Anglia Ruskin University

Faculty of Science and Engineering

An Intelligent Image-based Colourimetric Test Framework for Diagnosis

A thesis in partial fulfilment of the requirements of Anglia Ruskin University for the degree of Doctor of Philosophy

MARZIA HOQUE TANIA

November 2018

Abstract

The global imbalance between the healthcare provider and patient ratio, an increasingly elderly population and resource-limited settings have triggered the demand for point-of-care (POC) platforms, prompting the growth of personalised healthcare and homecare solutions. This thesis presents an investigation into an AI-enabled image-based system to perform automatic colourimetric tests in real-time. The case study of wet-chemical-based enzyme-linked immunosorbent assay (ELISA) and dry-chemical-based lateral flow assay (LFA) were utilised to design and develop an intelligent framework for chromaticity analysis with minimal user intervention or additional hardware attachments.

The proposed system was designed by exploring state-of-the-art solutions for each component of an image-based colourimetric test, trial and error, and domain knowledge. At first, a reaction phase and time-dependent approach was proposed to track the dynamic changes in a colourimetric reaction by calculating the Euclidean distances. Subsequently, the final static stage of the reactions were considered and the images were pre-processed and segmented before applying vigorous noise removal techniques. The 10-fold cross-validated classifiers were trained with the optimum number of features using supervised machine learning. A completely separate testing dataset was utilised while testing the model. Additionally, a pre-trained model of deep learning was deployed to determine the type of colourimetric test, which can be integrated into the system where feasible.

Based on our study, the reaction phase and time-dependent scheme was found to be more suitable for wet-chemical-based assays, particularly for low concentration samples. In addition to classification, the approach can assist in optimising the reaction time. However, due to the requirement of significant memory space by the video frames, the final system consisted of an alternative approach - considering only the reaction phase and time-independent scheme. On an ideal condition, the later approach provided more than 98% accuracy for colourimetric decision. Furthermore, the exploration of a pre-trained deep learning model revealed its strength in the test-type detection, instead providing the colourimetric classification. Therefore, deep learning was deployed to initiate the system based on the assay type (i.e. ELISA or LFA), which provided 100% accuracy.

The system we demonstrated complies with the ASSURED criteria. As compared to the existing systems, the proposed intelligent and robust system with real-time processing capabilities has experienced a more extensive course of validation to enumerate the reliability of the system. Unlike most of the works in the literature, the proposed system provided the colourimetric prediction without any opto-mechanical attachment. Such an easy-to-use and computationally efficient system can be integrated into a server or deployed on a mobile platform to create better harmony between biochemical and computational complexity and eliminate the subjectivity of interpretation.

Keywords: Colourimetric test, intelligent systems, image processing, feature selection, POC system, ELISA, LFA, RDT

Acknowledgements

All praise and thanks to the Allah the Most Gracious and the Most Merciful.

First of all, I would like to take this opportunity to express my gratitude to my supervisory team. I am grateful to my First supervisor, Prof Alamgir Hossain for his endless support, without who my ARU journey would not have been made possible. I would also like to thank my co-supervisors Dr Khin Lwin and Dr Antesar Shabut for their constant guidance.

I would like to express my deep appreciation to Team-ARITI; Faculty of Science and Engineering, especially the Research team of the faculty; to Dr Charlotte Nevison, Dr Ross Kemble and the Doctoral School; to the Student Services, Employment Bureau, Student Union, International Advisors, Subject Librarian and Study Skills Plus; and to Dr Jufen Zhang and the Faculty of Health, Education, Medicine and Social Care.

When I was struggling with the dataset, thanks to Dr Julian Priddle, Dr Nicholas Pugh and Dr Grisha Pirianov to come to the rescue. Charys Palmer and Ibrahim Dumbuya- thank you for your help on the quantitative colourimetric test. A special thanks belong to Dr Mohammad Najlah for his generosity to help me generating the dataset for the Case Study 2. Another person who I must mention is Mr Paul Cotton, who made me feel welcome in the SuperLabs.

I would also like to thank my former co-supervisor Dr Kamal Abu-Hassan, master's supervisor Prof Fahmida Gulshan and mentor Prof ASW Kurny. I thank Dr Fred Sherratt for making me realise it again that how much I enjoy research. I thank Dr Ziad Hunaiti and Prof Mohammad Patwary for their invaluable guidance during my PhD journey.

I would like to express my profound gratitude to Dr Linda Brown and Dr Doa'a Hatem D. Al-Jamal. I thank Dr Shamim Kaiser, especially for the critical reading. I thank Ms Julia Lane, the magician. I thank Dr Punnarumol Temdee for her insights when I first started to work on AI.

I cannot thank enough Jess Polaski, Amannda Wilsonium, the Islamic Society, the Muslim Chaplain and the International Café. I thank the Bangladeshi community of Chelmsford for making me feel at home. I am in lack of words to thank my family and friends in the UK and who are five thousand miles away.

Declaration

I hereby declare that this thesis has been genuinely carried out by myself and I confirm that it has not been submitted for the award of any other degree at Anglia Ruskin University or any other educational institution. I also declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. The invaluable participation of others in this thesis has been acknowledged where appropriate.

Table of Contents

List of Tables

List of Figures

List of Acronyms

A-criteria	Affordable
AI	Artificial Intelligence
ALP	Alkaline Phosphatase
ANN	Artificial Neural Network
ASSURED	Affordable, Sensitive, Specific, User-friendly, Rapid and Robust, Equipment-free and Deliverable to end users
AuNP	Gold Nanoparticles
BRANN	Bayesian Regularisation Artificial Neural Network
CI	Confidence Interval
CIE	International Commission on Illumination
CKD	Chronic Kidney Disease
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CRD	Congo Red Dot
DAG	Directed Acyclic Graph
D-criteria	Deliverable
DNG	Digital Negative
DSS	Decision Support System
E-criteria	Equipment-Free
ELISA	Enzyme-Linked Immunosorbent Assay
FC	Fully Connected
GEMM	GEneral Matrix Multiplication
HCA	Hierarchical Clustering Analysis
HIV	Human Immunodeficiency Virus
HCG	Human Chorionic Gonadotropin
HSA	Human Serum Albumin

ISO	International Standards Organisation (In digital photography ISO measures the sensitivity of the image sensor)
JPEG	Joint Photographic Experts Group
KNN	K-Nearest Neighbours
LDA	Linear Discriminant Analysis
LED	Light-Emitting Diode
LFA	Lateral Flow Assay
LPF	Low Pass Filter
LS-SVM	Least Squares Support Vector Machine
MLP	Multilayer Perceptron
NIST	National Institute Of Standards And Technology
PCA	Principle Component Analysis
POC	Point-Of-Care
PSA	Prostate-Specific Antigen
RBF	Radial Bias Function
R-criteria	Robust, Rapid
RDT	Rapid Diagnostic Test
ReLU	Rectified Linear Unit
RF	Random Forest
ROC	Receiver Operating Characteristic
ROI	Region of Interest
S-criteria	Specific, Sensitive
SDGM	Stochastic Gradient Descent with Momentum
SVM	Support Vector Machine
TML	Traditional Machine Learning
WHO	World Health Organisation
3D	Three-dimension

Deliverables

Segments of this doctoral research have been published and presented at several international and local events. A list of publications is provided based on the category of the event or publication, to include international conferences, journals, and presentations, as follows.

Journals

Shabut, A. M., **Hoque Tania, M.**, Lwin, K. T., Evans, B. A. Yusof, N. A., Abu-Hassan, K. J. & Hossain, M. A. (2018). An Intelligent Mobile-Enabled Expert System for Tuberculosis Disease Diagnosis in Real Time, Expert Systems with Applications, 114, pp. 65-77. https://doi.org/10.1016/j.eswa.2018.07.014

Hoque Tania, M., Lwin, K. T., & Hossain, M. A. (2018). Advances in Automated Tongue Diagnosis Techniques, Integrative Medicine Research, 8(1), pp. 42-56. https://doi.org/10.1016/j.imr.2018.03.001

Conferences

Hoque Tania, M., Lwin, K. T., Shabut, A.M. & Hossain, M. A.; Clustering and Classification of a Qualitative Colorimetric Test, in *IEEE International Conference on Computing, Electronics & Communications Engineering 2018 (iCCECE 2018)*, Essex, UK. 16-17 August 2018.

AbuHassan, K., Bakhori, N. E. M., Kusnin, N., Azmi, U. Z. M., **Hoque Tania, M.**, Evans, B. A., Yusof, N. A. , Hossain, M. A.; Automatic Diagnosis of Tuberculosis Disease Based on Plasmonic ELISA and Color-based Image Classification. In: *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2017)*, Jeju Island, Korea, 11-15 July 2017.

Hoque Tania, M., Lwin, K.T., AbuHassan, K, Bakhori N.M., Azmi, U.Z.M., Yusof, N.A., Hossain, M. A; An Automated Colourimetric Test by Computational Chromaticity Analysis: A Case Study of Tuberculosis Test. In: *11th International Conference on Practical Applications of Computational Biology & Bioinformatics. PACBB 2017)*, Springer, Porto, Portugal, 21-23 June 2017. Advances in Intelligent Systems and Computing, 616, pp. 313-320, Springer, Cham.

Hoque Tania, M., Lwin, K.T., Hossain, M. A; Computational Complexity of Image Processing Algorithms for an Intelligent Mobile Enabled Tongue Diagnosis Scheme. In: *10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA 2016)*, IEEE, Chengdu, China, 15 - 17 December 2016.

ARU Events

Hoque Tania, M.; AI-enabled Image-based Colourimetric Test for Medical Diagnosis. *The MEDiterranean Public HEALTH Alliance (MED-HEALTH)*, 8-9 August 2018, Chelmsford, UK. (Presentation)

Hoque Tania, M.; Deep Learning for pH test. *12th Annual Research Student Conference*, 26 July 2018, Chelmsford, UK. (Presentation)

Hoque Tania, M.; Intelligent Systems for ASSURED Colourimetric Tests. *AI for Medical Informatics, Engineering and Devices*. 28 June 2018, Chelmsford, UK. (Presentation)

Hoque Tania, M., Lwin, K.T., AbuHassan, K., Hossain, M. A; Chromaticity Analysis of Plasmonic ELISA based Tb Test and Impact of Concentration on Chemical Kinetics. *11th Annual Research Student Conference*, 7 July 2017, Cambridge, UK. (Poster presentation)

Hoque Tania, M., Lwin, K.T., AbuHassan, K., Hossain, M. A; Mobile Enabled Chromaticity Analysis for Testing TB in Real-time. *Faculty of Science & Technology 7th Annual Research Conference*, 10 May 2017, Chelmsford, UK. (Poster presentation)

Hoque Tania, M., Lwin, K.T., Hossain, M. A; Performance Analysis of Image Processing Techniques in the field of Tongue Diagnosis. *Faculty of Science & Technology 6th Annual Research Conference*, 6 July 2016, Chelmsford, UK. (Poster presentation)

(This page has been left intentionally blank.)

Chapter 1

Introduction

1.1 Background and Motivation

Recent advancements in the field of widely accessible (GSMA Intelligence, 2017) smart personal devices (e.g. Smartphones, Tablets, iPads) have been unprecedented and allowed scientists and healthcare professionals to improve our quality of life by leaps and bounds. However, the opportunities to exploit this widely accessible technology, paradoxically, have not been developed in large parts of the world population due to the lack of quality digital healthcare applications, particularly for diagnosis. This research aims to investigate the development of an image-based intelligent diagnostic test scheme using chromaticity analysis, suitable to be integrated on mobile devices. It particularly focuses on the design and development of a comprehensive framework which can perform the automatic colourimetric analysis based on user requirements for different applications. The proposed system uses hybrid image processing and machine learning algorithms to provide a faster solution for a wide range of colourimetric tests to automatically separate colours according to its relevance with relatively high accuracy, specificity and sensitivity. Two testbeds were chosen to contemplate the breadth of colourimetric tests. The proposed scheme has been evaluated, tested and validated through a set of experiments. The demonstrated system is capable of providing more comprehensive healthcare support where there are no (or limited) experts available, particularly in remote locations.

Socio-Economic Context

The early diagnosis facility, the disproportional ratio of health professionals (doctor, expert, staff, carer) to patients and the advancement of technology have an impact in the field of healthcare prompting the sector of mobile phone-based microscopy, assays, and sensing platforms for Point-Of-Care (POC) diagnostics (Contreras-naranjo, Wei and Ozcan, 2016; Rajan and Glorikian, 2009). There are <1 physician/1000 population for more than 44% of the World Health Organisation (WHO) member states (World Health Organization, 2017). Even in a developed country such as the UK, there are only 2.806 doctors for every 1000 people.

The longevity of human life has given rise to increasing the understanding of age-related disabilities and diseases, which can create significant burdens on healthcare systems. To support the elderly population, which is expected to increase to 1.91 billion in 2050 (P. D. of the Department of Economic and S. A. of the United Nations Secretariat, 2012), and limit the spread of pandemics, an intelligent, clearer and easier system with the least error-prone diagnosis results is required for both patients and clinicians.

The ratio of health professionals to patients cannot be dramatically changed. However, the number of mobile phones and tablets are growing five times faster than the human population (U.S. Census Bureau, 2017; GSMA Intelligence, 2017). Therefore, a mobile-enabled low-cost POC system could easily reach a wider population quickly. The POC devices have the potential to act as cost-effective, autonomous, portable diagnosis systems, as well as to detect biological warfare agents, pathogens in animals and plants, food and water contamination, environmental such as indoor air quality monitoring (Minogue et al., 2014; Liu, Lin and Lillehoj, 2014). Globally, the POC testing market is expected to reach over 23.15 billion GBP by 2021 (Research and Market and BCC Research, 2016).

In the absence of expert clinical staff, there is a requirement for systems that are easily operable. Such a system could be used by aid workers in remote places to support primary health care, in time of epidemic and environmental monitoring for many purposes such as identifying safe drinking water. The easy-to-use system could support the growing need of the elderly population as well. Moreover, the Artificial Intelligence (AI) for the healthcare solution is growing and expected to be 6.6 billion USD by 2021. By 2026, an association between health professionals and AI applications can possibly save 150 billion USD for the healthcare system in the USA (Accenture, 2017).

Mobile phone cameras are becoming more powerful every day. The integration of AI into mobile phones can administer a technically and financially feasible healthcare system. In this way, a smartphone-based expert system would be able to offer simplicity, immediacy, non-invasiveness and automatic health risk analysis, diagnosis and clinical decision-making process. In the future, temperature, tongue diagnosis, retinopathy, facial expressions, skin colour and surface could all be vital parameters for such expert systems. However, there are many existing challenges within the recent

advances in automatic disease diagnostic tools. The list includes- a requirement of blood serum collected by non-professionals, accuracy, reliability, detection time and requirement of second confirmatory tests.

Context of Colourimetric Test

WHO prefers the diagnostic system to be inexpensive, disposable and easy-to-use (Khademhosseini, 2011; Wang, Xu and Demirci, 2010). Such a diagnostic system should follow the criteria known as ASSURED (Affordable, Sensitive, Specific, User-friendly, Rapid and robust, Equipment-free, and Deliverable) (Kettler, White and Hawkes, 2004). The colourimetric tests can be considered as a strong candidate to accomplish this goal.

Adoption of colourimetric tests in clinical chemistry goes back to centuries. Newfound materials, advances in nanotechnology and optical sensing have opened the door for endless possibilities with colourimetric applications, from urinalysis to solving crimes, from soil testing to the detection of chemical weapons.

Acknowledging the broader scope of colour analysis, on close inspection concerning the resource-limited settings, however, can impeach the means of colourimetric analysis. If a diagnostic result is going to be determined by naked-eye measurements as shown by O'Connor, Paterson and de la Rica (2016), then one should also consider the issue of colour vision deficiency, which is widely known as colour blindness. Moreover, a complex colour representation such as Matthews et al. (2012) may rely on the proficiency of the user. Therefore, the colourimetric tests must be free from the subjectivity of interpretation.

Knowledge Gap and Challenges

The current point of view in the field is from a number of disciplines; dominated by biochemistry, nanotechnology and optoelectronics. Findings concerning such systems have presented the prospect of isolated individual colourimetric components but often lack in the rigorous detailing of how such a system can be and should be designed. Therefore, there is a need for an extensive study to deal with the inadequacy to perceive colourimetric tests from the frame of computer vision. The challenges include the quest for searching a suitable image processing technique for robust operation of colourimetric tests. There is a requirement of knowledge exploration for such

techniques to develop a better understanding of colourimetric test data, which can facilitate better management of the computational complexity of such data.

Analysing the extracted features from images of colourimetric tests could help to create domain-specific knowledge. Identifying the key features and how the features are being analysed can play a crucial role in the core model of a colourimetric Decision Support System (DSS). Thus, feature optimisation and feature analysis techniques would be a promising contribution.

Consideration of AI and at the same time resource-limited settings to fulfil the ASSURED criteria has compelled us to thoroughly appraise the hindrance to develop a low cost computer-aided colourimetric test suitable for mobile devices. The influence of mobile phones has drawn the attention of many researchers recently. However mobile devices have limited memory, processing and power capacity. Thus, there is a need for a system that is computationally inexpensive. The key challenge would be to develop a robust system for any colourimetric test with better accuracy, efficiency and flexibility. To design, develop and deploy an intelligent colourimetric test scheme in 'real time' without any additional hardware attachment is going to be the biggest challenge because the result has to be produced with the camera and processing capacity of mobile devices. The dominant paradigm in AI is the adoption of server-based approaches.

While it is true that a server-based colourimetric test can assist in telepathology and compensate the limited capacity of the mobile devices, it could be argued that such system can restrict anytime anywhere access and be vulnerable to cybercrimes. Consequently, it contradicts with 'deliverability' of the ASSURED criteria. Therefore, there is a need for a system that is efficient as a stand-alone system for the mobile device and does not conflict with the idea of server-based application either.

Homecare and resource-limited settings including field operation would require a system to be simultaneously easy-to-use, flexible, portable and robust. Many researchers have insisted on mobile-enabled rapid colourimetric tests. For the application in medical diagnosis, colourimetric tests are in vitro and not an auxiliary test, more often a confirmatory test. Real-time execution of a robust colourimetric system while maintaining high accuracy in the mobile device is a crucial impediment.

1.2 Research Questions

The key research question of this study is:

> **How to develop an intelligent image-based real-time colourimetric test framework for diagnosis?**

The key research question can be articulated as the following sub-questions:

Q.1: How to design and develop an intelligent, robust and real-time processing framework for the dynamic changes in a colourimetric test?

Q.2: How to design and develop an intelligent, robust and real-time processing framework using a static image of a colourimetric test?

1.3 Aims and Objectives

The broader objective of this thesis is therefore to address the above questions:

To develop an intelligent framework which can perform the automatic colourimetric analysis in real-time using their personal devices based on users' requirements for medical diagnosis.

The primary objective can be subcategorised into the following specific objectives:

Objective #1: To investigate and critically observe images of colourimetric tests and provide the colourimetric decision considering the change of colours throughout the reaction using video-frames analysis.

Objective #2: To investigate and strategise a completely automatic and computationally efficient image processing framework, perform feature selection and feature optimisation for wet-chemical and dry-chemical colourimetric assays.

Objective #3: To investigate the AI algorithms and their suitability to develop the proposed colourimetric test framework.

Objectives #4: To develop a decision-making scheme to provide anytime anywhere real-time colourimetric test results.

1.4 List of Contributions

The research contributions of this doctoral work are as follows:

1) The investigation into an automatic colourimetric detection framework that fulfils the ASSURED (Affordable, Specific, Sensitive, User-friendly, Robust, Rapid, Equipment-free and Deliverable) criteria to test wet-chemical-based Enzyme-Linked Immunosorbent Assay (ELISA) and dry-chemical based Lateral Flow Assay (LFA). It is worth mentioning that most of the other approaches in the literature attempted to address ASSURED criteria with some constraints and limitations.

2) The pseudo-control colours have been proposed as a novel feature-set to be considered along with the rest of the features, i.e. lower order colour moments to perform classification of paper-based assays.

3) A new approach is proposed for colourimetric classification using video-frame analysis by tracking the colourimetric reaction.

4) An innovative approach is proposed using advanced machine learning (deep transfer learning) algorithms for assay type detection.

5) The state-of-the-art technologies are explored to identify the research gap for colourimetric tests.

1.5 Case Studies for Validation of the Concept

Case Study 1

This thesis investigates an intelligent plasmonic ELISA based Tuberculosis (TB)-specific antibody detection as the first case study and refers to this case study as TB-test. To the best of our knowledge, this is the first attempt to use this case study for image-based automatic colourimetric detection[1]. The conventional plasmonic ELISA based techniques produce colours for positive and negative samples. However, making a final decision based on the colour appearance is not always accurate. Therefore, this thesis aims to provide a proof-of-concept for an intelligent wet-chemical-based

---

[1] This case study is part of the research project, '*TB-Test - A smart mobile enabled scheme for tuberculosis testing*'. This research was supported by British Council Newton Institutional Links and Newton-Ungku Omar Fund (Grant ID: 216385726) and was a collaboration between Anglia Ruskin University (UK) and Universiti Putra Malaysia (Malaysia). The final outcome of the project was devised as an Android application to perform TB-test. The mobile application itself is not part of this thesis.

qualitative colourimetric tests suitable for POC platform that takes the final decision based on the colour analysis using the case study of TB-test.

Case Study 2

This thesis considers a dry-chemical based or microfluidic LFA using universal pH indicator paper as the second case study. The case study of ELISA and LFA varies regarding the geometric features, result type, materials and colours. The second case study is chosen due to its stability as compared to the first case study. This case study can presumably assist in creating a more consistent dataset so that external factors can be analysed with more certainty and the result of the overall system can be claimed with more confidence. We acknowledge that due to the nature of the assay type this case study may require some corrections in the image processing scheme. Being a dry-chemical assay, the feature-set probably can be downsized, and we will also examine if the same classifier can be used or not. Moreover, the nonexistence of an analogous system to the first case study in the literature constricted this research to illustrate any comparative analysis. The prevalence of similar case studies as the second case study in the literature will enable us to perform a fair comparison with the existing works.

1.6 Organisation of the Thesis

Chapter 2 provides an overview of the colourimetric test from the perspective of computer vision by describing different aspects of a colourimetric test in the context of image processing. The state-of-the-art systems are inspected and described by order of associated components. Based on the literature, this chapter portrays the rationales behind the thesis.

Chapter 3 manifests the problem of image-based colourimetric test for POC platform. This chapter depicts a high-level illustration of the adopted research methodology and research design.

Chapter 4 evaluates the capacity of deep learning models to perform a colourimetric test on the POC platform in real-time. In the proposed framework, the pre-trained model of the deep neural network is a discretional layer to detect the type of an assay.

Chapter 5 presents the case study of an intelligent wet-chemical plasmonic ELISA to determine the presence or absence of TB bacteria. This chapter considers the above

mentioned specific objects and discusses each objective by analysing the experimental results. Considering the dynamic changes in the images of TB-test, this chapter aims to provide a unique solution as part of the research contribution. Furthermore, exploring the images of the end point of TB-test, this chapter examines the scope of delivering a high performing, easy-to-use, robust and rapid system that is suitable for the personal devices.

Chapter 6 presents the case study of an LFA using universal pH indicator paper. This chapter aims to provide a proof-of-concept for a dry-chemical based or microfluidic, stable and semi-quantitative assay using a larger dataset with diverse conditions. The purpose of this chapter is to deliberately examine the scope of an intelligent colourimetric test that fulfils ASSURED criteria and demonstrate the claim as well. The chapter also asserts the contribution regarding the pseudo-control colour.

Chapter 7 summarises the contributions of this doctoral research and indicates the future scope.

Chapter 2

Literature Review

2.1 Introduction

The world surrounded by colours is nothing but a characteristic of human visual perception. Thus, analysing colours on a mobile platform can open the door for colour painting by toddlers to the detection of oxidised proteins in an Alzheimer patient's brain. Such applications are exemplified in Section 2.2. In Section 2.3, this chapter presents an overview of the colourimetric test from the perspective of computer vision. The strengths and weaknesses of the computer-aided system using different approaches (e.g. hardware attachment, mobile) in the literature are discussed in Section 2.4. This chapter also explores the algorithms used in the literature for colourimetric tests to discuss the methods and their performances in Section 2.5. Based on our research goal, this chapter critically assesses the research gap in the state-of-the-art systems in the reported articles with the aid of the evaluation criteria (Section 2.6). Finally, a discussion is provided in Section 2.7 regarding the justifications for this research.

2.2 Applications of Colourimetric Test

When it comes to colourimetric assays for medical diagnosis, a wide range of rapid, visual readout, quantitative detection, low cost and robust systems have been utilised in the literature (Guo et al., 2016; Koo, Wee and Trau, 2016; Barbosa et al., 2015; Berg et al., 2015; Cate et al., 2015; Giavazzi et al., 2014; Yetisen, 2014; Yetisen, Akram and Lowe, 2013; You, Park and Yoon, 2013; Mavandadi et al., 2012; Lin and Scott, 2012b; Pollock et al., 2012; Wang, Xu and Demirci, 2010). These are the tests that comply with the research aspiration because bringing these tests on mobile-enabled POC platform can benefit millions of people and provide ease of use and portability. For visualisation of various assays to perform colourimetric tests, a few examples are shown in Fig. 2. 1.

The motivation of this thesis is related to resource-limited settings and personalised healthcare. Lateral flow assays are well-suitable for homecare settings and are widely used for a home pregnancy test, where a paper-based inexpensive device can determine the existence of a target analyte in a liquid sample. The assay type is frequently referred

to as a dry-chemical based assay or a microfluidic assay. An example is shown in Fig. 2. 1(c), which can also be referred to as a Rapid Diagnostic Test (RDT) reader.

(a)   (b)   (c)
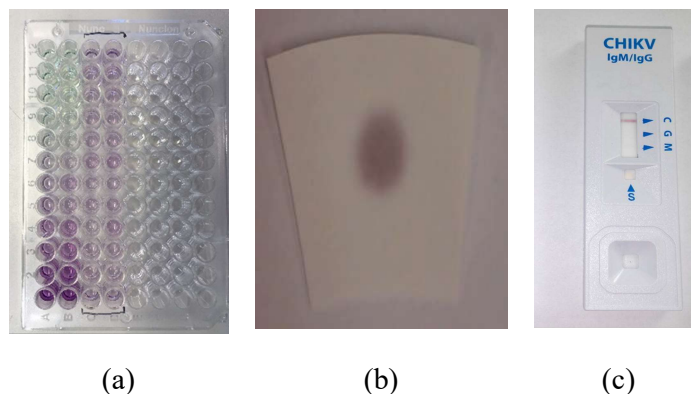
Fig. 2. 1: Example of colourimetric tests. (a) Bradford protein assay[2], (b) Reaction kinetics of alkaline phosphatase on paper[3], (c) Rapid diagnostic test reader for Chikungunya detection[4]

The RDT is an early, easy and fast diagnostic tool, which can promote effective monitoring of a chronic condition, assist in epidemic readiness and drug abuse (Banoo et al., 2006; Wongsrichanalai et al., 2007; Dinnes et al., 2007; Okeke et al., 2011; Cooper et al., 2012; Ozkan and Kayhan, 2016). It can be supervised by less experienced or trained medical personnel, but it can provide a precise diagnosis. However, the availability of the reader, know-how protocol, cost and device compatibility are some of the issues which require more progress. Thus, a computer-aided rapid test reader would be much more preferable.

For healthcare applications, the conventional system operate in the hospitals and diagnostic centres are currently bulky, expensive and expert-oriented. Therefore, this chapter does not include those conventional practices, instead focuses on the state-of-the-art systems and system components in the literature that matches with our research

---

[2] Courtesy: Dr Grisha Pirianov, Department of Biomedical and Forensic Science, Anglia Ruskin University, UK
[3] Courtesy: Dr Mohidus Samad Khan, Department of Chemical Engineering, Bangladesh University of Engineering and Technology, Bangladesh
[4] Courtesy: Dr Iqbal Kabir, Department of Epidemiology, National Institute of Preventive and Social Medicine, Bangladesh

aim. It should also be noted that colourimetric tests are well utilised for diverse applications beyond diagnosis; a broader context is exemplified in Table 2.1.

Table 2. 1: Potential field of application of intelligent colourimetric tests

Field	Application	Reference
Medical diagnosis and health status confirmation	Prostate Cancer	Koo, Wee and Trau (2016), Welch and Albertsen (2009), de la Rica and Stevens (2012)
	Urinalysis	Smith et al. (2016)
	Kidney diseases	Akraa et al. (2018)
	Diabetes	Arnett et al. (2016)
	Pregnancy and complications	Bu et al. (2018), Konnaiyan et al. (2017)
	HIV	de la Rica and Stevens (2012)
	Dengue	Matthews et al. (2012)
	Malaria	Wongsrichanalai et al. (2007)
Environmental monitoring and quality control	Mercury in water	Chen et al. (2016), Wei et al. (2014)
	Pesticides in water	Sicard et al. (2015)
	Water hardness	Bhattacharjee, Jiang and Behdad (2015)
	pH and nitrate in water	Lopez-Ruiz et al. (2014)
	Iron test in soil and water	Choodum, Sriprom and Wongniramaikul (2019)
	Indoor air quality	Qin et al. (2015)
Drug discovery	Structural relationship between material	Facchini et al. (2018)
Food industry	Meat quality	Magiati et al. (2019)
	Pasteurisation of milk	Yu et al. (2015)
	Antibiotics abuse in animals	Yan et al. (2018)
	Food allergen	Coskun, Ahmet F.Wong et al. (2012)
	Iron detection in white wine	Santos Neto et al. (2018)

	Bacteria in food	Zheng et al. (2018)
Forensic science and criminal investigation	Age of bloodstain	Shin et al. (2017)
	Driving under the influence of alcohol	Kim et al. (2017)
	Cocaine detection	Smith et al. (2014), Cooper et al. (2012)
Materials science	Arc-welding	Serrano et al. (2016)
Textile	Smart-textile	Promphet et al. (2019)

2.3 Colourimetric Assay Types

Conventionally, the assay types can be defined based on time point, number of analytes, signal amplification method, type of the substrate and format of the result. Regarding the detection method, this work considers the assay involving visible colour spectrum only. This thesis explores the assay types from a computer vision and machine learning perspective and how it relates to the nature of the assay process.

To the best of our knowledge, the existing works (mentioned in this chapter) on computer-aided colourimetric measurements consider only a single or set of specific applications for some particular domain (Karlsen, 2018; Yetisen, 2014). By computer-aided system, this thesis refers to full or partially automatic system utilising either a desktop, a server or a mobile application. This thesis aims to explore colourimetric test from a unified perspective, rather a single application (Table 2.1). Therefore, it is essential to understand that, within this domain, the assays can greatly vary in terms of purpose, materials, size, shape and colour. For a better visualisation of the context, this chapter mainly uses images from our case studies.

2.3.1 State of Matter

Based on the state of the matter, the colourimetric assays can be either dry or wet chemicals based. Both of these assays are well utilised in the reported articles (Alankus et al., 2018; Sicard et al., 2015; Lin and Scott, 2012b; Matthews et al., 2012). The state of matter is an important consideration because based on the critical review conducted in this thesis, the physical state of the sample can dictate the overall system regarding image processing algorithm and feature-set.

<div align="center">(a)　　　　　　　　(b)　　　　　　　　(c)</div>
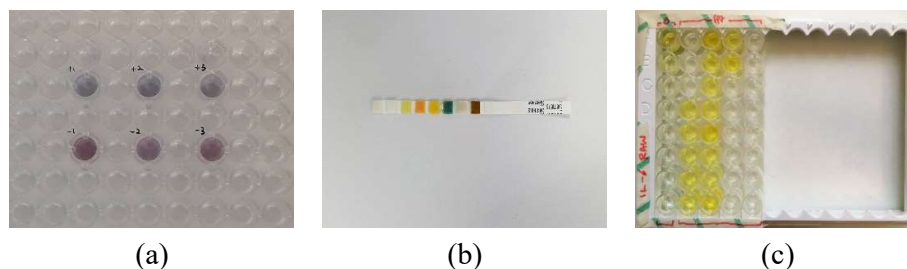
Fig. 2. 2: Example of assay types. (a) Plasmonic ELISA for TB-antigen specific antibody detection, (b) Urine dipstick, (c) ELISA for analysing pro-inflammatory protein production (IL-6)[5]

In Fig. 2. 2, (a) and (c) are wet-chemical-based ELISA images. It can be heavily affected by the ceiling light, which will act as noise during image processing. Sicasys Software GmbH (2017), a commercial mobile application on the quantitative colourimetric test, has highlighted this particular issue for wet-chemical-based assays. Conceptually, it can be well understood that liquid surface can easily be affected by surrounding light reflectance. Fig. 2. 2(b) is an image of a dry-chemical based lateral flow assay. Although the ambient condition should less affect the dry-chemical based assays, many of the recent works have utilised additional hardware attachments (Appendix B.1).

If the wet samples are placed in a transparent or translucent plate, the image of the assay would also have to deal with the shadow of the colours spread across the empty space of the plate. Sicasys Software GmbH (2017) referred to this issue as the smearing effect.

A wet-chemical-based assay can be subject to the shape of the container, i.e. well. In Fig. 2. 3(a), TB-test samples are placed in a plate, where the wells have a flat bottom. For better demonstration, natural food colours were placed in the same variety of assay plate in Fig. 2. 3(b). The samples in Fig. 2. 3(a) and (b) vary in viscosity. Intentionally, a minor variation was created within the food colouring as well. The variation is apparent from two rows of Fig. 2. 3(b).

---

[5] Courtesy: Charys Palmer and the Department of Biomedical and Forensic Science, Anglia Ruskin University.

(a)   (b)   (c)
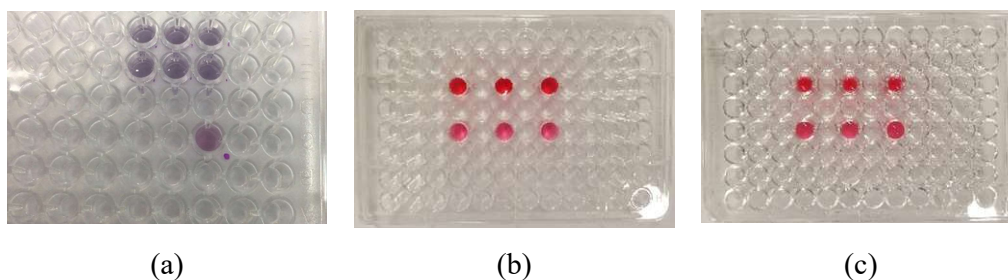
Fig. 2. 3: Variation in shape of the container. (a) TB-test in a flat plate, (b) Food colour in a flat[6] plate, (c) Food colour in a round plate[6]

On the other hand, by observing the images in the reported articles (Chen et al., 2016; Matthews et al., 2012) for various types of assay, this work assumes that, despite the materials used, the colour formation and brightness of the chroma component (Fig. 2. 4) of a sample can play a vital role in a computer-aided colourimetric decision.

(a)   (b)

(c)   (d)
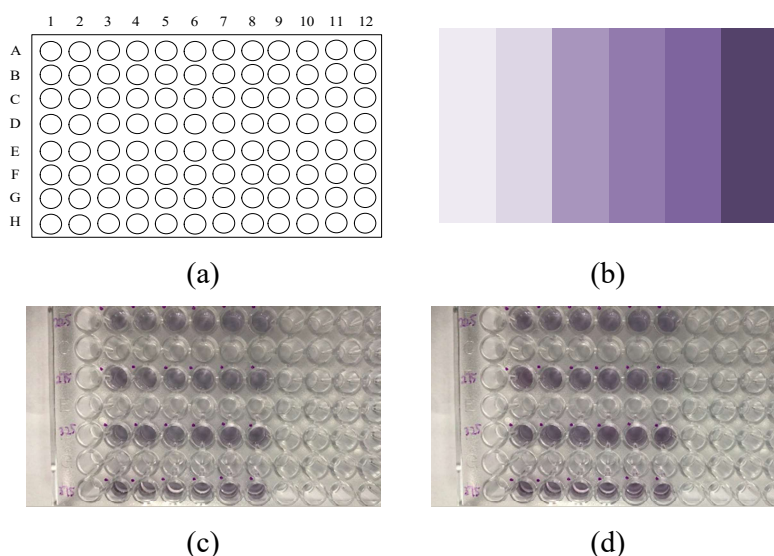
Fig. 2. 4: Visualisation of the impact of colour concentration. (a) Schematic diagram of an assay plate, (b) Variation of the brightness of the same colour or chroma, (c) Sample with light colour concentration, (d) Sample with a bright colour concentration

In both Fig. 2. 4(c) and (d), the concentration of hydrogen peroxide and gold ions were varied in A to H direction (see Fig. 2. 4(a)) of the assay plate. The difference between

---

[6] The images were generated as part of this thesis to visualise the impact of colour concentration and shape of the well.

Fig. 2. 4(c) and (d) is the brightness in colour or colour concentration. This research assumes that the issue of colour concentration can be more crucial for a computer-aided quantitative assay without any hardware attachment.

2.3.2 Colourimetric Result Evaluation Process

The conventional method defines the type of the result as the qualitative, semi-quantitative and quantitative assay, which are re-defined in this section for an intelligent system.

Binary Classification

A qualitative assay determines the presence or absence of a substance. The result obtained from a qualitative test would be in yes or no form. Therefore, it is a binary class classification problem. An example is shown in Fig. 2. 5(a).

(a)   (b)   (c)
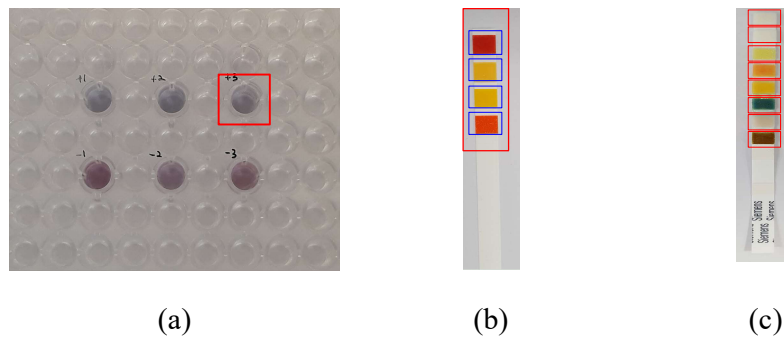
Fig. 2. 5: Object, target and sample in an image of a colourimetric test. (a) TB-test: image contains six filled wells. In other words, there are six samples. From the computer vision perspective, there are six objects. (b) pH indicator paper: four objects separated by blue boxes belong to the same pH test, (c) Multiplex assay: each object is directed to a different target

Although qualitative assays find widespread applications in the various domain, the most well-known application of this assay would be the pregnancy test strip to detect Human Chorionic Gonadotropin (HCG).

In theory, the two classes of a qualitative assay should use colour opponents. In this way, the assay would provide a naked-eye detection. For a person with no or less medical training, the conventional colourimetric result may be complicated to interpret, for example, the medical patch used by Matthews et al. (2012). Therefore, the use of

mobile-enabled dengue detection proposed by Matthews et al. can support non-medical personnel to understand the presence or absence of dengue virus from the binary classification provided by their system.

Multi-class Classification

The semi-quantitative assay is in between a qualitative and a quantitative test, yielding an approximation of the quantity or amount of a substance. The result can be compared with a colour chart by a naked-eye measurement. The clinicians use analysers such as CLINITEK Status+ Analyzer (Siemens Healthcare GmbH, 2018). This type of test can be seen as a multi-class classification problem as well as a regression problem. An example is shown in Fig. 2. 5(b).

Semi-quantitative assays have drawn a lot of attention for computer-aided colourimetric quantification (Kim et al., 2017a; Mutlu et al., 2017; Alankus et al., 2018; Solmaz et al., 2018). Mutlu et al. (2017) provided the semi-quantitative colourimetric quantification of fifteen pH levels (0-14) or classes.

Although the Quantofix Peroxide 100 is supposed to classify only six concentration levels of H_2O_2 (0, 1, 3, 10, 30 and 100 ppm), Solmaz et al. (2018) added five secondary concentration levels (0.5, 2, 6.5, 20 and 65 ppm) to increase the number of classes. In the absence of an elaborate description, we assume that the images of the test strips containing corresponding colours for these secondary classes can be considered as synthetic data, which may have reliability issue in a real-life scenario.

Research demonstrated in Kim et al. (2017) used paper-based assay to classify five standard concentrations of alcohol as well as nine classes by using intermediate enhanced concentration level. Due to the purpose and nature of the assay, the standard concentration is not an integer value, which is not a concern while performing classification, however, might be a point of interest for more elaborate research while using the intermediate values.

Similar to Kim et al. (2017), Alankus et al. (2018) also used class labels of fraction numbers to classify six concentration levels of phosphate in water. Nevertheless, the number of classes is related to the choice of the classifier and needs to be considered in the context of the curse of dimensionality (Bellman, 1961, 2003). Therefore, care

should be given while implementing machine learning algorithms for these colourimetric detections.

Regression

Quantitative assay helps in determining the amount of a substance, which can be seen as a pure regression problem such as mercury detection in water (Chen et al., 2016) and tracking HE4 biomarker to detect ovarian cancer (Lin and Scott, 2012b). The case studies of this thesis mainly focus on the qualitative and semi-quantitative colourimetric test. Therefore, while reviewing the literature in this chapter, more attention is provided towards the classification-based systems. Conceptually, the image processing framework used for a qualitative and semi-quantitative assay can also be used for a similar quantitative assay. An example of a quantitative assay is shown in Fig. 2. 2(c). In this case, instead of classification, the colourimetric decision has to be produced by regression due to the result-type.

2.3.3 Number of Objects per Sample

Based on the number of analytes, the assay can be single target or multiplex. However, this work considers the sample in image format. Therefore, it is essential for the decision making process to realise the number of objects associated with a single sample.

Single Object per Sample

In Fig. 2. 5, single sample is outlined with a red box, where the single object or well represents a single sample and a single test. In Fig. 2. 5(a), single sample implies sputum of an individual and single test indicates the plasmonic ELISA for TB-antigen specific antibody detection.

In the literature, most of the work considers the assay as a single object/ sample (Kim et al., 2017a; Solmaz et al., 2018). Despite being a single object/ sample kind of assay, an assay can hold multiple samples. For an image of a 96-well plate, there can be multiple samples (maximum 96 samples). The assay types are shown in Fig. 2. 5(a) and Fig. 2. 2(c). There are mobile applications available in the commercial app-stores which can act as a plate reader such as AssayColor (Alidans srl, 2015) and Spotxel® Reader (Sicasys Software GmbH, 2017b). Both of these mobile applications can

process multiple samples in an assay plate, treating each sample as a single object. However, these mobile applications utilise virtual plate to deal with a 96-well plate.

Multiple Objects per Sample

In Fig. 2. 5(b), multiple objects belong to a single sample as well as a single test, i.e. pH 3 and are collectively responsible for producing a decision. If these objects are to be detected individually (outlined in blue boxes), then the feature-set would require to be multiplied by the number of objects/ sample. Although the feature-set can be elongated, it is not logical to consider all four colour chambers as one object because it would contradict the purpose of the assay type. Multiple objects per sample type of assays are available in the literature as well (Matthews et al. 2012; Mutlu et al. 2017; Sicard et al. 2015).

Multi-test per Image and Single Object per Test

The multiplex assays are capable of performing multiple tests on a single sample (Devadhasan and Kim 2018; Lopez-Ruiz et al. 2014), e.g. urine of an individual. Therefore, each object, block or colour pad in Fig. 2. 5 (c) represents different test type such as glucose, pH, protein and ketone.

2.4 Computer-aided Colourimetric Test

The focus of this thesis is a computer-aided colourimetric test. For the convenience of describing different approaches taken in the reported articles, this work categorises the methods into the following four groups: systems with additional hardware attachments (Appendix B.1), proof-of-concept provided as a desktop application (Appendix B.2), mobile-enabled systems (Appendix B.3), and server-based approach (Appendix B.4). Each of these approaches has its own strengths and weaknesses.

By analysing the systems with hardware attachments in Appendix B.1, the following observations for colourimetric tests on POC platforms can be summarised.

The optomechanical attachments (Akraa et al., 2018; Yang et al., 2018; Kim et al., 2017a) for the personal devices present considerable advantages to provide a colourimetric test result. In some cases, the systems utilised small attachments which are capable of offering portability. However, regardless of being portable, in most cases, the additional attachments resulted in a bulky system.

18

Due to the position and size of these attachments (Appendix B.1), they can hold a compartment to place the sample and control, and components to facilitate minor heat-treatment. Many additional attributes of these optomechanical attachments are not directly associated with the colourimetric tests. However, these facilities could be beneficial for an extended research goal to develop an expert system which can benefit from fluorescence, spectroscopy, lens-free imaging and holography.

The optomechanical attachments often possess 3D-printed components (Chen et al., 2018a; Skandarajah et al., 2014; Ludwig et al., 2015). 3D-printing can offer precise measurement, and the cost of these printed components can be often downsized, however, it raises a concern regarding the accessibility of these printed components and the cost of a 3D-printer itself.

These optomechanical attachments (Appendix B.1) might be light-weight, but the convenience of attaching, as well as detaching, these components to and from the mobile devices by non-technical personnel and durability of these hardware components are yet be discovered.

The assays are often placed inside a box to protect the sample or biochemical reaction from the ambient air (Appendix B.1). Such attachments would be beyond the scope of this thesis. The biochemists or nanotechnologists may find a better solution in future to develop these assays for homecare solutions without requiring the samples to be placed in a box. With the advancement of label-free assays, it can be predicted that such assays would be more popular in the coming years. This thesis is more interested in investigating if the boxes along with the other hardware attachments are aiding the imaging conditions. For instance, the hardware attachments can channel the light, increase the brightness to ease the image processing mechanism and aid in noise cancellation so that the machine learning algorithms can perform well.

The images captured with the aid of these attachments (Appendix B.1) are supposed to be of high-quality, noise-free, correctly positioned and often in a magnified condition. It would be unfair to provide a direct comparison of the colourimetric images captured by a system without any hardware attachments. Mutlu et al. (2017) conducted a colourimetric classification to identify the pH level using a mobile phone with and without such an additional attachment. The result obtained by Mutlu et al. indicated that

it is possible to achieve equally good performance without any additional hardware attachments. However, no description was provided for the image processing algorithm with which Mutlu et al. achieved this performance. Therefore, it can be assumed that Mutlu et al. either manually cropped the image, or did not extensively consider the impact of the background other than the Region of Interest (ROI). Moreover, the study involved a paper-based assay. Hence, whether a good performance can be achieved for liquid samples without any optomechanical attachment or not— is yet to be explored.

A proof-of-concept for a computer-aided colourimetric test is often provided as a desktop application (Appendix B.2) rather on a mobile platform or any standalone device. Conceptually, if an algorithm is efficient enough, it can be developed as a mobile application. Our research suggests that, in such cases, the algorithm would have to comply with the limited capacity of the mobile devices.

Critically reviewing the mobile-enabled systems for colourimetric tests in Appendix B.3, the following observations were made.

Mobile-enabled systems possess enormous potential to provide a colourimetric decision. Studies (Appendix B.3) suggest that these systems can provide equal or even better performance than the conventional methods, which need to be validated by testing these claims on an ample amount of samples. It is quite common to repeat a scientific experiment for only three to five times for reliability. A computational system tested on such a small number of samples (Appendix B.3) while validating the system may not be a reliable system.

The literary works often take advantage of the available mobile applications to capture or process colours (Karlsen and Dong, 2017; Wu et al., 2015). On the other hand, for colourimetric tests, the use of a wide range of mobile phone camera resolutions is noticeable in the reported articles (3.2-20.7 megapixels). Shen, Hagen and Papautsky (2012) reported from the International Telecommunication Union that the majority of the mobile phone subscribers in 2011 possessed standard camera features of 3–8 megapixels. In the past few years, personalised devices including the mobile phone camera in terms of hardware and software has shown enormous growth; on the other hand, advanced mobile devices are now available at a cheaper price than in 2011.

The early research using low-resolution cameras have reported poor performance as compared to the recently published studies (Appendix B). However, the success attained in the recent articles could be due to better algorithms, feature selection and processing power, rather that only for the use of a high-resolution camera.

Studies have been conducted on various platforms including Android, iOS and cross-platform (Appendix B.3). Due to inter-disciplinary interest in the colourimetric test and prospect of commercialisation, the reported articles often did not disclose or provide any detailed discussion regarding the technicalities of these mobile-enabled schemes.

The reported articles (Appendix B) also often neglected to report or to evaluate the statistical performance in addition to accuracies, such as specificity and sensitivity. Moreover, the computational time was not taken into consideration by many.

The reported articles often utilise a server to process a colourimetric test result and send the result to others (Appendix B). Many researchers compensated for the drawback of limited storage, battery and processing capacity of the mobile devices by conducting partial or full processing in the server.

Data portability is one of the advantages of such a server-connected system because the test result can be easily shared among patients, their families and health professionals.

Reinforcement learning (Sutton and Barto, 2018) can be more easily integrated with server-based systems. Moreover, a server-based system can readily benefit from database and knowledge-base. The geo-tagging can also be evidently incorporated into such systems. In this way, a server-based system can offer more enhanced and robust performance and more conveniently trained for a new situation.

Health data could be sensitive information. A system with patient records requires protection from cyber-crimes. Hence, a server-based approach can be less secure and more expensive than the stand-alone approaches.

In 2012, a study (Matthews et al., 2012) showed that there is a trade-off between power consumption and computational complexity to perform a colourimetric test using a server and on mobile devices. The server-based approach can salvage power computation on the native devices by taking more time to process an image of a medical

patch for dengue detection. By comparing the computation time in a stand-alone as well as a server-based system, Kim et al. (2017a) reported the server-based system to be two times slower than the stand-alone system. The study (Kim et al., 2017a) involved paper-based test strips of the single object per sample for saliva alcohol detection on a mobile platform using an additional hardware attachment. Our research suggests that the computation time can be more affected for wet-chemical-based assays, an image containing multiple samples or if it is a multiplex assay and if any heavy image processing algorithm is integrated to compensate for the absence of the hardware attachment. Therefore, one needs to consider the aforementioned issues before entirely relying on a sever-based system.

2.5 Methods and Algorithms

Exploring different approaches for the computer-aided colourimetric test in Section 2.4, this section intends to provide an overview of the image processing and classification methods used in the literature for similar context.

2.5.1 Image Acquisition

Image acquisition is an essential step in an image-based system. The early stage and conventional studies often acquired images by a flatbed scanner (Suslick, Rakow and Sen, 2004; Mazzone et al., 2007, 2013; Zhong et al., 2018). Images acquired by such scanning involuntarily solves the issue with lighting condition and camera to sample position. However, flatbed scanning is not suitable for wet-chemical-based assays. This thesis assumes, the assays which can be analysed from the images acquired by flatbed scanners can also be analysed from the images captured by mobile phones.

The imaging condition can undoubtedly influence the image quality as well as colours. A wide range of smartphone camera specifications along with intrinsic and extrinsic parameters has been explored in the literature such as the International Organization for Standardization (ISO), flash and focus. Such explorations are exemplified in Table 2. 2.

Our research assumes, due to the advancement of hardware as well as software components of the phone camera, these parameters can be set to the automatic mode. For field operation, the systems would be exposed to the ambient illumination condition to capture images. Day by day in future, with an advancement of technology, high

powered cameras in the mobile devices will be more accessible, and the internal software of the phone camera will be more capable of adjusting the camera parameters according to the ambient condition.

Table 2. 2: Variability regarding imaging

Premise	Parameter	Specifics
Camera parameter	Resolution (MP)	Low: 3.2 (Cooper et al., 2012) High: 20.7 (Kim et al., 2017a)
	ISO	Varying from 50 (Alankus et al., 2018) to 800 (Karlsen 2018; Lopez-Ruiz et al., 2014). Auto (Karlsen 2018)
	Other parameters	Constant (Mutlu et al., 2017) Auto (Solmaz et al., 2018)
Camera to sample position	Distance	Low: 5cm (Yetisen et al., 2014) High: 2 feet (Feng et al., 2014)
	Exposure (Angle)	Parallel (Alankus et al., 2018) Tilted (Karlsen and Dong 2017)

The image acquisition techniques can influence the gap between training and testing datasets (Mutlu et al., 2017; Sicard et al., 2015; Wang et al., 2016b). Therefore, a computer-aided colourimetric test should aim for the requirement of minimum fine-tuning in case of such variations including the variations due to the change of platform.

2.5.2 Image Processing

This thesis refers to the term image processing as part of the data pre-processing using computer vision. Ideally, image processing would be comprised of image pre-processing, segmentation and post-processing. Not many works in the literature on colourimetric tests have reported to use automatic and intelligent image processing algorithms. Only a very few (Karlsen and Dong, 2017; Jonas et al., 2016; Lopez-Ruiz et al., 2014; Matthews et al., 2012) have utilised image segmentation techniques, such as Otsu's method (Otsu, 1979), Hough transformation (Fernandes and Oliveira, 2008), contour detection (Thevenaz and Unser, 2008) and custom made hybrid algorithms.

The task of image processing algorithm is often reduced or subsided as exemplified in Table 2. 3.

Table 2. 3: Alternatives to image segmentation

Application	Additional hardware	Full image	Referencing	Cropping	QR code	Virtual guideline	Reference
H_2O_2	No	No	No	Yes	No	No	Solmaz et al. (2018)
Urinalysis	Yes	Yes	No	No	No	No	Akraa et al. (2018)
	Yes	No	No	Yes	No	No	Rahmat et al. (2018)
pH test	No	No	No	Yes	No	No	Mutlu et al. (2017)
Saliva alcohol	Yes	No	No	Yes	No	No	Kim et al. (2017a)
Blood test	Yes	No	No	Yes	No	No	Kim et al. (2017c)
Water test	No	No	No	Yes	No	Yes	Sicard et al. (2015)
Wet-chemicals	Yes	No	No	No	No	Yes	Vashist et al. (2015)
Alkaline phosphate	Yes/No	No	Yes	No	No	No	Yu et al. (2015)
Blood test	Yes/No	No	No	No	Yes	Yes	Feng et al. (2014)

The reported articles often used colour calibration as part of the image processing algorithm (Akraa et al., 2018; Konnaiyan et al., 2017; Yetisen et al., 2014). The study performed by Yetisen et al. (2014) showed that the calibration requirement could cost a system to consume more time to produce a colourimetric decision.

The colourimetric test is not an auxiliary diagnosis, instead a confirmatory step for medical diagnosis. Thus, dependency on lighting environment condition and devices could be crucial for such cases. In the case of medical imaging, where the coloured

images are captured in ambient condition, the effect of illumination condition is studied elaborately. Within the context, a discussion is provided in our earlier study on tongue imaging (Tania, Lwin and Hossain, 2018). The doctoral research of Wang (2013) in a similar context showed the difference in the illumination condition can create a vast diversity in the imaging condition. Our research suggests that this diversity can result in the requirement of alteration in the image processing algorithm, feature-set, parameters of the classification and regression algorithm. One should expose the machine learning algorithm to a diverse dataset to make the system robust, but of course, cannot possibly train the system for all possible conditions. Therefore, in the absence of a universal colour correcting algorithm the aim should be to either include reinforcement learning so that the system can learn from new conditions or be retrained without compelling to fine-tune the algorithm.

Analysing the articles above on image-based colourimetric tests, it can be concluded that the computer-aided system can benefit from more advanced image processing algorithms, rather relying on hardware attachment, which will increase the cost of the system, limit the flexibility for the user and restrict the accessibility of the system. On the other hand, manual or guided cropping of the sample would require the user to possess some technical skills. The use of a virtual overlay on the assay can certainly simplify the image processing algorithm. However, these virtual plates with specific dimensions of the assay as well as sample positioning may restrict the user. In order to automate the complete image processing procedure, there is a scope of improvement to design a framework which is computationally efficient to be implemented on the mobile devices. Therefore, it is necessary to further explore the image processing algorithms used for other applications that deal with colours to evaluate the possible employment for colourimetric tests.

