa Ricci *et al.* proposed a method based on-line detectors and a linear support vector machine (SVM) for retinal vessel segmentation [60]. The proposed method achieved 0.9562 (± 0.0001) accuracy on the Digital Retinal Images for Vessel Extraction (DRIVE) [61] database. The DRIVE dataset is a database containing 40 retinal images. Of these images, 33 are healthy, and 7 show signs of certain pathologies.

The images were captured using a fundus camera with a field of view of 45 degrees [61]. This is one of the most famous retina image datasets and has been adopted by many researchers.

With the advance of deep learning-based methods, many researchers have utilised deep learning models for retinal image analysis. For example, Huazhu Fu *et al.* used a fully convolutional neural network and combined fully connected Conditional Random Fields (CRF). Their model achieved 94.72% accuracy on the DRIVE dataset [62]. Further, UNet-based models with higher sensitivity and specificity have been proposed in recent years [63], [64], [65]. This also leads to the creation of more retina image datasets, and an increasing number of annotated datasets are now available. For instance, the MESSIDOR database [66], [67], the ROC dataset [68] and EyePACs dataset [69].

Many studies analysing retina images using deep learning only focus on identifying vessels or classifying healthy/diseased eyes. There is limited effort on semantic artery/vein classification, although the classification could be helpful in detecting vascular diseases. The most common problem in DNN-based semantic classification is that the same vessel could be labelled as different classes [70], and this requires very complicated post/pre-processing steps to correct the vessel labels [71], [72]. Figure 1-5, (a) displays a retina image taken from the DRIVE dataset [61], (b) shows its corresponding label (b), (c) the artery and vein label [73], and (e) and (f) shows the incorrectly labelled vessels in the DNN prediction (d).

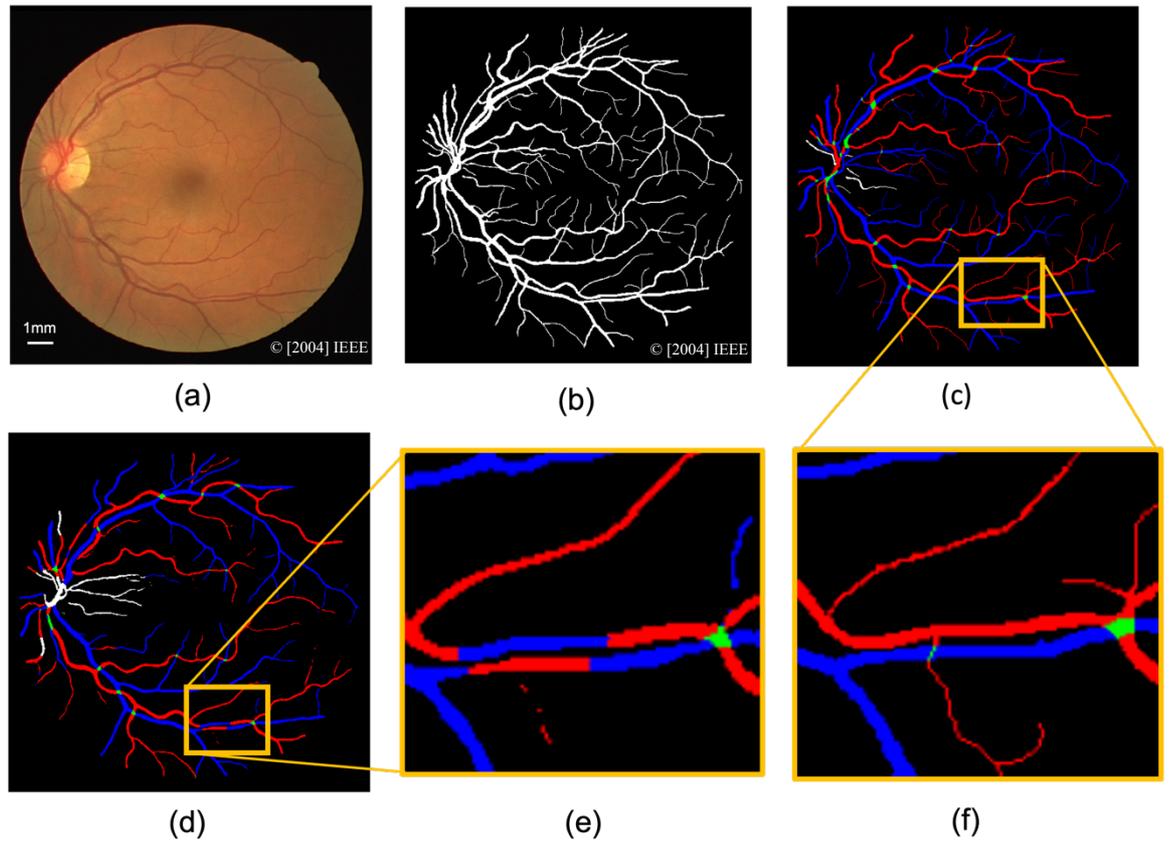


Figure 1-5 An example image from the DRIVE dataset and DNN prediction. (a) the input image (01_test) from the DRIVE dataset and its label (b). (c) the artery and vein label [73] of the same image. (d) is the prediction of a DNN, (e) a section with the mislabelled vessels and (f) the corresponding ground truth.

In this thesis, we now describe how different DNNs can be used for vessel segmentation and classification tasks using ultrawide field retina images, resulting in the development of an approach that combined DL and vessel profiling was developed for automated optical density (OD) analysis.

1.4.3 Blockchain

With the development of the Internet and the computing ability of the hardware, the modern world has become data driven. Data security, privacy, and reliability have become a great concern, particularly where the data collected from diagnostic tests or other healthcare scenarios require heightened privacy concerns.

The concept of Blockchain was put forward by Satoshi Nakamoto in 2008 [74], and the first application built with Blockchain technology is known as Bitcoin [74]. Blockchain has been regarded as a revolution of technology. It is a secured, decentralised system that can record transactions and digital events in the network. Unlike traditional digital transaction systems, no third party is needed. For example, in the traditional banking system, banks are required to maintain and monitor the ledger to prevent invalid transactions or fraud. On a Blockchain network, the ledger is owned by all participants of the network. For example, the transactions are packed in blocks in the Bitcoin system. Each block contains information, including its hash number, the previous block's hash, timestamp, and transaction details. If somebody modifies the ledger, the system will quickly recognise this infringement.

Additionally, the Bitcoin system has a consensus mechanism called proof-of-work, which consumes computers' computational power to solve a cryptographic puzzle, that gives the block's hash the required zero bits [74]. The process of generating a new block usually takes about 10 mins for the miners to solve the puzzle and validate their commitment. This is called the blocktime of the Bitcoin network, the time between two successfully generated blocks. The difficulty of the puzzle could be dynamically adjusted to keep the blocktime stable. This mechanism further increases the network's reliability and makes the ledger tamper-proof.

Blockchain also has shown its usefulness in healthcare industries in recent years. Its characteristics include immutability, consistency, and a high level of trust, perfectly matching the modern healthcare system requirements [75]. From an IBM's study, many healthcare leaders believe that Blockchain technology will have a tremendous impact in improving clinical trial management and regulatory compliance and providing a decentralised framework for sharing electronic health records (EHR) [76], [77]. Moreover, Blockchain could become part of the solution to many important health-related problems, such as forged health records and counterfeited drugs [78]. Most users just need a computer or smartphone with an internet connection to get access to a Blockchain network. No extra hardware is required, with this low-cost feature makes it possible to apply the Blockchain-based system to the whole world.

The Blockchain network proposed in Chapter 3 is based on Hyperledger [79], a Blockchain developing platform created by Linux Foundation and IBM. Although there are decentralised application (DApp) platforms such as Eos [80] and Ethereum [81], Hyperledger provides a more straightforward solution for developers to set up their Blockchain, and no cryptocurrency is required. Additionally, Hyperledger supports REST API and OAuth 2.0, which provides excellent flexibility in developing the end App. There are multiple sub-projects under the Hyperledger framework, and in this study Hyperledger Composer and Fabric were utilised.

1.4.3.1 Types of Blockchain networks

A large variety of different types of Blockchain networks have been reported for use in different applications [82]. All these networks can be divided into three categories, public Blockchain, consortium Blockchain and private Blockchain [83]. The main difference between these is in the extent of the level of trust. Public Blockchain networks are open to everyone with an internet connection, and all the participants are anonymous and untrusted. Consortium and private Blockchains require some level of trust to be shared by participants on the network; invitations are required to join these two kinds of networks, so only permitted users or organisations can acquire access. Table 1-1 shows the comparisons among these three types of Blockchains. Consequently, private and consortium Blockchains are more suitable for companies and organisations.

The public Blockchain network is fully decentralised and highly accessible. Everyone can join the Blockchain network and become part of it, for example, Bitcoin and Ethereum networks. The public Blockchain networks are fully decentralised and are built on strong consensus because no trusted third-party monitors and verifies each transaction. To prevent invalid users from getting access to the network and manipulate transaction data stored on the Blockchain, algorithms like proof of work (PoW), and proof of stake (PoS), have been proposed and adopted by many Blockchains [84].

The private Blockchain is usually utilised in a closed network, e.g., a company or organisation's local network. Only the authorised user can access a private Blockchain network, and external users will not be able to get access to the network or visualise the data on the chain. This type of Blockchain is centralised and built on strong trust. In some

cases, the participants know each other in real life, such as employees within the same company. However, it still operates on multiple nodes as a distributed ledger, all of which are owned by the same entity [85].

The consortium Blockchain network lies somewhere between a public and a private Blockchain network. Unlike public Blockchains with no trust or private Blockchains, all the participants are fully trusted. The consortium Blockchain is usually adopted by several different companies, institutes, or organisations. Only the participants from the consortium are allowed to join the network. Compared to private and public Blockchains, the consortium Blockchain is partially trusted [86], [87], [88]. The transactions in the network will need to be verified by participants from different organisations, which means it could offer better efficiency than the public chain and better security than the private chain.

Table 1-1 Comparisons among public Blockchain, consortium Blockchain and private Blockchain [89] © [2017] IEEE

| Property | Public Blockchain | Consortium Blockchain | Private Blockchain |
|-------------------------|-----------------------------|-------------------------------|-------------------------------|
| Consensus determination | All miners | Selected set of nodes | On organization |
| Read permission | Public | Could be public or restricted | Could be public or restricted |
| Immutability | Nearly impossible to tamper | Could be tampered | Could be tampered |
| Efficiency | Low | High | High |
| Centralised | No | Partial | Yes |
| Consensus process | Permission less | Permissioned | Permissioned |

1.4.3.2 Smart contract

The idea of a smart contract was initially introduced by the Ethereum Blockchain network [81] as a simple program stored on the Blockchain network that can automatically execute transactions when all the predefined requirements are met [90]. Usually, the smart contract is a group of “if...else...” commands running in a network, e.g., in a banking Blockchain

system, if the customer wants to submit a transaction to someone else, the smart contract will check both sender's and receiver's identity, account number, the balance in his/her account for example. Only when all the information matches will the transaction be granted and recorded. The smart contract is the core of a Blockchain network, especially in a public Blockchain network; any flaw in the smart contract might lead to the whole system failure.

1.4.3.3 Challenges and limitations of using Blockchain

Although the Blockchain network has better security compared to the traditional databases and is hard to be tampered with, it is not perfect. The major threat to the safety of a Blockchain network includes the double spending problem, byzantine faults (51% attack), phishing attacks and/or routing attacks. The double spending problem is a system flaw in which the same token has been spent more than once simultaneously. Byzantine fault occurs in some decentralised system, when more than 51% nodes have been taken over, the system will be controlled by the attacker. It is related to the consensus algorithm defined in the system. Instead of attacking the Blockchain network, some hackers will choose to attack the users' wallets. Usually, hackers use fake emails or hyperlinks to fraud Blockchain users. In routing attack, the hacker intercepts the data transferred between the nodes and the internet provider, the user cannot notice any abnormality until their asset been stolen. Consequently, the consensus and smart contract algorithms need to be designed carefully, and other data security protection approaches are needed to further secure the system and improve the trustworthiness.

1.4.4 IoT and edge computing

The Internet of Things (IoT) aims to connect objects in the real world with the Internet. In the context of diagnostics and healthcare, IoT devices have now shown great potential [44]. Prabal Verma and Sandeep K. Sood proposed an IoT-based framework for diagnostics; it could continually screen patients' health and automatically analyse the data collected from the IoT device in the cloud [45]. Alternatively, Priyan Malarvizhi Kumar *et al.* proposed a system that enables IoT and cloud computing for diagnostic purposes. They applied several different classifiers, e.g., K-Nearest Neighbour (KNN), Decision Tree (DT), and a fuzzy classifier, to provide decision support [91]. Compared with cloud-based solutions, an edge computing/fog computing strategy could be another option. As the IoT

device usually does not need to collect complex data, most calculations and analyses can be done locally. This can not only release the calculation power demand from the server but also could benefit privacy protection [92].

1.5 Aims and objectives

This PhD aims to develop a low-cost, easy-to-use diagnostic platform for researchers and healthcare organisations to better understand, monitor and control infectious diseases, such as malaria and schistosomiasis, in low-resource settings.

Current diagnostic approaches often require expensive and bulky equipment to achieve high sensitivity and specificity. For example, in malaria diagnostics, PCR is used as a gold standard. However, a PCR system usually costs over £5,000 and weighs more than 10kg, and well-trained operators are required to perform the test. Although Rapid Diagnostic Tests (RDTs) are widely used as a low-cost alternative, their sensitivity and specificity are lower compared to DNA and RNA-based diagnostic tests.

The Biomedical Engineering group at the University of Glasgow previously developed a LAMP-based microfluidic device for detecting malaria [93], [94]. It achieved a high sensitivity of 98%. The crucial challenges of the provision of using point-of-care medical devices have been highlighted in previous field research [94], including:

1. Bulky and expensive equipment.
2. Complex operating procedures and need for skilled operators.
3. Counterfeited and tampered devices.
4. Lack of trust between local test operators and researchers or healthcare institutes.
5. Data security during collection and sharing.
6. Continuous monitoring and analysing of the disease based on the testing outcomes.

To address the first two challenges, a new hardware device that could perform the test process and record the test result was needed. AI algorithms could help reduce the complexity of performing the test and the operator training cost.

In this PhD, a portable, user-friendly, mobile platform which can process molecular diagnostic assays, e.g., DNA amplification (LAMP), has been proposed and demonstrated. In the list above, Challenges 3 to 4 require approaches that can enhance trust and provide quality assurance to the diagnostic process. Blockchain technology is well-suited for achieving the goals, as it is hard to be tampered with, and the consensus mechanism ensures the trustworthiness of every participant in the network. This technology can help address issues such as fake test records where some tests were not conducted, provide a more secure and real-time method for data sharing, and build trust between local operators and remote researchers and doctors.

Blockchain and AI also have the potential to address Challenges 5-6. Clinical data are always highly sensitive and require a high level of protection. Blockchain systems can provide secure and safe storage for these data, and its immutable feature also maintains trust among the participants. AI could continually analyse the diagnostic test results and identify the outbreak of the disease at the early stage without the need for breaks.

In this study, a platform, including a portable heater for the LAMP-based malaria diagnostics (see Chapter 2), a Blockchain network for secured data storage (see Chapter 3) and a neural network for automated result readout (see Chapter 4), was developed. This platform allows running diagnostic tests in remote areas with limited (to no) resources and enhances the trust between the participants in a remote diagnostic process, e.g., the local test operators and researchers or doctors who might live in another city or country. With the support of AI, the training needs of the local operator are reduced, and it becomes possible to monitor the disease prevalence continually.

As AI shows increased performance in the context of remote diagnostics and telemedicine, it could greatly reduce the difficulty and cost of running diagnostic tests in low-resource environments. In this PhD, AI was also used in the retina image analysis (see Chapter 5). Models were trained for automated retina vessel segmentation and classification, which is a fundamental step of many retinal disease diagnoses.

1.6 Thesis structure

This thesis includes seven Chapters; the first Chapter introduces the background of infectious diseases, especially malaria, in low-resource settings, the commonly used point-of-care diagnosis approaches and the feasibility of adopting technologies such as deep learning and Blockchain in the context of medical diagnostics.

Chapters 2 to 4 propose a mobile-based diagnostic platform for malaria diagnostics. The platform consists of a mobile-controlled heater, an Android App and a Blockchain network. Also, deep neural networks were trained and integrated into the App for decision support.

Chapter 5 proposes an additional use of deep learning in automated retina image analysis. In this Chapter, the models were trained for retina vessel detection and artery vein (A/V) classification.

Chapter 6 introduced a group project during the COVID-19 pandemic. In this project, a web App for sample size calculation in diagnostic tests was developed and deployed. This work was performed during the “lock-down” period when there was limited access to the laboratories.

Chapter 2 Mobile phone-controlled heater for LAMP test

2.1 Introduction

Traditional clinical diagnosis of malaria is based on the patient's symptoms, such as fever, dizziness, and diarrhoea. Although this is the cheapest and most widely used method in the world, the most critical problem with this symptom-based diagnostic is that these are nonspecific and overlap with other common febrile diseases- leading to misdiagnosis, over-diagnosis [31], overtreatment, and ultimately the spread of drug resistance through the prescription of incorrect treatments and/or wrong doses.

In this Chapter, an instrument was designed as a low-cost device controlled by an Android mobile phone. Integral to the delivery of the DNA-based diagnostic assay was LAMP amplification, performed by integrating paper microfluidics within low-cost disposable cartridges, as described, and validated in the previous work of our group [95]. In this previous work, the heating was controlled manually, using a dry bath connected to power provided by a diesel generator or by solar power. Although the latter is available everywhere in remote settings in SSA, it requires expensive equipment and cannot be used at night, whilst the former is more inconvenient still and has a higher carbon footprint. However, it enables to perform multiple assays in a single device.

Consequently, in this thesis, a standalone heater was designed and demonstrated, powered, and controlled by a mobile phone. The development of the heater, including the software and control circuit, was carried out by myself as part of this PhD study. The temperature was maintained using a control circuit (circuit diagram provided in Figure 2-2), a micro-controller unit, two temperature sensors (one acting as reference, one measuring the cartridge temperature) and a heating unit. The heater was enabled either using the mobile phone On-To-Go (OTG) functionality or with a standard 5V battery power pack (through a micro-USB port and a voltage regulator LM317T). The body of the device was designed by using Autodesk Inventor and 3D printed using acrylonitrile butadiene styrene (ABS) plastic. This formed a small, lightweight (overall ~ 500g), low-cost(the prototype costs around £50. See Appendix Table 1), and long-lasting instrument (Figure 2-1). Relevant files, including code, 3D models, and circuit diagrams, were published and are available

on Zenodo (<https://doi.org/10.5281/zenodo.4429293>) and GitHub (<https://github.com/XGuoo/BlockchainDiagnostics>).

2.2 Multiplex LAMP system

The primer sets used for the LAMP assay were based on previously published [95] primer sequences for *P. falciparum*, whilst primers for a BCRA1 gene fragment were used to implement an assay which served as a positive control. The primers were all purchased from Eurofins Genomics. The reactions were amplified for no more than 45 min at 65°C.

Field testing was carried out in Uganda and followed the same protocol as previously used [95] to demonstrate the platform's functionality in the field. Briefly, the local technicians collected and tested blood samples from 50 school children from Kocoge Primary School in Tororo District. The mobile heater mentioned in this chapter had been tested by me and used for LAMP-based malaria diagnostic tests by my colleagues. This study was conducted as part of the activity undertaken by the Vector Control Division (VCD) of the Ministry of Health (MoH) in Kampala, Uganda, on neglected tropical diseases and was approved by the Vector Control Division (MOH) Research and Ethics Committee, VCDREC/078 and Uganda National Council for Science and Technology (HS 2193). No personal data were revealed to the investigators. Written informed consent was also obtained from the children's parents and the head teacher (see protocols in the previous study for details). All samples were also tested retrospectively by RDT and microscopy in the field, as well as by PCR, on our return to the UK [95].

Ethics approval was upon the basis of presumed positive, given the high prevalence of the disease, which is endemic in the region. All individuals were treated accordingly following the approved tests (RDT and microscopy) under the MoH of Uganda's guidance. Analysis was always double-blinded between on-site field testing and reference tests. After testing at the school, all used paper devices and small plastic consumables were incinerated for their disposal. In contrast, glass slides and RDTs, used as reference techniques, were stored in a biohazard container for safe disposal at the VCD, Kampala.

Analysis was performed in the children's classrooms, where there was no access to power or running water. For each individual, a finger-prick (~5 µL) of whole blood was used,

with sample processing including sample lysis, DNA extraction and amplification performed using the paper ‘origami’ protocol, as previously published [95].

2.3 Material and methods

The mobile-controlled portable heating system for the LAMP test consists of two main parts: a 3D-printed portable heater and a smartphone App which can run on any phone with an Android operating system (Android version > 5.0). The heater includes both hardware and software. It used a microcontroller unit (MCU) to control the heating process and transfer data with the mobile phone. The circuit diagram of the heater can be found in Figure 2-2. The software of the MCU was written in C language and compiled by Arduino IDE.

2.3.1 Mobile heater

The heater used a 10000mAh (38Wh, 5V, 1A output) battery pack as its power supply, and a Bluno Beetle (from DF Robot) was used as the MCU. Figure 2-1 shows the overall design of the heater. The enclosure of the heater comprised three parts, namely, the main body, a slide cover with a mobile phone holder and a top cover. There are two layers in the main body of the enclosure. The bottom layer was designed to contain the battery pack, and above that is the space for placing the circuit. The top cover was used to enclose the heater to reduce heat loss. Additionally, a mobile phone holder was integrated into the cover for easy monitoring of the heater’s status during testing. Simply place the mobile phone on top of the holder and keep an eye on the temperature and timer. Figure 2-1 provides an overview of the overall design of the heater.

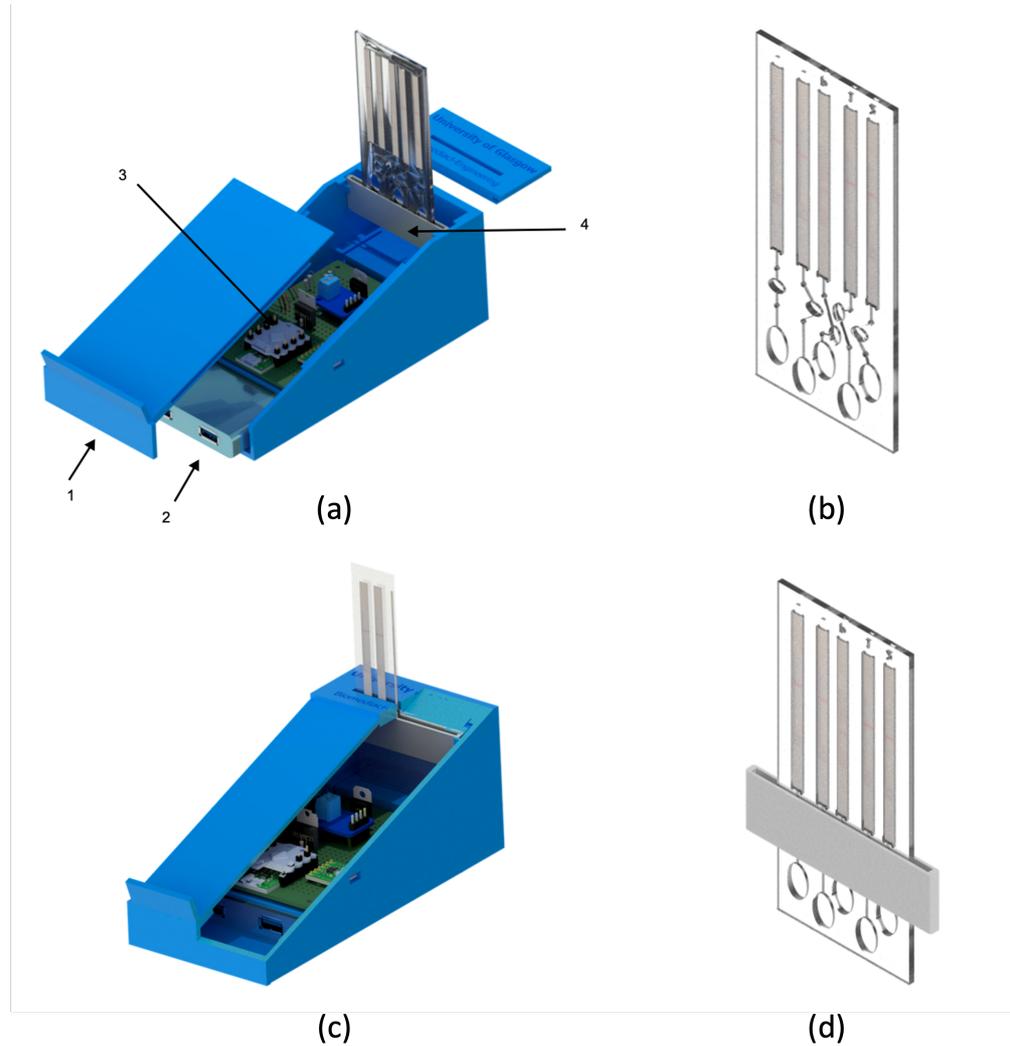


Figure 2-1 Heater design[96]. (a) Exploded view of the heater. The numbered parts are: 1 the cover of the heater with mobile phone holder; 2 the power bank; 3 the internal circuit; 4 the bent aluminium board. (b) Three-quarters view of the heater. (c) Multiplex diagnostic LAMP device. (d) Multiplex diagnostic LAMP device with aluminium loop attached onto the reaction chambers.

2.3.1.1 Material

The enclosure of the heater was 3D printed by Stratasys F170 printer with acrylonitrile butadiene styrene (ABS) copolymer.

The circuit used the following components: a solderable breadboard, Arduino, N-channel metal-oxide-semiconductor field-effect transistor (MOSFET)((IRLB8721), light-emitting diodes (LEDs), micro -USB female connector, an ultra-low-power thermoelectric

generator (TEG) (Peltier Module, 0.76 W, 600 mA, 2.5 V, 15 x 15 mm), an aluminium board, an AD8495 Thermocouple Amplifier (from Adafruit), K Type thermocouple wire, voltage regulator (LM2595/LM317T), and other small electronic parts.

2.3.1.2 Method

The enclosure of the heater was designed with Autodesk Inventor 2019. The heater used an aluminium band around the LAMP reaction chambers in the cartridge (numbered 4 in Figure 2-1 (a) and Figure 2-1 (d), to enhance thermal transfer and ensure homogeneity of temperature across the device.

Chapter 2 Mobile phone-controlled heater for LAMP test

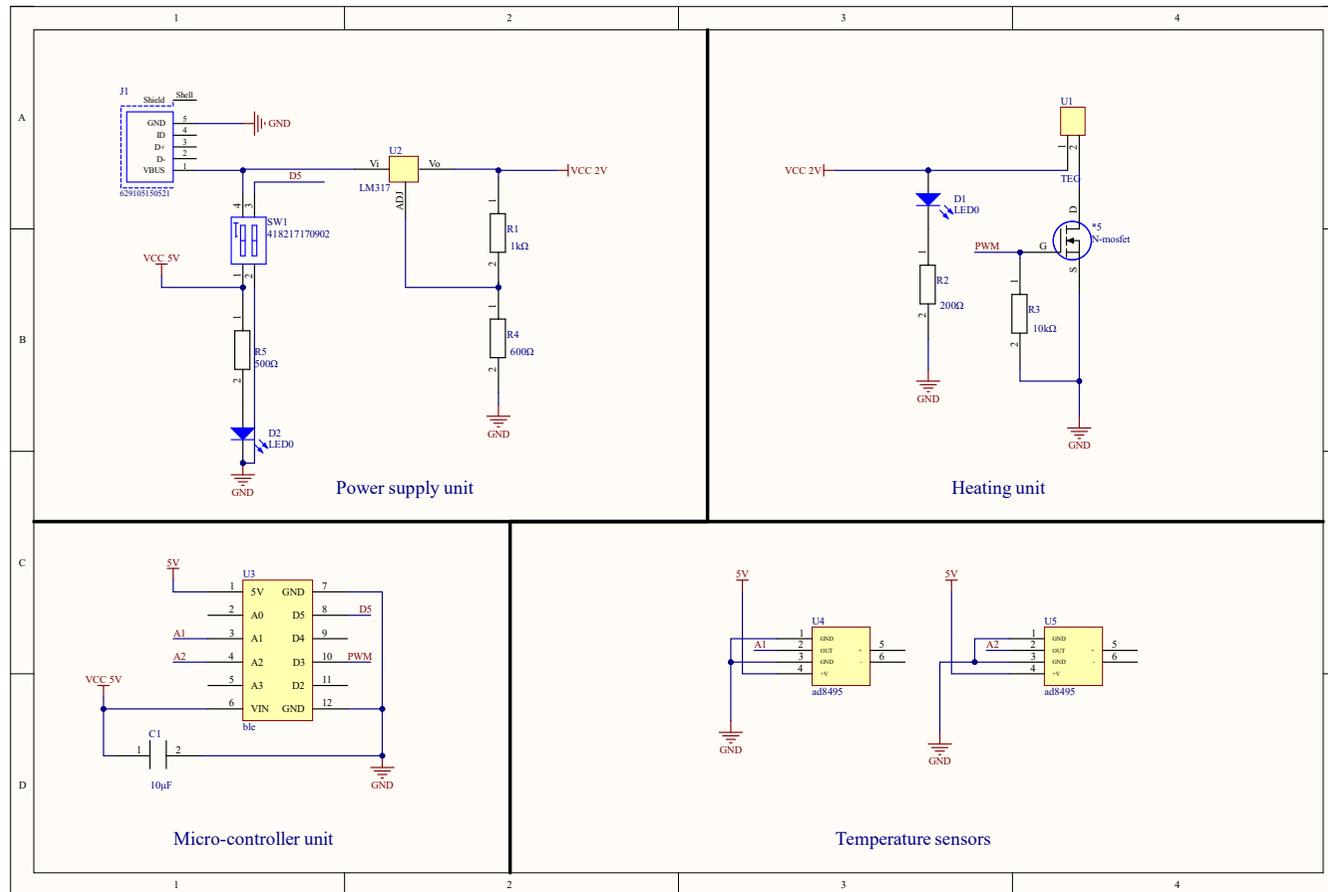


Figure 2-2 Circuit schematic diagram [96] The heater's circuit consists of four components: power supply, heating, microcontroller, and temperature sensor unit. Functionality is detailed in the main text.

2.3.1.2.1 Power supply unit

To maximise the operation time and reduce the complexity, the heater used a standard power bank (5V output) as its power source. The power supply unit of the circuit includes a female micro-USB port for connecting to the power bank. There are five pins in the typical micro-USB port: GND, IO, D+, D- and VBUS. Because there was no data transfer, the IO, D+, and D- pins were left empty. The VBUS pin could be regarded as the positive electrode of the power supply, and it provided a constant 5V voltage. The ground (GND) provided a 0V reference for the whole system.

The power supply unit needed to provide two different voltages for the MCU and TEG, the MCU requires a 5V input, and the input voltage of the TEG should be <2.5V. Two different voltage converters had been used for those purposes (Figure 2-3).

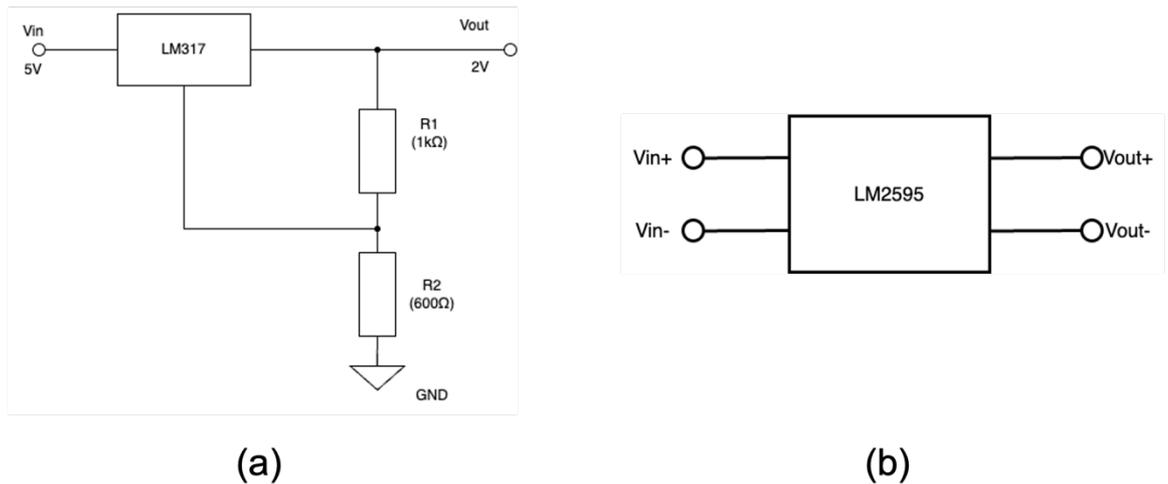


Figure 2-3 Voltage converters. (a) LM317(T) LDO convertor (b) LM2595 based Buck convertor.

LM317 is an adjustable 3-terminal positive voltage regulator [97], providing up to 1.5A current. The output voltage can be easily adjusted by the resistors that connect with its ADJ and the output pin (R1, R2). The output can be calculated by Equation 1.

$$V(out) = 1.25 * \left(1 + \frac{R_2}{R_1}\right) \quad \text{Equation 1}$$

In this study, the output voltage was set to 2V. In order to prevent any significant current flow through the resistors, the resistance of R1 and R2 were set to 1k Ω and 800 Ω , respectively.

Although LM317 performs well in the circuit, it generates too much heat, as has been seen in other LDO regulators. Consequently, the LM317 was replaced by a step-down (buck) regulator. The buck regulator converts the voltage by switching the circuit on and off to generate a PWM signal and uses a diode, a capacitor, and an inductor to regulate the signal to a stable output. However, the output voltage ripple could slightly affect the stability of the circuit. Figure 2-4 shows the basic principle of the buck converter.

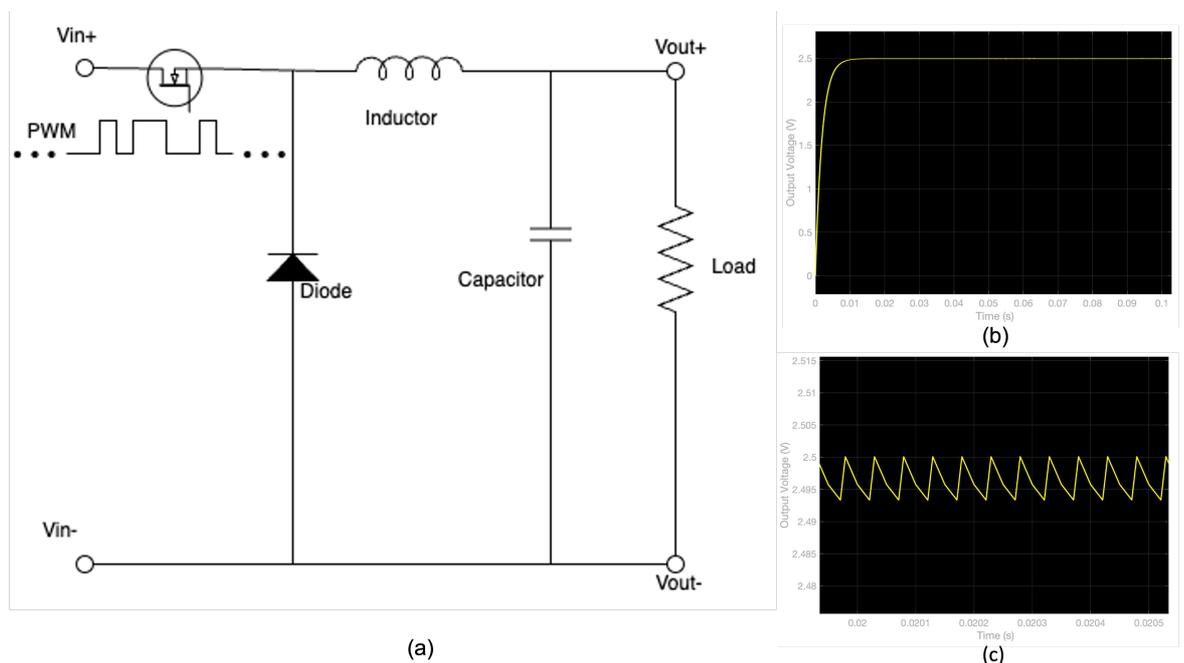


Figure 2-4 The basic principle of the buck converter. (a) the buck converter, (b) the simulated output, (c) the waveform of the output voltage ripple.

2.3.1.2.2 Heating unit

The heating unit used a Peltier as the heat source. Its compact size and flat surface make it easy to attach to the aluminium surface, allowing for better heat conduction. At low power, the energy consumption of a Peltier and a resistive heating element, the Peltier device can also be utilised for cooling, which allows for an upgrade of the heater for PCR-based tests in the future. A N-channel MOSFET performed as the switch. The drain and source terminals were connected to the TEG and ground, respectively, and the gate

terminal was connected to a digital pin from the MCU. The TEG will be switched on/off based on the Pulse Width Modulation (PWM) signal generated by the MCU. A 10k Ω resistor was placed between the gate terminal and the ground for circuit protection.

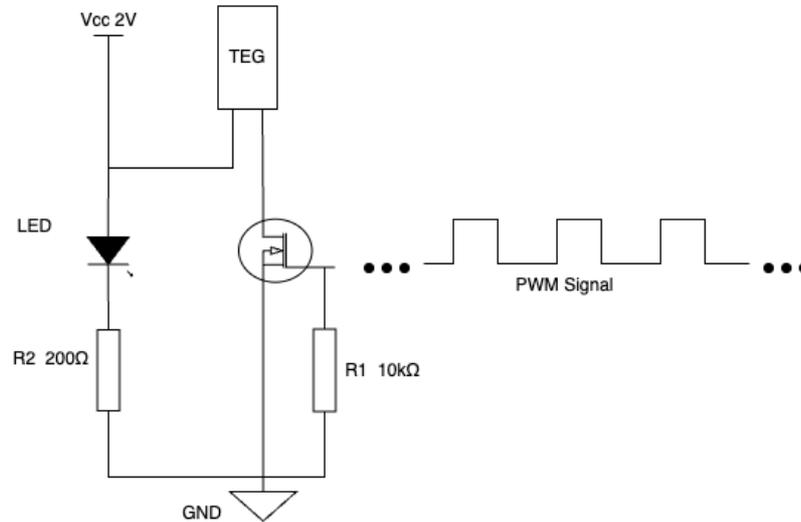


Figure 2-5 The TEG control circuit. The Vcc is the output of the voltage converter. The LED is used for displaying the work status of the TEG power supply. The n-channel MOSFET was used as a switch to turn on/off TEG. The PWM signal is provided by the MCU (Bluno).

2.3.1.2.3 Temperature sensor unit

The portable heater used two K-Type thermocouples and an AD8495 thermocouple amplifier as temperature sensors. Compared with other types of thermocouples, the advantages of K-Type thermocouples include low-cost, tolerance of oxidised environments, fast response, and reliability.

In the heater circuit, one thermocouple was used for reading the temperature of the TEG in real time and another for self-calibration. The mobile heater used a 1mm thick aluminium board for heat conduction. If compared with the heating block in the traditional LAMP heater used in DNA diagnostics, the heating area is limited, and the temperature in the channel could be lower than the TEG. Consequently, a calibration device was made to calibrate the target temperature and compensate for the heat loss (Figure 2-6 (a)).

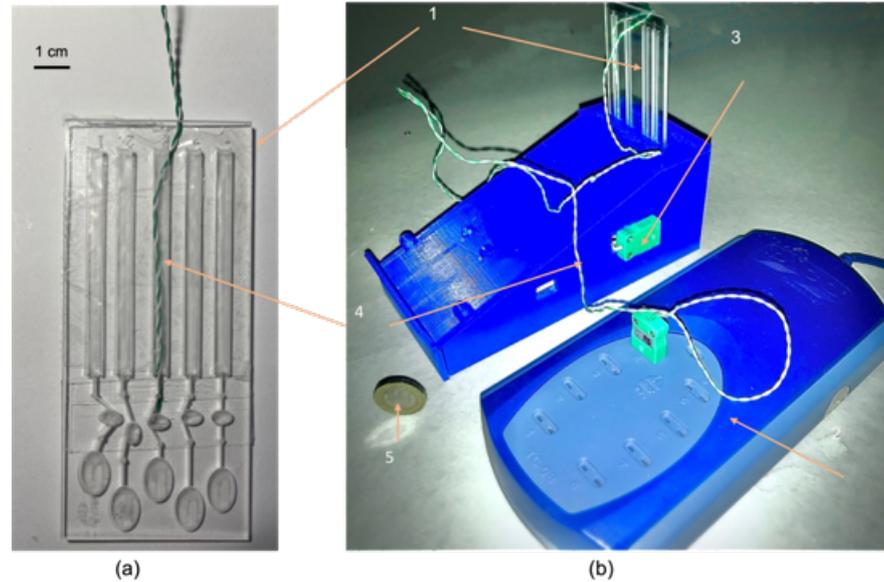


Figure 2-6 Temperature calibration device. (a) The thermocouple was inserted into the microfluidic device (1) chamber (4), where the heat will be supplied during amplification. (b) The calibration device in use. (2) is the Pico Logger, (3) is the female thermocouple connector for heater self-calibration, and (5) is a 1-pound coin for showing the device size.