2.5.3 Colourimetric Classification and Decision

An image-based colourimetric detection algorithm would require to consider the relevant attributes to be analysed. For the systems incorporating machine learning algorithms, one would require to make certain considerations regarding the dataset as well. A quantitative colourimetric test will require to perform regression, rather classification.

Dataset and Sample Selection

A biochemical procedure is more conventionally repeated three times for reliability and reproducibility to avoid anomalies. A colourimetric sample could be expensive data and often requires a controlled environment; the experiments could be time-consuming as well. On the other hand, an intelligent system using machine learning techniques would require a larger number of data to train, test and validate the model. To the best of our knowledge, all relevant reported articles used small dataset (<1000), and for most cases, the number of the independent tests was even smaller (Table 2. 4).

Table 2. 4: Dataset enhancement

Reference	Image replication	Variable			
		Illumination	Device	Orientation	Region
Alankus et al. (2018)	N/a	N/a	N/a	N/a	50
Solmaz et al. (2018)	N/a	7	5	N/a	N/a
Mutlu et al. (2017)	N/a	3	N/a	6	N/a
Feng et al. (2014)	50	N/a	N/a	N/a	N/a
Lopez-Ruiz et al. (2014)	27	N/a	N/a	N/a	N/a

N/a: Not applicable

The articles above incorporated balanced datasets comprising equal samples from the associated classes. Hence, data balancing techniques such as Weiss (2004) were not observed. The aforementioned articles also resampled the data using imaging conditions, rather using any predominant resampling techniques as implemented in a number of disciplines (Estabrooks, Jo and Japkowicz, 2004).

Feature-set

Understandably, the most important features for a colourimetric test would be colours, which was reflected in the reported articles as well (Table 2. 5). An inadequate feature-set can lead towards under-fitting, whereas an elongated one would result in a higher

dimension of the feature-set which would cost the system in higher computation time and occupying more memory space of the personal devices.

Table 2. 5: Feature analysis methods

Attribute	Description	Reference
Colour moment	Mean	Rahmat et al. (2018); Mutlu et al. (2017)
	Multiple	Kim et al. (2017a)
Colour difference	Initial ~ end point	Wang et al. (2016); Vashist et al. (2015)
Colour space	RGB	Rahmat et al. (2018); Mutlu et al. (2017); Soni and Jha (2017); Wang et al. (2016b); Sicard et al. (2015)
	HSL	Akraa et al. (2018)
	HSV	Lopez-Ruiz et al. (2014)
	LAB	Alankus et al. (2018); Konnaiyan et al. (2017)
	Multiple	Kim et al. (2017a)
Grey value	Weighted mean	Khan and Garnier (2013)
	Green channel	Barbosa et al. (2015)
	Rule-based	Sicard et al. (2015)

Algorithms Used for Colourimetric Tests

In general, the conventional qualitative test should be presented using colour opponents or at least by visually distinct colours, whereas the semi-quantitative tests involve reference colours which can be represented by a colour chart. The concentration calculation of a quantitative test engages statistical calculation from a standard curve using the optical density. In other words, it is possible to obtain a reference colour or a colour-ground truth for a colourimetric test despite its result type, even if it applies to a specific case study only. Therefore, reported articles often utilised the colour difference (ΔE) and colour referencing as the sole or partial major component of a colourimetric

decision scheme (Alankus et al., 2018; Rahmat et al., 2018; Jonas et al., 2016; Sicard et al., 2015). The prevalence of ΔE calculation can also be perceived from Table 2. 6.

Table 2. 6: Colourimetric classification algorithms and their performances

Reference	Classes	ΔE	ML	Performance metrics	Accuracy (%)	Execution time (s)
Alankus et al. (2018)	6x4	Yes	No	Confusion matrix; detection accuracy; precision; recall; F1-score	76-100	Rapid
Rahmat et al. (2018)	10*	Yes	No	Accuracy	95.45	N/m
Kim et al. (2017a)	5,9	No	Yes	Cross validated accuracy, PPV, NPV	80-100	~9.3
Mutlu et al. (2017)	15	No	Yes	Accuracy; ROC curve	100	N/m
Wang et al. (2016)	13x6	Yes	Yes	Cross-validated accuracy	100	Rapid
Feng et al. (2014b)	2	No	Yes	k-1 cross validated accuracy, correct vs incorrect	100	8

ΔE: Colour difference; ML: Machine learning; N/m: Not mentioned; PPV: Positive predictive value; NPV: Negative predictive value; ROC: Receiver operating characteristics s: second
*Each of the classes can be further categorised into different concentration levels or subclasses.

The conventional semi-quantitative test via RDT often offers a colour chart, for example, the colour chart for urine dipstick. Rahmat et al. (2018) provided the colourimetric measurement using only ΔE calculation. On the other hand, Vashist et al. (2015) provided the colourimetric decision by plotting the pixel intensity against the concentration of the analyte. Konnaiyan et al. (2017) utilised correlation graph. The use of PCA is also quite prevailing (Akraa et al., 2018). Jonas et al. (2016) performed a colour comparison from its ratio. Garg et al. (2014) utilised a calibration curve to determine the concentration of glucose. Although a colour chart was available for the dipstick used by Shen, Hagen and Papautsky (2012), the study used a calibration curve

to measure the pH level. A binary decision for substance tracing in Smith et al. (2014) was availed from the known cocaine standards. However, our research was more focused to study the prevalence of machine learning, especially classification algorithms to perform colourimetric tests (Table 2. 6). Therefore, in this section, more attention is provided towards the performance of the classifiers.

Support Vector Machines (SVM) is one of the most widespread classifiers applicable to various disciplines. Feng et al. (2014) used SVM (Burges, 1998; Noble, 2006; Theodoridis, Sergios; Koutroumbas, 2009) to provide a qualitative assessment of RDT based HIV detection. SVM is capable of performing both classification and regression. While quantifying the PSA concentration, instead of SVM Feng et al. (2014) opted for the predetermined calibration curve approach.

From the ΔRGB, Wang et al. (2016) used Linear Discriminant Analysis (LDA) to perform a qualitative analysis of catechols and partial least squares for quantification of the concentration. Inspired by Manuel A. Palacios et al. (2007), Wang et al. (2016) also used Hierarchical Clustering Analysis (HCA) using SPSS to cluster the catechols based on their similarity. Later, Wang et al. attained good discrimination among the catechols through LDA. The performance of LDA was validated using leave-one-out cross-validation. Wang et al. claimed 100% accuracy from concurrent use of PCA, HCA and LDA. However, the system needed to be exposed to an adequate number of dataset, rather using only five samples per class.

Similar to Wang et al. (2016), Kim et al. (2017a) also used LDA as one of the classification algorithms to classify saliva alcohol concentration. Despite using a hardware attachment, Kim et al. (2017a) used relatively elongated feature-set and three different classifiers simultaneously, namely LDA, SVM and Artificial Neural Network (ANN) to provide a mobile-enabled colourimetric decision. Using five independent datasets to train LDA, Kim et al. (2017a) aimed to reduce the variances for within-class alcohol concentration, while maximising the variances for between-classes. Based on our understanding, Kim et al. used 'svmtrain' (MATLAB, 2013) in the MATLAB platform and LIBSVM (Chang and Lin, 2011) in the mobile platform. ANN was utilised using backpropagation. Based on the user input, the proposed system of Kim et al. used either of these three classifiers to provide the classification. Among these classifiers, discussed the enhanced capabilities of ANN, especially in LAB colour space. For the

enhanced concentration levels, Kim et al. reported SVM to consume longer time than ANN, typically which should not be the case (Zanaty, 2012).

In the absence of detail of the parameters, we assume Kim et al. (2017a) utilised the standard parameters of NN toolbox and SVM. Hence, the observations by Zanaty (2012) should apply to Kim et al. as well, which is visible from the performance on the benchmark dataset in Kim et al. (2017a). While comparing the system using the MLR datasets (benchmark dataset), Kim et al. (2017a) found SVM to be faster, while ANN to provide higher accuracy, which is more plausible. Using the benchmark datasets, a detailed comparison was also provided for different mobile devices (Kim et al., 2017a). However, further investigation is required regarding the poor performance for the secondary dataset containing enhanced alcohol concentrations, which may have occurred due to the mechanism of class enhancement. A context regarding class enhancement is provided in Weiss (2004).

Least Squares Support Vector Machine (LS-SVM) is found to be one of the strongest classifiers to perform colourimetric tests (Solmaz et al., 2018; Mutlu et al., 2017). SVM possess inequality constraints and encounters quadratic programming, while LS-SVM has equality constraints. LS-SVM (Suykens and Vandewalle, 1999) was created to harness the advantages of SVM and at the same time simplifying it by using a set of linear equations. Unlike convex optimisation in SVM, the use of robust statistics aids weighted LS-SVM to present more robustness. Acknowledging the shortcoming of LS-SVM regarding sparseness, Suykens et al. (2002) suggested to use pruning techniques.

The conventional MLPs learn from the pattern of input and output. The parallel network architecture can have multiple inputs and outputs. Additionally, it uses a universal approximation of continuous nonlinear functions. Although MLPs can often outperform SVM, it has the disadvantage of hosting many local minima. The requirement of a number of neurons for a given task can also end up in the list of disadvantages of MLP. Similar to MLP, LS-SVM can take advantage of Bayesian interference and regularisation. Mutlu et al. (2017) achieved 100% accuracy by feeding the LS-SVM classifier with the mean RGB values of the universal pH indicator papers.

Using both LS-SVM and RF, Solmaz et al. (2018) achieved 95% accuracy via cross-validation to classify peroxide strips. For both primary and secondary concentration

levels, LS-SVM outperformed RF. We can perceive the performance from the correlation of Random Forest (RF) and rarity (Weiss, 2004).

For the primary concentration levels, Solmaz et al. (2018) reported LS-SVM to be independent of the feature-set. Based on the critical examination of the features and classifier, we anticipated a better performance from the overall system of Solmaz et al. (2018). Without re-implementing the system or critically exploring the dataset, it is difficult to conclude why their system degraded the performance from its initial cross-validated classification accuracy of 99%. Moreover, the degradation continued when the system was deployed on the mobile platform.

Based on the classification accuracy, required number of samples, model complexity, computation time, generalisation, trade-offs and performance of the classifiers above on the images of colourimetric tests, one must explore LS-SVM to design a computer-aided colourimetric scheme. However, similar to any other machine learning algorithms, LS-SVM is also not free from weaknesses. LS-SVM can struggle with robustness, input selection, on-line learning and large-scale problems as identified by its creator (Suykens, 2002). Therefore, it is important to validate the performance of LS-SVM to provide a colourimetric decision using a number of performance metrics and statistical analysis.

2.6 Evaluation Criteria

The WHO popularised the ASSURED criteria (Peeling et al., 2006) as an evaluation criteria to detect sexually transmitted infections. Originally, it was not directed towards computer-aided diagnostic tools on POC platform. While reviewing the paper-based microfluidic assays, Yetisen, Akram and Lowe (2013) stressed on the ASSURED criteria for mobile-enabled systems. The doctoral research of Karlsen (2018) was a smart-diaper that fulfils ASSURED criteria. In this section, the ASSURED criteria are explored from the computational context for image-based colourimetric tests.

2.6.1 Affordable Technology

The mobile phones have a high penetration rate (GSMA Intelligence), making it widely accessible and affordable technology to resource-limited settings. By 2019, the number of mobile phone users is expected to reach 5.07 billion (Statista, 2015). From the

computational context, the use of a mobile phone can act as an affordable-ASSURED technology.

Recently, the use of smartphones for chemical detection and analysis, and medical testing has gained popularity. Mobile-enabled sensors and biosensors have also made enormous advancements. In general, paper-based assays such as Sicard et al. (2015) are more affordable and suitable for less trained personnel. A mobile-enabled paper-based assay can enhance the processing of a result (Roda et al., 2016; Lopez-Ruiz et al., 2014), ease the effort to interpret the result and make the result conveniently communicable (Sicard et al., 2015). The objective of such a system is not aimed to replace the biochemical systems but instead to assist (Kim et al., 2017a), simplify (Lopez-Ruiz et al., 2014) or accelerate (Jonas et al., 2016) the process. For example, when it is challenging to provide visually distinguishable colours, such systems can aid to provide a better decision.

Modern smartphones are fully automated and equipped with high-resolution cameras, powerful processors with large storage capacity, wireless connectivity, real-time geo-tagging, secure data management, and cloud computing. These capabilities allow smartphones to be used as small, conveniently carried portable devices for convenient on-site diagnostic/chemical analysis and testing with rapid reproduction of results. Thus, smartphones have proven to be an attractive alternative for the on-site collection, imaging, and analysis of data. However, Zhang and Liu (2016) argued to minimise the requirements for additional optical attachments. No additional hardware requirement means there is no cost associated with the test reader. In this way, the mobile devices including phones can more effectively eliminate the operating cost by minimising the requirement of plate readers and analysers. Therefore, there is a need for a system which is computationally efficient to be deployed on the mobile platform, making it an affordable system.

2.6.2 Specific and Sensitive Performance

The next ASSURED criteria are specificity and sensitivity, which require the system to have low false negatives and false positives. Although, it is a common practice for the computation systems to present the result in the form of accuracy (Table 2. 6), evaluating the performance of the system only with accuracy can be misleading.

2.6.3 User-friendly System

The ASSURED criteria emphasise the minimum requirement of training from the users. The ratio between health professionals and patients are imbalanced worldwide. The global understaffed health systems can benefit from technologies that provide ease of use. These easy-to-use systems can support the associated need of growing elderly population, provide more autonomy to users for personalised healthcare at home settings, and more importantly in remote locations where trained medical personnel is scarce.

Exploring the existing literature, this research suggests, the user-friendly system should not only require less medical training but also should not demand high technical skills from the user. Therefore, the system should require minimum user interventions with the system in order to provide a decision. In the reported articles on colourimetric tests, some systems require users to assist with the data pre-processing techniques, e.g. cropping and seed points by the user (Rahmat et al., 2018; Mutlu et al., 2017; Solmaz et al., 2018). No detail description was provided in the article (Mutlu et al., 2017) regarding the cropping mechanism. Therefore, we assume that the cropping was performed manually without any intelligent image processing method. Solmaz et al. (2018) and Mutlu et al. (2017) utilised smartphones, whereas Rahmat et al. (2018) used a scanner for image acquisition. The cropping techniques used in these articles eliminated the segmentation process. However, the technique compromised the ease of use, compelling the user to possess technical skills.

2.6.4 Rapid and Robust System

The next criteria are to make the system rapid and robust. This work mainly focuses on the machine learning based computational systems. To evaluate the rapidness, the training time should be taken into account so that the system can support in time of any sudden outbreak such as climate change-related diseases (Kabir, Rahman and Milton, 2014). However, the testing time should be considered as the execution time.

On the other hand, robustness is a vast term and has a direct impact on the ease of use of the system. To provide reproducible and accurate results, if there are many rigid guidelines for the user to follow, then the system preserves lower robustness as well as

reduced user-friendliness. In this case, the robustness could be regarding the format of the data or the environmental settings.

The lighting condition is one of the biggest concerns for image processing, especially medical imaging. The ambient light can have a considerable impact on the RGB value. Therefore, lighting condition is considered as a critical factor of the robustness (Solmaz, 2018; H. Kim, Awofeso, Choi, Jung, & Bae, 2017; Mutlu et al., 2017). Solmaz et al. (2018) considered seven illumination conditions to train the model. The light condition can be channelled through additional optomechanical attachments such as H. Kim et al., 2017. The algorithm proposed by Mutlu et al. (2017) for the classification of pH test strips showed equal performance with and without such hardware attachments. The study (Mutlu et al., 2017) considered three different lighting conditions while experimenting without any hardware.

The camera to sample position can influence the image processing requirements as well as the classification performance. The mobile applications available in the app-stores utilise a guideline or virtual plate to limit the location of the sample position as well as the distance of the camera (Sicasys Software GmbH, 2017b; Enzo Life Sciences inc., 2015; Alidans srl, 2015). Mutlu et al. (2017) utilised six different orientations of the pH test strips (Merck, Germany), effectively varying the position of the sample from the camera lens as well as from the source of the light. The purpose of the variety in the rotation was to train the system for robust orientation of the sample, resulting in robustness within the dataset. The images were then manually rotated before training to maintain the same alignment. However, the system did not include any automatic image processing technique, instead relied on the user to crop the image.

Another parameter for robustness is interoperability (e.g. hardware compatibility, application programming interface or API). A mobile-enabled algorithm was tested on the paper strips to present a system with inter-phone repeatability by Yetisen, Martinez-Hurtado, Garcia-Melendrez, da Cruz Vasconcellos, & Lowe (2014). The downside of the system was longer calibration time. The performance of another smartphone application for paper-based saliva-alcohol testing was evaluated by Kim et al. (2017a) on five different smartphones, effectively varying the hardware and software components including phone-camera. However, the system involves additional hardware for illumination and imaging consistency.

Therefore, the criteria of robustness might be immeasurable, because how much autonomy a single system can provide without compromising the rest of the parameters of the ASSURED criteria is an optimisation issue.

2.6.5 Equipment-free System

The next ASSURED criterion is to make the diagnostic system equipment-free. This criterion is closely associated with the cost of the system. If a system can turn the personal device such as a mobile phone and tablet into a colourimetric test reader, necessarily there is no need for additional equipment. The standard practice for colourimetric tests involves plate reader, e.g. 96-well plate reader utilising light absorbance for wet-chemical-based quantitative tests, analysers for paper-based test strips and naked-eye tests for qualitative tests. In the literature, substantial amount of the reported articles on mobile-enabled systems utilised additional hardware attachment to channel the lights, enhance the image or ease the image processing technique (Kim et al., 2017a; Hussain et al., 2017; Kim et al., 2017c; Masawat, Harfield and Namwong, 2015). The 'additional' aid to make the smartphone act as a reader can be an obstacle to user-friendliness. Moreover, in some cases, these additional attachments are 3D printed, which is not available anytime, anywhere.

2.6.6 Deliverable System

The last criterion is that the system should be deliverable implying it should be accessible by those who will use the system. There is a close relation between affordable and deliverable systems. According to the World Bank, more households are likely to possess mobile phones than toilets (World Bank Group, 2016). Therefore, a computationally efficient system to be deployable as a mobile-enabled system is supposed to be more accessible than the specialised devices.

With the advancement of cloud computing, Yetisen et al. (2014), along with many, favoured web-based mobile applications. Undoubtedly such web or cloud-based applications would help with the dispute about limited storage, power and processing capacity of mobile devices. However, regarding accessibility (e.g. network connectivity, mobile device's hardware and native features) and data security, the native application without cloud dependence would be preferable.

On the other hand, a system compatible to be deployed using a server-based approach can make the system more suitable for telemedicine facilities. Therefore, a computational system should be efficient to be deployed on the mobile device but at the same time needs to be compatible with a server-based approach.

2.7 Justification of Research

The colourimetric analysis has been used for diversified fields for years. New methods and materials are continually being added to the current global research to make the tests more feasible on the POC platform. Unlike the aforementioned reported articles, this chapter critically reviewed the image-based colourimetric tests from the broadest possible latitude and entirely from the perspective of computer vision and intelligent systems.

Due to the vastly different approaches taken in the literature, it was essential to set evaluation criteria to harmonise with our research goal as well as to establish a fair comparison, which justifies the use of ASSURED criteria. Using this ASSURED criteria, it is evident from the above discussion that no existing study has explored the colourimetric tests from computer vision with a broader lens regarding size, shape, state of matter, result type and purpose, and for its diverse capability. Therefore, this work aims to explore and provide a proof-of-concept as to how a computer-aided system can achieve this goal.

The colourimetric test can be rapid as well as a prolonged process. To the best of our knowledge, no existing work in the literature considers the dynamic changes of colours throughout the chemical reaction of an end point assay. Few works have considered only the initial and final phase of the reaction. Understanding the overall biochemical reaction through computer vision by tracking the dynamic changes in the image can help us to understand more details about the data through the eyes of machines.

The state-of-the-art approach for POC based colourimetric tests involves optomechanical attachments using a 3D printer, which constrains the applicability to anywhere anytime. Due to the socio-economic context, the motivation of this research is to investigate a high-performing image-based colourimetric test using as few resources as possible.

The predominant use of hardware attachments and restricted ease of use in the literature led to incompetency in complete automation. Considering inter-disciplinary interest on the colourimetric test, there is a gap in the literature in the form of a critical exploration of a complete and automatic image processing framework. Firstly, there is a requirement for an image processing framework which will not require any support from additional hardware to channel the light to reduce the effect of ambience, magnify the image with a lens, improve the brightness of samples using LED or rely on a dark room. The image processing framework should also require minimum user intervention.

Few other works in the literature require calibration. The long computation time for calibration for test type detection is time-consuming. The computation time and computation complexity need to be critically considered with respect to accuracy.

The challenge of this work also includes one robust solution to fit into diversified colourimetric problems. The literature reflected the need for parallel observation of wet and dry samples and scrutinisation of the similarity and dissimilarity between diverse ranges of assays. This inspection will support future researchers in developing a universal or adaptive model to provide any colourimetric decision. Therefore, it is essential to understand the images of the assay in terms of image processing and to optimise the number of features to develop a computationally efficient algorithm.

Advanced visual recognition algorithms such as deep learning are capable of dealing with lighting condition, the distance between camera and sample, the angle of the camera, and the orientation of sample, making a system suitable for robust environments. Therefore, deep learning can be a better choice for image processing and assay detection for a colourimetric test. On the other hand, deep learning can come across as a heavy algorithm to classify colours for a colourimetric test. Moreover, scale in terms of dataset, training time and computational device, is an issue for deep learning. Hence, deep learning clearly contradicts the ASSURED criteria. However, an innovative approach may liberate deep learning to be used for a colourimetric test while fulfilling ASSURED criteria. To the best of our knowledge, there is no existing work on colourimetric tests using deep learning.

Unlike the image processing algorithms and deep learning, the reported articles explored a number of classification algorithms to perform colourimetric tests. There is a clear indication in the literature regarding the high performing classification algorithm for binary and multi-class classification of the colourimetric test. However, due to inadequacy in sample selection, lack of depth in feature exploration and optimisation, and consideration of only a limited scope of the colourimetric test, there is the need for an elaborate and systematic investigation of the colourimetric test using in-contrast case studies.

A prerequisite of assigning deep learning or Traditional Machine Learning (TML) for the colourimetric test would be an adequate amount of data. The reported articles failed to generate a significant amount of data and also precluded its availability from any open access database, which could have gradually turned into a sizeable dataset. The lack of independent tests and incredulous repeatability might be due to the primacy of stable assays, however in such cases exercising data resampling may offer a solution for the data-hungry approaches. At the same time, a study should also consider variability and rarity due to the nature of the application such as matrix effect and sensitivity to environmental factors. Hence, it is essential to explore different aspects of a computational system to deliver a colourimetric decision.

Our research goal includes a system that is intelligent to make it automatic and user intervention free; computationally efficient to be deployed in the mobile environment and can be integrated to the cloud as well; adaptable to comply with variability; rapid, and the specificity and sensitivity should be balanced and high to produce a reliable accuracy. Therefore, this work aims to investigate a colourimetric system that fulfils ASSURED criteria.

2.8 Summary

This chapter presented an interpretation of the colourimetric test from the eyes of machines and then provided a categorisation of necessary components. The state-of-the-art image-based colourimetric systems were explored in this chapter for each of these key components, such as data pre-processing and classification algorithms. This chapter also examined the recently reported articles using ASSURED evaluation criteria. Based on the research gap and prospect of the image-based colourimetric test,

this chapter reflected the rationale behind our research goal, which was finally described in a concise form in Section 2.7.

(This page has been left intentionally blank.)

Chapter 3

Research Design and Methodology

3.1 Introduction

This chapter presents an overview of the system framework to perform a colourimetric test using image analysis. At first, the interpretive operations of the problem are defined, followed by the key elements of the concept. The methodology developed in this chapter provides a unique view to explore the various stages of the colourimetric reaction of end point assays using two case studies. Using the case studies, the system can either reflect upon a collection of a number of phases of the colourimetric reaction or only the final stage of the reaction. This chapter presents a novel approach to performing the colourimetric test considering the continuous colour change during the chemical reaction. Alternatively, in consideration of the nature of the assays under investigation, i.e. end point assays, the concept exploits the strength of both TML and deep learning focusing on only the final state of the reaction; the key elements for this approach include the methods for image processing, feature extraction and classification. The evaluation criteria for the overall system is also discussed in this chapter.

3.2 Research Design

This research aims to explore how to design an automatic colourimetric classification scheme to function without a user intervention and any additional hardware attachment, suitable for personalised healthcare and resource-limited settings. In order to offer a solution to this problem, a computer-aided system is required to provide a decision based on the user's inputs (Fig. 3. 1).
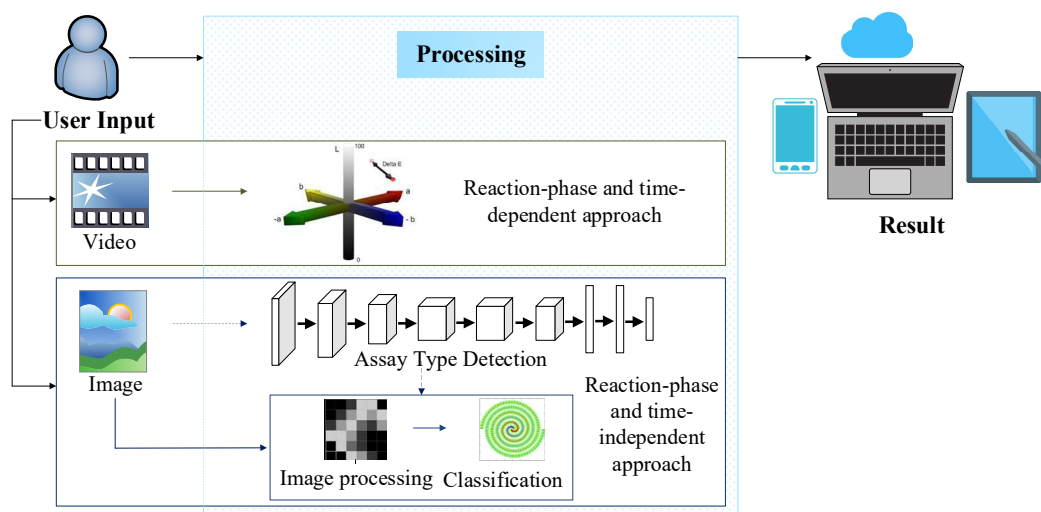
Fig. 3. 1: Research design considering dynamic and static changes in the image

This work aims to consider the diverse category of inputs, which is one of the key aspects of our research design. Based on the phase of the biochemical reaction, the problem can be depicted under two different scenarios:

1) Reaction phase and time-dependent approach considering the initial, key intermediates and final stage of the colourimetric reactions utilising video frames. The context is allied with the first research question, hence the first specific objective of the thesis.

2) Reaction phase and time-independent approach utilising the image of the final state of the colourimetric test. This context is associated with the second research question, which is to be addressed by the second, third and fourth specific objectives of the thesis.

   This approach or context can be further sub-categorised into the following:

- Using traditional machine learning techniques
- Using the pre-trained models of state-of-the-art deep learning techniques

Based on the type of input, the problem can be categorised from the computer vision perspective into the following two categories.

i) Input containing a single object for each sample, however multiple samples per image. Example: ELISA (Case Study 1)

ii)   Input containing multiple objects for each sample to classify any individual classes. Example: LFA (Case Study 2).

3.3 Data Preparation

The experiments on the case studies, mentioned in Section 3.2, were conducted in controlled environments, one in the UK and the other in Malaysia. This work does not involve any direct contact with human participants. The term dataset in this work mainly implies images and video frames converted as images. The dataset contains images in Joint Photographic Experts Group (JPEG) format.

The image capturing method comprises as diverse scenarios as possible (e.g. use of the different camera, lighting). Therefore, the image capturing, data preparation and sample selection method is elaborately discussed in the experimental chapters for each of the case studies. However, an example of data preparation at the laboratory environment is shown in Fig. 3. 2.

(a)   (b)
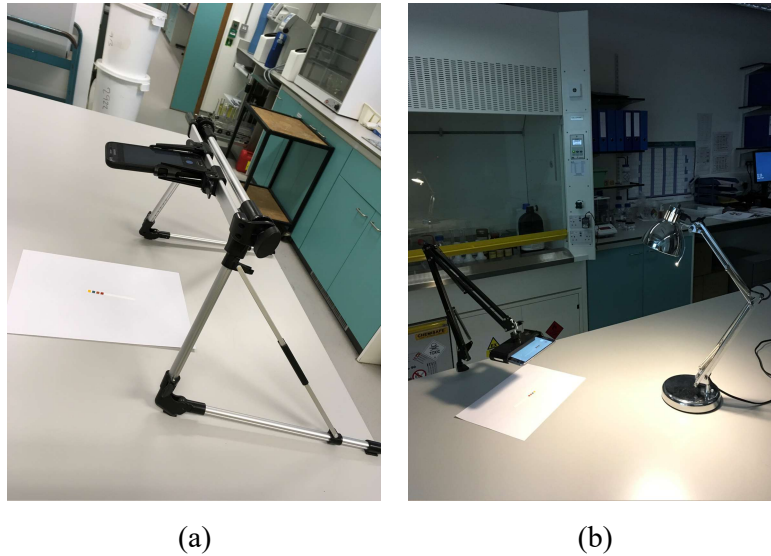
Fig. 3. 2: Example of laboratory environments. (a) Image being captured in a typical setting of laboratory light, (b) Image being captured in semi-controlled lighting environment

The use of stand shown in Fig. 3. 2, is to maintain a consistent and measured height between the camera and the sample. This can also aid in reducing the effect of shaky hands which may result in blurred images. However, it is not a mandatory procedure for the system to capture images using a mobile-stand.

The detail of instrumentation, i.e. the choice of platforms is provided in Appendix D.

3.4 Reaction Phase and Time-dependent Approach

A biochemical reaction of a colourimetric test often takes time to produce or transform colours. As mentioned in Section 3.2, considering the overall biochemical reaction, different states of a reaction can be recorded using personal devices. Due to the rapid nature of the lateral flow assays, this approach is more suitable for our first case study with wet-chemical-based ELISA, where meaningful information can be extracted from the different phases of the reaction.

In this work, the system input in the form of video, consisting initial to the final stage of the reaction is converted to images, i.e. frames taken from the video. To the best of our knowledge, this is the first attempt to provide a colourimetric classification based on the reaction phases of a wet-chemical ELISA, already published as Tania et al. (2017).

During the complete reaction taking t seconds, the instances (S) of the reaction is divided in t_i time interval for video sampling.

$$S= \{1, 2\ldots N\} \tag{3. 1}, \text{(Tania et al., 2017)}$$

where $S \in \mathbf{Z}^+$, $\mathbf{Z}^+$ denotes the set of positive integers.

For chromatic analysis, the colour difference is measured by delta E (ΔE), to be specific CIE76. With known colour space coordinates, the CIE 1976 formula delivers the colour difference. It was the first formula to provide ΔE in the LAB colour space. In LAB, if (L_1^*, a_1^*, b_1^*) and (L_2^*, a_2^*, b_2^*) are two colour coordinates at t_1 and t_2 seconds respectively, then colour difference is given by the following formulae (International Organization for Standardization, 2007).

$$\Delta E_{ab}^* = \sqrt{\{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2\}} \tag{3. 2}$$

Later, a high pass filter is applied on the calculated colour difference from Eq. 3.2 for each instance, which can be expressed as Eq. 3.3 (Tania et al., 2017).

$$\Delta E (S)_{ab}^* > JND \text{ , where } \Delta E (S)_{ab}^* = \sqrt{(\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2})} \tag{3. 3}$$

44

An important parameter for ΔE calculation is the just-noticeable difference (JND) or differential threshold. According to experimental psychology, "*it is the amount something must be changed for a difference to be noticeable, detectable at least half the time.*" As an intuitive value, some suggested ΔE to be 1.0, but a widely acceptable value is 2.3 (Mahy, Van Eycken and Oosterlinck, 1994; Sharma, 2003). A good example regarding the use of JND in computer vision is provided in Watson (1993).

The maximum amount of colour change at any instant can be calculated as Eq. 3.4 (Tania et al., 2017).

$$\alpha = \max(\Delta E(S)_{ab}^*) \qquad\qquad (3.\ 4)$$

Then finally, the cumulative sum of the colour difference after each t_i interval needs to be calculated. In order to calculate the total colour difference at any given instance S, the finite series of triangular numbers (Abramowitz and Stegun, 1972) can be utilised as Eq. 3.5.

$$\Delta E_S = \sum_{k=1}^{S} \Delta E_{ab\,k}^* = \Delta E_{ab\,1}^* + \Delta E_{ab\,2}^* + \Delta E_{ab\,3}^* + \cdots + \Delta E_{ab\,S}^* \qquad (3.\ 5)$$

$$= \frac{\Delta E_{ab\,S}^*\left(\Delta E_{ab\,S}^* + 1\right)}{2}$$

One can also find out the binomial coefficient as $\binom{S+1}{2}$ from Eq. 3.5, which represents the number of distinct pairs that can be selected from S + 1 instances.

In the case of colour transformation based studies, this work assumes, the number of the transition phase (ϕ), obtained from Eq. 3.5, can be represented as Eq. 3.6 (Tania et al., 2017).

$$\phi = \{0, 1, 2,\ldots, N\} \qquad\qquad (3.\ 6)$$

where $\phi \in Z^*$, $Z^* = \{0\} \cup \mathbf{Z}^+$. $\mathbf{Z}^+$ denotes the set of positive integers.

This thesis utilised JND as a threshold value. The captured colour changes in Eq. 3.2 is powerful enough to trace even a small change in the colourimetric reaction. However, the aim of the reaction phase and time-dependent approach is to track ϕ (Eq. 3.6). Therefore, this thesis assumes that it is not necessary to consider the changes below JND for Eq. 3.5 to identify ϕ smoothly.

Considering the complete reaction process, this approach can compensate for the following factors, which needs to be validated by the experiments.

- Image processing of low concentration wet-chemicals
- Provide classification without utilising machine learning techniques
- Optimising the reaction time

3.5 Reaction Phase and Time-independent Approach

The case studies utilised in the thesis are end point assays. Therefore, considering the final state of the colourimetric test, a reaction phase and time independent approach is investigated utilising TML as well as deep learning techniques.

3.5.1 System Framework

In this thesis, the computer-aided system for the colourimetric test includes both TML and Convolutional Neural Network (CNN) based deep learning. A proof-of-concept is delivered in the desktop environment using two case studies, as mentioned in Section 3.2. The system can be trained offline in the desktop environment as well as in the cloud. The trained system can be deployed as a desktop, mobile or web-based application. An overview of the workflow can be perceived from the system architecture illustrated in Fig. 3. 3.
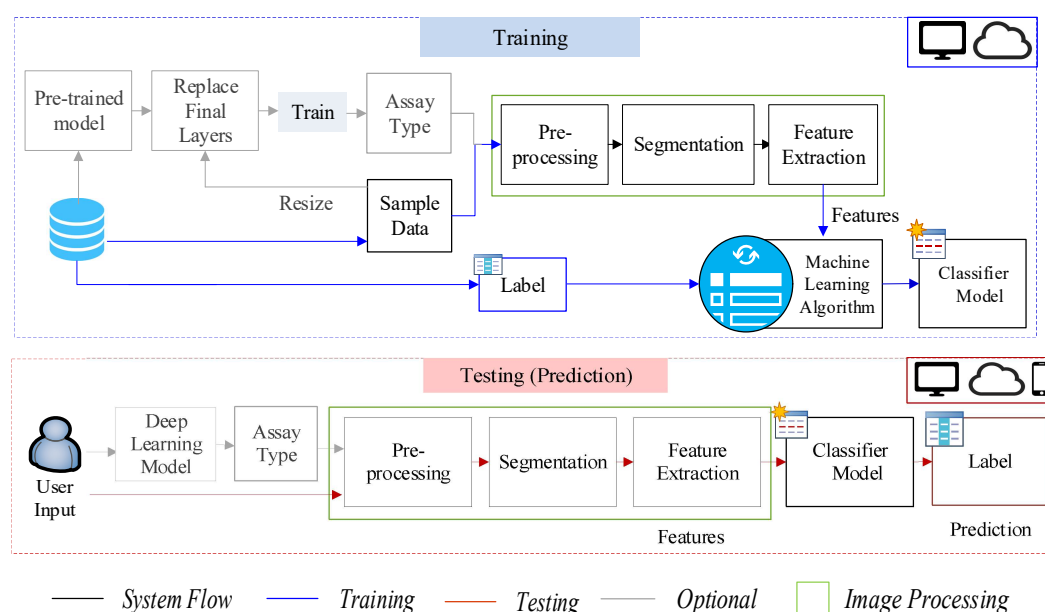
Fig. 3. 3: Reaction phase and time-independent framework for the proposed system

3.5.2 Deep Learning

In this work, the state-of-the-art pre-trained models (ImageNet, 2016) for visual recognition problems are investigated to recognise the assay type utilising inductive transfer (Karpathy, 2018; Soria Olivas, 2010). When there is a substantial amount of large data and similarity between the task is also high, the pre-trained models are supposed to exhibit good performance. A more detailed concept is described in Goodfellow, Bengio and Courville (2016).

To the best of our knowledge, this is the first attempt to process any colourimetric test using any deep learning technique including CNN. The assay type detection is an optional step of the overall system architecture (Fig. 3. 3). This thesis involves two case studies, each of which represents a distinct assay type. Therefore, the assay type determination is a binary classification problem.

<div align="center">(a)   (b)</div>
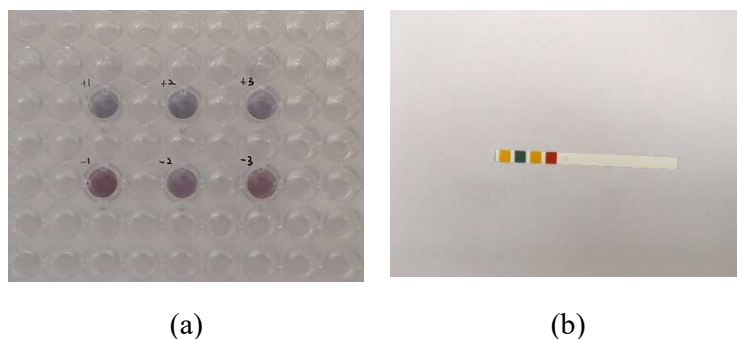
Fig. 3. 4: Assay type to be determined by the pre-trained model. (a) ELISA, (b) LFA

This work explores AlexNet (Krizhevsky, Sutskever and Hinton, 2012), inception modules (Szegedy et al., 2015, 2016) and ResNets (He et al., 2015) to determine the assay type by using our case studies (Fig. 3. 4). These case studies are different in size, shape, colour and materials. Therefore, we assume that the classification task by these pre-trained models should be performed effectively with high accuracy. It is also assumed that the classification can be provided without using a large dataset (<1000).

In addition to providing the broader classification of assay type, these pre-trained models should also be effective to fine-tune the system (Fig. 3. 3) for a brand to brand and purposeful variation among the assays, which may draw the attention of the future researchers (Fig. 3. 5).

(a)	(b)	(c)

Flat	Round	V-shape

(d)	(e)	(f)
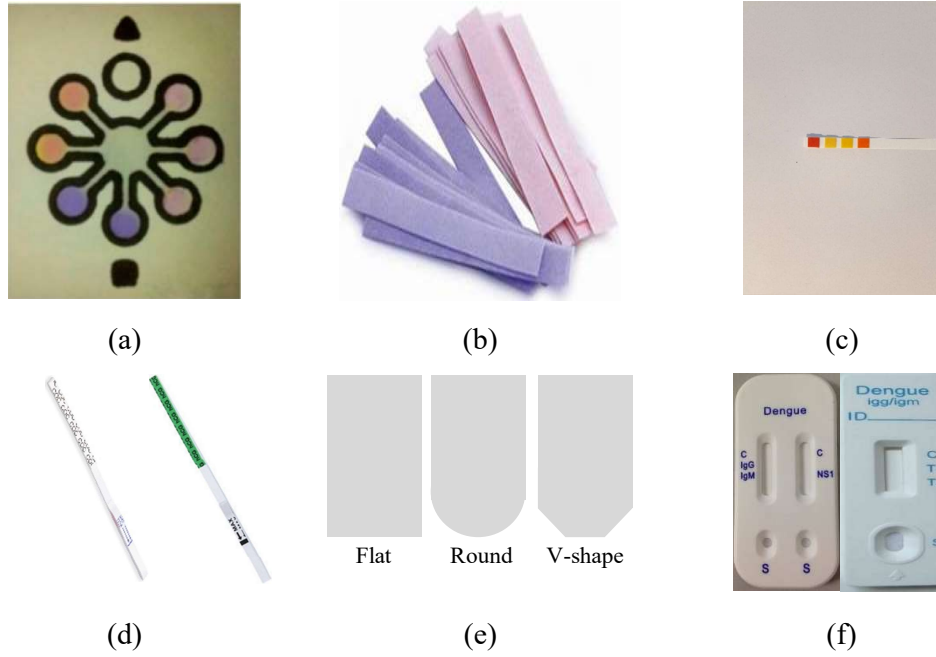
Fig. 3. 5: Variations in the assay. (a) pH and nitrite detection by (Lopez-Ruiz et al., 2014), (b) conventional litmus paper, (c) universal pH indicator used in this work, (d) brand to brand minor colour variation for HCG detection (pregnancy test), (e) variation in the well shape, (f) brand to brand variation in the geometric shape in dengue detection assay

Fine-tuning Pre-trained Models

As mentioned earlier, the input images require re-sizing to comply with the pre-trained model (Fig. 3. 6). After loading the pre-trained CNN, the final layers are replaced to fine-tune the model. After training the model, the result can be passed on to the subsequent stages.
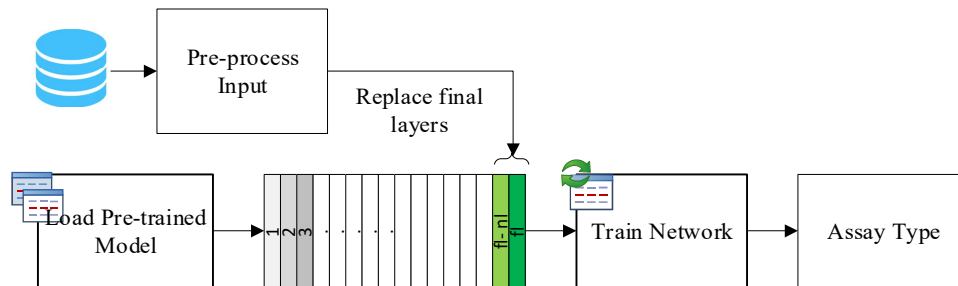
Fig. 3. 6: Deep learning using transfer learning function. fl= final layer and nl= nth layer before the final layer, $n \in \mathbf{Z}^+$ and n< fl

The AlexNet (Krizhevsky, Sutskever and Hinton, 2012) is trained on more than a million images. It contains five convolutional layers and three Fully Connected (FC) layers which are marked in Fig. 3. 7. The concept of convolution is well described in Ng (2018).
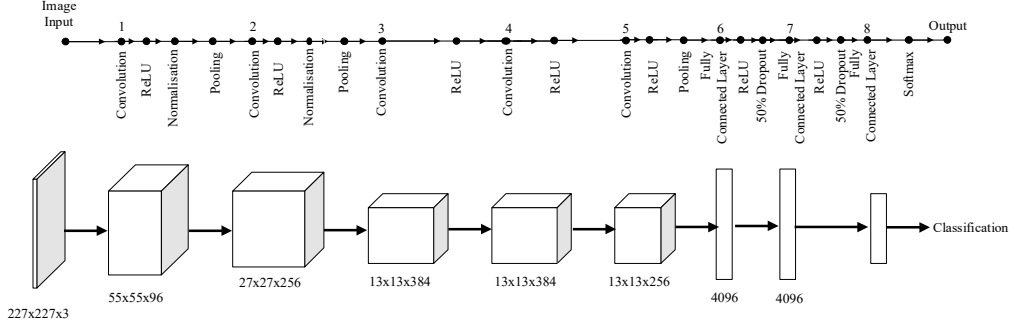
Fig. 3. 7: Eight layers of AlexNet

The fully connected layer utilises the output of the convolutional, activation or pooling layer (Scherer, Müller and Behnke, 2010) as the input volume and provides an N-dimensional vector, where N= number of classes. In the case of the softmax approach, FC considers the output of the previous layer and estimates which of the high-level features of the previous layer are the mostly correlated features for a specific class. By multiplying the input and the weight matrix and adding a bias vector, FC seeks to attain the correct probability for the classes. Using AlexNet in Fig. 3. 6, fl=25 and nl=2.

In the Large Scale Visual Recognition Challenge (ILSVRC2012), the AlexNet, containing two parallel convolutional neural network lines, showed 15.3% top-5 error rate. AlexNet uses ReLU function (Nair and Hinton, 2010).

$$f(x) = \max(0, x) \qquad (3.\ 7)$$

In Eq. 3.7, $f(x) = 0$, when $x < 0$ and $f(x) = x$, when $x \geq 0$. Hence, any negative input fed to ReLU function turns the value instantly into zero. Due to ReLU activation function after every fully connected layer, AlexNet can be trained faster than sigmoid function based TML techniques.

Few other pre-trained models were also explored in this work such as two versions of Inception (Szegedy et al., 2015, 2016) and residual connections based ResNets (He et

al., 2015) allowing us to train the model with less number of samples and faster computation.

The inception layer aims to scan a bigger area while retaining a fine resolution for small information of the images. The network architecture (Szegedy et al., 2015, 2016) presented more freedom regarding the specifics of the convolution. Each inception module can perform a number of parallel convolutional computation and concatenating. With a hope that a series of Gabor filters of different sizes will be able to provide better management of multiple object scales, the parallel convolution of different sizes starts from the most accurate detailing, i.e. 1x1, continues to the bigger ones (Fig. 3. 8).

(a)

(b)
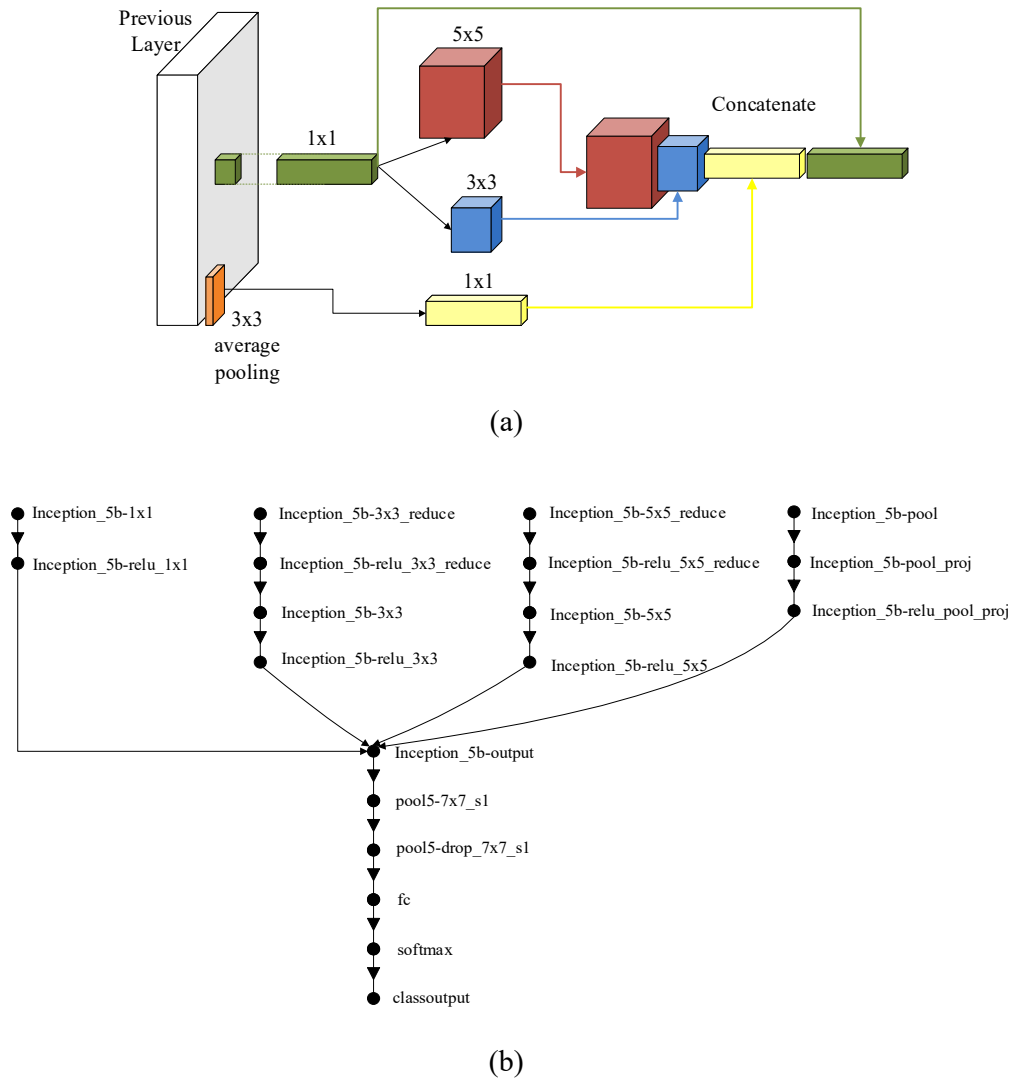
Fig. 3. 8: Architecture of GoogLeNet. (a) Basic idea of inception modules, (b) Final layers of fine-tuned GoogLeNet

A 1x1 convolution maps an input pixel along with the associated channels to an output pixel, not looking at anything around itself. It is often used to reduce the number of depth channels, since it is often very slow to multiply volumes with extremely large depths. Theoretically, the neural network can 'choose' which input 'colours' to look at while using this mapping, instead of multiplying everything, which would be extensively expensive.

Unlike the traditional Gabor filters, the filters of the inception layers are learnable. In order to prevent overfitting due to a huge number of parameters involved, the Inception exploited bottleneck approaches while rebuilding the inception module with more non-linearities and fewer parameters, followed by a max pooling layer to summarise the content of the previous layer. The results obtained are concatenated one after the other, and passed on to the next layer (Fig. 3. 8).

To fine-tune the model, in the case of inception modules based GoogLeNet, the final three layers of the network have to be replaced to retrain the model with our dataset. These three layers are 1000 fully connected layer called 'loss3-classifier', softmax and the output layer. These layers embrace the mechanism to train the model from the extracted features into class probabilities and labels. The final layers can be visualised in Fig. 3. 8.

In the case of an advanced version of GoogLeNet, i.e. Inception-v3, the final three layers are replaced and connected to the 313th connection that performs average pooling.

This thesis also explored ResNets. The ResNet (He et al., 2015) investigated that if larger function classes contain the smaller ones, then can it be guaranteed that increasing them strictly increases the expressive power of the network. The basic idea of ResNet includes the identity function as one of the elements while progressing through the network (Fig. 3. 9). Therefore, when a newly added layer is trained into an identity mapping, $f(x) =$input $(= x)$, the new model is supposed to be as effective as the original model.

The new model can potentially provide a better solution to fit the training dataset, hence the added layer may aid in reducing training errors. Using ResNets, the gradients can

flow directly through the skip-connections (Fig. 3. 9) backwards from later layers to initial filters.
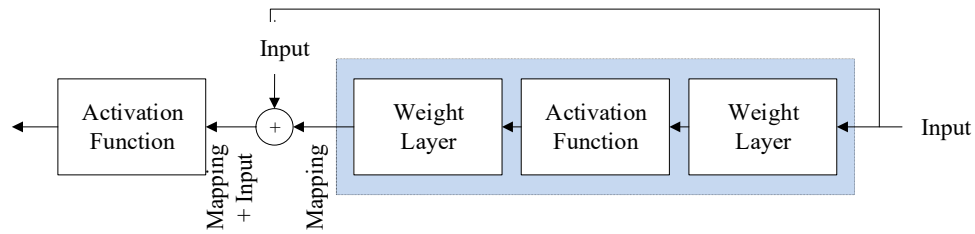
(a)
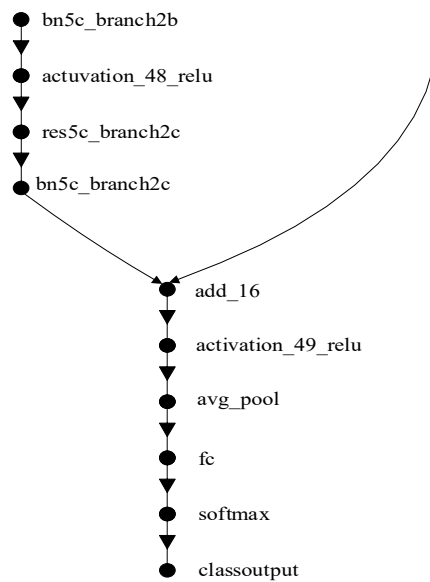
(b)

Fig. 3. 9: Architecture of ResNet. (a) Basic idea of residual network, (b) Final layers of fine-tuned ResNet-50

Similarly to GoogLeNet, the final layers have to be replaced to fine-tune the model of 50 layers deep ResNet-50 (Fig. 3. 9) and 101 layers deep ResNet-101.

However, unlike AlexNet, the learning rate of the intermediate layers of Inception models and ResNets are set to zero to minimise the training time and prevent overfitting. For example, the GoogLeNet comprises 22 layers with 144 connections. Computation of selected gradients is prevented by freezing the initial 110 connections up to inception_5a module in Fig. 3. 8, effectively making the process faster. Freezing the initial layer is more particularly helpful when the dataset is small and the similarity with the original model (e.g. GoogLeNet, ResNet) is low. In this way, the initial layers

would retain what it has been trained on, using a larger dataset, previously. The weights of these initial layer are chosen to be frozen. While freezing the layers, one should remember that only the high-level features incline towards the class label. Therefore few of the layers are required 'not' to be frozen.

In this work, the pre-trained models are explored for the task of assay type detection, not the final colourimetric decision. The pre-trained models are useful when the task is similar. The colourimetric test of a specific assay would have only one dissimilarity, i.e. colour. Therefore, it is not logical to use the deep neural network for such intra-class classification, especially when one of our goals is resource-limited settings.

The CNN architectures are used for learning detailed and abstract geometric features in relation to the position of those objects. The deeper layer is capable of learning complex features[7]. Colour is one of the features to be looked at the earlier layers of the network. In order to learn the assay type as well as to provide the colourimetric classification, the network would require a large-scale dataset and/or a different fine-tuning. A large-scale dataset and high computational resources would enable the scope of a developed model approach as well.

Fine-tuning of the pre-trained models in this work has been done using two approaches: feature extraction, where the final layers were replaced, and the overall network of the pre-trained models are used as a fixed feature extractor for the case studies used in this thesis; and freezing initial layers, while allowing the network to be trained for few layers.

On the other hand, the fine-tuning approach is logical for the assay type detection because it is a similar task, i.e. visual recognition. Therefore, there is no need for a large-scale dataset. It can save time by not requiring to continuously update the parameters for the entire network, thus less requirement of substantial computational power. However, for fine-tuning the colourimetric classification process, one possible approach could be- using the architecture of the pre-trained model by initialising all the weights randomly, train the model up to certain layers before FC layer, and use the output of this intermediate layer of the pre-trained model and then feed it into a linear

---

[7] The complexity would depend on the context.

classifier. This would complex the model unnecessarily, and a good model is usually the simpler ones. Moreover, due to a small number of data, this approach can still lead towards overfitting problem, which might not be addressable by data-augmentation. Therefore, in this work, we investigated TML techniques to seek a computationally efficient solution for the colourimetric classification.

3.5.3 Traditional Machine Learning

The basic model for image-based classification is well-established in the literature, consisting of data pre-processing and training the model with the class label. This work carefully explored the existing systems, to design each component and sub-section of these components to perform data pre-processing and classification techniques to solve our particular problem. Many of these techniques will require trial and error and experimentation to finalise the methods. The workflow of the finalised method is provided in Appendix E.