2.3.1.2.4 Microcontroller Unit

In order to carry out a LAMP-based diagnostic test, it is important to maintain a stable temperature. Typically, the required temperature range for LAMP is between 60°C and 65°C [98], Deviation away from this predefined regime can cause inefficiencies in the amplification and poor diagnostic performance.

The heater enabled the control of the temperature of the LAMP reaction chambers embedded within the plastic microfluidic chip by using the proportional, integral, and derivative (PID) control mechanism in the Arduino code. The temperature control was achieved by adjusting the duty cycle of the output PWM signal and using an N-MOSFET as the switch. The PID control mechanism can be formulated as follows:

$$u(t) = K_p e(t) + K_i \int_0^t e(t^2) dt^2 + K_d \frac{de(t)}{dt} \quad \text{Equation 2}$$

The $e(t)$ represents the error between the actual value (temperature) and the target, P stands for the proportional gain of the system, I is the integral gain, and D is the deviate gain. The K_p , K_i , and K_d are the coefficients of each type of gain. It is necessary to set appropriate coefficients to ensure the heater functions correctly. Coarse tuning can be done by adjusting K_p (proportional gain), a higher K_p can increase the system response speed; however, if it is too high, it might lead to overshoot and oscillation. K_i (integral gain) can reduce the steady-state error of the system by integrating past errors over time. A proper K_d (deviate gain) can help the system stabilize faster by taking the error changing rate into account. In this study, the PID gain coefficients of each heater were manually adjusted.

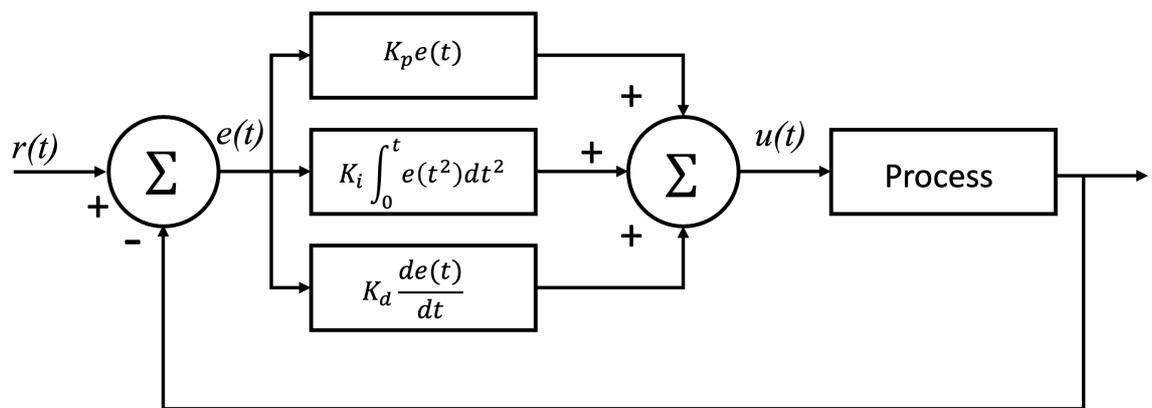


Figure 2-7 The PID control mechanism. The $r(t)$ represents the error between the target and real-time temperature. The P, I and D will contribute to the PID value (0-255) based on the $e(t)$, and the PWM signal's duty cycle will be adjusted.

A timer was integrated into the MCU program, in order that the system could automatically switch off the heating unit when the countdown to zero. The timer could be set up through the mobile App, described in next section.

2.3.2 Android app

The Android App was designed with Android Studio in Java. The minimum requirements for installing this App are Android version >5.0 and support for Bluetooth 4.0. Bluetooth Low Energy (BLE) was used for establishing connections with the Bluno chip, as part of Bluetooth 4.0. The phone with a lower Android version may be unable to install the App properly because some dependencies require an Android API level >21, and this Android version improved the support of BLE [99]. The App only has one page, as Figure 2-8 shows.

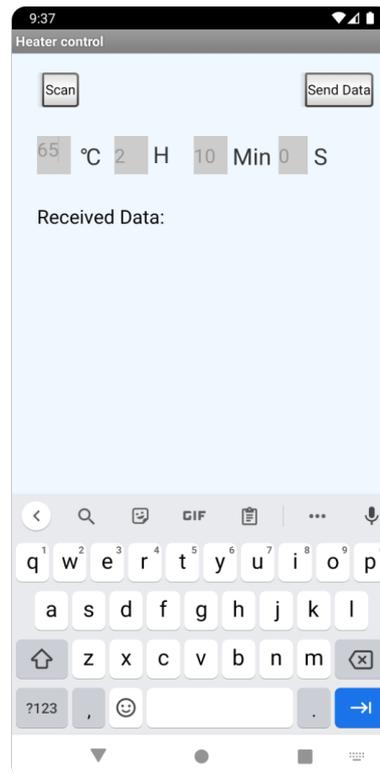


Figure 2-8 The user interface of the mobile App. There are two buttons on the screen. By clicking the “Scan” button, users can see the Bluetooth devices around them. Click the one named “Bluno,” and the App will connect with the mobile heater. Then, the user can enter the target temperature, set up a timer from the edit box, and click “Send Data”, and the settings will be sent to the heater. Once the App is successfully connected to the heater, the real-time temperature and a timer will be displayed in the received data area.

The Blockchain Diagnostics App was developed based on the official demo App BlunoBasicDemo from DFRobot [100] by customising the input data format. By using the input boxes, users can set the target temperature and create a timer, which can then be sent to the heater with a click of the top right button. The App used a Bluno Library, which includes Bluetooth low energy (BLE) services to manage the Bluetooth activities, and a ring buffer was used to handle the data flow (timing and temperature). More functionalities were developed and integrated to this App in subsequent updates, see Chapter 3 and 4.

2.4 Results

2.4.1 Laboratory testing

The heater was tested in the laboratory with different target temperatures (40°C, 65°C, 75°C and 90°C). The test results are shown in Figure 2-9. The mobile heater can supply heat for over 12 hours at 65°C, with a 10000mAh (36Wh) battery pack. The error of the actual temperature and target was controlled within 0.5°C.

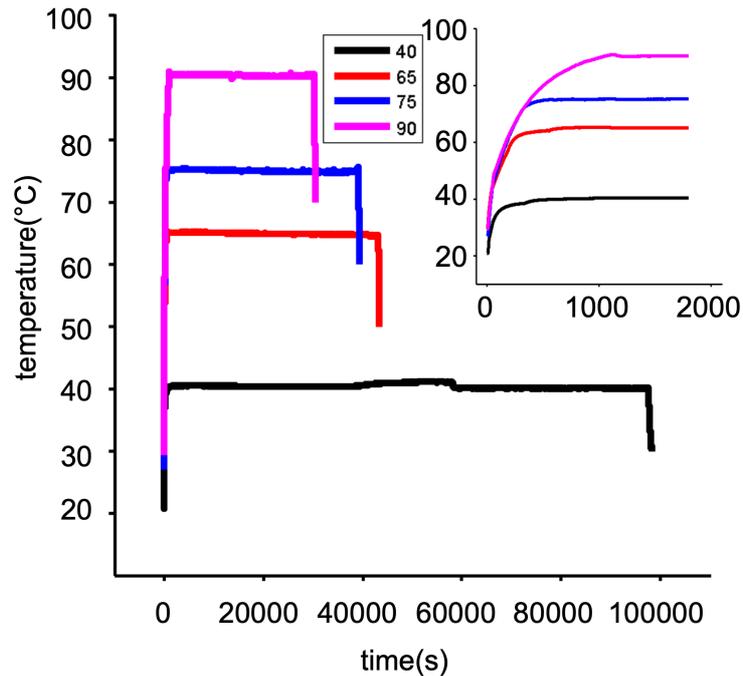


Figure 2-9 The mobile heater was tested at different target temperatures [96]: 90°C (purple), 75°C (blue), 65°C (red, the temperature most commonly used for LAMP) and 40°C (black). The X-axis represents time (in seconds, over 24h), and the Y-axis represents temperature. The temperature decreases when the battery becomes limited, providing the capability for >12h of LAMP reactions. The inset shows the temperature ramping up, demonstrating the control of the PID algorithm. Heating to 65°C took 10 minutes, providing the ability to run a full LAMP assay in under 1h (including sample processing [95]).

2.4.2 Field testing

The mobile heater was tested in Kocoge Primary School, Uganda. The heater successfully provided 64°C ($\pm 0.5^\circ\text{C}$) in the field environment shown in Figure 2-10. The temperature

was recorded by the temperature logger (PicoLog TC-08).

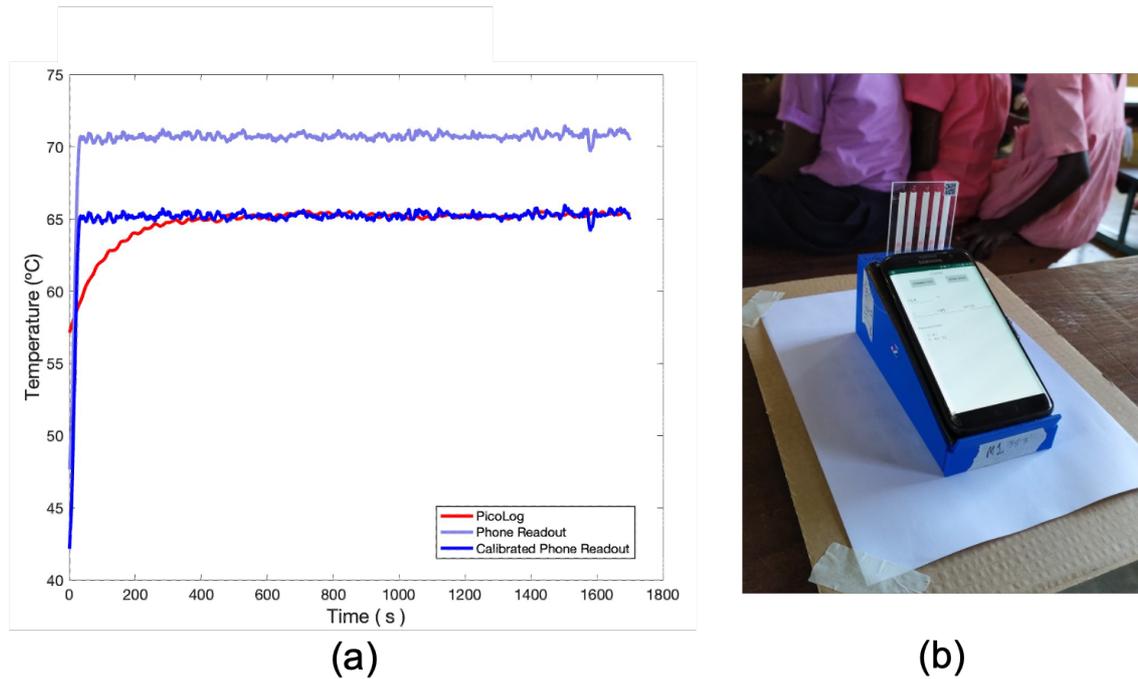


Figure 2-10 Field test result. (a) The temperature log obtained from the field. The x-axis shows the time in seconds, and the y-axis is the temperature. The red line is the temperature measured by the PicoLog temperature logger, the blue line with 50% transparency shows the temperature reading from the App without calibration, and the solid blue line represents the calibrated temperature reading from the phone. (b) The heater performs the LAMP test in the field.

2.5 Conclusion

This Chapter presented a 3D-printed heater for the LAMP microfluidic malaria test, as well as the development of an App to control the testing. It meets the requirements of being portable, affordable, and user-friendly. The heater used ABS material for the body, and a simple circuit was used to achieve its functionality. Also, an Android App was developed for controlling the heater.

The overall cost of the heater was estimated to less than £50, with these found in Appendix Table 1. Compared with the bulky traditional heater and heating blocks the weight was reduced significantly, only about 500g (including the power bank), which can be easily transported and held in the hand.

The heater used an Arduino-based microcontroller with Bluetooth as the central processor. Through the PID control mechanism, it can maintain a stable temperature within 1°C accuracy for the LAMP malaria test, whether in a laboratory or field environment.

The portable heater was designed to work without an external AC power source, which makes it a useful option for areas without access to electricity. It can be powered by a standard power bank with 5V output or a mobile phone with an OTG function. A 10,000mAH power bank can provide over 12-hour battery life.

Chapter 3 Blockchain network for diagnostic data management

3.1 Introduction

Since the concept of Blockchain and the first decentralised application was proposed in 2008 by Satoshi Nakamoto [74], Blockchain technology has changed many industries. Today, many companies have adopted Blockchain in their daily business, for instance, IBM, Nvidia, Citi Bank, and Coca-Cola. Companies can use Blockchain networks for different purposes, such as quality assurance, supply chain management and cybersecurity.

Compared to the traditional database system, the Blockchain has some significant advantages, including being hard to tamper with, trustworthiness, safety, and privacy. The data in a Blockchain network are kept in blocks, and each block includes four parts, header hash, the hash of its previous block, transaction details and a timestamp. Figure 3-1 shows the structure of the Blockchain.

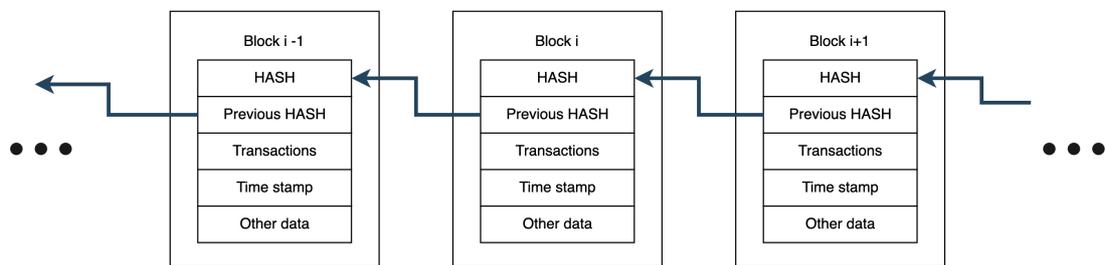


Figure 3-1 The Blockchain architecture. Each block has its own hash number, the hash number of its previous block, a timestamp, transactions, and other data such as the Node version or geolocation.

The immutability, safety, trustworthiness, and decentralised nature of Blockchain make it an ideal storage solution for sensitive clinical diagnostic data. And it can benefit scenarios such as remote diagnosis and data sharing between organisations. Data security and safety are critical concerns in remote healthcare and diagnostics. The Blockchain-based database perfectly meets these needs. Moreover, Blockchain can enhance trust when multiple organisations and participants are involved. Every transaction can be traced back, which also makes it suitable for continuous disease surveillance.

The Blockchain application mentioned in this Chapter was developed using Hyperledger Composer, which offers a simple way for developers to define the function and logic of the Blockchain application. The application operated on a Hyperledger Fabric Blockchain network.

3.1.1 Central concepts in Hyperledger Composer

There are some basic concepts and terms in Hyperledger Fabric and Composer Blockchain networks, which are defined below, including:

- Business network archive (BNA): the application running on the Blockchain, including a model file, a JavaScript file, a query file, and an access control file (Figure 3-2).
- Business network card: it provides all the necessary information for connecting to the business network, such as connection profile, private key, and certificate, in an encrypted format.
- Asset: in a business network, the asset could be anything that has value, e.g., goods and services. In this study, the asset is the diagnostic assay.
- Transactions: the interaction between participants and assets, such as buying and selling products. In this project, the transaction changes the assay information, e.g., the manufacturer uploads new device records to the Blockchain network.
- Chaincode: a program running in a secured docker container [101] to manage the ledger states and transactions just like the smart contract (introduced in Chapter 1).

3.2 Methods

3.2.1 Blockchain network

The Blockchain network, shown in Figure 3-2, was based on the open development toolkit Hyperledger Composer and Fabric, and the network was hosted on a Google Cloud server. The core of the Hyperledger composer Blockchain network was a business network archive (BNA), including a model, script, access control, and query file. The BNA was deployed to an existing Hyperledger Fabric runtime (inc. Fabric ordering service, certificate authority and peer nodes). Users needed to use a peer card containing public and private keys to obtain access to the Blockchain network.

Users can access the Blockchain from their desktop or mobile phone through a web browser or a customised Android app. This is made possible by the support of REST API and GitHub OAuth authentication. The database service in the cloud allowed a central point to collect information, enabling later analysis on geo-tagged disease propagation in the communities, with a certain point accessible by healthcare providers across the hierarchy of the healthcare system. The anonymisation of information in this database ensured privacy. At the same time, trust in the recorded data was always maintained, significantly improving the endorsement and privacy aspects, compared with either the manual or email transfer of records.

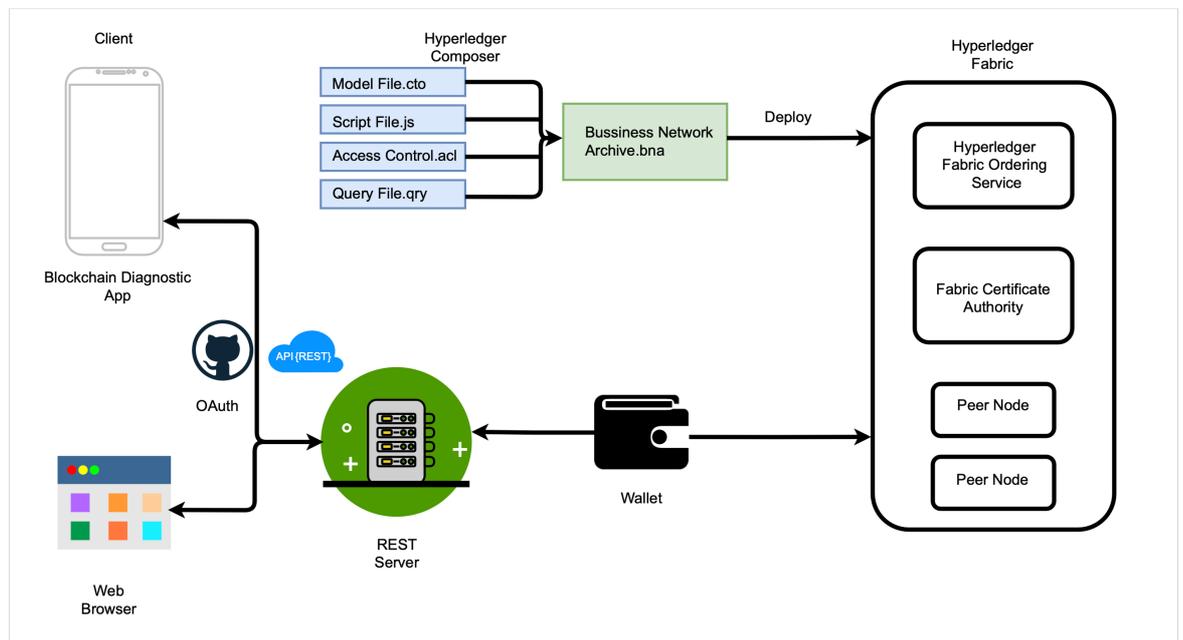


Figure 3-2 The system architecture of the diagnostic Blockchain. [96] The client can use the mobile App or a web App to access the Blockchain. GitHub OAuth was used for authentication. The business network was defined in a Hyperledger Composer business network archive file (BNA) and deployed on the Hyperledger Fabric network. The authenticated user can use their private and public key to interact with the Blockchain via REST requests.

3.2.1.1 Blockchain model

The Hyperledger composer model file defines the essential components of the business network, such as participants, transactions, assets, and events. The assets (which in our case is the microfluidic diagnostic device), the participants (manufacturers), the operators (as the healthcare workers involved in the delivery of the diagnostic assays and their

analysis) and the transactions (as connections) were all defined in the BNA file (Figure 3-2).

The diagnostic device was addressed with a unique identifier (ID), and the related information (including, e.g., date of manufacture) was printed as a quick response (QR) code on the device (Figure 4-8). Participants have their ID and username stored on the Blockchain (i.e., the ledger). The role they can play is limited by access control, although they can create a new device record or update a piece of device information.

3.2.1.2 Transaction

The assay tracking network has two defined transactions: one for adding new device information to the Blockchain (ProduceDevice) and another for updating diagnostic test results (DoTheTest). Once each transaction is successfully submitted, a corresponding event will be emitted. The transaction algorithms are shown below:

| Algorithm 1. ProduceDevice [96] | Algorithm 2. DoTheTest [96] |
|--|---|
| <p>Input: device ID, test name, manufacturer, date of manufacture, expire date, batch number, production place, status</p> <p>Result: Add new device record to the Blockchain</p> <p>If device exists, then return</p> <p>else set test name, participant (manufacturer), date of manufacture, expire date, batch number, production place, status to device attribute</p> <p>get asset registry</p> <p>emit ‘Produce Device’ event</p> | <p>Input: device ID, status, operator, test date, patient ID, gender, weight, URL (link to image of device after test), result, geo-location</p> <p>Result: Add test information to existing device</p> <p>If device does not exist, then return</p> <p>else update status set operator, test date, patient ID, gender, weight, URL, result, test place to device attribute</p> <p>emit ‘Do The Test’ event</p> |

The “Produce the device” function allows the manufacturer users to add new device records to the Blockchain network. The information, including the device ID, manufacturer, batch number, and device status (new / used / invalid), will be stored. The “Perform the test” function was designed for the operators to upload the test details and patient information.

3.2.1.3 Access control and authentication

Access control and authorisation strategies were used to improve the system’s security. In the access control file, which is a part of the BNA, different participants were given a different level of access to the Blockchain resources. The manufacturers can only submit “ProduceDevice” transactions, the operators of the diagnostic test were only allowed to submit “DoTheTest” transactions, and analysts could visualise all the transactions and data that had been stored on the Blockchain network but could not add new information to the Blockchain.

3.2.1.4 Authentication

GitHub OAuth was used for authenticating the user’s identity. A token (e.g., f9cdd11f6b7e4fcd6c8316ae87bd5e1d3459314e) could be generated for each participant, and this token is linked with the user’s network card (including certificate and private key), which been used to launch the rest server.

3.2.2 Android app

The Hyperledger Blockchain network supports REST API, which allows users can access to the Blockchain network by using standard HTTP [102] requests. Thus, the Android App used for heater control was updated to enable communication with the Blockchain network. The data flow of the new App can be found in Appendix Figure 2

Figure 3-3 shows the updated UI of the App. A login page (Figure 3-3 (a)) was added to the App to enable access control, and the users’ login details (User ID and Password) are associated with the GitHub OAuth token. Thus, different users can easily access the Blockchain network via mobile phones.

Multiple screens and more complex logic were implemented in this mobile App version. After logging in, different users will see different pages, depending on their role in the network (manufacturer, operator, analyst). Manufacturers can only add new devices to the Blockchain network, while operators can use the App for heater control and upload test information to the Blockchain. It is important to note that operators must scan the QR code on the device before conducting any tests. This step was used to verify the assay status and prevent counterfeited devices. Analysts can only view and search for data on the Blockchain via the App.

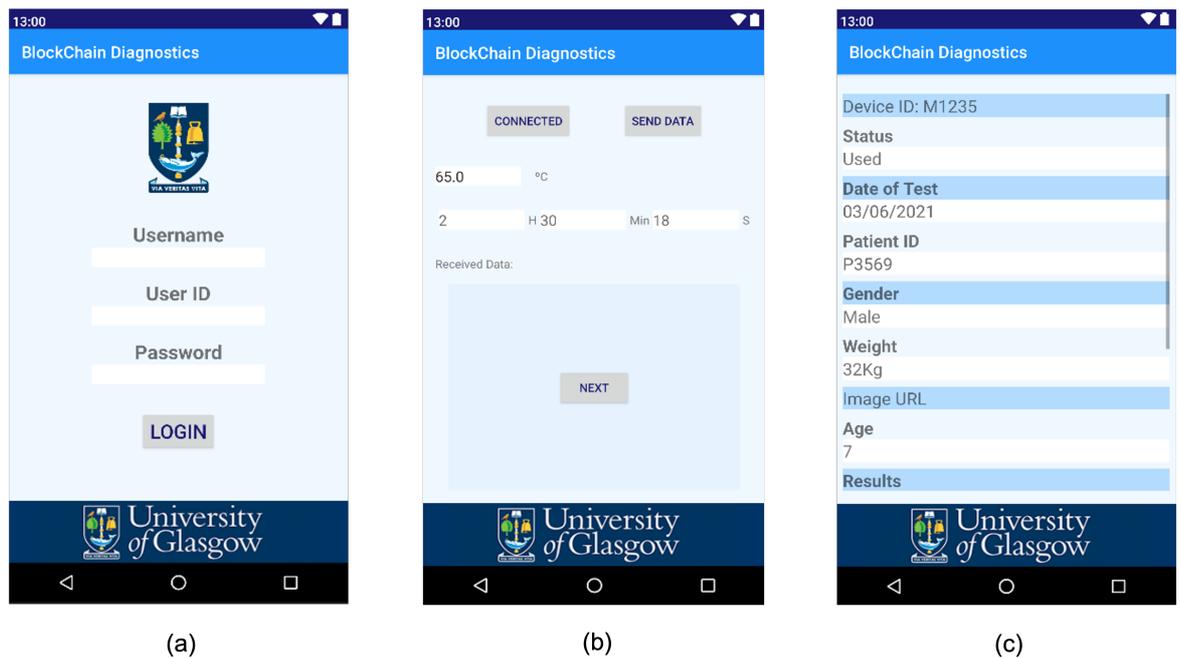


Figure 3-3 The Blockchain diagnostics app. [96] The App was developed with Android Studio in JAVA language. (a) the login page, (b) the heater control page, and (c) the test information page.

For each sample, the person running the test scanned the QR code of the device to be used (already ‘created’ by the manufacturer) and entered the required information on the test before inserting the device into the heater, controlled by the phone for amplification. The QR code was scanned, and a picture of the results was taken. The phone then returned the results for interpretation by the ‘analyst’, who could provide decision support to the person in charge of treatment. All testing steps (including derivatives) were also recorded manually to ascertain the results’ validity. When network connectivity was unavailable, the transactions were stored in the phone until connectivity was restored.

3.3 Results

Hyperledger Composer provided a helpful tool called Composer Playground. It is a web-based App that can be used to test and program the business network locally or in a cloud environment (Figure 3-4).

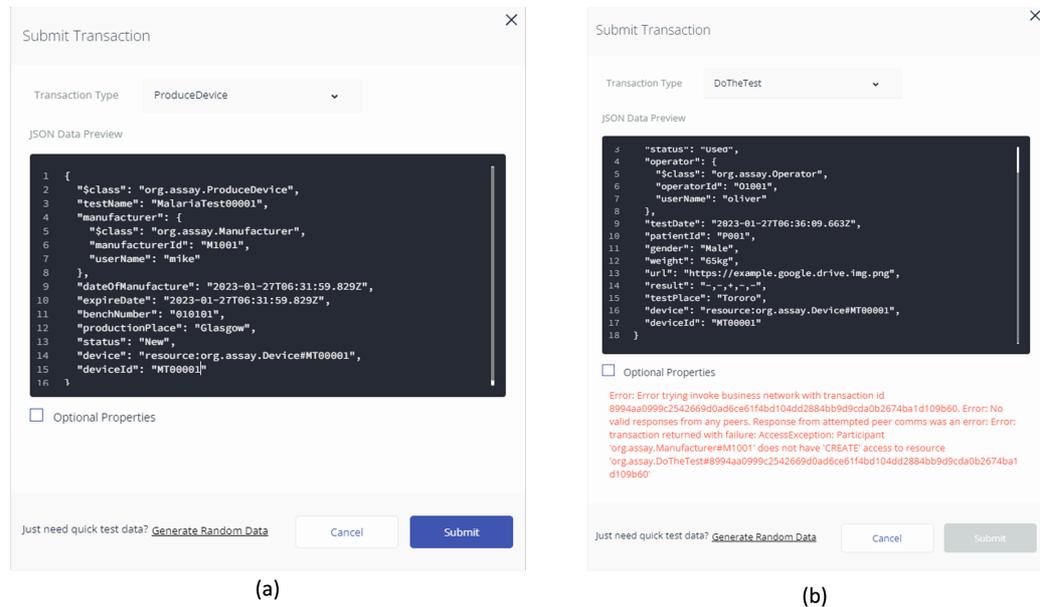


Figure 3-4 Transactions tested with manufacture network card in composer playground. (a) the manufacturer (id M1001) submits a “ProduceDevice” transaction. (b) the manufacturer tried to submit a “DoTheTest” transaction but got rejected because he has no access to this resource.

The Hyperledger Blockchain network has successfully been deployed on a Google server (nodes are running in Docker), whilst the Blockchain network was tested in Uganda in 2019 and 2020 with 79 device record successfully stored on the Blockchain.

All the tests were performed by using the mobile heater that has been proposed in Chapter 2. In the second field test, the system was upgraded. New features include geolocation data and automated result readout (see Chapter 4) was enabled.

3.3.1 Blockchain performance test

Hyperledger Caliper [103] was used as the benchmark, and both transaction functions were tested in six rounds. The test proceeded on an Ubuntu 16.04 virtual machine with four core CPUs, 4GB RAM and a 20GB hard drive.

The performance test results of the Blockchain are shown in Table 3-1. During the first round, 50 transactions were transmitted to the Blockchain network, resulting in a relatively low average latency of just 0.41 seconds. However, the network latency increased as more transactions were sent during the same period. Once the TPS reached 500, the average latency for all transactions exceeded 36 seconds. However, it is worth noting that none of the transactions failed, indicating that the network can handle high TPS surges with good tolerance.

Table 3-1 Performance evaluation results

| Transaction name | Succeed | Failed | Send Rate (TPS) | Max Latency (s) | Min Latency (s) | Average Latency (s) | Throughput (TPS) |
|------------------|---------|--------|-----------------|-----------------|-----------------|---------------------|------------------|
| ProduceDevice | 50 | 0 | 5.1 | 0.60 | 0.19 | 0.41 | 5.0 |
| | 100 | 0 | 10.1 | 3.04 | 0.42 | 1.71 | 9.2 |
| | 200 | 0 | 20.1 | 14.30 | 0.77 | 10.97 | 10.2 |
| | 300 | 0 | 30.1 | 23.10 | 6.02 | 22.12 | 10.3 |
| | 400 | 0 | 40.1 | 33.22 | 4.40 | 28.57 | 10.3 |
| | 500 | 0 | 50.1 | 46.43 | 9.42 | 36.99 | 10.3 |
| DoTheTest | 50 | 0 | 5.1 | 0.63 | 0.22 | 0.42 | 4.8 |
| | 100 | 0 | 10.1 | 4.36 | 0.27 | 2.52 | 8.7 |
| | 200 | 0 | 20.1 | 13.60 | 0.97 | 10.87 | 10.2 |
| | 300 | 0 | 30.1 | 23.73 | 1.17 | 18.44 | 10.6 |
| | 400 | 0 | 40.1 | 33.30 | 3.36 | 26.70 | 10.6 |
| | 500 | 0 | 50.1 | 43.02 | 17.05 | 36.02 | 10.3 |

3.4 Conclusion and Discussion

In this Chapter, a Blockchain network has been created to manage data effectively and securely. This network provides a tamper-proof solution for handling the diagnostic data gathered from the field and for continually monitoring the disease.

The Blockchain was developed using Hyperledger Composer and Fabric. Different functions were designed for specific users, such as the diagnostic device manufacturer, test operator, and analyst. An access control file and GitHub authentication regulated

access to the Blockchain. The communication between the user and the Blockchain network can be done by using HTTP requests from computer or mobile devices, e.g., get and post, from the web browser or a mobile App. Appendix Figure 2 shows the workflow of the presented Blockchain system.

Hyperledger Fabric can be used as a private (single organisation) or consortium (multiple organisations) Blockchain. Compared to another popular Blockchain network Ethereum, it can have lower latency and higher throughput, especially when facing large volumes of transactions[104]. The proposed Blockchain network was successfully tested in both lab and field environments. Although the network latency can increase noticeably when facing high throughput, no data loss was incurred during the test. However, it is possible to enhance the performance by using high-performance servers and optimising the transaction algorithm and network structure.

The use of Blockchain technology has significantly improved the potential and effectiveness of utilising smartphones for diagnostic testing in resource-limited areas. It helps to enhance trust by addressing issues encountered in past field studies, including the use of counterfeit test devices and unreliable test records [94], [105]. In this study, the information stored on the Blockchain does not contain any personally identifiable details. Each plastic microfluidic device has a unique QR code [105], [106], [107], [108]. Data security was further improved by implementing access controls. Analysts can trace and verify every test record, making the system reliable and trustworthy. Although the QR code is currently falsifiable (it can be removed from the device easily), in future, it could be embedded into the device material (using e.g., embossing techniques).

The Android App, also mentioned in Chapter 2, was updated to incorporate the Blockchain network. The updates included a user login interface, and multiple operation interfaces customised for different users. The App development was performed through collaboration with an undergraduate developer, Ivo Domingos (see also Acknowledgements).

A visual representation of the App's workflow can be found in Appendix Figure 3. The smartphone has potential to be used as a diagnostic tool, as it integrates with multiple

sensors, and its connectivity makes it an excellent interface of other technologies such as cloud computing, Blockchain and machine learning [109]. This makes the smartphone-based diagnostic solutions highly extendable and give us the opportunity to improve our platform even further.

Chapter 4 Convolutional neural network (CNN) for automated malaria test result interpretation

4.1 Introduction

Deep neural networks have shown great potential in many fields, such as computer vision, robotics, self-driving cars, and medical image processing. In this work, a convolutional neural network (CNN) was trained to read the result from the microfluidic LAMP test automatically, which was implemented in the Android App.

4.1.1 Deep learning

Deep learning is a branch of artificial intelligence (AI) and is now one of the most promising technologies in the field [110]. It uses deep neural networks (DNNs) to handle multiple types of tasks, for example, image classification, object detection, natural language processing (NLP), and semantic segmentation. DNNs comprise multiple layers of neurons that perform multiple nonlinear processes for feature extraction, classification, and transformation [111]. The training of DNNs includes supervised and unsupervised learning. In supervised learning approaches, the models need to be trained with labels, and the models learn how to generate the right prediction from the input based on the labels given in the training data. In unsupervised learning, the labels are not provided, which means the model does not know what the right answer is. The model needs to explore the data and find out the relationships within it. In this study, all the networks were trained using supervised approaches.

4.1.1.1 Convolutional layers

The convolutional layers are usually used to extract features from the training data and feed the neural network. The convolution process is shown in Figure 4-1. A convolutional layer's main component is convolutional kernels (filters), such as in Figure 4-1 (b).

Convolutional kernels are typically square matrices, but in certain scenarios, such as processing 3D data, a 3D kernel may be utilised [112]. Thomas H et al. employed a sphere convolutional kernel in their study [113], while Luo J et al. proposed the use of a hexagonal kernel in 3D shape classification [114]. Since the number of parameters and

kernel sizes have an exponential relationship, using a large convolutional kernel may not be cost-efficient. Therefore, the kernel size in spatial terms is not large.

The most commonly used kernel size is 3*3, and some other sizes were used in typical networks. For example, in AlexNet, 5*5 and 11*11 kernels have been used to have a wider receptive field. The receptive field is a crucial concept in deep convolutional neural networks. This refers to the region of the input space that a neuron or unit is sensitive to, and anything else outside the receptive field will not be able to affect that unit or neuron [115]. Expanding the receptive field enables neurons to identify broader regions and acquire more comprehensive characteristics from the input. However, simply increasing the kernel size will result in a significant rise in computational power.

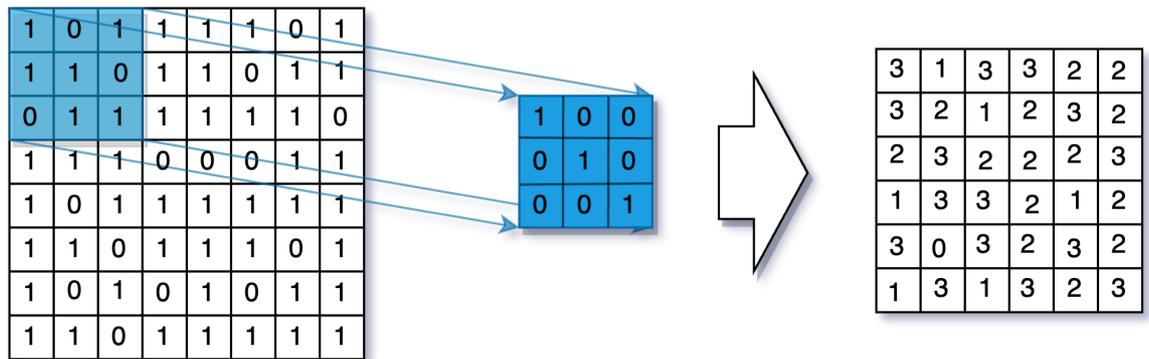


Figure 4-1 The convolutional layer. (a) The input matrix. (b) A 3*3 convolutional kernel. (c) The output of the convolutional layer.

4.1.1.2 Dense layers

The dense layer is also known as the hidden layer or fully connected layer. Each neuron in the dense layer connects with the neurons in its preceding layers, and the classification job was done in the dense layers. Figure 4-2 shows how a dense layer connects with its preceding layer and the following layer.

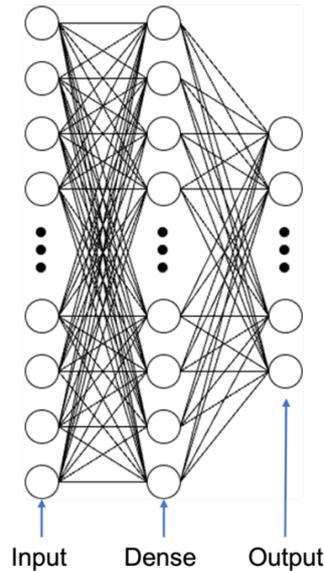


Figure 4-2 Dense layer. The input of the dense layer could be any layer with one-dimension output. Each neuron from the input layer connects with every neuron in the dense layer. In a classification network, the dense layer is usually used as the output layer as well.

4.1.2 Activation function

The activation function is an essential component of convolutional neural networks (CNNs). It can add nonlinearity to the system, reduce the complexity and accelerate the model running speed. There are different activation functions, with the most commonly used ones include ReLU, SoftMax, Sigmoid, and hyperbolic tangent function (Tanh).

4.1.2.1 ReLU

The rectified linear activation function (ReLU) can be expressed as follows:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

The ReLU function improves the nonlinearity and efficiency of the network by setting all negative values to 0.

In this Section, a ReLU function was applied to every layer of the sequential CNN model except the last one, the output layer.

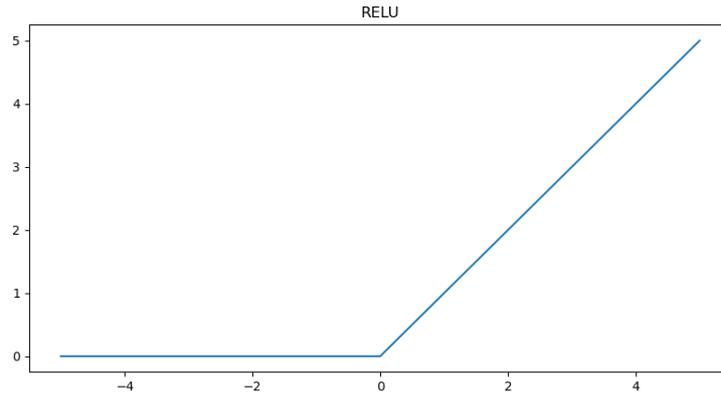


Figure 4-3 The ReLU function. It sets the input values that are less than 0 to 0.

Although there are many advantages of using ReLU, the biggest challenge in its use is “dying ReLU”, which means that some neurons could never be activated (died) because the ReLU does not have negative output.

To address the dying ReLU problem, some variants of the ReLU were developed, for example, Leaky-ReLU $f(x) = \begin{cases} \alpha x, & x < 0 \\ x, & x \geq 0 \end{cases}$ and ELU $(0 < \alpha) f(x) = \begin{cases} \alpha(e^x - 1), & x < 0 \\ x, & x \geq 0 \end{cases}$

[116]Figure 4-4.

Instead of setting all the negative values to 0, the Leaky-ReLU and ELU still retains some negative values, preventing the dying ReLU “problem”. However, as a result the computational complexity increased.