Image Processing

The image processing technique is a part of the data pre-processing step of the overall framework, and is outlined in Fig. 3. 10.
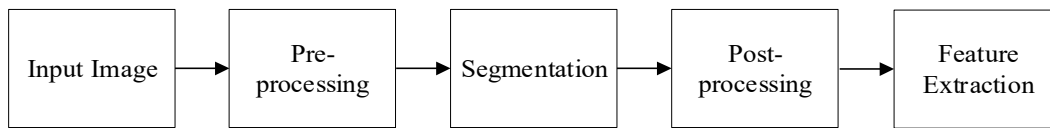
Fig. 3. 10: Overview of image processing technique

This process can be further sub-categorised into pre-processing, segmentation and post-processing before extracting the features from the ROI.

Image Pre-processing

    i)   Dimension Reduction

The images are dynamically scaled to increase the processing speed. The algorithm is provided in Appendix E.

For colour quantisation, quantisation was carried out at each colour channel as well as on the entire image to examine the impact on the overall image processing technique.

Another approach under consideration was quantisation at super-pixel level (Ren and Malik, 2003). The idea is to define a polygonal part of the image, which is larger than a normal pixel. This number is varied as 300, 500, 700 and 1000. This polygonal segment is rendered with uniform colour and brightness. This technique can merge close colours. An effective reduction of the number of colours can ease the segmentation process while creating masks, which could be particularly useful for wet-chemical-based images.

JSEG is a region-growing segmentation technique, proposed by Deng and Manjunath (2001), finds its usage to merge close clusters and to reduce the colours. Therefore, the method was also considered as a pre-processing technique.

ii)   Colour-space transformation:

The images captured by available mobile phone cameras are more commonly in sRGB format. The impact of colour space transformation for segmentation task on colour images was explored in our earlier work (Tania, Lwin and Hossain, 2018, 2016). Few of the segmentation techniques require colour space transformations to ease the process. For example, RGB to the LAB for k-medoids segmentation for the TB test case study.

The impact of colour spaces was evaluated at the feature analysis stage as well, where the RGB images were converted into a LAB, HSV, YCrCb, XYZ and L-RGB.

iii)   Image quality enhancement:

Sharpening: The RGB image is converted into grayscale and enhanced by sharpening the image features (e.g. edges) using unsharp masking method (MATLAB & Simulink, 2013).

CLAHE: The grayscale of the input image is enhanced by Contrast-Limited Adaptive Histogram Equalization or CLAHE (Zuiderveld, 1994). For colour image enhancement, the RGB image is converted into LAB colour space. The L channel value is scaled to keep the value within [0 1], to match with the range of equalisation. Then, CLAHE is performed on L channel only. The output is multiplied by the range used by LAB space. The resultant image is converted back to RGB for better visualisation. Later on, the tile

number and clip limit were varied. Then, super-pixel processing (Ren and Malik, 2003) was carried out.

As opposed to enhancement, few image degradation techniques were explored in the research. The Gaussian Blur filtering is utilised as a negative image enhancement technique. In an image, x being the distance from the origin in the horizontal axis, y being the distance from the origin in the vertical axis, 2D Gaussian or normal distribution can be written as Eq. 3.8 (Shapiro and Stockman, 2001).

$$\text{Gaussian, } G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.8}$$

where $\sigma = standard\ deviation$

As a separate technique, the Gaussian noise using Eq. 3.8 is added to each colour channel of the image. This additive noise with a variance of 0.01 is introduced to the image to increase the number of data.

Segmentation of ROI

In order to separate the ROI, at first, the foreground of the image needs to be segmented as Eq. 3.9.

$$\text{Segmented image, } g(x,y) = \begin{cases} 1 & if\ f(x,y) = foreground\ pixel \\ 0 & if\ f(x,y) = background\ pixel \end{cases} \tag{3.9}$$

Various segmentation techniques were carried out to explore the best suitable segmentation method, e.g. thresholding, clustering and region-based methods.

Thresholding is one way to find $g(x,y)$ in Eq. 3.9. A more detailed concept can be found in (Gonzalez and Woods, 2018).

Among the clustering techniques, this work utilised k-means for colour separation using k++ seeding (Arthur and Vassilvitskii, 2007). This work also explored k-medoids clustering which looks for centroid or medoid, which can be seen as that object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal. The k-medoid minimises the absolute distance between the points and the selected centroid, unlike k-means which minimise the squared distance. The value of enlarging criterion $\sum_{k=1}^{K}\sum_{C(i)} || X_i - c_k ||_2^2$ is commonly higher for k-medoid than k-means (where

observation $X_1, \dots X_n$, $X \in \mathbb{R}^p$ and initial assumption for centres $c_1, \dots c_k$; a cluster of points $X_1, \dots X_n = f(C)$, which assigns each observation X_i to a group k, $k = 1, \dots K$). Thus, k-medoid can be computationally expensive.

Post-processing

After segmentation, post-processing is required to clean the image as well as to separate the ROI. The steps include morphological operation, object detection and noise filtering. In a simple term, the morphology implies the study of form or structure. In the case of image processing, it is the mean to identify and extract meaningful image descriptors based on properties of form or shape within the image.

Performance Evaluation of the Image Processing Step

The conventional method to evaluate the image processing performance is based on the area of the ground truth. However, this work assumes that the area of segmented ROI would be less crucial due to the feature-set to be utilised from the extracted ROI. Therefore, the performance of image processing is evaluated using the quality of segmentation (Table 3. 1).

Table 3. 1: Segmentation performance evaluation by conventional and proposed method

Metric	Conventional	Proposed
True Positive (TP)	Ground truth $\cap$ Segmentation	ROI $\cap$ Segmentation
False Positive (FP)	Segmented Image $\cap$ (Ground truth)'	Segmented Image $\cap$ (ROI)'
False Negative (FN)	Ground truth $\cap$ (Segmented Image)'	ROI $\cap$ (Segmented Image)'
True Negative (TN)	(Segmented Image)' $\cap$ (Ground truth)'	(Segmented Image)' $\cap$ (ROI)'
Sensitivity (%)	$\dfrac{TP}{TP + FN} \times 100$	
Specificity (%)	$\dfrac{TN}{TN + FP} \times 100$	
Accuracy (%)	$\dfrac{TP + TN}{TP + TN + FP + FN} \times 100$	$\dfrac{Correctly\ classified\ samples}{Total\ number\ of\ samples} \times 100$

In this case, ROI would refer to the ROI of a single sample in case of TB-test (Case Study 1) and ROI of the objects in case of multi-objects per sampled type assay (Case Study 2).

The performance of the overall image processing framework is evaluated based on the correctly separated ROI. The success of the colourimetric classification will depend on the accuracy as defined in Table 3. 1.

Feature Extraction

Once the samples (ROI) are separated, the remaining noise is filtered (if any), and the characteristics of these samples are analysed. In this thesis, the feature analysis involves the measurement of colour moments (c_m) in various colour spaces (cs). A description of the features for any colour channel in any colour space $\{ft_{cs}(c_m)\}$ is provided in Table 3. 2. For each colour channel (c_c) of a colour space, let's assume, $N(g) =$ Number of pixels at any colour channel g; $L =$ Number of intensity or values available; $M =$ Number of pixels $= N_l \times N_w$; $I =$ Value of pixel (e.g. intensity); $r =$ Corresponding row; $c =$ Corresponding column; $n =$ Number of observations and $P(g) =$ First-order histogram probability $= \frac{N(g)}{M}$.

Table 3. 2: Feature set of colour moments

Feature	Equation	Description
Mean	$$\bar{g} = \sum_{g=0}^{L-1} g\, P(g) = \sum_r \sum_c \frac{I(r,c)}{M}$$	Average colour or brightness of the image
Standard deviation	$$\sigma_g = \sqrt{\frac{1}{1-n} \sum_{g=0}^{L-1} (g - \bar{g})^2 P(g)}$$	Contrast of the image
Mode	$mode$	Most frequent value in the distribution
Skewness	$$SKEW = \frac{\bar{g} - mode}{\sigma_g}$$	Asymmetry of colour distribution
Energy	$$E = \sum_{g=0}^{L-1} [P(g)]^2$$	Reciprocal to the number of intensity level
Entropy	$$ENTROPY = -\sum_{g=0}^{L-1} P(g)\, log_2\, [P(g)]$$	Bit information

The colour histogram of the segmented image, represented as a probability distribution, $P(g)$, signifies the number of pixels at each intensity level of the colour channels. More detail description of these features (Table 3. 2) in terms of image processing, are provided in Sergyan (2008).

Regarding colour spaces, this work considers RGB, LAB, HSV, YCrCb, XYZ and L-RGB colour spaces to begin with. The images captured by mobile phone cameras are usually in sRGB or standard Red, Green, Blue colour space. In CIE 1931 XYZ colour space, Y symbolises the luminance, whereas the XZ plane holds all possible chromaticities at that luminance. Both HSV and LAB are closer to human colour perception. HSV separates the intensity of the colour information. Therefore, for a robust system, HSV can help to deal with lighting conditions and shadows.

Similarly, the 'a' and 'b' channel of LAB colour space signifies the colour. The LAB-gamut is a superset of the RGB-gamut, and it is more perceptually linear, which provides it with an advantage while performing computation using Euclidean distance. The YCbCr also separates luma and chroma. Due to the advantage of chroma subsampling in YCbCr to make the image and video smaller in size, it finds usefulness in image processing, mobile application and game development. Lastly, the linearisation of L-RGB or linearised gamma-corrected RGB values is conducted using sRGB standard (Adobe, 1998).

In addition to colour moments, this work also considers the colour difference between the control and sample. This feature is denoted as $ft(\Delta E_{LAB})$ in this thesis. This feature, $ft(\Delta E_{LAB})$, is obtained by the colour difference calculation introduced earlier for reaction phase and time-dependent approach as Eq. 3.2. As explained in section 3.3, this difference is computed in LAB colour space only.

The use of the control is not a new thing in colourimetry (Smith et al., 2014). However, the meaning of control in this work merely refers to the colour of a typical sample. Regarding image processing, this control colour is the ground truth for the samples. For the case study of TB-test, the variation regarding colour is noteworthy. In real-life scenario, it can be more diverse due to climate, geo-location or the sample itself. Continuous update of the dataset using the TB-test at a different condition and an enormous amount of data can assist in finding the correct control colour, which is not

the circumstance of this study. Therefore, control colour is not used for the case study of TB-test. Using control colours as part of the feature would potentially make the system unreliable to completely unseen data, tested in a different environment.

For the LFA case study of universal pH indicator paper, the conventional method involve comparing the test strip against a colour chart. This is a widespread method for colourimetric tests, particularly for the semi-quantitative tests. This colour chart can be thought of as the control colours, which can be considered as a reliable feature because unlike the assay type of TB-test, these assays are stable.

This work also considers pseudo-control colours as part of the feature-set. This is a novel feature. This feature is applicable for the assay containing multiple objects such as urine dipstick and the pH indicator paper used in this study (Case Study 2). Therefore, more detail is provided in Chapter 6, where the feature is being used.

As similar to the control colour, pseudo-control colour is also not suitable for the TB-test. First of all, the TB-test is not multiple-objects per sample type assay; rather it can have multiple samples per image. Moreover, the assay plate could contain only one class. Therefore, this control colour would have to generated from each plate as shown in Smith et al. (2014). Hence, it would be control colour, not pseudo-control colour.

In order to optimise the number of features, the impact of the features is analysed by univariate analysis as part of the exploratory data analysis or EDA (Tukey, 1977). As illustrated in Fig. 3. 11, the feature optimisation is conducted using the process of elimination technique (Burden and Faires, 2001).
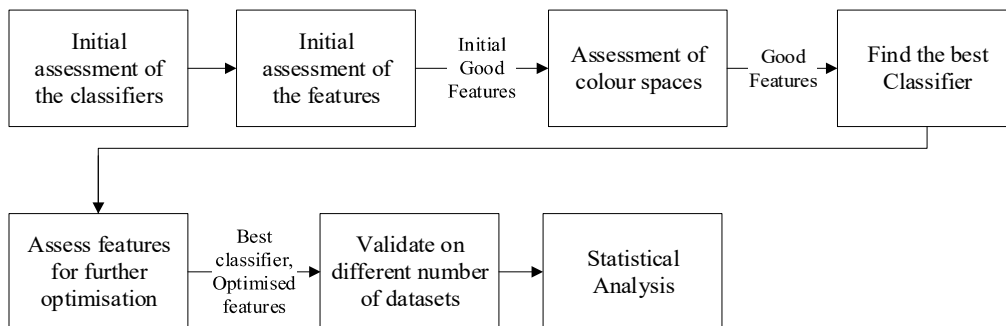
Fig. 3. 11: Feature selection, optimisation and classifier selection

Classification

After extracting the features, the classifiers are trained using supervised learning. In this work, the case study of TB-test is a binary classification problem, whereas the pH test strips are to be classified in eight different classes. Although this work investigates a wide range of conventional machine learning techniques for classification, in this chapter, more attention is provided towards LS-SVM due to its performance in the literature for similar applications (Solmaz et al., 2018; Mutlu et al., 2017).

LS-SVM, utilised in this work, was proposed by Suykens et al. (2002) and was originally introduced in Suykens and Vandewalle (1999). This work utilises Radial Bias Function (RBF) as the kernel. This work also takes advantage of Bayesian interference and regularisation proposed by Suykens et al. (2002) to take full benefit of probabilistic interpretations, automatic hyperparameter tuning and relevance determination. However, simplification of the model using least square and equality constraints has few shortcomings as well such as the issue with sparseness and robustness. Suykens et al. (2002) compensated for these issues by introducing weights to the LS-SVM algorithm and trade-offs between efficiency and robustness. Finally, a robust cross-validation (Brabanter et al., 2002) is employed as suggested by Suykens (2002).

Performance of the Classifier

The performance of the classifiers is primarily evaluated based on the confusion matrix. The performance was sequentially evaluated to finalise an optimal solution as illustrated in Fig. 3. 11. The Receiver Operating Characteristic (ROC) curve is also presented for the finally selected classifier utilising the optimised feature-set to demonstrate the quality of the classifier.

The data-partitioning mechanism for the overall thesis is illustrated in Fig. 3. 12.
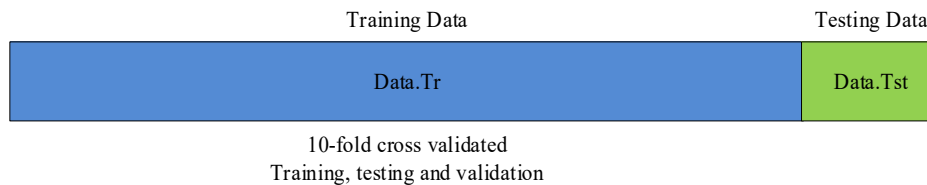
Fig. 3. 12: Training and testing data division

The training dataset (Data.Tr) in Fig. 3. 12 is k-fold cross-validated. In k-fold cross-validation, the dataset or sample-set is randomly partitioned into k equal sized subsamples. In each case (of k), a single subsample is retained as the validation data for testing the model, and the remaining k − 1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results from the folds (k) can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. In this work, k was chosen to be 10, which is the most conventional value.

A small dataset (Data.Tst) is reserved till the very end to test the final system. In this way, the system can easily be tested for the entirely unseen data, which can assist in identifying the overfitting problem, if any. Hence, the overall cross-validation can also be seen as a k-1 cross-validation method.

3.6 ASSURED Evaluation Criteria

As mentioned in Chapter 1 and 2, this work evaluates the final system to comply with the ASSURED criteria. Therefore, the proposed system does not involve any additional hardware or operational cost to comply with the ASSURED-affordability. The methods presented in this chapter aspired to necessitate minimal user intervention. The system is supposed to act as an expert to provide a colourimetric decision. This work aims to present a computationally efficient system, which can be easily deployable to mobile devices to make the system accessible to the target audience.

All the initial computational performance assessment in this work is based on the accuracy. However, a machine learning model with a known level of accuracy may have higher predictive power than the models with higher accuracy (Zhu and Davidson, 2007). The formal terminology of this phenomena is known as accuracy paradox. The accuracy of a system can be a misleading evaluation metric because it cannot differentiate between the impact of the majority and minority class. Weiss and Provost (2003) showed that accuracy could lead towards a low minority-class performance. Therefore, this work utilises specificity and sensitivity, supported by the ASSURED criteria as well. The performance in terms of true positive and false positive rates can

also be explained by the ROC curve. The area under the curve expresses how much a model is capable of distinguishing between classes.

This work also considers the Confidence Intervals (CI) of accuracy, specificity and sensitivity using Clopper-Pearson's confidence intervals (Clopper and Pearson, 1934) in short known as the 'exact' method.

Few other evaluation metrics such as precision, F-measure and Cohen's kappa coefficient is presented in our published research (Abuhassan et al., 2017; Shabut et al., 2018), however is not included in this thesis, because the ASSURED criteria requires only specificity and sensitivity; more focus of this thesis is drawn towards the data-balancing issue.

For predictive models such as the work presented in this thesis, a balanced dataset would be preferable to explain the performance of the classifier in terms of specificity and sensitivity. However, data-balancing can be tricky due to definite bias and permanent bias issue (Matloff, 2017). It should also be mentioned that the imbalance in the classes should not be perceived by the ratio of classes. This is a particular concern when there is no substantial amount of dataset to learn the pattern of the minority class.

In the medical image processing, finding an adequate amount of balanced dataset has always been a challenge. In the case of imbalance data, it is not always possible to generate more dataset for the minority-class. Hence, it is necessary to deploy a resampling technique such as Chawla et al. (2002), Estabrooks, Jo and Japkowicz (2004).

Before performing any balancing operation, this study considers two types of rarity: i) rare class and ii) rare case. In theory, the rarity should be considered in the context of underlying unknown distribution (Weiss, 2004). The rare class refers to the issue of class imbalance, whereas the rare case is a domain-specific problem, which can be in the form of both labelled and unlabelled data. With consideration of both scenarios, data-balancing is described in each of the experimental chapters before implementing it.

The robustness of the system can be perceived by the performance of k-fold and k-1 cross-validation as shown in Fig. 3. 12. Moreover, the system is going to be tested on the different proportion of the dataset.

This work explores the optimum number of features and the best suitable classifiers to reduce the computational complexity (Fig. 3. 11). This work defines a computationally efficient system in terms of accuracy, required memory size and computational time (Appendix E).

Finally, the performance of the system is evaluated by statistical analysis (Appendix F) to quantify the reliability of the claim.

3.8 Summary

The aim of this chapter was to investigate and develop a model for both of the case studies with as much similarity as possible regarding the image processing, features and classification algorithm while presenting a computationally efficient system. The rationale for the research design and methods described in this chapter are heavily dependent on our research objectives, domain knowledge for the type of data and the evaluation criteria. The next three chapters present the result of the experiments based on the research design and the methods described in this chapter. The first experimental chapter is based on the optional layer (step), as shown in Fig. 3. 3, to detect the type of the assay. Based on the type of the assay, the system will initiate the rest of the model, described as either of the case studies. In the absence of the optional layer or step using a pre-trained model, the start point of the system will be the case study itself (the workflow is illustrated in Appendix E).

Chapter 4

Assay Type Detection using Deep Learning

4.1 Introduction

Deep learning has demonstrated incredible success in recent years for image classification including object identification reducing the exasperation of the image processing (Krizhevsky, Sutskever and Hinton, 2012; Szegedy et al., 2015; He et al., 2015). Despite the advantages of deep learning of advanced analysis and feature extraction, it is a resource-demanding technology, which contradicts with our attention towards resource-limited settings. Under the circumstances, transfer learning function is utilised, where the model is trained on one task and re-used on a second similar task. Two most widely used transfer learning approaches are the develop model and the pre-trained model. In the case of a developed model approach, one would require an enormous amount of data while selecting the source-task. On the other hand, the success of the ongoing ImageNet project (ImageNet, 2016), trained on over 14 million images, has opened the door for faster implementation of deep learning on a smaller dataset using the state-of-the-art pre-trained model to classify images through magnificent features. In this chapter, we have utilised the pre-trained model of AlexNet (Krizhevsky, Sutskever and Hinton, 2012), inception module based GoogLeNet and Inception v-3 (Szegedy et al., 2015, 2016) and residual network based ResNets (He et al., 2015) to determine the assay type. For proof-of-concept, the following two assay types were selected to be determined: 1) ELISA and 2) LFA.

4.2 Dataset for the Pre-trained Models

4.2.1 Data Resampling Technique

For assay type determination, the dataset called D.ELISA. 1 and D.LFA.1 were utilised. An overview of the dataset can be seen from Table 4. 1. Each of class in Table 4. 1 contains 800 images. The LFA dataset contains one sample per image, whereas most of the images in the ELISA dataset contain multiple samples.

Table 4. 1: Dataset for assay type determination

Class label	Name of the dataset	Number of images before resampling	Resampling technique	Number of images
ELISA	D.ELISA. 1	264	Gaussian noise, CLAHE	800
LFA	D. LFA. 1	800	Not applicable	800
			Total	1,600

The dataset of LFA (D.LFA.1) is consist of an equal number of tests strips from pH level 3-10. Therefore, resampling techniques were not required to balance the LFA dataset (Table 4. 1).

The ELISA (D.ELISA.1) dataset contains both positive and negative samples from TB-test. The assay plate contains an uneven distribution of positive and negative samples. Moreover, the original dataset of ELISA is too small to be trained by the CNN. Therefore, in addition to the best images, video frames were included to increase the volume of data to 264. In order to create more diversity in the dataset as well as to create an ample amount of data to be trained by CNN, the dataset was resampled by changing the quality of the image as shown in Fig. 4. 1.

In this chapter, two techniques were deployed to change the image quality as part of the resampling technique. At first, the image quality was degraded by adding Gaussian noise. Another technique was carried to generate more data by enhancing the original images by using CLAHE as mentioned in Section 3.5.3. The change in the image quality is apparent from the histogram (Fig. 4. 1).
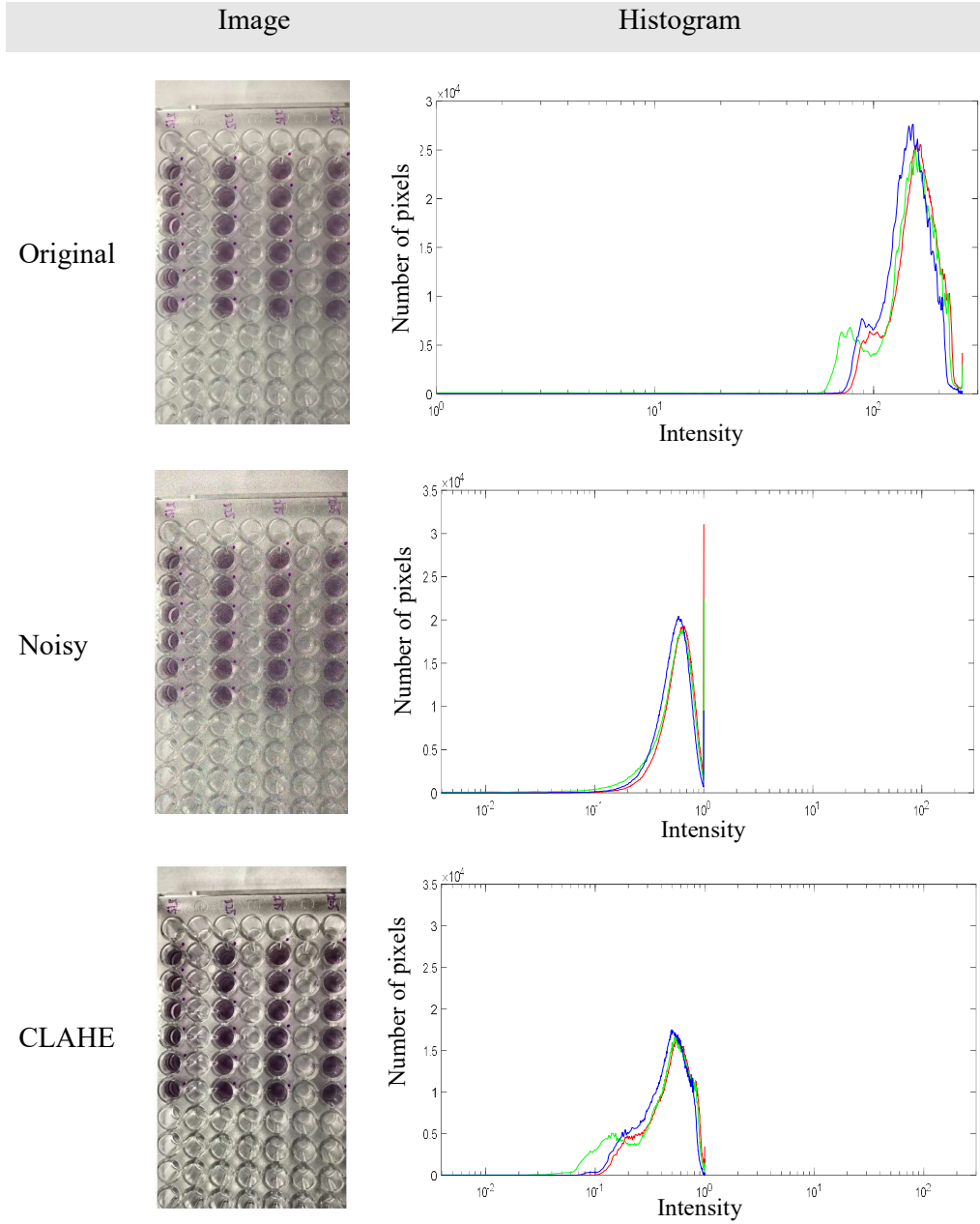
Image	Histogram

Fig. 4. 1: Image resampling by adding Gaussian noise and enhancement by CLAHE. In the first row, the histogram is shown for the original image before deploying any image quality alteration. The second row shows the noisy image and its histogram. The change in the image quality and its colour components is visible from the histogram. The last row shows the enhanced image and its histogram. The histogram shows that CLAHE enhancement resulted in higher pixel intensities in the enhanced image than the original image. For better visualisation, the horizontal axis is shown in logarithmic scale.

While further extending our experiment with deep learning, for proof-of-concept, only one of the two assay types (Table 4. 1) was considered for intra-class classification. For intra-class classification of LFA, D.LFA.1 was inapplicable due to the need of partitioning based on the intermediate classes. The dataset for intra-class classification of LFA to determine pH levels is summarised in Table 4. 2.

Table 4. 2: Dataset for intra-class classification of pH test

Name	Samples/ Class	Classes	Total samples	Total images	Resampling technique
original	65	8	520	520	Not applicable
D. LFA. 2	100	8	800	8000	Gaussian noise, CLAHE
D. LFA. 3	100	8	800	4472	Not applicable
D. LFA. 4	100	8	800	3160	Not applicable

The dataset D.LFA.2 is the final dataset after enhancing the volume of LFA data (Table 4. 2), and {original, D.LFA.3, D.LFA.4} ∈ D.LFA.2. In Table 4. 2, the dataset named 'original' contains 520 images, captured in the laboratory environment using a Samsung Galaxy S6. The original dataset contains the consistent image of independent test samples without any duplication. It is worth mentioning that deep learning is a data-hungry approach. Therefore, in order to feed the network a large amount of data, the volume of LFA dataset was increased by several techniques considering a minimal violation of 'rarity' (Weiss, 2004).

The reported articles often increase the volume of the data by capturing the same sample in different illumination conditions, several devices, various orientations and converting the image into different file formats (Mutlu et al., 2017; Solmaz et al., 2018; Kim et al., 2017a). In this chapter, we applied such techniques to increase the number of images to be fed to CNNs.

Before reaching to 8000 images (D.LFA.2), the dataset was gradually enhanced by capturing the image of 15-30% of the samples under four illumination conditions and using different devices (D.LFA. 3-4). Up to this point, the dataset can considered as

the clean data (Goodfellow, Bengio and Courville, 2016). Then, data resampling technique was applied on the LFA dataset to further enhance the volume of the data (D.LFA.2) by image degradation and enhancement techniques as described earlier in this section.

4.2.2 Data Pre-processing

The pre-trained models require the input images to possess a specific size. Resizing thousands to millions of images is computationally expensive. However, the required image size is defined by the model itself. The reduction in size may result in a loss of information, which is already expected to happen in the successive layers of the network. Therefore, the images were down-scaled before initiating the model.

In theory, the pre-processing including image resizing can be performed while data-augmentation. In this chapter, the datasets were separately resized for the pre-trained models to enable a faster execution (Appendix D). The images were converted to the size of 227x227x3 for AlexNet. The GoogLeNet and ResNets require a relatively smaller size of the image, i.e. 224x224x3. The Inception v-3 expects the image to be of relatively larger size (299x299x3).

In general, larger image size would mean more computational processing at each layer, which would demand more memory. However, it should be mentioned that the CNNs above do not deal with the whole image as a single input, instead work on fixed windows sliding over the full image. During the training process, the models operate in mini-batches comprising many tensors (DoITPoMS, 2010). To ensure better performance, the tensor is supposed to fit into the memory properly. In addition to the overall faster performance due to a smaller data load, the resizing aids the pre-trained models to perceive the key features at earlier layers. The networks may require later layers to learn the key features of a bigger image, eventually requiring longer computational time.

Data Augmentation

Data-augmentation is a popular technique to increase the dataset. In this work, data augmentation is utilised to handle the data partitioning for training, validation, testing, and prediction. Data augmentation process in MATLAB is illustrated in Fig. 4. 2 and depicted from MATLAB & Simulink (2018c).
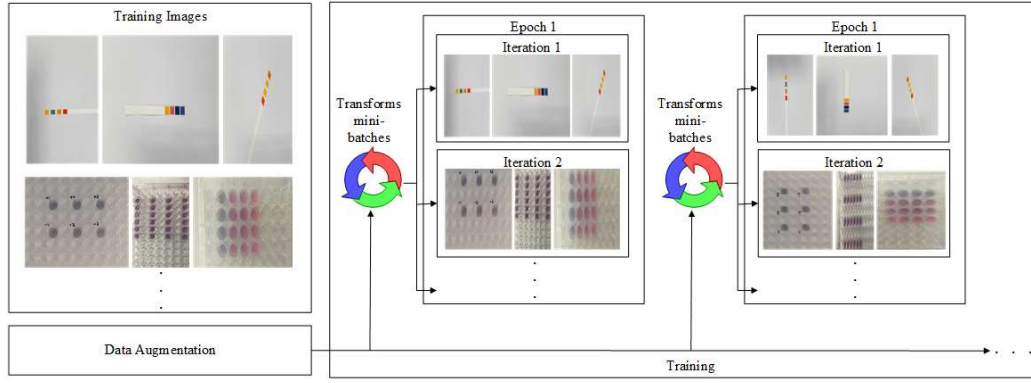
Fig. 4. 2:: Data-augmentation in MATLAB (MATLAB & Simulink, 2018c)

Pre-processing while data augmentation such as rotation and reflection needs to be randomised by perturbing the training data for every epoch (Appendix E). As shown in Fig. 4. 2, every epoch uses the dataset with some variation. This process assists in preventing the network to remember the exact details of the training image, which ultimately assists in preventing the overfitting problem. Then, normalisation is performed at the image input layer, which is the first layer of the pre-trained models used in this work. The normalisation is conducted by using the mean of the augmented images, rather using the mean of the original input images. The mean is calculated only once, i.e. for the 1st epoch. Understandably, all the epochs will have the same mean. Otherwise the average image would be altered during the training process.

Table 4. 3: Data augmentation for advanced image pre-processing

Property	Specification	Description
Reflection	Direction: Left→ Right; Amount: With 50% probability	Random reflection; Each image is reflected horizontally.
Translation	Direction: Horizontal and Vertical; Specified by the pixel range	Each augmented image is translated by a distance, in pixels, picked randomly from a uniform distribution within the range.
Scaling	Direction: Horizontal and Vertical; Specified by a 2-element vector of positive numbers	Each image is scaled by a factor picked randomly from a uniform distribution within the range.

This chapter utilised random affine geometric transformations, namely resizing, reflection, and translation while building the mini-batches. The specifics of the augmentation for training deep neural networks is provided in Table 4. 3.

4.3 Assay Type Determination

4.3.1 AlexNet

In terms of top-5 error rate, the best performing pre-trained model of ImageNet challenge considered in this work chronologically are ResNet, GoogLeNet and AlexNet (ImageNet, 2016). Among these models, AlexNet is computationally the least complex one.

For the large-scale data (ImageNet, 2016), AlexNet can take about six days to be trained, if stochastic gradient descent or SGD learning rate is 0.01, momentum, $\gamma = 0.9$ and weight decay is 0.0005 using two GTX 580 GPUs. After reaching the accuracy plateau, the learning rate is divided by 10. The learning rate decreases three times during the overall training.

Table 4. 4: Training parameter of AlexNet

Input Arguments	Description
Solver training network	SDGM (Murphy, 2012): $$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l) + \gamma(\theta_l - \theta_{l-1})$$ $\theta =$ parameter vector $l =$ iteration number $E(\theta) =$ cost function $\gamma = 0.9$
	Initial learning rate=0.001
Mini-batch options	Maximum number of epochs= 20

The training parameters of AlexNet utilised are listed in Table 4. 4. The gradient of loss or cost function, $\nabla E(\theta)$ uses the whole training set, whereas θ_{l+1} utilises the whole dataset. The stochastic gradient descent algorithm is used with a momentum to reduce the oscillation effect by controlling the influence of prior gradient step to the current one (Murphy, 2012). The Stochastic Gradient Descent with Momentum (SDGM)

algorithm assesses the gradient followed by updating the parameters utilising a mini-batch i.e. subset of the training set.

The detailed structure, including the weights and bias at each layer of the network, can be perceived from Fig. 4. 3. The total number of learnables utilised in our study can be calculated from the learners, which is briefly discussed in this section.

ANALYSIS RESULT

↑	NAME	TYPE	ACTIVATIONS	LEARNABLES	
1	data 227x227x3 images with 'zerocenter' normalization	Image Input	227×227×3	-	
2	conv1 96 11x11x3 convolutions with stride [4 4] and padding [0 0 0 0]	Convolution	55×55×96	Weights Bias	11×11×3×96 1×1×96
3	relu1 ReLU	ReLU	55×55×96	-	
4	norm1 cross channel normalization with 5 channels per element	Cross Channel Nor...	55×55×96	-	
5	pool1 3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	27×27×96	-	
6	conv2 256 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2]	Convolution	27×27×256	Weights Bias	5×5×48×256 1×1×256
7	relu2 ReLU	ReLU	27×27×256	-	
8	norm2 cross channel normalization with 5 channels per element	Cross Channel Nor...	27×27×256	-	
9	pool2 3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	13×13×256	-	
10	conv3 384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13×13×384	Weights Bias	3×3×256×384 1×1×384
11	relu3 ReLU	ReLU	13×13×384	-	
12	conv4 384 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13×13×384	Weights Bias	3×3×192×384 1×1×384
13	relu4 ReLU	ReLU	13×13×384	-	
14	conv5 256 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13×13×256	Weights Bias	3×3×192×256 1×1×256
15	relu5 ReLU	ReLU	13×13×256	-	
16	pool5 3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	6×6×256	-	
17	fc6 4096 fully connected layer	Fully Connected	1×1×4096	Weights Bias	4096×9216 4096×1
18	relu6 ReLU	ReLU	1×1×4096	-	
19	drop6 50% dropout	Dropout	1×1×4096	-	
20	fc7 4096 fully connected layer	Fully Connected	1×1×4096	Weights Bias	4096×4096 4096×1
21	relu7 ReLU	ReLU	1×1×4096	-	
22	drop7 50% dropout	Dropout	1×1×4096	-	
23	fc 2 fully connected layer	Fully Connected	1×1×2	Weights Bias	2×4096 2×1
24	prob softmax	Softmax	1×1×2	-	
25	classoutput crossentropyex with classes 'ELISA' and 'LFA'	Classification Output	-	-	

Fig. 4. 3: Detailed layers of AlexNet

As shown in Fig. 4. 3, the input image is of 227x227x3. Considering 11x11x3 blocks of pixels within the input image, each block was stretched into a column vector sized of

(11*11*3=) 363. In other words, the local regions of the image were stretched and converted into columns. There are 96 filters in this layer, which can also be thought of as the depth of the first convolutional layer. Therefore, the weight of the matrix was 96x363, which is denoted as W. The first convolutional layer operated with 11x11x3 filters and the stride was 4. There are many approaches to perform convolution.

Let us assume,

O_w = Width of the output image

I_w = Width of the input image

K = Size or width of the kernels used in the convolutional layer

N = Number of kernels

S = Stride of the convolutional operation

P = Padding

The size of the output tensor of a convolutional layer can be calculated as $O_w = \frac{I_w - K + 2P}{S} + 1$ and the number of channels of this output tensor or image is equal to N. This is a typical convolutional operation. Using this method, the output of the first convolutional layer would be [{(227-11)/4} +1] x [{(227-11)/4} +1] x 96=55x55x96, where P=0. However, for an image-based task including AlexNet, this operation is perform by GEneral Matrix Multiplication (GEMM)[8]. If M is the result matrix, where M= 396 x (55*55), then GEMM would provide $M \times W$.

The next step was to stop all the negative values to propagate through the network by activating the ReLU layer. After normalising the output of the ReLU layer with five channels per element, max pooling was conducted. One can calculate the output as $O_w = \frac{I_w - P_s}{S} + 1$, where P_s = Pool size. In this case, the input was 55x55x96 and max pooling signifies sub-sampling operation using a maximum of the four values in a 2x2 window. Therefore, the outcome of this operation was {(55-3)/2+1} x {(55-3)/2+1} x 96= 27x27x96. The step is supposed to reduce the computation and to deal with the

---

[8] Many researchers have favoured GEMM for image-based convolutional operation (Hadjis et al., 2015). GEMM has the disadvantage of overlapping blocks of pixels, which consumes a huge memory. Due to GEMM operation, the matrix is required to be reshaped after multiplication. Therefore, it is not clear that why researchers have favoured this operation.

overfitting problem. There are 256 filters in the second convolutional layer and the filter size is 5x5x48. The output of this layer was 27x27x256. This is to draw the attention of the readers again to the fact that these CNNs work on volume. Within the second convolutional layer, the dimension (plane or length x width x height) altered from 227x227x3 to 27x27x256. This work utilised a single CPU. If there were two GPUs, then this computation task would have distributed as 27x27x128 for each of GPU.

The similar convolutional operation continued for 3rd, 4th and 5th convolutional layer. In both third and fourth convolutional layer, there are 384 kernels, varying in the size. The size of the kernels in the third layer is 3x3x256, whereas 3x3x192 in the fourth one. The final convolutional layer has 256 kernels. The size of the kernels is similar to the fourth convolutional layer, i.e. 3x3x192.

The deep learning model does not rely on the domain experts to process the images, extract the ROI and feed the selected features to the machine learning algorithm. Instead, these convolutional layers act as the enhanced feature extractors. Therefore, in a conventional sense, the FC layer is the actual network of the AlexNet (Fig. 4. 3).

The first FC layer is the sixth layer of the eight-layer deep AlexNet. In this case, the 13x13x256 input is transformed into a vector, and max-pooling operation is performed before flattening the output to 4x4x256=4096 for the FC layer. The weight of the FC layer is 9216x4096. In order to map 6*6*256= 9216 neurons to 4096 neurons, this much weight is needed for the FC layer. After connecting each of the 9216 neurons to 4096 neurons to create a full connection using general matrix vector multiplication, the output would become 4096x1. In a more proper form, the output can be expressed as $[9216x4096]^{T}\cdot$ [9216x1]=[4096x1]. In this way, the FC layer will have a sufficient amount of neurons to apprehend the variability of the large-scale dataset.

One can also calculate the number of parameters involved in this layer by combing the number of weights of the first FC Layer connected to the convolutional layer, 6*6*256-4096= 3,77,48,736 and the bias, 4096. Therefore, the first FC layer has 3, 77, 52,832 parameters connected to the convolutional layer.

As most of the computation is conducted in the initial convolutional layers, the computational cost is much lower in the FC layers. The FC layer combines all the features detected from the patches of the images to produce the decision. In other words,

the layer contains most of the model parameters. Instead of placing all the burdens on a single FC layer, there are three FC layers in this structure, as shown in Fig. 4. 3. In this way, more flexibility can be attained to overcome memory constraints.

As explained in Chapter 3, the final layers have to be replaced with our particular problem to fine-tune the pre-trained model; the final FC layer had only two fully connected layers to detect the assay type. Selecting the final FC layer and two class indices as the channels, Fig. 4. 3 illustrates detailed images that strongly activate these classes. However, it is difficult to perceive any significant information from Fig. 4. 4.
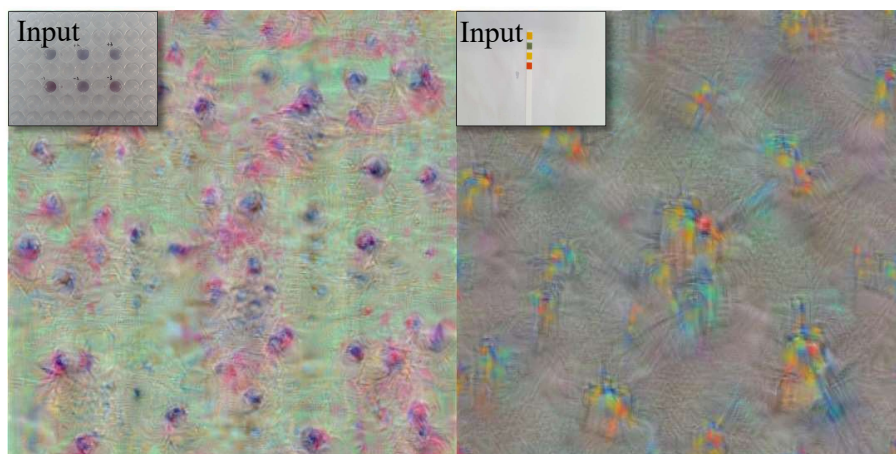
Fig. 4. 4: Visualisation of features at FC layer

From the training progress in Table 4. 5, the mini-batch accuracy and mini-batch objective function can be observed.

Table 4. 5: Training progress of AlexNet for assay type determination

Epoch	Iteration	Mini-batch accuracy (%)	Mini-batch loss
1	1	48.44	0.9782
3	50	100.00	1.7509e-07
6	100	100.00	7.8231e-08
9	150	100.00	5.1100e-05
12	200	100.00	1.7659e-06
15	250	100.00	4.1723e-07
18	300	100.00	1.2871e-06
20	340	100.00	1.6391e-07

The training was stopped after the 340th iteration. In this case, the number of iteration implies the number of passes, where a single pass signifies one forward-pass and one backward-pass, and each pass utilises the 'batch-size-number' of examples. The number of training examples in either forward or backward-pass is the batch size. Hence, iteration can be phrased as one forward pass and one backward pass of each batch size. There were 20 epochs, where each epoch represents one forward pass and one backward pass of all the training examples. In the case of 70:30 proportion of training and testing data, there were 17 iterations per epoch[9].

The overall accuracy revealed that the learning rate and maximum number of iteration were adequate for the binary classification of the selected assay types. The training progress per epoch can also be visualised in Fig. 4. 5.
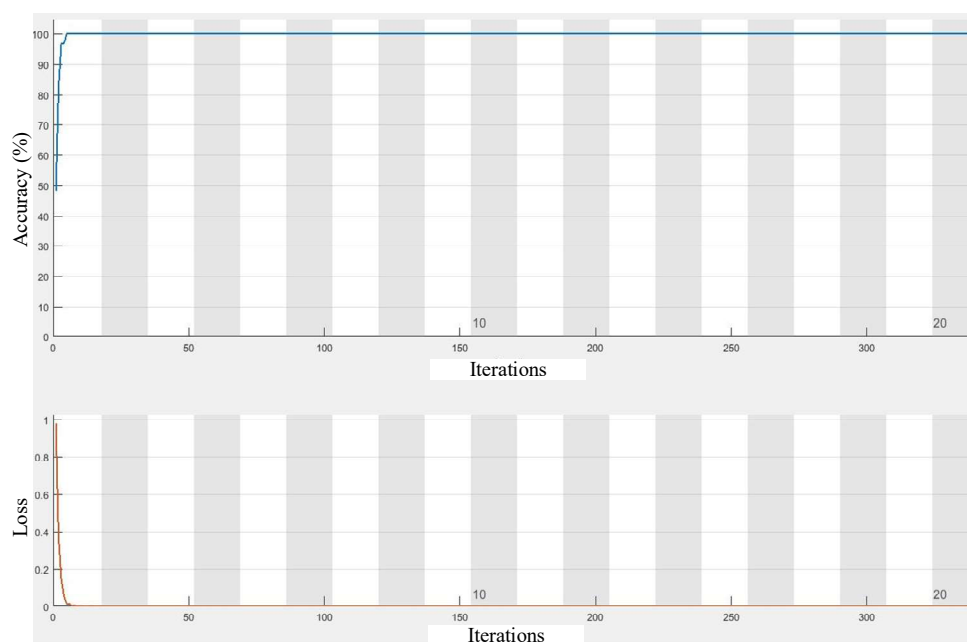
Fig. 4. 5: Training progress of AlexNet for assay type detection

Unlike the rest of the chapters, this chapter excluded k-fold cross-validation, which is the gold standard for TML. The k-fold cross-validation can assist in achieving a robust estimate of the performance of any model, including deep learning, on unseen data. However, the common notion of deep learning models to subside k-fold cross validation

---

[9] There is no thumb-rule regarding the best or right number of epochs, however, the number of epoch is highly dependent on the diversity and rarity within the dataset.

is due to higher computational expense. Instead of training k different models, in this case, a random subset of the training data is retained as the hold-out dataset for validation. At first, 70% of the data (Table 4. 1) was used for training and 30% for testing. The confusion matrix of Fig. 4. 6 is an illustration of the testing performance using 30% of the data.

Fig. 4. 6: Confusion matrix of the testing dataset using AlexNet

Later, the ratio between training and testing was re-partitioned to 80:20 and 90:10. For verification of the system against potential random variation, the training and testing were repeated three times. The performance of AlexNet was consistent throughout the process.

From Table 4. 6, it can be seen that the accuracy of assay type detection by AlexNet is between 99.77-100% using 95% CI. Fig 4.5-4.6 and Table 4.6- all suggest that AlexNet can be used for assay type detection.

Table 4. 6: Result of AlexNet for assay type detection

Metric	Value (%)	95% CI (%)
Sensitivity	100	99.54-100
Specificity	100	99.54-100
Accuracy	100	99.77-100

In Fig. 3.2, the layer of assay type detection acts as an object identification layer. It can be helpful to automatically fine-tune the system for different assay type and mitigate any brand-to-brand difference. Then, the successive layers of Fig. 3.3 can be initiated with reliability, based on the output of the assay type.

The overall model was 621MB including the training and testing images. The fine-tuned core model was 202MB, which might not be suitable for standalone native operation on the currently available and widely accessible mobile devices.

The training on 1,600 images took 15 min 42 seconds on a single CPU, whereas the prediction was provided within a fraction of a second. This trained model is heavy for the mobile devices, however, can be easily uploaded to a server. An example is illustrated in Fig. 4. 7, where the trained model is uploaded to the MATLAB server and then called on MATLAB Mobile (MATLAB & Simulink, 2018b).
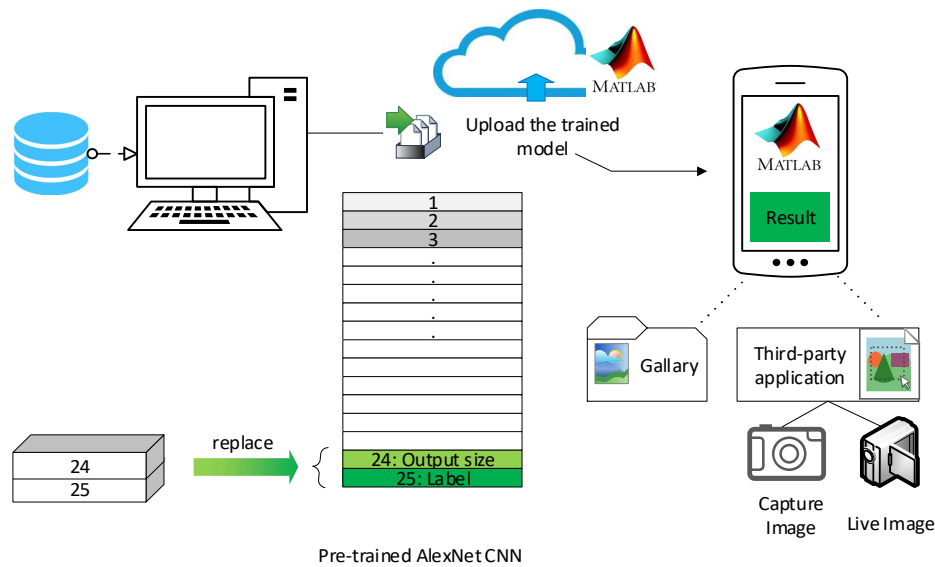
Fig. 4. 7: Mobile-enabled server-based pH test using AlexNet

To provide a prediction of assay type of an entirely unseen image, a third party application[10] was utilised to capture the new image on-site by MATLAB mobile. The server takes less than a minute to be connected with the mobile phone, depending on the specification of the mobile device, server and internet connection. The prediction time on the mobile platform was slightly lower than the desktop application when the trained model was uploaded to the server and connection was established with the mobile due to a fast internet connection and advanced mobile device, which may not be the case all the time. As discussed in Chapter 2, a detailed study of Kim et al. (2017)

---

[10] At this moment, MATLAB Mobile cannot pull input image without using any third party application.

showed that a server-based approach could make a colourimetric classification system two times slower.

4.3.2 Inception Modules

The performance of GoogLeNet and Inception v-3 was evaluated using the same computational platform with the same dataset as AlexNet (Table 4. 1). At first, for GoogLeNet the input images were converted to 224x224x3. In the first layer, these images were normalised by subtracting the mean image of the training set, which is known as 'zerocentre' data transformation. The model itself is 22 layers deep. However, the Directed Acyclic Graph (DAG) network architecture considers all the intermediate layer-array as layers. Therefore, it can be stated that the network is consist of 144 layers and 170x2 connections. The initial 110 layers of GoogLeNet were chosen to be frozen. The filter output at different layers can be visualised in Fig. 4. 8.

(a)   (b)

(c)   (d)
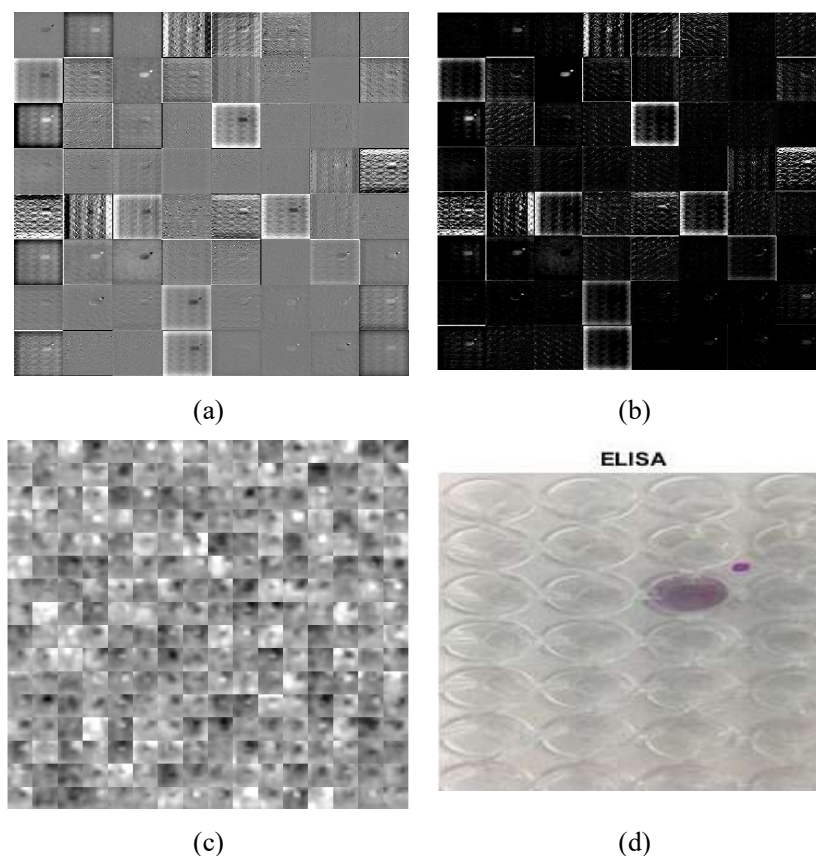
Fig. 4. 8: Random testing data at a different layer of GoogLeNet. (a) Layer 2: 'conv1-7x7_s2', (b) Layer 3: 'conv1-relu_7x7', (c) Layer 112: 'inception_5a-1x1', (d) Output with predicted label

A random image from the testing dataset was chosen by random permutation. In layer 2, which is the first convolution layer as well, 64 filters were used. The output of this operation can be seen from Fig. 4. 8(a). Each filter was 7x7x3, and the filter output was down-sampled by a factor of 2 in each direction. The immediate successive layer utilises the ReLU function to nullify the negative values in the previous layer. The outcome is shown in Fig. 4. 8(b). In Fig. 4. 8(c), the output of 256 filters at the final convolution layer called 'inception_5a-1x1' can be visualised. There were 832 channels with zero padding.

The training progress can be visualised in Fig. 4. 9. There were six epochs, and the maximum iteration was 672. The learning rate was 0.0001. In this chapter, we have utilised pre-trained models, and these pre-trained weights are supposed to be better than randomly initialised weights. Therefore, one should aim at retaining the essence of the traits of the original models, so that the pre-trained weights are not highly altered instantly. Therefore, the common practice is to use initial learning rate ten times smaller than the actual model. Although a smaller learning rate would elongate the convergence time, this hyperparameter can assist in tracking all the local minima by controlling the weights-adjustment with regarding the loss gradient.
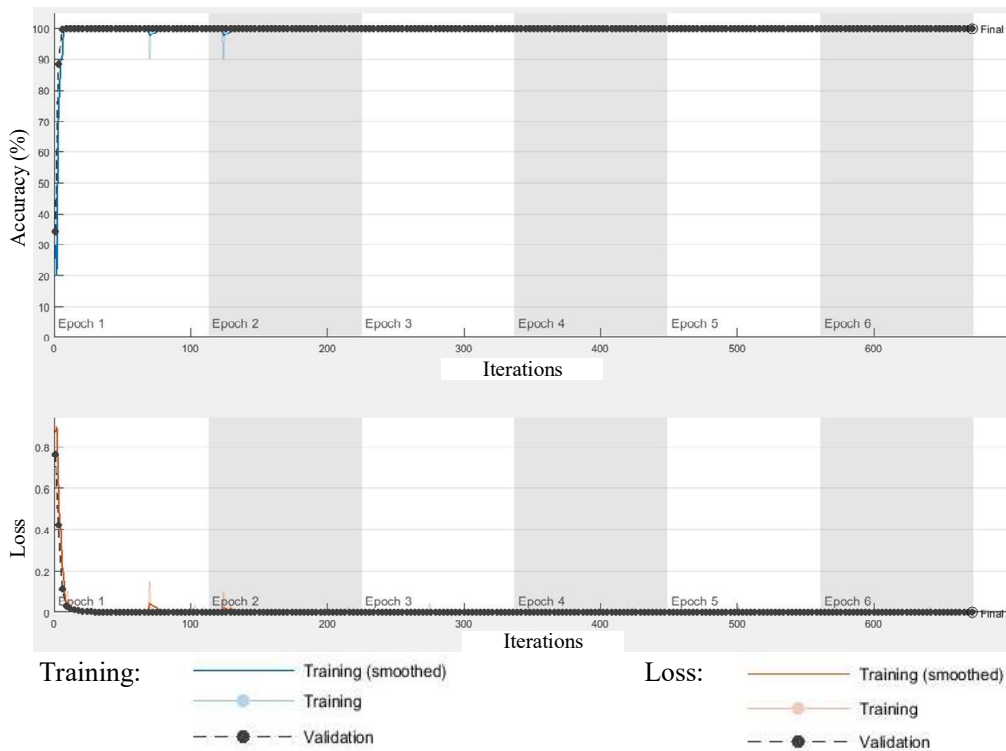
Fig. 4. 9: Training progress of GoogLeNet for assay type detection

The training accuracy (Fig. 4. 9) is the classification accuracy on each mini-batch. The initial error is visible from the smoothened training accuracy. The loss function was calculated by cross-entropy loss (Bishop, 2006). The overall accuracy of the model was 100%. The statistical output (Table 4. 6) and the confusion matrix (Fig. 4. 6) of AlexNet and GoogLeNet were identical. However, regarding training time, the model was 6.7 times more expensive than the AlexNet.

On the other hand, Inception models have several advantages over AlexNet such as batch normalization and image distortion. The later versions of Inceptions aimed at reducing the representational bottleneck. An advanced version of GoogLeNet, called Inception v-3 was deployed to evaluate the impact on the overall performance. This network includes batch normalisation in the 'Auxiliary Classifiers'. It also holds advanced regularising components to the loss formula that prevents the network from becoming too confident about a class to prevent overfitting.

Although the final accuracy by Inception v-3 was 100%, the training progress in Fig. 4. 10 revealed, the trend of accuracy and loss were nosier than GoogLeNet and the training time was more than three times higher as well.
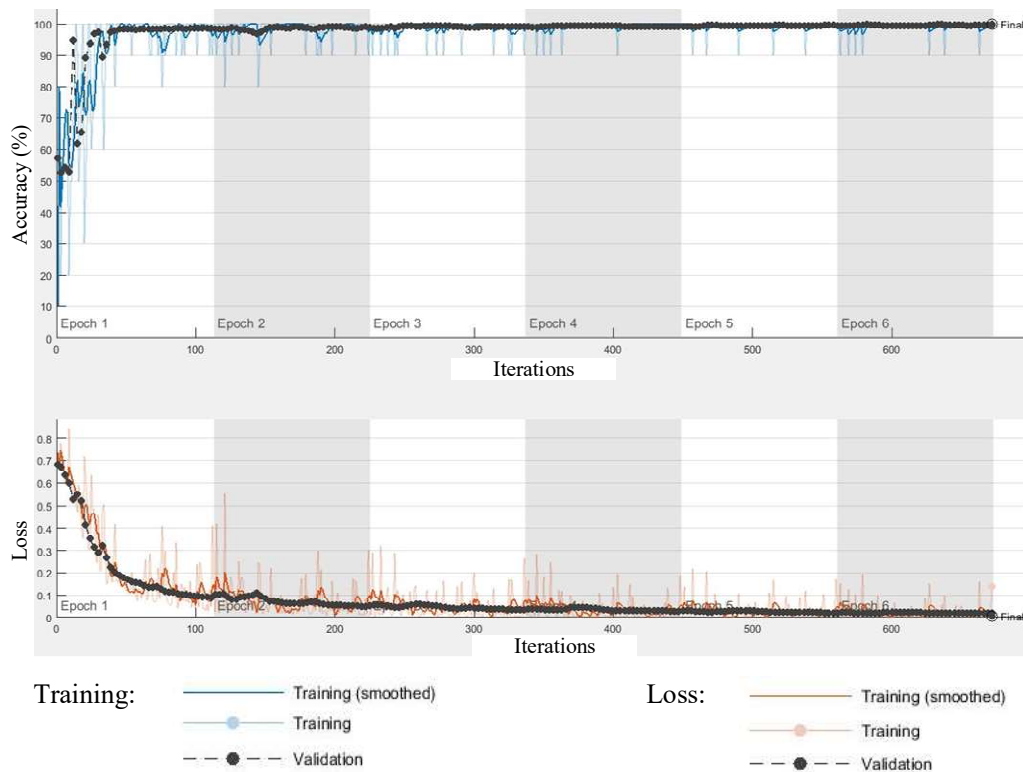
Fig. 4. 10: Training progress of Inception v-3 for assay type detection

The spikes shown in Fig. 4. 10 were due to the local minima, which could be thought of as the saddle points. The empirical study in Goodfellow, Bengio and Courville (2016) showed that the parameter optimisation of high-dimensional NN could circumvent any global minimum. Therefore, the saddle points occurred during Inception v-3 (Fig. 4. 10) are the local minima for the mini-batches which were applicable to certain dimensions. The model often was often trapped in these minima, and it can happen in many dimensions at the same time. However, the model recovered and managed to escape the local minima as the loss function was approaching towards the global minimum. On the other hand, based on the convergence of the model as shown in Fig. 4. 10, it appears that the training should have continued for a more extended period, to evaluate if the network can circumvent its pattern of entrapping in the local minima.

4.3.3 ResNets

The residual networks, ResNet-50 and ResNet-101, were explored in this thesis for assay type classification. In the ImageNet challenge, the top-5 error rate of ResNet-50 was 3.6%, which is lower than the inception models (ImageNet, 2016).
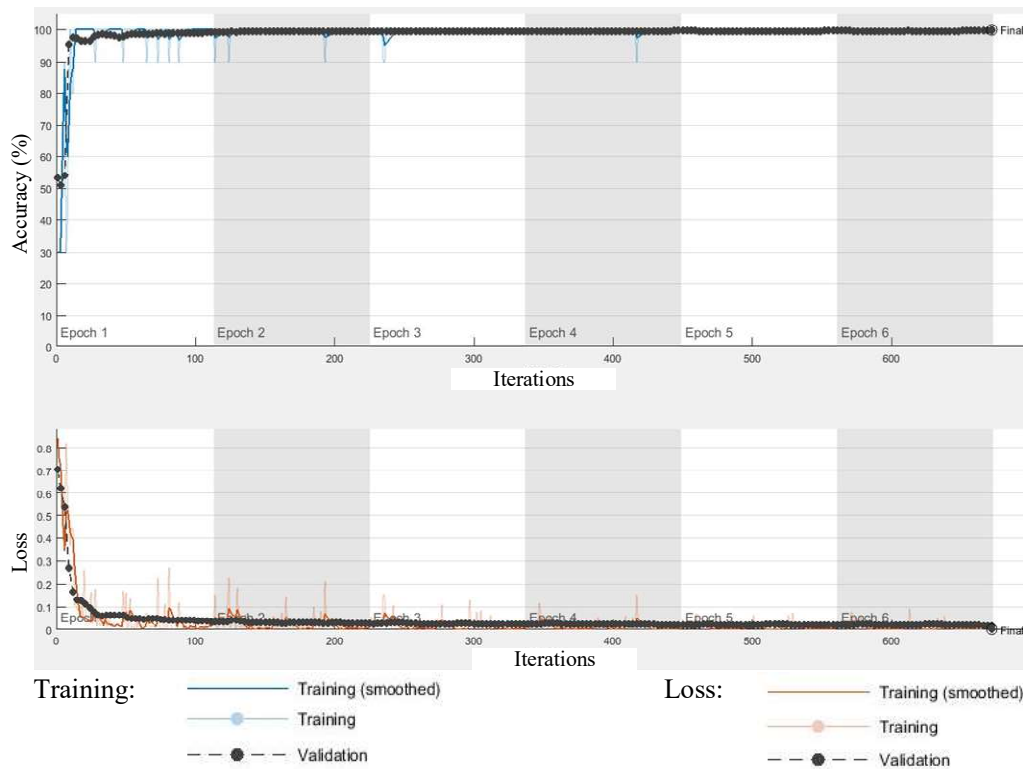
Fig. 4. 11: Training progress of ResNet-50 for assay type detection

82

The performance of the residual networks for assay type detection can be perceived from Fig. 4. 11 and Fig. 4. 12.

The number of iterations, epochs and learning rate for ResNets were kept as same as the inception models. Both ResNet-50 and ResNet-101 produced 100% accuracy. The initial training was noisy for both of them. However, the overall training progress was smoother for ResNet-101 (Fig. 4. 12). Therefore, based on the convergence on our dataset for assay type detection, ResNet-101 performed better than ResNet-50 (Fig. 4. 11). On the other hand, both of the ResNets performed better than Inception v-3, however less than GoogLeNet.
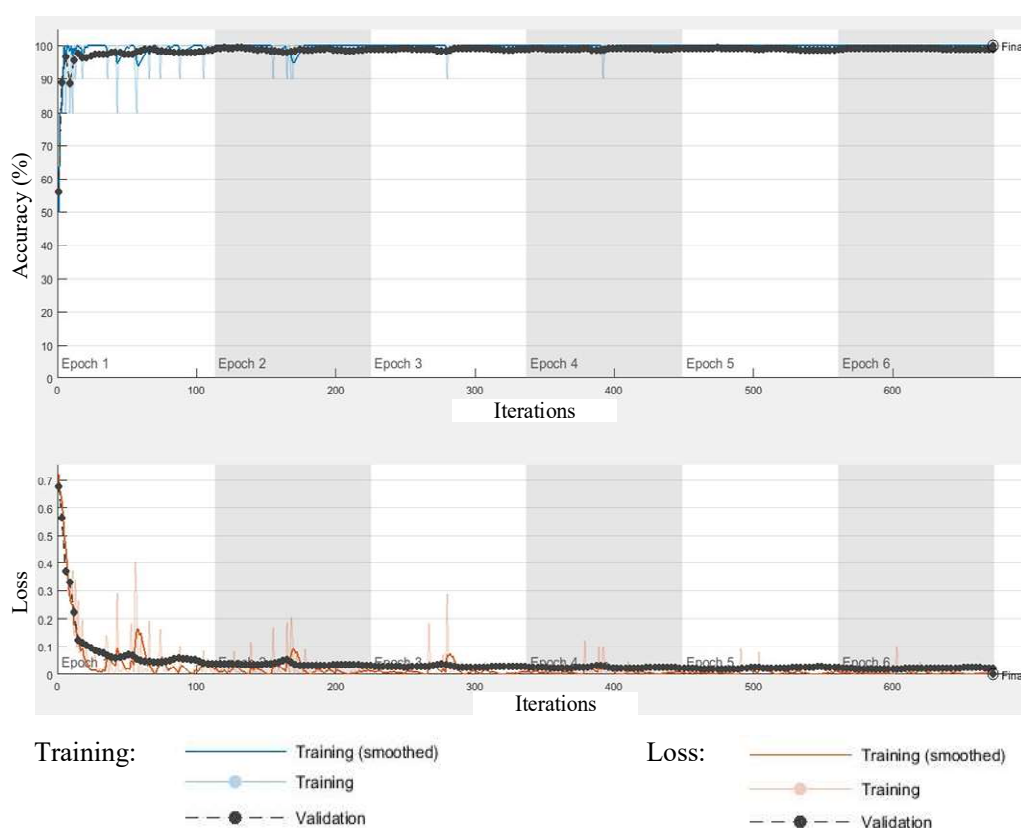
Fig. 4. 12: Training progress of ResNet-101 for assay type detection

4.4 Comparative Performance of the Pre-trained models

For assay type determination, all the pre-trained models mentioned above exhibited 100% accuracy. It is worth mentioning that the accuracy mentioned in Section 4.3 as well as in Table 4. 7, is the final validation accuracy after completing the training.

Table 4. 7: Comparative performance for the assay type detection

Model	Accuracy (%)	Comparative time	Total Size (MB)	Core Model Size (MB)	Network
AlexNet	100	1	621	207.2	Series
GoogLeNet	100	x 6.71	65.6	22.2	DAG
Inception v-3	100	x 22.65	232	79.4	DAG
ResNet-50	100	x 17.72	253	86.2	DAG
ResNet-100	100	x 30.38	456	155.5	DAG

x reads as times, e.g. the training time of GoogLeNet was 6.71 times higher than AlexNet

From Fig. 4. 9- 4.12, this accuracy is identified as the Final point. The training accuracies and losses are shown in Fig. 4. 9- 4.12 were only for the individual mini-batches. The validation accuracy, as well as loss for both inception models and ResNets, are the classification accuracy using the entire validation dataset, which was held-out while initiating the training procedure.

The confusion matrix and statistical output provided the identical result as the models above. Regarding prediction time, a new image can be classified within a fraction of a second. Therefore, these models were evaluated based on the model complexity, convergence time, smoothening and size of the model.

From Table 4. 7, it can be observed that AlexNet required the lowest training time, which is easily understandable that the deeper layers and more complexity of the rest of the models resulted in higher computational expense. The AlexNet converged more smoothly within the earlier epochs than the other pre-trained models. Although the initial layers for rest of the models were chosen to be frozen, the training time for 1600 images (of relatively smaller sized input) was significantly higher than AlexNet.

Comparing the trend of training progress, AlexNet and GoogLeNet were found to be more stable models for the assay type detection using the chosen parameters. The size of the models mentioned in Table 4. 7 contains all the variables including the training and testing dataset. The core model is much smaller in size.

Regarding total memory size, the AlexNet is bigger than the others. As shown in Table 4. 7, the GoogLeNet occupied the least amount of memory space. Unlike the rest of the models, the gradients in AlexNet were computed at each layer during the re-training process, which justifies the larger size. However, the memory size of different networks can be better explained with the number of parameters.

One of the biggest successes of the inception module based network is its capability to operate on multiple sized filters at the same level, effectively making the network wider rather than deeper (Szegedy et al., 2015, 2016). A lower dimensional embedding may contain many useful information about a relatively large image patch. In the inception module architecture (Fig. 3. 8(a), in Chapter 3), 1x1 convolutions are used to compute reductions before the expensive 3x3 and 5x5 convolutions. Due to the network structure performing several small convolutions to drastically reduce the number of parameters, the memory occupancy by the Inceptions was downsized (Table 4. 7). The memory occupancy of the ResNets in Table 4. 7 can be understood from the higher number of parameters originated from the 'network-in-network' architecture of ResNets (Fig. 3. 9, in Chapter 3).

Regarding model complexity, the AlexNet utilised a series network for deep learning, where all the layers are arranged one after the other. Among the 25 layers within the network, there is a single input layer, eight layers with learnable weights consisting convolutional layers and fully connected layers, and a single output layer. The rest of the pre-trained models mentioned in this chapter utilises the DAG network. Therefore, in addition to having longer layers, these model possess a much more complex network than AlexNet.

4.5 Intra-class Classification

The intention behind using pre-trained models was to determine the type of the assay, where the type of the assay signified the class label. In this section, we extended our experiments on the pre-trained models for intra-class classification of the assays. In other words, in Section 4.3, the class was defined as the assay type; in Section 4.5, further classification within the class (i.e. assay type) is provided, which is referred to as the intra-class classification. This intra-class classification is the colourimetric classification that provides a decisive analysis of the present elements in a qualitative or semi-quantitative manner. This experiment aims to evaluate the scope of using pre-

trained models for a complete colourimetric classification, rather than using the framework proposed in Fig. 3.3.

4.5.1 AlexNet for pH Test

At first, the LFA was chosen to be classified for intermediate classes within LFA using the same universal paper strips of pH indicator as a case study. In this case, there were eight classes for pH level ranging from 3.0 to 8.0. Keeping the rest of the model-parameters as given in Table 4. 4, the AlexNet was trained as described in Section 3.5.2, replacing the final layers only. The detail of the final layers is provided in Table 4. 8.

Table 4. 8: Details of the final layers of AlexNet for intra-class classification of pH test

Layer	Name of the layer	Type	Activation	Learnable
23	FC 8 fully connected layer	Fully Connected	1x1x8	Weights 8x4096 Bias 8x1
24	prob softmax	Softmax	1x1x8	-
25	Classoutput Crossentropyex	Classification Output	-	-

The convolutional layers embody many two dimensional arrays. These arrays are called channel. In order to explore deep learning-features by examining the activations within convolutional layers, each of the five convolutional layers was critically inspected. Initial layers of a convolutional network are supposed to learn basic features such as colours and edges, whereas the deeper features should be able to learn complex features. Each of the successive layers structures the features by relating to features of the preceding layers as well.

The first convolutional layers consist of 96 11x11x3 convolutions with [4 4] stride and [0 0 0 0] padding, as shown earlier in Fig. 4. 3. The output of the activations can be represented by a three-dimensional array. In this case, the third dimension indexes the channel of the first convolutional layer. For illustration purpose, an arbitrary image

from the dataset D. LFA. 2 was chosen by random permutation. The class label of the image is pH 7. The image was passed on to the first convolutional layer of AlexNet, and Fig. 4. 13 (a) represents the montage of the images for each channel, by taking it into a four-dimensional shape, where the 3rd dimension in the input to montage is for colours, and fourth dimension is the index of the channel.

(a)

2nd Convolutional Layer

(b)

(c)

4th Convolutional Layer
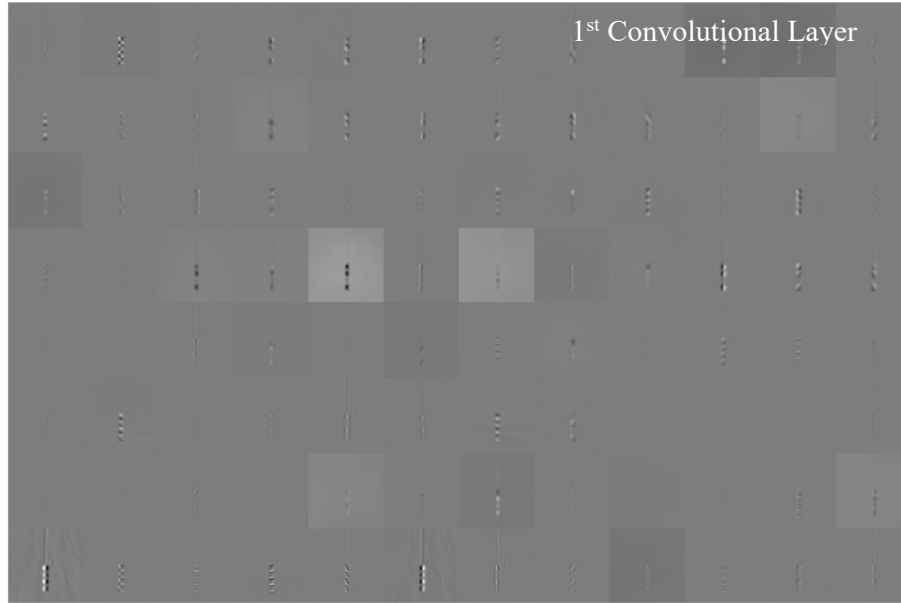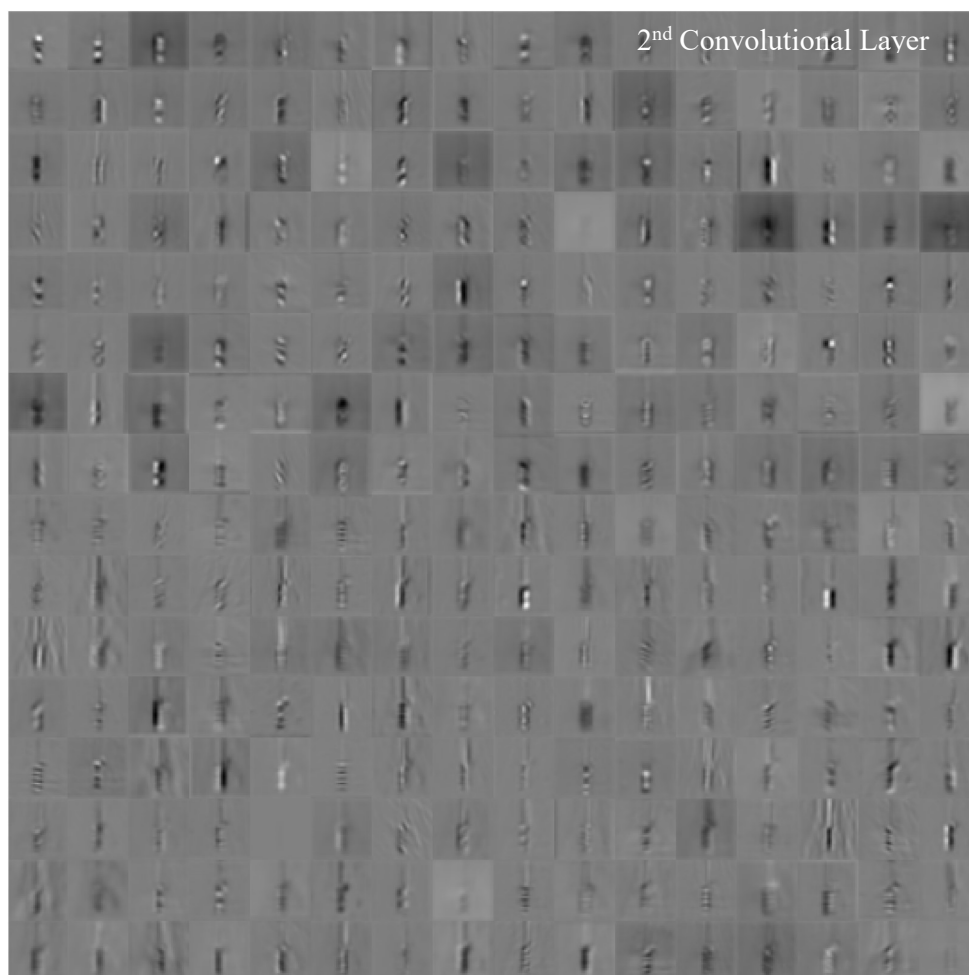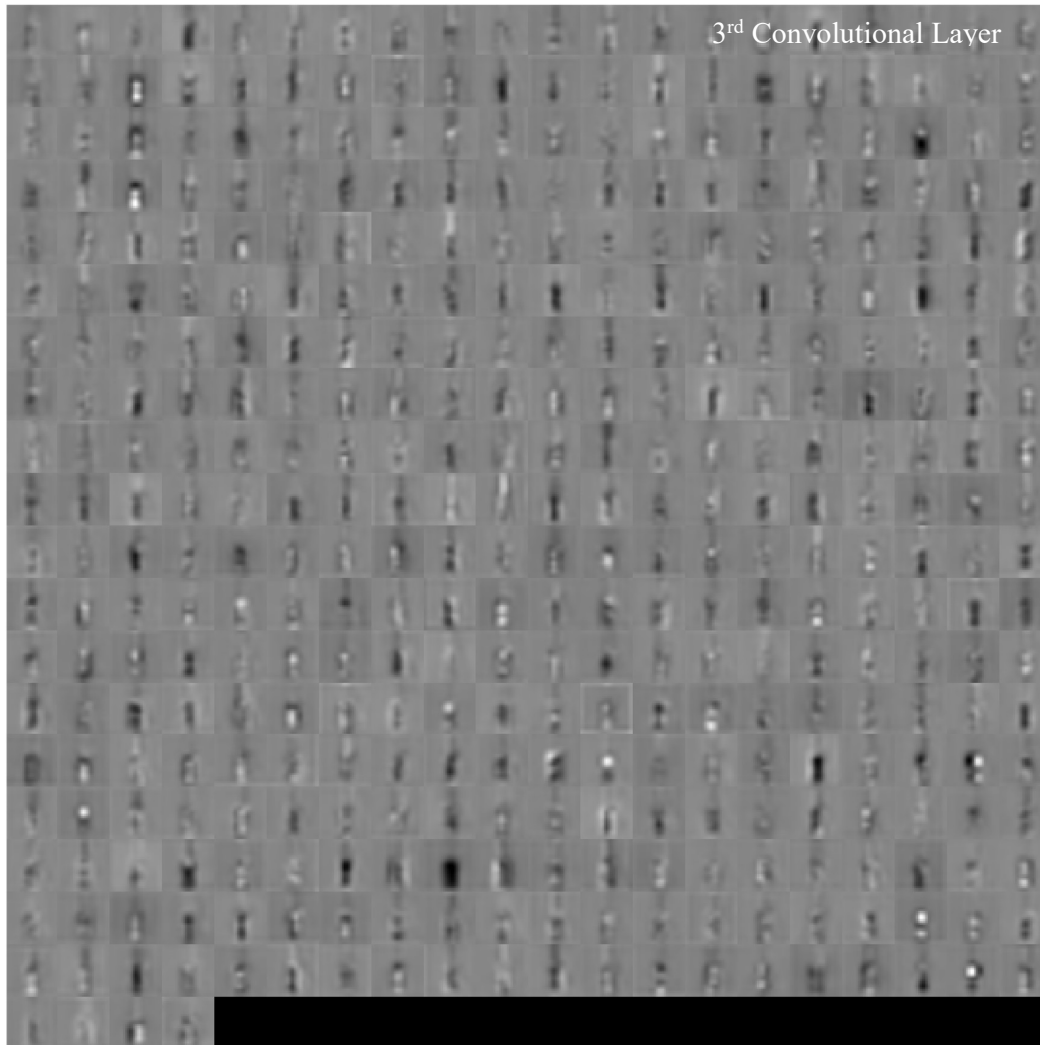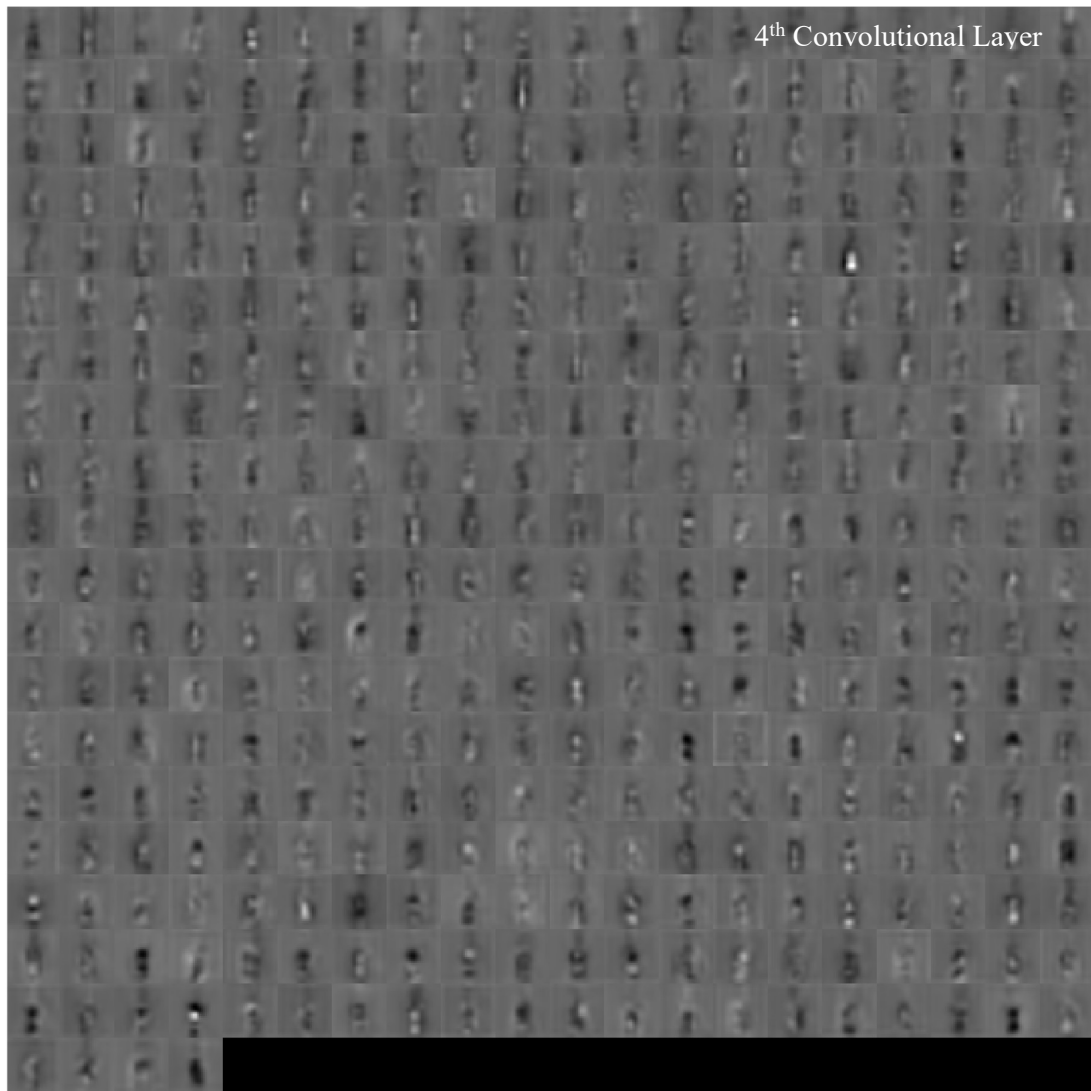
(d)
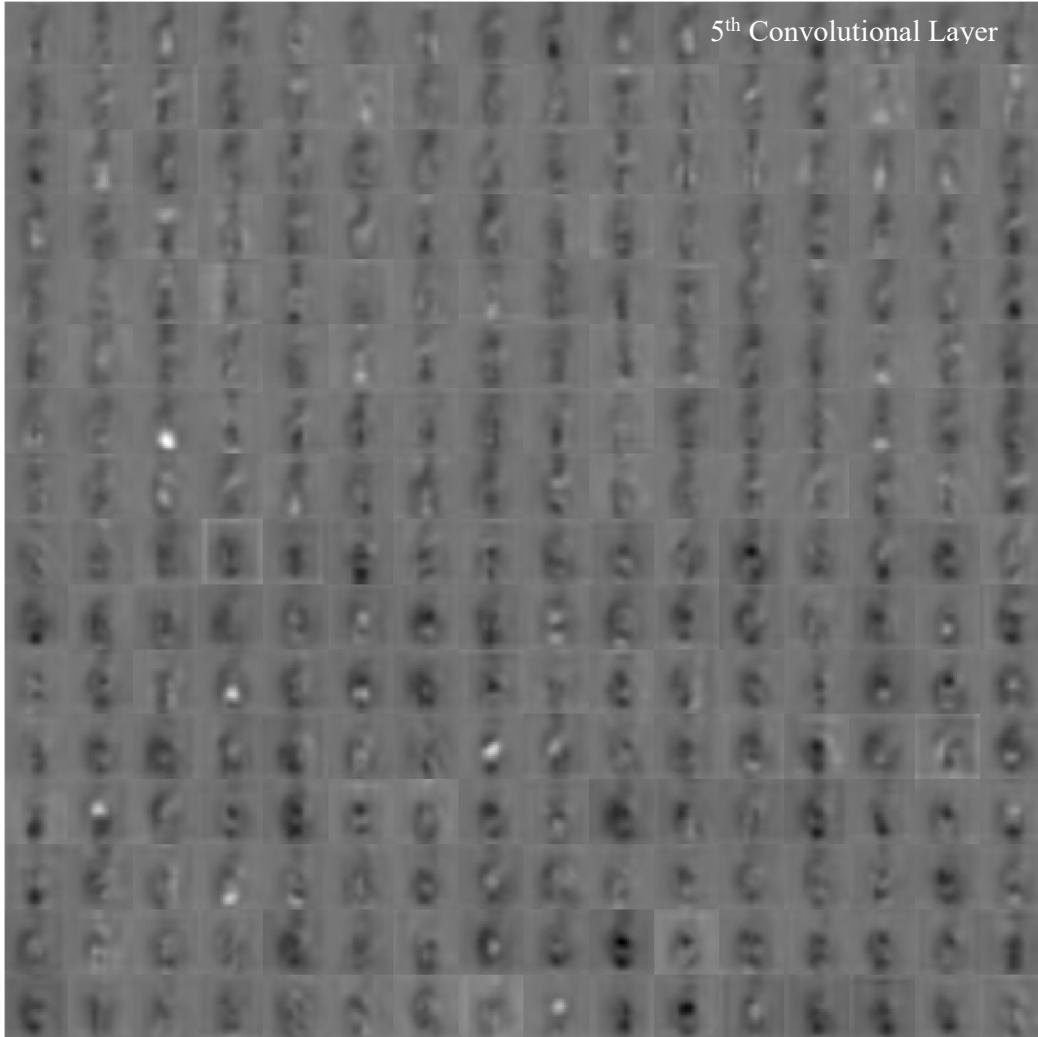
(e)

Fig. 4. 13: Activations of the convolutional layer for pH test. (a) First convolutional layer, (b) Second convolutional layer, (c) Third convolutional layer, (d) Fourth convolutional layer, (e) Fifth convolutional layer

The activations are colour blind. Therefore, the images were normalised and scaled, and the outputs are shown as grey images. Due to the choice of the activation function, in this case, minimum activation=0 and maximum activation=1. Similar to the activations in the first convolutional layer in Fig. 4. 13 (a), each of the convolutional layer along with its channels were investigated as shown in Fig. 4. 13(b-e). The maximum activations in each of the layers of Fig. 4. 13 is separately shown in Fig. 4. 14 for better understanding.
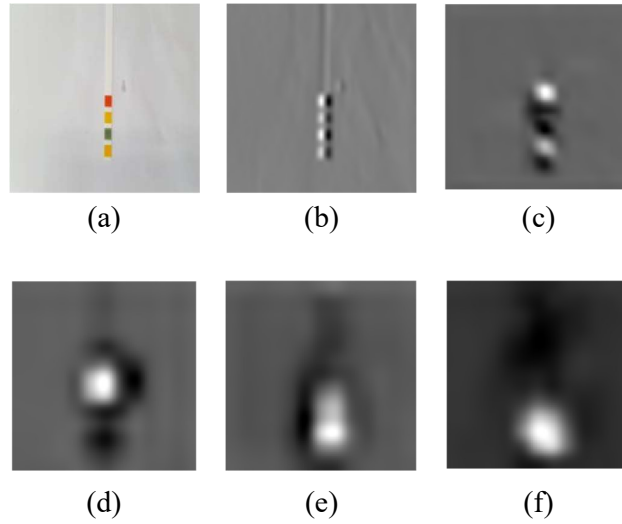
Fig. 4. 14: Comparison of the strongest activations in each convolutional layer of AlexNet. (a) Input Image. The strongest activation in the (b) first, (c) second, (d) third, (e) fourth and (f) fifth convolutional layer

It appears from the Fig. 4. 14 that the most massive activations in the deeper layers may not be the point of interest. The point of interest may have been buried in some other channels. Moreover, the later layers of Fig. 4. 14 is suggesting an indication of both positive and negative activations, which was rectified in the ReLU layers.
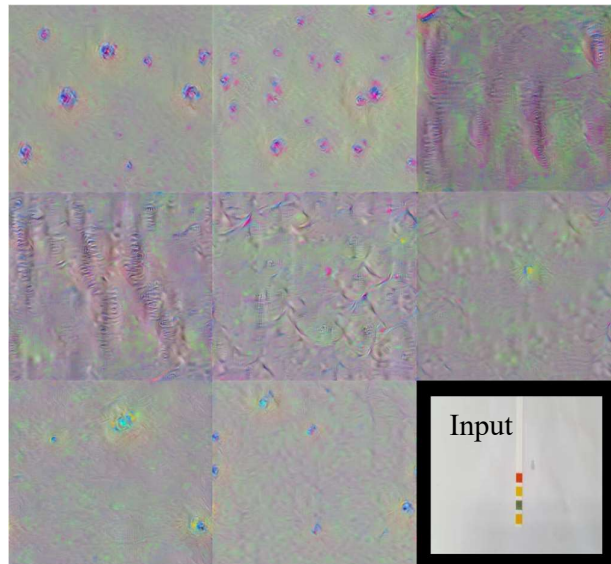
Fig. 4. 15: Visualisation of features at final FC layer for pH test

Using Deep Dream Images (Mordvintsev, Alexander; Olah and Tyka, 2015; Mordvintsev, Alexander; Olah, Christopher; Tyka, 2015; MATLAB & Simulink,

2018a), the output of the 23rd layer is shown in Fig. 4. 15 for each of the class label. The pH test strip was chosen by random permutation. There were eight pH levels; therefore, the last position for an image in Fig. 4. 15 appeared blank. For the convenience of the readers, the blank place is filled with the input image.

Using the same arbitrary image of class label pH 7 as the previous figure, the detailed output images is shown in Fig. 4. 16 that strongly activated the mentioned class label.

(a)   (b)
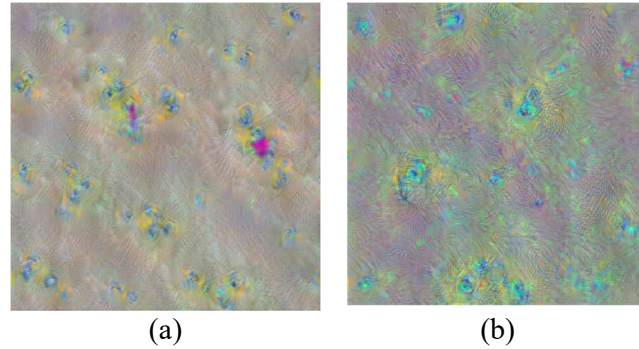
Fig. 4. 16: Detailed image with strong activation at (a) layer 23rd and (b) layer 24th

The larger output was generated by increasing the number of pyramid levels and iterations for each pyramid levels (Fig. 4. 16). To perform a colourimetric classification, these layers were supposed to search for only colours of each block of the pH test strip and block to block colour difference. However, as mentioned in Section 4.3.1, it is difficult to gather any significant information for the classification from these images.

Table 4. 9: Training progress of AlexNet for intra-class classification of pH test

Epoch	Iteration	Mini-batch accuracy (%)	Mini-batch loss
1	1	7.81	2.4580
1	50	62.50	0.8779
2	100	85.94	0.3926
2	150	82.81	0.4963
3	200	89.06	0.2573
3	250	90.63	0.2136
4	300	95.31	0.1276
5	350	89.06	0.2425
5	400	93.75	0.1701
6	450	93.75	0.1719

6	500	87.50	0.2094
7	550	93.75	0.1145
7	600	96.88	0.0713
8	650	96.88	0.1075
9	700	98.44	0.0948
9	750	89.06	0.1822
10	800	96.88	0.0853
10	850	100.00	0.0371
11	900	92.19	0.1696
11	950	95.31	0.0781
12	1000	93.75	0.0914
13	1050	90.63	0.1451
13	1100	98.44	0.0725
14	1150	96.88	0.0544
14	1200	90.63	0.1097
15	1250	96.88	0.0643
15	1300	98.44	0.0386
16	1350	95.31	0.0958
17	1400	90.63	0.1295
17	1450	92.19	0.1311
18	1500	96.88	0.0544
18	1550	98.44	0.0304
19	1600	96.88	0.1187
19	1650	93.75	0.1247
20	1700	93.75	0.0970
20	1740	93.75	0.0721

From the training progress in Table 4. 9, the mini-batch accuracy and mini-batch objective function can be analysed. Within the first epoch, the network managed to learn more than 62% regarding the data.

There were 87 iterations for each of 20 epochs with a constant learning rate of 0.001. The trend of training accuracy and loss is showed in Fig. 4. 17.
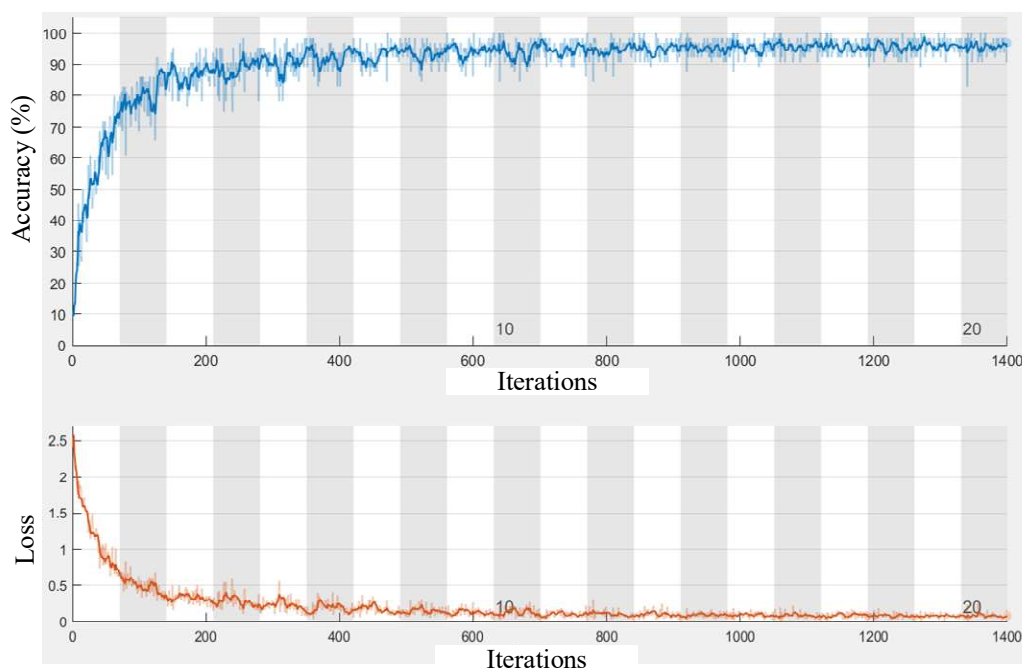
Fig. 4. 17: Training Progress of AlexNet for pH test

The overall accuracy of the model was ~93%. The accuracy of intra-class classification is considerably lower than the classification conducted for assay type detection. For a stable assay such as universal pH indicator paper, the classification accuracy obtained from the pre-trained model of AlexNet is not acceptable. The general perception regarding CNNs is larger dataset produces a better result. Therefore, this chapter explores if the poor accuracy was due insufficient amount of data or some other factors (Fig. 4. 18).

The AlexNet required only 15 minutes and 42 seconds while training on 1600 images for the assay type detection (Section 4.3.1). Using 8000 images, the pre-trained model of AlexNet, consisting series network, was trained by a considerable amount of time. Further exploration of the computation time of AlexNet confirmed that larger dataset takes longer time (Fig. 4. 18). Therefore, it was necessary to explore the impact of the size of the dataset on AlexNet. The dataset D. LFA. 2 utilised for pH test is already resampled. Hence, the investigation of the impact of the size of the dataset did not involve any further resampling.
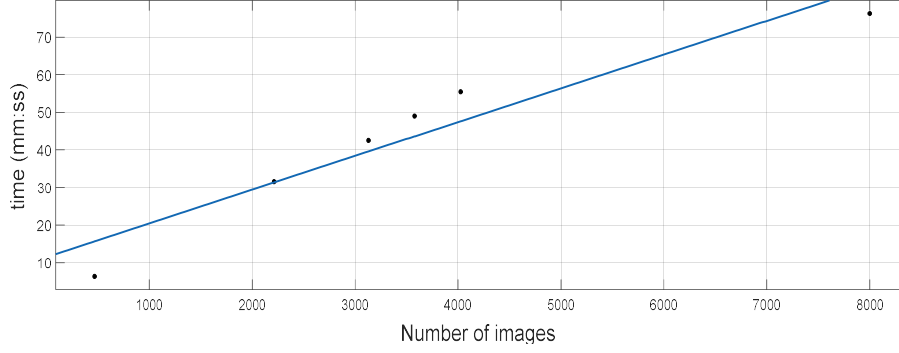
Fig. 4. 18: Increase in elapsed time with respect to increment in data size

The model trained on the original dataset consisting 520 images (Table 4. 2), took 6 minutes and 34 seconds as shown in Fig. 4. 18. The dataset containing 3,160 images (D.LFA.4), took 31 minutes and 54 seconds. In order to generate more data-points to attain an interpolation of the trend of computational time with respect to the number of images, the dataset containing 4,472 images (D.LFA.3) was is shown with training and testing data-partitioning of 70:30 (42 minutes and 52 seconds), 80:20 (49 minutes and 02 seconds) and 90:10 (55 minutes and 43 seconds). The experimental analysis shows how the computational time is affected by the number of training dataset, even if the total dataset remains the same.

■ original (70|30)	■ original (80|30)	■ original (90|30)
■ D. LFA. 3 (70|30)	■ D. LFA. 3 (80|30)	■ D. LFA. 3 (90|30)
■ D. LFA. 2 (70|30)	■ D. LFA. 2 (80|30)	■ D. LFA. 2 (90|30)
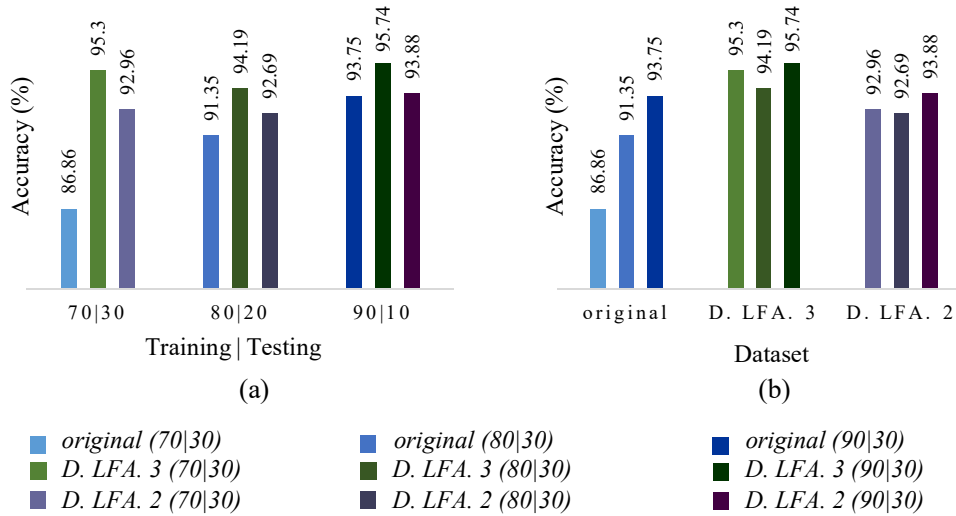
Fig. 4. 19: Performance of AlexNet for pH test. (a) Ratio of training and testing datasets, (b) With respect to datasets

In addition to the computation time, a detail inspection of the impact of the size of the dataset is shown in terms of classification accuracy. When the model was trained on the

original dataset consisting of 520 images, the classification accuracy was below 90% (Fig. 4. 19).

With an 8.6 times larger dataset, the accuracy increased to 95.3%. On 8000 images, the accuracy decreased to ~93%. One possible explanation of this degradation of performance could be resampling technique itself. The LFA dataset is consistent in terms of samples per class, exposure of the plane and image capturing technique. As explained in Section 4.2.1, resampling was used to feed the network a bigger dataset. Instead of using an exact duplicate, the resampling technique involved image enhancement techniques and induced noise (D.LFA.2). Although Goodfellow, Bengio and Courville (2016) showed that adding computer-generated noise to the training dataset can improve the overall accuracy of the model, because in this way the model is exposed to a more adversarial scenario and more generalisation can be achieved, the experimental result obtained in this section was not consistent with the claim. Therefore, a bigger dataset without excessive resampling may improve the classification accuracy of AlexNet.

Another noticeable factor regarding the size of the dataset was revealed by varying the proportion of training and testing data. This variation was created from changing the parameters of augmentation. The data augmentation by creating batches of training, validation, test, and prediction data, aids the model to deal with overfitting. It also prevents the model to memorise the exact specifications of the training images. The ratio was varied from 70:30 to 90:10. In Fig. 4. 19(b), the classification performance on the original dataset reveals prompt increase with an increase in the training dataset. On the larger datasets, the larger proportion of training dataset had a slower impact. This phenomenon indicates potential saturation of the classification accuracy, even if the dataset is further extended.

From the critical analysis of the features of each layer of AlexNet and the training progress from Table 4. 9 and Fig. 4. 17, another explanation of poor classification accuracy for pH test could be the training cycles. The number of iterations and epoch were inadequate to provide an accurate colourimetric classification for the pH levels. It is already mentioned that Fig. 4. 19 suggests a slower increase in the accuracy with the higher number of the dataset, which could be due to the resampling technique. In addition to the need of significant amount a larger dataset, Table 4. 9 indicates a need

for longer training cycle which would effectively require more enhanced resources, e.g. higher computational power[11], which contradicts with our research goal.

4.5.2 Rest of the Models for pH test

This section presents an investigation for the rest of the models mentioned in Table 4. 7 to explore the possibility of better classification accuracy using a pre-trained model. Unlike Table 4. 7, the performance of GoogLeNet degraded for pH test. It provided only 67.58% accuracy on D.LFA.2 as shown in Fig. 4. 20.
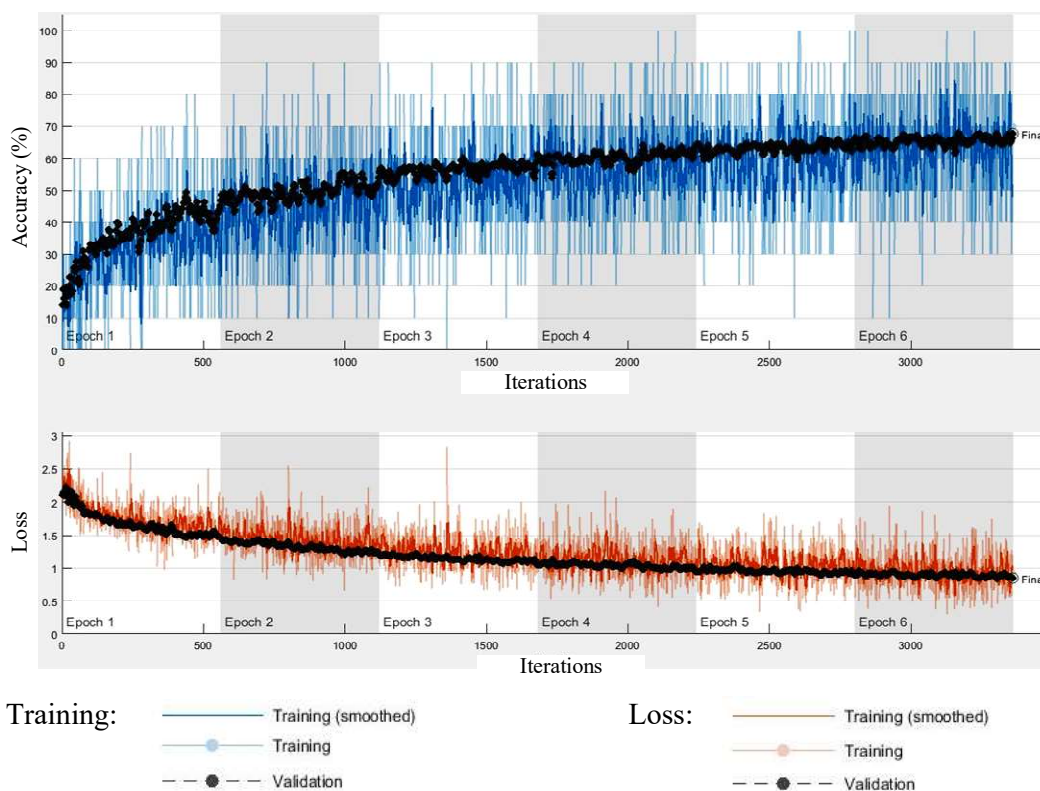
Fig. 4. 20: Training progress of GoogLeNet for pH test

There were six epochs and 560 iterations during epoch (Fig. 4. 20), which resulted in a total of 3360 iterations. The learning rate was kept as same as for the assay type detection, i.e. 0.0001. Based on the trend of the convergence, more extended training

---

[11] Another alternative would be developed model approach, which would be contradicting our research aim due to more computationally expensive scheme.

period probably would have provided a better accuracy, which was avoided to prevent over-training.

With the same number of epochs and iterations as GoogLeNet, Inception v-3 provided 76.17% accuracy to classify eight pH levels as shown in Fig. 4. 21.
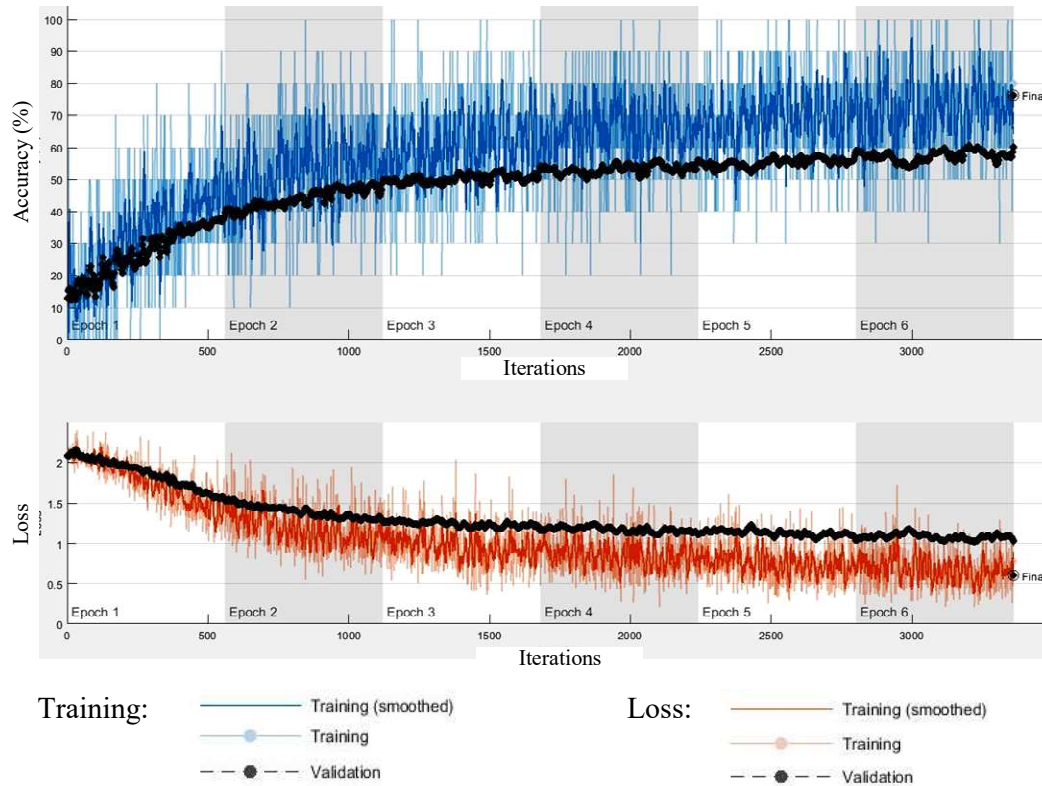
Fig. 4. 21: Training progress of Inception v-3 for pH test

The performance of Inception v-3 was slightly higher than GoogLeNet, however considerably lower than AlexNet and its own performance in Table 4. 7. Based on the convergence, it is difficult to say if only a longer training cycle and a bigger dataset would have provided a high accuracy or not.

From AlexNet to Inception v-3, the layers grew deeper for each of the models. One may wonder if only the deeper layers are the solution for every problem. He et al. (2015) discussed in their paper on ResNet that because of the vanishing gradient problem, the model weights of the initial layers are unable to be updated adequately using backpropagation of the error gradient.