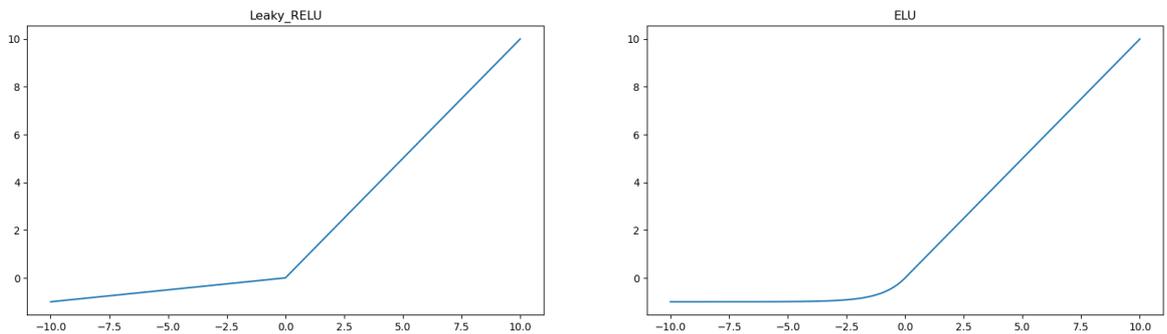


Figure 4-4 Leaky ReLU (left) and ELU (right) activation function.

4.1.2.2 Sigmoid

The sigmoid function can be formulated as follows:

$$S(x) = \frac{1}{1 + e^{-x}}$$

Sigmoid has usually been used as the activation function for the output layer. Because the output ranges from 0 to 1, that can be considered as the possibility of the prediction.

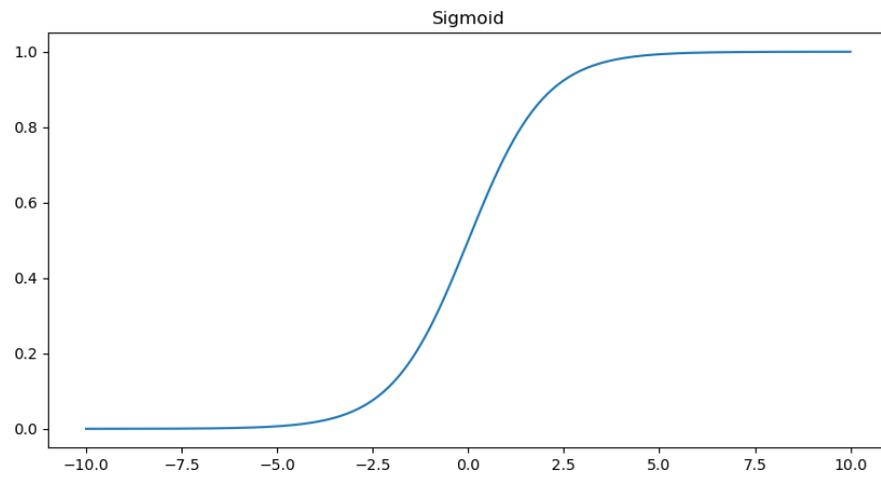


Figure 4-5 The Sigmoid function.

4.1.2.3 SoftMax

SoftMax is usually used in the output layer for multi-class classification models. It can be formulated as follows:

$$S(x_i) = \frac{e^{x_i}}{\sum_j^K e^{x_j}}$$

The SoftMax function can be considered as a combination of multiple sigmoid functions. The $S(x_i)$ represents the SoftMax value of the input vector x_i , e^{x_i} represents a standard exponential function for input vector, the e^{x_j} represents the standard exponential function for output vector, and K stands for the number of classes.

The SoftMax function could map the output and the sum of the output of each class will be 1, and the outputs of the SoftMax function could be interpreted as possibilities of every class.

4.1.3 Optimizer

The optimizer is a function or algorithm for adjusting the attributes of the neural network, such as weight and learning rate, to help the model converge. The commonly used optimizers include gradient descent (GD), stochastic gradient descent (SGD), and Adaptive Moment Estimation (Adam).

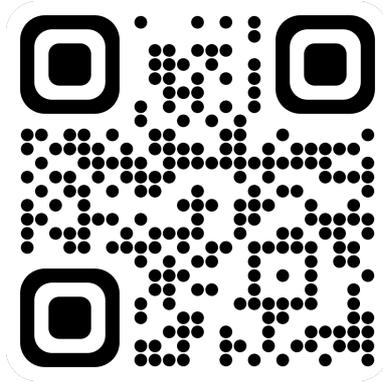
The GD is the most basic and widely adopted optimizer and is based on the first-order derivative of the loss function. It uses backpropagation to decide which direction how weighting should be tuned to minimize the loss. The advantage is that GD is easy to compute and implement, although, it could be trapped in the local minima, and when the dataset is too large, it will take a long time to converge. The GD will only update the weight matrix after the whole dataset has been processed (every epoch).

The SGD is a variant of GD. The calculation of SGD is similar to GD, but it could update the parameters more frequently. Instead of updating the weights and biases after the whole dataset has been processed, SGD randomly picks a batch of data to calculate and update the model parameters. This could help the model converge much faster, and requires less RAM for calculating the gradient, compared to the GD.

Adam was first proposed by Diederik P. Kingma In 2015 [117], and has shown better performance compared with SGD [117], [118]. The Adam algorithm is provided as Algorithm 3.

Algorithm 3. Adam [117] Details removed due to copyright restrictions.

QR code below links to the original paper



In the Adam function, the goal is to minimise the value of the function $\mathbb{E}[f(\theta)]$. Noisy objective function $f(\theta)$ is a differentiable w.r.t. parameter θ stochastic scalar function. And $g(t) = \nabla_{\theta} f_t(\theta)$ represents the partial derivatives of f_t , w.r.t. θ evaluated at time step t .

Compared to the GD and SGD, Adam could converge more rapidly, and the loss does not change sharply because of the data variance. In this project, the Adam optimizer was used for all models.

4.1.4 Deep CNNs

There are several models that have been used for different use cases and are considered milestones because of their outstanding performance, such as AlexNet, ResNet, VGG, and MobileNet.

4.1.4.1 AlexNet

The AlexNet was designed by Alex Krizhevsky in 2012 [119], it was considered as a milestone of Deep neural networks in image recognition. It achieved a 15.3% error rate in ILSVRC-2012 competition, compared to the second-best which was 26.2% [119]. It proved that the depth of the neural network could greatly influence the model performance.

4.1.4.2 ResNet

The ResNet, a deep residual network, has been widely adopted as the backbone for many deep neural networks due to its exceptional ability to extract and classify features. By leveraging the backpropagation mechanism, ResNet has taken the depth of DCNN to new

heights [120]. Unlike the DCNNs that have been used before, the ResNet has several residual blocks that consist of convolutional layers [120], and “shortcut connections” were used.

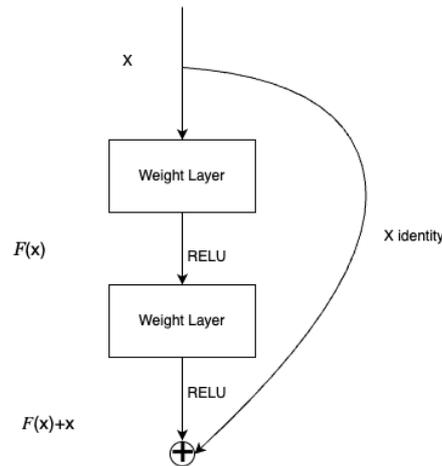


Figure 4-6 A residual block in Resnet. [120] The network is feedforward and with a shortcut connection.

The architecture of the ResNet allows the model to have better depth and tackle the gradient vanishing problem of the traditional CNN. In this Chapter, a ResNet-based object detection model was trained and used as a reference.

4.2 Methods

In this study, a CNN classifier and an object detection network were trained to classify the LF results automatically. The models were trained on both local and cloud environments. The local environment was a desktop with RTX3060 12G graphic card, 32G RAM, and Intel i5-12400 CPU. The cloud environment is Google Colab, which provides CPU, GPU and TPU for neural network training. All the training scripts were written in Python, and the models were based on Tensorflow 2.0 and Keras.

4.2.1 Data preparation

The training dataset included 5 classes, and 100 images in each class. The training data was obtained by carrying out targeted tests on synthetically prepared samples from the laboratory. Positive samples were obtained from the LAMP amplification of a Pf target sequence (WHO DNA standard obtained from the National Institute for Biological Standards) at 10^5 copies/reaction. Negative samples were obtained by LAMP

amplification using Pf primers and probes without any target (in this case, using de-ionised water).

This comprised two test strips for detecting *Plasmodium falciparum* (*Pf*) and *Plasmodium pan* (*Ppan*), with “pan” comprising a test of all species that cause malaria. There was also one positive control channel (using a test for the presence of *BRCA1* human gene, always present in humans) and two negative control strips (one for each of the two test strips), Figure 4-7.

The test strip in each channel therefore had three possible outcomes: negative, positive, and blank (invalid). Thus, using all results combinations across the five lateral flow strips gave 243 different possible result scenarios, including operator errors. To reduce the classification complexity, the outputs were subdivided into five clinically meaningful categories described in Table 4-1.

Table 4-1 Classification criteria

| .Channel Class | Negative control 1 | Negative control 2 | Positive control | Test 1 (Pf) | Test 2 (Ppan) |
|-----------------|--------------------|--------------------|------------------|-------------|---------------|
| 1N2P | - | - | + | + | - |
| 1P2N | - | - | + | - | + |
| Double positive | - | - | + | + | + |
| Negative | - | - | + | - | - |
| Invalid | +, / | +, / | -, / | / | / |

To increase the range of intensities in the bands available for training, amplicons were used at different amplification times (5, 10, and 15 mins), leading to a balanced dataset with 100 images in each class. This was for simulating the possible outcomes in actual

tests. Because in actual experiments, variations in the number of target DNA copies may influence the intensity of test bands.

To reduce the training time and improve the accuracy, the images were cropped in the App to a small 16:9 picture that contained results (Figure 4-8). All training images were resized to 128*128*3 before sending to the model.

Keras image generator (provided by TensorFlow) was used to augment the input data in real-time by adjusting both brightness/shade and zoom range as well as the orientation during the training process. This enlarged the training dataset and mimicked various possible conditions that may occur when taking pictures, such as different exposure, white balance, contrast, alignment, and so on. This can help improve the generalization ability of the model and reduce over-fitting.

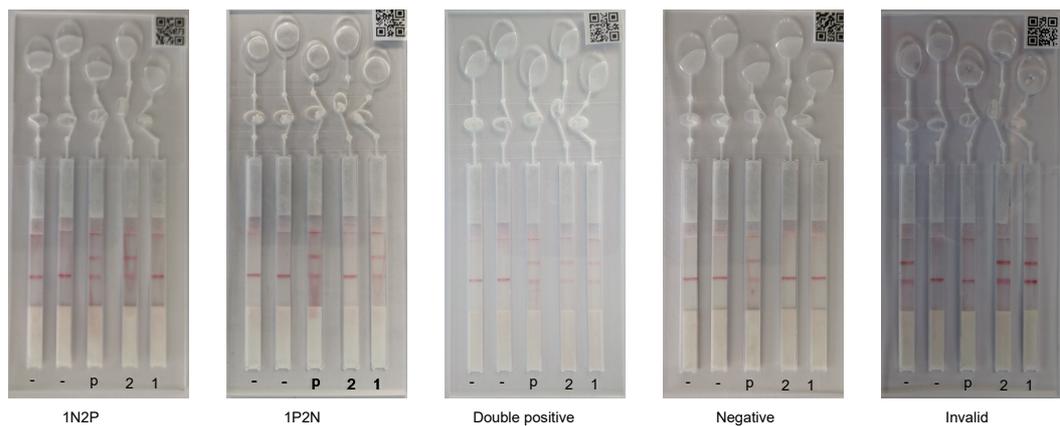


Figure 4-7 Examples of images from different categories. The channels marked with "1" and "2" are the test channels, the channel marked with "p" is the positive control channel and the channels marked with "-" are negative control channels.

A separate dataset with five categories containing 92 test images (including 11 "1N2P", 13 "1P2N", 23 "double-positive", 15 "negative", and 30 "Invalid") was created to test the model performance (the test set). The microfluidic devices were prepared with the help of Dr Xiaoxiang Yan and Dr Shantimoy Kar. For training the CNN, instead of using the full image as the input, the images were cropped into a smaller size (Figure 4-8 (b)) to reduce the complexity of the input data.

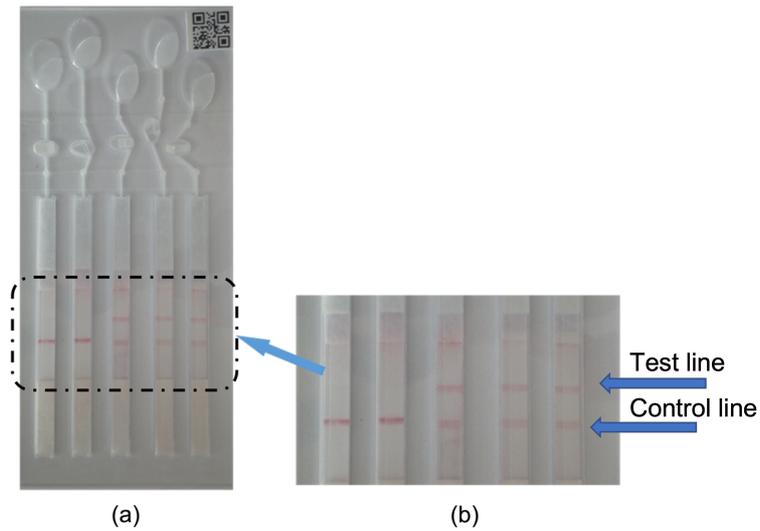


Figure 4-8 The Plastic cartridge [96] includes a microfluidic circuit with chambers for LAMP amplification reaction and lateral flow strips for readout, as well as a QR code for traceability. The dashed lines represent the cropped area for analysis by AI.

The same dataset without cropping was labelled using LabelImg software [121] to train the object detection model. There are five labels in the label map: negative, positive, empty, device, and QR code, where negative, positive, and empty indicate the outcome of each strip. After labelling, the images were divided into two separate sub-sets, 90% for training and 10% for testing, and corresponding tf-record (a format of TensorFlow dataset) files were created.

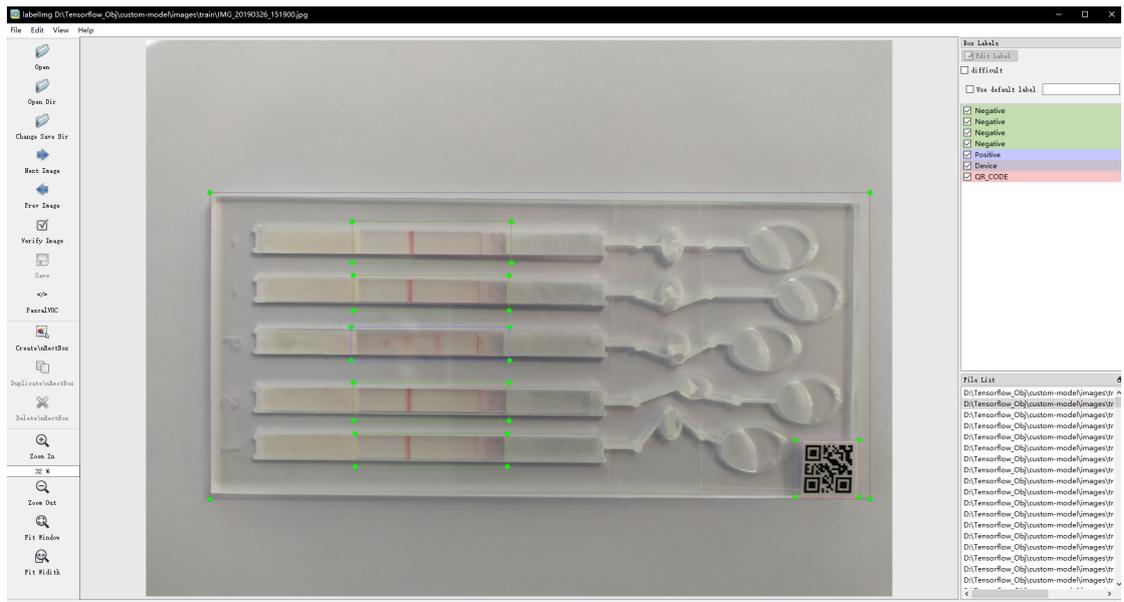


Figure 4-9 An example of image labelling using LabelMe [121].

4.2.2 Model training

Two neural networks were trained to automatically analyse the results of diagnostic tests [94]. An image of the device after the test is shown in Figure 4-8 as an example. One was a convolutional neural network (CNN) model, while the other was an object-detection model. Both models are based on Keras TensorFlow 2.0. The CNN classifier is a small-scale network, making it easier to use on devices with limited computational power.

The CNN was incorporated into the Android App discussed in earlier Chapters, while the object detection model, which was based on a Faster Region-based Convolutional Neural Network (R-CNN) ResNet50 model, was not included in the App. but instead was used as a reference for post-analysis to independently verify the results.

4.2.2.1 Classification network

The CNN model was developed and integrated into the App to classify the paper-based microfluidic diagnostic test images automatically. The five-plex DNA diagnostic strips, including species-specific diagnostics for *Plasmodium sp* as well as controls, were used as designed previously [95], based on lateral (capillary) flow showing a control line and a test line (Figure 4-8). The CNN was trained using both Google Colab and a local PC.

4.2.2.1.1 Architecture of CNN

In this project, a CNN with four convolutional layers and three dense layers was designed and used to automatically classify the test result from the paper-based LAMP device. The graphical architecture of the network is shown in Figure 4-10. There are two main tasks in CNN, namely feature extraction and classification. Several convolutional and pooling layers were used for extracting features from the input image. Each layer generated feature maps with varying levels of feature expression. The classification includes 3 dense layers, and between each dense layer, a dropout layer was applied to prevent overfitting.

It takes a 128*128*3 image that contains the test result of the LAMP test as the input. Then in the convolutional layers, the image will be scanned by several convolutional kernels, and a set of feature maps will be generated and sent to the dense layers. Between every two convolutional layers, a pooling layer was applied to reduce the number of parameters in the network.

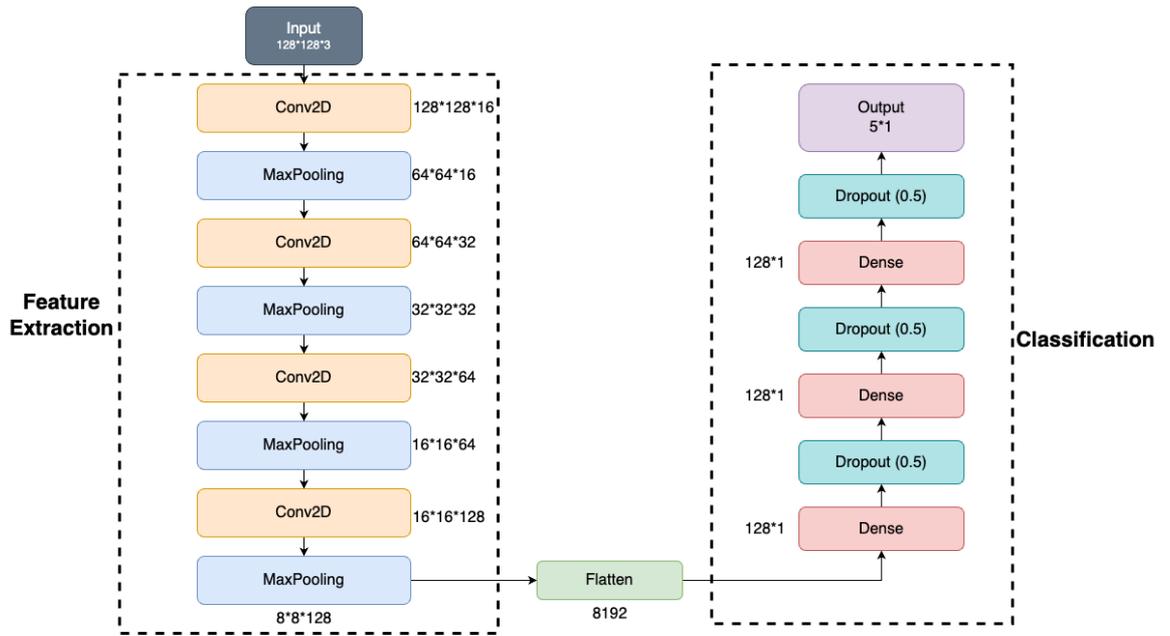


Figure 4-10 The architecture of CNN. It used four convolutional layers for feature extraction. After each convolutional layer, a MaxPooling layer was used to reduce the number of parameters. The extracted feature maps will be flattened to a 2D tensor and processed by the dense (hidden) layers.

The CNN was based on the TensorFlow Keras sequential model, which is “a plain stack of layers where each layer only has one input tensor and one output tensor” [122]. The structure of the sequential model was simple, allowing us to build it in a shorter time with Keras API, which generated computationally lightweight models suitable for smartphone deployment [123].

The model hosted sixteen layers, including four convolution layers, four max-pooling layers, a flatten layer, three dropout layers and four dense layers. The depth (number of layers) of the model was finetuned by multiple tests. The convolutional layer extracted features from the input images by scanning the input with a weighted matrix (convolution kernel). The process of generating a single feature map could be presented as follows:

$$A_j = f(\sum_{i=1}^N I_{i,j} * K_{i,j} + B_j) \quad \text{Equation 3}$$

Every input matrix I_i was convolved with kernels $K_{i,j}$, and a bias B_j was added to every element in the sum of convoluted matrices. The non-linear activation functions were applied to the matrix. All convolution layers used the activation function ReLU to

improve the learning speed and non-linearity of the model, setting all negative values of input matrices to zero. The Max-Pooling layer reduced the dimension of the output matrices of the previous convolution layer, using a 2*2 kernel with stride 2 to scan its input and take the largest number from four adjacent elements.

To extract sufficient features and detail whilst reducing the number of parameters in the training process, four convolutional layers were implemented, with a pooling layer following each convolutional layer. The output of the last pooling layer was then flattened to a 1D tensor and sent to the fully connected dense layers by the flattened layer. As the training dataset was relatively small and only had three categories, the model needed to have more fully connected (FC) layers and relatively fewer neurons [124]. Consequently, three dense layers (size 128) and one dense layer (output layer) with size 5 (number of class) were used to obtain better accuracy.

Between every dense layer, a dropout layer was utilised to prevent overfitting. The first three dense layers also used ReLU as their activation function. The last dense layer, the output layer, used SoftMax ($S(x_i)$) as its activation function to provide the predictions and their probability.

The loss function used in training the CNN is sparse cross-entropy, which can be formulated as follows:

$$L = -\frac{1}{N} \sum_i \sum_{j=1}^M y_{ij} \log(p_{ij}) \quad \text{Equation 4}$$

In the equation above, N represents the number of samples, while $y_{i,j}$ indicates whether sample i belongs to category j, and $p_{i,j}$ indicates the predicted possibility that sample i belongs to category j.

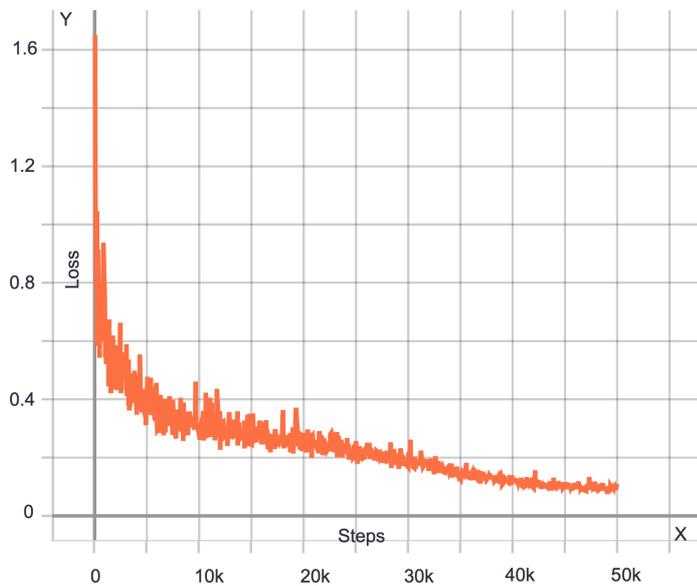
During model training, an Adam optimizer with an initial learning rate of 0.0001 was used, while the batch size was set to 32.

4.2.2.2 Object detection network

To evaluate the CNN, we used an object detection model as a comparison benchmark. The objective detection model was based on a faster RCNN ResNet50 model provided in the

Tensorflow model Zoo [125]. Tensorflow offers object detection APIs and pre-trained models to train object detection models. Useful Python scripts for model training were also provided. To start training, the training pipeline only requires a few configurations and can be started using the provided training script.

TensorBoard was used to monitor the training process, Figure 4-11(a) shows the training loss curve of the object detection model and an example of its prediction. The loss function used for this model is also sparse cross-entropy.



(a)



(b)

Figure 4-11 Performance of the object detection model. [96] (a) The loss curve of the training process of the object detection model. The model was based on SSD resnet50. The x-axis was the training steps and y-axis was the loss. (b) A sample image with predictions, all the targets were detected successfully and classified correctly with high confidence.

4.2.3 App implementation

The Android App now has the CNN feature available, which was tested and found to have an accuracy of over 98% and a loss of less than 0.1. Since mobile phone devices have weaker processing power when compared to desktops, the CNN model needed to be compressed. TensorFlow offers a technique known as “model quantization” that reduces the model size while maintaining performance.

4.3 Results and discussion

The CNN was tested with both laboratory collected data and field data. Figure 4-12(a)(b) shows the model accuracy and loss changes during the training process, respectively. The model converged after 300 training steps, with a batch size of 32. A training step denotes the model successfully processed a batch of data and updated the model parameters. The model achieved 98% accuracy and less than 0.1 loss. Figure 4-12(c) is the confusion matrix of the CNN, and it is based on the test set created using the lab data, which includes 92 images. Only two images were wrongly classified, including one 1P2N and one Double positive.

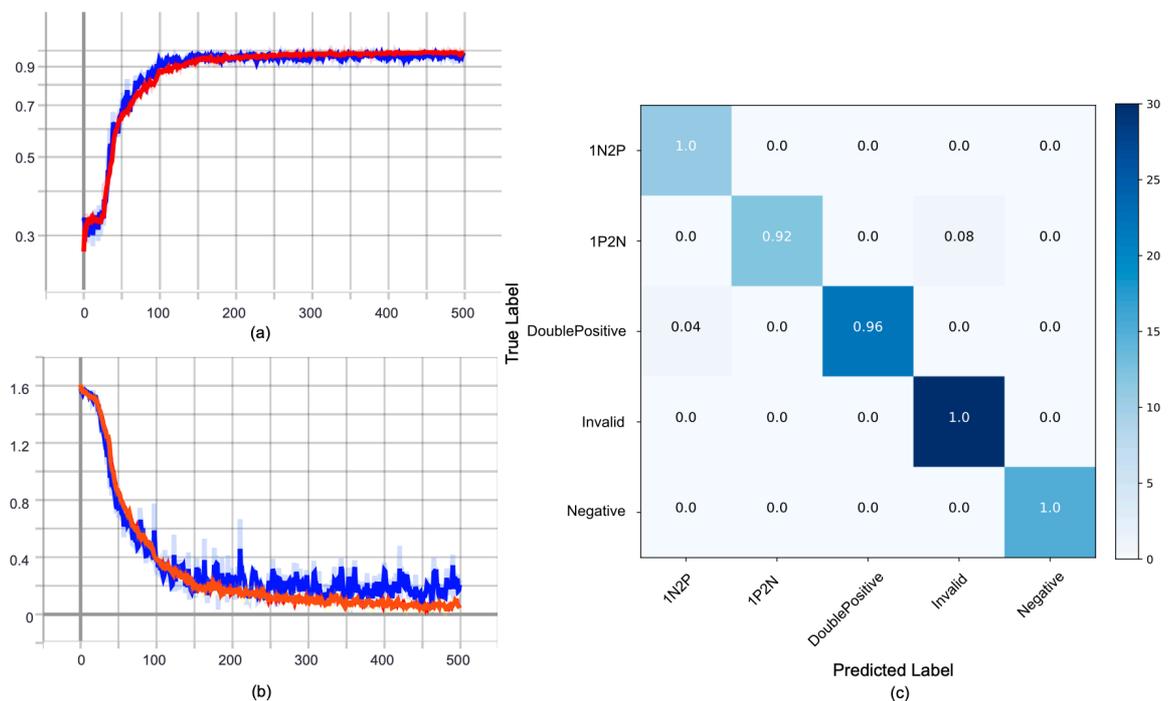


Figure 4-12 The performance of the CNN. [96] (a) accuracy and (b) loss of AI training process (x-axis shows epochs). Red represents the training set and blue represents the test set. (c) Confusion matrix of the test results of the CNN model. The horizontal axis of the matrix represents the predicted label, and the vertical axis represents the actual label of every test image. The background colour of each grid of the matrix represents the number of images that were classified into that case (darker means there were more images), and the number on each grid is the percentage of predictions in that case.

Our CNN has an overall sensitivity of 0.976 with a standard deviation (SD) of 0.035 and a specificity of 0.994 (SD=0.0079). In comparison to other similar systems that use

smartphones and CNN to interpret the lateral flow test results, our CNN has achieved a state-of-the-art level of accuracy. Table 4-2 provides a comparison of the performance of other readers and our CNN. Our CNN was implemented in the Android App after quantisation. Tested on a Nokia 4.2 mobile phone with Android version 12.0, processing one image only took $20\text{ms} \pm 5\text{ms}$.

Table 4-2 Model performance comparison

| Method | Accuracy | Sensitivity | Specificity |
|----------------------------|----------|-------------|-------------|
| Fuhad K al. (2020) [126] | 0.968 | 0.976 | 0.957 |
| Mendels D al. (2021) [127] | 0.993 | 0.995 | 0.999 |
| Mujtaba D al. (2021) [128] | 1.000 | 1.00(0.997) | 1.00(0.996) |
| Wong N al. (2022) [129] | 0.983 | 0.971 | 0.994 |
| Lee S al.(2023) [130] | 0.99 | 0.987 | 0.978 |
| Our CNN | 0.978 | 0.976 | 0.994 |

To better evaluate the CNN's classification ability, we compared the classification ability of the ResNet50-based object detection model and our CNN. Figure 4-13 is the precision-recall curve of the CNN (blue) and ResNet50(red). The test was carried out by using the images collected from the field test. A total of 40 images (29 valid results and 11 invalids) were used for testing the models. The CNN was able to correctly classify all the invalid results and 28 of the valid results into correct categories. Only one positive result has been wrongly labelled as invalid.

Because the object detection model has both localisation accuracy and classification accuracy, the SSD ResNet50 model only considers the classification results from the strips that have been successfully detected. For example, if there is one strip that has not been detected by the model, the result will be skipped. Our CNN and the ResNet-based model both achieved a high area under the precision-recall curve (AUC), but the CNN had a slightly better classification performance.

Our CNN has proven to be highly accurate and efficient, but it has limitations when it comes to scalability. It was specifically trained for use with the five-chamber microfluidic device [94], so any modifications to the design would require the model to be retrained.

Using the object-detection model, like the SSD ResNet50 we used in this study, can be a viable solution for enhancing scalability, but it requires additional computational power and time for both training and prediction [131]. Additionally, preparing the dataset for training an object-detection model can be time-consuming.

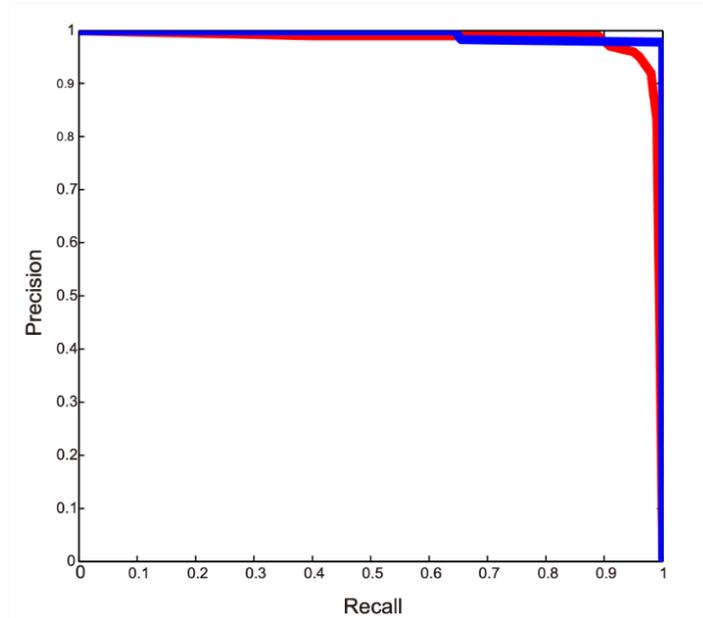


Figure 4-13 The precision-recall curves [96] for the CNN (blue) and the SSD ResNet 50 (red) to compare their predictive abilities, with recall measured as $TP/(TP+FN)$, and precision as $TP/(TP+FP)$, where TP are the true positive predictions, FN is false negative predictions and FP are false positive predictions. If the confidence level of a prediction exceeds a threshold (e.g. 0.8), the result is deemed a positive case if not, it is a negative case. If the prediction matches the true label of the input, then the output is true. If not, it is false. The AUC of the CNN and ResNet50, respectively, are 0.993 and 0.983.

As the object detection models are usually more complex and have more parameters, running them on mobile devices can be another challenge, especially in low-resource settings. Nevertheless, many mobile CPUs now come with integrated neural processing units (NPUs), which significantly enhance the capacity of mobile devices to execute deep learning models [132]. Moreover, an increasing number of networks have been designed specifically for mobile devices, allowing them to run machine-learning tasks more efficiently [133], [134].

The increase in the number of mobile network services and the increasing number of smartphone users in low- and middle-income countries (LMICs) has made smartphones the most accessible technology in many resource-limited regions. In Sub-Saharan Africa and South Asia, for instance, the growth rate of smartphone users has exceeded 20% [135]. With the improvement of network connections, cloud computing services are becoming more accessible in some LMICs, allowing complex deep learning models to be deployed on cloud servers instead of local devices.

Cloud-based services can provide greater computational power and a wider range of services compared to edge computing devices. Whilst edge-computing devices can provide lower latency, energy consumption and user experience [136]. In the previous Chapter, we explored the possibility of utilising a cloud-based Blockchain network in low-resource settings, and the results were positive. In the future, it would be advantageous to investigate the integration of cloud-based solutions and edge computing to enhance services in resource-constrained environments and enhance the platform's overall capabilities and reliability.

4.4 Conclusion

The smartphone-based system for malaria tests mentioned in Chapters 2 and 3 now includes a neural network based on CNN. The CNN enables automated result analysis and has shown high accuracy in classifying positive and negative test outcomes.

The CNN was designed as a 5-class classification model with 16 layers, and the input size was set to 128x128x3. It is small in scale and can run on a normal smartphone after quantisation. It was integrated with the Android App mentioned in the previous Chapters, allowing the test operator to take a photo of the microfluidic device after each test. and have the App interpret the results. This not only eliminates the potential for human interpretation errors but also reduces the operator's training requirements.

Cloud storage was also used to store the images captured after the tests. Thus, other participants in the Blockchain network can verify each record that has been stored on the Blockchain system. This enabled the entire system to meet the WHO ASSURED

(Affordable, Sensitive, Specific, User-friendly, Rapid and Robust, Equipment-free and Deliverable to end-users) criteria [137], [138], [139].

Additionally, the test records kept on the Blockchain are now geo-tagged, allowing for constant disease surveillance. However, to protect privacy, the exact geolocation was not included.

Chapter 5 Retina vessel identification and segmentation using deep learning

5.1 Introduction

Deep learning has proven to be useful in many medical image analysis tasks due to its capability to perform tasks such as image classification, segmentation, and object detection. Based on the previous experience of using CNN as an interpreter of the lateral flow test result, this Chapter investigated the possibility of using deep learning to assist in diagnosing retinopathy.

Retinal diseases place a heavy burden on public health. There are approximately 422 million people who have diabetes mellitus (DM) [140], [141] and diabetic retinopathy (DR) as one of the common complications of DM and the leading cause of blindness [142]. Thus, diagnosing and treating diabetic retinopathy in the early stage is necessary to prevent blindness in patients. However, diagnosing retinal diseases requires a high level of expertise and heavily relies on expensive imaging systems, which are normally inaccessible in remote villages.

Retinal image analysis is a widely adopted and helpful tool for diagnosing retinal pathology. For example, the abnormal tortuosity of the retina vessel and the presence of exudates can be noticed in fundus images. The abnormalities in the retina image could reveal lesions, such as retina vascular diseases, cardiovascular disease, DR, and hypertension in the early stage [143]. To perform the analysis, retinal vessel segmentation and classification are fundamental and crucial steps [143], [144].

Traditionally, retinal vessel segmentation and classification require manual processing and substantial professional knowledge. The annotating process can be very time-consuming and tedious [62], [145]. Ophthalmologists need to carefully identify each vessel (identifying the arteries and veins) from the background. The analysis based on observation can sometimes lead to false results due to uneven illumination of the imaging system, artefacts, glares, and various interpretations from different specialists [145]. Thus, many researchers endeavoured to develop techniques for automated retinal image analysis.

In the past ten years, many studies have introduced automated retinal vessel segmentation. For example, Huang *et al.* proposed a framework for artery vein (A/V) classification based on a feature selection algorithm [146], Liangzhi Li *et al.* [147] designed a novel deep learning model called IterNet, which consists of a UNet and iterations of several mini-UNets, and Zhun Fan *et al.* [143] introduced an octave convolutional neural network that extracts frequency information from the retinal images to differentiate vessels from the background. Many studies have reported over 0.95 accuracy and specificity.

However, in these studies, the data were mainly from public datasets such as “DRIVE” [61], “INSPIRE” [148] and “ROC” [68], which only consist of pictures that have been captured by professional ophthalmoscopes. Additionally, compared to vessel segmentation, there are fewer studies on A/V identification, as most studies focus only on vessel extraction [149]. The data in Table 5-1 shows the outcomes of several recent studies on retina vessel segmentation and A/V classification using deep learning approaches. Networks that solely focus on segmentation have higher accuracy than those that also perform the A/V classification.

Table 5-1 Performance comparison of retina vessel segmentation and A/V classification

| Method | Accuracy | Sensitivity | Specificity | AUC |
|---|----------|-------------|-------------|-------|
| Huang F <i>al.</i> (2018) [146] (A/V) | 0.720 | 0.709 | 0.738 | 0.78 |
| Galdran A <i>al.</i> (2019) [150] (A/V) | 0.890 | 0.890 | 0.900 | - |
| Kang H <i>al.</i> (2020) [70] (A/V) | 0.906 | 0.883 | 0.927 | - |
| Liagngzhi Li <i>al.</i> (2020) [147] | 0.957 | 0.774 | 0.984 | 0.982 |
| Zhun Fan <i>al.</i> (2020) [143] | 0.976 | 0.867 | 0.984 | 0.991 |
| Wang D <i>al.</i> (2020) [151] | 0.958 | 0.799 | 0.981 | 0.982 |
| Uysal E <i>al.</i> (2020) [152] | 0.942 | 0.755 | 0.968 | - |
| Saranya P <i>al.</i> (2022) [153] | 0.960 | 0.950 | 0.990 | - |

The goal of this Chapter was to build upon previous expertise developed in earlier Chapters to develop systems that could be used to segment and classify vessels in photos captured by a low-cost ultrawide field retinal vessel imaging system developed by Professor Andy Harvey (subsequently analysed by Dr Victor Ochoa, see Acknowledgements). Different from the CNN proposed in Chapter 4, semantic

segmentation models have a larger scale due to pixel-wise labelling of images. As each pixel needs to be classified, more parameters are required, and it is challenging to identify small and similar features accurately.

Compared to public datasets, these ultrawide field images revealed more details of the retina but also contained more glares, artefacts, and background noise. The most challenging part of the project was the scarcity of data. Due to the limited access to the camera and lack of ophthalmologists to label the data, only six ultrawide fundus images were available, with only two of them having ground truth labels. This was significantly less than the public datasets, which typically contain more than 30 images. Thus, the DRIVE dataset was used as a reference in this study.