In an ANN using gradient-based learning and backpropagation, the weights of each network receive an update. This update is proportional to the partial derivative of the error function corresponding to the weight for the current iteration, which continues for each iteration during training. For some cases, the gradient can be 'vanishingly' small. This phenomenon can prevent the weights to be updated. He et al. (2015) was motivated to retain the gradient, which is one of the key advantages of ResNet. Using the pre-trained model of the ResNet-50 on the D. LFA. 2, 76.17% accuracy was attained for pH level determination (Fig. 4. 22).
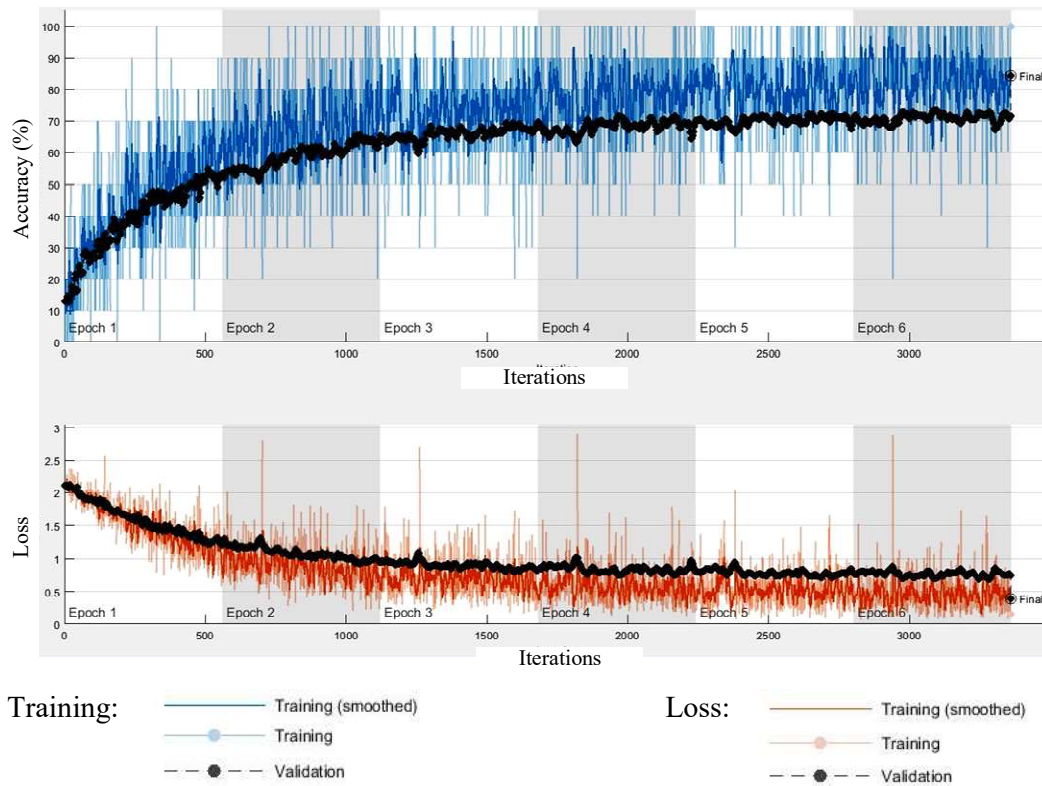
Fig. 4. 22: Training progress of ResNet-50 for pH test

The ResNet-50 evidently showed better performance than the inception modules on D.LFA.2. The computation time was higher than Inception v-3 but lower than GoogLeNet. The ResNet-101 showed slightly better performance on D.LFA.2 than ResNet-50 (Fig. 4. 23).
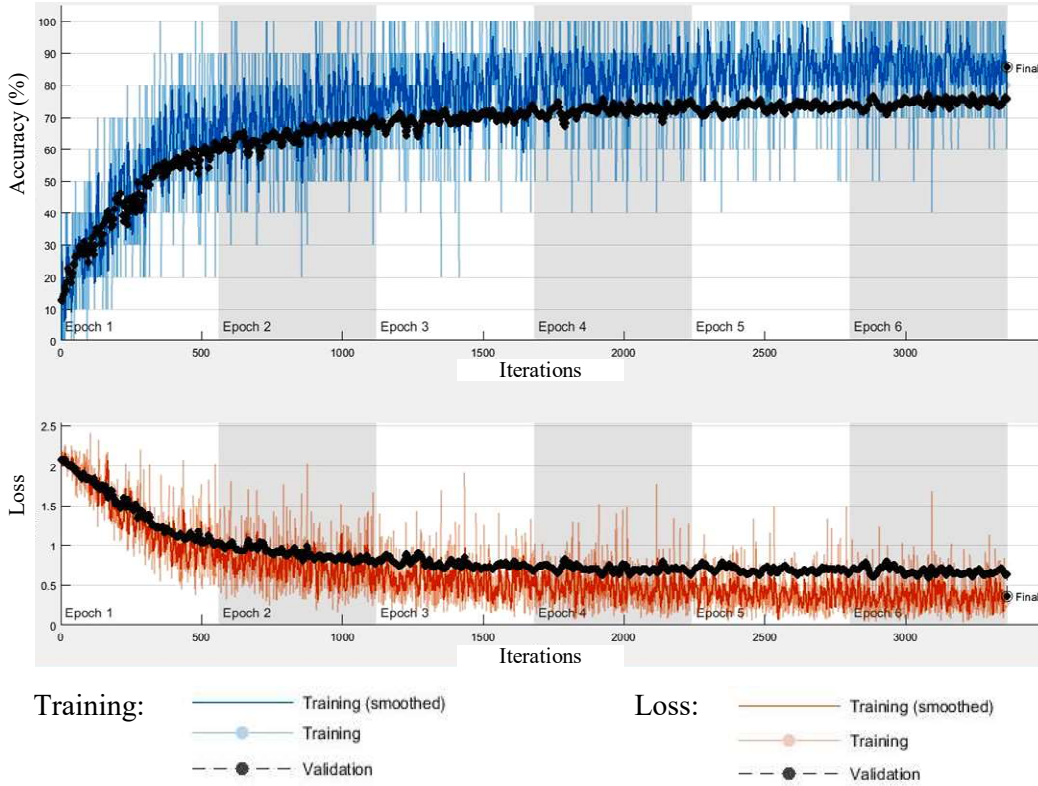
Fig. 4. 23: Training progress of ResNet-101 for pH test

The computation time for ResNet-101 was significantly higher than any other models. However, the accuracy was still low (67.11%). Without more fine-tuning the model including data cleaning (Reed et al., 2015), incremental feature construction and hyperparameter tuning such as cyclical learning rate, activation function and more experimenting with weight initialisation, which may require more resources, it is difficult to claim if a longer training period can assist in the convergence rate of the ResNet-101.

Performance of these pre-trained models with 70:30 training and testing data ratio can be summarised as Table 4. 10. The model sizes were similar to Table 4. 7. Therefore, the discussion regarding memory occupancy is not repeated for D.LFA.2.

Table 4. 10: Comparative performance for the assay type detection

Model	original		D.LFA.2	
	Accuracy (%)	Comparative time	Accuracy (%)	Comparative time
AlexNet	86.86	1	92.92	x 12.03
GoogLeNet	50.00	x 2.08	67.58	x 383.66
Inception v-3	55.26	x 7.46	76.17	x 1216.75
ResNet-50	63.82	x 5.58	84.42	x 996.15
ResNet-100	67.11	x 10.59	85.58	x 1633.04

x *reads as times, e.g. the training time of GoogLeNet is 2.08 times higher than AlexNet*

In Table 4. 10, the accuracy of pH level recognition or determination was higher for AlexNet than the other pre-trained models. Rest of the models with deeper networks with frozen layers are insufficient to provide the colourimetric classification using the chosen parameters. For a better illustration of the accuracy achieved using original dataset and the final dataset containing resampled images, Fig. 4. 24 is separately showing the accuracy per model for two datasets, where the trend can be better perceived.
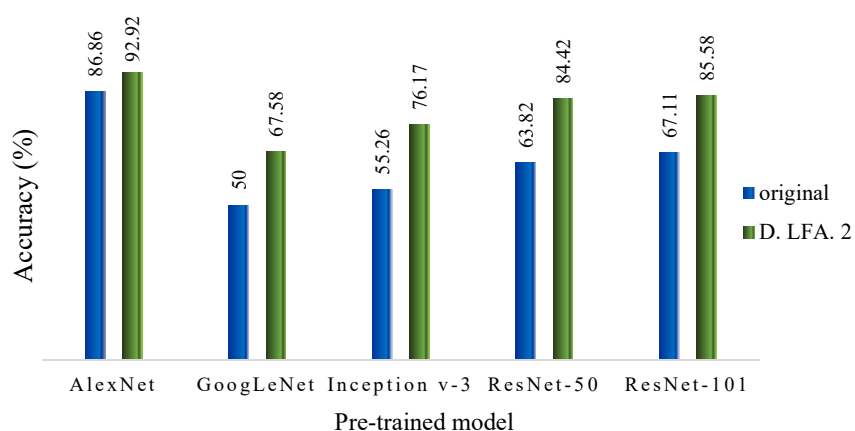
Fig. 4. 24: Performance of the pre-trained models with respect to datasets for pH test

The comparative performance of the models trained on 520 images of the original dataset and 8000 images of D. LFA. 2, illustrated in Fig. 4. 24, suggests the scope of improved performance from Inception modules and ResNets using a bigger dataset and

longer training period. Based on the comparative performance (Fig. 4. 24) and the trend of convergence, these models are more capable of learning from these resampled data than AlexNet. However, in order to preserve 'rarity' (Weiss, 2004), the dataset was not further resampled. Moreover, these models were computationally more expensive even with the frozen layers.

4.5.3 Intra-class Classification for ELISA

The intra-class classification was conducted for LFA only, not ELISA due to the following reasons.

- For proof-of-concept, we found only one intra-classification would be sufficient because the intra-class classification accuracy obtained by the pre-trained models was unsatisfactory. Therefore, the approach would have had to be rejected for the final colourimetric classification, despite the outcome of the case study of ELISA.
- The case study of LFA is a stable assay. The dataset possesses consistency. A large number of images were fed to the network. The classification accuracy was still poor. The case study of ELISA contains less stable samples. Therefore it would be more difficult to understand and explain the underlying reasons of the hypothetical poor performance of the pre-trained models.
- The resampling techniques assisted in generating two times larger volume of data for LFA. The final dataset contained 1000 images per class (D.LFA.2). The analysis of the classification accuracy suggested the requirement of samples with less resampled data. The ELISA dataset is already smaller than the LFA dataset. Therefore, the number of images would be insufficient for these data-hungry pre-trained models.

4.6 Summary

The determination of assay type can aid a less medically trained person to begin the colourimetric test without any prior knowledge about the nature of the assay. It can also assist to initiate the rest of the framework based on the output of the pre-trained model. Moreover, the variation regarding size, shape and colour of the assay by different commercial brands (Fig. 3. 5) can be mitigated to certain extent using these deep neural networks.

In the overall framework, the assay type determination is indicated as an optional step. Considering one of the focus of the research being resource-limited settings, the present state of the processing capacity of the mobile phones would require the pre-trained models to utilise a server-based approach.

Based on the analysis of the result obtained by the pre-trained models, considering the training time, AlexNet would be a better option. Moreover, utilisation of the series network makes the model less complex as well. However, a trained system does not require to be trained each time when a user provides an image to be tested. Therefore, considering the requirement of memory size, GoogLeNet can be chosen for initiating the rest of the framework based on the output regarding the assay type.

The performance of the models was evaluated for intra-class classification as well. The result indicates that these pre-trained models are more suitable for assay type detection rather than binary or multi-class classification within the assay type. The geometric shape of assay and location of the coloured samples are easily distinguishable for these pre-trained models using the standard kernels with a reasonable amount of dataset. However, the intra-class classification would require the model to search for only the colours even when the rest of the geometric features are similar, and occurrence of those colours are in the same location. Therefore, it is more logical to use a simpler machine learning model for the colourimetric classification, instead of building more deep layers which would require more processing capacity, memory size, larger dataset and more dependency on the cloud-based approach.

Chapter 5

Experiments and Results: Case Study 1

5.1 Introduction

This chapter explores the data structure and number of algorithms to provide an automatic colourimetric decision for the wet-chemical-based assay. The case of Gold Nanoparticles (AuNP) based plasmonic enzyme-linked immunosorbent assay for tuberculosis antigen-specific antibody detection is the main focus of the chapter. The case study is referred to as TB-test in the thesis, as mentioned earlier in Section 1.5 (Chapter 1). This chapter sequentially discusses the impact of the case study, several methods to solve the problem using existing knowledge of computer vision and machine learning, and based on the performance of several alternative approaches, this study outlines an optimal solution for the presented problem.

5.2 Background of the Case Study

TB is a communicable disease, infecting one-third of the world's population. In 2015, 1.8 million TB-related deaths were reported (Centers for Disease Control and Prevention, 2017). On the other hand, every year about 244 million migrants cross international borders (Department of Economic and Social Affairs, 2016). The carriage of TB in a mobile population is a global challenge, which is a particular concern for the border agencies (Posey, Marano, & Cetron, 2017). However, TB is curable with appropriate early diagnosis. The most common diagnostic procedure for TB is a skin test[12] or a blood test[13] (Centers for Disease Control and Prevention.; NHS). Despite many commercial test schemes, there is still a need for an easy-to-use, effective and feasible point-of-care TB diagnostic tool, particularly for the remote community where there are very limited or no diagnostic facilities. Such a tool should possess the following features: low-cost mobile solution, anytime anywhere access, low energy consumption, ease of use, fast and automatic identification of TB.

---

[12] Mantoux test
[13] Interferon-Gamma Release Assay (IGRA)

There are commercial and endorsed mobile applications for TB in the popular mobile application stores (Appendix C). When it comes to diagnosis, the applications are for screening purposes only. The available mobile applications can ensure the data portability and can be used for the diagnostic decision. However, they lack automation to produce a diagnostic result from the specimen. Thus, there is a need for a system that does not require any additional hardware (e.g. microplate photometer) and can produce laboratory scale test results.

To the best of our knowledge, there is no existing mobile, desktop or server-based system for plasmonic ELISA based detection of TB antigen-specific antibodies. In the literature, only a few studies employed machine learning techniques to assist in the diagnosis and monitoring of TB to offer a low-cost, simple, rapid and portable platform. For example, Tracey et al. (2011) utilised acoustic signals to track the recovery of pulmonary tuberculosis patients. The MLP showed 88.2% accuracy for ambulatory cough analysis.

Osman, Mashor, & Jaafar (2010) proposed a tuberculosis bacteria detection technique from the tissue sample by Ziehl-Neelsen staining method. The prepared sample image from an optical microscope was segmented by k-mean clustering for tuberculosis bacteria extraction. Both RGB and C-Y colour were utilised to acquire a robust and improved segmentation under various staining conditions. The hybrid MLP selected the features among the geometrical features of Zernike moments to detect tuberculosis bacteria. The result showed 98.0%, 100% and 96.19% of accuracy, sensitivity and specificity respectively to find the class of definite and possible TB.

Sutherland et al. (2016) presented an affordable and easy-to-use system for TB diagnosis on POC platform. Although the system is not image based, the system is based on LFA. The system aimed at a primary indication of TB and similar respiratory diseases. The classification accuracy was only 92%, hence it heavily relies on a secondary confirmatory test.

Tsai, Shen, Cheng, & Chen (2013) developed colourimetric sensing using unmodified gold nanoparticles and single-stranded detection oligonucleotides for a TB test. The focus of the work was salt-induced AuNP colourimetric diagnosis for sensing target TB DNA sequences without multiple PCR cycles to amplify specific MTB target DNA

sequences from extracted sputum or tissue samples. A smartphone was utilised just to collect the multiple detection results of colour variation from the concentration on cellulose paper and transmit the data to the cloud.

Most of the articles mentioned above do not involve colourimetric test, even if they comprise intelligent systems. Therefore, in the absence of a similar technology to diagnose TB, this chapter does not include any comparative analysis with the proposed system.

5.3 Reaction Phase and Time-dependent Approach

This section explores the first research question of this thesis using the case study of TB-test.

5.3.1 Video Frames Acquisition

The case study of sandwich ELISA presented in this chapter is an end point assay. The reaction time of TB-test was typically from 14 to 50 minutes (N.A. Yusof, 2018; Tania et al., 2017). The reaction time and the colours produced can be affected by geo-location and climate. For a conventional naked-eye measurement, a single measurement is performed after a fixed incubation period. In this work, we have taken a novel approach to monitor the entire reaction to provide a computer-aided classification, instead of considering only the end point.

Regarding the biochemical protocol, a batch of 96-well plates went through a series of processes as illustrated in Fig. 5. 1. However, the focus of this thesis is not the biochemical components.
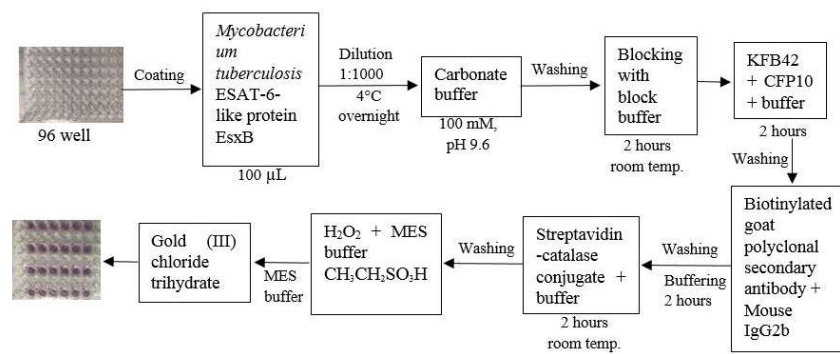
Fig. 5. 1: Stepwise plasmonic ELISA based TB-test

Three separate experiments were conducted on three sample plates (Pl_1, Pl_2, Pl_3) and all of the experiments were recorded. There were 31 samples in total- 1 in Pl_1, 6 in Pl_2 and 24 in Pl_3. The videos were recorded using iPhone 7 plus (12MP, wide-angle: $f/1.8$ aperture, telephoto: $f/2.8$ aperture) and iPhone 4 (5MP) in laboratory environment in the Universiti Putra Malaysia. The videos were converted from MOV to JPEG based images using a conversion software taking time interval (t_i) as 0.0332 (every frame), 1, 5, 10, 20 and 50 seconds. The conversion along with the processing can be performed by direct video acquisition using the MATLAB toolbox as well.

The plasmonic ELISA based TB-test shown in Fig. 5. 2(a), (b) and (c) were conducted in 14:26, 13:10 and 14:09 minutes respectively. The video frames shown in Fig. 5. 2 were sampled at ten seconds interval. The readers can visually track the dynamic changes in colour for all three plates in Fig. 5. 2. However, interpretation of the colour can be subjective. In Fig. 5. 2, the blank frames appeared as black.
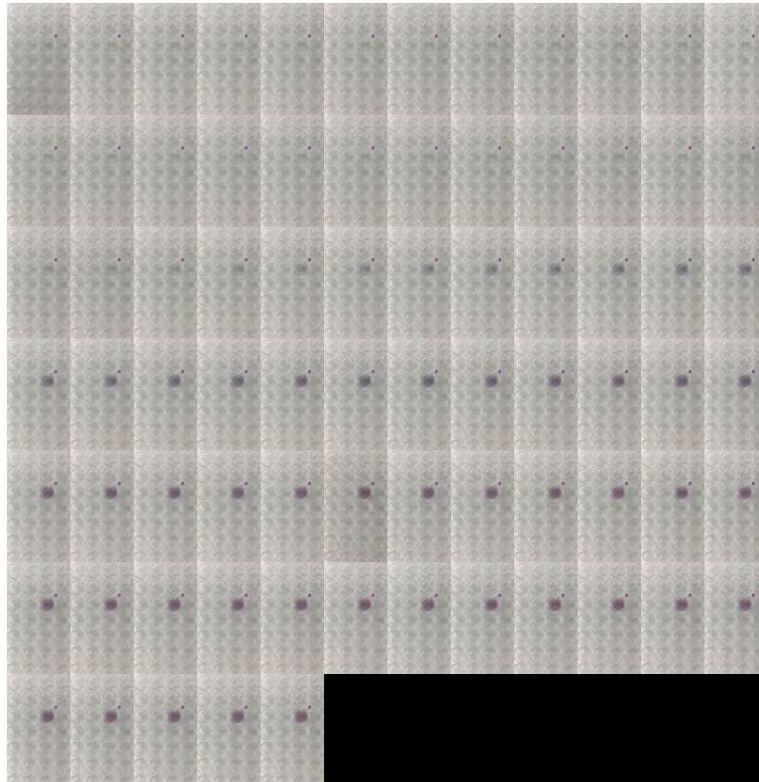
Original video: 14:26 min; Frame width: 1920; Frame height: 1080;
Frame rate: 29.98 frames/ second. Plate contains only one sample. Initial frames do not show
any visible colour.

(a)
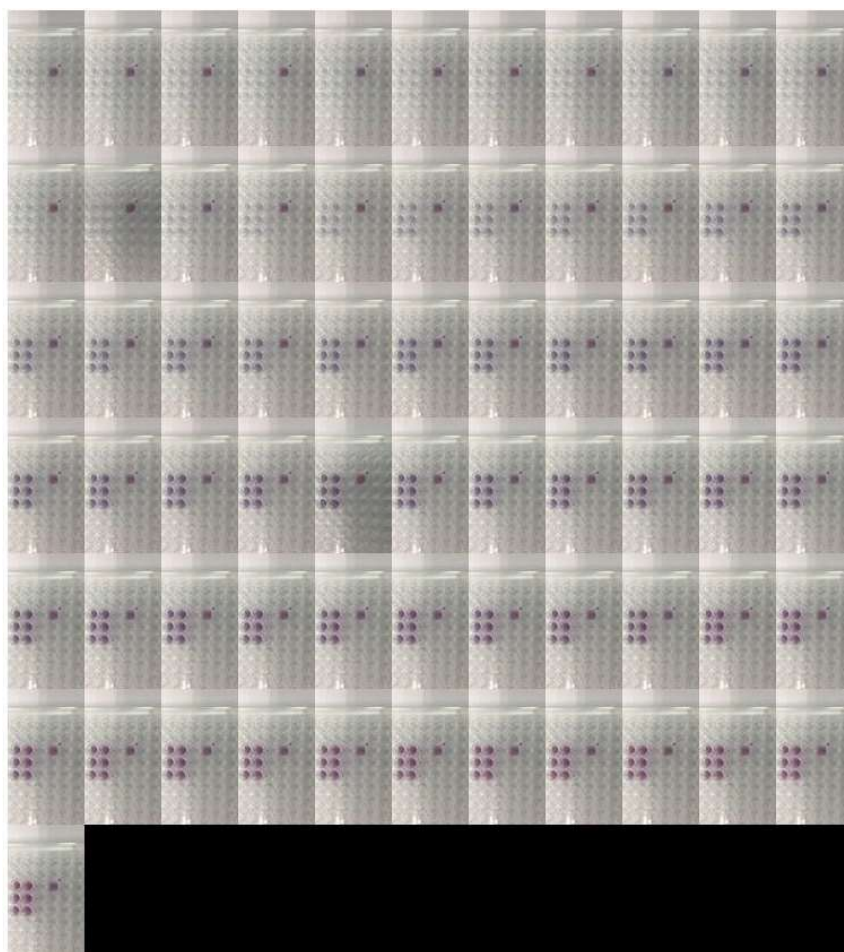
Original video: 13:10 min; Frame width: 1920; Frame height: 1080;
Frame rate: 29.98 frames/ second. Six more samples were added to the plate shown in (a).
Therefore, initial frames only show the sample from (a).

(b)
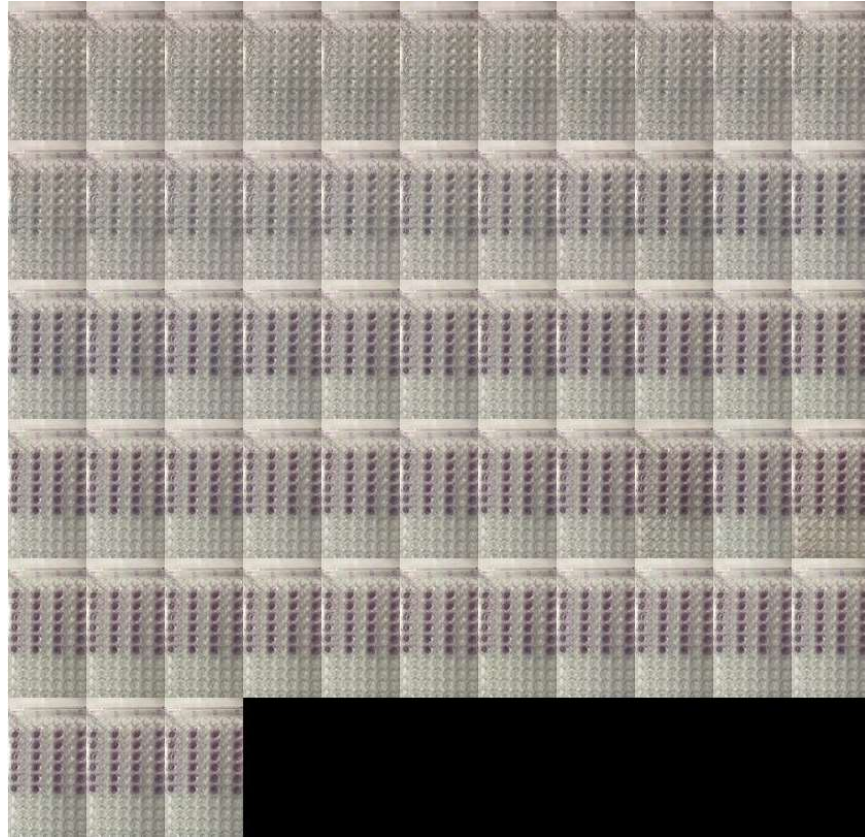
Original video: 14:09 min; Frame width: 1920; Frame height: 1080;
Frame rate: 30.00 frames/ second. Plate contains twenty-four samples. The concentration
varies in each column of the assay plate.

(c)

Fig. 5. 2: Video frames of overall reaction sampled at 10 seconds interval (a) Pl_1 , (b) Pl_2 and (c) Pl_3

The frame rate per second signifies the quality of the video. Due to the involvement of the number of devices, there is a minor variation in the quality of the videos, which can be perceived from the frame rates shown in Fig. 5. 2.

5.3.2 Image Segmentation

For a sample plate containing a mixture of the positive and negative specimen, one would require to perform image segmentation before classification. In Fig. 5. 2, other than Pl_1 , the plates contained multiple samples. Therefore, utilising Eq. 3.5, the images were segmented for each instances.

At first, a mask was created using the colour difference of initial and the end point images (Appendix E). This mask served the same purpose as the virtual plates described in Section 2.3.3 (Chapter 2). The images at each instance were segmented using this mask, followed by separating each sample as the smallest rectangle of 1xQ, containing the ROI, where Q is the number of dimensions of the image. However, the mask is created using the experiment itself. Therefore, it provides more flexibility over the virtual plates regarding the geometric features of the assay plate.

Another advantage of this segmentation technique is its strength to segment the samples with lower colour concentration[14]. Therefore, the segmentation technique can be used to segment the whole assay plate without taking any additional time. On the other hand, this technique would compel the system to segment all the frames defined by the interval, which would lead towards more memory requirement and higher processing capacity. The impact of the power storage of the personalised device would also play its role.

5.3.3 Classification

All of the assay plates in Fig. 5. 2 contained only one class. In order to determine the class of the sample, Eq. 3.6 can be expressed as Eq. 5.1 (Tania et al., 2017).

$$\phi_{TB} = \begin{cases} 1, & TB - ve \\ 2, & TB + ve \end{cases}, \text{ where } \phi_{TB} \neq f(\alpha) \tag{5. 1}$$

Without any stopping agent, the positive sample exhibits two colour transitions (ϕ=2), whereas the negative sample gradually turns into pink without any intermediate colour transition phase. For a better illustration of the concept, the colour transition is shown in Fig. 5. 3. A hypothetical line is drawn for TB negative (Fig. 5. 3). The TB positive sample initially shows a gradual transition towards blue, and then eventually segregates to pink, if there is no stopping agent. $\Delta\phi$ is the time taken for the specimen to turn from blue to pink.

---

[14] For example, Fig. 2. 2(c). This chapter focuses on the TB-test. Therefore, such segmentation result is not included in this chapter to avoid confusion. However, the issue was highlighted in Section 3.4 (Chapter 3) as well.
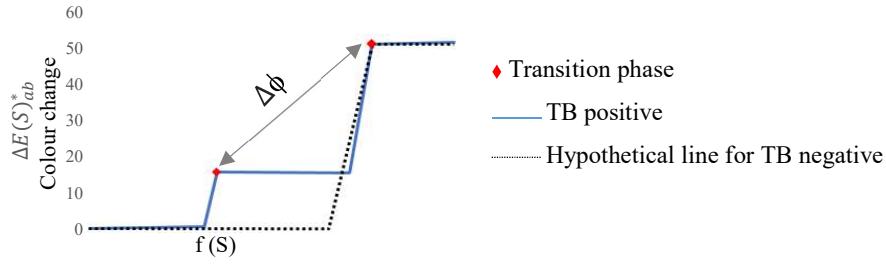
Fig. 5. 3: Concept of colour transition for both positive and negative samples

In Fig. 5. 2(a), there is only one sample. Therefore, the process did not require any segmentation to provide the classification. The same plate was filled with six more new samples, which is denoted as Pl_2 and shown in Fig. 5. 2(b). Both (b) and (c) contain more than one sample, hence required segmentation. At first, the concept is explained with Pl_1 for its simplicity.

For Pl_1 with t_i =10 seconds, the total number of images, f(S) = 86. Eliminating the number of images where the wells were being filled, the number of images was downsized to 77, which are shown in Fig. 5. 2(a). Let us denote it as $f(S_{new})$= 77. Ideally, the video should be recorded after filling the wells with samples. Therefore, the instances should be S_{new}. The colour difference for Pl_1 at each instance is shown in Fig. 5. 4(a).

At each instance, there is 'some' colour change traced by the system. Hence, JND required to be taken into account, which assisted in eliminating a certain part of the noise using Eq. 3.3. One of the conventional value for JND is 1.

As explained in Section 3.4 (Chapter 3), the colourimetric classification will depend on Eq. 3.6, and on Eq. 5.1 in the case of TB-test. Therefore, changes in Fig. 5. 4 (a) below the red reference line (JND), occurred mainly due to changes[15] in the intensity can be considered as noise, and was rectified before further computation.

---

[15] This is a result of change in the colour concentration, not the chroma. Please see Fig. 2. 4(b) in Chapter 2.
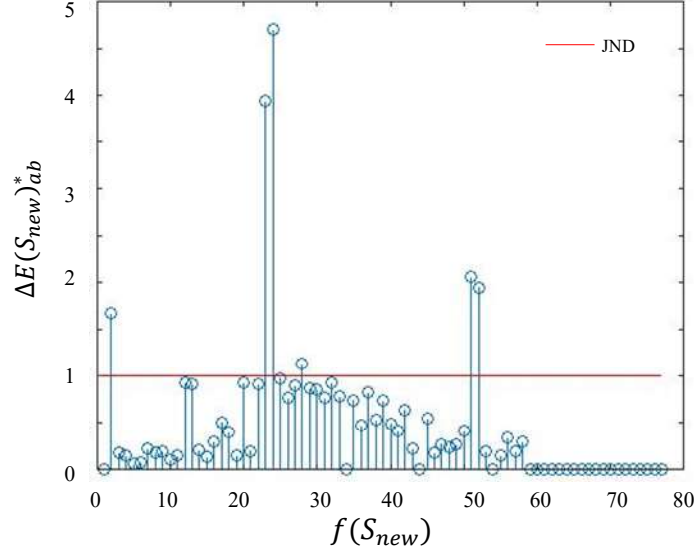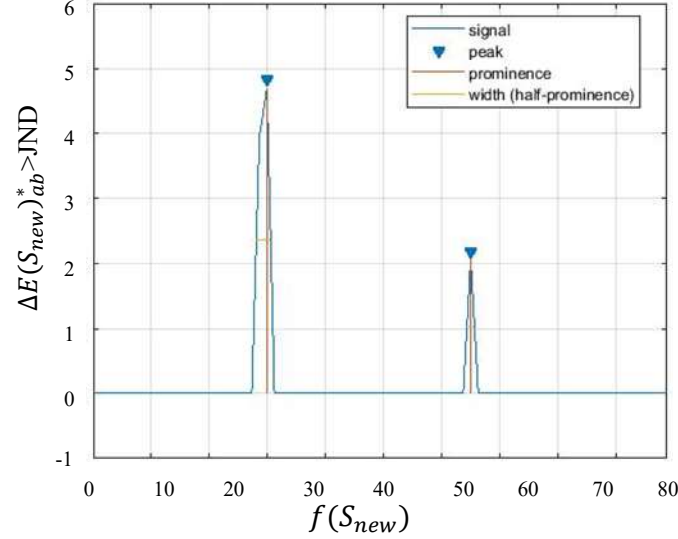
Fig. 5. 4: Dynamic changes in the colour of the images. (a) Time response analysis for Pl_1 with t_i = 10 seconds, (b) Peak detection after considering JND

During the colour transition period, there were number of instances where the change was apparent in the video frames of Fig. 5. 4 (a). During the first transition phase, there was an abrupt increase at S_{new}=23, which continued till S_{new}=24, where the change was the maximum (Eq. 3.4).

During the next transition phase, when the sample started to turn pink at S_{new}=51, similar to the previous transition phase, the change continued till the next instance. In

order to reflect the true colour transition, not just any abrupt colour change, the prominence of peaks was determined as shown in Fig. 5. 4 (b). It is a relative peak based on the intrinsic height and location as compared to other peaks. Hence, spike at 23rd and 24th instances were detected as one peak in Fig. 5. 4 (b), where the events of colour change were clustered (gathered) together. This is the peak when the sample turned into blue due to the presence of the bacteria. Although, the amount of change of colour at S_{new}=24 was higher than S_{new}=52, it was not considered as a separate transition phase; because the spike at S_{new}=24 is the continuation of the change in colour started at S_{new}=23. Similarly, when the sample was turning into pink, the clustered peaks were considered as the second transition phase, confirming the sample to be TB positive.
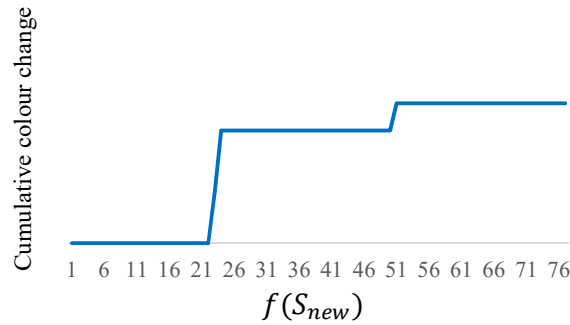
Fig. 5. 5 Detection of colour transition in Pl_1

After tracing the prominence (Fig. 5. 4), the transition was tracked in Fig. 5. 5, which followed the trajectory of the class of TB positive as theorised in Fig. 5. 3. The trajectory was consistent even if the assay plate contained multiple samples and required image-segmentation to calculate the number of transition (Fig. 5. 6).
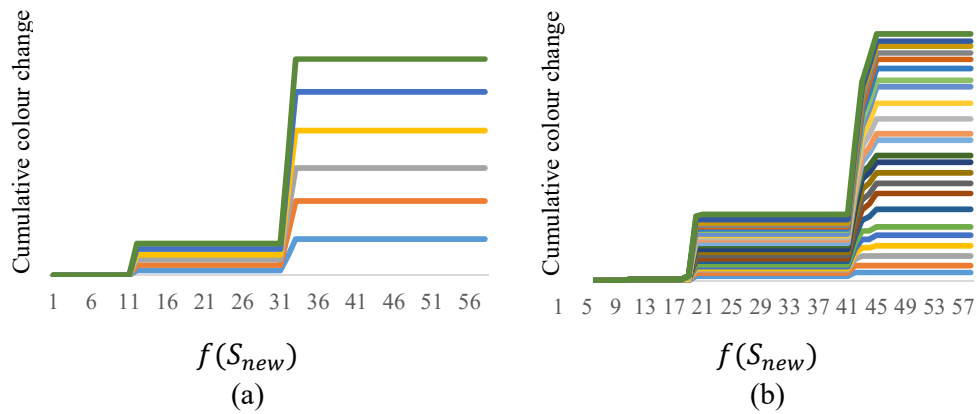
Fig. 5. 6 Detection of colour transition in Pl_2 and Pl_3. (a) Pl_2 containing 6 samples and (b) Pl_3 containing 24 samples

114

Varying t_i, it was observed that the transition is smoother for lower t_i. However, a consideration of $t_i < 1$ second is irrational because then the TB-test would become more time consuming without providing any useful new information.

Each column of Pl_3 varied in concentration, which can be observed from Fig. 5. 2 (c). From the experiment conducted for the reaction phase and time-dependant approach, it was also observed that a higher concentration provides a faster result.

Tracking the dynamic changes of the overall reaction, this approach aided us to understand the data from the eyes of the machine. One of the key findings of the approach was to confirm the need for a reduced reaction time. In other words, the machines can perceive a better meaning of colour at a $f(S_{new})$ lower than the human vision.

In the literature, the ΔE calculation found its usage for various purpose (Table 2. 6 in Chapter 2). However, this is the first attempt to provide such a colourimetric classification using ΔE calculation. The method presented has the advantage of using a lightweight algorithm. The computation can be completed within ~34 seconds. As the system works based on the difference between the colours in LAB space with time variation, the dependency on camera quality was reduced, which can be seen in Fig. 5. 2. Thus, the dependency on high configuration camera can be avoided. The work being independent of lighting environment is another paramount advantage. This method is applicable to a wide range of colourimetric examinations such as on-site cocaine detection (Smith et al., 2014). On the other hand, the outcome of the lightweight algorithm to process multiple video frames led towards the issue with memory occupancy and heavy processing for the mobile devices, contradicting the ASSURED criteria. One possible alternative could be more rigid sampling rate. The sampling rate was set by an empirical study. Therefore, excessive reduction of samples may impose unreliability of the approach. Another possibility is to fix the sampling rate from the anticipated time of the change of the phase. In a control environment at the same geo-location, it may be possible. For such a case, one would require to explore ϕ and $f(S_{new})$ for that particular test before setting a rigid sampling rate.

5.4 Reaction Phase and Time-independent Approach

A detail description of the biochemical protocol is provided in Bakhori et al. (2018), inspired by de la Rica and Stevens (2012). As described earlier, in a conventional sense, the method is supposed to be treated as an end point assay. From this point forward, the chapter deals with the data as an end point assay. In this way, only the static image of the final stage of the reaction is taken into consideration to investigate the second research question of this thesis.

5.4.1 Materials Preparation

A brief description of the protocol is described in Appendix D for the readers to perceive the overall process. One major difference with the biochemical process described in Section 5.3.1 is the use of the stoping agent. However, as mentioned earlier (Section 5.3.1), the biochemical assay preparation is beyond the scope of this thesis.

In this work, the presence of TB-specific antibodies can be confirmed if the sample turns into blue in the ELISA plate. In Fig. 5. 7(a), gold ions were reduced when H_2O_2 was present. The bottom 3 samples in Fig. 5. 7(a) were free from TB-specific antibodies. In the presence of H_2O_2, non-aggregated nanoparticles are formed, turning the solution pink. In the top three samples in Fig. 5. 7(a), the concentration of H_2O_2 was decreased, turning the samples into blue, confirming the presence of TB-specific antibodies.

Positive sample (Blue)

Negative sample (Pink)

Positive sample (Blue)

Negative sample (Pink)

(a)   (b)
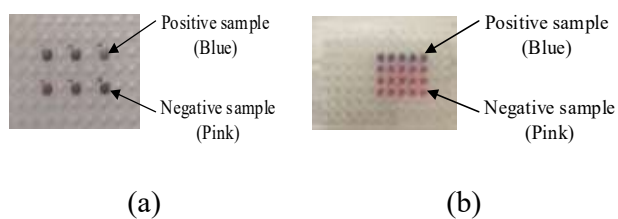
Fig. 5. 7: Samples in a plasmonic ELISA plate. (a) Samples are hard to visually distinguish, (b) Samples are visually distinguishable

5.4.2 Dataset of TB-test

The dataset of this case study contains images of 96-wells, which were partially filled, which means the plates contained both empty wells in addition to wells filled with the

sample. The final selection of 27 images[16] from 22 independent observations was conducted in the laboratory lighting environment. These images contained 266 samples - 81 of them were positive for TB-specific antibody, 181 were negative, and three of the samples failed to produce any indicative result. Thus 263 samples were finally selected. These samples from the 27 images are referred to in this thesis as the UPM-selected samples.

The acquired images varied regarding well size, camera to ELISA plate position, light exposure and mobile phone. Considering a robust application, this variation is expected in real life incoming images.

Let us assume, the assay plate, $A = f(X_x, Y_y, Z_z)$, where $\{X, Y\} \in \mathbf{Z}^+$ and $\mathbf{Z} \in \mathbb{R}$ and, $\mathbf{Z} > 0$. In this work, $X = \{1, 2, \dots, 12\}$ and $Y = \{A, B, \dots, H\}$. In order to explain the sample and assay position, an illustration is provided in Fig. 5. 8.
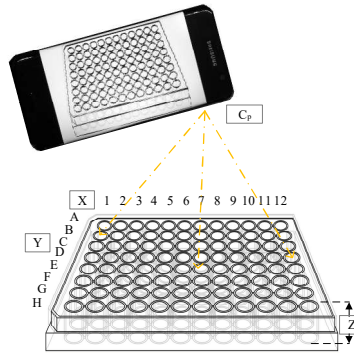
Fig. 5. 8: Impact of sample and camera position with respect to ELISA plate. X and Y are the length and width of the ELISA plate respectively. Z= volume of sample in the well and C$_P$= camera position.

For the commercially available 96-well plates, X and Y will maintain such positions in rows and columns. The space between these wells can vary from plate to plate. Thus, the wells are signified in (x, y, z) coordinates. Each well denoted by $w_{X,Y} \in w_{x,y}$ in the plate and $s_{X,Y} \in w_{X,Y} =$ sample, i.e. the well is filled with the sample. Both the shape and depth of the well can vary, depending on the specification of the assay plate. Due to the dimension of the well itself, the distance between these wells can differ from

[16] A description of the data sorting method is provided in Appendix D.

plate to plate. Depending on the biochemical protocol, the amount of sample to fill these wells can vary as well. All this information has a direct impact on the imaging. However, the colour of each sample, $s_c(r, g, b) \neq f(x, y, z)$.

As opposed to Karlsen (2018), we have maintained the camera position (C$_p$) parallel to the A, giving the wells a uniform exposure to the camera (Fig. 5. 8). For a static C$_p$, the distance between C$_p$ to each $w_{X,Y}$ was not equal. Thus, the sample to camera exposure was not equal. In theory, it would make $s_c(r, g, b)$ appear as $s_{cn}(r, g, b)$. The best exposure would be attained by the median $w_{X,Y}$.

The $s_c(r, g, b)$ can potentially differ due to the ambient conditions such as temperature, weather and geo-location, and certainly for the sample itself. However, this work was conducted in the laboratory environment.

After acquiring the data, it is important to understand the underlying structure of the data and maximise the insight regarding the presented problem to develop an intelligent colourimetric scheme. Exploring the dataset visually, there were some initial assumptions regarding the background and foreground colours as illustrated in Fig. 5. 9. The speculations are closely associated with the issue of signal vs. noise during image processing and classification stages.
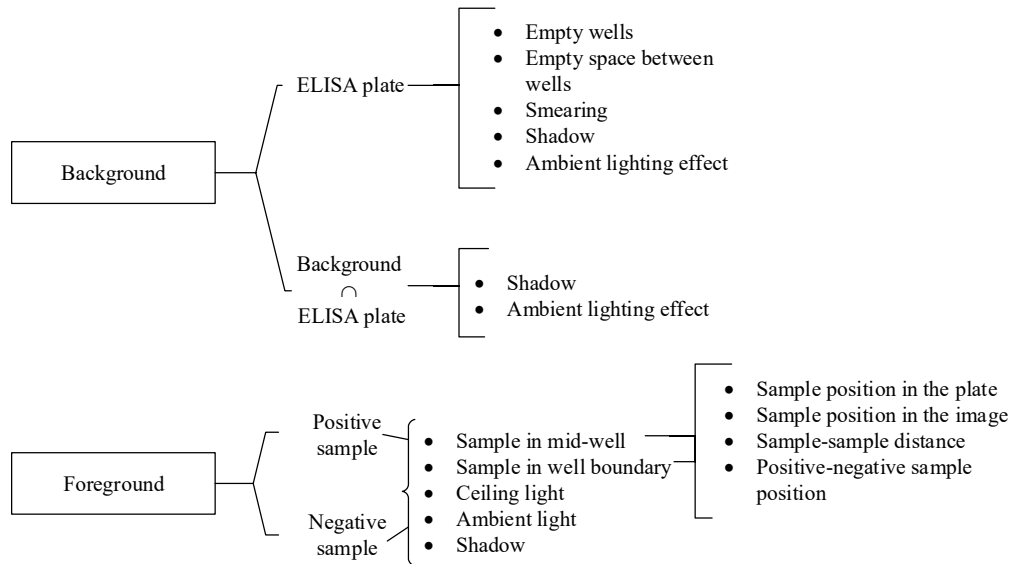
Fig. 5. 9: Observation of the associated colours and key variables in the image

The critical observations from the detailed inspection of the dataset are listed below.

Obs. 1: In the presented dataset, the sample-to-sample distance was not constant (Fig. 5. 8). If the wells are filled within a close neighbourhood, there is an unavoidable smearing effect. Thus, the background cluster holds many pixels which are close to the foreground pixels. With varying position (x, y), depending on the class of $s_{X,Y}$, the background cluster would be difficult to separate from the foreground clusters.

Obs. 2: In some cases, the positive and negative samples are hardly visually distinguishable. For example, in Fig. 5. 7(b), the samples are adequate for naked-eye measurements. For sample image, e.g. Fig. 5. 7(a), the indicated sample pair are hard to differentiate. This issue can worsen if the plate contains only one sample and the colour is as ambiguous as in Fig. 5. 7(a), which can lead to subjective interpretation. Moreover, there is a conscious variation in the sample colour, $s_c(r, g, b)$ on independent A.

Obs. 3: In the dataset, the value of Z (the volume of sample in a well) had an impact on the size of the sample ($s_{x,y}$). It implies that the 1st order colour moment can vary based on how the wells are filled. $s_{cn}(r, g, b) = f(Z)$. A well filled up to the surface would have a better exposure even if they are positioned at the far edge of the plate.

Obs. 4: This case study is comprised of the wet sample, which is not immune to light reflection from its surroundings (Fig. 5. 9).

Initially, our hypothesis was: the $w_{X,Y}$ with median (x, y) would be the ideal position for the samples, which might be immune to the ceiling light. However, the critical observations suggested that even a well filled up to the surface (Obs. 3) in the median position can suffer from the ceiling light reflection.

Obs. 5: The impact of 'camera to well position' (Fig. 5. 8) is aligned with our prediction in this section (Section 5.4.2). The SKEW can analyse such influence.

The observations Obs. 1, Obs. 3 and Obs. 4 have a definite impact on the image processing measures. The Obs. 2 works in our favour. The qualitative colourimetric tests are usually suitable for naked-eye detection, which necessitates (i) adequate biosensors to produce visually distinguishable colours and (ii) a user who has appropriate colour vision. Firstly, the use of intelligent systems can reduce the biochemical complexity without compromising the accuracy, specificity, sensitivity

and reliability. For instance, the positive and negative samples do not require to be visually distinguishable. Secondly, an intelligent system such as the system we presented in this work can eliminate the subjectivity of interpretation. A robust system should be able to handle the variation of sample colour, as mentioned in Obs. 2.

5.4.3 Data Resampling Technique

From Section 5.4.2, it can be observed that the dataset is imbalanced, where the number of negative samples is double that of the number of positive samples. The imbalanced dataset is a particular concern for the machine learning based classification algorithms. This imbalance is neither uncommon nor unexpected (He and Garcia, 2009). In this chapter, the data balancing was conducted by under-sampling, over-sampling as well as resampling, particularly for classification (Table 5. 1).

Table 5. 1: ELISA Dataset for TB-test

Sl.	Name	Number of samples	Balanced	Sampling
1	UPM-selected	266	x	x
2	D-E.1	186	√	Under-sampling
3	D-E.2	254	x	x
4	D-E.3	348	√	Modified over-sampling

Before beginning with any balancing operation, three samples, which produced to fail colour, were discarded from the classification-dataset. As mentioned earlier in Section 5.4.2, the assay plate contained multiple samples with an uneven distribution. Therefore, it is not possible to discard these samples during the image processing stage. Thus, there were 263 images in the UPM-selected dataset. Then, nine samples were taken out of the UPM-selected dataset for final testing and validation as shown in Fig. 3. 12 (Chapter 3). This dataset is denoted as D-E. 2 in Table 5. 1, which also does not contain a balanced dataset.

Based on the UPM-selected dataset, we introduced a partial selective resampling technique in Shabut et al. (2018). The resampling technique (Shabut et al., 2018) has

been excluded from this thesis to maintain consistency in the research methodology of this thesis. With an aim to create a balanced dataset in this chapter, at first, the UPM-selected samples were down-sampled and analysed (D-E.1). Then, we have balanced the minority class of the D-E. 2 by duplicating the samples, varying the density of data points (pixels); the dataset is denoted as D-E.3 in Table 5. 1.

5.5 Image Processing of the End point Assay

The proposed image processing framework is illustrated in Fig. 5. 10. This section sequentially discusses each component of the framework. If the output of the framework (Fig. 5. 10) is erroneous (Table 3.1), then the final colourimetric classification will also contain error.
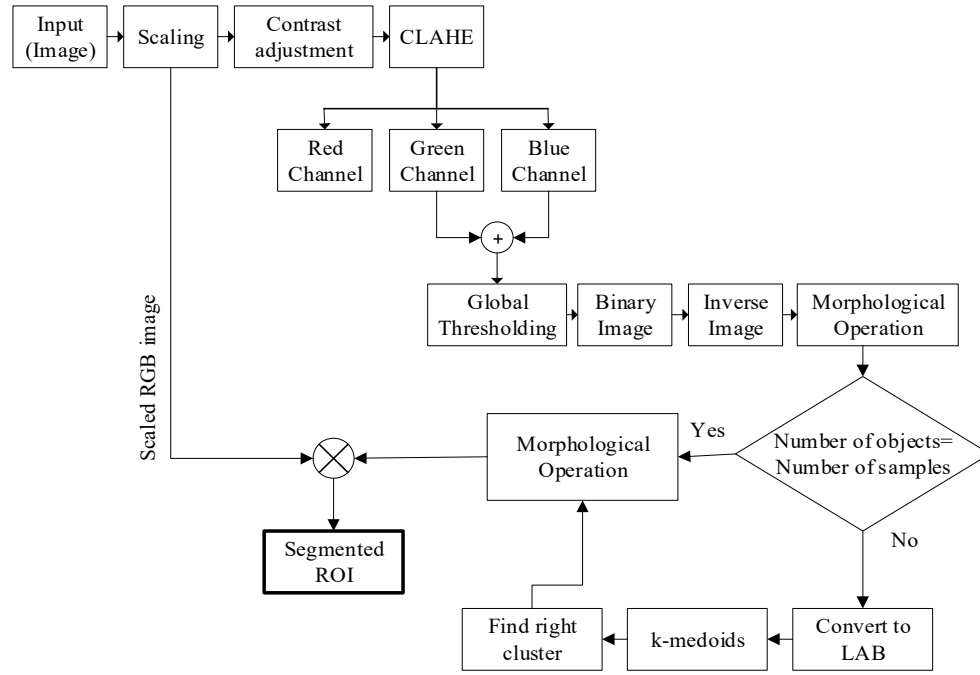
Fig. 5. 10: Image processing framework for TB-test

5.5.1 Image Pre-processing

At first, the acquired images were scaled and quantised to reduce the size of the image. For a simple method such as Otsu, the impact of scaling on processing time is negligible (Fig. 5. 11). On the other hand, to perform numerous iterations, the aid of scaling is obligatory for a heavy segmentation technique. The execution of algorithms such as superpixel, watershed and k-means can be considerably faster if the full-sized image (100%) is resized to a smaller dimension (Fig. 5. 11).

	100%	75%	50%	25%	Dynamic Scaling
----◆---- Adaptive Thresholding	0.420872	0.299939	0.26956	0.246409	0.224484
—■— Watershed	5.926099	3.346057	1.751006	0.822585	0.621361
—▲— Otsu Multilevel thresholding	0.199969	0.266984	0.22578	0.21463	0.222837
—✕— Global histogram thresholding by Otsu's method	0.184661	0.229378	0.221198	0.216305	0.214806
—✳— k-means	7.31429	4.354801	2.270487	0.893477	0.525735
—●— Superpixel	4.730296	1.947345	1.104111	0.59062	0.438966

Fig. 5. 11: Execution time for different segmentation method using various size reduction

Bourouis et al. (2014) utilised 32x32 pixels retinal images, which is not substantial to analyse the colour features of the presented dataset. Moreover, the resizing in Bourouis et al. (2014) was not dynamic. For a known condition, the height and the width of the image will not vary to a great extent. However, it may vary due to factors such as the position of the camera, size of the plate, and camera configuration. Thus, the size reduction in this work was performed dynamically and proportionally so that the geometry of the ROI was not deformed (Appendix E).

Initially, a wide range of pre-processing techniques was explored to aid the image segmentation process, as illustrated in Fig. 5. 12.
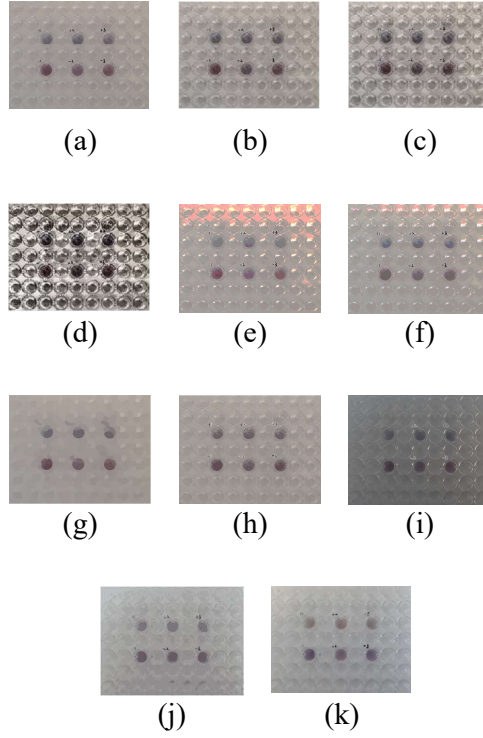
Fig. 5. 12: Initially explored image pre-processing techniques. (a) Samples in a plasmonic ELISA plate. Gradual enhancement of the image: (b) sharpened, (c) smoothened, and (d) final enhancement before colour space transformation. Quantisation input: (e) full size quantisation, (f) plane-by-plane quantisation, (g) Superpixel, (h) JSEG (i) Gabor filtering, (j) k-means[17], and (k) Gaussian filtering[17]

After exploring the effect of the pre-processing methods (Fig. 5. 12), the proposed pre-processing technique is comprised of scaling and sequential image enhancements (Fig. 5. 13). In Fig. 5. 13(b), the resized image was enhanced by the edge-aware local contrast operation. The edge threshold was carefully varied from 0.1 to 1.0 before setting the threshold value to uphold the strongest sharp edges. The higher value would sharpen the edge of the empty wells as well. Then the contrast was increased by 50%. This step improved the image quality to identify the filled wells and considerably enhanced the colours of the filled wells for both positive and negative samples.

---

[17] The displayed image was produced in OpenCV

(a)	(b)	(c)	(d)
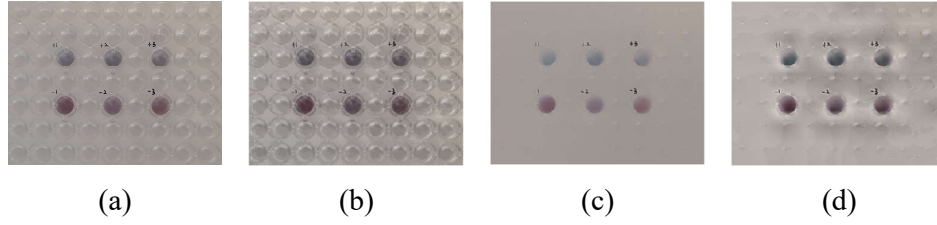
Fig. 5. 13: Image pre-processing of TB-test. (a) Input image, (b) Enhancement of the scaled image, (c) Smoothened image, (d) Enhanced contrast of the smoothened image

Exploring several smoothing techniques, e.g. Wiener Filter, Gaussian Filter, in Fig. 5. 13(c), the image was smoothened by further manipulating the edge-aware local contrast to reduce the detail of the image with an aim to diminish the edges of empty wells. This process weakened the contrast of the image, which required to be improved again. Therefore, finally, CLAHE (Zuiderveld, 1994) was utilised in Fig. 5. 13(d).

5.5.2 Image Segmentation

Regarding segmentation techniques, a number of processes (Wang et al., 2016a) were explored. This work deals with colour; therefore at first, colour thresholding was explored as a segmentation technique. Analysing the histogram, it was revealed that thresholding the colours would not provide an adequate segmentation, however, separating the colour channels can assist in the overall segmentation process.