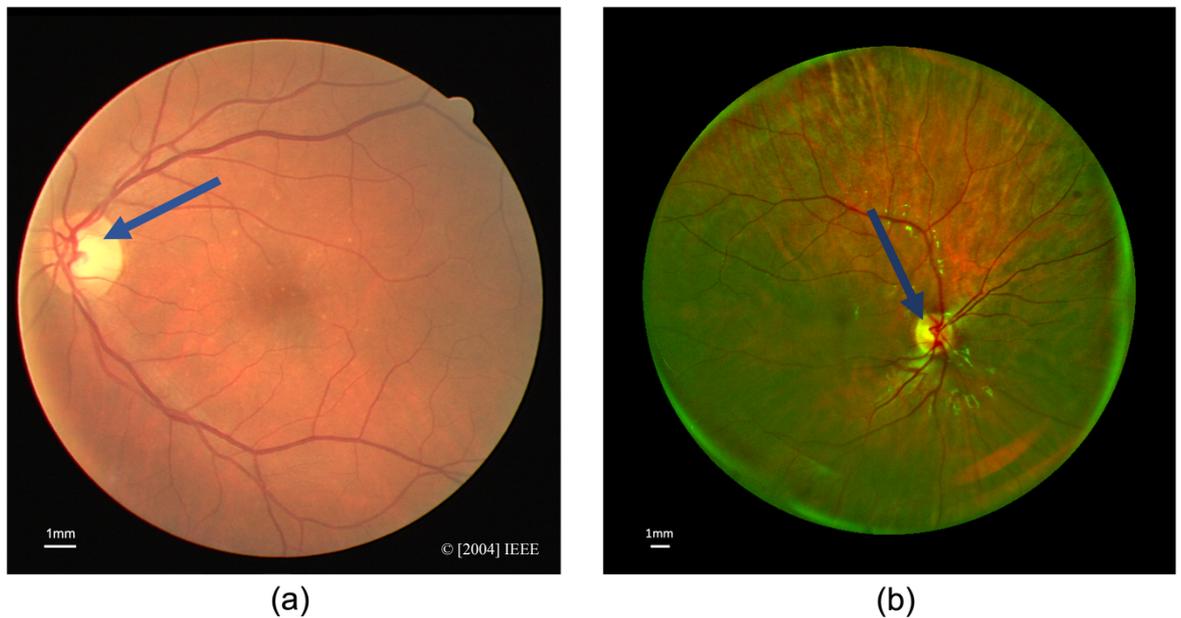


Figure 5-1 An example of images from the DRIVE dataset (a) and our dataset (b) which shows more glare spots and many red flocculent parts. The bright area pointed by the blue arrow is the optic disc.

Figure 5-1 (a) shows an example of the retina image captured from a standard fundus camera (DRIVE 21_training.tif), and Figure 5-1 (b) shows a UWF retina image captured from a low-cost handheld fundus camera. The blue arrows point to the optic discs. It can be noticed that the optic disc area in the UWF image was smaller, which means the UWF image has a wider field of view and could cover more area of the fundus than the standard

fundus cameras. Thus, the UWF imaging system could detect some early lesions that cannot be seen from the typical fundus cameras.

5.1.1 Challenges in retina vessel segmentation and classification

Vessel segmentation is a challenging task due to the difficulty of distinguishing small vessels from the background noise. This noise may come from the illumination system or imaging sensor, resulting in bright spots and fuzzy areas in Figure 5-2. From observation, these noises have very similar features compared with some small vessels, and in most of the published studies, the DNNs were struggling with identifying tiny vessels and backgrounds from the features [154], [155]. It could be more challenging in our UWF fundus images because it covers a wider area of the retina, and more information will be included.

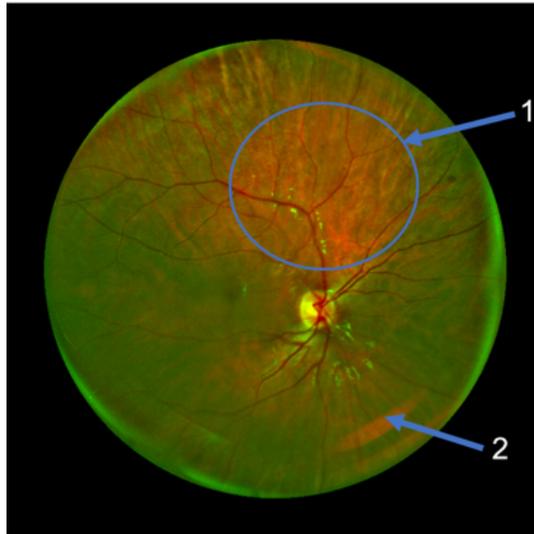


Figure 5-2 The noise in the UWF fundus image. 1 The red flocculent parts and 2 glares.

Another challenge of applying deep learning in medical image analysis is the data availability of medical images. Although, as mentioned previously, there are some public retinal image datasets available, compared to other public datasets, medical image datasets are generally much smaller. For example, the ADE20K [156] dataset comprised >20,000 labelled images, but retina image datasets normally only contain less than 200 images.

A further challenge is that medical images are usually complex and difficult to label. For instance, the retina image contains numerous small vessels, which are essential in detecting lesions during the early stage [157]. Some blood vessels are challenging to

distinguish from the background and other features, including for example, optic nerves. The shape of the vessels and lesions is also often irregular; consequently, labelling the fundus image on the pixel level can be very time-consuming.

5.1.2 Diabetic retinopathy

The diagnosis of DR is usually based on the early treatment diabetic retinopathy study (ETDRS) grading system (Appendix Figure 4), which is the gold standard of DR diagnosis [158], [159], [160]. DR can be divided into two types based on the stage, which are non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR), [159], [160]. There are four major types of lesions in DPDR: hard exudates, soft exudates, microaneurysms, and haemorrhages. According to the severity of the lesions, the NPDR can be classified as mild to moderate, moderate to severe and severe to very severe.

In practice, the grading system could be slightly different. For example, DR could be divided into three [160] or six stages [159], according to NHS England (Appendix 1-4) and Scotland (Appendix 5) diabetic retinopathy screening guidelines, respectively.

5.2 Methods

In this study, several different DNNs were trained for two tasks: vessel segmentation only and artery vein segmentation + classification. Both tasks were done using supervised learning approaches. Different models, including UNet, PsPNet, and Attention UNet, have been trained and tested using the DRIVE dataset and UWF retina images.

5.2.1 Data preparation for vessel classification

5.2.1.1 DRIVE dataset

The DRIVE dataset was used for training and evaluating different model performances. It includes 20 hand-labelled retina images for training and 20 labelled photos for testing. The 20 images in the training set were divided into two subsets: a training set that includes 16 images and a validation set with four images. The test set was only used in model performance evaluation.

The retina images are RGB with size 565*584*3, and the label files are single-channel images with size 565*584. The image segmentation approach used in this study is semantic segmentation. It requires the data to be labelled pixel-wise. There are two labels in the DRIVE dataset, background and vessel, and each class were represented by pixel value 0 (background) and 255(vessel), respectively.

The label files of the DRIVE dataset were provided in 3 channel Graphics Interchange Format (.gif), and the image files are provided with tag image format (.TIF) format. Images and labels are converted to Portable Network Graphics (.PNG) format to simplify the data loading process. And the labels were converted to single-channel images, and pixel values were encoded to 0 and 1 to represent the background and vessel.

For the A/V classification task, a dataset called RITE [73] was used. The RITE dataset was created by the University of Iowa, and it can be considered a variant of the DRIVE dataset with A/V information. Figure 5-3 shows an example of the data from the DRIVE [61] and RITE datasets.

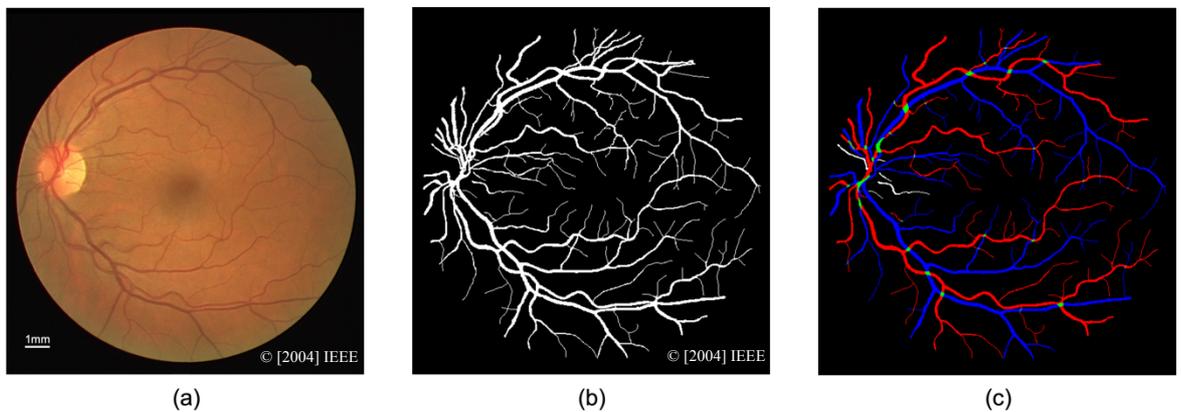


Figure 5-3 Sample images from the DRIVE dataset. (a) original vessel annotation from the DRIVE dataset. (b) the extracted vessels. (c) The label used for AV classification (red for arteries, blue for veins, the overlapping of arteries and veins are labelled in green; the uncertain vessels are labelled in white.

5.2.1.2 UWF retina images

The annotation of the image (see Figure 5-4) was created using Lableme [161] software based on the ground truth provided by Dr Victor Ochoa. Figure 5-4 (a) shows the original

image used for training, and Figure 5-4 (b) is the ground truth of the vessels overlaid on the raw image.

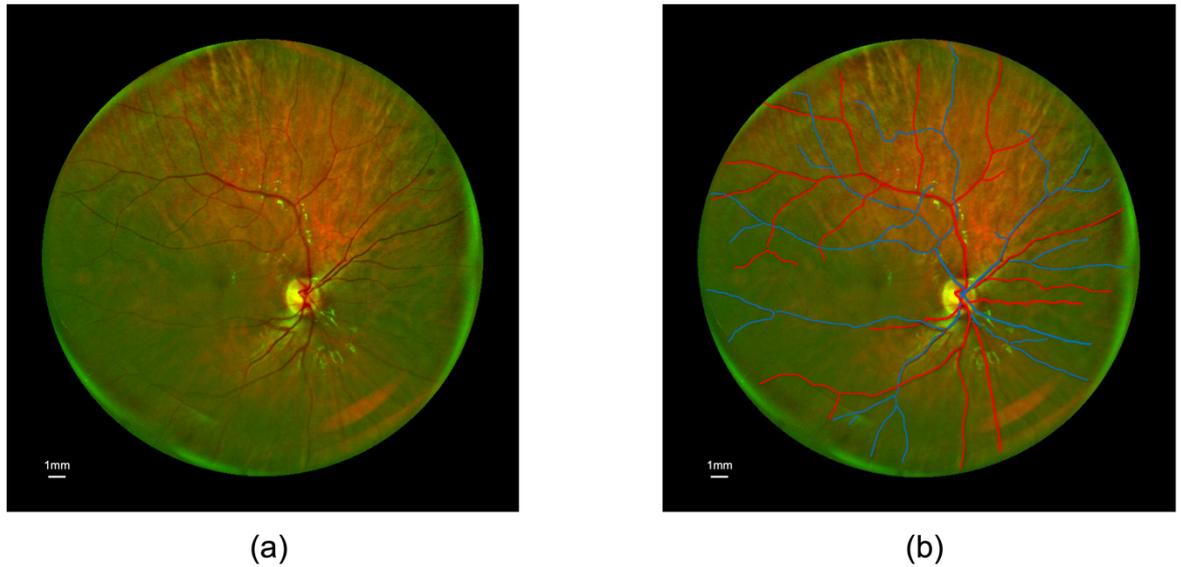


Figure 5-4 Example of the retinal image used for training. (a) The original ultrawide field retina Image for training. (b) hand labelled ground truth provided by Dr Victor Ochoa, the red represents veins, and the blue is the arteries.

5.2.1.3 Image labelling

The hand-labelled vessels in Figure 5-4 (b) show the category of the vessels of the fundus image; however, some of the annotations cannot accurately cover the vessel area, and a separate label file was needed for model training. Thus, the images were relabelled using a Python software called Labelme [161] (Figure 5-5). In Labelme software, the label can be accurately created by clicking the edge of the vessel. The vessels were pixel-wise labelled as arteries and veins by following the ground truth shown in Figure 5-4 (b).



Figure 5-5 The user interface of LabelMe [161] software. The mask is created by creating polygons that follow the edges of the vessels. The coordination and label information will be stored in a JSON file.

After labelling, Labelme [161] software could save the label information in JSON format, including the coordinates and category of each labelled vessel. The JSON file can be converted to an image by using the built-in script (*labelme_json_to_dataset*) of Lableme. Figure 5-6 shows an example of the label file, (a) is the hand-labelled ground truth, and (b) is the label for model training.

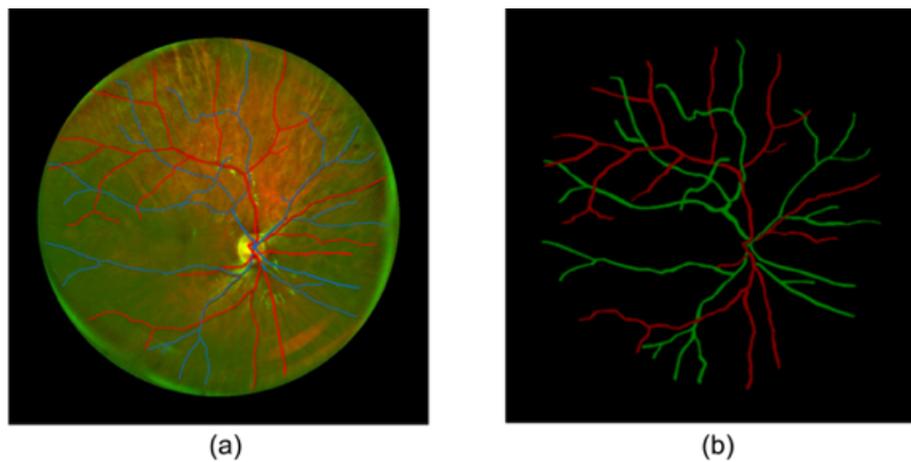


Figure 5-6 The labelled retina image. (a) the ground truth provided by Victor (red for vein and blue for artery), and (b) the label map used for model training (red for vein and green for artery).

5.2.1.4 Data augmentation

Both DRIVE and our UWF datasets are relatively small for training the deep neural network, and it could easily lead to overfitting. Thus, data augmentation was necessary for improving the model's generalisation ability.

The data augmentation was applied in two stages. Firstly, the training images were augmented to 800 images by applying random rotation, flip, distortion, brightness, and contrast. This part was performed using the Augmentor software [162], and could simultaneously be applied to the same transformations to the image and its corresponding mask.

It should be noted that it is important to verify that the labels' category index (the pixel value that indicates the category) was not changed during the augmentation process. During the study, the Augmentor could occasionally apply unwanted adjustments to the labels. For instance, when adjusting the brightness and contrast of the image, its corresponding label should not be modified, but sometimes the result is the opposite. This could lead to errors when starting the training because the label's pixel value indicates each pixel's classes.

To address this issue, after the augmentation, a checking step was added to check if any label images had been changed mistakenly. The checking was performed using the NumPy unique function. It could return the unique values in the label matrixes, and it could be easy to notice if there are any mistakes, as all values in the label matrixes should be integers, and the range of the values is set.

The second stage of data augmentation was performed when the images were loaded into the dataset. Transformations include randomly adjusting contrast, brightness, gamma, image quality and random flip applied to the training images.

To prevent overfitting and ensure that the original images were still used for training, a possibility of 0.5 was set for each adjustment. This was achieved by generating a random number between 0 and 1 using the *tf.random.uniform* function. If the number is greater

than 0.5, then the corresponding transformation will be applied. The transformation includes random brightness, contrast, and flip.

In the beginning, full-size fundus images were used for training. The model however could not perform well in detecting very thin vessels (see Chapter 5.3). To address this, there could be two options, namely either increasing the input size of the model or cropping the image into small patches and restoring the full picture afterwards. Increasing the input size could greatly increase the model complexity, and introducing more parameters will take more time for training. Thus, the latter option was chosen.

To perform this, random cropping and zooming transformation were implemented at the first data augmentation step. The probability of cropping was set to 1 with a cropping area ranging from 0.2 to 0.5, and the zooming possibility was set to 0.5 with a zoom range from 1.2 to 2. Figure 5-7 shows an example of the small patch of the retinal image.

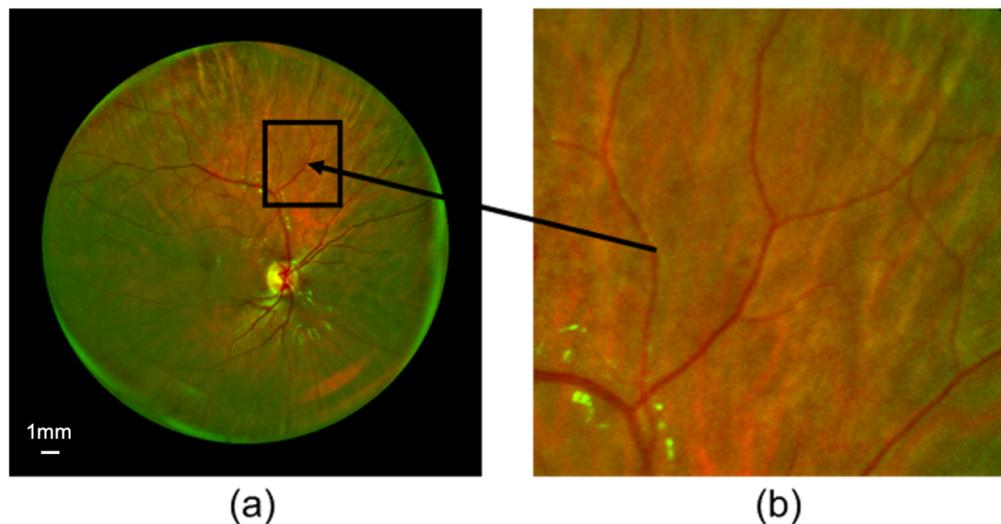


Figure 5-7 An example of the cropped image. (a) the original image. The black box was the cropped area (b) the small section of the retina image.

5.2.2 DNN Models for Semantic Segmentation

5.2.2.1 UNet

The UNet model was the first model that was used in this project for vessel segmentation and classification. Different from the typical convolutional neural networks used for

image classification, such as the one described in Chapter 4. The UNet was a fully convolutional network for semantic image segmentation [163]. Figure 5-8 shows the architecture of the UNet.

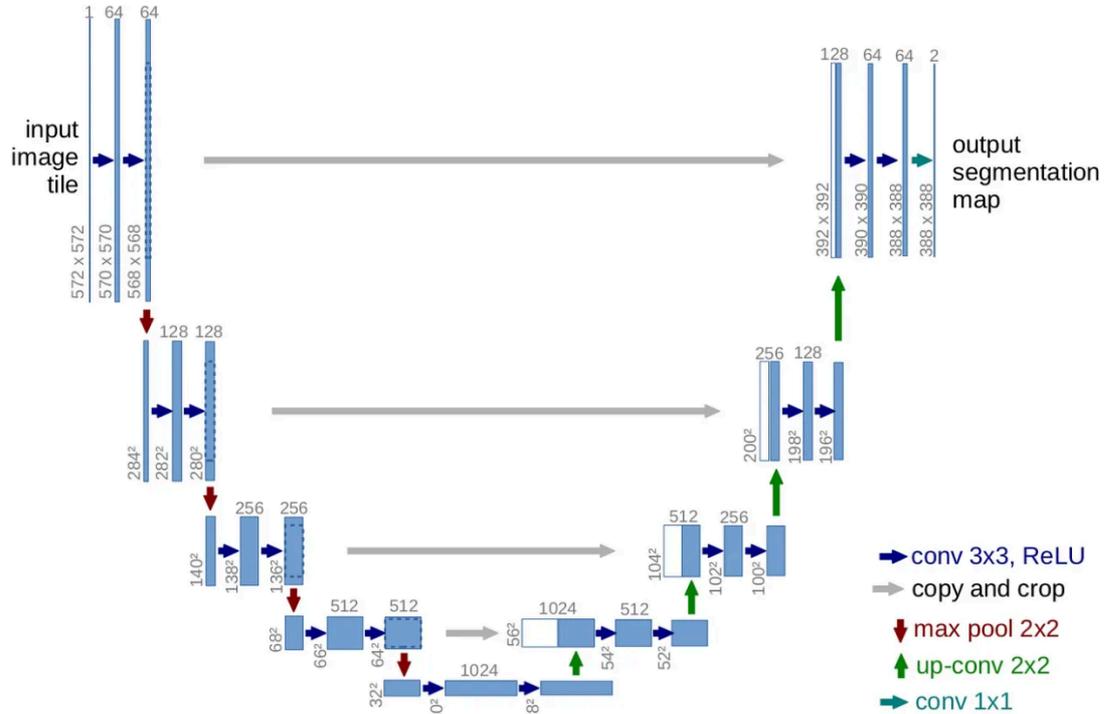


Figure 5-8 UNet architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows indicate the different operations [163].

Some modifications were made based on the original UNet architecture, namely:

1. The input size was changed to 512*512 and 256*256.
2. The copy and crop process were simplified to copy.
3. The depth of the model was adjusted.

The model underwent modifications to make training and prediction easier. Padding was added after each convolutional process, resulting in input and output images of the same size. Changing the size to either 512*512 or 256*256 can prevent odd-sized matrices when sending the input to a pooling layer. The Figure 5-9 shows the architecture of the

UNet models that have been used in this project. The UNet in Figure 5-9 (a) is a UNet with 31.12 million parameters, it has a $512 \times 512 \times 3$ input size and more layers.

Thus, we were able to preserve more details from the input images. Figure 5-9 (b) depicts a UNet with fewer parameters(7.73 million) and depth, enabling it to achieve faster training speeds. When using a computer with an RTX3060 12 GB graphics card and the batch size was set to 6, the smaller UNet completes an epoch in only 14 seconds, whereas the larger model takes 64 seconds per epoch.

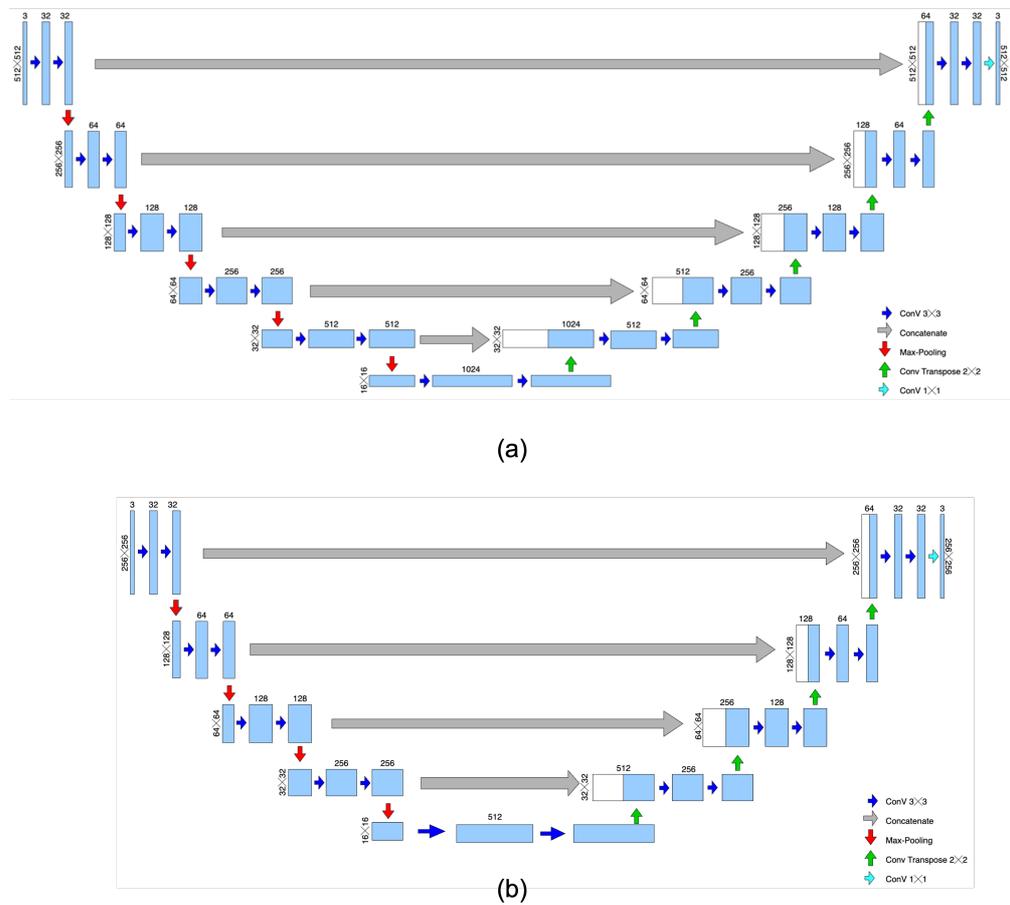


Figure 5-9 The architecture of the optimised UNet. (a) UNet with input size $512 \times 512 \times 3$, and the model includes five down-sampling processes and five up-sampling steps. (b) the UNet with reduced layers and a smaller input size ($256 \times 256 \times 3$).

The UNet has two parts: the left part was a typical convolutional neural network (CNN) for feature extraction, which can be regarded as a down-sampling process (encoder), and the right side was up-sampling (decoder). There are several connections (the grey lines in

Figure 5-9) between the encoder and decoder layers. This operation could address the context-missing issue caused by down-sampling. In this process, the feature maps from the encoder were copied and connected with the feature maps in the decoder to maximise the preserved features.

A conditional random field (CRF) layer was implemented to a UNet as the output layer for the classification task. This model architecture was first proposed by Shuai Zheng *et al.* in 2015 [164]. The idea of using CRF is because the convolutional process and max pooling in a CNN model could result in unclear boundaries and noisy dots in the predicted label mask.

The CNN provided predictions only based on considered regional features of the image, but in semantic segmentation tasks, the global features also need to be considered. For example, one vessel should only belong to one class (vein or artery). CRF could solve this issue by turning the label assignment issue into a probability inference problem.

As TensorFlow does not provide a CRF layer, a customised CRF RNN layer designed by S. Jayasumana *et al.* [165] was implemented to test the feasibility of using CRF in this study.

5.2.2.2 PsPNet

The Pyramid Scene Parsing Network (PsPNet) was proposed by Hengshuang Zhao *et al.* in 2017 [166]. The PsPNet aimed to capture multi-scale contextual information from the input image by leveraging a pyramid pooling module. Like UNet, the PsPNet also uses a CNN encoder-decoder structure. The encoder was used for extracting high-level features from the input images, and the decoder could restore the low-resolution feature maps to the original image size.

Figure 5-10 shows the PsPNet that has been used in this project. In this study, a ResNet50 model was used as the CNN of the PsPNet.

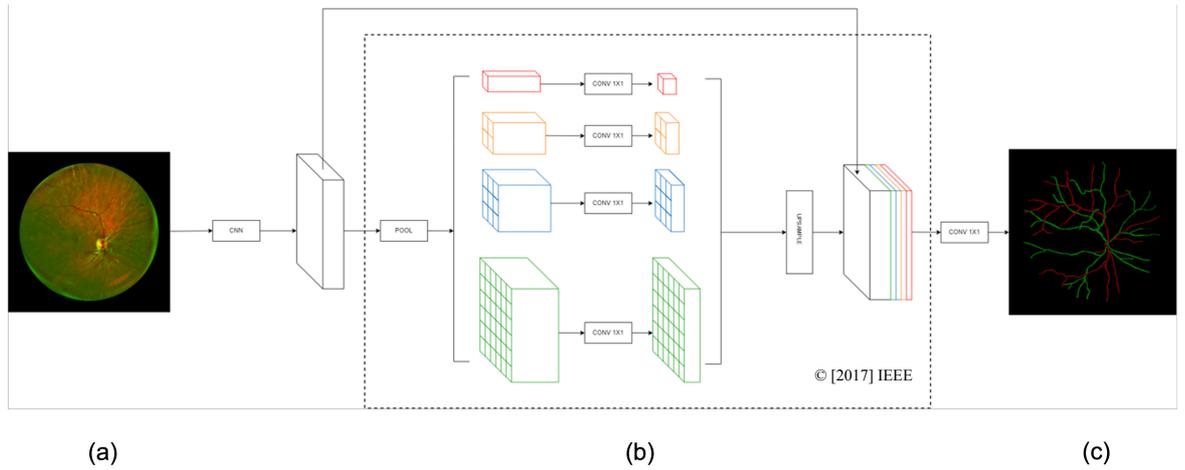


Figure 5-10 Overview of the PsPNet used for retina vessel segmentation and classification. (a) the input image, (b) the pyramid pooling module [166], and (c) the predicted label.

5.2.2.3 Attention UNet

The attention mechanism has been proved that could improve the model performance by enhancing the weight of some crucial areas of the input and reduce the weight of the regions that the model is not paying attention to. In the retina vessel segmentation task, the attention mechanism can reduce the background noise and emphasis the relevant features.

The architecture of the Att-UNet is shown below:

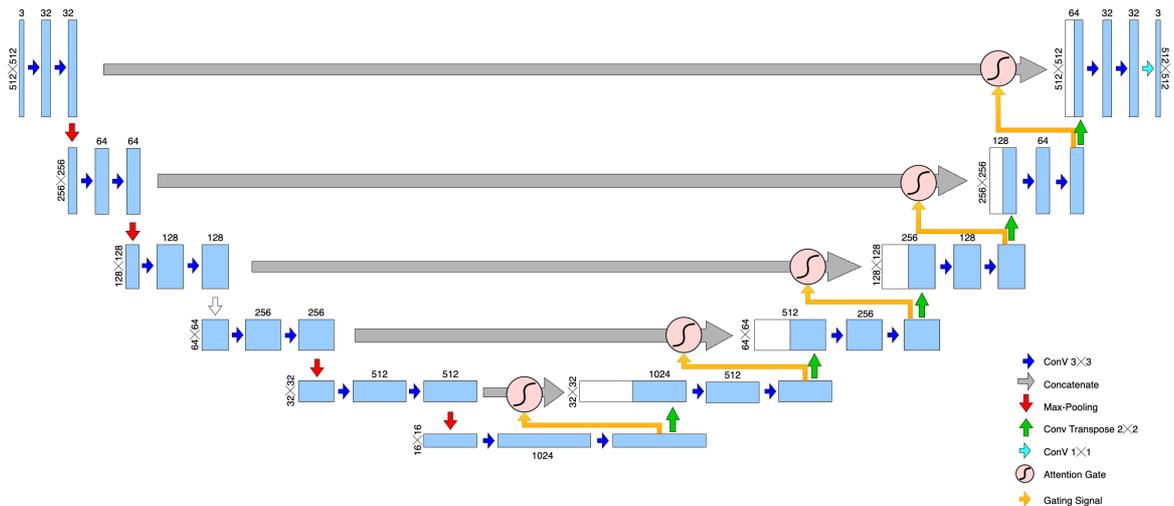


Figure 5-11 The architecture of the Attention UNet. An attention gate was added to the connections between the upsampling and down-sampling steps.

Using the attention mechanism in UNet has first been proposed by Oktay *et al.*[167]. They implemented an attention gate (AG) (Figure 5-12) to the UNet to enhance the features that are relevant to the target. The AG takes two input vectors g ($H_g \times W_g \times D_g$) and x^l ($H_x \times W_x \times D_x$), the x^l is the feature maps of the layer l , and g is the feature maps from the corresponding lower layer from the up sampler ($H_x = H_g \times 2$, and $W_x = W_g \times 2$). The two vectors go through a $1 \times 1 \times 1$ convolution (the W_g and W_x) to have the same dimension.

Subsequently, the two vectors do the element-wise sum and go through a ReLU activation function, during this process, the aligned features will gain more weight. After, a $1 \times 1 \times 1$ convolution ψ , the vector dimension will be reduced to $1 \times W_g \times W_g$, and the sigmoid function produces the attention coefficients by scaling the vector to the $[0,1]$ range. Finally, the resampler up sample the attention coefficient matrix to $W_x \times W_x$ with trilinear interpolation, and multiple with the original x^l on elementwise.

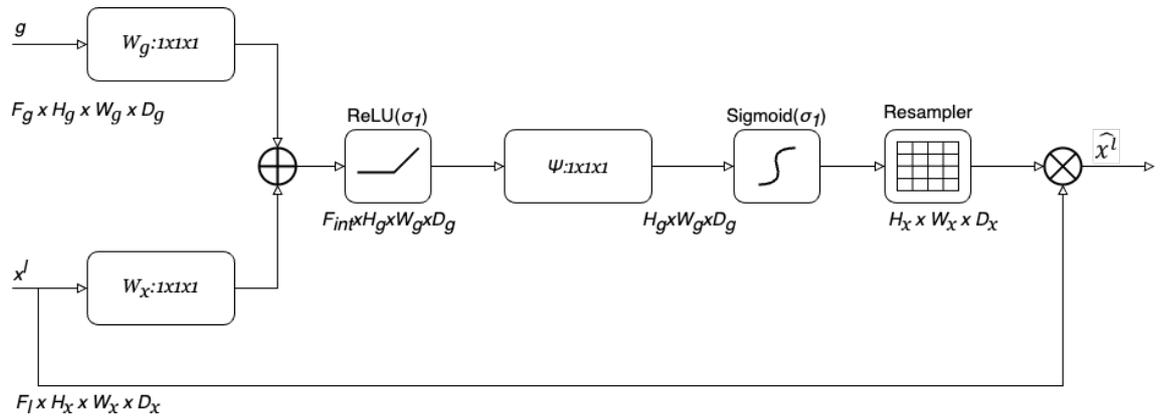


Figure 5-12 Attention Gate [167]. The AG can produce an attention coefficient matrix α and contribute to the feature map x^l . The AG takes two feature maps (x and g) as its input, the gating signal g has a smaller scale and better feature representation

5.2.3 Model training

The training environment includes a Windows desktop with an RTX3060(12G) graphic card and Google Colab. The batch size was set to 4 due to the limitation of the graphic RAM, and 50 epochs were enough for the model to converge. The sparse categorical cross-entropy loss function was used for all models.

To assess the performance of the trained Attention UNet, standard evaluation metrics were used for semantic segmentation, including pixel classification accuracy and mean Intersection over Union (MIoU). Pixel accuracy measures the percentage of correctly classified pixels, while MIoU calculates the average overlap between the predicted and ground truth segmentation maps.

Figure 5-13 illustrates how IoU is calculated. IoU is a value that ranges from 0 to 1 and indicates the accuracy of the prediction. A value closer to 1 means that the prediction is more accurate. MIoU is the average of all class IoU values in multi-class semantic segmentation.

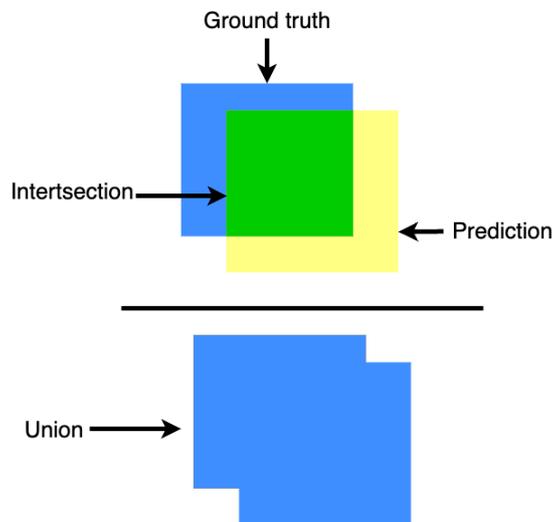


Figure 5-13 Intersection over union(IoU). IoU is the ratio of the intersection area (the green square) over the union area.

5.2.4 Vessel tracking

Dr Ochoa developed an accurate retina vessel classification approach, which used the vessel's optical density (OD) to do the A/V classification [168], because the vein and artery have different light absorption characteristics so that we can differentiate arteries and veins from the retina vessel images by measuring the OD of the blood vessels.

The differential light absorption characteristics between arteries and veins enabled the differentiation of arterial and venous vessels by analysing the absorption patterns of different colours of light. By drawing a perpendicular line along the vessel, a profile vector can be extracted, and then the profile vectors will be fitted by a Gaussian function

$g(x)$ and a linear function $l(x)$ (Equation 5). Figure 5-14 demonstrates the idea of getting the profile and calculating OD.

$$f(x) = g(x) + l(x) = a * e^{-\frac{(x-c)^2}{b}} + p * x + q \quad \text{Equation 5}$$

However, this method was time-consuming as it required manual drawing and extracting the profile. This process can be accelerated by using the DNN since the DNNs have good performance in vessel segmentation.

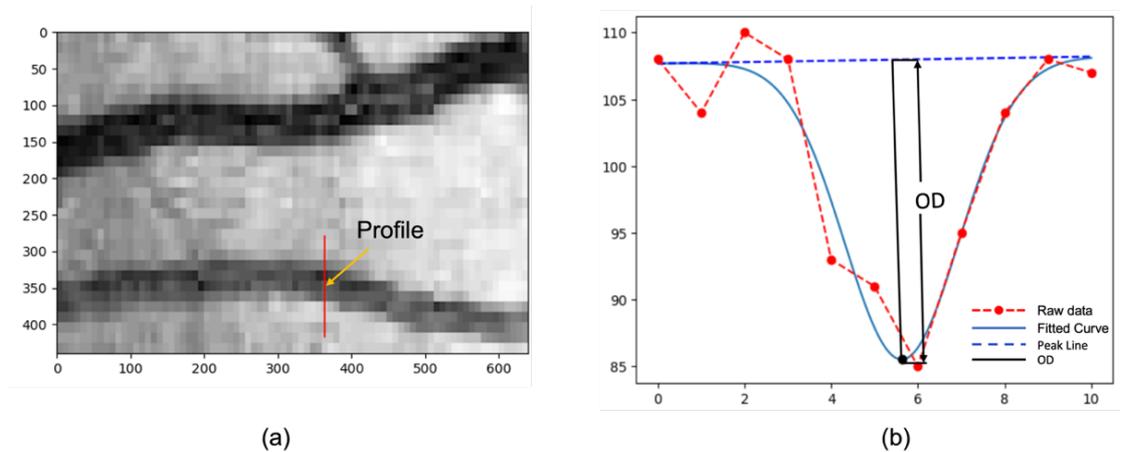


Figure 5-14 Vessel profiling and OD calculation. (a) shows a segment of the retina image. The red line is a hand-drawn profile. In (b), the red dashed line is the data of the profile, the blue line is the fitting result, and the black line shows the OD value.

5.2.5 Profile extraction using semantic segmentation

The output of the DNN is a binary image that contains all the detected vessels (Figure 5-15 (a)). To obtain the vessel profiles, the coordinates of each vessel are needed. Thus, a vessel tracking algorithm was designed for this task.

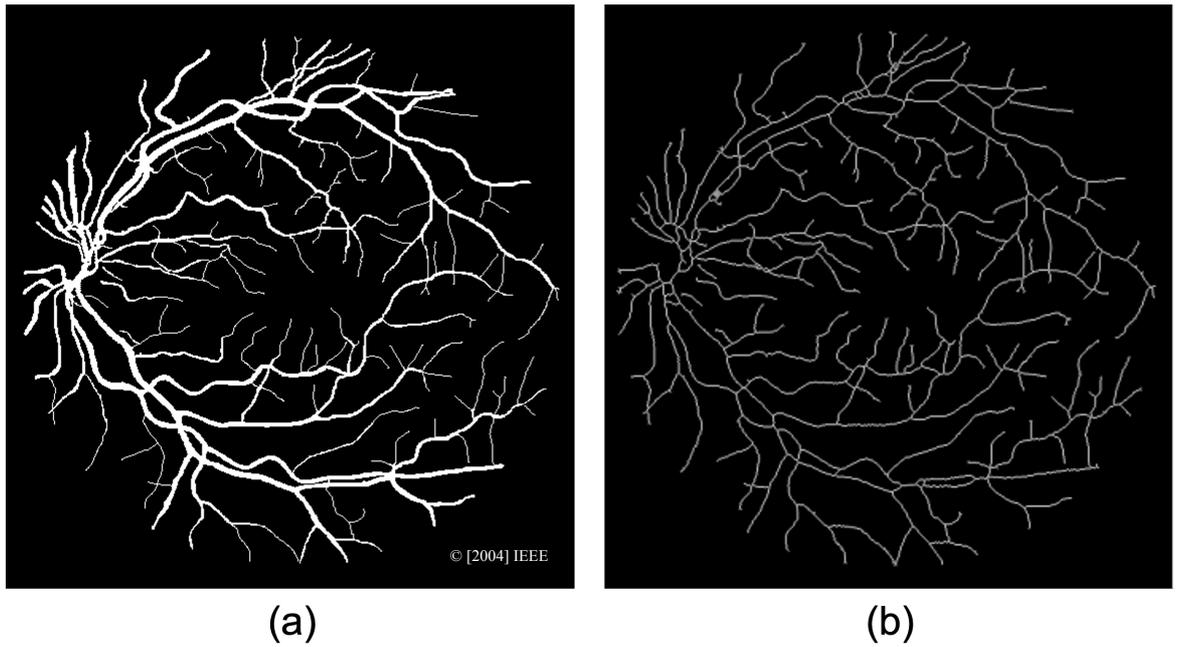


Figure 5-15 The vessel label(a) and the vessel skeleton(b).