In the field of image processing, Otsu (Otsu, 1979) is one of the most popular methods for image segmentation due to its simplicity and effectiveness. The method performed well for only a certain number of images (Fig. 5. 14). Therefore, more advanced clustering techniques were explored using unsupervised machine learning.

In our preliminary study, we performed the k-means with k=6 without any pre-processing and with minimum resizing (Abuhassan et al., 2017). Later we downsized it to k=4 (Shabut et al. 2018). In this chapter, based on the silhouette (Rousseeuw, 1987) suggested optimum number of clusters[18], the hybrid algorithm[19] (Appendix E) utilising

---

[18] An example is shown in Fig. 5. 14 (b).

[19] A proof-of-concept is already provided in Shabut et al. (2018) that heavy clustering algorithm such as k-means utilising four clusters can be handled in the mobile devices in real time. The computational complexity of k-medoids is very close to k-means. Therefore, conceptually, k-medoids using only two clusters should be deployable on mobile devices in real time as well.

k-medoids aimed at minimising the required number of the cluster to k=2. However, the random selection of cluster centroid position at initial stage compelled a requirement for a rule-based best cluster persuasion (Appendix E).

(a)

(b)

(c)   (d)
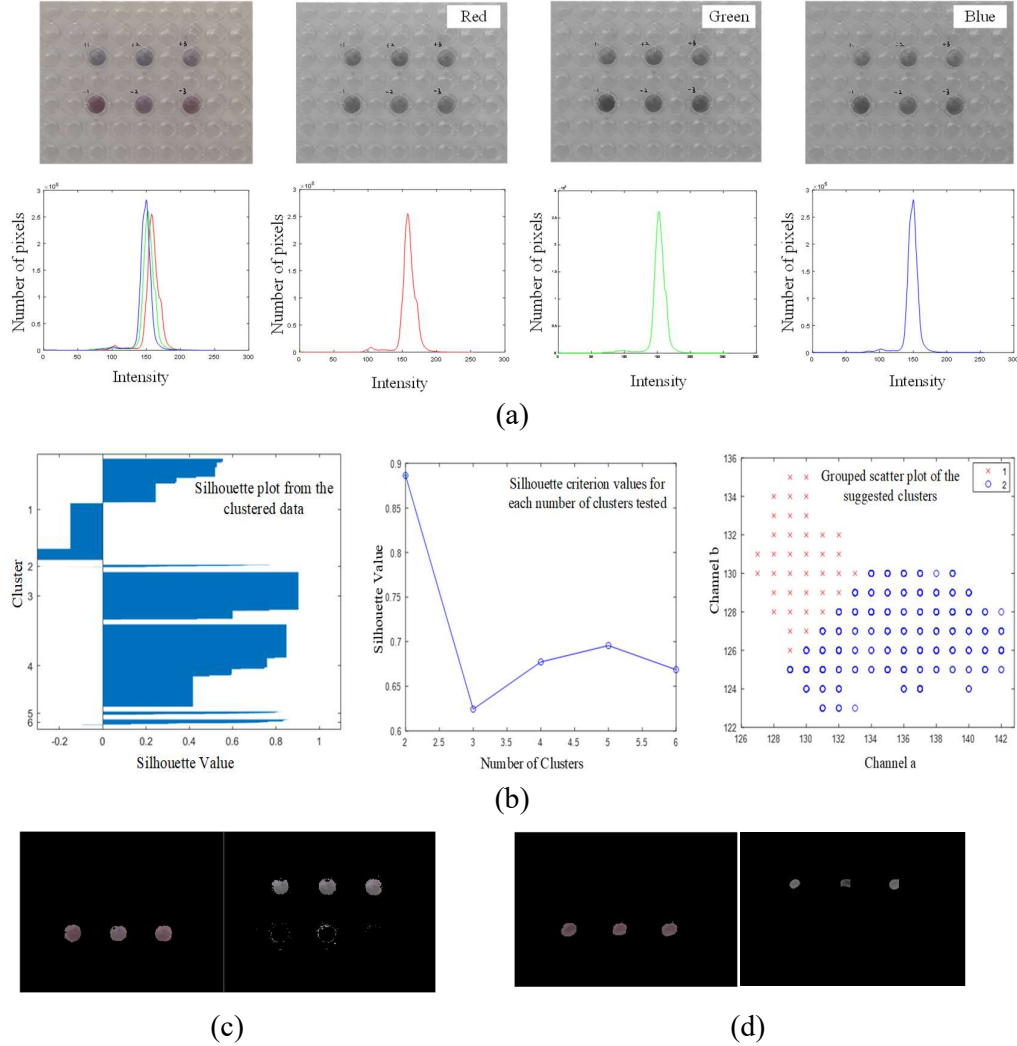
Fig. 5. 14: Image segmentation and post-processing of TB-test. (a) Colour histogram, (b) Investigation into the required number of clusters, (c) k-medoids on the pre-processed image, and (d) post-processed image.

5.5.3 Image Post-processing

After segmenting the image, irrespective of the technique, the image requires cleaning to ensure the minimal noise level at feature extraction stage. The overall process can be summarised as the hybrid algorithm (Appendix E).

The image processing framework (Fig. 5. 10) provided 99.62% accuracy to segment UPM-selected 266 samples (Table 5. 2). The accuracy was obtained from the number of correctly identified samples (Table 3. 1, in Chapter 3), rather than using the area of ROI.

Table 5. 2: Performance of image processing framework for the TB-test

Metric (%)	Value	95% CI
Accuracy	99.62	98.65- 99.95
Sensitivity	99.62	97.92- 99.99
Specificity	99.62	97.92- 99.99

Fig. 5. 15 represents segmented wells concerning the number of filled wells in the input image. In other words, Fig. 5. 15 shows how many samples were in the original input image and how many samples were segmented using the image. For example, the first image contained only one samples, whereas the 27th image contained 24 samples (Fig. 5. 15). After processing the images, there was only one segmented well for the first image, which is correct. On the other hand, there were 25 segmented wells for the 27th image. Therefore, the false positive and false negatives can also be tracked from the chart (Fig. 5. 15).
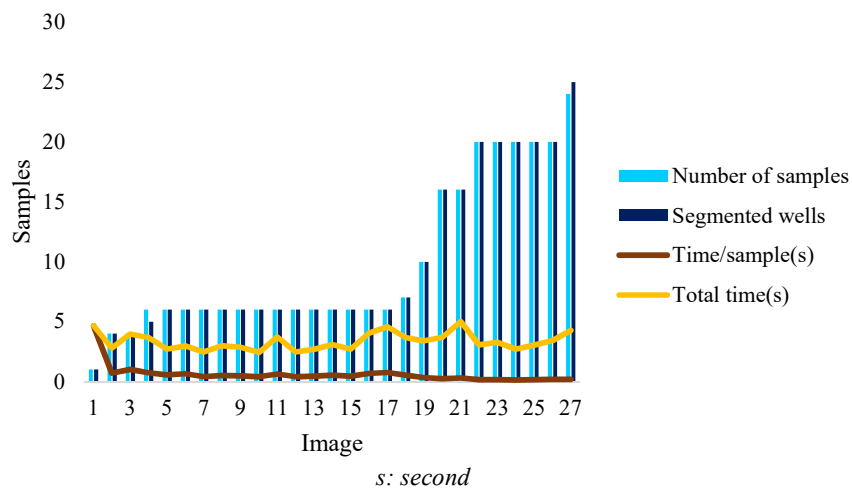
s: second

Fig. 5. 15: Performance of image processing framework with respect to computation time per sample

126

Fig. 5. 15 also shows the time taken to process the whole image, which is represented as 'total time' in seconds. Time taken to provide the segmented ROI as the output is also presented in the figure, as time/sample. The average time for segmenting a single well was found to be 0.6 seconds. With three times of repetition, the computation time was divided by the number of segmented samples per image, which provided the average segmentation time.

Analysis Fig. 5. 15, it can be stated that the segmentation time was not entirely dependent on the number of samples per image, instead was more inclined towards the observations made in Section 5.4.2. Based on the image quality, well to well difference, the impact of different colours as shown in Fig. 5. 9, the image processing algorithm was often forced to use more than one clustering techniques, which elongated the image segmentation process. Among 266 samples, 18.05% of the samples were processed using both Otsu and k-medoids clustering.

5.6 Classification and Feature Analysis

5.6.1 Feature Selection

This chapter utilises lower order colour moments from the extracted pixels of the segmented ROIs to be analysed for the classification purpose (Appendix E). Our preliminary study (Abuhassan et al., 2017) provided a good result using the RF classifier. Therefore, the colour moments and colour spaces were evaluated (as proposed in Fig. 3. 11) using RF. Considering only the mean colours of extracted ROIs in three colour channels of six colour spaces, tested on D-E. 1, YCrCb, LAB and HSV showed good performance (Fig. 5. 16).
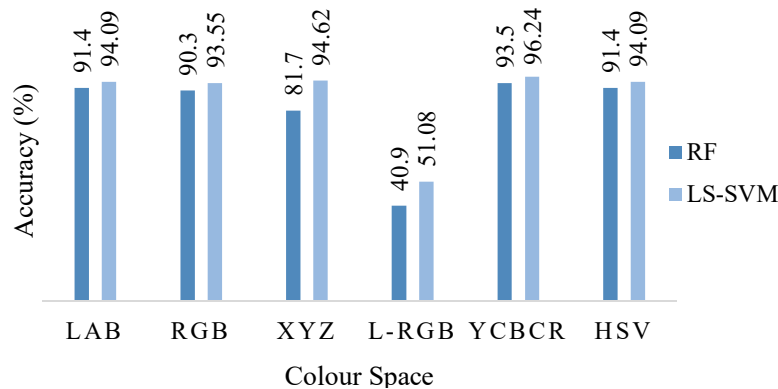
Fig. 5. 16: Performance of colour spaces for the TB-test

It is beneficial to use HSV colour space in time for colour correction, colour picking and artistic colour application, neither of which is necessary for the case study of this chapter. Originally, HSV was created to compensate for the issue with low CPU overhead, which was eventually overcome with the advancement of technology and LAB colour space (Margulis and Dan, 2005; Smith, 1978). Moreover, between LAB and HSV, LAB is preferable due to its ability to perceive relative perceptual differences between colours. The use of the cylindrical model in HSV is another downside.

Table 5. 3: Performance of features in LAB colour space for the TB-test (% accuracy)

Feature	Algorithm	D-E. 1	D-E. 2	D-E. 3
Mean	RF	91.4	94.9	96.3
	LS-SVM	94.09	96.46	98.28
Standard Deviation	RF	89.8	89.4	92
	LS-SVM	94.62	96.06	97.41
Mode	RF	88.2	70.9	78.2
	LS-SVM	94.09	86.61	80.7
Skew	RF	88.7	83.5	80.7
	LS-SVM	91.4	95.67	96.26
Energy	RF	80.1	94.1	96
	LS-SVM	86.56	96.85	97.13
Entropy	RF	78.5	85	89.7
	LS-SVM	73.66	87.01	93.68
Colour Moments	RF	94.6	95.7	97.4
	LS-SVM	98.92	100	100

Y, in YCrCb, holds the luma or luminance, can be directly computed from nonlinear RGB and Cr and Cb are the chrominance components. The Cb is the difference between blue and components of luminance, and Cr is the difference between red and luminance component. Due to the advantage of chroma subsampling, this colour space finds its particular usefulness in the computer gaming industry and video streaming. In MATLAB, the colour space is a better fit, if the computation is conducted in uint8 (8-bit unsigned integer arrays) than a double precision value. Moreover, YCrCb is a device dependent colour space. The case study utilised in this chapter has a comparatively smaller dataset. Although the images were captured using various mobile phone cameras, the proportion of these devices was not uniform. The images captured by different devices were not of the same assay plate either. Therefore, the LAB colour space was chosen over YCrCb. The performance of colour moments in LAB colour space considering all three colour channels is outlined in Table 5. 3.

The RF exhibited better performance when all colour moments in LAB colour space were considered (Table 5. 3). From the univariate analysis, mean, energy and standard deviation were found to be the more valuable features, whereas entropy was identified as a less important feature. This phenomenon was noted, however entropy as a feature was not fully discarded.

Table 5. 4: Levene's test F for the feature selection

Number of features	Mean	Standard Deviation	Standard Error Mean		Degree of Freedom	58
					Equal variances assumed	
					Levene's test F	107.46
15	99.72	0.254	0.046		Mean Difference	-0.28
					T	-5.794
18	100	0.000	0.000		p	0.000

The symbols carry usual meaning.

An inferential statistical test was conducted to explore the impact of inclusion and exclusion of entropy (Table 5. 4). The statistical test utilising the Student's t-test involved two groups, one containing the accuracies obtained using fifteen features,

excluding entropy and the other considered all eighteen features including the entropy. Levene's test in Table 5. 4 suggested that there is a statistically significant difference in the variances between these two groups. Therefore, the result of the intelligent colourimetric test for Case Study 1 would be more reproducible if all 18 features are considered (Table 5. 4). Due to this reason, entropy was not discarded from the feature-set.

5.6.2 Performance of Classification Method

Unsupervised Machine Learning

For a qualitative colourimetric test which can provide a naked-eye detection, the close points in data space are supposed to be in the same class, which implies, the positive and negative samples should be distinguishable without training the system with the corresponding class label. For such a case, the decision boundaries should lie in low-density regions. Thus, the subjectivity of colour interpretation for a binary classification can be easily removed by an unsupervised machine learning technique.

As the image segmentation techniques described in Section 5.5.2, involves clustering techniques, i.e. k-medoids clustering using colours in AB space, we used k-medoids and kmeans to determine the presence or absence of TB antigen-specific antibodies. The two-phase iterative algorithms were used to minimise the sum of point-to-centroid distances for 18 features, summed over 2 clusters present in here.

(a)

(b)
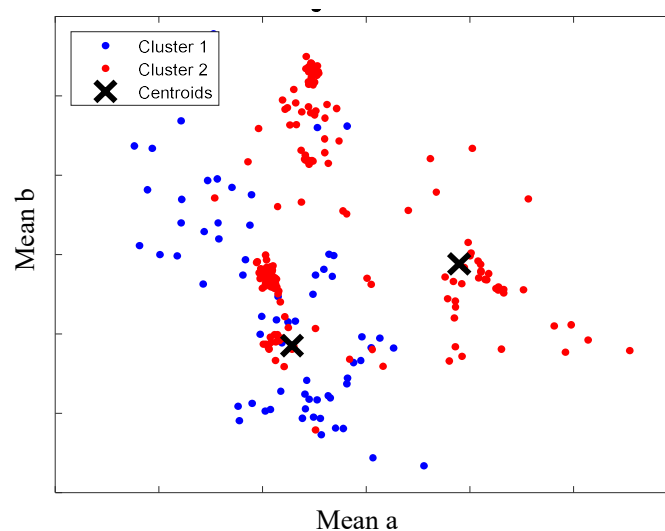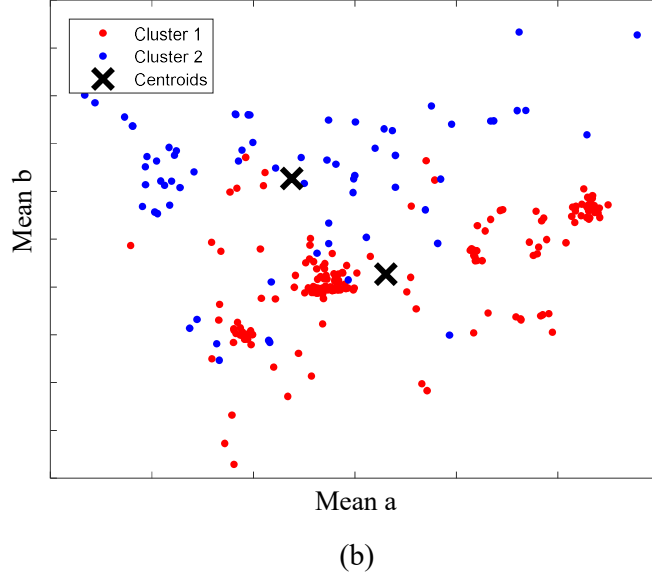
Fig. 5. 17: Cluster assignments and centroids using mean colour values in 'a' and 'b' channel (a) k-medoids and (b) kmeans

It is difficult to produce 18-dimensional illustration on paper; thus, the distribution of only mean a and b colour value of the samples are shown in Fig. 5. 17. Based on the same features, k-means and k-medoids showed the same performance. Therefore, the result is explained using only one of the algorithms, i.e. kmeans. At first, the squared Euclidean distance was used, later replaced by other proximity measures such as correlation, cosine and Manhattan distance, which played no difference in the result.

A dataset of n x p matrix can be presented as $X_{ij} \in \mathbb{R}^{nxp}$, cluster centroid location $C_{ij} \in \mathbb{R}^{kxp}$, where, k=2 and row j implies the centroid of cluster j. Using k-means++ algorithm (Arthur and Vassilvitskii, 2007), ten different sets of seeds were selected. In this way, the clustering process was repeated ten times using new initial cluster centroid positions. In return, k-means provided the lowest within-cluster sums of point-to-centroid distances, which is a 2x1 vector, where element j is the sum of point-to-centroid distances within cluster j.

From the ratio between correctly classified samples and classified samples, the correct rate for k-means clustering was found to be 85.43%. The overall accuracy was below 50% while considering 15 features, excluding entropy in three colour channels. Considering the full feature-set, 18 features including entropy, the accuracy improved

to 68.1%, which is still low. The conceivable reason for such poor performance is the use of k-means using dependent variables and the translation invariance.

The colour channels l, a and b are independent variables. Therefore, the use of k-means (Abuhassan et al., 2017; Shabut et al., 2018) and k-medoids (Section 5.5.2) clustering for image segmentation is justified. The use of Euclidean distance in the image processing framework satisfies the underlying assumptions (including independence of the variables). In this chapter, the presented dataset does not provide easily distinguishable naked-eye detection of TB antigen-specific antibodies. Therefore, the decision boundary does not lie within the mean colour value of the low-density regions in l, a and b.

As mentioned earlier, the meaningful attributes were extracted from the pixel lists in LAB colour space. Thus, variables derived from the histogram in l, a and b are not independent. Moreover, some of the features have a more direct dependency on the other features as well. Neither k-means nor k-medoids is able to comprehend such dependencies.

Supervised Machine Learning

Using 18 features in LAB colour space, both parametric and non-parametric classifiers were trained, and 10-fold cross-validated on the datasets. The supervised machine learning techniques showed significantly better performance than the unsupervised techniques. The only disadvantage in the case of supervised learning is its ability to function only using labelled data.

Considering the original dataset for ELISA containing 254 images (D-E.2), the top performing classifiers were LS-SVM (Suykens and Vandewalle, 1999; Suykens, 2002), Bayesian Regularisation Artificial Neural Network or BRANN (Hagan and Menhaj, 1994) and fine K-Nearest Neighbours (KNN). The performance is shown in Fig. 5. 18.

The KNN showed 96.29% accuracy (Fig. 5. 18), without any hyperparameter tuning. The most common distance metric for KNN is the Euclidean distance. For the fine KNN, which uses medium prediction speed in MATLAB, i.e. ~1 second, the number of neighbours (k) was chosen to be 1. After standardising the data, such prevalent parameters provided the mentioned accuracy. The KNN tends to perform better with Euclidean distance when a higher number of neighbours are chosen. However, the

increment in k degraded the performance. The medium KNN with k=10 and the coarse KNN with k=100 provided 92.1% and 69.7% accuracy, respectively. Thus, other distance metrics were explored.
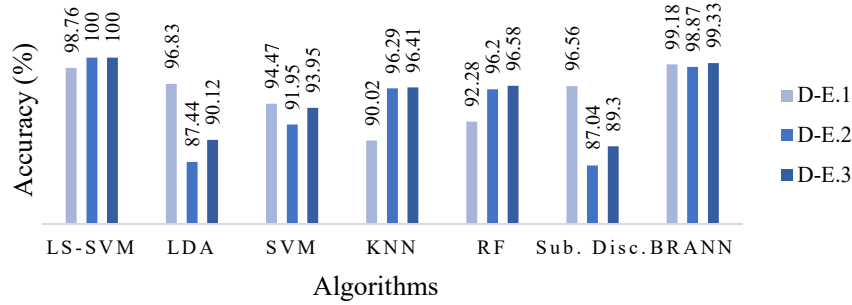
Fig. 5. 18: Performance of different classifiers for the TB-test

Among the distance metrics, in addition to Euclidean distance, some other Minkowski distances were explored in this work, e.g. Hamming and Manhattan distance. These two are equivalent when all features are in binary form, which is not the case in here.

In this work, both Hamming and Jaccard distance provided poor performance. In the case of Jaccard, the number of common attributes is divided by the number of attributes that exist in at least one of the two objects. The performance from Jaccard distance was predictable from the clustering performance.

The Chebyshev, correlation and cosine distance showed similar performance. The cosine (or angular) distance is proportional to the squared Euclidean distance, which dissatisfies the triangular inequality. The KNN cosine works well when the subset of original data is consistent with the original data. From this context, a large dataset, maintaining the same camera to sample distance and ambient condition may work well. The magnitude of the vector is another point of concern for the presented dataset, for which cosine KNN is not a good choice for our case.

The best performance from KNN was achieved using Manhattan (also known as the city block) distance. This is another case of Minkowski distance, where p=1. It has L^1 norm, whereas the Euclidean distance is of L^2-norm. The Manhattan distance usually performs better in case of high dimensional data. The 10-fold cross-validated accuracy was 97.6%, which is still lower than the close performing classifiers LS-SVM and BRANN.

In this work, we employed an MLP with a feedforward neural network (Rosenblatt, 1961). The hidden layers used hyperbolic tangent sigmoid transfer function (tanh(N), N= input), whereas the output layer was implemented by a linear transfer function. The feedforward network provided 94.9% accuracy.

Using error calculation from the training output and the target, the backpropagation can adjust the weights and biases of the input and hidden layers. To improve the performance furthermore, Bayesian regularisation was deployed. A combination of squared errors and weights were minimised to produce a corrected combination. Thus a better generalisation is produced. Using BRANN, the weights and bias values were updated by Levenberg-Marquardt optimisation (Hagan and Menhaj, 1994). The network architecture is shown in Fig. 5. 19.
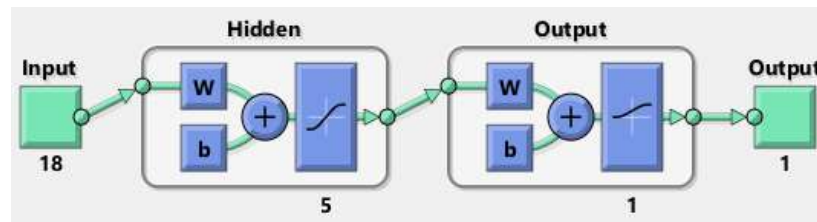
Fig. 5. 19: Network architecture of MLP with Bayesian regularisation backpropagation

The hidden layers were varied from 2 to 10. The best result was achieved when the number of hidden layers is either 5 or 6. For a more stable network, six hidden layers were chosen. No validation stopping was used, as regularisation was utilised. In this way, the training was carried out until an optimal combination of errors and weights was obtained. Regularisation aims to build a more generalised model. The performance of training and testing stage can be visualised from the mean squared normalised error with respect to epochs (Fig. 5. 20).

The best training performance was $3.3411e^{-09}$ at epoch 83. The error tends to reduce after more epochs of training. However, it affects the training time. In an ideal case, the testing graph in Fig. 5. 20 was supposed to follow the training one. Although the error was low for both cases, with more epochs, the minimisation of testing error stopped earlier, which may have occurred due to overfitting. This issue needs to be further investigated in future (Section 7.3).

(a)   (b)
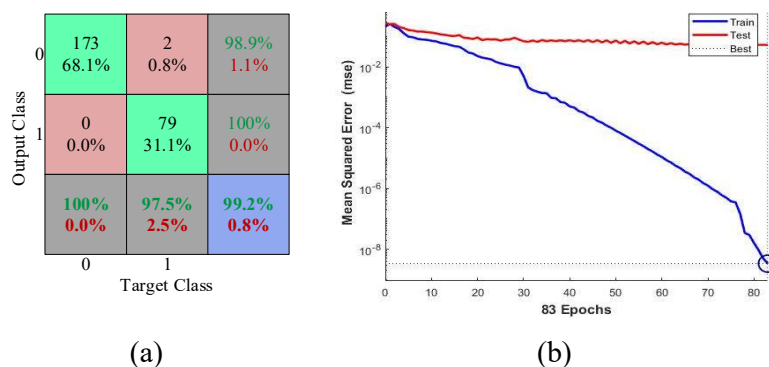
Fig. 5. 20: Result of MLP with Bayesian regularisation backpropagation. (a) Confusion matrix, (b) Training and testing performance

The parametric methods (e.g. LDA) are independent of the number of samples, easier to understand and interpret and fast to learn. On the original dataset, without under or oversampling, the non-parametric methods (e.g. LS-SVM, KNN) performed better than the parametric models (Fig. 5. 21). This chapter focuses on colourimetric tests of wet-chemicals, where the colour of samples can vary due to ambient condition, the position of the sample in the assay plate, bio-chemicals and shape of the well. Therefore, for robust operation, the system should be flexible while making assumptions regarding the form of the mapping function. More generalisation can be achieved for the unseen data by such algorithms. Hence, the better performance of non-parametric models is conceivable and more logical.

The parametric methods showed good performance in D-E.1 (Fig. 5. 21). The linear models such as linear SVM and LDA showed good performance on the downsized balance dataset (D-E.1) of 186 samples. However, the performance was not well maintained when exposed to rare data (Fig. 5. 21).

Among the discriminant analysis, the quadratic discriminant analysis provided a better result than ensemble subspace discriminant analysis. The performance of different tree models including RF was also good (>90-96.5%).
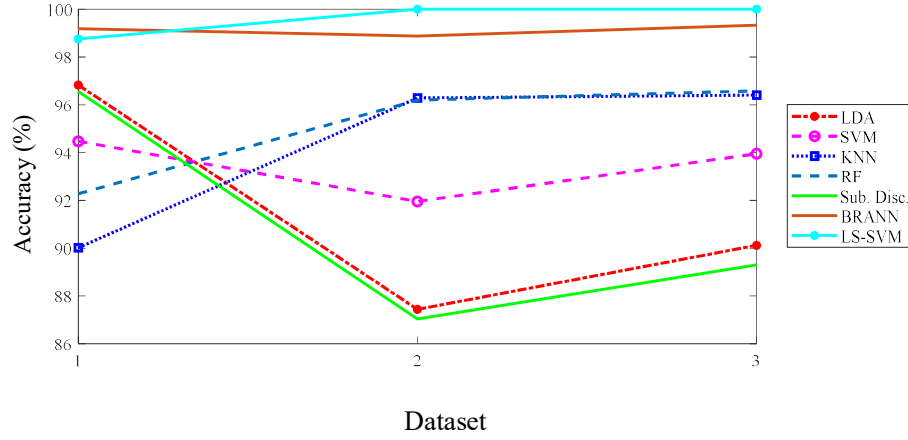
Fig. 5. 21: Performance of different classifiers with respect to the number of samples and data-balance

Based on the performance of all the classification models mentioned above on all three datasets, the LS-SVM was found to be the top performing classifier[20]. The closest performance was achieved by BRANN. Hence, further analysis is presented for LS-SVM only. These analyses include final optimisation of features using all the dataset (Table 5. 1) and statistical analysis.

Table 5. 5: Performance of LS-SVM on the TB-test datasets[21]

Dataset	Sensitivity (%)		Specificity (%)		Accuracy (%)	
	Value	95% CI	Value	95% CI	Value	95% CI
D-E.1	100	96.11-100	97.89	92.60-99.74	98.94	96.21-99.87
D-E.2	100	95.44-100	100	97.91-100	100	98.56-100
D-E.3	100	97.90-100	100	97.90-100	100	98.95-100

Using all 18 features in LAB colour space, LS-SVM showed 100% accuracy on the 10-fold cross-validated UPM-selected dataset, D-E.2. When the execution was repeated 30 times, the average accuracy was maintained at 100%. The overall dataset is small, and D-E. 2 is not balanced. Therefore, the classifiers required to be examined for the balance

---

[20] The LS-SVM was not considered in our preliminary study (Abuhassan et al., 2017).
[21] The performance was obtained in a random execution of the classifier.

dataset. The average accuracy of Fig. 5. 18 is within the 95% confidence interval shown in Table 5. 5.

The performance can also be visualised from the ROC curve in Fig. 5.22. The ROC for LS-SVM was found to be identical for D-E.2 and D-E.3, which can be seen in Table 5. 5 as well.

area: Area under the ROC curve; std: Standard deviation
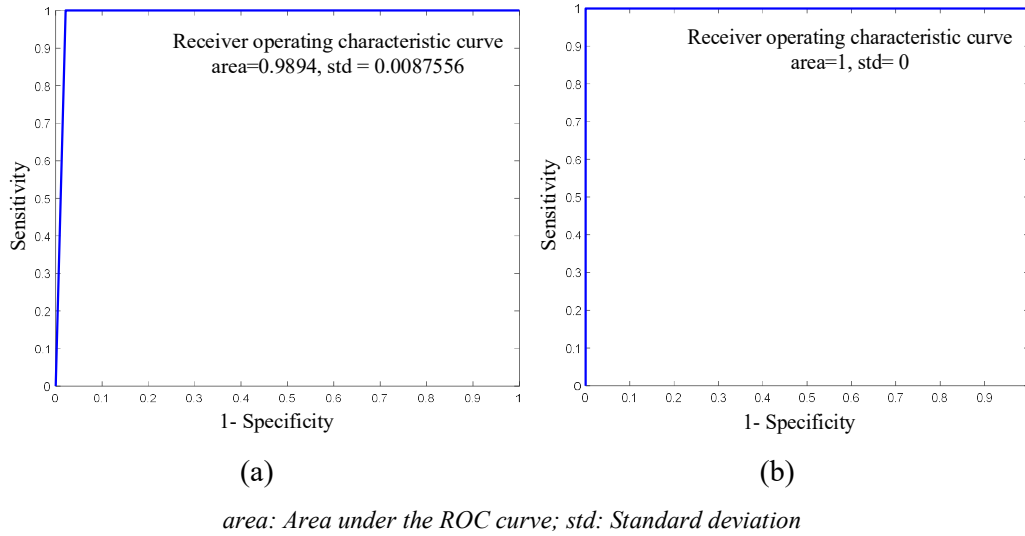
Fig. 5. 22: ROC of LS-SVM for the TB-test on (a) D-E.1, and (b) D-E.2 and D-E.3

LS-SVM showed the average accuracy of 98.76% on the balanced dataset of under-sampled D-E. 1 (Fig. 5. 21). The non-parametric LS-SVM was able to learn from the newly introduced 'rare' data; the accuracy increased to 100% when exposed to D-E. 2. This classifier showed consistent performance for the resampled dataset (D.E.-3) with altered data-points distribution. Finally, the system was tested and validated by using unseen samples; the system provided a correct classification for all of them.

As rapidness is one of the key criteria of this thesis, Fig. 5. 23 presents the average computational time of top performing classifiers for each of the datasets. For better visualisation, time in second is presented in logarithmic scale.

With a higher number of data, the classifiers were supposed to take a longer time to compute. The repeated execution (three times) matches of prediction as shown in Fig. 5. 23. For D-E.1, D-E.2 and D-E.3, the computation time by LS-SVM were 0.0219, 0.0265 and 0.0289 seconds, respectively.
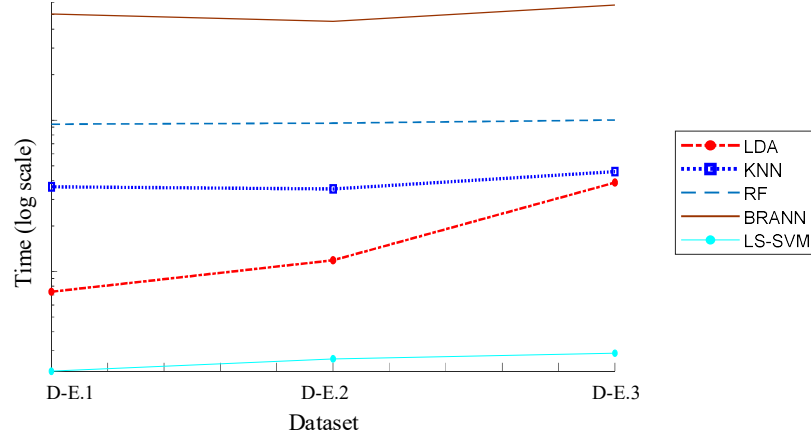
Fig. 5. 23: Computation time by different classifiers with respect to the number of samples and data-balance

For D.E-1, the close performing classifiers were LDA and LS-SVM (Fig. 5. 23). Due to the computation process, one can assume that LDA would be the fastest classifier among the five classifiers shown in Fig. 5. 23, and BRANN would take a longer time to compute. BRANN was found to be the slowest one, as predicted; however, for all the datasets, LS-SVM was found to be the fastest classifier.

Statistical Analysis

As mentioned earlier in this section (Section 5.6.2), the performance of the classifier varied based on the balanced dataset as well as the ability to learn from the bigger dataset. The statistical analysis was conducted to portray a clear picture of the overall performance.

Based on the results obtained by different classifiers including supervised and unsupervised, parametric and non-parametric, and neural network, Student's t-test was conducted in SPSS for the top performing classifiers (Lwin, 2015; Hatem Al-jamal, 2017).

The statistical results obtained by a two-tailed t-test at a 0.05 level of significance are given in Table 5. 6. The result of Algorithm 1 ↔ Algorithm 2 stated as "+", "−", or "~" can be read as- Algorithm 1 is significantly better than, worse than, or statistically equivalent to Algorithm-2, respectively.

The performance obtained by Student's t-test is consistent with our analysis described in Section 5.6.2. To summarise the overall performance achieved by different classifiers on balanced and unbalanced data, and the ability to learn from new data, the LS-SVM was found to be the best performing classifier for the presented problem. The performance of LS-SVM for qualitative classification of TB-test is supported by the performance of colourimetric classification in the literature as well (Section 2.5.3, in Chapter 2).

Table 5. 6: Student's t-test results of the top performing classifiers for the TB-test

Algorithm 1 ↔ Algorithm 2	D-E.1	D-E.2	D-E.3
LS-SVM ↔ LDA	+	+	+
LS-SVM ↔ KNN	+	+	+
LS-SVM ↔ RF	+	+	+
LS-SVM ↔ BRANN	-	+	+
LDA ↔ KNN	+	-	-
LDA ↔ RF	+	-	-
LDA ↔ BRANN	-	-	-
KNN ↔ RF	-	-	~
KNN ↔ BRANN	-	-	-
RF ↔ BRANN	-	-	-

5.7 Summary

This chapter inspected the case study of TB-test to deliver the specific objectives. Considering the ELISA reaction as a continuous process, a novel classification approach was presented using video-frames analysis. A detail discussion is provided for this method as reaction phase and time-dependent approach. In addition to providing classification, this Euclidean distance-based method, without using any complex machine learning technique, indicated that the computational system could recognise the colour at an earlier 'end point' than the conventional end point designated for naked-eye measurement. However, treating an end-point assay under time-dependent scheme resulted in handling multiple images, whereas consideration of only end-point would require only one image as an input of the system. Therefore, the reaction phase and

time-dependent approach will be a better fit for the kinetic assays (Khan and Garnier, 2013) where the chemical reaction can easily be tracked using the proposed method.

Considering the case study of ELISA as an end point assay, this chapter also demonstrated an intelligent image-based expert system to perform automatic detection of TB antigen-specific antibodies with the integration of machine learning techniques. Using a robust image processing technique comprised of histogram analysis including clustering, the system can detect samples (wells) without any guide or virtual plate. The decision components facilitated the selection of the right cluster among the clusters, detection of wells and transcending the samples from noise. Therefore, unlike the reported articles, the system does not require the user to provide seed points or perform cropping. Moreover, the system is capable of reading multiple samples and classifying them as positive or negative in real time.

This work incorporated supervised machine learning to free the TB test result from the colour perception of individuals and its subjectivity of interpretation. Utilising 18 histogram features, we achieved 100% accuracy with LS-SVM classifier. The performance of k-fold cross-validated LS-SVM was verified through statistical analysis as well as by entirely unseen samples to the classifier.

Although the performance of the system for TB-test was outstanding, images were captured as robustly as possible, plate to plate colour variation was also present, which is supposed to aid the classifier to learn the data as dynamically as possible and overall system was trained, tested and validated as critically as possible- the dataset was small. Additionally, plate to plate sample variation was immensely noticeable, because it is not a stable type of assay. The image capturing process was also not consistent, and variation in the image capturing process was not proportionally distributed. Considering more dynamic real-life application, the colours in the filled wells, i.e. sample colour can vary due to climate and geo-location. Based on the experimental analysis, it is difficult to prove how much variation was caused by the sample, due to the randomness, or lack of proficient use of the protocol of the biochemical assay; all of them are grounds to the rarity (Weiss, 2004). Therefore, in order to support our claim regarding the performance achieved in this chapter, another case study was considered (Chapter 6). We will further extend the scope to deal with multi-class classification and multi-objects per sample in the next chapter.

Chapter 6

Experiments and Results: Case Study 2

6.1 Introduction

This chapter utilises a case study of a lateral flow assay. The LFA is mainly popular for POC platforms since they are easy-to-use, fast and low-cost. However, they often suffer from limited specificity and sensitivity due to the limitation of materials including biochemical components (Koczula and Gallotta, 2016). As a POC system, integration of computational system to LFA can enhance the overall diagnosis experience such as the research conducted by Ozkan (2017). This chapter mainly utilises paper-based assays, using universal pH test strip as one of the test studies. In general, paper-based assays such as filter paper-based water-quality monitoring scheme by Sicard et al. (2015) are more affordable and suitable for less trained personnel. This chapter aims to explore a stable assay at diverse conditions, mechanism to perform multiclass colourimetric tests using histogram thresholding based image processing and machine learning algorithm without any user intervention, and propose an intelligent, high performing, robust and rapid method to perform colourimetric test for the assays similar to the case study.

6.2 Experimental Selection

6.2.1 Materials Preparation

As a paper-based LFA, this chapter mainly focuses on pH indicator universal test strips. The pH ranging from 3.0 to 10.0 was considered for this work. As the objective is to deliver a proof-of-concept, the whole range (0-14) was not considered. In this work, we used buffer solutions to ensure more stability and longevity of the solution, therefore increasing the reliability of the overall experiment.

For pH 3.0-5.0, the citrate buffer was prepared as x ml $A + y$ ml B diluted in deionised water, where $A=$ X citric acid ($C_6H_8O_7$) and $B=$ Y sodium citrate ($Na_3C_6H_5O_7$), where X and Y represents the concentration in molar or M (Gomori Buffers). Traditionally 0.1M is used. However, a higher concentration would result in longevity of the solution. Similarly, for pH 6.0-7.0, the phosphate buffer was prepared from NaH_2PO_4 and

Na$_2$HPO$_4$. Using NaHCO$_3$ as the weak base and Na$_2$CO$_3$ as the strong base, buffer solution for pH 9 was prepared. The pH 8 and 10 were tested by commercially available National Institute of Standards and Technology (NIST) traceable borate pH buffer solution (Fisher Scientific, UK). The pH level measurements were controlled with calibrated pH meter (HI 208, Hanna Instruments). The calibration was conducted with standard buffer of pH 4 and pH 7.

The Fisherbrand® pH test strips were immersed in the prepared pH solutions. The test strips instantly form the colour and change it quickly after drying. Therefore, images were captured rapidly. Some of the test strips were allowed to dry on tissue paper, while some of them were purposefully not allowed to dry the residue on tissue paper (Appendix D). The purpose of it was to make the dataset more diverse.

6.2.2 Experimental Setup

The ambient condition can be a vital influential factor for colour image processing using mobile devices. Therefore, the experimental condition is described in detail to deliver a reproducible system (Fig. 6. 1). All the experiments were conducted in a controlled environment (Biosafety level 2), where the room temperature was maintained at approximately 20±2˚ C.

1	Observer
2	Camera position
3	Sample plane
4	Lamp position
5	Natural light through window
6	Fluorescent ceiling light
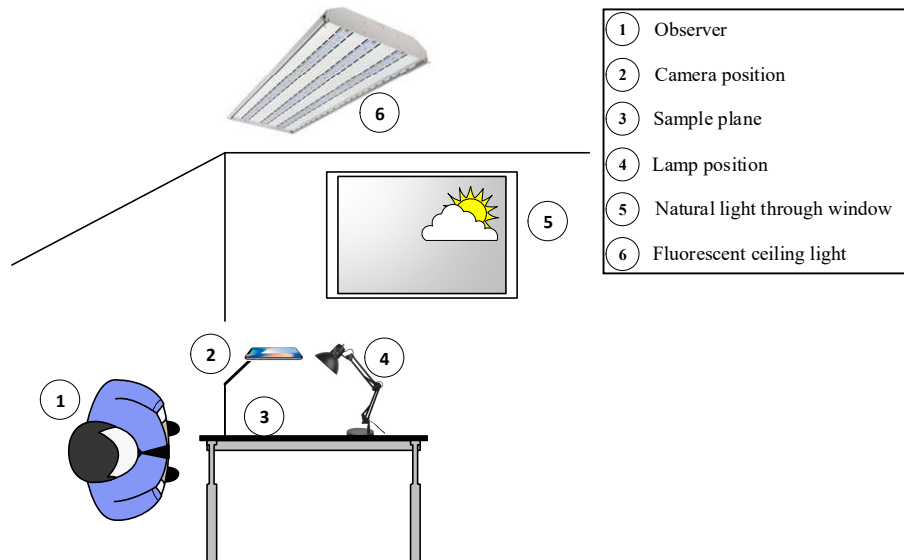
Fig. 6. 1: Experiment condition

Lighting condition, $\tilde{A} = A_i + A_c$ (6. 1)

A_i= Natural light coming through the window

$$A_i = f(E_{ext}) \tag{6.2}$$

$$E_{ext} = E_{sc} \cdot (1 + 0.033412 \cdot \cos(2\pi \tfrac{d_n-3}{365})) \qquad \text{(6.3), (Kandilli and Ulgen, 2008)}$$

E_{sc}= Solar Illuminance constant

d_n= Number of day in the year

The experiments were conducted during winter and spring in the UK. Therefore E_{ext} in Eq. 6.3 potentially varied from < 1 to 11.1×10^4 lx.

A_c= Controlled ambient light

In this work, three different controlled lights were added to A_i in Eq. 6.1.

A_c (l)= Office environment with fluorescent ceiling lights

A_c (dl)= Daylight lamp, 10 Watt, 6500 K

A_c (jl)= Warm white lamp, 5 Watt, 2700K

$\tilde{A}_e$ = Light exposure on sample plane

$$\tilde{A}_e(t) = \begin{cases} \sim 4200K, \ \tilde{A}_e = A_i \\ \sim 3700K, \ \tilde{A}_e = A_i + A_c\,(l) \\ \sim 3000K, \ \tilde{A}_e = A_i + A_c\,(jl) \\ \sim 5300K, \ \tilde{A}_e = A_i + A_c\,(dl) \end{cases}$$

-- Light exposure on sample plane ($\tilde{A}_e$)
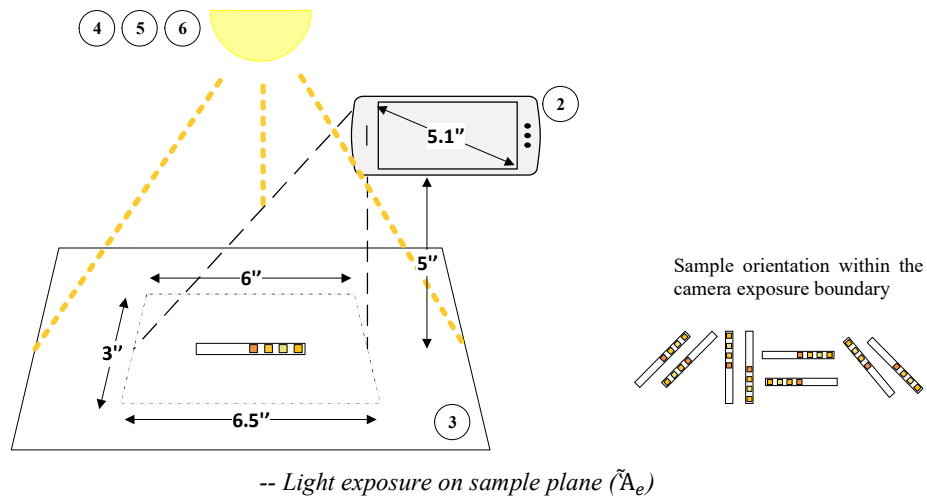
Fig. 6. 2: Camera and light exposure on sample plane. The numbers carry the same meaning as in Fig. 6. 1

The average colour temperature was measured on the sample plane with an Android application (QuArt Studio, 2018). An elaborate illustration of the exposure plane in Fig. 6. 1 is sketched in Fig. 6. 2.

The sensitivity rating of the camera is defined by the International Organization for Standardization (ISO). Agudo Acemel (2017) drew attention towards the impact of parameters such as pixel resolution, ISO number, and white calibration on colour quantification. In the reported articles, works have been done keeping these parameters constant (Mutlu et al., 2017) as well as variants (Solmaz et al., 2018).

In this chapter, the SM-G920F camera (f/1.9, exposure time 1/50 second, focal length 4mm, maximum aperture 1.85, 35mm focal length =28, normal exposure program, without flash) was set on its default mode. The idea is to utilise the strength of the available camera with minimum user interaction. The camera automatically adjusted the white balance, ISO speed, metering mode and set the brightness by the internal software of the smartphone (G920FXXU6ERC9).

The metering mode signifies the camera exposure method. The quantity of light/unit area to imaging sensor estimated by the shutter speed, aperture and scene luminance is expressed as the camera exposure. The dataset contains images where the camera measured 1–5% of the viewfinder area (spot metering) as well as the images where the meter concentrated ~60–80% of the sensitivity towards the central part of the viewfinder (centre-weighted average metering).

The native and extended ISO of the mentioned device can vary from 100-800. Lopez-Ruiz et al. (2014) conducted their experiments by setting the ISO at 200 using LED flash at dark environment. Their study found lower noise in the image at ISO 200 than higher ISO. This work observed that the camera selects lower ISO speed at the bright ambient condition. For example, ISO speed 64 in case of $A_i + A_c$ (dl) and 160 at $A_i + A_c$ (jl) and 200 at A_i (varied for image to image).

In addition to the SM-G920F camera, this work also utilised iPad Pro ((f/2.2, exposure time 1/50 second, focal length 4mm, without flash) and SM-J327T (f/1.9, exposure time 1/50 second, focal length 4mm, maximum aperture 1.85, without flash) to generate diversity in the dataset.

6.2.3 Dataset for LFA

The original dataset in Table 6. 1 is utilised to investigate the required image processing framework, classification algorithm and an optimum number of features. The original dataset contains 520 images.

Table 6. 1: LFA dataset for intelligent image-based colourimetric test

Sl.	Name	Description	Brand	Lighting Environment	Device	Samples/class	Total
1	original			$\tilde{A}_e = A_i + A_c\ (l)$	Samsung Galaxy S6	65	520
2	new_pH10			$\tilde{A}_e = A_i + A_c\ (l)$	Samsung Galaxy S6	65	65
3	D-A_i			Images captured in without any controlled light	Samsung Galaxy S6	15	120
4	D-dl	Universal pH indicator paper	Fisherbrand® pH	$\tilde{A}_e = A_i + A_c\ (dl)$	Samsung Galaxy S6	15	120
5	D- jl			$\tilde{A}_e = A_i + A_c\ (jl)$	Samsung Galaxy S6	15	120
6	D-i			$\tilde{A}_e = A_i + A_c\ (l)$	iPad Pro	20	20
7	D-j			$\tilde{A}_e = A_i + A_c\ (l)$	Samsung Galaxy J3 Prime	20	20
8	$\aleph\_pH$		Hicarer-pH Test Strips-01	$\tilde{A}_e = A_i + A_c\ (l)$	Samsung Galaxy S6	65	65
9	untested urine dipstick	Reagent strips for urinalysis	Multistix® GP, Siemens	$\tilde{A}_e = A_i + A_c\ (l)$	Samsung Galaxy S6	5	5
Total							**1,055**

$$D\text{-}A_i + D\text{-}dl + D\text{- }jl = D\text{-}lights$$

Let us denote the samples of pH 3-9 in the original dataset as 'D-o1-9'. Once the model was developed, the rest of the datasets were utilised to explore the merits and capabilities of the proposed scheme.

As already discussed in Section 2.5.3 (Chapter 2), Mutlu et al. (2017) used single test strip per class and extended the dataset by changing the format of the file, the orientation of the test strip and capturing the image of the same strip for five times. Their study did not repeat the test on pH strips to consider potential anomalies. They tried to compensate for the issue by capturing the image of the same test strip using different controlled-lighting conditions. We acknowledge that the presented case study is a stable assay which is not subject to as much variations as the TB-test. However, the dataset of Dhar, Mehta, & Sit (2017) indicated that even a stable paper assay could have nonconformities when the solution is not held within the colour pads, and it is spread across the base paper. Our research suggests that this can potentially increase the false positive ROI (region of interest) area during image processing and act as a noise during classification. Consideration of such noise level at different test attempt would have asserted statistical likelihood and demonstrated a more reliable system by Mutlu et al. (2017). In order to rectify the issue, in our research, the original dataset contains 65 images of individual pH indicator strips for each of the mentioned levels. The samples were allowed to have any arbitrary orientation as well as position within the sample plane exposed to the camera (Fig. 6. 2).
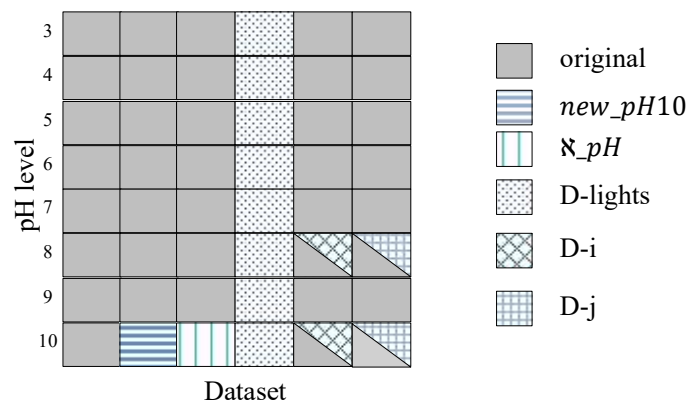
Fig. 6. 3: Proportion and distribution of the dataset

The original dataset was utilised to develop the system from image pre-processing to feature extraction and classification. Using 10-fold cross-validation, the dataset assisted in optimising the features and identifying the best suitable classifier. Later, 535 images

were considered as the extended dataset to evaluate the concept using a wider latitude. The proportion of data distribution is illustrated with a heat-map (Fig. 6. 3).

The purpose of using these extended datasets was to evaluate the robustness of the image processing algorithm, the stability of the classifier and reliability of the overall system by investigating the following questions.

i)   How the system performs on an unseen data under the similar condition?
     - (new_pH10)
ii) Can the system be trained for a similar assay from a different commercial brand?
     - ($\aleph\_pH$)
iii) Does the image process algorithm adaptable to analogous assay type?
     - (untested urine dipstick)
iv) Is the system trainable for images under different illumination condition without fine-tuning the image processing framework, feature set and hyper-parameters?
     -(D-lights)
v)   Is the system trainable for images captured by different devices without fine-tuning the image processing framework, feature set and hyper-parameters?
     - (D-i and D-j)

Extended Dataset

Among these extended samples, aside from the urine dipstick, the rest of the samples were immersed in pH buffer solutions.

i)   At first, a dataset (new_pH10) was created using the same assay brand and same ambient condition, placing the test strips in an arbitrary position within the sample plane with the aim to create more diverse conditions. This new_pH10 dataset was used as a testing dataset to validate the system on the entirely unseen data.

ii) In order to validate the robustness of the image processing algorithm, universal pH indicator paper of a different brand (Hicarer-pH Test Strips-01) was utilised, denoted as $\aleph\_pH$ papers. The dimension, including the thickness of the colour pads of $\aleph\_pH$ papers is slightly different than the original dataset. The colour pads in $\aleph\_pH$ papers are densely situated and the base papers are more

147

hydrophilic than our original test strips. Therefore, the image-processing technique would have to deal with more noise for these test papers.

iii) In the reported articles (Yetisen et al., 2014; Rahmat et al., 2018; Smith et al., 2016; Chen, Wu and Dong, 2014; Wirth et al., 2018), the urine dipstick is a well-utilised example of LFA using with and without additional hardware. Due to the resemblance of the assay, a urine dipstick was included in the extended dataset. Similar to these multi-object assays, e.g. original dataset, the urine dipstick has multiple colour pads. Although these assays are different in terms of the number of targets, the image processing framework should be able to segment the ROIs. Image of a single untested urine dipstick was captured five times with a slight variation in the position for repeatability. Due to different targets, the dataset was not used for classification.

iv) The reported articles emphasize creating a diverse dataset by considering different illumination conditions and mobile devices (Kim et al., 2017a; Solmaz et al., 2018). Therefore, we have further extended our dataset (Table 6. 1) to include the additional three different illumination conditions as described in the experimental setup. The original dataset was generated in an indoor laboratory environment. Without the ceiling lights (in the presence of A_i), for each pH level, 15 samples were generated- a) without any controlled light, using natural daylight only (D-A_i) b) using warm light (D-jl) and c) using cold light (D-dl). Therefore, from 120 independent pH tests, 360 images were produced using these lighting conditions (D-A_i, D-jl, D-dl). From this point forward, these images are collectively denoted as 'D-lights'.

v) The properties of images captured with different devices may vary due to the camera, optics and imaging software (Solmaz et al., 2018), even when the interoperability issue (Yetisen et al., 2014) of the mobile application is not considered. Hence, a small dataset[22] (D-i and D-j) was generated from images captured using different devices to explore the impact (Fig. 6. 3).

---

[22] The dataset of D-i and D-j are relatively small (<30 samples for 2 classes). Rest of the pH samples per class, e.g. original dataset, contain 65 images per class. Therefore, D-i and D-j were not considered for statistical analysis in order to maintain the statistical assumptions.

To summarise the overall dataset, this work includes an LFA dataset of 1,055 images in total.

6.3 Data Processing of the Paper Strips using Computer Vision

The use of an intelligent image processing scheme can make the system user-friendly by reducing the required user intervention to produce the result. A robust image processing algorithm can also eliminate the necessity of additional hardware attachments. In our wet-chemical-based case study (Abuhassan et al., 2017; Shabut et al., 2018) (Chapter-5), we segmented the images using extensive processing including iterative method such as k-medoids clustering. The colour of the samples of our qualitative test were bluish-pink and pinkish-blue. The R and B channels had an overlap. Therefore, k-medoids provided a better performance among the histogram thresholding based image processing techniques. In this chapter, the associated colour of the samples are mostly colour opponents. Using clustering algorithms such as k-means and k-medoids would create more matrices to accommodate the images of multiple clusters before selecting the best cluster. Moreover, due to the multiple colour opponent objects of the same sample, it would be difficult to keep these colour pads in the same cluster. Being a 'multi-object/ sample' type colourimetric test, the ROIs divided into different clusters would occupy more memory space and the feature extraction stage would require more iterations, essentially taking more time to process. Therefore, this work presents an image processing algorithm circumventing the iterative approach, making it more suitable for mobile environments of limited storage and processing capacity.
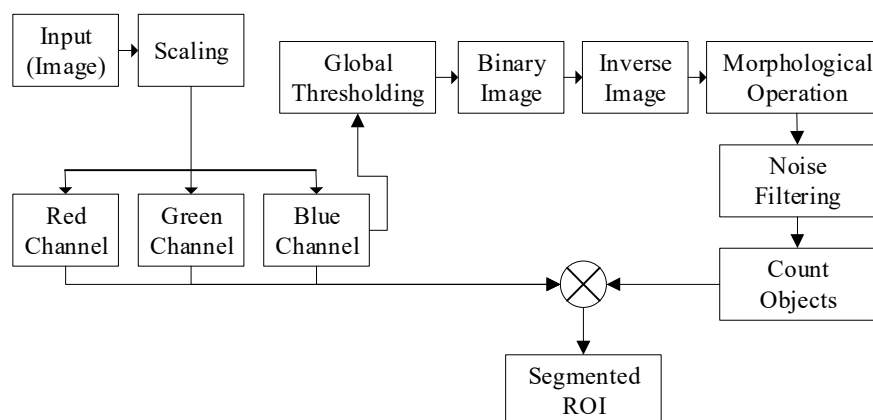
Fig. 6. 4: Image processing framework for LFA based colourimetric assay

149

The key steps of the algorithm are illustrated as a framework in Fig. 6. 4.