The vessel tracking algorithm includes three steps:

1. Skeletonise the label map (Figure 5-15(b)) of vessels and get the distance between the central line and the vessel edge.
2. Scanning the picture, find the vessel skeleton and extract the coordination of each pixel that contains the vessel skeleton.
3. Drawing the profile line based on the vessel skeleton and the distance.

Figure 5-16 shows an example of a binarised vessel skeleton. This step was performed by using the scikit-learn package [169]. The “medial_axis” function could return both the central line and the distance between the central line and the edge.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

Figure 5-16 An example of the binarized skeletonised vessel segment. “1” represents the vessel skeleton “0” for the background.

After the vessel label image has been skeletonised, a 3*3 sliding window will scan the padded image using the method shown in Figure 5-17. When the centre of the sliding window is 1 and the number of ones ≤ 3 , the sliding window will follow the direction of the vessel skeleton and record the coordinates.

Once the sliding window approaches the crossing point (Figure 5-17(d)), the function will return the coordinate array of the vessel segment and remove the scanned vessel from the image. This could avoid scanning the same vessel multiple times. Then, the algorithm will restart the scanning until all the vessels have been scanned.

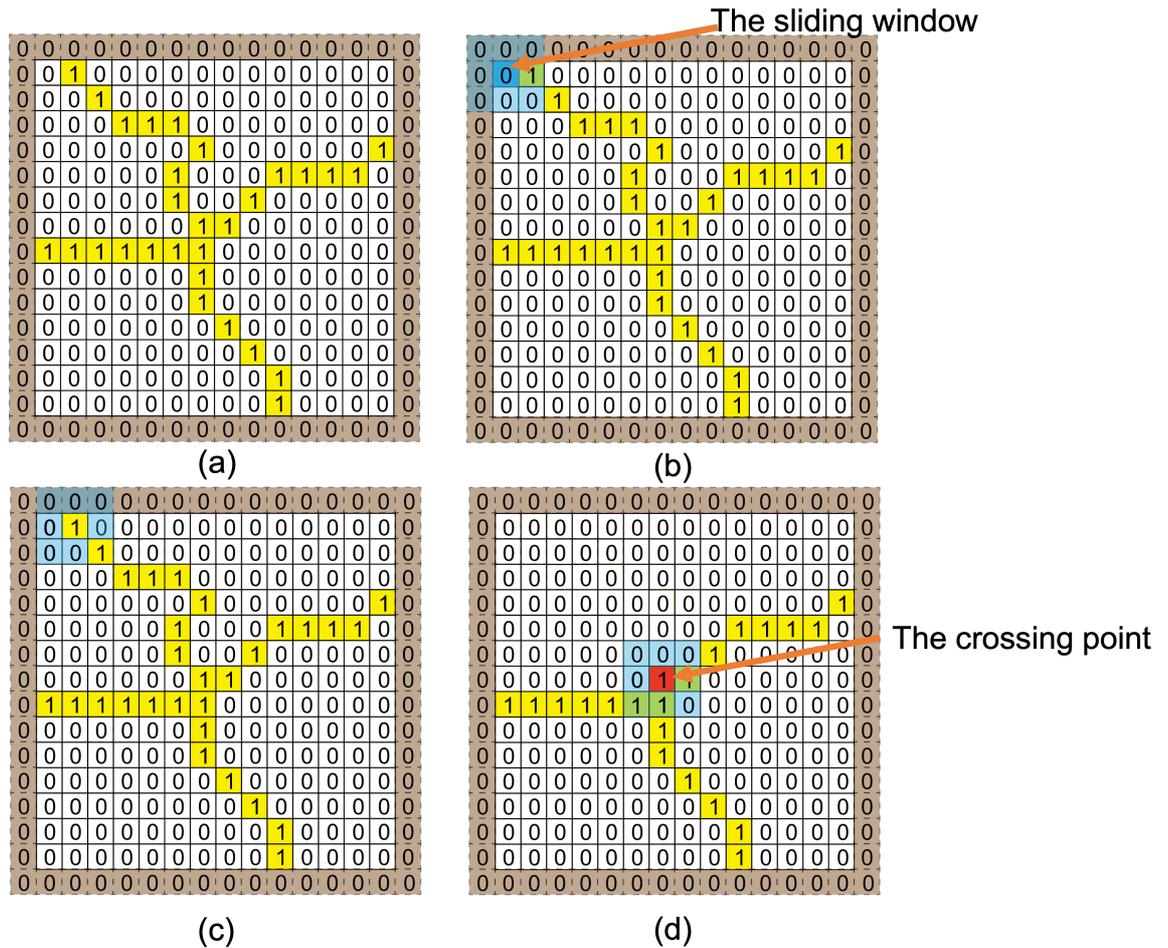


Figure 5-17 Vessel tracking process. (a) Adding 1-pixel padding to the skeletonised vessel picture. (b) A 3*3 slide window at the starting point. (c) The sliding window will scan every pixel alone on the x-axis until the centre pixel value of the sliding window is 1. (d) The sliding window will follow the non-zero values until approaching the crossing point.

The most crucial aspect of the algorithm was determining the direction in which the sliding window would move. To locate the starting point of the vessel, the sliding window will scan the image pixel by pixel. Once the centre of the sliding window detects a 1 and there are two additional ones within the sliding window, that point will be considered the starting point(Figure 5-17(c)). The sliding window then shifts to the "1" that is off-centre and has not been scanned. There are 8 possible directions that the sliding window can go, as shown in Figure 5-18.

The tracking algorithm performed the count of ones within the sliding window to determine the scanning status. To ensure that the sliding window moves in the correct

direction, recording the previous movement of the sliding window is necessary. A notation of right, top right, top, top left, left, bottom left, bottom, and bottom right, was utilised to document the sliding window's most recent movement direction. For example, if the last move of the sliding window is bottom left, then the next position can only be top left, left, bottom left, bottom, or bottom left.

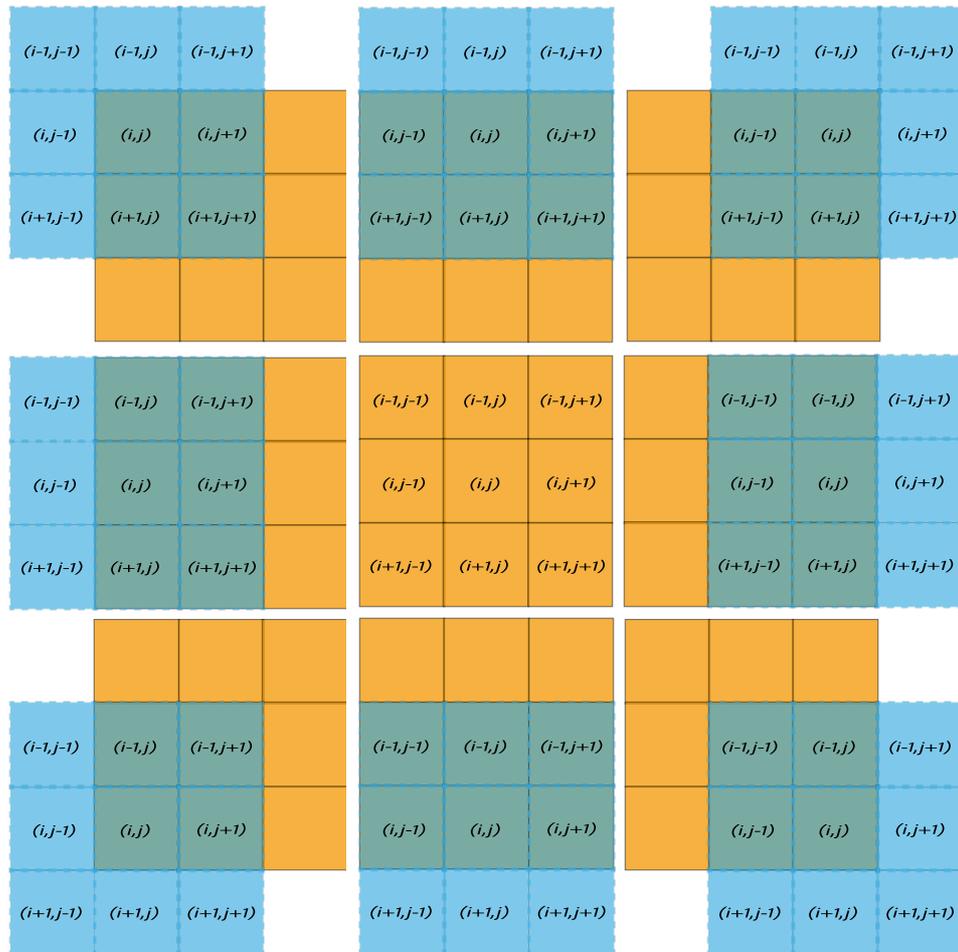


Figure 5-18 The tracking function at pixel (i, j) , the yellow squares are the position of the current sliding window, and the blue squares show the direction of the sliding window movement. If the sliding window approaches pixel (i, j) and its value is 1, then check the values of the other eight pixels around pixel (i, j) to decide the direction of scanning.

After the scanning, the coordinates of the vessels' central lines will be saved as a list, and the position of the profiles will be calculated based on these coordinates. Figure 5-19

illustrates the four scenarios in calculating the coordinates of the profiles. In the case of Figure 5-19 (a) (b) and (c), the coordinates of p3 and p4 can be calculated with the following equations:

$$k = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{Equation 6}$$

$$x_3 = \frac{(x_1 + x_2)}{2} - k * \sqrt{\frac{(2d * m)^2}{4(1 + k^2)}} \quad \text{Equation 7}$$

$$y_3 = \frac{(y_1 + y_2)}{2} + \sqrt{\frac{(2d * m)^2}{4(1 + k^2)}} \quad \text{Equation 8}$$

$$x_4 = \frac{(x_1 + x_2)}{2} + k * \sqrt{\frac{(2d * m)^2}{4(1 + k^2)}} \quad \text{Equation 9}$$

$$y_4 = \frac{(y_1 + y_2)}{2} - \sqrt{\frac{(2d * m)^2}{4(1 + k^2)}} \quad \text{Equation 10}$$

In the above formula, “k” represents the slope of line segment P1P2, m represents the multiple by which it extends relative to “d”, and it is typically chosen as 1.2. In the case of Figure 5-19 (d), the k does not exist, and $x_2 = x_1$. The p3 and p4 coordinates were:

$$x_3 = \frac{(x_1 + x_2)}{2} + d * m \quad \text{Equation 11}$$

$$x_4 = \frac{(x_1 + x_2)}{2} - d * m \quad \text{Equation 12}$$

$$y_3 = y_4 = \frac{(y_1 + y_2)}{2} \quad \text{Equation 13}$$

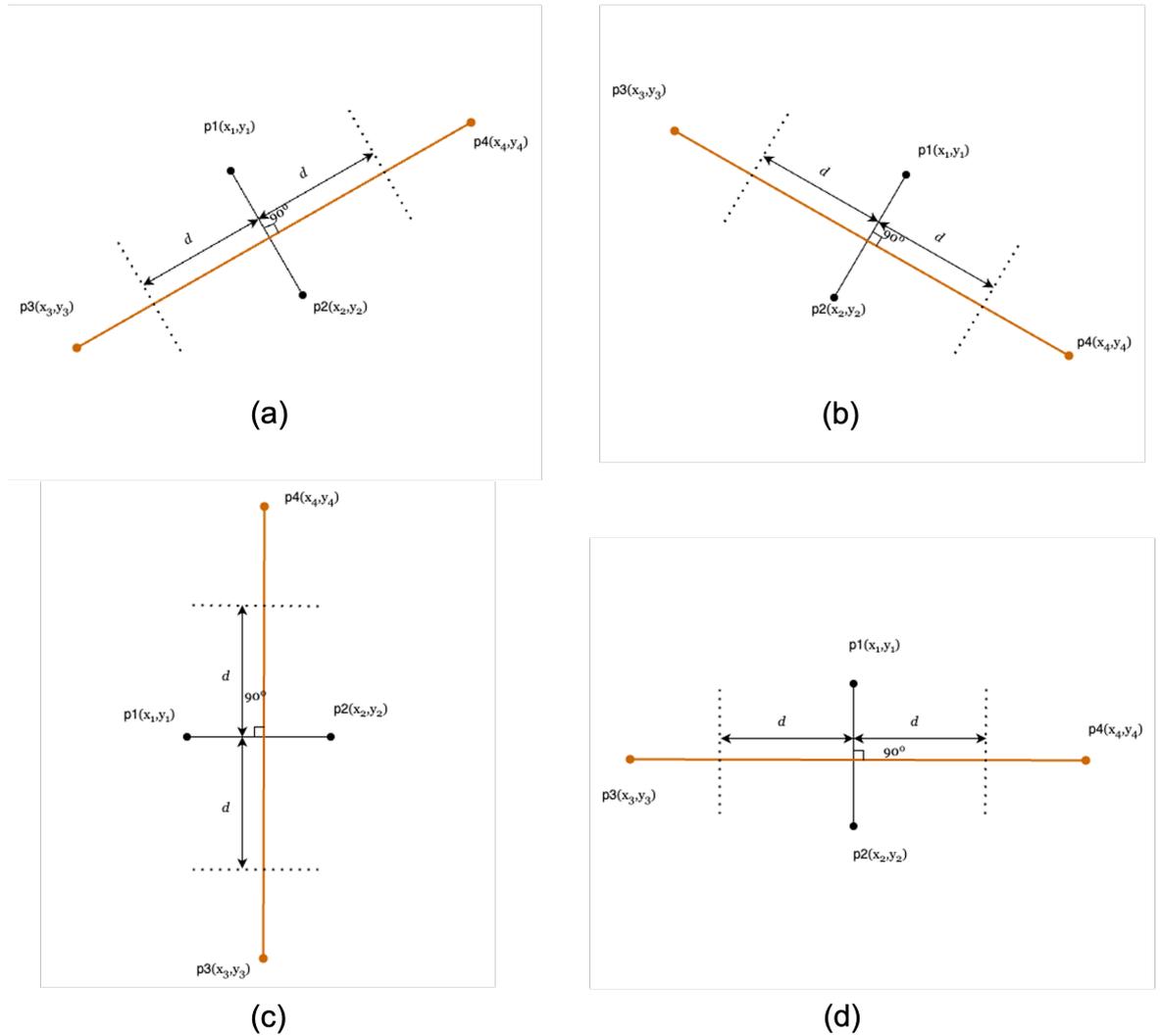


Figure 5-19 Profile calculation. $P1$ and $P2$ are two points on the central line of the blood vessel, and the distance between these two points can be adjusted based on the specific circumstances, typically by selecting adjacent or spaced points. The dashed line represents the edge of the blood vessel, and “ d ” is the distance between the centre and the edge of the blood vessel. $P3$ and $P4$ are two endpoints of a cross-section of the blood vessel.

5.3 Results and discussion

5.3.1 Vessel classification and segmentation

The deep neural networks, including UNet PspNet and Att-UNet, show outstanding performance in vessel segmentation, but the classification performance needs to be improved. The major problems are the discontinuity of tiny vessels and the vessel’s

mislabelling. The DNNs were evaluated using two different test datasets, one from the DRIVE dataset and another from the UWF image captured by the handheld retinopathy of prematurity (RoP) camera.

5.3.1.1 Vessel segmentation

Table 5-2 provides the classification accuracy of the models in doing the vessel segmentation task, and Table 5-3 provides the MIoU of the models. The performance of the UNet with 256*256 input size and fewer layers was comparable to the larger one in terms of accuracy and MIoU, despite having fewer parameters and layers. By incorporating the attention mechanism, the accuracy of the model was improved. The Attention UNet model outperformed the other models. It achieved an accuracy of 0.96, 0.88 sensitivity, 0.97 specificity, and MIoU 0.86 on the Drive dataset, which is comparable with the related work in Table 5-1. However, the PsPNet model did not perform well and only achieved an MIoU score of 0.715 on the DRIVE dataset.

It was observed that all models performed better on the DRIVE dataset, which suggests that more images are needed if we want to use deep neural networks in processing images captured by a low-cost camera. This suggests that more images are needed if we intend to use deep neural networks to process images captured by a low-cost camera. Therefore, it would be beneficial to gather more data using the RoP camera developed by Professor Andy Harvey's team in future work.

Table 5-2 The Vessel segmentation accuracy of UNet, Att-UNet, and PsPNet

| | UNet(256) | UNet(512) | Att-UNet | PsPNet |
|-------|-----------|-----------|----------|--------|
| DRIVE | 0.965 | 0.967 | 0.963 | 0.942 |
| UWF | 0.879 | 0.890 | 0.906 | 0.899 |

Table 5-3 The MIoU of UNet, Att-UNet, and PsPNet in vessel segmentation

| | UNet(256) | UNet(512) | Att-UNet | PsPNet |
|-------|-----------|-----------|----------|--------|
| DRIVE | 0.810 | 0.815 | 0.816 | 0.715 |
| UWF | 0.573 | 0.638 | 0.669 | 0.663 |

Table 5-4 Sensitivity and specificity of different models on the DRIVE dataset

| | Sensitivity | Specificity |
|-----------|-------------|-------------|
| UNet(256) | 0.778 | 0.982 |
| UNet(512) | 0.875 | 0.970 |
| Att-UNet | 0.881 | 0.970 |
| PsPNet | 0.752 | 0.980 |

5.3.1.2 Artery/Vein (A/V) classification

Table 5-5 displays the outcomes of the model's accuracy in conducting vessel segmentation and artery-vein classification at the same time. and Table 5-6 shows the MIoU of different models in A/V classification.

When compared to the result of segmentation only, the A/V classification result is a little worse. It can be seen that the Att-UNet accuracy and sensitivity dropped to 0.940 and 0.489, respectively. This result is similar to the result from other's studies, as it appears to be more difficult for the DNNs to perform the A/V classification while segmenting the vessels from the background.

Table 5-5 The Vessel classification accuracy of UNet, Att-UNet, and PsPNet

| | UNet | Att-UNet | PsPNet |
|-------|-------|----------|--------|
| DRIVE | 0.932 | 0.940 | 0.923 |
| UWF | 0.928 | 0.945 | 0.879 |

Table 5-6 The MIoU of UNet, Att-UNet, and PsPNet

| | UNet | Att-UNet | PsPNet |
|-------|-------|----------|--------|
| DRIVE | 0.692 | 0.677 | 0.570 |
| UWF | 0.646 | 0.677 | 0.632 |

One effective way to improve the accuracy of A/V classification is by incorporating additional techniques into the deep learning model instead of relying on just one model. For example, Sojun Go *et al.* proposed an approach that combined DNN with fluorescein angiography to achieve an accuracy of 0.948, with 0.950 sensitivity and 0.945 specificities [170]. Another study by Luo S *et al.* developed a two-stage deep learning framework that utilised topological information and enhanced A/V classification accuracy from 0.957 to 0.970 [171]. Additionally, Yi J *et al.* proposed a method that utilised a multi-channel module (MM) to enhance feature input images, as well as a multi-scale transformation (MT) module and multi-feature fusion (MF) module to expand the network's receptive field and enhance the representability of deep features [172].

For this research, I utilised deep learning to segment the blood vessels and obtain the coordinates for enhancing Dr Victor Jalil Ochoa Gutierrez's proposed method. By employing deep learning, the process of drawing profiles along the vessels, which previously took days, was reduced to approximately 5 minutes.

5.3.2 Vessel tracking and profiling.

The vessel tracking algorithm was tested with both the UWF retinal image and the DRIVE dataset. The algorithm can automatically draw profiles along the vessels and export the coordinates of each profile. Figure 5-20 shows the result of applying the vessel tracking algorithm to a segment of the UWF retina image. Most of the profiles can be drawn correctly on both arteries and veins.

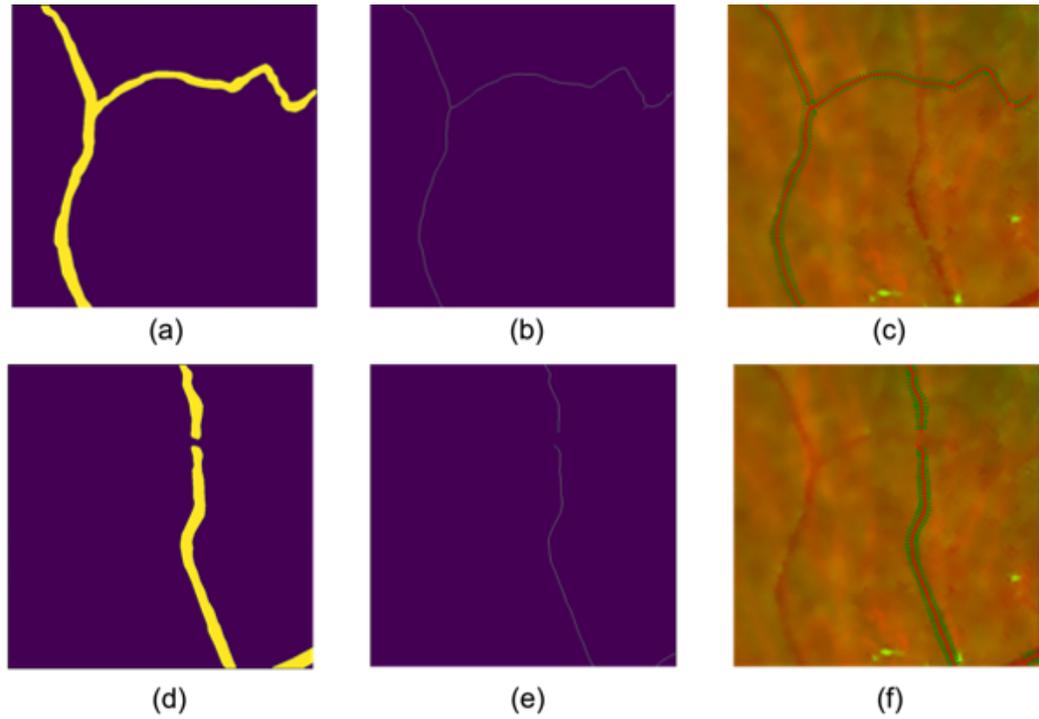


Figure 5-20 The outcome of vessel tracking and profile algorithm (a) and (d) are the binarized labels of a vein and artery, respectively. (b) and (e) is the skeleton of the corresponding vessel. (c) and (f) shows the profiles (green) drawn on the original image.

Software with a graphic user interface (GUI) was also developed using PyQt5, as illustrated in the figure below. To use this software, users only need to upload the retina image and its label, adjust the parameters, and click on the run button. The output that includes the coordinates of the profiles can be saved as a .csv file.

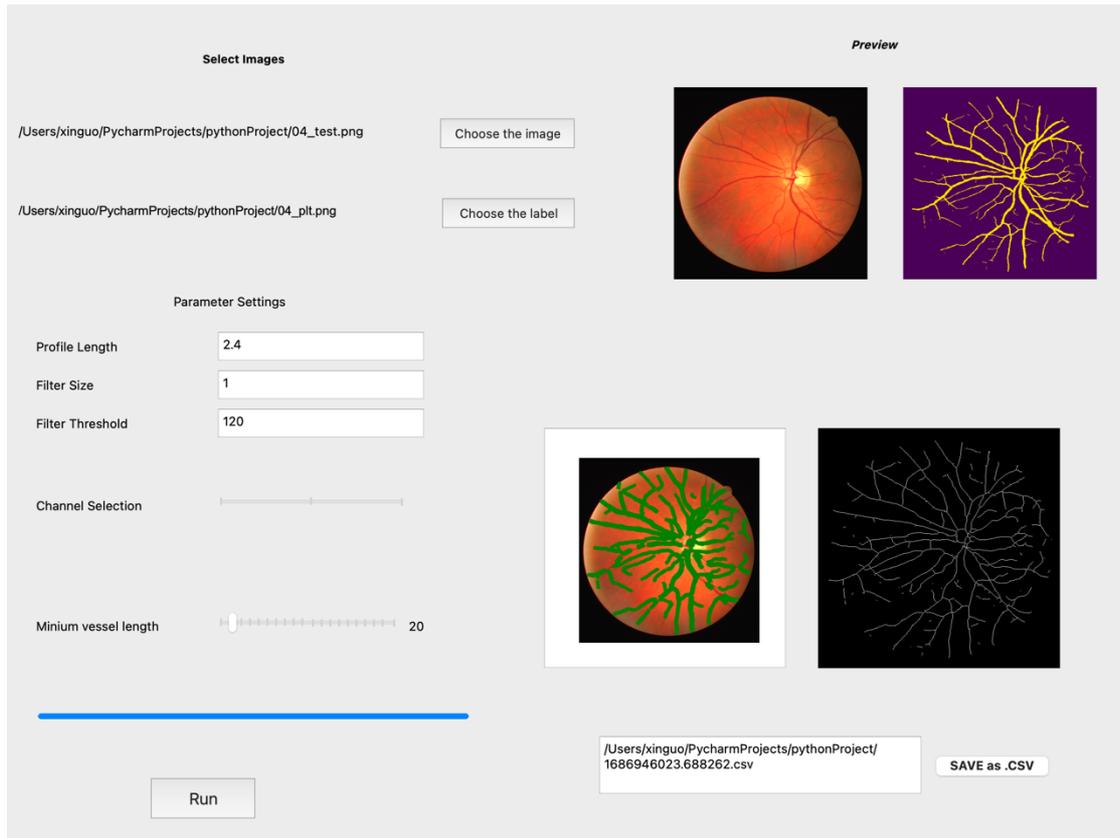


Figure 5-21 The software GUI. Users can select the retina image and its label from the computer and set the parameters, including profile length (multiple of the distance from the centreline to the edge), filter size, threshold, channel selection (the R, G, B channels), and minimum vessel length (ignore the segments that are too short).

It appears that the machine learning-based algorithms used for tracking and profiling vessels are quite precise. These algorithms are able to provide accurate coordinates for most vessels.

5.4 Conclusion

This Chapter focuses on developing neural networks, including UNet, attention UNet, and PsPNet, for identifying blood vessels and classifying arteries and veins. These models are different from the CNN multi-class classification network discussed in the previous Chapter, as they are designed for semantic segmentation tasks, which involve pixel-wise classification. All those model structures were tested for both tasks using different data types, including a public dataset (DRIVE) and a private UWF dataset [168].

Amongst the different models, Att-UNet has exhibited the most impressive performance in both tasks. However, none of the models has been able to attain an MIoU of over 0.8 in the classification of arteries/veins for both datasets. This phenomenon also appeared in other studies (Table 5-1), the models are less accurate in A/V classification than in segmentation only. The main issue is the inconsistent prediction on identical vessels. This implies that although the models can recognise blood vessels from retinal images, they have some difficulty in identifying the type of some blood vessels.

A software was developed to improve our group's vessel extraction method [168], as the models were found to provide a satisfactory performance in vessel segmentation with 0.88 sensitivity and 0.97 specificity. Although this approach was found accurate in previous studies, it was also time-consuming, linked to many manual steps. By incorporating deep learning and a vessel tracking algorithm, the processing time has been significantly reduced from days to just a few minutes. However, more testing is needed to determine if combining the two methods will enhance the accuracy of A/V classification in future research.

Chapter 6 Sample size calculator for diagnostic study

6.1 Introduction

The UK has been under lockdown due to COVID-19 for around two years during 2020-2021, resulting in limited access to labs and suspension of field research, leading to the necessity to adapt our research plans, in particular around field testing and experimental.

A group project resulted from these circumstances in June 2020 aimed to develop a web-based calculator to determine a diagnostic study's minimum required sample size. The team comprised PhD students of a wide range of backgrounds (from epidemiology to optical, software and mechanical engineering). Most students in the team were involved in other research that would have led to diagnostic surveys or field studies that require a rational determination of sample size [173]. Testing a large number of subjects in studies typically requires a significant amount of effort and using a too-large sample size could lead to many issues, such as being time-consuming, resource-wasting and potentially unethical. On the other hand, a limited sample size could reduce the study's statistical power and, consequently, its conclusions and impact. The term power in statistics represents the ability to avoid a type II error (accept a false null hypothesis). In this project, I was responsible for implementing the web page functionality, building upon the statistical theories that were researched and validated by other team members. Additionally, I designed the web page layout in accordance with the storyboard (Appendix 7) developed by our team.

Many studies have been conducted on sample size estimation in diagnostic studies. E.g., Nancy M *et al.* introduced the method of incorporating the prevalence into the sample size calculation in 1996 [174]. They proposed the method of sample size estimation with and without conditions in 1998 [175], David *et al.* proposed a method of estimating the minimum sample size required for a positive likelihood ratio with its respective confidence interval [176]. S Carley *et al.* developed a nomogram to estimate sample size based on the determination of sample size precision [177].

Statistical analysis is necessary for many studies, but not all researchers are proficient in mathematics or statistics [173]. Statistics tools, such as SPSS [178] and PASS [178], can

be expensive and geared towards professionals, making them inaccessible to many potential users. Installing software specifically for specificity and sensitivity analysis may not be necessary. Therefore, there is a need for a simpler, user-friendly, and free tool to calculate sample size. Moreover, this project can be carried out without the need for laboratory equipment, enabling us to work from home amidst the lockdown.

6.2 Methods

6.2.1 Waterfall methodology

This project was completed using the widely used waterfall software engineering methodology [179] waterfall model can aid in organising tasks, minimising human errors in team collaborations, and boosting overall productivity. Figure 6-1 illustrates the five steps of the waterfall: requirements analysis, system design, coding, testing, and maintenance. To promote knowledge sharing and shorten the software development cycle, different project steps were assigned to different team members.

All of the software designs and coding tasks were performed as part of this thesis.

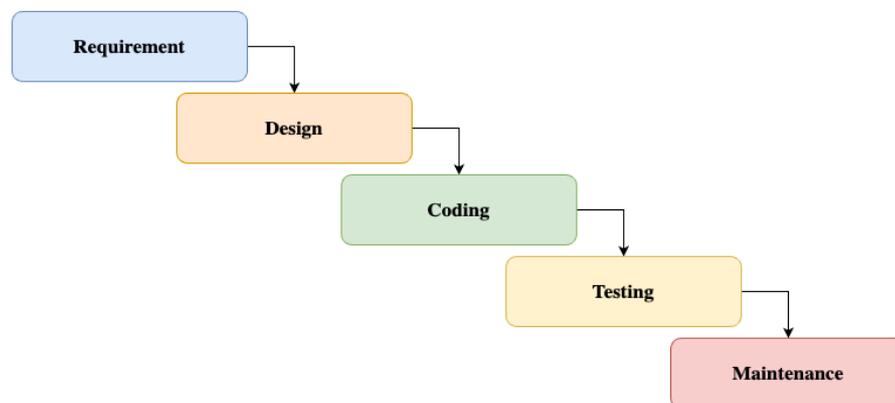


Figure 6-1 The waterfall methodology. The App development process was divided into five stages: analysing requirements, designing, coding, testing, and maintaining the code.

Various methods, including storyboarding and fast code prototyping, were also utilized to enhance collaboration and continuously improve the software. A detailed storyboard was created that outlines the statistical theory and web interface design. It includes a comprehensive description of both aspects. The fast-prototyping process required an Excel file to validate the calculator. The storyboard can be found in Appendix 7.

6.2.2 Webpage development

The web App was developed based on Angular 9, one of the most popular web development frameworks. It provided a solid foundation for creating a dynamic and interactive user interface. Additionally, various Angular libraries and components were leveraged to enhance the functionality and user experience of the web tool.

Angular uses TypeScript and HTML to create applications for computers and mobile devices. The platform is constructed with a modular architecture that includes different building blocks called components. Each component comprises a TypeScript class, an HTML template, and styles. This modular approach gives developers the freedom and flexibility to create applications according to their unique needs during the development process.

6.2.3 User interface design

The user interface (UI) of the web tool was designed to be intuitive and user-friendly. Users are presented with input fields where they can enter the desired sensitivity and specificity values. Clear instructions and tooltips are provided to guide users in entering valid inputs. The user interface design was based on the storyboard which can be found in the Appendix 7.

The UI of the initial version of the sample size calculator is displayed in Figure 6-2. It is divided into four sections: title, introduction, calculation, and references. There were three calculation modes available: sensitivity, specificity, and a combination of both to estimate sample size. At the bottom of the page, the corresponding formulas and helpful references are provided.

Sample Size calculator for diagnostic accuracy

The identification of the proper sample size in evaluating the accuracy of a diagnostic test in medicine is of crucial importance.

This is a calculator that helps you to determine the number of samples/subjects you need to recruit in order to define the sensitivity and specificity of your diagnostic test with the desired level of accuracy

Select a mode

Based on Sensitivity
 Based on Specificity
 Based on Sensitivity&Specificity

Prevalence, $Prev$ %
 Sensitivity, Se %
 Specificity, Sp %
 Maximum marginal error, M %
 Confidence level, $(1-\alpha)$ ▼
 Likelihood, q %

Calculate

Clear

Estimated sample size N_{Se} (based on sensitivity)
 Estimated sample size N_{Sp} (based on specificity)
 Estimated Positive Predictive Value(PPV)
 Estimated Negative Predictive Value(NPV)

Formula for the computation of sample size based on sensitivity: $N_{Se}' = \frac{Z_{\alpha/2}^2 Se(1-Se)}{M^2 \times prev}$ $N_{Se} = \frac{N_{Se}'}{(1-q)}$

Formula for the computation of sample size based on specificity: $N_{Sp}' = \frac{Z_{\alpha/2}^2 Sp(1-Sp)}{M^2 \times (1-prev)}$ $N_{Sp} = \frac{N_{Sp}'}{(1-q)}$

PPV(Estimated positive predicted value): $PPV = \frac{Se \times Prev}{Se \times Prev + (1-Sp) \times (1-Prev)}$

NPV(Estimated negative predicted value): $NPV = \frac{Sp \times (1-Prev)}{(1-Se) \times Prev + Sp \times (1-Prev)}$

[1] Hess, A. S., et al. "Methods and recommendations for evaluating and reporting a new diagnostic test." *European journal of clinical microbiology & infectious diseases* 31.9 (2012): 2111-2116.

[2] Bartlett, II, J. E., et al. "Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research." *Information technology, learning, and performance journal* 19.1 (2001): 43.

[3] Hajian-Tilaki, K. "Sample size estimation in diagnostic test studies of biomedical informatics." *Journal of biomedical informatics* 48 (2014): 193-204.

Figure 6-2 The user interface of the sample size calculator V1.0. There was a brief introduction to the calculator on the top of the page, and under that, three radio buttons were provided for different calculation modes. The inputs include the prevalence of a disease ($Prev$), sensitivity (Se) and specificity (Sp) of a diagnostic test, the maximum marginal error (M), the confidence level of the statistic test, and likelihood($1-\alpha$) of the test failure. The results include the sample size based on sensitivity and specificity, positive predicted value (PPV), and negative predicted value (NPV). The formulas used for calculation were rendered by a node module named MathJax and displayed under the result section. Also, the references were provided at the bottom.

Following the waterfall methodology, some issues were identified in the first version during the test phase, including the compatibility problem on mobile devices and limited description of some functions. New requirements were suggested by our group members, including adding more calculation modes and hypothesis tests.

In the second development phase, the webpage incorporated responsive design principles, allowing it to adapt to different screen sizes and devices for a seamless user experience. The UI colour scheme was enhanced, and animated effects were incorporated.

The UI of different functional areas were placed in separate components to enhance the program’s scalability. This not only simplifies the process of adjusting the UI but also prevents any impact on other components when adding new ones. Figure 6-3 shows the updated UI on both laptops and mobile phones.

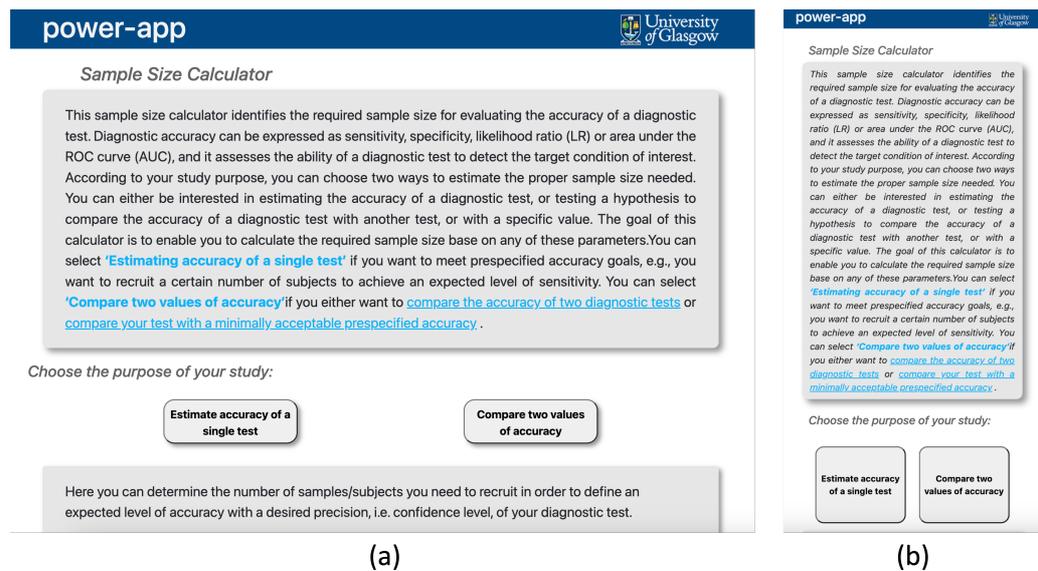


Figure 6-3 The UI with responsive design. (a) a screenshot captured from PC, (b) a Screenshot captured from a mobile phone.

The modularised design of Angular allows the implementation of new components to the project easily by using the “@Component ()” decorator. To integrate a new component with the existing project, only their relationship in the App module file needs to be declared. Additionally, a routing module was added to the app, it allows each page to have its own URL.

6.2.4 Implementing website functionality

The first version of the calculator was designed as a single-page website with only one Angular component. It could be used for estimating the sample size based on sensitivity and specificity (as shown in Figure 6-2). The inputs include the desired sensitivity (Se) and specificity (Sp), the prevalence (Prev) of the disease, Maximum marginal error (M), confidence level (1- α) and failure rate in sampling (q). The formula below shows the calculation of the estimated sample size based on the sensitivity.

$$N_{se}' = \frac{Z_{\frac{\alpha}{2}}^2 \times Se(1-Se)}{M^2 \times Prev} \quad \text{Equation 14}$$

$$N_{se} = \frac{N_{se}'}{1-q} \quad \text{Equation 15}$$

The minimum sample size needed to achieve the desired sensitivity, without considering the failure rate of sampling, is denoted as N_{se}' . $Z_{\frac{\alpha}{2}}$ represents the Z-score with confidence interval α , a table of commonly used Z-score is provided in Appendix Table 5. Similarly, the following equations were used for estimating the sample size based on specificity.

$$N_{sp}' = \frac{Z_{\frac{\alpha}{2}}^2 \times Sp(1-Sp)}{M^2 \times (1-Prev)} \quad \text{Equation 16}$$

$$N_{sp} = \frac{N_{sp}'}{1-q} \quad \text{Equation 17}$$

When considering both sensitivity and specificity, the positive predictive value (PPV) and negative predictive value (NPV) will also be calculated to show the estimated proportion of true positives and negatives of the diagnostic test. The calculation of PPV and NPV can be formularized as follow:

$$PPV = \frac{Se \times Prev}{Se \times Prev + (1-Sp) \times (1-Prev)} \quad \text{Equation 18}$$

$$NPV = \frac{Sp \times (1-Prev)}{(1-Se \times Prev) + Sp \times (1-Prev)} \quad \text{Equation 19}$$

The updated web page implemented hypothesis tests and sample size estimation based on the area under the receiver operating characteristic (ROC) curve (AUC). The sample size estimation based on AUC was implemented as a new function of the previous sample size calculator. This function was designed to calculate the required number of samples or subjects for estimating the AUC of the ROC curve of a diagnostic test.

The sample size for estimating the AUC of ROC curve consists of two parts, the sample

size with $(N_D = \frac{Z_{\alpha}^2 \times V(AUC)}{M^2 \times (1-q)})$ and without conditions of interest ($N_{ND} = R \times N_D$) [175].

The overall sample size needed is $N = N_D + N_{ND}$. In the calculation of N_D , the $V(AUC)$ can be calculated by:

$$V(AUC) = N_D \text{Var}(\widehat{AUC}) = (0.0099 \times e^{-a^2/2}) \times \left((5a^2 + 8) + \frac{a^2+8}{R} \right) \text{ Equation 20}$$

The $\text{Var}(\widehat{AUC})$ denotes the variance of \widehat{AUC} [180], $a = \varphi^{-1}(AUC) \times 1.414$ and φ^{-1} is the inverse of standard cumulative normal distribution, the φ^{-1} value can be found in Appendix Table 7.

The hypothesis test comprises two functions, testing if the accuracy of a single diagnostic test achieved the desired accuracy and comparing if the accuracies of two tests are equivalent. The following figure displays the required inputs for testing the accuracy of a single diagnostic test.

Chapter 6 Sample size calculator for diagnostic study

| | |
|--|------------------------|
| Sensitivity/Specificity of reference, P_0 | <input type="text"/> % |
| Sensitivity/Specificity of the new test, P_1 | <input type="text"/> % |
| Significance level, α ⓘ | <input type="text"/> % |
| Power, $(1-\beta)$ ⓘ | <input type="text"/> % |
| Failure rate in sampling, q ⓘ | <input type="text"/> % |

Calculate

Clear

Figure 6-4 Testing the accuracy of a single test.

In testing the accuracy of a single test, the null hypothesis assumes that the test sensitivity or specificity p_1 is the same as the pre-specified one p_0 . The $\alpha\%$ represents the likelihood of mistakenly rejecting the null hypothesis, which serves as evidence for demonstrating the presence of the phenomenon [12]. The power $(1-\beta)\%$ refers to the likelihood that the study will identify a variation in accuracies, assuming that the phenomenon being studied is indeed present [181]. The calculator uses a two-tailed z-score test under the null and alternative hypothesis, and normal approximation is the general rule [180]. To determine the sample size, Equation 21 was used for the conditional approach, and Equation 22 for the unconditional approach. The Z-scores can be found in Appendix Table 6.

$$N = \frac{\left[\frac{Z_\alpha \sqrt{p_0(1-p_0)} + Z_\beta \sqrt{p_1(1-p_1)}}{2} \right]^2}{(p_1 - p_0)^2 \times (1 - q)} \quad \text{Equation 21}$$

$$N = \frac{\left(\frac{Z_\alpha + Z_\beta}{2} \right)^2 p_1(1 - p_1)}{(p_1 - p_0)^2 \times (1 - q)} \quad \text{Equation 22}$$

When comparing the accuracies of two tests, there is an additional input factor to consider: the allocation ratio ($k = \frac{n_1}{n_2}$). The n_1 and n_2 are the determined sample size of the two independent diagnostic tests. The sample size for the reference test (n_2) can be obtained

by using Equation 23, in which $p = \frac{kp_1 + p_2}{1+k}$, with the conditional approach and Equation 24 with the unconditional approach. The sample size needed for the reference group $n_1 = k * n_2$, and the total number of subjects $N = n_1 + n_2$.

$$n_2 = \frac{N_2}{(1-q)} = \frac{\left[\frac{Z_\alpha}{2} \sqrt{\left(1 + \frac{1}{k}\right) p(1-p)} + Z_\beta \sqrt{\frac{p_1(1-p_1)}{k} + p_2(1-p_2)} \right]^2}{(p_1 - p_2)^2(1-q)} \quad \text{Equation 23}$$

$$n_2 = \frac{N_2}{(1-q)} = \frac{\left(\frac{Z_\alpha + Z_\beta}{2} \right)^2}{(1-q)(p_1 - p_2)^2} \left[\frac{p_1(1-p_1)}{k} + p_2(1-p_2) \right] \quad \text{Equation 24}$$

A continuity factor was applied to address the discrepancy between the continuous and discrete nature of the data. The formula for determining the sample size of group 2 with the continuity factor is as follows:

$$N_2 = \frac{N_2}{4} \left[1 + \sqrt{1 + \frac{2(k+1)}{kN_2|p_1 - p_2|}} \right]^2 \quad \text{Equation 25}$$

$$n_2 = \frac{N_2}{(1-q)} \quad \text{Equation 26}$$

6.2.5 Validation

To ensure the calculation results' accuracy by utilising the Excel prototype and G*Power software [182], [183]. We tested various combinations of target sensitivity, specificity, and other parameters to confirm that the results obtained from our website matched those from the other two sources.

6.3 Results

A prototype of the sample size calculator was developed and hosted on the Firebase in June 2020. This can be accessed from <https://sample-size-calculator-69531.web.app/>. The first version is straightforward and easy to use. Users can simply estimate the minimum sample size that is required for their study based on sensitivity and specificity. However, this version was not compatible with some mobile phone screens. The accuracy of our

calculator has been verified by an Excel version prototype created by other group members.

The updated version of the webpage allows for dynamic display adjustment. This new feature has been tested on computers and mobile devices, including Android and iPhone. In this update, the project's architecture was modularised. Thus, the extendibility of the calculator was improved, and adding new functions to the App became easier. New functions, including ROC-based and hypothesis test calculations, were implemented, and validated by using G*Power software [182], [183]. The latest version of the website can be accessed at <https://power-app2.web.app/sample-size>.

Chapter 7 Discussion

Advances in technology, such as deep learning, Blockchain and IoT, have eliminated the boundaries of space and time in many aspects, including healthcare and diagnostics. The WHO recognised the potential of developing an eHealth system back in 2005 [183], and today, with the progress of communication and digital technology, eHealth has become a reality. Embracing these technologies can significantly aid medical diagnostic testing, particularly in areas with limited resources. WHO has emphasised that there is a high demand for affordable, high-quality health technologies [184].

As important components of health technologies, medical devices, assistive devices, and eHealth solutions play an important role in saving lives and improving patients' quality of life and well-being [184]. However, there are still many people suffering from the lack of access to medical facilities and equipment.

The smartphone is the most accessible technology in today's world, even in many Low- and Middle-income Countries (LMICs). The number of smartphone users has increased from 3.6 billion to 5.8 billion between 2016 and 2020, and it was predicted to be approximately 7.86 billion in 2028 [185]. Notably, Sub-Saharan Africa is expected to have the highest growth rate in mobile subscriptions globally, with an estimated 613 million subscribers by 2025, representing a growth of approximately 5% [186]. Smartphone adoption is also increasing in sub-Saharan African countries, with Uganda reaching a penetration rate of 23%.

The accessibility and versatility of smartphones provide an opportunity to deliver appropriate and timely medical services to people living in remote areas with limited access to modern healthcare facilities. For example, the Ministry of Health in Uganda uses smartphone apps to provide frontline health workers access to patient healthcare records [187]. And most of the applications that have been mentioned in this thesis can be run on smartphones and tablets.

As the number of mobile connections continues to increase, so do IoT connections, which will approach 15.14 billion devices globally in 2023 [188]. IoT is now predicted to be one

of the leading solutions for enhancing global healthcare services [189]. Increasing the versatility of smartphones will greatly reduce the costs of digital interventions compared to conventional methods, as illustrated in the work on retinal imaging of Chapter 5.

The diagnostic platform for malaria combined the smartphone into the context of IoT, which demonstrated the capability, capacity, and opportunity of edge computing in the area where internet connection can be intermittent. This advancement improved upon the existing cloud-based diagnostics, such as the IoT solution for e-health presented by Kumar et al. [190]. The system showed significant tolerance for internet traffic.

Clinical data is highly sensitive and requires top-notch protection. Fortunately, the tamperproof nature of Blockchain technology offers a secure means of storing this data in the proposed diagnostic platform. The use of wireless connectivity in cloud-based diagnostic testing also raises ethical concerns regarding the transmission of personally identifiable information. Therefore, it is essential to prioritise the preservation of privacy for private and individual data. To address this concern, privacy preservation frameworks have been developed, such as the BeepTrace network proposed by Xu H et al. [191], they designed a framework using mobile phones and the Ethereum Blockchain network for COVID-19 surveillance and tracing. The BeepTrace framework is a secure and trustworthy system and is more complex than the Blockchain we used in Chapter 3. However, like many Ethereum-based applications, it requires cryptocurrency for transactions (gas fees), which can be an additional expense.

It's worth noting that the Blockchain system is not failproof, and some security risks remain to consider. For instance, before the data is fully uploaded to the Blockchain network, it will be temporarily stored on the phone, which opens up the possibility of theft and access. Android phones are vulnerable to phishing links and malicious programs, which may lead to serious security issues, such as private key disclosure [192]. Besides the problems such as double spending, the 51% attack, and the cost can also be obstacles to integrating Blockchain into the existing system [193]. Moreover, in systems like the proposed diagnostic platform, the distribution of app and network cards can be vulnerable. Additional measures may need to be put in place to further improve the security of the

system, for instance, we need to ensure that the keys are provided to the correct individuals.

The energy consumption of Blockchain, especially those using the PoW mechanism, is another major concern in wide adoption of Blockchain technology. In 2021, the equivalent of about 30% of Germany's or 35% of Ireland's energy was consumed by Bitcoin mining [194]. Other types of Blockchain networks can also consume more energy than traditional data centres [195]. The Blockchain system is decentralised, meaning it has more redundancy than traditional centralised approaches. According to a study by Sedlmeir J et al., a public Blockchain system can consume 10^4 to 10^9 times more energy for handling a single transaction than traditional approaches, while even enterprise Blockchain (private or consortium) systems can consume 10 times as much energy [195].

However, the environmental sustainability and cost of Blockchain cannot simply be concluded as how much energy it consumes. It is worth finding a balance between its benefits and drawbacks when considering whether Blockchain technology should be used. The advantage of this technology, like security and transparency, was also found helpful in reducing the carbon footprint in industry by supplying chain management, reducing waste, and providing quality assurance [196]. Additionally, the development of Blockchain has resulted in significantly lower energy demands, compared to the first generation Blockchain networks, by utilising enhanced consensus algorithms and reducing network redundancy [195]. Furthermore, the use of clean energy in Blockchain is becoming increasingly important, with more policies and legislation guiding its development towards greater sustainability [197], [198].

The healthcare industry is utilising AI more frequently in diagnostics to improve the current system, and it has been authorised for use as a medical device [199], [200]. When interpreting results from immunodiagnostic lateral-flow rapid diagnostic tests (RDTs) devices, AI can offer a more consistent, accurate performance than visual interpretation [201], as confirmed by a revision of ASSURE criteria [137]. The malaria diagnostic platform that has been proposed in chapters 2 to 4 is not only ASSURED but also meets the WHO target product profile (TPP) for readers of RDTs [201].

The use of AI in diagnostics also raises concerns around trustworthiness, which is crucial in healthcare [202]. In healthcare, inaccurate predictions can lead to fatal subsequent effects. For example, a false negative could cause a patient to miss out on receiving timely treatment, while a false positive could lead to healthy individuals taking unnecessary medication (or further testing with economic impacts as well as potential health impacts when radiation is involved, for example). Thus, using AI in diagnostics requires caution since most deep neural networks can easily get overfitted. The CNN introduced in Chapter 3 has proved its trustworthiness by its high accuracy and combination of Blockchain technology. During the test, all false predictions will lead to repeating the test, which ensures the trustworthiness of the entire platform and each step in the process.

After successfully using AI to assist in reading test results from microfluidic tests, further investigation was conducted into its capabilities. In Chapter 5, several deep neural networks were utilised for retina image processing. Although the performance of AI was found to not be on par with that of human expertise, it still has shown great promise and potential when combined with other approaches. The advances in model architecture and algorithms can improve the precision of the predictions. For instance, state-of-the-art deep neural networks, diffusion networks and transformer mechanisms recently became popular in computer vision and have been proven to improve the accuracy and stability of medical image analysis significantly [203].

In a low-resource environment, it can be difficult to redo the test. Diabetic retinopathy screening strategy is being transformed by AI, leading to greatly improved cost-effectiveness [206]. Early treatment and continuous screening can effectively prevent blindness caused by asymptomatic diabetic retinopathy (DR) since it can be asymptomatic until advanced stages [207]. However, non-communicable diseases like DR are frequently overlooked in LMICs, unlike communicable diseases.

It has been proven that incorporating other types of image processing in developing automated diagnostic tools can improve overall accuracy and reliability than relying solely on AI predictions [204]. For instance, the system proposed by Xingzheng Lyu et al. attained a higher accuracy by combining pre and post-processing steps, an improved neural network for completing the retina vessel analysis, and a fractal dimension to

indicate the image quality [205]. As some fundus images can be hard to analyse even by experienced ophthalmologists, a typical solution in clinical practice is to retake the photo with better quality.

In conclusion, digital technologies such as AI, Blockchain, IoT, and edge computing offer opportunities to enhance healthcare systems by introducing innovative diagnostic approaches, providing more efficient and secured data management systems, and engaging the public to be aware and better understand the challenges that remain in healthcare. These technologies can provide people in remote rural areas with access to advanced healthcare services at a lower cost while also improving the healthcare experience for those in urban cities.

This study proposes possible solutions to overcome challenges in conducting diagnostics in low-resource settings and improving current diagnostic methods. However, introducing new technologies into diagnostics always requires caution and consideration of potential issues such as trust, regulation, and ethics.

Chapter 8 Conclusion and future works

8.1 Thesis conclusion

Digital technologies have greatly impacted people's lives both positively and negatively. While it has improved our daily lives, it has also brought challenges such as data security, privacy concerns, system bugs, and dependence on electricity. This research focuses on using emerging technologies in the context of medical diagnostics, with the aim to address these challenges whilst improving current diagnostic methods.

Chapters 2-4 showcase a smartphone-based diagnostic platform for malaria that offers a low-cost, user-friendly, secure, and trustworthy solution for conducting diagnostic tests in low-resource settings. The platform tackled the challenges pointed out in Chapter 1.6. By combining low-cost hardware, such as a heater and a mobile phone, with software, such as Blockchain, a mobile app, and a neural network, the proposed platform allows running sensitive and specific DNA malaria diagnostic tests in the environment with very limited resources.

In Chapter 5, AI was examined for its potential use in analysing retina vessels. Although current neural networks may not be able to accurately classify arteries and veins, they are capable of identifying blood vessels in fundus images with human-level expertise. When combined with other image processing techniques, AI's performance can improve significantly. Chapter 5 also proposes a vessel tracing algorithm, reducing processing time from days to just a few minutes.

The COVID-19 pandemic significantly impacted this PhD study. Chapter 6 outlines a research project on mitigation that was carried out by a team of group members during this period, although all work described in this thesis, including coding and creation of web-based tools, was performed by the author. The project focused on exploring the statistical models required for determining the appropriate sample size to achieve the necessary power and accuracy when validating new diagnostic tests. As part of this study, a web-based tool was developed and deployed with open access.

The purpose of diagnostics is to identify the cause of the disease, provide timely and appropriate treatment to patients, and prevent illness. Educating and engaging the public could also be essential to accomplish these goals. An example is the study conducted by Kaushik M *et al.*, which demonstrates that public awareness regarding disease prevention can significantly decrease the likelihood of contracting COVID-19 during the pandemic [184]. As part of this study, an interactive website with a map and media was created to make people aware of the risks of contracting diseases like malaria. And the objective was to increase awareness about the actions being taken to enhance health and wellness in isolated African villages with limited resources. Additionally, the aim was also to motivate more people to join in the efforts to eradicate infectious diseases such as malaria and provide safe drinking water, sanitation, and hygiene (WASH) [185] to those living in resource-limited areas. The webpage can be accessed from <https://vr-360-bme.firebaseio.com/>.

In conclusion, this PhD thesis mainly focused on utilising technologies to address the problems existing in doing diagnostic tests and providing a low-cost, easy-to-use, and robust solution to minimise the requirements, such as equipment and human expertise in running diagnostic tests in low-resource circumstances. Also, collaborations with people who have different background knowledge has helped me better understand and find the gap between biological study and software hardware engineering.

8.2 Future works

Using Blockchain technology is an effective way to safeguard sensitive medical information, such as diagnoses and patient data, through its secure features. However, there are still challenges that need to be addressed. Security threats, like cyber-attacks and ethical concerns, such as data ownership, can impact the safety and privacy of this information. Additionally, since the Blockchain is decentralised, numerous ledger copies will be stored on multiple nodes, which could require significant storage space, especially for large files like images and videos that are common in medical data. Moreover, the amount of energy consumed by using Blockchain is also a concern that needs to be addressed.

Studies have shown that AI has the potential to aid medical decisions, but there is still a long way to go. During this PhD, the ability of AI to handle complex medical images was found to be insufficient compared with human experts. However, AI, as an emerging technology, is rapidly improving.

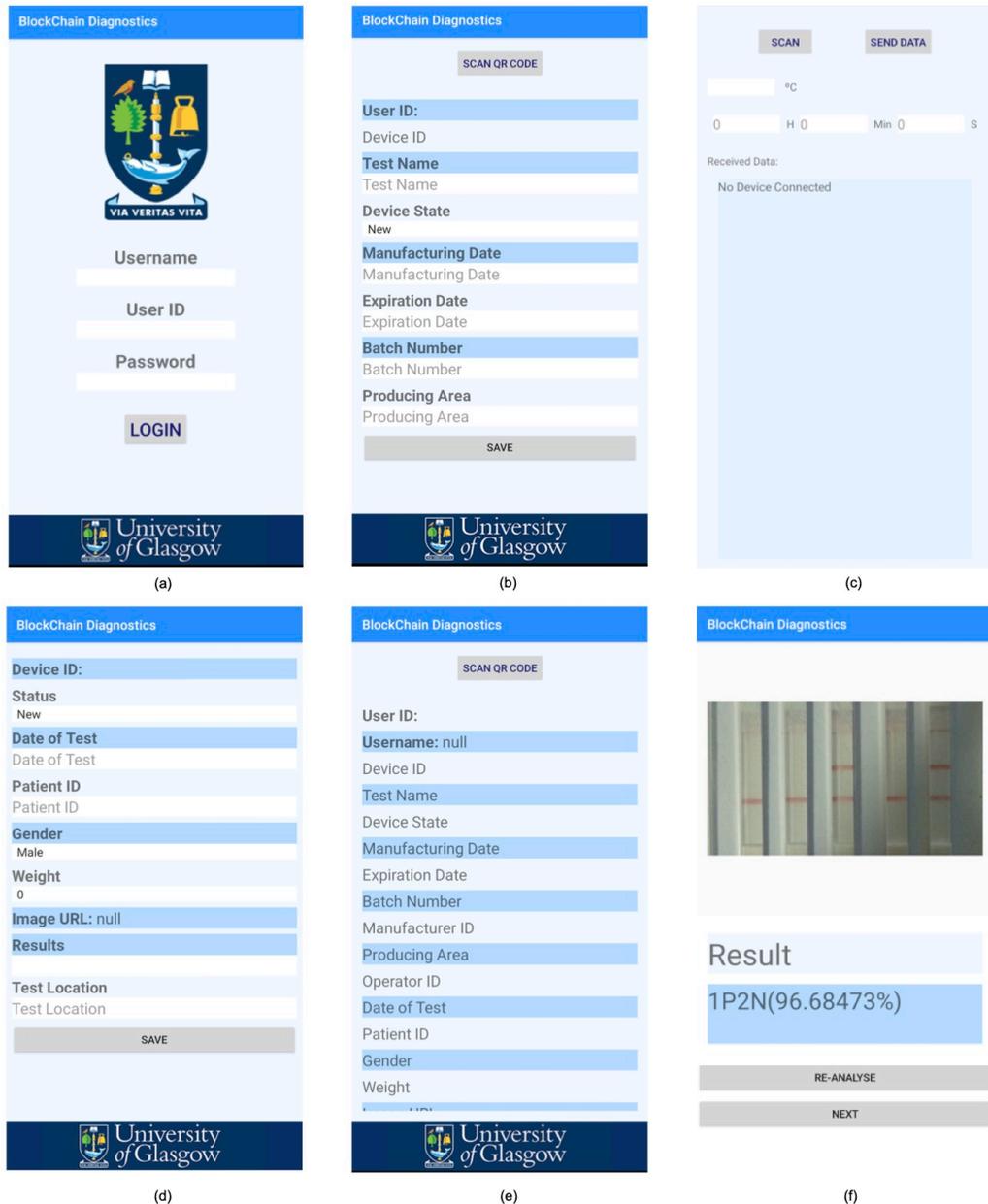
Recent studies have demonstrated the excellent performance of AI in handling image-related tasks, such as the stable-diffusion network [186] and transformer [187]. Numerous scaled neural networks with optimised architecture and algorithms have been developed in recent years [188], especially the generative models, such as ChatGPT [189] and Stable Diffusion [186] have shown impressive performance. In the future, AI's ability to handle various types of medical data should improve [190], and regulations should be established to address ethical concerns surrounding the use of AI in healthcare decision-making [191].

One of the major challenges identified in using AI as a diagnostic tool in this study is the limited data availability. While neural networks have demonstrated excellent performance, the quality and quantity of data are crucial to their success. Unfortunately, most of the medical image datasets are significantly smaller than those for general use, e.g., the ADE20K [156] dataset, which was designed for segmenting the objects in people's daily lives. This is particularly problematic for supervised learning, which requires high-quality data that has been labelled by specialists. Fortunately, the development of machine learning has led to the availability of more open-access annotated databases. For example, the EyePACS dataset [192] includes 88,702 images for detecting diabetic retinopathy, and the FIVES dataset [193] has 800 pixel-wise labelled high-resolution retina photographs.

Furthermore, as mobile devices become more powerful and neural processor units (NPUs) are integrated into more smartphones. In the coming future, it will be possible to perform many AI-powered diagnostic tests using just a smartphone or with basic external hardware. We have noticed a significant increase in commercially-available [194] products involving AI, e.g., for interpreting LFT results (such as those from Audere [195], [196], the UKHSA [197] or Abingdon Health [198] as a promising step into the future.

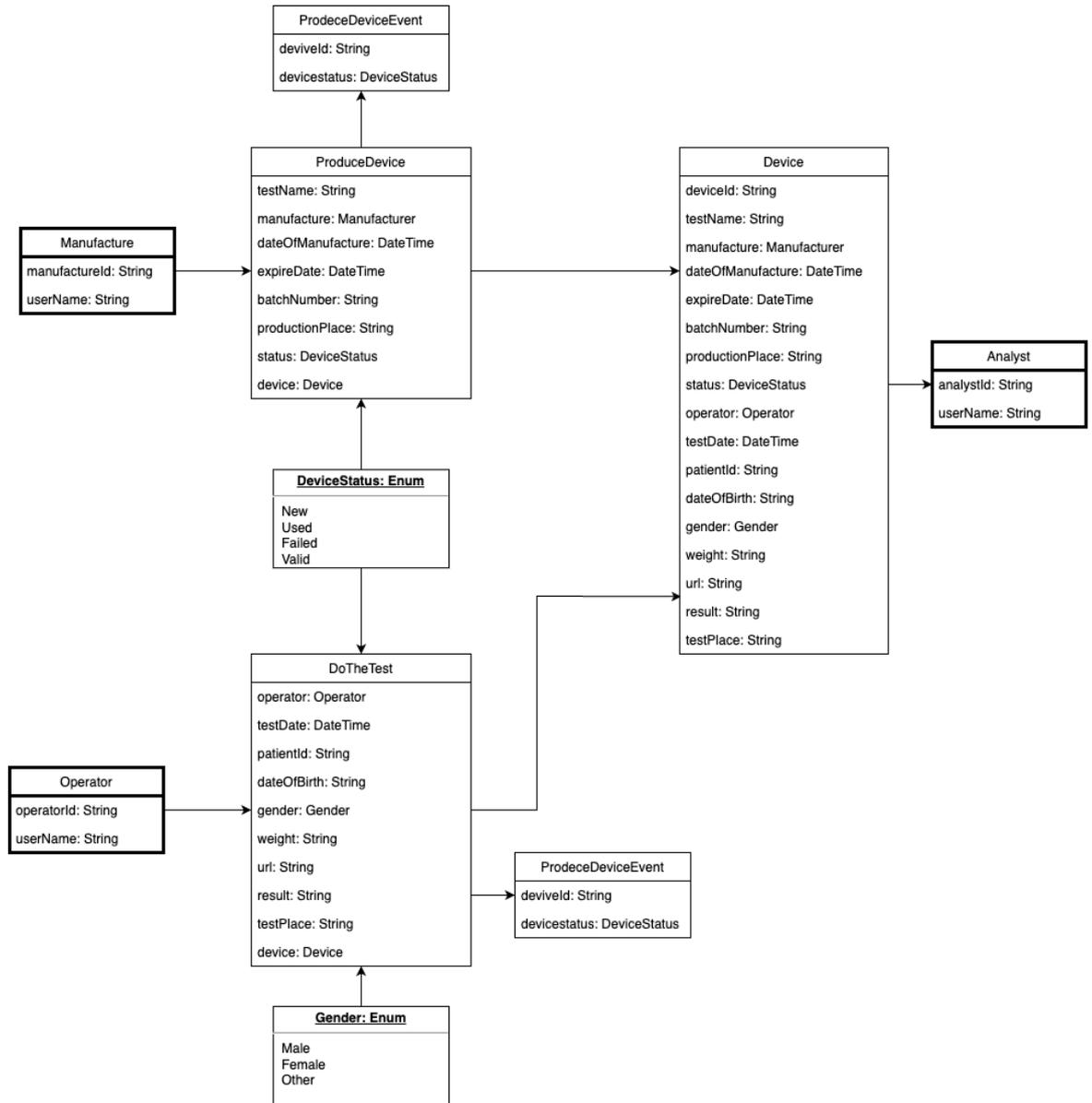
Appendix

Appendix figures



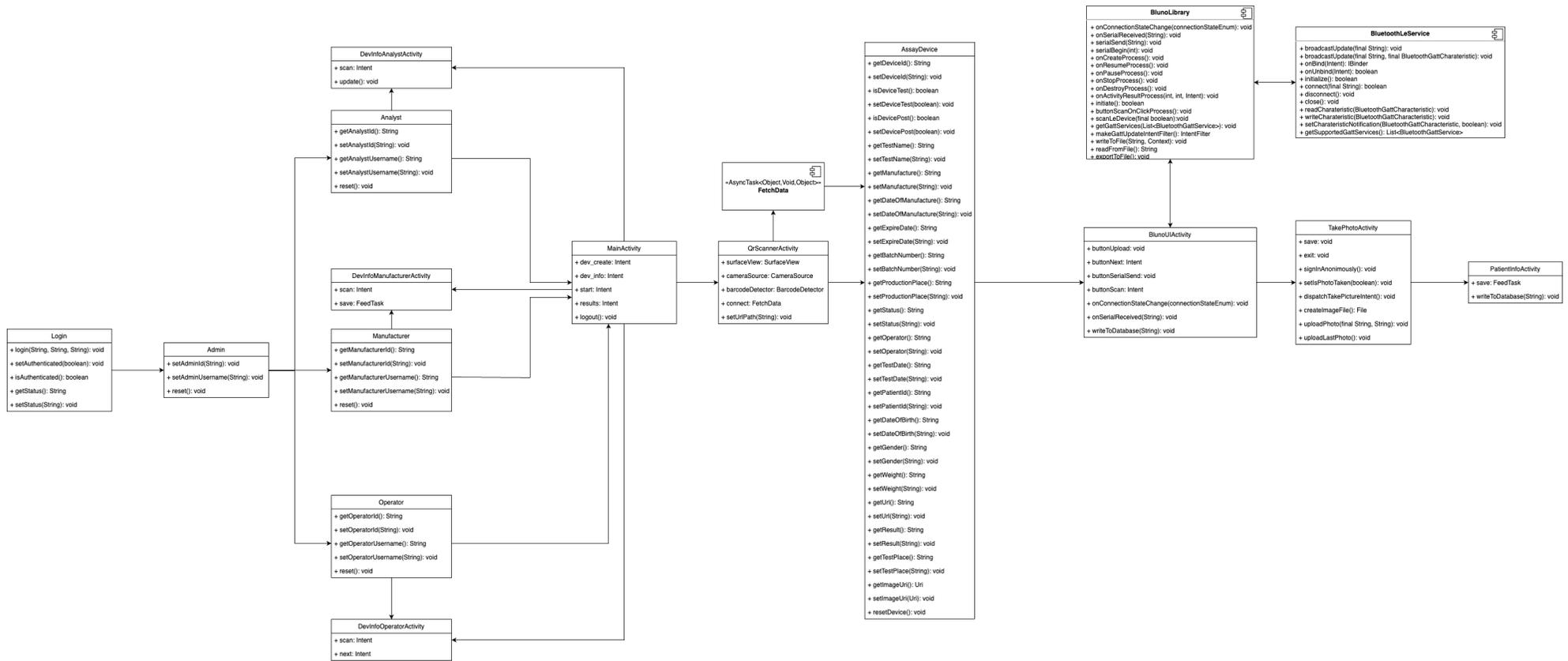
Appendix Figure 1. Screenshots of Android App (latest version). (a)The login page of the app. (b)The screen for manufacturers to add new device information to the system. (c)The heater control page for the operator. (d)The screen for operators to upload the test information. (e)Screen for analysts to read device information. (f). The screen for result analysis.

Appendix



Appendix Figure 2. Workflow of the Blockchain network. The information of manufacturer, operator, and analyst includes their ID and username, which could be recoded with other data required by transactions to the diagnostic device properties. Each block provided the information of data required and its type in different classes.

Appendix

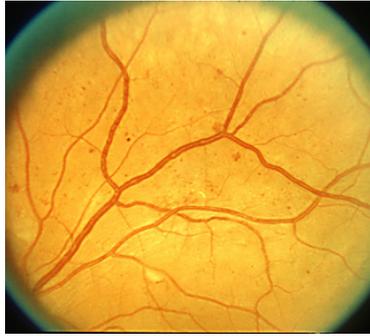


Appendix Figure 3. Workflow of the Android App.

Appendix



Standard photograph 2A, the standard for hemorrhages/microaneurysms. Eyes with severe NPDR have this degree of severity of hemorrhages and microaneurysms in all four midperipheral quadrants.



Standard photograph 6A, less severe of two standards for venous beading. Two main branches of the superior temporal vein show beading that is definite but not severe.



Standard photograph 8A, the standard for moderate IRMA. Patients with severe NPDR have moderate IRMA of at least this severity in at least one quadrant.

Reprinted with permission from the Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10. *Ophthalmology* 1991;98:786-806.

Appendix Figure 4. ETDRS standard photography 2A, 6A and 8A.[158], [160]

Appendix tables

Appendix Table 1 The estimated cost of a mobile heater

| Estimated cost of a mobile heater | |
|---|--------|
| Micro-controller | £11 |
| Power bank | £10-20 |
| Thermoelectric generator (TEG) | £14 |
| Thermocouple | £5 |
| Other electronic parts (resistors etc.) | £5 |
| 3D printing | £5 |
| Total | £50-60 |

Appendix

Appendix Table 2 System resource usage during the Blockchain Performance Test

| Node Name | Memory (avg) MB | CPU% (avg) | Traffic In (MB) | Traffic Out (MB) | Disc Read (KB) | Disc Write (MB) |
|---|-----------------------|---------------|--------------------|------------------------|----------------------|-----------------------|
| dev- peer0.org1.example.com- assay-tracking-network- 0.1.0 | 118.9 | 42.80 | 9.70 | 4.5 | 0.0 | 0.0 |
| dev- peer0.org2.example.com- assay-tracking-network- 0.1.0 | 174.5 | 9.47 | 1.80 | 1.3 | 0.0 | 0.00 |
| peer0.org2.example.com | 363.8 | 6.92 | 5.30 | 11.0 | 380.0 | 1.40 |
| peer0.org1.example.com | 318.4 | 26.0 | 21.2MB | 26.2 | 16.0 | 1.40 |
| orderer.example.com | 53.8 | 0.30 | 0.67 | 1.4 | 0.0 | 0.94 |
| couchdb.org1.example.com | 185.0 | 58.47 | 4.60 | 13.3 | 128.0 | 2.70 |
| ca.org2.example.com | 17.9 | 0.01 | 0.0003 | 0.0 | 0.0 | 0.00 |
| ca.org1.example.com | 7.1 | 0.00 | 0.0003 | 0.0 | 0.0 | 0.00 |
| couchdb.org2.example.com | 174.0 | 18.82 | 1.00 | 3.1 | 48.0 | 2.70 |

Appendix

Appendix Table 3 Performance evaluation results

| Transaction name | Succeed | Failed | Send Rate (TPS) | Max Latency (s) | Min Latency (s) | Average Latency (s) | Throughput (TPS) |
|------------------|---------|--------|-----------------|-----------------|-----------------|---------------------|------------------|
| ProduceDevice | 50 | 0 | 5.1 | 0.60 | 0.19 | 0.41 | 5.0 |
| | 100 | 0 | 10.1 | 3.04 | 0.42 | 1.71 | 9.2 |
| | 200 | 0 | 20.1 | 14.30 | 0.77 | 10.97 | 10.2 |
| | 300 | 0 | 30.1 | 23.10 | 6.02 | 22.12 | 10.3 |
| | 400 | 0 | 40.1 | 33.22 | 4.40 | 28.57 | 10.3 |
| | 500 | 0 | 50.1 | 46.43 | 9.42 | 36.99 | 10.3 |
| DoTheTest | 50 | 0 | 5.1 | 0.63 | 0.22 | 0.42 | 4.8 |
| | 100 | 0 | 10.1 | 4.36 | 0.27 | 2.52 | 8.7 |
| | 200 | 0 | 20.1 | 13.60 | 0.97 | 10.87 | 10.2 |
| | 300 | 0 | 30.1 | 23.73 | 1.17 | 18.44 | 10.6 |
| | 400 | 0 | 40.1 | 33.30 | 3.36 | 26.70 | 10.6 |
| | 500 | 0 | 50.1 | 43.02 | 17.05 | 36.02 | 10.3 |

Appendix

Appendix Table 4 Species of Plasmodium that cause malaria [199]

| | | P falciparum | P vivax | P malariae | P ovale |
|--|--------------|-------------------------|---|-------------------------|---------------------------------|
| Distribution | | Global tropics | Global tropics and some temperate zones | Global tropics | Africa, southeast Asia, Oceania |
| Disease burden (number of people infected) | | 400 million | 80-400 million | Minor | Probably minor |
| Number of people at risk | | 5 billion | 3 billion | Unknown | Unknown |
| Type of malaria | | | | | |
| Asexual blood | Asymptomatic | Many endemic settings | Many endemic settings | Almost always | Unknown |
| | Therapies | Many | Many | Many | Many |
| Sexual blood | Asymptomatic | Always | Always | Always | Always |
| | Therapies | Primaquine, single dose | Primaquine, hypnozoitocide | Primaquine, single dose | Primaquine, hypnozoitocide |
| Asexual liver | Asymptomatic | Always | Always | Always | Always |
| | Therapies | Not indicated | Primaquine, 14 daily doses | Not indicated | Primaquine, 14 daily doses |

Appendix

Appendix Table 5 $Z_{\frac{\alpha}{2}}$ value and Significance level corresponding to confidence level

| Confidence level | $Z_{\frac{\alpha}{2}}$ | Significance level |
|------------------|------------------------|--------------------|
| 80% | 1.28 | 20% |
| 85% | 1.44 | 15% |
| 90% | 1.645 | 10% |
| 95% | 1.960 | 5% |
| 98% | 2.326 | 2% |
| 99% | 2.576 | 1% |

Appendix Table 6 $Z_{1-\beta}$ value under different power

| Power | $Z_{1-\beta}$ |
|-------|---------------|
| 0.70 | 0.52 |
| 0.75 | 0.67 |
| 0.80 | 0.84 |
| 0.85 | 1.03 |
| 0.90 | 1.28 |
| 0.95 | 1.64 |
| 0.99 | 2.33 |

Appendix

Appendix Table 7 Normal distribution – inverse cumulative distribution function

| Z | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|------|--------|--------|--------|--------|--------|
| 0.00 | 0.0000 | 0.2533 | 0.5244 | 0.8416 | 1.2816 |
| 0.01 | 0.0252 | 0.2793 | 0.5534 | 0.8779 | 1.3408 |
| 0.02 | 0.0502 | 0.3055 | 0.5828 | 0.9154 | 1.4051 |
| 0.03 | 0.0753 | 0.3319 | 0.6128 | 0.9542 | 1.4758 |
| 0.04 | 0.1004 | 0.3585 | 0.6433 | 0.9945 | 1.5548 |
| 0.05 | 0.1257 | 0.3853 | 0.6745 | 1.0364 | 1.6449 |
| 0.06 | 0.1510 | 0.4125 | 0.7063 | 1.0803 | 1.7507 |
| 0.07 | 0.1764 | 0.4399 | 0.7388 | 1.1264 | 1.8808 |
| 0.08 | 0.2019 | 0.4677 | 0.7722 | 1.1750 | 1.9600 |
| 0.09 | 0.2275 | 0.4958 | 0.8064 | 1.2265 | 2.0537 |

Other appendices

Appendix 1 Standard Feature Based Grading Form to be used in Routine Digital Screening [159]

| R | NHS Diabetic Eye Screening Programme Feature Based Grading Classification | | |
|----|---|---|---|
| R0 | None | No DR | |
| R1 | Background | 1 microaneurysm * | |
| | | 2 microaneurysms | |
| | | 3 microaneurysms | |
| | | 4 microaneurysms | |
| | | 5+ microaneurysms | |
| | | or | |
| | | microaneurysm(s) | |
| | | retinal haemorrhage(s) | |
| | | venous loop | |
| | | any exudate in the presence of other features of DR | |
| | | any number of cotton wool spots (CWS) in the presence of other features of DR | |
| R2 | Pre-proliferative | venous beading | |
| | | venous reduplication | |
| | | multiple blot haemorrhages | |
| | | intraretinal microvascular abnormality (IRMA) | |
| R3 | Proliferative | Stable pre-retinal fibrosis + peripheral retinal scatter laser | R3s (Stable post treatment) |
| | | Stable fibrous proliferation (disc or elsewhere) + peripheral retinal scatter laser | (If discharged from the Hospital Eye Service a photograph should be taken at or shortly after discharge from the Hospital Eye |
| | | Stable R2 features (from feature based grading) + peripheral retinal scatter laser | discharge from the Hospital Eye |

Appendix

| | | | |
|----|--|---|---|
| | | R1 features (from feature based grading) + peripheral retinal scatter laser | service (HES) that records these features) |
| | | | |
| | | new vessels on disc (NVD) | R3a (Active Proliferative Retinopathy) |
| | | new vessels elsewhere (NVE) | |
| | | New pre-retinal or vitreous haemorrhage | |
| | | New pre-retinal fibrosis | |
| | | New tractional retinal detachment | |
| | | Reactivation in a previous stable R3s eye | |
| | | | |
| M | | Maculopathy | |
| M0 | | No maculopathy | absence of any M1 features |
| M0 | | any microaneurysm or haemorrhage within 1DD of the centre of the fovea if associated with a best VA of $\square\square/12$ | |
| | | is known and is not diabetic macular oedema. | |
| | | | |
| M1 | | exudate within 1 disc diameter (DD) of the centre of the fovea | |
| | | group of exudates within the macula | A group of exudates is an area of exudates that is greater than or equal to half the disc area and this area (of greater than or equal half the disc area) is all within the macular area |
| | | retinal thickening within 1DD of the centre of the fovea (if stereo available) | |
| | | any microaneurysm or haemorrhage within 1DD of the centre of the fovea only if associated with a best VA of $\square/12$ (if no stereo) | |
| | | | |
| P | | Photocoagulation | |

Appendix

| | | | |
|---|--------------------------|---|--|
| P0 | | No evidence of previous photocoagulation (default) | |
| P1 | | focal/grid to macula or peripheral scatter | |
| | | | |
| U | | Ungradable | |
| | | An image set that cannot be graded | |
| | | | |
| | Other Lesions (optional) | See later section 'The Grading Pathway Non DR Features' | Lesions resulting in referral to ROG or finalisation of grade are programme definable. |
| | | | |
| * dot haemorrhages should be included in the count as it is often difficult to tell the difference between a microaneurysm and a dot haemorrhage. | | | |
| Outcomes | | | |
| | | Outcome choice | How outcome is decided |
| | | Routine Digital Screening Annual Recall | See DESP Grading Criteria and Actions Version 1.0 07 March 2012 |
| | | Refer to Digital Surveillance | Primary / secondary / arbitration / referral outcome / grader |
| | | Refer to SLB Surveillance | Primary / secondary / arbitration / referral outcome / grader |
| | | Refer DR to Ophthalmology urgently | See DESP Grading Criteria and Actions Version 1.0 07 March 2012 |
| | | Refer DR to Ophthalmology routinely | See DESP Grading Criteria and Actions Version 1.0 07 March 2012 |
| | | Exclude | Referral Outcome grader can select but admin staff have to approve |
| | | Off register? | Admin staff only |
| | | Refer non DR to Ophthalmology/GP urgently | Referral Outcome Grader, refer according to local protocol |
| | | Refer non DR to Ophthalmology/GP routinely | Referral Outcome Grader, refer according to local protocol |

*Appendix 2 Feature Based Grading Forms to be used in Surveillance [159]***Background**

As part of a series of pathway modifications the NHS Diabetic Eye Screening Programme (DESP) wishes to introduce Surveillance clinics in all programmes which will encompass fundal visualisation by digital camera, slit lamp biomicroscopy (SLB). These clinics will be implemented and recorded as a separate stream of activity to normal screening but the call and recall will be within screening management software.

This operational guidance is intended to clarify the additional minimum requirements for recording grading and associated workflows in both Digital Surveillance and SLB Surveillance clinics.

The grading pathway used in surveillance should be capable of being configured flexibly so that grading can be used in a full arbitration type grading, by technician followed by ophthalmologist in a hierarchical grade or finalised by a single user with the appropriate access type. This will enable programmes to use technician, optometric or ophthalmologist graders in ways which they consider both efficient and appropriate.