At first, the dimensions of the images were reduced by dynamic scaling and then the colour channels were separated.

x_r = 8-bit unsigned integer array of filtered image in red channel

x_g = 8-bit unsigned integer array of filtered image in green channel

x_b = 8-bit unsigned integer array of filtered image in blue channel

The examples used in this chapter, both universal pH strip and urine dipstick contain multi-objects per sample. Using the same concept as Eq. 3.9 (Chapter 3), the relevant colours of our original dataset can be placed in 5 clusters or groups in theory. Therefore, the segmented image can be expressed as Eq. 6.4[23].

$$g(x,y) = \begin{cases} g_1 & \textit{if } f(x,y) = \textit{foreground pixel} \\ & (\textit{colour pad 1}) \\ g_2 & \textit{if } f(x,y) = \textit{foreground pixel} \\ & (\textit{colour pad 2}) \\ g_3 & \textit{if } f(x,y) = \textit{foreground pixel} \\ & (\textit{colour pad 3}) \\ g_4 & \textit{if } f(x,y) = \textit{foreground pixel} \\ & (\textit{colour pad 4}) \\ g_0 & \textit{if } f(x,y) = \textit{background pixel} \end{cases} \qquad (6.4)$$

In order to separate these clusters, clustering algorithm such as k-means, k-medoids and c-means can be used. However, as explained earlier, for rapid execution, the relevant colours should be categorised as the foreground region of interest (ROI) and the background colours (e.g. paper, tricyte) to minimise the required number of clusters, instead of using five clusters as Eq. 6.4.

From the colour histogram analysis (Fig. 6. 5), it can be observed that the higher band of G (green channel) and B (blue channel) holds the background pixels from g_0. However, in the case of green channel, the threshold level overlaps with the foreground

---

[23] There are four colour pads in the universal pH test strip used in this chapter. Therefore, considering the background as well as the foreground (consisting four colour pads), there are five colour groups in theory. If the segmentation is performed based on the 'colour groups', then it would result in five clusters as expressed in Eq. 6.4.

pixels of one of the four groups in Eq. 6.4. Therefore, the background separation can be performed more convincingly in the blue channel.
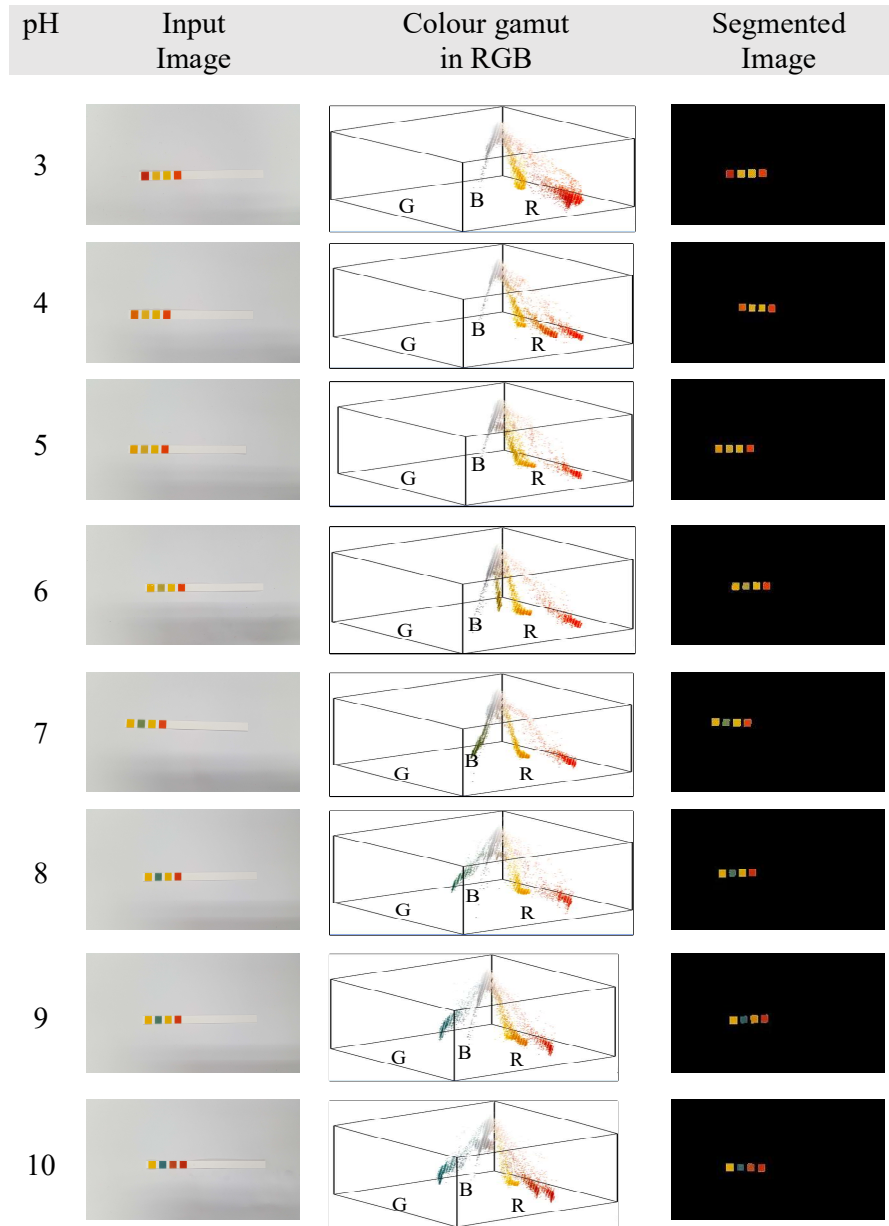
Fig. 6. 5: Histogram-based image segmentation

For better visualisation, Fig. 6. 6 (a) illustrates one of the input images and in Fig. 6. 6 (b) the full colour gamut is represented as point clouds in RGB colour space. When the channels are separated as it is mentioned in the framework (Fig. 6. 4), the R channels holds both background as well as foreground pixels, shown in Fig. 6. 6 (c) and (d).

From Fig. 6. 6 (g), it is visible that the background is better perceivable in the blue channel.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)
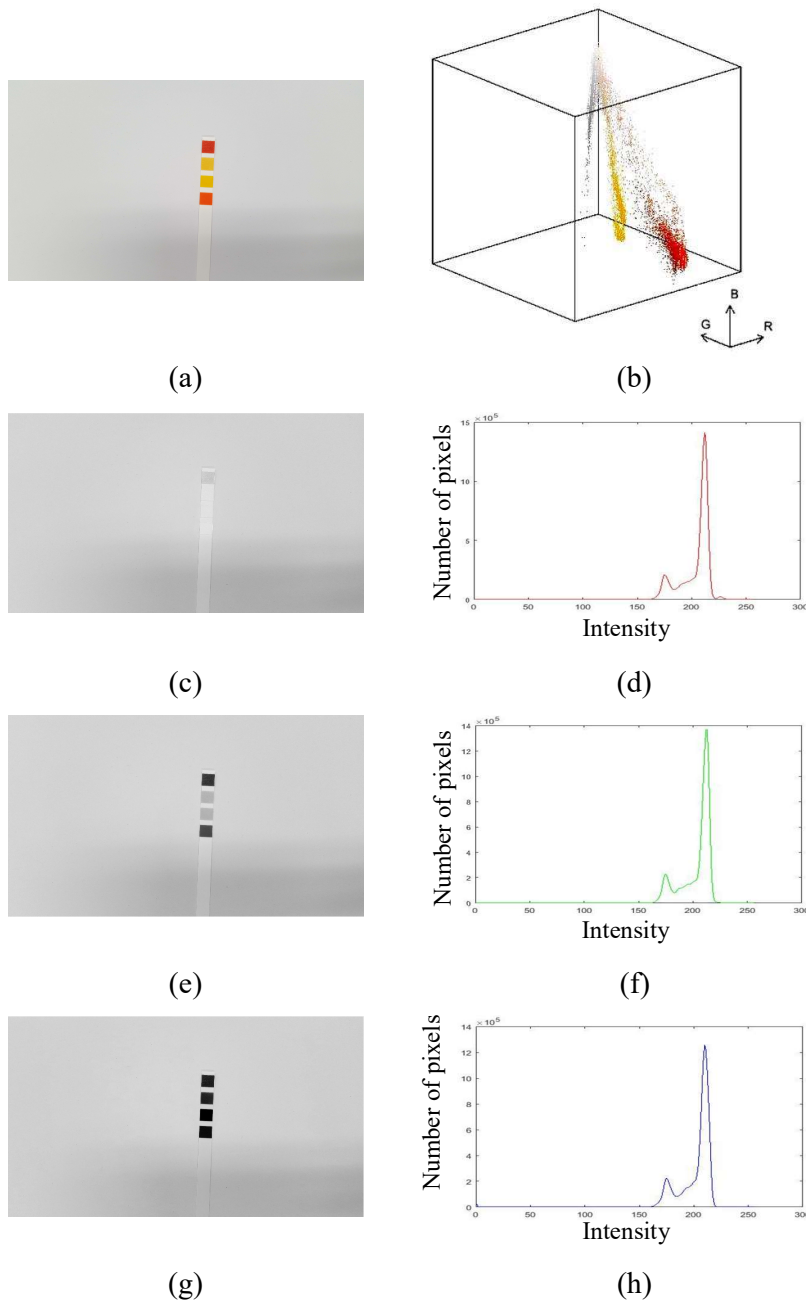
Fig. 6. 6: Histogram of a pH test strip. (a) Paper strip tested on pH 3, (b) three dimensional colour gamut as point cloud in RGB, (c) image 'a' in red (R) channel, (d) histogram of R channel, (e) image 'a' in green (G) channel, (f) histogram of G channel, (g) image 'a' in blue (B) channel and (d) histogram of B channel

The intensity of x_b was then normalised by Otsu (1979), prior converting it into a binary image. At this point, the processed image only contained g_0. The logical negative of the matrix is capable to provide the foreground pixels.

To extract all the blocks of the colour pad, Euler[24] number property was utilised.

I= Binary foreground pixels

$O_i(n)$ = Objects in I by 8-connectivity, where n = 1, 2, ... N

h_o = Number of holes in $O_i(n)$

$e_{L,H}$ = $number\ of\ O_i - number\ of\ h_o$, where L and H are low and high value, respectively.

C = Connected components by 2-D Euler Number $(e_{L,H})$

All the pixels in C lower than the threshold value were eliminated to remove the noise in the binary image. Therefore, the image contained the pixel position of ROI only. As it is mentioned earlier, the main focus of this chapter is the assays that contain multi-objects per sample where the number of objects per sample would vary from assay to assay. In our original dataset as well as the $\aleph\_pH$ dataset, the universal pH test papers hold four colour pads and the decision of the pH level depends on the combination of these colour pads. Therefore, an object counting rule was deployed to increase the reliability of the system, discarding the false positive ROIs.

I_c= 8-bit unsigned integer array from logical array containing C

$$Y = \begin{cases} x_r \times I_c \\ x_g \times I_c \\ x_b \times I_c \end{cases}$$

The final output obtained by the AND gate operation can be expressed as Y and illustrated in Fig. 6. 5 as the segmented image. The algorithm is provided in Appendix E. The performance of the algorithm or framework (Fig. 6. 4) is illustrated in Fig. 6. 7.

---

[24] Although conventionally, Euler number refers to the base of the natural logarithm, $e_{L,H}$ computes the morphological Euler number of regions in a binary image. The Euler numbers are defined in terms of Euler polynomials: $e(n) = 2^n \left(n, \frac{1}{2}\right)$, where n = Index of the Euler number or polynomial. In plain words, the Euler number is the total number of objects in the image minus the total number of holes in those objects.
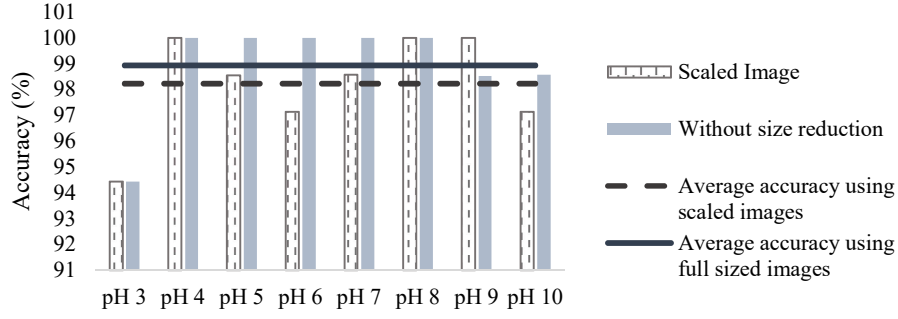
Fig. 6. 7: Performance of the image processing algorithm for pH test

The overall image processing performance showed >98% accuracy. Considering the fact that the light source '5' in Fig. 6. 2 varied during winter and spring in the UK while the images were taken, the image processing method exhibited good accuracy and reliable for the indoor environment.

After image segmentation, due to rule-based noise filtering, false positive ROI area was minimum. The image processing framework showed >97% accuracy for pH 4-10. Among the pH levels, the algorithm showed comparatively poor performance on pH 3. Further investigation revealed that the light source of Eq. 6.2 had more variation on the days when dataset of pH 3 was generated, which created more shadows on the images. The position of the sample plate (Fig. 6. 2) was kept constant for all the experiments. There were shadows due to the ambient objects that could have been avoided by moving the sample plane based on '5' (Fig. 6. 2).

In this chapter, we also evaluated the possibility of performance degradation of the histogram-based image processing algorithm due to scaling operation tested in a desktop environment. Without scaling and keeping the rest of the framework same, the full-sized images provided 98.94% accuracy on average. If the framework (Fig. 6. 4) is followed, then the mean accuracy of the image processing algorithm was 98.23% (Fig. 6. 7). Therefore, it can be stated that the dynamic scaling had a negligible impact on the image quality. There is an emphasis on scaling in the literature for mobile-enabled medical image processing (Bourouis et al., 2014). As the image processing algorithm in this work was not affected by resizing, after applying dynamic scaling, the maximum dimension of the image was 300 on average. The dimension reduction at a different stage, without compromising the performance, effectively reduced the overall memory occupancy of the system (Fig. 6. 8), making it suitable for the mobile environment.
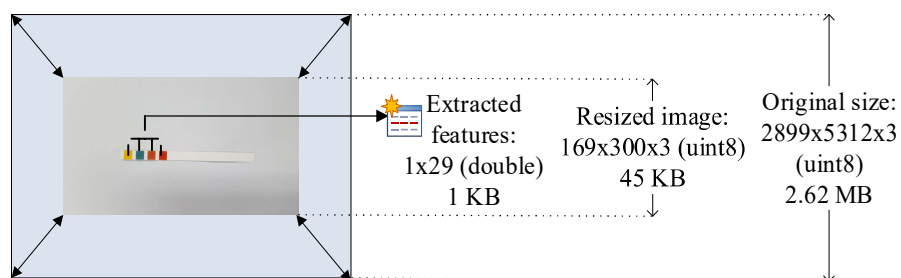
Fig. 6. 8: Example of dimension reduction at different stage. The memory occupany for each image was reduced to ~60% by downscaling the image. After extracting the ROIs, the final feature-set of the whole dataset was less than 1KB.

As shown in Fig. 6. 3, the א_*pH* papers were immersed in pH 10 buffer solution, maintaining rest of the parameters similar to Fig. 6. 2. The image processing framework (Fig. 6. 4) was found to be robust enough to successfully separate the ROI of א_*pH* papers without any additional fine-tuning (Fig. 6. 9).

Sample	Input Image	Segmented image
pH 10		
Untested urine dipstick		

Fig. 6. 9: Image processing of א_*pH* paper using pH 10 buffer solution and untested urine dipstick

In order to further examine the adaptability of the image processing algorithm, another example (Multistix® GP, Siemens) of a lateral flow assay was utilised, consisting eight pads for different indicators (such as glucose, ketone, pH). There are two additional blocks for referencing. These pads or blocks have similar length and width as our original dataset. However, the thickness and block-to-block distance are different. Moreover, the base is made of trycite, which are more hydrophobic than paper and has different reflectance. These test strips are multi-objects/ sample assays. However, each of these objects is designated or targeted for different test (e.g. pH, glucose) performed on the same sample (e.g. urine of a subject). Although each image contains multiple

objects for a single test strip, they are single object/ target, known as multiplex assay. As illustrated in Fig. 6. 9, the image processing framework (Fig. 6. 4) was successful to separate the ROI of the multiplex assay of urine dipstick.

Different illumination conditions can influence the performance of an image processing algorithm (Smith et al., 2014). However, the proposed image processing framework showed consistent performance for D-A_i, D-jl and D-dl. The framework provided 353 correct segmentation out of 360 images with 98.06% accuracy.

The same pH test strip under different illumination conditions is visibly showing different histogram pattern using the same mobile phone camera in Fig. 6. 10.

$$- \tilde{A}_e = A_i + A_c \ (l) \quad \ldots \ \tilde{A}_e = A_i \quad \Delta \ \tilde{A}_e = A_i + A_c \ (jl) \quad * \ \tilde{A}_e = A_i + A_c \ (dl)$$
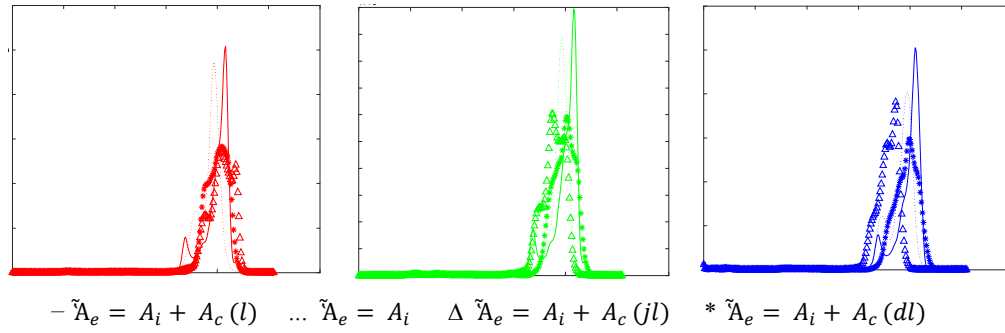
Fig. 6. 10: Histogram of an image of a pH test strip under different illumination across three colour channels. In the colour histogram, the horizontal and vertical axis represent the intensity and the frequency or number of pixels, respectively.

The histograms in Fig. 6. 10 confirm the need to include diverse illumination condition to develop a robust system, which is supported by the reported articles as well (Kim et al., 2017a; Solmaz et al., 2018).

In Fig. 6. 11, the impact of capturing an image of the same pH test strip using same the illumination condition can be observed while incorporating the camera of a number of personal devices. Comparing the colour histogram of Fig. 6. 10 and Fig. 6. 11, it appeared that the images can be more affected by the illumination condition than the variation of the mobile devices.

In addition to Samsung Galaxy S6 (original dataset), the images were captured by two other devices: iPad Pro and Samsung Galaxy J3 Prime. The sizes of these devices are different, which effectively varied the exposure plane shown in Fig. 6. 2.

| — Samsung Galaxy S6 | ⁻⁻ iPad Pro | ☆ Samsung Galaxy J3 Prime |
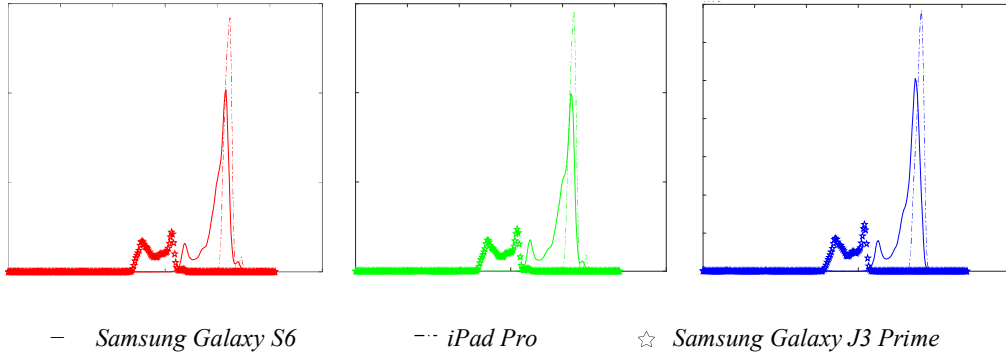
Fig. 6. 11: Histogram of an image of a pH test strip across three colour channels using different mobile phone cameras. In the colour histogram, the horizontal and vertical axis represent the intensity and the frequency or number of pixels, respectively.

Therefore, the focal length of these devices had to be adjusted for the consistency of the overall system. The dimension of Samsung Galaxy J3 Prime is similar to Samsung Galaxy S6. Therefore, no adjustments were conducted on the focal length for this device. The iPad Pro was held at 11 inches away, varying the 35mm focal length within 66-68. It should also be mentioned that the images captured by the different devices varied in size due to the system-defined standardisation. Nevertheless, the image processing algorithm (Appendix E) showed consistent performance (Fig. 6. 12) to segment the images captured by different devices. It was also observed that the presented system is capable of handling such variation in the focal length.

Device	Input Image	Segmented Image
iPad Pro		
Samsung Galaxy J3 Prime		

Fig. 6. 12: Image segmentation by different devices

6.4 Feature Extraction

Once the ROIs were segmented, the characteristics of these samples were analysed from its colour moments. However, it should be noted that the colourimetric classification based on the colour moments depends on the successful separation of the segmented ROIs (Table 3. 1, in Chapter 3).

In our earlier study (Abuhassan et al., 2017), we have utilised lower order colour moments, c_m (Sergyan, 2008) in only LAB colour space (cs) for all the colour channels (c_c) to classify wet-chemical-based qualitative colourimetric tests. For stable paper assays such as pH indicator strips and urine dipsticks, the required feature-set may vary. Therefore, the impact of different attributes is required to be analysed. This thesis expresses the required features (ft) for colourimetric tests as Eq. 6.5.

$$ft = \{ft_{cs}(c_m) \times ft_{cs}(c_c)\} + ft(\Delta E_{LAB}) \tag{6.5}$$

where ΔE_{LAB} represents the colour difference with respect to the control colours.

The feature-set in Abuhassan et al. (2017) can be described as Eq. 6.6, where six colour moments where considered in L, a and b channel, discarding the entropy in L channel. As the images involved single objects per sample, there were 17 features in total.

$$ft = ft_{LAB}(c_m) \times ft_{LAB}(c_c) - Entropy_L \tag{6.6}$$
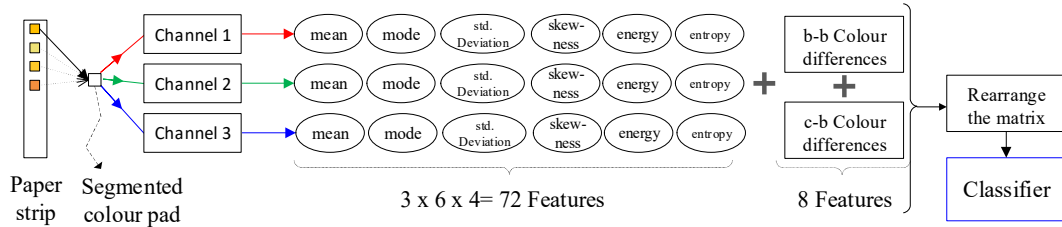
Fig. 6. 13: Feature extraction at initial stage before feature optimisation

The feature extraction framework is illustrated in Fig. 6. 13. In this work, the colour differences in LAB colour space are calculated in a closed loop as additional features (Tania et al., 2017). If the colour block is b_p and position of the block is p; $p \in \mathbf{Z}^+$, then let's measure the colour difference for each block calculated from its previous block using the following equation.

$$\Delta E_{b_p}^* = \sqrt{\{(mean\ L_{b_p}^* - mean\ L_{b_{p-1}}^*)^2 + (mean\ a_{b_p}^* - mean\ a_{b_{p-1}}^*)^2 + (mean\ b_{b_p}^* - mean\ b_{b_{p-1}}^*)^2\}} \tag{6.7}$$

The features generated from Eq. 6.7 are novel features, where a pseudo-control colour set is generated for each test using the user input image itself. However, in real life situation, one would require to compare the tested paper assay with a colour chart. This colour chart holds the control colour for the block or sample. In the literature, the colour of the block is often tracked against the colour chart (Rahmat et al., 2018; Solmaz et al., 2018). In this work, we have also used the colour difference of each colour block from the corresponding control colour block (c_p) as part of the feature-set (Appendix D).

$$\Delta E_{c_p}^* = \sqrt{\{(mean\ L_{c_p}^* - mean\ L_{b_p}^*)^2 + (mean\ a_{c_p}^* - mean\ a_{b_p}^*)^2 + (mean\ b_{c_p}^* - mean\ b_{b_p}^*)^2\}} \tag{6.8}$$

In Eq. 6.5, $ft(\Delta E_{LAB})$ was obtained from Eq. 6.7 and Eq. 6.8, which could be stated as Eq. 6.9.

$$ft(\Delta E_{LAB}) = \Delta E_{b_p}^* + \Delta E_{c_p}^* \tag{6.9}$$

The case study involves multi-objects for single sample per image, which elongated the feature set. Therefore, initially 440 features[25] in total were considered to train the classification model.

6.5 Classification and Feature Analysis

6.5.1 Feature Selection

At first, the classifiers were trained using 440 colour features. Among the standard classifiers including discriminant analysis, SVMs, KNNs and ensemble methods, the ensemble method called Subspace discriminant analysis showed the best performance.

---

[25] Eq. 6.5: (6 colour moments x 4 blocks x 3 colour channels x 6 colour spaces) + Eq. 6.9 = **432** features + Eq. 6.9;
Eq. 6.7: No of features: 1 x4=**4**;
Eq. 6.8: No of features: 1 x4=**4**;
Eq. 6.9: No of features 4+4=**8**.
Therefore, Eq. 6.5= 432+8= **440 features**

Using 30 base learners and subspace dimension of 220, the classifier provided 99.4% accuracy with three misclassifications. The training was conducted in 33.418 seconds, and the prediction speed was in 350 samples/seconds. However, more features would mean more model complexity and requirement of more storage and processing capacity. Therefore, the number of features was required to be optimised. The optimisation was conducted using univariate analysis at different stages. For the convenience of faster training, the analysis was executed via MATLAB classification learner application.

The use of histogram features in the reported articles can be observed from Table 6. 2.

Table 6. 2: Region of interest, feature-set and classification algorithm for paper-based assays

Reference	Object/ sample	Histogram Features	Colour Spaces	Algorithm	Result Accuracy
H. Kim et al. (2017)	Single	Mean, Median, Mode, Bin median and Bin centroid	RGB, HSV, YUV, LAB	LDA, SVM, ANN	80%; 100%
Solmaz et al. (2018)	Single	With and without grey-world corrected mean	RGB, HSV, LAB	LS-SVM, RF	90.3%; 95%
Mutlu et al. (2017)	Multi	Mean of JPEG, RAW and RAWc	RGB	LS-SVM	100%
Rahmat et al. (2018)	Multi*	Mean	RGB	Euclidean distance	95.45%

Multi-object per sample but single object per target (Multiplex assay)
LDA: Linear Discriminant Analysis; SVM: Support Vector Machine; ANN: Artificial Neural Network;
LS-SVM: Least squares support vector machine; RF: Random Forest

This chapter systematically investigated the impact of the histogram features including colour moments and colour spaces. Additionally, the impact of control colours was studied as well. We have explored the colour moments in LAB colour space to begin with (Fig. 6. 14).

Based on the univariate analysis conducted on $ft_{LAB}(c_m)$, the best performing features were found to be mean and energy (Fig. 6. 14). These two features comprise a good

signal, performing as proficient (97.1% accuracy) as the combined features of $ft_{LAB}(c_m)$.
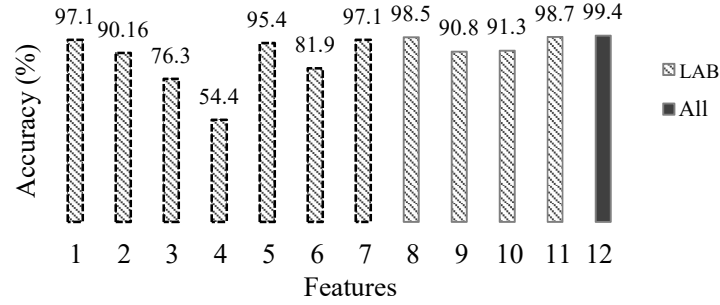
1: Mean; 2: Standard deviation; 3: Mode; 4: Skew; 5: Energy; 6: Entropy; 7: Colour moments; 8: All features in LAB (Eq. 6.5); 9: Pseudo-control colours (Eq. 6.7); 10: Control colours (Eq. 6.8); 11: Both category of control colours (Eq. 6.9); 12: 440 features in 6 colour spaces. The features with dotted outlines are functions of $ft_{LAB}(c_m)$

Fig. 6. 14: Performance of the features for the pH test

Considering the fact that we are analysing the colourimetric tests, the average colour or brightness of the ROI is a key feature. The mean colour value is considered to be the most important features in the reported articles as well (Mutlu et al., 2017; Rahmat et al., 2018). On the other hand, the energy in L, a and b are the amplified brightness level. Therefore, energy is directly linked with the performance of the mean colours. Due to the reflectance on the two dimensional surface, the unbiased standard deviation within each colour pad did not provide a significant contrast to improve the classification performance. The mode of colours on the solid surface of individual ROI without any opto-mechanical attachment can be misleading. Because, the system would be susceptible to the ambient lighting environment. The mode is supposed to provide effective information, if the ROI is scaled such as hardware based systems of CLINITEK Status® + Analyzer and the wavelength is filtered (Siemens Healthcare GmbH, 2018). On an ideal condition, there should not be any asymmetry of colour distribution, therefore the SKEW can also be discarded from the feature-set. Hence, among the features from $ft(c_m)$, only mean and energy were considered to train the classifier.

As illustrated in Fig. 6. 14, the control colours were also found to be influential features. Only the control colours (both categories) are sufficient enough to provide 98.7% accurate classification using Eq. 6.9, which was studied further while exploring and fine-tuning the classifier. Similar to our earlier work (Tania et al., 2017) on wet-

161

chemical-based colourimetric test, the colour differences (Eq. 6.7-6.9) are considered in LAB space only. The L, a and b, imitating the nonlinear response of the human eyes, can also resemble the uniform changes in the perceived colour facilitated by the uniform changes in the LAB- components. Therefore, the control colour related calculations were computed in a colour space which has more advantages at Euclidean space.

The standard devices capture the image in sRGB colour space. After the initial assessment of the colour moments and control colours as features, the colour spaces were appraised (Fig. 6. 15).
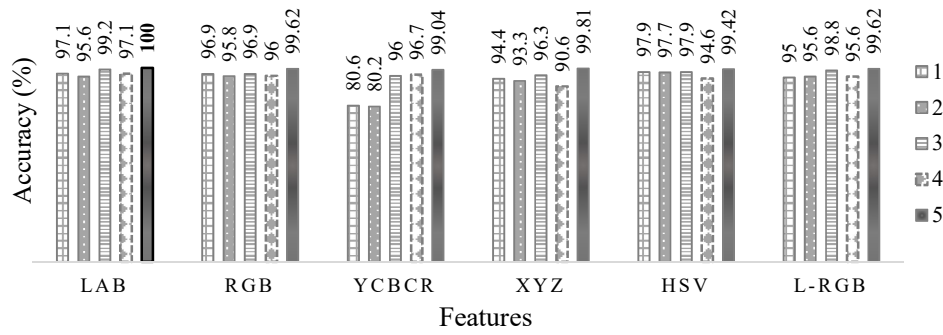
1: Mean; 2: Energy; 3: Mean, energy and pseudo-control colour; 4: Colour moments;
5: Mean, energy and pseudo-control colour using LS-SVM

Fig. 6. 15: Performance of colour spaces using top-performing classifiers with respect to top performing features identified at the previous stage. 1-4 are the performances by the best performing classifier excluding LS-SVM and 5 is the performance using LS-SVM.

H. Kim et al. (2017) utilised four colour spaces to provide a colourimetric decision, whereas Solmaz et al. (2018) used three (Table 6. 2). Mutlu et al. (2017) used RGB images in JPEG along with capturing the images at the RAW format. In this work, we have studied the performance of six colour spaces on the original dataset. Among the colour spaces, the LAB was found to be the most influential colour space. The strength of the LAB is the perceptual uniformity property.

Another popular dimension reduction technique, PCA, combining correlated attributes to create superior new features, did not improve the overall performance. Therefore, based on the performance of different features and colour spaces, mean, energy and the

control colours in three channels of the LAB were chosen to be feature-set (32 features) to explore the performance of the classifiers.

6.5.2 Performance of Classification Methods

Exploring 440 features, the selected 32 features were identified as the good features to ensure that the classifiers are trained with signals, not noise. Different supervised learning techniques were evaluated to provide the semi-quantitative colourimetric classification. The list of classifiers includes LS-SVM, LDA and RF which provided good accuracy in the reported articles for similar classification task (Mutlu et al., 2017; Solmaz et al., 2018; Dhar, Mehta and Sit, 2017). The classifiers were trained and 10-fold cross-validated by the original dataset, followed by a re-evaluation of the good features (mean, energy and control colours).

After careful selection of 32 features, SVM, KNN and discriminant analysis exhibited similar performance in Fig. 6. 16 (>98% accuracy).
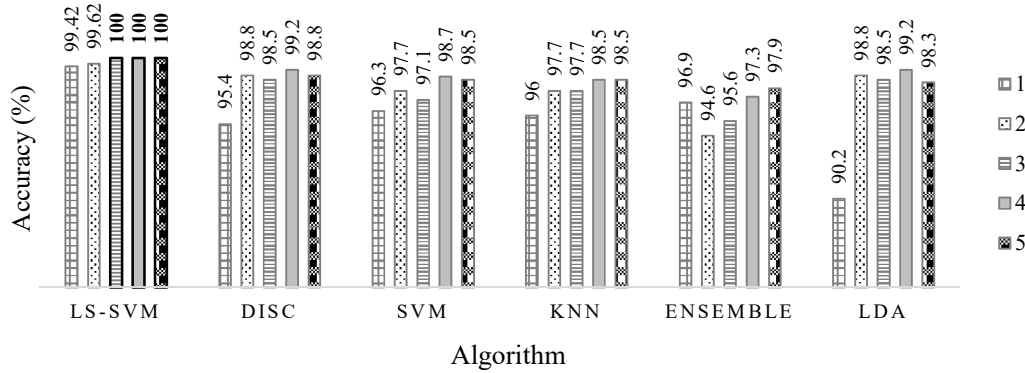
1: Mean; 2: Mean and energy; 3: Mean, energy and pseudo-control colours; 4: Mean, energy and control colours; Mean, energy and both category of control colours

Fig. 6. 16: Performance of different classifiers

The overall performance from the discriminant analysis was good. For a different combination of the good features, the LDA and quadratic discriminant analysis outperformed each other. Therefore, subspace discriminant also exhibited good performance as an ensemble method. The performance of RF was also notable. However, the LS-SVM (Suykens and Vandewalle, 1999; Suykens, 2002) showed the best performance using the selected features as illustrated in Fig. 6. 16. Due to the weighted function with a modified cost function, LS-SVM is more robust than SVM.

The performance of LS-SVM was consistent for any good features in any colour space (Fig. 6. 15 and Fig. 6. 16).

After identifying the best performing classifier using good features or signals (as proposed in Fig. 3. 11 in Chapter 3), the selected 32 features were explored again which revealed that the features can be further reduced to 28. The selected 32 features, including both feature-set of control colours (i.e. Eq. 6.9), LS-SVM provided 100% accuracy. As shown in Fig. 6. 16, in addition to mean and energy, the use of only one feature-set of control colours, i.e. either Eq. 6.7 or Eq. 6.8, is capable of providing 100% accurate colourimetric classification for LFA. Therefore, instead of Eq. 6.9, only one set of control colours can be used. The final performance can be seen in terms of ROC and area under the ROC curve (=1) in Fig. 6.17 as well.

area: Area under the ROC curve; std: Standard deviation
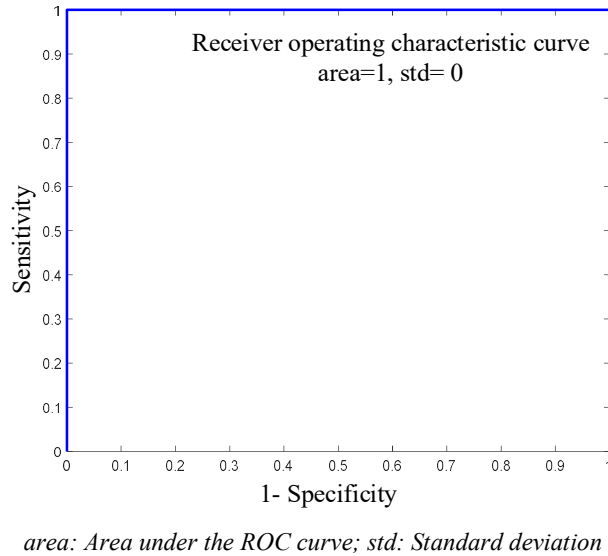
Fig. 6. 17: ROC curve of LS-SVM for the pH test using the optimised 28 features[26]

Furthermore, in order to specify between the choices of control colours, this work suggests, it is better to use the pseudo-control colours (Eq. 6.7), especially for multi-object single-target colourimetric tests. The control colours (Eq. 6. 8) generated from the colour chart are the features generated one time, acting as a ground truth may widely vary from the condition where the user input image is captured. Thus, the user input can appear as a noisy image. The pseudo-control colours (Eq. 6.7) are generated each

---

[26] Due to the equal and perfect classification, all 8 classes are clustered together.

time using the user input itself. Therefore, these features are more reliable. Finally, the optimised feature-set can be summarised as mean, energy and pseudo control colours in three colour channels of LAB colour space.

From Table 6. 3 as well as from Fig. 6.17, it can be observed that the specificity and sensitivity achieved by LS-SVM is also 100%, which are our key evaluation criteria. Among the reported articles in Table 6. 2, Mutlu et al. (2017) performed the colourimetric classification utilising the same application, i.e. paper-based pH test strip. Similar to Mutlu et al., this research also found LS-SVM to be the best performing classifier for the presented classification problem. Mutlu et al. showed that, with and without apparatus, the system can exhibit the same performance. The experiment of Mutlu et al. without any apparatus included 270 images to provide the classification of pH strips, which explains the confidence interval percentage in Table 6. 3. The 95% CI for accuracy, specificity and sensitivity works in favour of our proposed scheme due to the choice of the dataset.

Table 6. 3: Comparative performance of LS-SVM with the reported article

Ref.	Number of independent tests/ class	Number of classes	Number of samples/ class	Sensitivity/ class (%)		Specificity/ class (%)		Accuracy/ class (%)	
				Value	95% CI	Value	95% CI	Value	95% CI
Mutlu et al. (2017)	1	15	18	100	81.47-100	100	98.56-100	100	98.66-100
Proposed method	65	8	65	100	94.48-100	100	99.19-100	100	99.29-100

Mutlu et al. (2017) utilised images saved in different file formats which increased the volume of the data, however did not carry any significance regarding features or classifiers. Therefore, the effective feature-set of the experiment of Mutlu et al. can be considered as the mean colours at RGB colour space. Using the feature-set of Mutlu et al. [27] on our original dataset, the performance degraded from 100% to 98.85% accuracy

---

[27] In case of literary work re-implemented using our dataset, it is cited with the symbol 'ŧ'. For an example, Mutlu et al. (2017) as Mutluŧ.

(Fig. 6. 18), which can be perceived from the data generation point of view. Due to the light source '5' in Fig. 6. 2 and use of 65 independent test strips for each class, the original dataset contains much robust data, whereas Mutlu et al. used the same pH test strip at different orientation using three different light sources. The absence of rarity (Weiss, 2004) generated from independent test strips constrained Mutlu's approach regarding robustness and reproducibility.
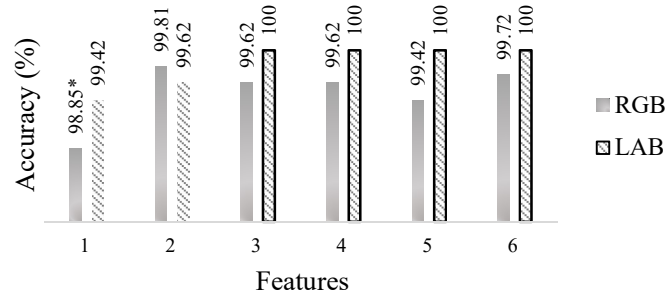
Using original dataset- 1: Mean; 2: Mean and energy; 3: Mean, energy and pseudo-control colour; 4: Using new_pH10 and 28 features (mean, energy and pseudo-control colour); 5: Using ℵ_pH and 28 features (mean, energy and pseudo-control colour); 6: Using D-lights and 28 features (mean, energy and pseudo-control colour)
*Using the feature-set of Mutlu[t]

Fig. 6. 18: Comparative performance of selected featured in LAB and RGB using LS-SVM

Mutlu et al. (2017) used RGB colour space and our research suggested LAB to be a better choice for the presented problem. For the chosen case study, the same 12 features of Mutlu[t], i.e. mean colours in the LAB, instead of RGB colour space showed better performance due to its strength of colour separation and handling shadows. The RGB is good at modelling the output of the phone camera, but LAB is closer to the 'human colour perception' and the presented problem deals with mimicking the naked-eye measurement of the colourimetric tests using computer vision. Therefore, LAB is a better choice of colour space.

To further explore the performance of the classifier as well as the optimised 28 features on unseen images and to validate the system using a similar lateral flow assay, we extended our experiment using *new_pH*10 and ℵ_*pH* from Table 6. 1. The extended experiment can validate the reliability, adaptability and robustness of the system. The *new_pH*10 contains 65 images of pH test strips of level 10 at various random

orientation. The classifier was not effected by more variation in the orientation, as it provided 100% accuracy using 10 fold cross validation (Fig. 6. 18).

In order to validate the system's adaptability to similar lateral flow assays, the א_*pH* dataset was utilised. A similar dataset is available in GitHub (Dhar, Mehta and Sit, 2017). However, there are only 10 samples per class available in the open source domain. The pH indicating colours of the א_*pH* dataset are different than our original dataset, therefore the classifier was re-trained using the optimised 28 features, keeping rest of the hyper-parameters same. The 10 fold cross-validation showed the consistent result with 100% accuracy (Fig. 6. 18).

Using a pH indicator paper of different brand (א_*pH* dataset) effectively changed the colours of each block for the same class label. There was a slight variation in the block size and block to block distance as well. As the performance of the system was as good as the original dataset, it justifies our choice of the classifier as well as the feature set, endorsing the reliability and adaptability of the system.

After finalising the classifier, optimising the features and cross-sectional performance evaluation, the extended dataset (D-lights) was utilised. As mentioned earlier, due to significant shift in the colour histogram (Fig. 6. 10 and Fig. 6. 11), consideration of such diversity would enhance the reliability of the system. After including these 360 images with the original dataset, the optimised features and LS-SVM showed consistent performance for the dataset (D-lights). The effect was further analysed separately without including the original dataset (Fig. 6. 18).

6.5.3 Statistical Analysis

Based on the results obtained by different classifiers in Fig. 6. 16 and the classifiers mentioned in the reported articles (Table 6. 2), top five classifiers were selected for further statistical analysis. These classifiers are: LS-SVM, LDA, SVM, and two ensemble algorithms RF and subspace discriminant. In Section 6.5.2, the obtained accuracy is relatively close. Hence, the statistical analysis was necessary to claim the performance and establish reliability. In this research, we compared the performance of the best performing classifiers by using Student's t-test. As this chapter includes 30 observations of the best performing algorithms, normality is assumed by the central limit theorem.

The performances mentioned in Fig. 6. 14-6.17, are the mode of accuracies. Therefore, at random execution of the algorithm, these are the accuracies more likely to be achieved. However, for statistical analysis, each of the algorithms were executed 30 times; the average accuracy achieved by LS-SVM on three different datasets are shown in Fig. 6. 19.

1: original dataset; 2: new_pH10; 3: ℵ_pH;
4: D-lights using 28 features (mean, energy and pseudo-control colours)
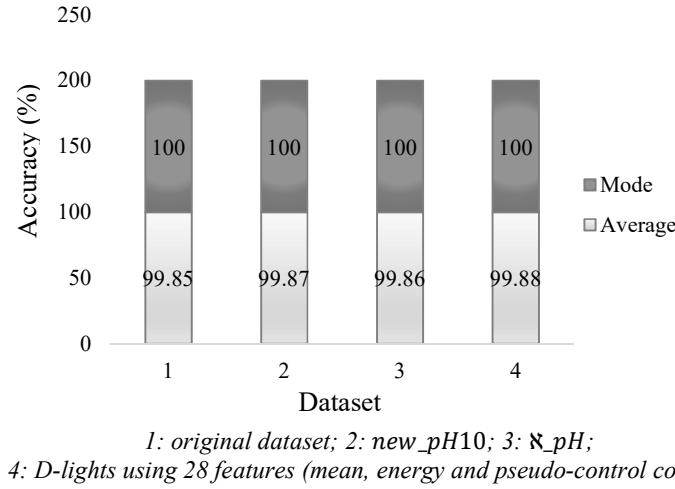
Fig. 6. 19: Average performance of LS-SVM on different dataset

The statistical results obtained by a two-tailed t-test at a 0.05 level of significance are given in Table 6. 4 (Lwin, 2015; Hatem Al-jamal, 2017). The result of Algorithm 1 ↔ Algorithm 2 stated as "+", "−", or "~" can be read as- Algorithm 1 is significantly better than, worse than, or statistically equivalent to Algorithm-2, respectively. Among the listed algorithms in Table 6. 4, none of them was statistically equivalent[28]. Based on the observations and Table 6. 4, it can be stated that LS-SVM outperformed the rest of the algorithms for all the dataset, which justifies our choice of the algorithm.

Although LDA performed well on the original dataset, the performance was significantly reduced as compared to SVM and RF when the classifiers were exposed to more variations in terms of colours and illumination. LDA being a parametric method is highly constrained to the specific form and exhibited less robustness as compared to SVMs and RF.

---

[28] We acknowledge that a substantial amount of observations has an impact on the significance level, i.e. significance is not equivalent to importance.

Among the ensemble methods, subspace discriminant analysis requires less memory space and showed good performance at initial stage. This random subspace ensemble performs well for multiclass classification problem when there are more features. Therefore, after feature optimisation, the subspace discriminant showed significantly poor performance among the classifiers mentioned in Table 6. 4.

Table 6. 4: Student's t-test results of the top performing classifiers for the pH test

Algorithm 1 ↔ Algorithm 2	original	new_pH10	$\aleph\_pH$	D-lights
LS-SVM ↔ LDA	+	+	+	+
LS-SVM ↔ SVM	+	+	+	+
LS-SVM ↔ RF	+	+	+	+
LS-SVM ↔ Sub. Disc.	+	+	+	+
LDA ↔ SVM	+	-	-	-
LDA ↔ RF	+	+	-	-
LDA ↔ Sub. Disc.	+	+	+	+
SVM ↔ RF	+	+	+	+
SVM ↔ Sub. Disc.	+	+	+	+
RF ↔ Sub. Disc.	+	+	+	+

6.6 Regression

The qualitative colourimetric test can be seen as a pure classification problem, whereas the quantitative colourimetric test can be presented as a regression problem. The semi-quantitative colourimetric tests such as pH test can be described with a classification as well as a regression model (Fig. 6. 20).

Among different regression model such as linear, support vector machine and ensemble methods, the Gaussian Process regression (GPR) (Rasmussen et al., 2006) showed better performance using the optimised 28 features (Fig. 6. 20). The performance was compared by the square root of the mean squared error (RMSE). It is an estimation of the standard deviation of the error distribution. The coefficient of determination (R^2) value suggested that the model can explain approximately 99% of the variability in the

selected response variables. It took 11.51 seconds to train the model and the prediction speed was approximately 7300 observations per second. We have utilised an isotropic kernel. The covariance function, $k(x_i, x_j)$ estimates the course of response at point x_i effecting the response at a further point x_j, where x_i = predictor values, $i = 1 + 2 + 3 + \cdots + n$ and $i \neq j$. The Euclidean distance between x_i and x_j, $r = \{(x_i + x_j)^T(x_i - x_j)\}^{1/2}$. If y_j = target, σ_f = signal standard deviation, $\sigma_f > 0$ and σ_l = characteristic length scale, $\sigma_l > 0$, then the kernel function having same length scale for each predictor utilised in this work can be expressed as Eq. 6.10 (Rasmussen et al., 2006).

$$\text{Covariance function, } k(x_i, x_j) = \sigma_f^2 \left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2} e^{\left(-\frac{\sqrt{5}r}{\sigma_l}\right)}\right) \tag{6.10}$$
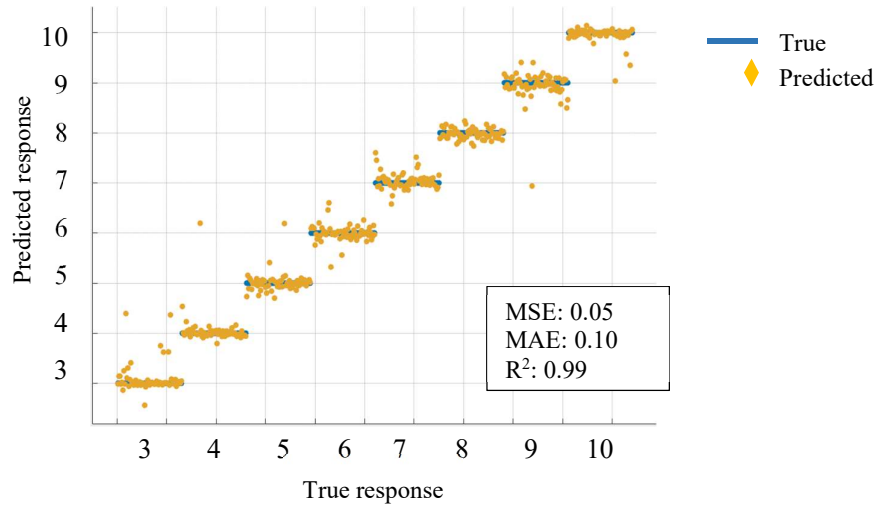
Fig. 6. 20: Performance of Gaussian Process Regression. RMSE: 0.22

In this work, as the system was trained on <1000 samples using the best performing features only, the GPR performed better than the other models due to well-balanced bias and variance, smoothing, optimised hyper-parameters and local generalisation. The kernel scale parameter attained the random basis for random feature expansion utilising sub-sampling-based heuristics. The model may require further rectification using continuous pH levels (quantitative) as well as a larger dataset to substantiate the performance.

6.7 Comparative Performance and Evaluation by ASSURED Criteria

This work evaluates the performance of analogues computational systems for lateral flow assays using ASSURED criteria. In the absence of the exact dataset, in this work, we have compared the performance based on the concept of recently reported high accuracy systems (Table 6. 5).

Table 6. 5: Comparative performance using ASSURED criteria

Reference	Application	A	S	S	U	R	E	D
H. Kim et al. (2017)	Alcohol saliva test	↓	PPV-NPV: >95%		↓	Robust, <30s	No	Yes
Solmaz et al. (2018)	H_2O_2	↑	N/A	N/A	↓	Robust, N/A	Yes	Yes
Mutlu et al. (2017)	pH test	↑	100%	100%	↓	Robust, N/A	Yes	Yes
Rahmat et al. (2018)	Urine dipstick	↑	>98.25%	>98.25%	↓	Not robust, N/A	Yes	Yes
Proposed method	pH test	↑	100%	100%	↑	Robust, Real-time	Yes	Yes

N/A: Information not available

6.7.1 Affordable, Equipment-free, User-friendly and Accessible System

As mentioned in Section 3.6 (Chapter 3) regarding the evaluation criteria, the affordable systems are more likely to be equipment-free and more accessible. In Table 6. 5, the expense of the systems are comparative, e.g. although the additional hardware attachment is low-cost and straightforward (Kim et al., 2017a), it would be still more costly than a system which does not require such attachments at all. The presented system in this work utilises the built-in camera of the personal devices such as mobile phone without enhancing or channelling the light with any additional hardware such as in Kim et al. (2017a), making the system convincingly more portable and easily operable.

6.7.2 Specificity and Sensitivity

The accuracy of the reported articles is presented in Table 6. 2. The specificity and sensitivity (or similar metric such as precision-recall, type I- type II error, PPV-NPV) are often not described in the articles. Based on the available information, a comparison is provided in Table 6. 5. It should be also taken into account that the dataset of Table 6. 5 varied in terms of number of test strips, images of the same sample for repeatability and variation within dataset. The data were often pre-processed manually or with an aid of additional hardware.

Table 6. 6: Comparison of accuracy (%) conducted on the original dataset[27]

Features	LDA	SVM	ANN	RF	Sub disc	LS-SVM
Kim[t]	Failed	97.7	68.8	97.9	99.6	100
	= 97.7					
Solmaz[t]	98.7	97.5	57.8	**96.3**	98.3	**99.62**
Multu[t]	88.1	96.9	64.2	91.2	76	**98.85**
Proposed method	98.5	97.1	83.1	96.5	92.7	**100**

Sub disc: Subspace discriminant. Highlighted cells represent the algorithms used in the original articles

Using the original dataset, we have trained the better performing classifiers affirmed in the reported articles using the same feature-set as mentioned in Table 6. 2. In the case of few missing hyper-parameter values, we have utilised the default values in MATLAB. Rahmat et al. (2018) of Table 6. 2 is excluded from Table 6. 6, as it involves the same feature-set as Mutlu et al. (2017). Analysing the performance from these tables, the justification behind the choice of our classifier, i.e. LS-SVM is well supported by Table 6. 2 and evident from Table 6. 6. Among the classifiers, the ANN with ten hidden layers (Kim et al., 2017a) showed a poor performance, which requires further investigation.

A detailed comparison with Mutlu et al. (2017) is already presented earlier, in Section 6.5. Both paper test strips of H. Kim et al. (2017) and Solmaz et al. (2018) has only one colour pad. Therefore, average binning of four colour pads could not aid the classifier in Table 6. 6 while reproducing Kim's work using our dataset. H. Kim et al. (2017)

utilised hardware attachment as well, which helped to discard noise from the colour signals. In this chapter, using the feature-set and classifier of H. Kim et al. (2017) on our original dataset, the attained accuracy was 97.7%, whereas the same feature-set of Kim[t] provided a higher accuracy using LS-SVM as well as the ensemble classifiers (Table 6. 6).

Solmaz et al. (2018) used LS-SVM and RF utilising with and without grey-world corrected mean in RGB, HSV and LAB colour space. The performance of the combination of the features of Solmaz et al. (2018) in Table 6. 6 can be perceived from the performance of mean colours in different colour spaces in Fig. 6. 15.

Our research has managed to optimise the number of features (both colour moments and colour spaces) and proposed pseudo-control colours as an additional feature-set, which demonstrated better accuracy, specificity and sensitivity than the existing systems as shown in Table 6. 6.

6.7.3 Robust and Rapid System

In Table 6. 5, different works included a different element of robustness. As described earlier in Section 2.6.4 (Chapter 2), the robustness of the system can be represented by adaptability. In this work, we utilised analogous separate data set to evaluate the adaptability of the image processing algorithm, selected optimised feature-set and classifier. The system was found to be adaptable to the new sample sets.