The form allows recording of outcomes from OCT. However, OCT is not part of the national screening programme pathway.

| R | NHS Diabetic Eye Screening Programme Feature Based Grading Classification | | |
|----|---|------------------------|--|
| R0 | None | No DR | |
| R1 | Background | 1 microaneurysm * | |
| | | 2 microaneurysms | |
| | | 3 microaneurysms | |
| | | 4 microaneurysms | |
| | | 5+ microaneurysms | |
| | | or | |
| | | microaneurysm(s) | |
| | | retinal haemorrhage(s) | |
| | | venous loop | |

Appendix

| | | | |
|----|-------------------|---|--|
| | | any exudate in the presence of other features of DR | |
| | | any number of cotton wool spots (CWS) in the presence of other features of DR | |
| | | | |
| R2 | Pre-proliferative | venous beading | |
| | | venous reduplication | |
| | | multiple blot haemorrhages | |
| | | intraretinal microvascular abnormality (IRMA) | |
| | | | |
| R3 | Proliferative | Stable pre-retinal fibrosis + | R3s (Stable post treatment) |
| | | Stable fibrous proliferation (disc or elsewhere) + peripheral retinal scatter laser | (If discharged from the Hospital Eye Service a photograph should |
| | | Stable R2 features (from feature based grading) + peripheral retinal scatter laser | be taken at or shortly after |
| | | R1 features (from feature based grading) + peripheral retinal scatter laser | discharge from the Hospital Eye service (HES) that records these features) |
| | | | |
| | | new vessels on disc (NVD) | R3a (Active Proliferative Retinopathy) |
| | | new vessels elsewhere (NVE) | |
| | | New pre-retinal or vitreous haemorrhage | |
| | | New pre-retinal fibrosis | |
| | | New tractional retinal detachment | |
| | | Reactivation in a previous stable R3s eye | |
| | | | |
| M | | Maculopathy | |
| M0 | | No maculopathy | absence of any M1 features |
| M0 | | any microaneurysm or haemorrhage within 1DD of the centre of the | |
| | | is known and is not diabetic macular oedema. | |
| | | | |
| | | exudate within 1 disc diameter (DD) | |

Appendix

| | | | |
|---|--------------------------|---|--|
| | | group of exudates within the macula | A group of exudates is an area of exudates that is greater than or equal to half the disc area and this area (of greater than or equal half the disc area) is all within the |
| | | retinal thickening within 1DD of the centre of the fovea (if stereo available) | |
| | | any microaneurysm or haemorrhage within 1DD of the centre of the fovea only if associated with a best VA of \square 6/12 (if no stereo) | |
| | | | |
| P | | <u>Photocoagulation</u> | |
| P0 | | No evidence of previous | |
| P1 | | focal/grid to macula or peripheral | |
| | | | |
| U | | <u>Ungradable</u> | |
| | | An image set that cannot be graded | |
| | | | |
| | Other Lesions (optional) | See later section 'The Grading Pathway Non DR Features' | Lesions resulting in referral to ROG or finalisation of grade are programme definable. |
| | | | |
| * dot haemorrhages should be included in the count as it is often difficult to tell the difference between a microaneurysm and a dot haemorrhage. | | | |
| Outcomes | | | |
| | | Outcome choice | How outcome is decided |
| | | Routine Digital Screening Annual Recall | See DESP Grading Criteria and Actions Version 1.0 07 March |
| | | Refer to Digital Surveillance | Primary / secondary / arbitration / referral outcome / grader |
| | | Refer to SLB Surveillance | Primary / secondary / arbitration / referral outcome / grader |
| | | Refer DR to Ophthalmology urgently | See DESP Grading Criteria and Actions Version 1.0 07 March |
| | | Refer DR to Ophthalmology routinely | See DESP Grading Criteria and Actions Version 1.0 07 March |
| | | Exclude | Referral Outcome grader can select but admin staff have to |
| | | Off register? | Admin staff only |
| | | Refer non DR to Ophthalmology/GP urgently | Referral Outcome Grader, refer according to local protocol |
| | | Refer non DR to Ophthalmology/GP routinely | Referral Outcome Grader, refer according to local protocol |

Appendix 3 Standard Feature Based Grading Form to be used in Digital Surveillance

[159]

| | | | Right | | Left | |
|---|-----------------|---|------------------|------|-----------------|------|
| Visual Acuity (best of unaided, best corrected with spectacles and with pinhole) | | | | | | |
| | | | Screener Grading | | Ophthalmologist | |
| | | | Right | Left | Right | Left |
| OCT* | | | | | | |
| Adequate | | | | | | |
| Inadequate | | | | | | |
| OCT positive | | | | | | |
| OCT borderline | | | | | | |
| OCT negative | | | | | | |
| Photographs | | | | | | |
| Adequate | | | | | | |
| Inadequate | | | | | | |
| R0 | None | No DR | | | | |
| | | | | | | |
| R1 | Background | microaneurysm(s) | | | | |
| | | retinal haemorrhage(s) | | | | |
| | | venous loop | | | | |
| | | any exudate in the presence of other features of DR | | | | |
| | | any number of cotton wool spots (CWS) in the presence of other features of DR | | | | |
| | | | | | | |
| R2 | Pre-proliferati | venous beading | | | | |

Appendix

| | | | | | | |
|--|--|----------------------------|--|--|--|--|
| | | venous reduplication | | | | |
| | | multiple blot haemorrhages | | | | |

| | | | | | | |
|----|---------------|--|--|--|--|--|
| | | intraretinal microvascular abnormality | | | | |
| R3 | Proliferative | | | | | |
| | R3s | Stable pre-retinal fibrosis + peripheral | | | | |
| | | Stable fibrous proliferation (disc or elsewhere) + peripheral retinal | | | | |
| | | Stable R2 features (from feature based grading) + peripheral retinal scatter | | | | |
| | | R1 features (from feature based grading) + peripheral retinal | | | | |
| | R3a | new vessels on disc (NVD) | | | | |
| | | new vessels elsewhere (NVE) | | | | |
| | | New pre-retinal or vitreous | | | | |
| | | New pre-retinal fibrosis | | | | |
| | | New tractional retinal detachment | | | | |
| | | Reactivation in a previous stable R3s eye | | | | |
| M | Maculopathy | | | | | |
| | M0 | No maculopathy | | | | |
| | | any microaneurysm or haemorrhage within 1DD of the centre of the fovea if associated with a best VA of $\leq 6/12$ where the cause of the reduced vision is known and is not diabetic macular oedema | | | | |
| | M1 | exudate within 1 disc diameter (DD) of the centre | | | | |
| | | group of exudates within the macula | | | | |

Appendix

| | | | | | | |
|--|--|--|--|--|--|--|
| | | retinal thickening within 1DD of the centre of the fovea (if stereo available) | | | | |
|--|--|--|--|--|--|--|

| | | | | | | |
|---|--------------------------|--|--|--|--|--|
| | | any microaneurysm or haemorrhage within 1DD of the centre of the fovea only if associated with a best VA of $\square\square/12$ (if no stereo) | | | | |
| | | | | | | |
| P | Photocoagulation | | | | | |
| | P0 | No evidence of previous | | | | |
| | P1 | focal/grid to macula or peripheral scatter | | | | |
| | | | | | | |
| U | Ungradable | | | | | |
| | U | An image set that cannot be graded | | | | |
| | | | | | | |
| | Other Lesions (optional) | See later section 'The Grading Pathway Non DR Features' | Lesions resulting in referral to ROG or finalisation of grade are programme definable. | | | |
| | | | | | | |

Outcomes

NOTE: outcomes are chosen by examiner and are not defined by business rules

| | | | | | | |
|--|--|---|--|--|--|--|
| | | Refer DR to Ophthalmology urgently | | | | |
| | | Refer DR to Ophthalmology routinely | | | | |
| | | Refer to Slit Lamp biomicroscopy clinic routinely | | | | |
| | | Review in Digital Surveillance clinic in 3/12 | | | | |
| | | Review in Digital Surveillance clinic in 6/12 | | | | |
| | | Review in Digital Surveillance clinic in 9/12 | | | | |

Appendix

| | | | | | | |
|--|--|--|--|--|--|--|
| | | Review in Digital Surveillance clinic in 12/12 | | | | |
| | | Discharge back to screening | | | | |
| | | Exclude | | | | |
| | | Refer non DR to Ophthalmology/GP urgently | | | | |
| | | Refer non DR to Ophthalmology/GP routinely | | | | |

*OCT definitions

Appendix

Appendix 4 Standard Feature Based Grading Form to be used in SLB Surveillance[159]

| | | | Right | Left |
|---|-------------------|--|-------|------|
| Visual Acuity (best of unaided, best corrected with spectacles and with pinhole) | | | | |
| Slit lamp biomicroscopy | | | | |
| View # | | Retinal view adequate | | |
| | | Retinal view inadequate | | |
| R0 | None | No DR | | |
| | | | | |
| R1 | Background | microaneurysm(s) | | |
| | | retinal haemorrhage(s) | | |
| | | venous loop | | |
| | | any exudate in the presence of other features of DR | | |
| | | any number of cotton wool spots (CWS) in the presence of other features of DR | | |
| | | | | |
| R2 | Pre-proliferative | venous beading | | |
| | | venous reduplication | | |
| | | multiple blot haemorrhages | | |
| | | intraretinal microvascular abnormality (IRMA) | | |
| | | | | |
| R3 | Proliferative | | | |
| | R3s | Stable pre-retinal fibrosis + peripheral retinal scatter laser | | |
| | | Stable fibrous proliferation (disc or elsewhere) + peripheral retinal scatter | | |
| | | Stable R2 features (from feature based grading) + peripheral retinal scatter laser | | |
| | | R1 features (from feature based grading) + peripheral retinal scatter | | |
| | | | | |
| | R3a | new vessels on disc (NVD) | | |

Appendix

| | | | | |
|---|------------------|---|--|--|
| | | new vessels elsewhere (NVE) | | |
| | | New pre-retinal or vitreous haemorrhage | | |
| | | New pre-retinal fibrosis | | |
| | | New tractional retinal detachment | | |
| | | Reactivation in a previous stable R3s eye | | |
| | | | | |
| M | Maculopathy | | | |
| | M0 | No maculopathy | | |
| | | any microaneurysm or haemorrhage within 1DD of the centre of the fovea if associated with a best VA of $\square \square 6/12$ where the cause of the reduced vision is known and is not diabetic macular oedema | | |
| | | | | |
| | M1 | exudate within 1 disc diameter (DD) of the centre of the fovea | | |
| | | group of exudates within the macula | | |
| | | retinal thickening within 1DD of the centre of the fovea (if stereo available) | | |
| | | any microaneurysm or haemorrhage within 1DD of the centre of the fovea only if associated with a best VA of $\square 6/12$ (if no stereo) | | |
| | | CSMO - Retinal thickening at or within 500 microns of the centre of the | | |
| | | CSMO - Hard exudates at or within 500 microns of the centre of the macula, if associated with thickening of the adjacent retina (not residual hard exudates remaining after disappearance of retinal thickening)hard exudates remaining | | |
| | | CSMO - A zone or zones of retinal thickening one disc area or larger, any part of which is within one disc diameter of the centre of the macula. | | |
| | | | | |
| P | Photocoagulation | | | |
| | P0 | No evidence of previous photocoagulation (default) | | |

Appendix

| | | | | |
|---|--------------------------|--|--|--|
| | P1 | focal/grid to macula or peripheral scatter | | |
| | | | | |
| U | Ungradable | | | |
| | U | Retinal status cannot be determined by slit lamp biomicroscopy | | |
| | Other Lesions (optional) | See later section 'The Grading Pathway Non DR Features' | Lesions resulting in referral to ROG or finalisation of grade are programme definable. | |
| | | | | |
| Outcomes | | | | |
| NOTE: outcomes are chosen by examiner and are not defined by business rules | | | | |
| | | Refer DR to Ophthalmology urgently | | |
| | | Refer DR to Ophthalmology routinely | | |
| | | Refer to Digital Surveillance clinic | | |
| | | List for cataract surgery (if conducted within HES) | | |
| | | Review in slit lamp biomicroscopy in 6/12 | | |
| | | Review in slit lamp biomicroscopy in 12/12 | | |
| | | Discharge back to screening | | |
| | | Exclude | | |
| | | Refer non DR to Ophthalmology/GP urgently | | |
| | | Refer non DR to Ophthalmology/GP routinely | | |

SLB Examination Quality

Adequate: Practitioner is confident that referable pathology can be assessed or there is referable DR present

Inadequate: Practitioner cannot visualise majority of fundus and no referable DR is visible.

The Grading Pathway Non DR Features

The Diabetic Eye Screening Programme has the primary function of detecting retinal disease secondary to diabetes.

Other non DR features may be observed during the screening process which may result in the need for a referral or which may need to be recorded:

to refer in other sight threatening conditions

to have an audit trail back to see if confounders for DR (drusen / AMD, asteroid) were detected

It is the responsibility of the Clinical Lead of the screening programme to decide what 'other lesions' are recorded within the diabetic retinopathy screening programme, train the graders to detect these lesions and, if any referral is made from the screening programme for other eye lesions, to decide the referral pathway.

Programmes should determine their own local policy for recording the features and handling the outcomes of Non DR related pathologies. This includes the number of grading stages where only non DR is observed.

The lesions may either be recorded and require no action or, if they require a second opinion or referral, the image set should be sent to the ROG.

The ROG will determine what outcome is indicated. The status of the patient from that point is described in the current exclusions and suspensions guidance.

Local policy will determine whether the referral is directly from the service or routed through the GP.

Appendix

Appendix 5 The different grades of diabetic retinopathy (DR) in the Scottish Grading Protocol: features and outcomes[160]

| Retinopathy | Description | Outcome |
|-------------------|---|--|
| R0 | No diabetic retinopathy anywhere | Rescreen 12 months |
| R1(mild) | Background diabetic retinopathy BDR – mild <ul style="list-style-type: none"> ▪ At least one dot haemorrhage or microaneurysm with or without hard exudates | Rescreen 12 months |
| R2(moderate) | BDR – moderate <ul style="list-style-type: none"> ▪ Four or more blot haemorrhages (\geqAH standard photograph 2a) in one hemi-field only (Inferior and superior hemi-fields delineated by a line passing through the centre of the fovea and optic disc) | Rescreen 6 months(or refer to ophthalmology if this is not feasible) |
| R3(severe) | BDR – severe Any of the following features: <ul style="list-style-type: none"> ▪ Four or more blot haemorrhages (\geqAH standard photograph 2a) in both inferior and superior hemi-fields ▪ Venous beading (\geqAH standard photograph 6a) ▪ IRMA (\geq AH standard photograph 8a) | Refer ophthalmology |
| R4(proliferative) | Proliferative diabetic retinopathy PDR Any of the following features: <ul style="list-style-type: none"> ▪ New vessels ▪ Vitreous haemorrhage | Refer ophthalmology |
| R5(enucleated) | Enucleated eye | Rescreen 12 months (other eye) |
| R6 (inadequate) | Not adequately visualised <ul style="list-style-type: none"> ▪ Retina not sufficiently visible for assessment | Technical failure. Arrange alternative screening examination |

Appendix 6 Documentation of the medial_axis function from Skimage

The `medial_axis` function defined in the scikit-learn package. The documentation and example of this package was provided below:

```
medial_axis(image, mask=None, return_distance=False, *, random_state=None):
```

```
    """Compute the medial axis transform of a binary image.
```

Parameters

```
-----
```

```
image : binary ndarray, shape (M, N)
```

```
    The image of the shape to be skeletonized.
```

```
mask : binary ndarray, shape (M, N), optional
```

```
    If a mask is given, only those elements in `image` with a true
    value in `mask` are used for computing the medial axis.
```

```
return_distance : bool, optional
```

```
    If true, the distance transform is returned as well as the skeleton.
```

```
random_state : {None, int, `numpy.random.Generator`}, optional
```

```
    If `random_state` is None the `numpy.random.Generator` singleton is
    used.
```

```
    If `random_state` is an int, a new ``Generator`` instance is used,
    seeded with `random_state`.
```

```
    If `random_state` is already a ``Generator`` instance then that
    instance is used.
```

```
.. versionadded:: 0.19
```

Returns

```
-----
```

```
out : ndarray of bools
```

```
    Medial axis transform of the image
```

```
dist : ndarray of ints, optional
```

```
    Distance transform of the image (only returned if `return_distance`
```

is True)

See Also

skeletonize

Notes

This algorithm computes the medial axis transform of an image as the ridges of its distance transform.

The different steps of the algorithm are as follows

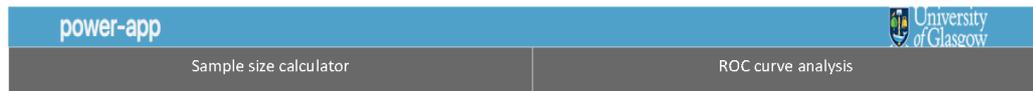
- * A lookup table is used, that assigns 0 or 1 to each configuration of the 3x3 binary square, whether the central pixel should be removed or kept. We want a point to be removed if it has more than one neighbor and if removing it does not change the number of connected components.
- * The distance transform to the background is computed, as well as the corneriness of the pixel.
- * The foreground (value of 1) points are ordered by the distance transform, then the corneriness.
- * A cython function is called to reduce the image to its skeleton. It processes pixels in the order determined at the previous step, and removes or maintains a pixel according to the lookup table. Because of the ordering, it is possible to process all pixels in only one pass.

Examples

```
>>> square = np.zeros((7, 7), dtype=np.uint8)
>>> square[1:-1, 2:-2] = 1
>>> square
```

```
array([[0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 1, 1, 1, 0, 0, 0],
       [0, 0, 1, 1, 1, 0, 0, 0],
       [0, 0, 1, 1, 1, 0, 0, 0],
       [0, 0, 1, 1, 1, 0, 0, 0],
       [0, 0, 1, 1, 1, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0]], dtype=uint8)
>>> medial_axis(square).astype(np.uint8)
array([[0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 1, 0, 1, 0, 0, 0],
       [0, 0, 0, 1, 0, 0, 0, 0],
       [0, 0, 0, 1, 0, 0, 0, 0],
       [0, 0, 0, 1, 0, 0, 0, 0],
       [0, 0, 1, 0, 1, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0]], dtype=uint8)
```

Appendix 7 Storyboard



If “Sample size calculator” is selected:

Sample Size Calculator

This sample size calculator identifies the **required sample size** for **evaluating the accuracy of a diagnostic test**. Diagnostic accuracy can be expressed as sensitivity, specificity, likelihood ratio (LR) or area under the ROC curve (AUC), and it assesses the ability of a diagnostic test to detect the target condition of interest.

According to your study purpose, you can choose two ways to estimate the proper sample size needed. You can either be interested in estimating the accuracy of a diagnostic test, or testing a hypothesis to compare the accuracy of a diagnostic test with another test, or with a specific value. The goal of this calculator is to enable you to calculate the required sample size base on any of these parameters. You can select ‘**Estimating accuracy of a single test**’ if you want to [meet prespecified accuracy goals](#), e.g., you want to recruit a certain number of subjects to achieve an expected level of sensitivity. You can select ‘**Compare two values of accuracy**’ if you either want to [compare the accuracy of two diagnostic tests](#) or [compare your test with a minimally acceptable prespecified accuracy](#).

Choose the purpose of your study:



If “Compare two values of accuracy” is selected:

Test accuracy of a single test or compare accuracies of two tests

Testing the accuracy of a single test or comparing two different tests accuracies implies testing a hypothesis. A **comparison of the accuracies of two tests** refers to [test the equivalence of the accuracies of two tests](#). In both scenarios, the hope is to reject the hypothesis of equivalence and thus prove that the accuracies are different. **Testing the accuracy of a single test** means [studying the equivalence of your test to a predetermined level of accuracy](#).

Here is a calculator that helps you in testing your hypothesis with a desired precision, *i.e.* confidence level, and power.



- If “Test accuracy of a single test” is selected:

You can determine the number of samples/subjects needed in order to ensure that your diagnostic test meets an acceptable pre-specified performance and that the study has high probability (power) to detect that your test is better than the minimal fixed accuracy [1].

Appendix

The null hypothesis states that your test accuracy p_1 is equivalent to the pre-specified one p_0 , *i.e.* the phenomenon under test is absent.
 $\alpha\%$ is the probability of falsely rejecting the null hypothesis and it gives a basis to prove the existence of the phenomenon to be demonstrated [2].
 The power $(1-\beta)\%$ is the probability that the study will detect a difference in accuracies, given that the phenomenon actually exists [3].
 This calculator uses a **two-tailed z-score test under the null and alternative hypothesis (conditional approach)** and normal approximation is the general rule [4]. One-sided tests are often possible but rarely used in clinical research.

Please enter either the sensitivity or specificity values.

Sensitivity/Specificity of reference, p_0 : % i Pre-specified accuracy value, i.e. reference test.

Sensitivity/Specificity of the new test, p_1 : % i

Significance level, α : % i 5% significance level is considered as the standard in the literature. Higher significance levels lead to an increase in power. However, as the significance level increases, the probability of making an error in claiming a difference when in fact none exists becomes higher [3].

Power, $(1 - \beta)$: % i The minimum power value is usually set as 80% [7].

Failure rate in sampling, q : % i The sample processing, stretching from patient identification, to sample collection and result, might not work as planned due to technical or procedural problems (e.g. contamination, failure to get the test performed in time, subjects withdrawal, missing data). It is necessary to account for these issues when determining the sample size and an additional 10-20% subjects are recruited.

Compute

Sample size using a conditional approach, N :

Sample size using an unconditional approach, N :

If "Estimate Accuracy" is selected:

Estimate accuracy of a diagnostic test

Here you can determine the number of samples/subjects you need to recruit in order to define an expected level of accuracy with a desired precision, *i.e.* confidence level, of your diagnostic test.

Please select the accuracy of interest

Sensitivity
 Specificity
 Sensitivity & Specificity

Likelihood ratio (LR)
 Area under the ROC curve (AUC)

- For Sensitivity/Specificity/Sensitivity & Specificity see storyboards of phase 1 (no changes)

Appendix

- If Area Under the ROC curve is selected:

You can determine the number of samples/subjects needed in order to estimate the area under the ROC curve of a diagnostic test. This takes into account the trade-off between the sensitivity and the specificity of the test. AUC can be estimated using parametric (binormal) and nonparametric (Wilcoxon statistic) approaches. Here we use the **binormal approach**.

Area under the curve, AUC:

Allocation ratio, $R = N_D/N_{ND}$:



The allocation ratio represents the sample size of subject with (N_D) the condition of interest over the ones without (N_{ND}). $R = N_D/N_{ND}$

Margin of error, M :

Confidence level, $(1 - \alpha)$:

Failure rate in sampling, q :

%



The sample processing, stretching from patient identification, to sample collection and result, might not work as planned due to technical or procedural problems (e.g. contamination, failure to get the test performed in time, subjects withdrawal, missing data). It is necessary to account for these issues when determining the sample size and an additional 10-20% subjects are recruited.

Compute

The sample sizes for each group of diseased and nondiseased people is:

Sample size for the group with the condition, N_D :

Sample size for the group without the condition, N_{ND} :

Sample size, N :

Formulas used:

$$N_D' = \frac{Z_{\alpha/2}^2 \times V(AUC)}{M^2} \quad \text{and} \quad N_{ND}' = R \times N_D'$$

$$N_D = \frac{N_D'}{(1-q)} \quad \text{and} \quad N_{ND} = \frac{N_{ND}'}{(1-q)}$$

$$N = N_D + N_{ND}$$

$$V(AUC) = N_D \text{Var}(\widehat{AUC}) = (0.0099 \times e^{-a^2/2}) \times \left((5a^2 + 8) + \frac{a^2 + 8}{R} \right)$$

Where $\text{Var}(\widehat{AUC})$ denotes the variance of \widehat{AUC} , $a = \varphi^{-1}(AUC) \times 1.414$ and φ^{-1} is the inverse of standard cumulative normal distribution

Intermediate outcomes:

$\varphi^{-1}(AUC)$

a

$e^{-a^2/2}$

$6a^2 + 16$

$V(AUC)$

If “Compare two values of accuracy” is selected:

Test accuracy of a single test or compare accuracies of two tests

Testing the accuracy of a single test or comparing two different tests accuracies implies testing a hypothesis. A **comparison of the accuracies of two tests** refers to [test the equivalence of the accuracies of two tests](#). In both scenarios, the hope is to reject the hypothesis of equivalence and thus prove that the accuracies are different. **Testing the accuracy of a single test** means [studying the equivalence of your test to a predetermined level of accuracy](#).

Here is a calculator that helps you in testing your hypothesis with a desired precision, *i.e.* confidence level, and power.

Please select a mode

Compare accuracy of two tests
 Test accuracy of a single test

- If “Test accuracy of a single test” is selected:

You can determine the number of samples/subjects needed in order to ensure that your diagnostic test meets an acceptable pre-specified performance and that the study has high probability (power) to detect that your test is better than the minimal fixed accuracy [1].

The null hypothesis states that your test accuracy p_1 is equivalent to the pre-specified one p_0 , *i.e.* the phenomenon under test is absent.

$\alpha\%$ is the probability of falsely rejecting the null hypothesis and it gives a basis to prove the existence of the phenomenon to be demonstrated [2].

The power $(1-\beta)\%$ is the probability that the study will detect a difference in accuracies, given that the phenomenon actually exists [3].

This calculator uses a **two-tailed z-score test under the null and alternative hypothesis (conditional approach)** and normal approximation is the general rule [4]. One-sided tests are often possible but rarely used in clinical research.

Please enter either the sensitivity or specificity values.

Sensitivity/Specificity of reference, p_0 : % Pre-specified accuracy value, i.e. reference test.

Sensitivity/Specificity of the new test, p_1 : %

Significance level, α : % 5% significance level is considered as the standard in the literature. Higher significance levels lead to an increase in power. However, as the significance level increases, the probability of making an error in claiming a difference when in fact none exists becomes higher [3].

Power, $(1 - \beta)$: % The minimum power value is usually set as 80% [7].

Failure rate in sampling, q : % The sample processing, stretching from patient identification, to sample collection and result, might not work as planned due to technical or procedural problems (e.g. contamination, failure to get the test performed in time, subjects withdrawal, missing data). It is necessary to account for these issues when determining the sample size and an additional 10-20% subjects are recruited.

Sample size using a conditional approach, N :

Sample size using an unconditional approach, N :

Appendix

Formula for the computation of sample size [4]:

$$\text{Conditional approach: } N' = \frac{\left[\frac{Z_{\alpha} \sqrt{p_0(1-p_0)} + Z_{\beta} \sqrt{p_1(1-p_1)}}{2} \right]^2}{(p_1 - p_0)^2}$$

$$\text{Unconditional approach: } N' = \frac{\left(\frac{Z_{\alpha} + Z_{\beta}}{2} \right)^2 p_1(1-p_1)}{(p_1 - p_0)^2}$$

$$N = \frac{N'}{(1 - q)}$$

References:

- [1] B. Biswas, "Clinical Performance Evaluation of Molecular Diagnostic Tests," *J. Mol. Diagnostics*, vol. 18, no. 6, pp. 803–812, 2016, doi: 10.1016/j.jmoldx.2016.06.008.
- [2] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Second Edi. Lawrence Erlbaum Associates, 1988.
- [3] A. Beam, "Strategies Radiology," no. September, 1992.
- [4] K. Hajian-Tilaki, "Sample size estimation in diagnostic test studies of biomedical informatics," *J. Biomed. Inform.*, vol. 48, pp. 193–204, 2014, doi: 10.1016/j.jbi.2014.02.013.
- [5] J. L. Fleiss, A. Tytun, and H. K. Ury, "A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions Author (s): Joseph L . Fleiss , Alex Tytun and Hans K . Ury Published by : International Biometric Society Stable URL : <https://www.jstor.org/stable/2529990> Intern," *Int. Biometric Soc.*, vol. 36, no. 2, pp. 343–346, 1980, [Online]. Available: http://zulsidi.tripod.com/pdf/DeterminingSampleSizes.pdf%0Ahttps://www.jstor.org/stable/2529990?seq=1#metadata_info_tab_contents.
- [6] S.-C. Chow, B. Jones, J. Liu, and K. E. Peace, *Sample size estimation in diagnostic test studies of biomedical informatics*, Second Edi. Chapman & hall/crc biostatistics series, 2008.
- [7] N. Lu *et al.*, "Power analysis for cross-sectional and longitudinal study designs," *Shanghai Arch. Psychiatry*, vol. 25, no. 4, p. 259, 2013, doi: 10.3969/j.j.issn.1002-0829.2013.04.009.

Appendix

- If "Compare accuracy of two tests" is selected:

You can determine the number of samples/subjects needed in order to ensure that the two diagnostic test have different diagnostic abilities and that the study has high probability (power) to detect the actual difference.

The null hypothesis states that the two tests are equivalent in their ability to diagnose, where p_1 and p_2 are the accuracies of each test.

$\alpha\%$ is the probability of falsely rejecting the null hypothesis and it gives a basis to prove the existence of a difference in the phenomenon under test [2].

The power, $(1-\beta)\%$, is the probability that the study will detect a difference in diagnostic abilities, given that the phenomenon actually exists [3].

This calculator uses a **two-tailed z-score test under the null and alternative hypothesis (conditional approach)** for a **two-sample test of equality**. Normal approximation is the general rule and the **two study groups, n_1 and n_2 , are independent** [4], meaning that the two tests are evaluated on different subjects.

One-sided tests are often possible but rarely used in clinical research.

Please enter either the sensitivity or specificity values.

Sensitivity/Specificity of reference, p_1 : % Pre-specified accuracy value, i.e. reference test.

Sensitivity/Specificity of the new test, p_2 : %

Allocation ratio, $k = n_1/n_2$:

Significance level, α : % 5% significance level is considered as the standard in the literature. Higher significance levels lead to an increase in power. However, as the significance level increases, the probability of making an error in claiming a difference when in fact none exists becomes higher [3].

Power, $(1 - \beta)$: % The minimum power value is usually set as 80% [7].

Failure rate in sampling, q : % The sample processing, stretching from patient identification, to sample collection and result, might not work as planned due to technical or procedural problems (e.g. contamination, failure to get the test performed in time, subjects withdrawal, missing data). It is necessary to account for these issues when determining the sample size and an additional 10-20% subjects are recruited.

Sample size using conditional approach:

Sample size, N :

Sample size group 1, n_1 :

Sample size group 2, n_2 :

Sample size with the **continuity correction factor**: i

Sample size, N :

Sample size group 1, n_1 :

Sample size group 2, n_2 :

The continuity correction factor accounts for the lack of continuity in the binomial distribution [5].

Sample size using unconditional approach:

Sample size, N :

Sample size group 1, n_1 :

Sample size group 2, n_2 :

Sample size with the **continuity correction factor**: i

Sample size, N :

Sample size group 1, n_1 :

Sample size group 2, n_2 :

The continuity correction factor accounts for the lack of continuity in the binomial distribution [5].

Formula for the computation of sample size [6]:

Appendix

$$k = \frac{n_1}{n_2}$$

$$p = \frac{kp_1 + p_2}{1 + k}$$

$$\text{Conditional Approach: } N_2 = \frac{\left[Z_{\alpha} \sqrt{\left(1 + \frac{1}{k}\right)p(1-p)} + Z_{\beta} \sqrt{\frac{p_1(1-p_1) + p_2(1-p_2)}{k}} \right]^2}{(p_1 - p_2)^2}$$

$$\text{Unconditional approach: } N_2 = \frac{\left(\frac{Z_{\alpha} + Z_{\beta}}{2}\right)^2}{(p_1 - p_2)^2} \left[\frac{p_1(1-p_1)}{k} + p_2(1-p_2) \right]$$

$$n_2 = \frac{N_2}{(1 - q)}$$

$$N_1 = k * N_2$$

$$n_1 = \frac{N_1}{(1 - q)}$$

$$N = n_1 + n_2$$

When the continuity factor is applied the last 4 formulas change as follow:

$$N_2 = \frac{N_2}{4} \left[1 + \sqrt{1 + \frac{2(k+1)}{kN_2|p_1 - p_2|}} \right]^2$$

$$n_2 = \frac{N_2}{(1 - q)}$$

$$N_1 = k * N_2$$

$$n_1 = \frac{N_1}{(1 - q)}$$

$$N = n_1 + n_2$$

To compute the sample size taking into account the exact test and unconditional approaches refer to G*Power 3.1 is another powerful calculator to compute the sample size and can be found at <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html>. Several types of power analysis can be explored along with the 'a priori' one, that is the one performed by this sample calculator.

References:

- [1] B. Biswas, "Clinical Performance Evaluation of Molecular Diagnostic Tests," *J. Mol. Diagnostics*, vol. 18, no. 6, pp. 803–812, 2016, doi: 10.1016/j.jmoldx.2016.06.008.
- [2] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Second Edi. Lawrence Erlbaum Associates, 1988.
- [3] A. Beam, "Strategies Radiology," no. September, 1992.
- [4] K. Hajian-Tilaki, "Sample size estimation in diagnostic test studies of biomedical informatics," *J. Biomed. Inform.*, vol. 48, pp. 193–204, 2014, doi: 10.1016/j.jbi.2014.02.013.
- [5] J. L. Fleiss, A. Tytun, and H. K. Ury, "A Simple Approximation for Calculating Sample Sizes for Comparing Independent Proportions Intern," *Int. Biometric Soc.*, vol. 36, no. 2, pp. 343–346, 1980.
- [6] S.-C. Chow, B. Jones, J. Liu, and K. E. Peace, *Sample size estimation in diagnostic test studies of biomedical informatics*, Second Edi. Chapman & hall/crc biostatistics series, 2008.
- [7] N. Lu *et al.*, "Power analysis for cross-sectional and longitudinal study designs," *Shanghai Arch. Psychiatry*, vol. 25, no. 4, p. 259, 2013, doi: 10.3969/j.issn.1002-0829.2013.04.009.

References

- [1] ‘WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data’. Accessed: Jul. 12, 2022. [Online]. Available: <https://covid19.who.int/>
- [2] ‘Maps: Track how full ICUs are with Covid patients across the U.S.’ Accessed: Oct. 04, 2022. [Online]. Available: <https://www.nbcnews.com/health/health-news/hospital-icu-stress-level-tracker-n1287375>
- [3] A. K. Singh and A. Misra, ‘Impact of COVID-19 and comorbidities on health and economics: Focus on developing countries and India’, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 6, pp. 1625–1630, Nov. 2020, doi: 10.1016/J.DSX.2020.08.032.
- [4] Z. Wu and J. M. McGoogan, ‘Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention’, *JAMA*, vol. 323, no. 13, pp. 1239–1242, Apr. 2020, doi: 10.1001/JAMA.2020.2648.
- [5] P. K. Bhatraju *et al.*, ‘Covid-19 in Critically Ill Patients in the Seattle Region — Case Series’, *New England Journal of Medicine*, vol. 382, no. 21, pp. 2012–2022, May 2020, doi: 10.1056/NEJMOA2004500/SUPPL_FILE/NEJMOA2004500_DATA-SHARING.PDF.
- [6] G. Grasselli *et al.*, ‘Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy’, *JAMA*, vol. 323, no. 16, pp. 1574–1581, Apr. 2020, doi: 10.1001/JAMA.2020.5394.
- [7] Y. Wang *et al.*, ‘Clinical course and outcomes of 344 intensive care patients with COVID-19’, *Am J Respir Crit Care Med*, vol. 201, no. 11, pp. 1430–1434, Jun. 2020, doi: 10.1164/RCCM.202003-0736LE/SUPPL_FILE/DISCLOSURES.PDF.
- [8] L. C. Myers, S. M. Parodi, G. J. Escobar, and V. X. Liu, ‘Characteristics of Hospitalized Adults With COVID-19 in an Integrated Health Care System in California’, *JAMA*, vol. 323, no. 21, pp. 2195–2198, Jun. 2020, doi: 10.1001/JAMA.2020.7202.
- [9] G. Grasselli and A. Zanella, ‘Critically ill patients with COVID-19 in New York City’, *The Lancet*, vol. 395, no. 10239, pp. 1740–1741, Jun. 2020, doi: 10.1016/S0140-6736(20)31190-9.

- [10] E. Sherrard-Smith *et al.*, ‘The potential public health consequences of COVID-19 on malaria in Africa’, *Nature Medicine* 2020 26:9, vol. 26, no. 9, pp. 1411–1416, Aug. 2020, doi: 10.1038/s41591-020-1025-y.
- [11] D. J. Weiss *et al.*, ‘Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: a geospatial modelling analysis’, *Lancet Infect Dis*, vol. 21, no. 1, pp. 59–69, Jan. 2021, doi: 10.1016/S1473-3099(20)30700-3.
- [12] S. Bhatt *et al.*, ‘The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015’, *Nature* 2015 526:7572, vol. 526, no. 7572, pp. 207–211, Sep. 2015, doi: 10.1038/nature15535.
- [13] ‘Malaria’. Accessed: Dec. 04, 2022. [Online]. Available: <https://www.who.int/data/gho/data/themes/malaria>
- [14] Organização Mundial de Saúde, *World health statistics 2022 (Monitoring health of the SDGs)*. 2022. [Online]. Available: <http://apps.who.int/bookorders>.
- [15] ‘World malaria report 2021’. Accessed: Aug. 22, 2022. [Online]. Available: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2021>
- [16] ‘Key facts HIV’, 2022, Accessed: Sep. 04, 2023. [Online]. Available: https://cdn.who.int/media/docs/default-source/hq-hiv-hepatitis-and-stis-library/key-facts-hiv-2021-26july2022.pdf?sfvrsn=8f4e7c93_5
- [17] L. H. Miller, D. I. Baruch, K. Marsh, and O. K. Doumbo, ‘The pathogenic basis of malaria’, *Nature* 2002 415:6872, vol. 415, no. 6872, pp. 673–679, Feb. 2002, doi: 10.1038/415673a.
- [18] R. G. A. Feachem *et al.*, ‘Shrinking the malaria map: progress and prospects’, *The Lancet*, vol. 376, no. 9752, pp. 1566–1578, Nov. 2010, doi: 10.1016/S0140-6736(10)61270-6.
- [19] ‘World malaria report 2019’. Accessed: Sep. 04, 2023. [Online]. Available: <https://www.who.int/publications/i/item/9789241565721>
- [20] P. Yager, G. J. Domingo, and J. Gerdes, ‘Point-of-Care Diagnostics for Global Health’, <http://dx.doi.org/10.1146/annurev.bioeng.10.061807.160524>, vol. 10, pp. 107–144, Jul. 2008, doi: 10.1146/ANNUREV.BIOENG.10.061807.160524.
- [21] G. Snounou *et al.*, ‘High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction’, *Mol Biochem Parasitol*, vol. 61, no. 2, pp. 315–320, 1993, doi: 10.1016/0166-6851(93)90077-B.

- [22] S. Opoku Afriyie *et al.*, ‘Accuracy of diagnosis among clinical malaria patients: comparing microscopy, RDT and a highly sensitive quantitative PCR looking at the implications for submicroscopic infections’, *Malar J*, vol. 22, no. 1, pp. 1–11, Dec. 2023, doi: 10.1186/S12936-023-04506-5/FIGURES/5.
- [23] N. Tangpukdee, C. Duangdee, P. Wilairatana, and S. Krudsood, ‘Malaria Diagnosis: A Brief Review’, *Korean J Parasitol*, vol. 47, no. 2, p. 93, 2009, doi: 10.3347/kjp.2009.47.2.93.
- [24] N. H. Ahmed and J. C. Samantaray, ‘Quantitative Buffy Coat Analysis-An Effective Tool for Diagnosing Blood Parasites’, *J Clin Diagn Res*, vol. 8, no. 4, p. DH01, 2014, doi: 10.7860/JCDR/2014/7559.4258.
- [25] A. Clin Toxicol, ‘Advances in Clinical Toxicology Committed to Create Value for Researchers Malaria Tests-DNA PCR and Quantitative Buffy Coat (QBC) Malaria Tests-DNA PCR and Quantitative Buffy Coat (QBC)’, 2022, doi: 10.23880/act-16000232.
- [26] M. Kochareka, S. Sarkar, D. Dasgupta, and U. Aigal, ‘A preliminary comparative report of quantitative buffy coat and modified quantitative buffy coat with peripheral blood smear in malaria diagnosis’, *Pathog Glob Health*, vol. 106, no. 6, p. 335, 2012, doi: 10.1179/2047773212Y.0000000024.
- [27] L. Ochola, P. Vounatsou, T. Smith, M. Mabaso, and C. Newton, ‘The reliability of diagnostic techniques in the diagnosis and management of malaria in the absence of a gold standard’, *Lancet Infect Dis*, vol. 6, no. 9, pp. 582–588, Sep. 2006, doi: 10.1016/S1473-3099(06)70579-5.
- [28] F. Shujatullah, A. Malik, H. M. Khan, and A. Malik, ‘Comparison of different diagnostic techniques in Plasmodium falciparum cerebral malaria’, *J Vect Borne Dis*, vol. 43, pp. 186–190, 2006.
- [29] O. E. Lema *et al.*, ‘COMPARISON OF FIVE METHODS OF MALARIA DETECTION IN THE OUTPATIENT SETTING’, *Am. J. Trop. Med. Hyg*, vol. 60, no. 2, pp. 177–182, 1999.
- [30] ‘QBC Malaria Test System with ParaLens Advance and Centrifuge | HCE’. Accessed: Jul. 22, 2023. [Online]. Available: https://www.hce-uk.com/product/qbc-malaria-test-system-with-paralens-advance-and-centrifuge/?gclid=Cj0KCQjw_O2lBhCFARIsAB0E8B8rGYbeUyRRgwvAZLMKRKSCYBY3k9n7BfksHsjYBw6ioFd-KXqC3PwaAiPwEALw_wcB

- [31] D. S. Tarimo, J. N. Minjas, and I. C. Bygbjerg, 'Malaria diagnosis and treatment under the strategy of the integrated management of childhood illness (IMCI): relevance of laboratory support from the rapid immunochromatographic tests of ICT Malaria P.f/P.v and OptiMal', *Ann Trop Med Parasitol*, vol. 95, no. 5, pp. 437–444, 2001, doi: 10.1080/13648590120068971.
- [32] 'Global Malaria Programme'. Accessed: Aug. 14, 2023. [Online]. Available: <https://www.who.int/teams/global-malaria-programme/case-management/diagnosis/rapid-diagnostic-tests/how-malaria-rdts-work>
- [33] D. J. Kyabayinze, J. K. Tibenderana, G. W. Odong, J. B. Rwakimari, and H. Counihan, 'Operational accuracy and comparative persistent antigenicity of HRP2 rapid diagnostic tests for Plasmodium falciparum malaria in a hyperendemic region of Uganda', *Malar J*, vol. 7, 2008, doi: 10.1186/1475-2875-7-221.
- [34] 'New perspectives : malaria diagnosis : report of a joint WHO/USAID informal consultation, 25-27 October 1999'. Accessed: Sep. 04, 2023. [Online]. Available: <https://apps.who.int/iris/handle/10665/66321>
- [35] A. Mayor and Q. Bassat, "'Resistance" to diagnostics: A serious biological challenge for malaria control and elimination', *EBioMedicine*, vol. 50, pp. 9–10, Dec. 2019, doi: 10.1016/j.ebiom.2019.11.041.
- [36] B. Morassin, R. Fabre, A. Berry, and J. F. Magnaval, 'One year's experience with the polymerase chain reaction as a routine method for the diagnosis of imported malaria', *Am J Trop Med Hyg*, vol. 66, no. 5, pp. 503–508, 2002, doi: 10.4269/AJTMH.2002.66.503.
- [37] M. T. Makler, C. J. Palmer, and A. L. Ager, 'A review of practical techniques for the diagnosis of malaria', *Ann Trop Med Parasitol*, vol. 92, no. 4, pp. 419–433, 1998, doi: 10.1080/00034989859401.
- [38] D. P. Van Dijk, P. Gillet, E. Vlieghe, L. Cnops, M. Van Esbroeck, and J. Jacobs, 'Evaluation of the Palutop+4 malaria rapid diagnostic test in a non-endemic setting', *Malar J*, vol. 8, no. 1, 2009, doi: 10.1186/1475-2875-8-293.
- [39] T. Notomi *et al.*, 'Loop-mediated isothermal amplification of DNA', *Nucleic Acids Res*, vol. 28, no. 12, 2000, doi: 10.1093/NAR/28.12.E63.
- [40] N. Tomita, Y. Mori, H. Kanda, and T. Notomi, 'Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products', *Nature Protocols 2008 3:5*, vol. 3, no. 5, pp. 877–882, Apr. 2008, doi: 10.1038/nprot.2008.57.

- [41] M. Hema and N. C. Konakalla, 'Recent developments in detection and diagnosis of plant viruses', *Recent Developments in Applied Microbiology and Biochemistry: Volume 2*, pp. 163–180, Jan. 2021, doi: 10.1016/B978-0-12-821406-0.00016-3.
- [42] C. D. Kelly-Cirino *et al.*, 'Importance of diagnostics in epidemic and pandemic preparedness', *BMJ Glob Health*, vol. 4, no. Suppl 2, p. e001179, Feb. 2019, doi: 10.1136/BMJGH-2018-001179.
- [43] S. Sharma, J. Zapatero-Rodríguez, P. Estrela, and R. O'Kennedy, 'Point-of-Care Diagnostics in Low Resource Settings: Present Status and Future Role of Microfluidics', *Biosensors 2015, Vol. 5, Pages 577-601*, vol. 5, no. 3, pp. 577–601, Aug. 2015, doi: 10.3390/BIOS5030577.
- [44] S. Selvaraj and S. Sundaravaradhan, 'Challenges and opportunities in IoT healthcare systems: a systematic review', *SN Appl Sci*, vol. 2, no. 1, pp. 1–8, Jan. 2020, doi: 10.1007/S42452-019-1925-Y/TABLES/1.
- [45] P. Verma and S. K. Sood, 'Cloud-centric IoT based disease diagnosis healthcare framework', *J Parallel Distrib Comput*, vol. 116, pp. 27–38, Jun. 2018, doi: 10.1016/J.JPDC.2017.11.018.
- [46] 'The Mobile Economy 2023', 2023, Accessed: Feb. 21, 2024. [Online]. Available: www.gsmainelligence.com
- [47] 'The Mobile Economy Sub-Saharan Africa 2023', 2023, Accessed: Feb. 21, 2024. [Online]. Available: www.gsmainelligence.com
- [48] S. Majumder and M. J. Deen, 'Smartphone Sensors for Health Monitoring and Diagnosis', *Sensors 2019, Vol. 19, Page 2164*, vol. 19, no. 9, p. 2164, May 2019, doi: 10.3390/S19092164.
- [49] 'ARM Chips List'. Accessed: Feb. 07, 2023. [Online]. Available: <https://www.cs.umd.edu/~meesh/cmsc411/website/proj01/arm/armchip.html#Arm10>
- [50] 'Preparing for success at the next node with Calibre'. Accessed: Feb. 07, 2023. [Online]. Available: <https://www.techdesignforums.com/practice/technique/calibre-randd-node-introduction/>
- [51] A. K. Karlson, B. R. Meyers, A. Jacobs, P. Johns, and S. K. Kane, 'Working overtime: Patterns of smartphone and PC usage in the day of an information worker', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5538 LNCS, pp. 398–405, 2009, doi: 10.1007/978-3-642-01516-8_27/COVER.

- [52] B. G. Buchanan, 'A (Very) Brief History of Artificial Intelligence', *AI Mag*, vol. 26, no. 4, pp. 53–53, Dec. 2005, doi: 10.1609/AIMAG.V26I4.1848.
- [53] T. M. Cover and P. E. Hart, 'Nearest Neighbor Pattern Classification', *IEEE Trans Inf Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.
- [54] Y. Lecun, Y. Bengio, and G. Hinton, 'Deep learning', *Nature* 2015 521:7553, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [55] P. K. Mall *et al.*, 'A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities', *Healthcare Analytics*, vol. 4, p. 100216, Dec. 2023, doi: 10.1016/J.HEALTH.2023.100216.
- [56] Z. S. Ballard *et al.*, 'Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors', *npj Digital Medicine* 2020 3:1, vol. 3, no. 1, pp. 1–8, May 2020, doi: 10.1038/s41746-020-0274-y.
- [57] V. Turbé *et al.*, 'Deep learning of HIV field-based rapid tests', *Nature Medicine* 2021 27:7, vol. 27, no. 7, pp. 1165–1170, Jun. 2021, doi: 10.1038/s41591-021-01384-9.
- [58] L. Huang *et al.*, 'Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning', *Nature Communications* 2023 14:1, vol. 14, no. 1, pp. 1–14, Jan. 2023, doi: 10.1038/s41467-022-35696-2.
- [59] K. Mahato, A. Srivastava, and P. Chandra, 'Paper based diagnostics for personalized health care: Emerging technologies and commercial aspects', *Biosens Bioelectron*, vol. 96, pp. 246–259, Oct. 2017, doi: 10.1016/J.BIOS.2017.05.001.
- [60] E. Ricci and R. Perfetti, 'Retinal blood vessel segmentation using line operators and support vector classification', *IEEE Trans Med Imaging*, vol. 26, no. 10, pp. 1357–1365, Oct. 2007, doi: 10.1109/TMI.2007.898551.
- [61] 'Introduction - DRIVE - Grand Challenge'. Accessed: Apr. 19, 2021. [Online]. Available: <https://drive.grand-challenge.org/>
- [62] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, 'Retinal vessel segmentation via deep learning network and fully-connected conditional random fields', *Proceedings - International Symposium on Biomedical Imaging*, vol. 2016-June, pp. 698–701, Jun. 2016, doi: 10.1109/ISBI.2016.7493362.
- [63] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, 'DUNet: A deformable network for retinal vessel segmentation', *Knowl Based Syst*, vol. 178, pp. 149–162, Aug. 2019, doi: 10.1016/J.KNOSYS.2019.04.025.

- [64] C. Guo, M. Szemenyei, Y. Pei, Y. Yi, and W. Zhou, ‘SD-Unet: A Structured Dropout U-Net for Retinal Vessel Segmentation’, in *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, IEEE, Oct. 2019, pp. 439–444. doi: 10.1109/BIBE.2019.00085.
- [65] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, ‘IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks’, in *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, Institute of Electrical and Electronics Engineers Inc., Mar. 2020, pp. 3645–3654. doi: 10.1109/WACV45572.2020.9093621.
- [66] E. Decencière *et al.*, ‘FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE’, *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, Aug. 2014, doi: 10.5566/IAS.1155.
- [67] M. D. Abràmoff *et al.*, ‘Automated Analysis of Retinal Images for Detection of Referable Diabetic Retinopathy’, *JAMA Ophthalmol*, vol. 131, no. 3, pp. 351–357, Mar. 2013, doi: 10.1001/JAMAOPHTHALMOL.2013.1743.
- [68] M. Niemeijer *et al.*, ‘Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs’, *IEEE Trans Med Imaging*, vol. 29, no. 1, pp. 185–195, Jan. 2010, doi: 10.1109/TMI.2009.2033909.
- [69] J. Cuadros and G. Bresnick, ‘EyePACS: An Adaptable Telemedicine System for Diabetic Retinopathy Screening’, *Journal of diabetes science and technology (Online)*, vol. 3, no. 3, p. 509, 2009, doi: 10.1177/1932296809000300315.
- [70] H. Kang, Y. Gao, S. Guo, X. Xu, T. Li, and K. Wang, ‘AVNet: A retinal artery/vein classification network with category-attention weighted fusion’, *Comput Methods Programs Biomed*, vol. 195, p. 105629, Oct. 2020, doi: 10.1016/J.CMPB.2020.105629.
- [71] R. A. Welikala *et al.*, ‘Automated arteriole and venule classification using deep learning for retinal images from the UK Biobank cohort’, *Comput Biol Med*, vol. 90, pp. 23–32, Nov. 2017, doi: 10.1016/J.COMPBIOMED.2017.09.005.
- [72] F. Girard, C. Kavalec, and F. Cheriet, ‘Joint segmentation and classification of retinal arteries/veins from fundus images’, *Artif Intell Med*, vol. 94, pp. 96–109, Mar. 2019, doi: 10.1016/J.ARTMED.2019.02.004.
- [73] ‘RITE dataset | Department of Ophthalmology and Visual Sciences’. Accessed: Feb. 01, 2023. [Online]. Available: <https://medicine.uiowa.edu/eye/rite-dataset>

- [74] 'Bitcoin: A Peer-to-Peer Electronic Cash System'. Accessed: Apr. 15, 2020. [Online]. Available: <https://git.dhimmel.com/bitcoin-whitepaper/>
- [75] R. Kumar, N. Marchang, and R. Tripathi, 'Distributed Off-Chain Storage of Patient Diagnostic Reports in Healthcare System Using IPFS and Blockchain', in *2020 International Conference on COMMunication Systems and NETWORKS, COMSNETS 2020*, Institute of Electrical and Electronics Engineers Inc., Jan. 2020, pp. 1–5. doi: 10.1109/COMSNETS48256.2020.9027313.
- [76] J. Shin *et al.*, 'Mobile diagnostics: Next-generation technologies for: In vitro diagnostics', *Analyst*, vol. 143, no. 7. Royal Society of Chemistry, pp. 1515–1525, Apr. 07, 2018. doi: 10.1039/c7an01945a.
- [77] I. Institute for Business Value, 'Healthcare rallies for blockchains Keeping patients at the center Healthcare rallies for blockchains Keeping patients at the center Healthcare rallies for blockchains', 2017.
- [78] M. Nagaraj and S. Chaterji, 'Panel 3 Position Paper: Blockchain can be the Backbone of India's Economy', in *2019 11th International Conference on Communication Systems and Networks, COMSNETS 2019*, Institute of Electrical and Electronics Engineers Inc., May 2019, pp. 523–526. doi: 10.1109/COMSNETS.2019.8711052.
- [79] 'Hyperledger - The Open Global Ecosystem for Enterprise Blockchain'. Accessed: Sep. 04, 2023. [Online]. Available: <https://www.hyperledger.org/>
- [80] 'Documentation/TechnicalWhitePaper.md at master · EOSIO/Documentation · GitHub'. Accessed: Feb. 23, 2023. [Online]. Available: <https://github.com/EOSIO/Documentation/blob/master/TechnicalWhitePaper.md>
- [81] 'Ethereum Whitepaper | ethereum.org'. Accessed: Feb. 23, 2023. [Online]. Available: <https://ethereum.org/en/whitepaper/>
- [82] M. Xu, X. Chen, and G. Kou, 'A systematic review of blockchain', *Financial Innovation*, vol. 5, no. 1, pp. 1–14, Dec. 2019, doi: 10.1186/S40854-019-0147-Z/FIGURES/2.
- [83] O. Dib, A. Durand, K.-L. Brousmiche, E. Thea, and B. Hamida, 'Blockchain for Smart Transactions-IRT SystemX View project MIC (Modelling interoperability Communication) View project Consortium Blockchains: Overview, Applications and Challenges', *International Journal on Advances in Telecommunications*, no. 2,

- 2018, Accessed: Aug. 18, 2023. [Online]. Available: <http://www.iariajournals.org/telecommunications/2018>,
- [84] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, 'An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends', in *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, Institute of Electrical and Electronics Engineers Inc., Sep. 2017, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.
- [85] R. Yang *et al.*, 'Public and private blockchain in construction business process and information integration', *Autom Constr*, vol. 118, p. 103276, Oct. 2020, doi: 10.1016/J.AUTCON.2020.103276.
- [86] 'On Public and Private Blockchains | Ethereum Foundation Blog'. Accessed: Feb. 23, 2023. [Online]. Available: <https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains>
- [87] O. Dib, A. Durand, K.-L. Brousmiche, E. Thea, and B. Hamida, 'Consortium Blockchains: Overview, Applications and Challenges', *International Journal on Advances in Telecommunications*, no. 2, 2018, Accessed: Sep. 04, 2023. [Online]. Available: <http://www.iariajournals.org/telecommunications/2018>,
- [88] M. M. Merlec, M. M. Islam, Y. K. Lee, and H. P. In, 'A Consortium Blockchain-Based Secure and Trusted Electronic Portfolio Management Scheme', *Sensors (Basel)*, vol. 22, no. 3, Feb. 2022, doi: 10.3390/S22031271.
- [89] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, 'An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends', in *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, Institute of Electrical and Electronics Engineers Inc., Sep. 2017, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.
- [90] 'What are smart contracts on blockchain? | IBM'. Accessed: Feb. 23, 2023. [Online]. Available: <https://www.ibm.com/uk-en/topics/smart-contracts>
- [91] P. M. Kumar, S. Lokesh, R. Varatharajan, G. Chandra Babu, and P. Parthasarathy, 'Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier', *Future Generation Computer Systems*, vol. 86, pp. 527–534, Sep. 2018, doi: 10.1016/J.FUTURE.2018.04.036.
- [92] P. Verma, R. Tiwari, W. C. Hong, S. Upadhyay, and Y. H. Yeh, 'FETCH: A Deep Learning-Based Fog Computing and IoT Integrated Environment for Healthcare

- Monitoring and Diagnosis’, *IEEE Access*, vol. 10, pp. 12548–12563, 2022, doi: 10.1109/ACCESS.2022.3143793.
- [93] A. C. Garrett, ‘Development of a point-of-care molecular diagnostic tool for malaria and schistosomiasis in low-resource settings’, PhD thesis, University of Glasgow, Glasgow, 2021. Accessed: Jul. 26, 2023. [Online]. Available: <https://theses.gla.ac.uk/82700/>
- [94] J. Reboud *et al.*, ‘Paper-based microfluidics for DNA diagnostics of malaria in low resource underserved rural communities’, *Proc Natl Acad Sci U S A*, vol. 116, no. 11, pp. 4834–4842, Mar. 2019, doi: 10.1073/pnas.1812296116.
- [95] J. Reboud *et al.*, ‘Paper-based microfluidics for DNA diagnostics of malaria in low resource underserved rural communities’, *Proc Natl Acad Sci U S A*, vol. 116, no. 11, pp. 4834–4842, 2019, doi: 10.1073/pnas.1812296116.
- [96] X. Guo *et al.*, ‘Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security’, *Nat Electron*, vol. 4, no. 8, 2021, doi: 10.1038/s41928-021-00612-x.
- [97] Onsemi, ‘LM317 datasheet’. Accessed: Jan. 19, 2023. [Online]. Available: <https://www.onsemi.com/pdf/datasheet/lm317t-d.pdf>
- [98] J. Sattabongkot, T. Tsuboi, E. T. Han, S. Bantuchai, and S. Buates, ‘Loop-Mediated Isothermal Amplification Assay for Rapid Diagnosis of Malaria Infections in an Area of Endemicity in Thailand’, *J Clin Microbiol*, vol. 52, no. 5, p. 1471, 2014, doi: 10.1128/JCM.03313-13.
- [99] ‘Android Lollipop | Android Developers’. Accessed: Jul. 26, 2023. [Online]. Available: <https://developer.android.com/about/versions/lollipop>
- [100] ‘GitHub - DFRobot/BlunoBasicDemo: The basic demo for bluno’. Accessed: Jan. 17, 2023. [Online]. Available: <https://github.com/DFRobot/BlunoBasicDemo>
- [101] ‘Docker: Accelerated Container Application Development’. Accessed: Aug. 08, 2023. [Online]. Available: <https://www.docker.com/>
- [102] ‘REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces - Mark Masse - Google 图书’. Accessed: Dec. 10, 2022. [Online]. Available: <https://books.google.co.uk/books?hl=zh-CN&lr=&id=eABpzyTcJNIC&oi=fnd&pg=PR3&dq=rest+api&ots=vAVv46g8PE&sig=1vTKVeHXcpCN6sR3Gq-vyTnk0kU#v=snippet&q=http&f=false>
- [103] ‘Hyperledger Caliper’. Accessed: Aug. 08, 2023. [Online]. Available: <https://hyperledger.github.io/caliper/>

- [104] S. Pongnumkul, C. Siripanpornchana, and S. Thajchayapong, 'Performance analysis of private blockchain platforms in varying workloads', *2017 26th International Conference on Computer Communications and Networks, ICCCN 2017*, Sep. 2017, doi: 10.1109/ICCCN.2017.8038517.
- [105] X. Guo *et al.*, 'Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security', *Nature Electronics 2021 4:8*, vol. 4, no. 8, pp. 615–624, Aug. 2021, doi: 10.1038/s41928-021-00612-x.
- [106] K. Makanyadevi, S. Rithika, S. Biratheep, and S. Subanki, 'QR Code with Block Chain Technology for Medical Device Ownership', *2023 9th International Conference on Advanced Computing and Communication Systems, ICACCS 2023*, pp. 1760–1764, 2023, doi: 10.1109/ICACCS57279.2023.10112771.
- [107] N. Gupta and P. Bedi, 'E-waste Management Using Blockchain based Smart Contracts', *2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018*, pp. 915–921, Nov. 2018, doi: 10.1109/ICACCI.2018.8554912.
- [108] M. Lamichhane, O. Sadov, and K. Andersson, 'Pervasive Computing & COMmunications for sustainable development'.
- [109] S. Majumder and M. J. Deen, 'Smartphone Sensors for Health Monitoring and Diagnosis', *Sensors 2019, Vol. 19, Page 2164*, vol. 19, no. 9, p. 2164, May 2019, doi: 10.3390/S19092164.
- [110] D. Ravi *et al.*, 'Deep Learning for Health Informatics', *IEEE J Biomed Health Inform*, vol. 21, no. 1, pp. 4–21, Jan. 2017, doi: 10.1109/JBHI.2016.2636665.
- [111] P. Ongsulee, 'Artificial intelligence, machine learning and deep learning', *International Conference on ICT and Knowledge Engineering*, pp. 1–6, Jan. 2018, doi: 10.1109/ICTKE.2017.8259629.
- [112] Q. Dou *et al.*, 'Automatic Detection of Cerebral Microbleeds from MR Images via 3D Convolutional Neural Networks', *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1182–1195, May 2016, doi: 10.1109/TMI.2016.2528129.
- [113] H. Thomas, C. R. Qi, J.-E. Deschard, B. Marcotegui, F. Goulette, and L. J. Guibas, 'KPConv: Flexible and Deformable Convolution for Point Clouds'.
- [114] J. Luo, W. Zhang, J. Su, and F. Xiang, 'Hexagonal convolutional neural networks for hexagonal grids', *IEEE Access*, vol. 7, pp. 142738–142749, 2019, doi: 10.1109/ACCESS.2019.2944766.

- [115] W. Luo, Y. Li, R. Urtasun, and R. Zemel, ‘Understanding the Effective Receptive Field in Deep Convolutional Neural Networks’.
- [116] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, ‘FAST AND ACCURATE DEEP NETWORK LEARNING BY EXPONENTIAL LINEAR UNITS (ELUS)’.
- [117] D. P. Kingma and J. Lei Ba, ‘ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION’.
- [118] D. Choi, C. J. Shallue, Z. Nado, J. Lee, C. J. Maddison, and G. E. Dahl, ‘On Empirical Comparisons of Optimizers for Deep Learning’, Accessed: Feb. 16, 2023. [Online]. Available: <https://www.tensorflow.org/>
- [119] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘ImageNet Classification with Deep Convolutional Neural Networks’, Accessed: Jan. 09, 2023. [Online]. Available: <http://code.google.com/p/cuda-convnet/>
- [120] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.48550/arxiv.1512.03385.
- [121] ‘GitHub - tzutalin/labelImg:  LabelImg is a graphical image annotation tool and label object bounding boxes in images’. Accessed: Nov. 26, 2020. [Online]. Available: <https://github.com/tzutalin/labelImg>
- [122] ‘The Sequential model’. Accessed: Sep. 23, 2020. [Online]. Available: https://keras.io/guides/sequential_model/
- [123] ‘TensorFlow Lite | ML for Mobile and Edge Devices’. Accessed: Oct. 12, 2020. [Online]. Available: <https://www.tensorflow.org/lite>
- [124] S. H. Shabbeer Basha, S. Ram Dubey, V. Pulabaigari, and S. Mukherjee, ‘Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification’, 2019, doi: 10.1016/j.neucom.2019.10.008.
- [125] ‘Tensorflow 2 Model Zoo’. Accessed: Feb. 04, 2020. [Online]. Available: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md
- [126] K. M. F. Fuhad, J. F. Tuba, T. Rahman, and N. Mohammed, ‘CNN Based Model for Malaria Diagnosis with Knowledge Distillation’, *ACM International Conference Proceeding Series*, pp. 131–135, Jun. 2020, doi: 10.1145/3408127.3408159.

- [127] D. A. Mendels *et al.*, ‘Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation’, *Proc Natl Acad Sci U S A*, vol. 118, no. 12, p. e2019893118, Mar. 2021, doi: <https://www.pnas.org/doi/abs/10.1073/pnas.2019893118>.
- [128] D. F. Mujtaba and N. R. Mahapatra, ‘Lateral Flow Test Interpretation with Residual Networks’, *Proceedings - 2021 International Conference on Computational Science and Computational Intelligence, CSCI 2021*, pp. 1283–1286, 2021, doi: 10.1109/CSCI54926.2021.00261.
- [129] N. C. K. Wong *et al.*, ‘Machine learning to support visual auditing of home-based lateral flow immunoassay self-test results for SARS-CoV-2 antibodies’, *Communications Medicine 2022 2:1*, vol. 2, no. 1, pp. 1–10, Jul. 2022, doi: 10.1038/s43856-022-00146-z.
- [130] S. Lee *et al.*, ‘Sample-to-answer platform for the clinical evaluation of COVID-19 using a deep learning-assisted smartphone-based assay’, *Nature Communications 2023 14:1*, vol. 14, no. 1, pp. 1–11, Apr. 2023, doi: 10.1038/s41467-023-38104-5.
- [131] D. Kolosov, V. Kelefouras, P. Kourtessis, and I. Mporas, ‘Anatomy of Deep Learning Image Classification and Object Detection on Commercial Edge Devices: A Case Study on Face Mask Detection’, *IEEE Access*, vol. 10, pp. 109167–109186, 2022, doi: 10.1109/ACCESS.2022.3214214.
- [132] T. Tan and G. Cao, ‘Efficient Execution of Deep Neural Networks on Mobile Devices with NPU’, pp. 283–298, May 2021, doi: 10.1145/3412382.3458272.
- [133] X. Zhang, X. Zhou, M. Lin, and J. Sun, ‘ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices’. pp. 6848–6856, 2018.
- [134] D. Li, X. Wang, and D. Kong, ‘DeepRebirth: Accelerating Deep Neural Network Execution on Mobile Devices’, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, pp. 2322–2330, Apr. 2018, doi: 10.1609/AAAI.V32I1.11876.
- [135] A. Delaporte and K. Bahia, ‘THE STATE OF MOBILE INTERNET CONNECTIVITY 2022’, 2022, Accessed: Aug. 29, 2023. [Online]. Available: www.gsmainelligence.com
- [136] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, ‘Edge Computing: Vision and Challenges’, *IEEE Internet Things J*, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.

- [137] ‘Mapping the landscape of diagnostics for sexually transmitted infections : key findings and recommendations / Hannah Kettler, Karen White, Sarah Hawkes’. Accessed: Sep. 11, 2023. [Online]. Available: <https://apps.who.int/iris/handle/10665/68990>
- [138] K. J. Land, D. I. Boeras, X. S. Chen, A. R. Ramsay, and R. W. Peeling, ‘REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes’, *Nature Microbiology* 2018 4:1, vol. 4, no. 1, pp. 46–54, Dec. 2018, doi: 10.1038/s41564-018-0295-3.
- [139] R. W. Peeling, K. K. Holmes, D. Mabey, and A. Ronald, ‘Rapid tests for sexually transmitted infections (STIs): the way forward’, *Sex Transm Infect*, vol. 82, no. suppl 5, pp. v1–v6, Dec. 2006, doi: 10.1136/STI.2006.024265.
- [140] ‘How Many People Have Diabetes - Diabetes Prevalence Numbers’. Accessed: Feb. 01, 2023. [Online]. Available: <https://www.diabetes.co.uk/diabetes-prevalence.html>
- [141] E. J. Sussman, W. G. Tsiaras, and K. A. Soper, ‘Diagnosis of Diabetic Eye Disease’, *JAMA*, vol. 247, no. 23, pp. 3231–3234, Jun. 1982, doi: 10.1001/JAMA.1982.03320480047025.
- [142] ‘Common Eye Disorders and Diseases | CDC’. Accessed: Feb. 23, 2023. [Online]. Available: <https://www.cdc.gov/visionhealth/basics/ced/index.html>
- [143] Z. Fan *et al.*, ‘Accurate Retinal Vessel Segmentation via Octave Convolution Neural Network’. Accessed: Apr. 19, 2021. [Online]. Available: <https://github.com/JiajieMo/OctaveUNet>
- [144] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, ‘Deepvessel: Retinal vessel segmentation via deep learning and conditional random field’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9901 LNCS, pp. 132–139, 2016, doi: 10.1007/978-3-319-46723-8_16/FIGURES/4.
- [145] J. I. Orlando and M. Blaschko, ‘Learning fully-connected CRFs for blood vessel segmentation in retinal images’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8673 LNCS, no. PART 1, pp. 634–641, 2014, doi: 10.1007/978-3-319-10404-1_79/COVER.

- [146] F. Huang, B. Dashtbozorg, T. Tan, and B. M. ter Haar Romeny, 'Retinal artery/vein classification using genetic-search feature selection', *Comput Methods Programs Biomed*, vol. 161, pp. 197–207, Jul. 2018, doi: 10.1016/j.cmpb.2018.04.016.
- [147] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, 'IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks', in *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020*, Institute of Electrical and Electronics Engineers Inc., Mar. 2020, pp. 3645–3654. doi: 10.1109/WACV45572.2020.9093621.
- [148] 'Inspire Datasets | Department of Ophthalmology and Visual Sciences'. Accessed: Apr. 19, 2021. [Online]. Available: <https://medicine.uiowa.edu/eye/inspire-datasets>
- [149] M. R. K. Mookiah *et al.*, 'A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification', *Med Image Anal*, vol. 68, p. 101905, Feb. 2021, doi: 10.1016/J.MEDIA.2020.101905.
- [150] A. Galdran, M. Meyer, P. Costa, Mendonca, and A. Campilho, 'Uncertainty-aware artery/vein classification on retinal images', *Proceedings - International Symposium on Biomedical Imaging*, vol. 2019-April, pp. 556–560, Apr. 2019, doi: 10.1109/ISBI.2019.8759380.
- [151] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, 'Hard Attention Net for Automatic Retinal Vessel Segmentation', *IEEE J Biomed Health Inform*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020, doi: 10.1109/JBHI.2020.3002985.
- [152] E. Uysal and G. E. Güraksin, 'Computer-aided retinal vessel segmentation in retinal images: convolutional neural networks', *Multimed Tools Appl*, vol. 80, no. 3, pp. 3505–3528, Sep. 2020, doi: 10.1007/S11042-020-09372-W/FIGURES/11.
- [153] P. Saranya, S. Prabakaran, R. Kumar, and E. Das, 'Blood vessel segmentation in retinal fundus images for proliferative diabetic retinopathy screening using deep learning', *Visual Computer*, vol. 38, no. 3, pp. 977–992, Mar. 2022, doi: 10.1007/S00371-021-02062-0/FIGURES/13.
- [154] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, 'Comparative study of retinal vessel segmentation methods on a new publicly available database', <https://doi.org/10.1117/12.535349>, vol. 5370, pp. 648–656, May 2004, doi: 10.1117/12.535349.
- [155] M. Melinscak, P. Prentasic, and S. Loncaric, 'Retinal Vessel Segmentation Using Deep Neural Networks'.

- [156] ‘ADE20K dataset’. Accessed: Feb. 28, 2023. [Online]. Available: <https://groups.csail.mit.edu/vision/datasets/ADE20K/>
- [157] M. Gandhi and R. Dhanasekaran, ‘Diagnosis of diabetic retinopathy using morphological process and SVM classifier’, *International Conference on Communication and Signal Processing, ICCSP 2013 - Proceedings*, pp. 873–877, 2013, doi: 10.1109/ICCSP.2013.6577181.
- [158] ‘Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group’, *Ophthalmology*, vol. 98, no. 5 Suppl, pp. 786–806, 1991, doi: 10.1016/S0161-6420(13)38012-9.
- [159] ‘NHS Diabetic Eye Screening Programme: grading definitions for referable disease - GOV.UK’. Accessed: Jan. 28, 2023. [Online]. Available: <https://www.gov.uk/government/publications/diabetic-eye-screening-retinal-image-grading-criteria/nhs-diabetic-eye-screening-programme-grading-definitions-for-referable-disease>
- [160] ‘Diabetic Eye Screening Services in Scotland: A Training handbook – July 2003: page 17 |’. Accessed: Jan. 28, 2023. [Online]. Available: https://www.ndrs.scot.nhs.uk/?page_id=1609
- [161] ‘GitHub - wkentaro/labelme: Image Polygonal Annotation with Python (polygon, rectangle, circle, line, point and image-level flag annotation).’ Accessed: Aug. 31, 2023. [Online]. Available: <https://github.com/wkentaro/labelme>
- [162] M. D. Bloice, P. M. Roth, and A. Holzinger, ‘Biomedical image augmentation using Augmentor’, *Bioinformatics*, vol. 35, no. 21, pp. 4522–4524, Nov. 2019, doi: 10.1093/bioinformatics/btz259.
- [163] O. Ronneberger, P. Fischer, and T. Brox, ‘U-net: Convolutional networks for biomedical image segmentation’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
- [164] S. Zheng *et al.*, ‘Conditional Random Fields as Recurrent Neural Networks’.
- [165] ‘sadeepj/crfasrnn_keras: CRF-RNN Keras/Tensorflow version’. Accessed: Apr. 19, 2021. [Online]. Available: https://github.com/sadeepj/crfasrnn_keras
- [166] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, ‘Pyramid Scene Parsing Network’, Accessed: Jun. 08, 2023. [Online]. Available: <https://github.com/hszhao/PSPNet>

- [167] O. Oktay *et al.*, ‘Attention U-Net: Learning Where to Look for the Pancreas’.
- [168] Victor Jalil Ochoa Guterrez, ‘Optical Invasive & Non-invasive Monitoring for Healthcare’, Phd thesis, University of Glasgow, Glasgow, 2023.
- [169] ‘scikit-learn: machine learning in Python — scikit-learn 1.3.0 documentation’. Accessed: Aug. 30, 2023. [Online]. Available: <https://scikit-learn.org/stable/>
- [170] S. Go, J. Kim, K. J. Noh, S. J. Park, and S. Lee, ‘Combined Deep Learning of Fundus Images and Fluorescein Angiography for Retinal Artery/Vein Classification’, *IEEE Access*, vol. 10, pp. 70688–70698, 2022, doi: 10.1109/ACCESS.2022.3187503.
- [171] S. Luo, Z. Heng, M. Pagnucco, and Y. Song, ‘Two-Stage Topological Refinement Network for Retinal Artery/Vein Classification’, *Proceedings - International Symposium on Biomedical Imaging*, vol. 2022-March, 2022, doi: 10.1109/ISBI52829.2022.9761669.
- [172] J. Yi, C. Chen, and G. Yang, ‘Retinal artery/vein classification by multi-channel multi-scale fusion network’, *Applied Intelligence*, pp. 1–18, Aug. 2023, doi: 10.1007/S10489-023-04939-0/TABLES/5.
- [173] M. A. Bujang and T. H. Adnan, ‘Requirements for minimum sample size for sensitivity and specificity analysis’, *Journal of Clinical and Diagnostic Research*, vol. 10, no. 10. Journal of Clinical and Diagnostic Research, pp. YE01–YE06, Oct. 01, 2016. doi: 10.7860/JCDR/2016/18129.8744.
- [174] N. M. F. Buderer, ‘Statistical Methodology: I. Incorporating the Prevalence of Disease into the Sample Size Calculation for Sensitivity and Specificity’, *Academic Emergency Medicine*, vol. 3, no. 9, pp. 895–900, Sep. 1996, doi: 10.1111/j.1553-2712.1996.tb03538.x.
- [175] N. A. Obuchowski, ‘Sample size calculations in studies of test accuracy’, *Statistical Methods in Medical Research*, vol. 7, no. 4. Arnold, pp. 371–392, 1998. doi: 10.1177/096228029800700405.
- [176] D. L. Simel, G. P. Samsa, and D. B. Matchar, ‘Likelihood ratios with confidence: Sample size estimation for diagnostic test studies’, *J Clin Epidemiol*, vol. 44, no. 8, pp. 763–770, 1991, doi: 10.1016/0895-4356(91)90128-V.
- [177] S. Carley, S. Dosman, S. R. Jones, and M. Harrison, ‘Simple nomograms to calculate sample size in diagnostic studies’, *Emergency Medicine Journal*, vol. 22, no. 3, pp. 180–181, Mar. 2005, doi: 10.1136/emj.2003.011148.

- [178] ‘Sample Size Software | Power Analysis Software | PASS | NCSS.com’. Accessed: Aug. 31, 2023. [Online]. Available: <https://www.ncss.com/software/pass/>
- [179] P. S. Ganney, S. Pisharody, and E. Claridge, ‘Software engineering’, *Clin Eng*, pp. 131–168, Jan. 2020, doi: 10.1016/B978-0-08-102694-6.00009-7.
- [180] K. Hajian-Tilaki, ‘Sample size estimation in diagnostic test studies of biomedical informatics’, *Journal of Biomedical Informatics*, vol. 48. Academic Press Inc., pp. 193–204, 2014. doi: 10.1016/j.jbi.2014.02.013.
- [181] C. A. Beam, ‘Strategies for improving power in diagnostic radiology research’, *American Journal of Roentgenology*, vol. 159, no. 3, pp. 631–637, 1992, doi: 10.2214/AJR.159.3.1503041.
- [182] ‘Universität Düsseldorf: G*Power’. Accessed: Apr. 26, 2021. [Online]. Available: <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html>
- [183] ‘G*Power’. Accessed: Apr. 26, 2021. [Online]. Available: <https://stats.idre.ucla.edu/other/gpower/>
- [184] M. Kaushik, D. Agarwal, and A. K. Gupta, ‘Cross-sectional study on the role of public awareness in preventing the spread of COVID-19 outbreak in India’, *Postgrad Med J*, vol. 97, no. 1154, pp. 777–781, Dec. 2021, doi: 10.1136/POSTGRADMEDJ-2020-138349.
- [185] ‘Water, sanitation and hygiene (WASH)’. Accessed: Jun. 28, 2023. [Online]. Available: https://www.who.int/health-topics/water-sanitation-and-hygiene-wash#tab=tab_1
- [186] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, ‘High-Resolution Image Synthesis with Latent Diffusion Models’, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2022-June, pp. 10674–10685, Dec. 2021, doi: 10.1109/CVPR52688.2022.01042.
- [187] A. Vaswani *et al.*, ‘Attention Is All You Need’, *Adv Neural Inf Process Syst*, vol. 2017-December, pp. 5999–6009, Jun. 2017, Accessed: Sep. 12, 2023. [Online]. Available: <https://arxiv.org/abs/1706.03762v7>
- [188] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, ‘Ensemble deep learning: A review’, *Eng Appl Artif Intell*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/J.ENGAPPAI.2022.105151.
- [189] ‘Introducing ChatGPT’. Accessed: Sep. 12, 2023. [Online]. Available: <https://openai.com/blog/chatgpt>

- [190] M. Moor *et al.*, ‘Foundation models for generalist medical artificial intelligence’, *Nature* 2023 616:7956, vol. 616, no. 7956, pp. 259–265, Apr. 2023, doi: 10.1038/s41586-023-05881-4.
- [191] S. Harrer, ‘Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine’, *EBioMedicine*, vol. 90, Apr. 2023, doi: 10.1016/j.ebiom.2023.104512.
- [192] ‘Kaggle DR dataset (EyePACS)’. Accessed: Sep. 12, 2023. [Online]. Available: <https://www.kaggle.com/datasets/mariaherrerot/eyepacspreprocess>
- [193] K. Jin *et al.*, ‘FIVES: A Fundus Image Dataset for Artificial Intelligence based Vessel Segmentation’, *Scientific Data* 2022 9:1, vol. 9, no. 1, pp. 1–8, Aug. 2022, doi: 10.1038/s41597-022-01564-3.
- [194] T. A. LFD Consortium *et al.*, ‘Machine learning for determining lateral flow device results for testing of SARS-CoV-2 infection in asymptomatic populations’, *Cell Rep Med*, vol. 3, no. 10, p. 100784, Oct. 2022, doi: 10.1016/j.xcrm.2022.100784.
- [195] ‘Audere’. Accessed: Sep. 12, 2023. [Online]. Available: <https://www.auderenow.org/>
- [196] ‘HealthPulse Solutions’. Accessed: Sep. 12, 2023. [Online]. Available: <https://www.healthpulsenow.org/>
- [197] ‘UKHSA to deploy AI Digital Reader for lateral flow tests | UKAuthority’. Accessed: Sep. 12, 2023. [Online]. Available: <https://www.ukauthority.com/articles/ukhsa-to-deploy-ai-digital-reader-for-lateral-flow-tests/>
- [198] ‘Lateral Flow Market 2.0: Smartphones, the Lateral Flow Readers in Your Pockets’. Accessed: Sep. 12, 2023. [Online]. Available: <https://www.abingdonhealth.com/smartphones-the-lateral-flow-readers-in-your-pockets/>
- [199] J. K. Baird, ‘Eliminating malaria—all of them’, *The Lancet*, vol. 376, no. 9756, pp. 1883–1885, Dec. 2010, doi: 10.1016/S0140-6736(10)61494-8.