The elapsed time for training and prediction for the systems described the reported articles, evaluated using our original dataset is shown in Fig. 6. 21.
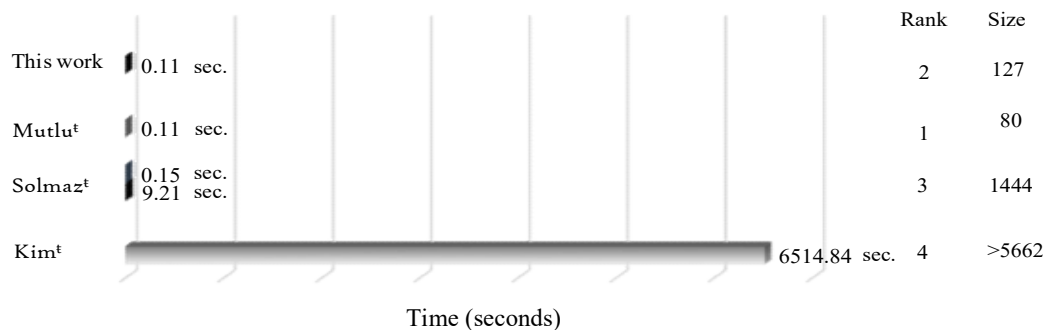
Fig. 6. 21: Comparative performance in terms of elapsed time for classification

A ranking in Fig. 6. 21 is provided based on the computational cost and model size. A higher number of features would undoubtedly increase the size of the model. The size can also get affected by the complexity of the classifier itself.

The system proposed in this chapter, does not involve heavy algorithm or extensive iterations, making it computationally efficient to be deployable on the mobile devices using native features without requiring it to process the image or analyse the features on the server (Solmaz et al. (2018). Therefore, the system is real-time and more secure (Appendix B.4).

For the other R-criteria, i.e. robustness, we have also included different orientation of the sample to vary the camera to sample position. The randomness of the light source '5' in Fig. 6. 2 created variation in the illumination condition.

Based on the elapsed time, our work showed similar rapidness as Mutlu[t]. The combined use of three different classifiers along with a larger feature-set by Kim[t] resulted in a considerably larger model size and higher computational time.

6.8 Summary

This chapter presented an intelligent system for paper-based lateral flow assays. Due to technical and economic feasibility, we have utilised universal pH indicator papers, possessing multi-objects/sample to demonstrate a proof-of-concept. Due to the intelligent histogram-based image processing technique, the system is user-friendly. Unlike the literature (Table 6. 2), the system does not require any user intervention. The automatic image processing technique provided 98.94% accuracy to separate the ROI. The developed system can automatically process the image of the assays without any additional hardware component. The system is equipment-free, does not have any operating cost and accessible and can be deployed to mobile devices.

The system was trained under the semi-controlled ambient condition on a balanced dataset using cross-validation. The performance was validated on completely unseen data. The system showed high accuracy, specificity and sensitivity for colour classification without compromising the degree of freedom.

The system possesses adaptability to work on similar assays without compromising the performance, confirmed by experiments conducted on the assay from a different brand

($\aleph\_pH$ dataset) as well as on the urine dipstick. To the best of our knowledge, there is no such evaluation for robustness performed in the literature.

Among the reported articles, the research conducted by Mutlu et al. (2017) is the most similar research to the case study of this chapter. Therefore, we have deployed Mutlu's method using our dataset to present a fair comparison. Mutlu et al. performed the colourimetric study using the mean RGB value of JPEG, RAW and RAW-corrected image formats in different illumination conditions on the same paper test strip. Mutlu reported the accuracy to be 100%. Re-implementation of Mutlu's work using our dataset revealed our system to be less dependent on user-intervention and to be more reliable, robust and reproducible due to the choice of bigger and more diverse dataset, feature-set and the validation process (Fig. 6. 18-6.19 and Table 6. 6).

The comparative performance analysis was further extended and reported as Table 6. 6. Similar works from the reported articles were re-implemented using the same features and classifiers on our dataset so that a direct comparison can be provided. The detailed study indicated the strength of our proposed scheme which demonstrated better accuracy and more compatibility with the ASSURED criteria.

Due to less iterative image processing, optimised feature-set and selection of the classifier, the computational complexity was optimised. The result can be produced in real-time, conceptually faster than the existing works. The rapidness of the system was quantified in Fig. 6. 21 while re-implementing the reported articles using our dataset.

(This page has been left intentionally blank.)

(This page has been left intentionally blank.)

Chapter 7

Conclusion and Future Work

7.1 Introduction

With an ambition to design and develop a computationally efficient image-based colourimetric system suitable to act as a standalone system on the POC platform, whether integrated to a server or not, this thesis demonstrated a high performing, robust and rapid system which can function without any additional hardware attachment and is suitable to be used at homes as well as within resource-limited settings. The workflow of the finalised system is illustrated in Appendix E (Fig. E. 1). This chapter presents a synopsis of the accomplishments with regard to the research questions using the case studies. The chapter includes the overall performance of the proposed framework while describing the contributions. This chapter also includes the strengths and limitations of this research and potential further research into intelligent colourimetric tests.

7.2 Contributions of the thesis

- **Identification of the Research Gap**

To the best of our knowledge, this is the first attempt to define each component of the colourimetric test using machine vision. In this research, the colourimetric tests were explored from the perspective of computer vision, followed by a critical inspection of the research gap to design and develop an intelligent image-based colourimetric system.

- **Assay Type Detection**

This thesis presented a novel approach to determine the assay type using a pre-trained model, whereas the literary works performed this step using time-consuming calibration with limited capability. The pre-trained model determined the assay type with 100% accuracy. More advantages of the pre-trained model over the calibration approach are robustness and readiness. As mentioned earlier in Chapter 4, recognising the type of assay can assist a novice user to begin the colourimetric test. This step can also enable better management of telemedicine and enhance the experience of telepathology.

- **Colourimetric Classification by tracking the Dynamic Images**

Tracking dynamic changes of an end point assay by video-frame analysis to produce a colourimetric decision is a novel approach. The presented method not only provided a correct classification for all 31 samples but also indicated the extended capability of computer vision to differentiate colours beyond the human visionary system. However, pursuing dynamic changes of the end point assay is computationally expensive and in contradiction of ASSURED criteria. Therefore, the approach should be reserved for minimising the reaction time and analysing the kinetics, instead of a colourimetric classification.

- **Pseudo Control Colours**

This thesis proposed an exclusive feature-set, i.e. pseudo-control colours to be part of the feature-set. Exploring 440 features for the LFA dataset, the optimised feature-set was found to be mean, energy and pseudo-control colours.

- **ASSURED Image-based Colourimetric Test**

This section briefly summarise the contribution towards the computational components to achieve ASSURED image-based colourimetric test using two case studies.

Case Study 1: TB-test

To the best of our knowledge, this is the first attempt to use this case study to provide an intelligent colourimetric solution. The presented hybrid image processing framework demonstrated 98.65-99.95% accuracy (95% CI) to separate the ROI. Using the lower order colour moments in LAB colour space, 98.95-100% classification accuracy (95% CI) was achieved by LS-SVM to determine the presence or absence of TB bacteria in the liquid sample. The supervised machine learning algorithm is capable of processing 18 features of 348 samples within 0.029 seconds. To the best of our knowledge, regarding accuracy and rapidness, this is the best performance achieved by an intelligent system to perform a qualitative colourimetric test on a liquid sample. Due to the efficient design with optimised dimension, the demonstrated system is suitable to be deployed on mobile devices and provide the colourimetric decision using the cloud as well. Therefore, the system is suitable for resource-limited settings.

Case Study 2: pH Test

This thesis investigated, designed and developed an immaculate image processing framework to separate multiple colour pads in the universal pH indicator paper with >98% accuracy, tested and validated by varying illumination, mobile devices, distances (within close proximity and clear exposure), and brand.

This thesis managed to deliver the colourimetric decision for both case studies using the same classifier, i.e. LS-SVM (Suykens and Vandewalle, 1999; Suykens, 2002). For the stated case study, the 10-fold cross validated training and testing for 520 samples was conducted within 0.11seconds with 99.29-100% accuracy (95% CI). The extensive analysis based on the evaluation criteria suggested our system to be more compatible with the ASSURED criteria than existing similar works. Due to the meticulously tested reliability of the system on an adequate amount of appropriate data, our claim is well supported by the precise experimental results.

7.3 Limitations

Despite the effort, any scientific study is not free from certain limitations. In this thesis, one of the major struggles was regarding the dataset. Lack of prior research using a similar approach to perform colourimetric test constrained us to a limited number of reliable data.

The initial assessment included the images of a quantitative test for the case study of drug response. The images were utilised to validate the image processing algorithm. Due to lack of adequate amount of data, this study excluded the regression analysis of the quantitative test, hence this was not presented as a case study.

This thesis considered dynamic changes in the colourimetric reaction. Our initial hypothesis suggested that this method can assist in analysing the chemical kinetics. Due to the lack of validation instruments (gold standard), this thesis excluded the claim.

Although we have considered diverse possible scenarios at indoor settings, the adaptability of the algorithm is hypothetical and validated for a defined condition. There is a scope for further enhancement of the image processing algorithm to induce more robustness and achieve a higher performance. The presented image processing scheme expects the user to insert the number of patients. In other words, the user would

have to define the number of samples present in an image of sample before initiating the TB-test system.

Even though the proposed system is intended to fulfil ASSURED criteria, the biomaterials, biosensing components and protocol is beyond the scope of this thesis. Therefore, the ASSURED criteria are applicable to the intelligent or digital components only.

In order to fulfil the ASSURED criteria, this work was constricted to the limited capacity of the mobile devices. Therefore, the empirical study was less applicable to the architecture of deep learning. Within this context, while deploying an intelligent system on the mobile platform, there is going to be trade-offs among accuracy, execution time and functionality based on the mobile applications targeted for native application suitable for specific devices, device-independent mobile web applications or native applications using HTML interface.

7.4 Future Work

In this section, some suggestions for the future work are given for the continuation of the work presented in this thesis.

- **Feature Optimisation Algorithm**

As compared to the existing research, this work presented more extensive feature exploration considering the distinct dimension of qualitative and semi-quantitative colourimetric tests. Future inclusion of an advanced feature optimisation algorithm on a larger dataset may enhance the efficiency of the system, even within a more robust environment.

- **Developed Model Approach**

With the aid of a larger dataset and a high performing computational system, our future research will include an empirical study using the developed model approach, instead of using pre-trained models using inductive transfer learning to evaluate the possibility of better performance from deep learning to produce a colourimetric decision. On the other hand, a real-time execution of such a system within the existing capacity of mobile devices, while fulfilling the ASSURED criteria, is going to a paramount challenge.

- **Adaptability and Universality**

To the best of our knowledge, this study included a larger dataset and diverse applications compared to the existing research on image-based colourimetric testing. A further extension on samples per class and variability will increase the reliability of the system. Furthermore, inducing more diversity in the assay will demonstrate and evident our claim which can ultimately result in a universal model for image-based colourimetric testing. In such a situation, consideration of the universal applicability can verify the adaptability of the presenting system as well. However, a large-scale problem of the colourimetric test, for example, a universal model, may compel us to reform the classifier, i.e. LS-SVM. Therefore, one may inquire more regarding the scope of ANN.

- **Reinforcement Learning**

Integration of reinforcement learning can update the system based on environmental factors (e.g. geo-location, climate) and change in the pattern of the assay. Associating the colourimetric tests with its context, i.e. symptom descriptions from the patient and health tracking, using cognitive computing can administer a polythetic approach in delivering a confirmatory primary diagnosis and better management of chronic conditions.

- **Commercialisation**

Realising the potential of the presented system, future work will aim for a further application using chromaticity analysis. Analysing the market for such an easy-to-use high performing intelligent system, the prospect of commercialisation would be another possible future direction of this thesis.

References

1. Abuhassan, K.J., Bakhori, N.M., Kusnin, N., Azmi, U.Z.M., Tania, M.H., Evans, B.A., Yusof, N.A. and Hossain, M.A., 2017. Automatic diagnosis of Tuberculosis disease based on plasmonic ELISA and color-based image classification. In: *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Jeju Island, South Korea, pp.4512–4515. 10.1109/EMBC.2017.8037859.

2. Accenture, 2017. *Artificial Intelligence in Healthcare*. [online] Available at: <https://www.accenture.com/us-en/insight-artificial-intelligence-healthcare> [Accessed 07 October 2017].

3. Adobe, 1998. *A Adobe® RGB (1998) Color Image Encoding*. [online] Available at: <http://www.adobe.com> [Accessed 28 July 2018].

4. Agudo Acemel, M., 2017. *Digitization of colorimetric measurements for quantitave analyses using a smartphone*, Masters, Universitat Oberta de Catalunya. Available at: <http://openaccess.uoc.edu/webapps/o2/handle/10609/59186> [Accessed 21 March 2018]

5. Akraa, S., Anh Pham, T.T., Shen, H., Tang, Y., Tang, B.Z., Li, J. and Walker, S., 2018. A smartphone-based point-of-care quantitative urinalysis device for chronic kidney disease patients. *Journal of Network and Computer Applications*. 10.1016/j.jnca.2018.04.012

6. Alankus, G., Horzum, N., Mutlu, A.Y., Bayram, A. and Solmaz, M.E., 2018. Single-image-referenced colorimetric water quality detection using a smartphone. *ACS Omega*, 3(5), pp.5531–5536. 10.1021/acsomega.8b00625.

7. Alidans srl, 2015. *AssayColor*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.alidans.assaycolor> [Accessed 10 January 2017].

8. Android Developers, 2018. *Overview of memory management*. [online]. Available at: <https://developer.android.com/topic/performance/memory-overview> [Accessed 29 May 2018].

9. Anon 2015. *CM-H505 App*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=anieng.com.colorimeter> [Accessed 10 January 2017].

10. Anon 2017. *Color Thesaurus*. [online] Available at: <http://colorthesaurus.epfl.ch/> [Accessed 09 January 2017].

11. App Annie, 2016. *Forecast Report. Forecast Intelligence: Store – Q2 2016*. [online] Available at: <https://www.appannie.com/%5Cnhttps://www.appannie.com/apps/ios/app/ifood-assistant-by-kraft/#>. [Accessed 09 January 2017].

12. Arnett, N., Vergani, A., Winkler, A., Ritter, S. and Mehta, K., 2016. Inexpensive urinalysis test strips to screen for diabetes in developing countries. In: *IEEE Global Humanitarian Technology Conference (GHTC)*. Seattle, pp.589–596. 10.1109/GHTC.2016.7857339.

13. Arthur, D. and Vassilvitskii, S., 2007. k-means++: The advantages of careful seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. New Orleans, pp.1027–1035. Available through: ACM Digital Library website <http://libraries.acm.org/digital-library> [Accessed 12 December 2017].

14. Auralisoft, 2016a. *Calibrated Photo Viewer*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.auralisoft.jpegviewer> [Accessed 10 January 2017].

15. Auralisoft, 2016b. *Calibrated Photo Viewer Demo*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.auralisoft.jpegviewerdemo> [Accessed 10 January 2017].

16. Auralisoft, 2016c. *Camera Colorimeter*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.auralisoft.colorimeter&hl=en_GB> [Accessed 09 January 2017].

17. Aymerich, J., Márquez, A., Terés, L., Muñoz-Berbel, X., Jiménez, C., Domínguez, C., Serra-Graells, F. and Dei, M., 2018. Cost-effective smartphone-based reconfigurable electrochemical instrument for alcohol determination in whole blood samples. *Biosensors and Bioelectronics*, 117, pp. 736-742. https://doi.org/10.1016/j.bios.2018.06.044.

18. Bakhori, N.M., Yusof, N.A., Abdullah, J. and Wasoh, H., 2018. Immuno nanosensor for ultrasensitive and affordable naked eye detection of Tuberculosis. *Sensors*, 18(6), pp.1–10. Available at: https://doi.org/10.3390/s18061932.

19. Balkenius, C., Johansson, A.J. and Balkenius, A., 2003. *Color constancy in visual scene perception*. [online] *LUCS Studies*, Lund, Sweden. Available through: Lund University

Libraries website <http://lup.lub.lu.se/record/776820> [Accessed 25 September 2018].

20. Banoo, S., Bell, D., Bossuyt, P., Herring, A., Mabey, D., Poole, F., Smith, P.G., Sriram, N., Wongsrichanalai, C., Linke, R., O'Brien, R., Perkins, M., Cunningham, J., Matsoso, P., Nathanson, C.M., Olliaro, P., Peeling, R.W. and Ramsay, A., 2006. Evaluation of diagnostic tests for infectious diseases: general principles. *Nature reviews. Microbiology*, 4(9), pp.S20–S32. https://doi.org/10.1038/nrmicro1523.

21. Barbosa, A.I., Gehlot, P., Sidapra, K., Edwards, A.D. and Reis, N.M., 2015. Portable smartphone quantitation of prostate specific antigen (PSA) in a fluoropolymer microfluidic device. *Biosensors and Bioelectronics*, 70, pp.5–14. 10.1016/j.bios.2015.03.006.

22. Barnes, L., Heithoff, D.M., Mahan, S.P., Fox, G.N., Zambrano, A., Choe, J., Fitzgibbons, L.N., Marth, J.D., Fried, J.C., Soh, H.T. and Mahan, M.J., 2018. Smartphone-based pathogen diagnosis in urinary sepsis patients. *EBioMedicine*, 36, pp. 73-82. 10.1016/j.ebiom.2018.09.001.

23. Bellman, R.E., 1961. *Adaptive control processes: A guided tour*. Princeton University Press.

24. Bellman, R.E., 2003. *Dynamic programming*. Republisheed. Courier Dover Publications.

25. Berg, B., Cortazar, B., Tseng, D., Ozkan, H., Feng, S., Wei, Q., Chan, R.Y.-L., Burbano, J., Farooqui, Q., Lewinski, M., Di Carlo, D., Garner, O.B. and Ozcan, A., 2015. Cellphone-based hand-held micro-plate reader for point-of-care testing of enzyme-linked immunosorbent assays. *ACS nano*, 9(8), pp.7857–7866. 10.1021/acsnano.5b03203.

26. Bhattacharjee, T., Jiang, H.R. and Behdad, N., 2015. A fluidic colorimetric sensor design for water hardness detection. *IEEE Sensors Journal*, 15(2), pp.819–826. 10.1109/Jsen.2014.2351813.

27. Bishop, C.M., 2006. *Pattern recognition and machine learning*. Springer.

28. Bourouis, A., Feham, M., Hossain, M. a. and Zhang, L., 2014. An intelligent mobile based decision support system for retinal disease diagnosis. *Decision Support Systems*, 59(1), pp.341–350. 10.1016/j.dss.2014.01.005.

29. Brabanter, J.D., Pelckmans, K., Suykens, J.A.K.. and Vandewalle, J., 2002. Robust cross-validation score function for non-linear function estimation. In: *Dorronsoro J.R. (eds) Artificial Neural Networks — ICANN 2002. ICANN 2002. Lecture Notes in Computer Science*, 2415. Springer, pp.713–719. https://doi.org/10.1007/3-540-46084-5_116.

30. Bu, S., Wang, K., Ju, C., Han, Y., Li, Z., Du, P., Hao, Z., Li, C., Liu, W. and Wan, J., 2018. A pregnancy test strip for detection of pathogenic bacteria by using concanavalin A-human chorionic gonadotropin-$Cu_3(PO_4)_2$ hybrid nanoflowers, magnetic separation, and smartphone readout. *Microchimica Acta*, 85(10), p.464. https://doi.org/10.1007/s00604-018-2968-2.

31. Burden, R.L. and Faires, J.D., 2001. *Numerical analysis*. Brooks/Cole.

32. Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), pp.121–167. https://doi.org/10.1023/A:1009715923555.

33. Calabria, D., Caliceti, C., Zangheri, M., Mirasoli, M., Simoni, P. and Roda, A., 2017. Smartphone–based enzymatic biosensor for oral fluid L-lactate detection in one minute using confined multilayer paper reflectometry. *Biosensors and Bioelectronics*, 94, pp.124–130. https://doi.org/10.1016/j.bios.2017.02.053.

34. Cate, D.M., Adkins, J.A., Mettakoonpitak, J. and Henry, C.S., 2015. Recent developments in paper-based microfluidic devices. *Analytical Chemistry*, 87(1), pp.19–41. 10.1021/ac503968p.

35. Centers for Disease Control and Prevention, 2017. *Tuberculosis (TB) | CDC*. [online]. Available at: <https://www.cdc.gov/tb/> [Accessed 18 September 2017].

36. Chang, C.-C. and Lin, C.-J., 2011. LIBSVM -- A Library for Support Vector Machines. [computer program] Available at: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/> [Accessed 26 September 2018].

37. Chawla, N. V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, pp.321–357. https://doi.org/10.1613/jair.953.

38. Chematics Inc., 2018. *Alco-Screen test strips*. [online] Available at: <https://chematics.com/> [Accessed 18 September 2018].

39. Chen, C., Wu, Y. and Dong, T., 2014. Dipsticks integrated on smart diapers for colorimetric analysis of urinary tract infections in the field. In: *Proceedings of the 16th International Conference on Mechatronics, Mechatronika*. Brno, Czech Republic, pp.423–427. 10.1109/MECHATRONIKA.2014.7018295.

40. Chen, G., Chai, H.H., Yu, L. and Fang, C., 2018a. Smartphone supported backlight illumination and image acquisition for microfluidic-based point-of-care testing. *Biomedical optics express*, 9(10), pp. 4604-4612. 10.1364/BOE.9.004604.

41. Chen, G., Fang, C., Chai, H.H., Zhou, Y., Yun Li, W. and Yu, L., 2018b. Improved analytical performance of smartphone-based colorimetric analysis by using a power-free imaging box. *Sensors and Actuators B: Chemical*. https://doi.org/10.1016/j.snb.2018.09.019.

42. Chen, W., Fang, X., Li, H., Cao, H. and Kong, J., 2016. A simple paper-based colorimetric device for rapid mercury(II) assay. *Scientific Reports*, 6(August), p.1-6. https://doi.org/10.1038/srep31948.

43. Cheng, W.-C. and Badano, A., 2018. Display for medical imaging and DICOM grayscale standard display function fundamentals. In: P. Russo, ed., *Handbook of X-ray imaging : Physics and technology*, New York, pp.1283–1296.

44. Chitade, A.Z. and Katiyar, D.S.K., 2010. Colour based image segmentation using k-Means clustering. *International Journal of Engineering Science and Technology*, 2(10), pp.5319–5325.

45. Choodum, A., Sriprom, W. and Wongniramaikul, W., 2019. Portable and selective colorimetric film and digital image colorimetry for detection of iron. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 208, pp.40–47. https://doi.org/10.1016/j.saa.2018.09.062.

46. Chromaflo Technologies, 2015. *Innovatint Tablet version*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=air.com.chromaflo.innovatint.tablet&hl=en_G B> [Accessed 10 January 2017].

47. Clopper, C.J. and Pearson, E.S., 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4), pp.404–413. https://doi.org/10.1093/biomet/26.4.404.

48. Colorix, 2016. *ProfiTec Colordesign*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.colorix.profiteccolordesign> [Accessed 10 January 2017].

49. Computer Graphics Technology P.L., 2015. *ArgyllPRO ColorMeterDemo*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.argyllpro.colormeterdemo&hl=en_GB> [Accessed 09 January 2017].

50. Contreras-naranjo, J.C., Wei, Q. and Ozcan, A., 2016. Mobile phone-based microscopy, sensing , and diagnostics. *IEEE Journal of Selected Topics in Quantum Electronics*, 22(3), pp.1–14. 10.1109/JSTQE.2015.2478657.

51. Cooper, D., 2012. Mobile image ratiometry for the detection of Botrytis cinerea (Gray Mold). In: *Nature Precedings*. http://dx.doi.org/10.1038/npre.2012.6989.1.

52. Cooper, D., Callahan, B., Callahan, P. and Burnett, L., 2012. Mobile image ratiometry: a new method for instantaneous analysis of rapid test strips. In: *Nature Precedings*. pp.2–3. http://dx.doi.org/10.1038/npre.2012.6827.1.

53. Coskun, Ahmet F.Wong, J., Khodadadi, D., Nagi, R., Teya, A. and Ozcan, A., 2012. A personalized food allergen testing platform on a cellphone. *Lab on a Chip*, 13(4), pp.636–640. 10.1039/c2lc41152k.

54. Coskun, A.F., Nagi, R., Sadeghi, K., Phillips, S. and Ozcan, A., 2013. Albumin testing in urine using a smart-phone. *Lab on a Chip*, 13(21), p. 4231-4238. 10.1039/C3LC50785H.

55. Datacolor, 2014. *Spyder4Gallery*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.datacolor.caprica> [Accessed 10 January 2017].

56. de la Rica, R. and Stevens, M.M., 2012. Plasmonic ELISA for the ultrasensitive detection of disease biomarkers with the naked eye. *Nature Nanotechnology*, 8(9), pp.1759–1764. https://doi.org/10.1038/nprot.2013.085.

57. Deng, L. and Yu, D., 2014. *Deep learning: Methods and applications*. [pdf] Available at: <https://www.microsoft.com/en-us/research/publication/deep-learning-methods-and-applications/?from=http%3A%2F%2Fresearch.microsoft.com%2Fpubs%2F209355%2Fdeepl earning-nowpublishing-vol7-sig-039.pdf> [Accessed 28 July 2018].

58. Deng, Y. and Manjunath, B.S., 2001. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8), pp.800–810. 10.1109/34.946985

59. Department of Economic and Social Affairs, 2016. *International Migration Report 2015*. [pdf] Available at: <http://www.un.org/en/development/desa/population/migration/publications/migrationreport/d ocs/MigrationReport2015.pdf> [Accessed 24 September 2017].

60. Devadhasan, J.P. and Kim, J., 2018. A chemically functionalized paper-based microfluidic

platform for multiplex heavy metal detection. *Sensors and Actuators B: Chemical*, 273, pp.18–24. https://doi.org/10.1016/j.snb.2018.06.005.

61. Dhar, A., Mehta, R. and Sit, M., 2017. *Colorimetric detection of pH strips*. [GitHub>mattsit> ph-machine-learning]. [online] Available through: GitHub, Inc. <https://github.com>.

62. Dinnes, J., Deeks, J., Kunst, H., Gibson, A., Cummins, E., Waugh, N., Drobniewski, F. and Lalvani, A., 2007. A systematic review of rapid diagnostic tests for the detection of tuberculosis infection. *Health Technology Assessment (Winchester, England)*, 11(3), pp.1–196.

63. DoITPoMS, 2010. What is a Tensor? *Tensors in Materials Science*. [online via TLP Library], University of Cambridge. Available at: <https://www.doitpoms.ac.uk/tlplib/tensors> [Accessed 14 September 2018].

64. Enzo Life Sciences Inc., 2015. *Enzo ELISA Plate Reader*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.enzo.elisaplatereader> [Accessed 21 September 2017].

65. Estabrooks, A., Jo, T. and Japkowicz, N., 2004. A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence*, [online] 20(1), pp.18–36. Available through: Wiley Online Library < https://onlinelibrary.wiley.com> [Accessed 05 September 2018].

66. Facchini, F.A., Zaffaroni, L., Minotti, A., Rapisarda, S., Calabrese, V., Forcella, M., Fusi, P., Airoldi, C., Ciaramelli, C., Billod, J.-M., Schromm, A.B., Braun, H., Palmer, C., Beyaert, R., Lapenta, F., Jerala, R., Pirianov, G., Martin-Santamaria, S. and Peri, F., 2018. Structure–activity relationship in Monosaccharide-based Toll-Like Receptor 4 (TLR4) antagonists. *Journal of Medicinal Chemistry*, 61(7), pp.2895–2909. 10.1021/acs.jmedchem.7b01803.

67. Falcon Solutions Co, 2016. *Color Meter*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.falcon.colormeter&hl=en_GB> [Accessed 10 January 2017].

68. Feng, S., Caire, R., Cortazar, B., Turan, M., Wong, A. and Ozcan, A., 2014. *Immunochromatographic diagnostic test analysis using google glass*. *ACS Nano*, 8(3), pp. 3069-3079. 10.1021/nn500614k.

69. Fernandes, L.A.F. and Oliveira, M.M., 2008. Real-time line detection through an improved Hough transform voting scheme. *Pattern Recognition*, 41(1), pp.299–314. https://doi.org/10.1016/j.patcog.2007.04.003

70. Fheldt, 2014. *colorStriker*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.colorStriker&hl=en_GB> [Accessed 09 January 2017].

71. Field, A.P., 2013. *Discovering statistics using IBM SPSS statistics : And sex and drugs and rock 'n' roll*. 4th ed. London: SAGE.

72. Flavio Gonzalez App-Entwicklung, 2016. *Camera FV-5: Professional camera application for Android*. [Google Play] Available at: <https://www.camerafv5.com/> [Accessed 20 September 2018].

73. Folkstedt, 2013. *Subtractive RYB Color Meter*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=cassiopeia.colormeter> [Accessed 10 January 2017].

74. FTLapps, 2012. *ColorAssist*. [iTunes] Available at: <https://itunes.apple.com/us/app/colorassist/id496585813?mt=8> [Accessed 16 January 2016].

75. Garg, S., Ramprasaath, R.S., Kapur, S. and Rao, K.M.M., 2014. Automated colorimetric analysis in paper based sensors. In: *IEEE International Conference on Image Processing (ICIP)*. Paris, pp.3607–3611. 10.1109/ICIP.2014.7025732.

76. Giavazzi, F., Salina, M., Ceccarello, E., Ilacqua, A., Damin, F., Sola, L., Chiari, M., Chini, B., Cerbino, R., Bellini, T. and Buscaglia, M., 2014. A fast and simple label-free immunoassay based on a smartphone. *Biosensors and Bioelectronics*, 58, pp.395–402. https://doi.org/10.1016/j.bios.2014.02.077.

77. Glass, 2017. *Glass*. [online] Available at: <https://x.company/glass/> [Accessed 07 August 2017].

78. Gonzalez, R.C. and Woods, R.E., 2018. *Digital Image Processing*. 4th ed. New York: Pearson.

79. Gonzalez, R.C., Woods, R.E. and Eddins, S.L., 2009. *Digital image processing using MATLAB®*. 2nd ed. Gatesmark Publishing.

80. Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep Learning*. London: MIT Press.

81. Google Brain Team, 2018. *TensorFlow*. [online] Available at: <https://www.tensorflow.org> [Accessed 18 September 2018].

82. GSMA Intelligence, 2017. *Definitive data and analysis for the mobile industry*. [GSMA

Intelligence > Global Data]. [online] Available through: GSMA Intelligence <https://www.gsmaintelligence.com/> [Accessed 07 August 2017].

83. Guo, L., Xu, S., Ma, X., Qiu, B., Lin, Z. and Chen, G., 2016. Dual-color plasmonic enzyme-linked immunosorbent assay based on enzyme-mediated etching of Au nanoparticles. *Scientific Reports*, 6(August). https://doi.org/10.1038/srep32755.

84. Hadjis, S., Abuzaid, F., Zhang, C. and Ré, C., 2015. Caffe con Troll: Shallow ideas to speed up deep learning. In: *Proceedings of the Fourth Workshop on Data analytics at sCale (DanaC 2015), Workshop on Data Analytics in the Cloud*. Melbourne. Available through: NIH Public Access < https://www.ncbi.nlm.nih.gov/>. [Accessed 11 October 2018].

85. Hagan, M.T. and Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6), pp.989–993.

86. Hatem Al-jamal, D., 2017. *Financial Instruments under the IFRSs: A comparative study between the aarly adoption of IFRS 9 and its precedent standards set in commercial banks in the Middle East*. Ph.D. Anglia Ruskin University. Available at: <https://core.ac.uk/download/pdf/153570146.pdf> [Accessed 23 August 2018].

87. He, H. and Garcia, E.A., 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp.1263–1284. 10.1109/TKDE.2008.239.

88. He, K., Zhang, X., Ren, S. and Sun, J., 2015. *Deep residual learning for image recognition*. [online] Cornell University Library. Available at: <https://arxiv.org/pdf/1512.03385.pdf> [Accessed 10 June 2018].

89. Healthy.io, 2018. *Clinically approved digital urinalysis*. [online] Available at: <https://healthy.io/product/> [Accessed 29 September 2018].

90. Hernández-Neuta, I., Neumann, F., Brightmeyer, J., Tis, T.B., Madaboosi, N., Wei, Q., Ozcan, A. and Nilsson, M., 2018. Smartphone-based clinical diagnostics: Towards democratization of evidence-based health care. *Journal of Internal Medicine*. [online] Available at: <http://doi.wiley.com/10.1111/joim.12820> [Accessed 11 August 2018].

91. Hussain, I., Das, M., Ahamad, K.U. and Nath, P., 2017. Water salinity detection using a smartphone. *Sensors and Actuators, B: Chemical*, [online] 239, pp.1042–1050. https://doi.org/10.1016/j.snb.2016.08.102.

92. ICC, 2004. *White paper #5 on occupational regulation.* [pdf] International Color Consortium. Available at: <http://www.color.org/ICC_white_paper5glossary.pdf>. [Accessed 21 September 2017].

93. ImageNet, 2016. *ImageNet*. [online] Available at: <http://image-net.org > [Accessed 28 July 2018].

94. Interactive Health Solutions, 2016a. *Global Fund TB*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.ihsinformatics.tbr3mobile_sa&hl=en> [Accessed 21 September 2017].

95. Interactive Health Solutions, 2016b. *MINE TB*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.ihsinformatics.tbr4mobile> [Accessed 21 September 2017].

96. Interactive Health Solutions, 2016c. *TB REACH 4 - Kotri*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.ihsinformatics.tbr4mobile_pk&hl=en> [Accessed 21 September 2017].

97. Interactive Health Solutions, 2017. *Childhood TB-Bangladesh*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.ihsinformatics.childhoodtb_mobile&hl=en> [Accessed 21 September 2017].

98. International Organization for Standardization, 2007. *ISO 11664-4:2008 (CIE S 014-4/E:2007) Colorimetry -- Part 4: CIE 1976 L*a*b* Colour space*. Vienna: CIE International Commission on Illumination, International Organization for Standardization.

99. Jonas, S.M., Deserno, T.M., Buhimschi, C.S., Makin, J., Choma, M.A. and Buhimschi, I.A., 2016. Smartphone-based diagnostic for preeclampsia: An mHealth solution for administering the Congo Red Dot (CRD) test in settings with limited resources. *Journal of the American Medical Informatics Association*, 23(1), pp.166–173. https://doi.org/10.1093/jamia/ocv015.

100. Kandilli, C. and Ulgen, K., 2008. Solar illumination and estimating daylight availability of global solar irradiance. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 30(12), pp.1127–1140.

101. Karlsen, H., 2018. *Smartphone-based urinary biomarker detection: an application-oriented device and algorithm*. Ph.D. University College of Southeast Norway. Available at: <https://brage.bibsys.no/xmlui/handle/11250/2480144> [Accessed 21 March 2018].

102. Karlsen, H. and Dong, T., 2017. Smartphone-based rapid screening of urinary biomarkers. *IEEE*

Transactions on Biomedical Circuits and Systems, 11(2), pp.455–463. 10.1109/TBCAS.2016.2633508.

103. Karpathy, A., 2018. *CS231n Convolutional neural networks for visual recognition*. [online via internet VLE]. Stanford University Available at: <http://cs231n.github.io/transfer-learning/> [Accessed 14 September 2018].

104. Kettler, H., White, K. and Hawkes, S., 2004. *Mapping the landscape of diagnostics for sexually transmitted infections: Key Findings and Recommendations*. [pdf] Special Programme for Research and Training in Tropical Diseases (TDR), UNICEF/ UNDP/ World Bank/ WHO. Available at: <http://www.who.int/tdr/publications/documents/mapping-landscape-sti.pdf?ua=1> [Accessed 23 April 2018].

105. Khademhosseini, A., 2011. Nano/microfluidics for diagnosis of infectious diseases in developing countries. *Advanced Drug Delivery Reviews*, 62(4–5), pp.449–457. https://doi.org/10.1016/j.addr.2009.11.016.

106. Khan, M.S. and Garnier, G., 2013. Direct measurement of alkaline phosphatase kinetics on bioactive paper. *Chemical Engineering Science*, 87, pp.91–99. https://doi.org/10.1016/j.ces.2012.09.022.

107. Kim, H., Awofeso, O., Choi, S., Jung, Y. and Bae, E., 2017a. Colorimetric analysis of saliva--alcohol test strips by smartphone-based instruments using machine-learning algorithms. *Applied Optics*, 56(1), pp.84–92. https://doi.org/10.1364/AO.56.000084.

108. Kim, H., Jung, Y., Doh, I.J., Lozano-Mahecha, R.A., Applegate, B. and Bae, E., 2017b. Smartphone-based low light detection for bioluminescence application. *Scientific Reports*, 7(December 2016), pp.1–11. https://doi.org/10.1038/srep40203.

109. Kim, S.C., Jalal, U.M., Im, S.B., Ko, S. and Shim, J.S., 2017c. A smartphone-based optical platform for colorimetric analysis of microfluidic device. *Sensors and Actuators, B: Chemical*, 239, pp.52–59. https://doi.org/10.1016/j.snb.2016.07.159.

110. Kiratiratanapruk, K., 2016. Silkworm egg image analysis using different color information for improving quality inspection. In: *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. Phuket. 10.1109/ISPACS.2016.7824731.

111. Koczula, K.M. and Gallotta, A., 2016. Lateral flow assays. *Essays in biochemistry*, 60(1), pp.111–20. 10.1042/EBC20150012.

112. Konnaiyan, K., Cheemalapati, S., Pyayt, A. and Gubanov, M., 2017. mHealth dipstick analyzer for monitoring of pregnancy complications. *IEEE Sensors Journal*, 17(22), pp.7311–7316. 10.1109/JSEN.2017.2752722.

113. Koo, K.M., Wee, E.J.H. and Trau, M., 2016. Colorimetric *TMPRSS2-ERG* gene fusion detection in prostate cancer urinary samples via recombinase polymerase amplification. *Theranostics*, 6(9), pp.1415–1424. 10.7150/thno.15250.

114. Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS)*. Lake Tahoe, Nevada. 1, pp.1097–1105.

115. Kuehni, R., 2010. *Color spaces*. [online] Scholarpedia, 5(3):9606. Available through: <http://www.scholarpedia.org/article> [Accessed 05 October 2018].

116. Lab4U, 2016. *Colorimeter by Lab4U*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=cl.lab4u.chemistrydemo> [Accessed 10 Januray 2017].

117. Lee, S., Oncescu, V., Mancuso, M., Mehta, S. and Erickson, D., 2014. A smartphone platform for the quantification of vitamin D levels. *Lab on a chip*, 14(8), pp.1437–42. 10.1039/c3lc51375k.

118. Leizer, R., 2016. *Color Analysis*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.leizersoft.coloranalysis&hl=en_GB> [Accessed 09 January 2017].

119. Levene, H., 1960. Robust tests for equality of variance. In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford, California: Stanford University Press, pp.278–292.

120. Li, F., Li, H., Wang, Z., Wu, J., Wang, W., Zhou, L., Xiao, Q. and Pu, Q., 2018. Mobile phone mediated point-of-care testing of HIV p24 antigen through plastic micro-pit array chips. *Sensors and Actuators B: Chemical*, 271, pp. 189-194. https://doi.org/10.1016/j.snb.2018.05.090.

121. Lin, G.G. and Scott, J.G., 2012a. Imaging without lenses: Achievements and remaining challenges of wide-field on-chip microscopy. *Nature Methods*, 9(9), pp.889–895. https://doi.org/10.1038/nmeth.2114.

122. Lin, G.G. and Scott, J.G., 2012b. Integration of cell phone imaging with microchip ELISA to detect ovarian cancer HE4 biomarker in urine at the point-of-care. *Lab on a Chip*, 100(2), pp.130–134. 10.1039/c1lc20479c.

123. Liu, X., Lin, T.-Y. and Lillehoj, P.B., 2014. Smartphones for Cell and Biomolecular Detection. *Annals of Biomedical Engineering*, 42(11), pp.2205–2217. https://doi.org/10.1007/s10439-014-1055-z.

124. Lopez-Ruiz, N., Curto, V.F., Erenas, M.M., Benito-Lopez, F., Diamond, D., Palma, A.J. and Capitan-Vallvey, L.F., 2014. Smartphone-based simultaneous pH and nitrite colorimetric determination for paper microfluidic devices. *Analytical Chemistry*, 86(19), pp.9554–9562. 10.1021/ac5019205.

125. Ludwig, S.K.J., Tokarski, C., Lang, S.N., van Ginkel, L.A., Zhu, H., Ozcan, A. and Nielen, M.W.F., 2015. Calling biomarkers in milk using a protein microarray on your smartphone. *Plos One*, [online] 10(8), p.e0134360. Available at: <http://dx.plos.org/10.1371/journal.pone.0134360> [Accessed 26 August 2018].

126. Lumley, T., Diehr, P., Emerson, S. and Chen, L., 2002. The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23(1), pp.151–169. https://doi.org/10.1146/annurev.publhealth.23.100901.140546.

127. Lwin, K.T., 2015. *Evolutionary approaches for portfolio optimization*. Ph.D. Univeristy of Nottingham Available at: <http://eprints.nottingham.ac.uk/29142/1/ktl-ethesis-corrected.pdf> [Accessed 22 August 2018].

128. Magiati, M., Myridaki, V.M., Christopoulos, T.K. and Kalogianni, D.P., 2019. Lateral flow test for meat authentication with visual detection. *Food Chemistry*, 274, pp.803–807. 10.1016/J.FOODCHEM.2018.09.063.

129. Mahy, M., Van Eycken, L. and Oosterlinck, A., 1994. Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV. *Color Research & Application*, 19(2), pp.105–121. https://doi.org/10.1111/j.1520-6378.1994.tb00070.x.

130. Manuel A. Palacios, Ryuhei Nishiyabu, Manuel Marquez, and and Pavel Anzenbacher, J.., 2007. Supramolecular chemistry approach to the design of a high-resolution sensor array for Multianion detection in water. 10.1021/ja0704784.

131. Margulis, D. and Dan, 2005. *Photoshop LAB color : The canyon conundrum and other adventures in the most powerful colorspace*. Berkeley, California: Peachpit Press.

132. Markets and Markets, 2015. *mHealth Solutions Market By Connected Devices & Services – 2020*. [online] Report Code: HIT 2104. Available at: <http://www.marketsandmarkets.com/Market-Reports/mhealth-apps-and-solutions-market-1232.html> [Accessed 01 March 2017].

133. Markets and Markets, 2016. *Immunoassay Market by Technology, Application & End User – 2021*. [online] Report Code: MD 4751. Available at: <http://www.marketsandmarkets.com/Market-Reports/immunoassay-market-436.html> [Accessed 01 March 2017].

134. Masawat, P., Harfield, A. and Namwong, A., 2015. An iPhone-based digital image colorimeter for detecting tetracycline in milk. *Food Chemistry*, 184, pp.23–29. https://doi.org/10.1016/j.foodchem.2015.03.089.

135. MathWorks, 2017. MATLAB. [computer program] The MathWorks, Inc. Available at: <https://uk.mathworks.com/products/matlab.html> [Accessed 18 September 2018].

136. MATLAB, 2013. Train support vector machine classifier. [computer program] The MathWorks, Inc. Available at: <https://uk.mathworks.com/help/stats/svmtrain.html> [Accessed 26 September 2018].

137. MATLAB & Simulink, 2018a. Deep Dream images using AlexNet. [computer program] The MathWorks, Inc. Available at: <https://uk.mathworks.com/help/nnet/examples/deep-dream-images-using-alexnet.html> [Accessed 09 September 2018].

138. MATLAB & Simulink, 2018b. MATLAB Mobile overview. [computer program] The MathWorks, Inc. Available at: <https://uk.mathworks.com/products/matlab-mobile.html> [Accessed 09 September 2018].

139. MATLAB & Simulink, 2018c. *Preprocess images for deep learning*. [online] MathWorks. Available at: <https://uk.mathworks.com/help/deeplearning/ug/preprocess-images-for-deep-learning.html;jsessionid=e43272072335d60313e2016301a5> [Accessed 15 September 2018].

140. MATLAB & Simulink, 2013. Sharpen image using unsharp masking. [computer program] The MathWorks, Inc. Available at: <https://uk.mathworks.com/help/images/ref/imsharpen.html#btuqhrj-4> [Accessed 16 September 2018].

141. Matloff, N., 2017. *Statistical regression and classification : From linear models to machine learning*. Boka Raton, Florida: CRC Press, Taylor & Francis Group.

142. Matthews, J., Kulkarni, R., Gerla, M. and Massey, T., 2012. Rapid dengue and outbreak detection with mobile systems and social networks. *Mobile Networks and Applications*, 17(2), pp.178–191. 10.1007/s11036-011-0295-5.

143. Mavandadi, S., Dimitrov, S., Feng, S., Yu, F., Sikora, U., Yaglidere, O., Padmanabhan, S., Nielsen, K. and Ozcan, A., 2012. Distributed medical image analysis and diagnosis through crowd-sourced games: A malaria case study. *PLoS ONE*, [online] 7(5). Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0037245> [Accessed 01 March 2017]

144. Mazzone, P.J., Hammel, J., Dweik, R., Na, J., Czich, C., Laskowski, D. and Mekhail, T., 2007. Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array. *Thorax*, 62(7), pp.565–8. 10.1136/thx.2006.072892.

145. Mazzone, P.J., Wang, X.F., Xu, Y., Mekhail, T., Beukemann, M.C., Na, J., Kemling, J.W., Suslick, K.S. and Sasidhar, M., 2013. Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. *Journal of Thoracic Oncology*, 7(1), pp.137–142. 10.1097/JTO.0b013e318233d80f.

146. Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. eds., 1983. *Machine learning*. Berlin, Heidelberg: Springer.

147. Minogue, T.D., Rachwal, P.A., Hall, A.T., Koehler, J.W. and Weller, S.A., 2014. Cross-institute evaluations of inhibitor-resistant PCR reagents for direct testing of aerosol and blood samples containing biological warfare agent DNA. *Applied and Environmental Microbiology*, 80(4), pp.1322–1329. 10.1128/AEM.03478-13

148. Mordvintsev, Alexander; Olah, Christopher; Tyka, M., 2015. Inceptionism: Going deeper into neural networks. *Google,* [blog] 17 June. Available at: <https://web.archive.org/web/20150703064823/http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html> [Accessed 09 September 2018].

149. Mordvintsev, Alexander; Olah, C. and Tyka, M., 2015. DeepDream. *Google,* [blog] 1 June. Available at: <https://web.archive.org/web/20150708233542/http://googleresearch.blogspot.co.uk/2015/07/deepdream-code-example-for-visualizing.html> [Accessed 09 September 2018].

150. Murphy, K.P., 2012. *Machine learning : A probabilistic perspective*. [e-book] MIT Press. Available through: <https://mitpress.mit.edu/books> [Accessed 29 July 2018].

151. Mutlu, A.Y., Kılıç, V., Özdemir, G.K., Bayram, A., Horzum, N. and Solmaz, M.E., 2017. Smartphone-based colorimetric detection via machine learning. *The Analyst*, 142(13), pp.2434–2441. 10.1039/C7AN00741H.

152. Mystrica, 2016. *Mystrica Colorimeter*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.xdesign.android.mystrica&hl=en_GB> [Accessed 09 January 2017].

153. Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: *27th International Conference on International Conference on Machine Learning*. Haifa, Israel: Association for Computing Machinery, Haifa, Israel. pp.807–814.

154. National Eye Institute, 2015. *Facts about color blindness*. [online]. Available at: <https://nei.nih.gov/health/color_blindness/facts_about> [Accessed 03 October 2018].

155. N.A. Yusof, 2018. *TB nanosense: Plasmonic ELISA part 2.* [video online] Available at: <https://www.youtube.com/watch?v=gzy_oL9mvRw> [Accessed 10 August 2018].

156. Ng, A., 2018. Deep learning specialization. [online visa VLE], Coursera. Available at: <https://www.deeplearning.ai/> [Accessed 15 September 2018].

157. NHS, 2017. *Tuberculosis (TB) - NHS Choices*. [online] 2016. Available at: <http://www.nhs.uk/Conditions/Tuberculosis/Pages/Introduction.aspx> [Accessed 18 September 2017].

158. Nicksdroid, 2016. *Color Meter*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.nicksdroid.colormeter&hl=en_GB> [Accessed 09 January 2017].

159. Noble, W.S., 2006. What is a support vector machine? *Nature Biotechnology*, 24(12), pp.1565–1567. 10.1038/nbt1206-1565.

160. O'Connor, E.F., Paterson, S. and de la Rica, R., 2016. Naked-eye detection as a universal approach to lower the limit of detection of enzyme-linked immunoassays. *Analytical and Bioanalytical Chemistry*, 408(13), pp.3389–3393. https://doi.org/10.1007/s00216-016-9453-8.

161. Okeke, I.N., Peeling, R.W., Goossens, H., Auckenthaler, R., Olmsted, S.S., De Lavison, J.F.,

Zimmer, B.L., Perkins, M.D. and Nordqvist, K., 2011. Diagnostics as essential tools for containing antibacterial resistance. *Drug Resistance Updates*, 14(2), pp.95–106. https://doi.org/10.1016/j.drup.2011.02.002.

162. Oliveira, R.B., Pereira, A.S. and Tavares, J.M.R.S., 2017. Skin lesion computational diagnosis of dermoscopic images: Ensemble models based on input feature manipulation. *Computer Methods and Programs in Biomedicine*, 149, pp.43–53. https://doi.org/10.1016/j.cmpb.2017.07.009.

163. Open Medicine Project, 2014. *FIND TB*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=tompsa.findtb&hl=en> [Accessed 21 September 2017].

164. Operation Asha, 2017. *eAlert Cambodia*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=org.opasha.eCompliance.ecomplianceLabCambodia> [Accessed 21 September 2017].

165. Osman, M.K., Mashor, M.Y. and Jaafar, H., 2010. Detection of mycobacterium tuberculosis in Ziehl-Neelsen stained tissue images using Zernike moments and hybrid multilayered perceptron network. In: *IEEE International Conference on Systems, Man and Cybernetics*. Istanbul, pp.4049–4055. 10.1109/ICSMC.2010.5642191.

166. Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp.62–66.

167. Ozkan, H., 2017. Rapid diagnostic lateral flow strip test reader. *Balkan Journal Of Electrical & Computer Engineering*, 5(2), pp.34–39. 10.17694/bajece.334337.

168. Ozkan, H. and Kayhan, O.S., 2016. A novel automatic rapid diagnostic test reader platform. *Computational and Mathematical Methods in Medicine*, [online] Available at: <http://dx.doi.org/10.1155/2016/7498217> [Accessed 09 January 2017].

169. P. D. of the Department of Economic and S. A. of the United Nations Secretariat, 2012. *World population prospects: The 2012 revision*. [pdf] UN. Available at: <http://www.un.org/en/development/desa/population/publications/pdf/trends/WPP2012_Wallchart.pdf> [Accessed 27 June 2017].

170. P.L., C.G.T., 2015. *ArgyllPRO ColorMeter*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.argyllpro.colormeter> [Accessed 09 January 2017].

171. Pal, S. and Gulli, A., 2017. *Deep learning with Keras*. Birmingham: Packt Publishing.

172. Peeling, R.W., Holmes, K.K., Mabey, D. and Ronald, A., 2006. Rapid tests for sexually transmitted infections (STIs): The way forward. *Sexually Transmitted Infections*, 82 (suppl 5), pp.v1-6. http://doi.org/10.1136/sti.2006.024265.

173. Phothisonothai, M., Tantisatirapong, S. and Aurasopon, A., 2016. Automated aetermination of watermelon ripeness based on image color segmentation and rind texture analysis. In: *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. Phuket. 10.1109/ISPACS.2016.7824766.

174. Pollock, N.R., Pollock, N.R., Rolland, J.P., Kumar, S., Beattie, P.D. and Jain, S., 2012. A paper-based multiplexed transaminase test for low-cost, point-of-care liver function testing. *Science Translational Medicine*, [online] 4, 152: 152ra129 Available through: Digital Access to Scholarship at Harvard <http://nrs.harvard.edu> [Accessed 20 September 2017].

175. Posey, D.L., Marano, N. and Cetron, M.S., 2017. Cross-border solutions needed to address tuberculosis in migrating populations. *The International Journal of Tuberculosis and Lung Disease*, 21(5), pp.485–486. https://doi.org/10.5588/ijtld.17.0187.

176. Promphet, N., Rattanawaleedirojn, P., Siralertmukul, K., Soatthiyanon, N., Potiyaraj, P., Thanawattano, C., Hinestroza, J.P. and Rodthongkum, N., 2019. Non-invasive textile based colorimetric sensor for the simultaneous detection of sweat pH and lactate. *Talanta*, 192, pp.424–430. https://doi.org/10.1016/j.talanta.2018.09.086.

177. Qin, X., Wang, R., Tsow, F., Forzani, E., Xian, X. and Tao, N., 2015. A colorimetric chemical sensing platform for real-time monitoring of indoor formaldehyde. *IEEE Sensors Journal*, 15(3), pp.1545–1551. 10.1109/JSEN.2014.2364142.

178. QuArt Studio, 2018. *Aperture Meter*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.quartstudio.aperture.meter&hl=en_GB> [Accessed 23 April 2018].

179. Rahmat, R.F., Royananda, Muchtar, M.A., Taqiuddin, R., Adnan, S., Anugrahwaty, R. and Budiarto, R., 2018. Automated color classification of urine dipstick image in urine examination. *Journal of Physics: Conference Series*, [online] 978(1), p.012008. Available at: <http://stacks.iop.org/1742-

6596/978/i=1/a=012008?key=crossref.88f0a381b969099cf5bf5e6b9c95e157> [Accessed 20 March 2018].

180. Rajan, A. and Glorikian, H., 2009. Point-of-care diagnostics: Market trends and growth drivers. *Expert Opinion on Medical Diagnostics*, 3(1), pp.1–4. https://doi.org/10.1517/17530050802651579.

181. Rasband, W., 2016. ImageJ. [computer program] National Institutes of Health. Available at: <https://imagej.nih.gov/ij/> [Accessed 18 September 2018].

182. Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian processes for machine learning. In: T. Dietterich, ed., *Adaptive Computation and Machine Learning*. Cambridge, Massachusetts: MIT Press.

183. Reed, S.E., Lee, H., Anguelov, D., Szegedy, C., Erhan, D. and Rabinovich, A., 2015. Training deep neural networks on noisy labels with bootstrapping. In: *International Conference on Learning Representations*. San Diego.

184. Ren and Malik, 2003. Learning a classification model for segmentation. In: *9th IEEE International Conference on Computer Vision*. Nice, pp.10–17 vol.1. 10.1109/ICCV.2003.1238308.

185. Research and Market and BCC Research, 2016. *Point of Care Diagnostics*. [online] Available at: <http://www.researchandmarkets.com/research/plmq7v/point_of_care> [Accessed 13 March 2017].

186. Roda, A., Michelini, E., Zangheri, M., Di Fusco, M., Calabria, D. and Simoni, P., 2016. Smartphone-based biosensors: A critical review and perspectives. *TrAC - Trends in Analytical Chemistry*, 79, pp.317–325. https://doi.org/10.1016/j.trac.2015.10.019.

187. Rosenblatt, F., 1961. *Principles of neurodynamics; perceptrons and the theory of brain mechanisms*. Washington DC: Spartan Books.

188. Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, pp.53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

189. Santos Neto, J.H., Porto, I.S.A., Schneider, M.P., dos Santos, A.M.P., Gomes, A.A. and Ferreira, S.L.C., 2018. Speciation analysis based on digital image colorimetry: Iron (II/III) in white wine. *Talanta*, 194, pp. 86-89. https://doi.org/10.1016/j.talanta.2018.09.102.

190. Satterthwaite, F.E., 1946. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6), pp.110–114. 10.2307/3002019.

191. Savardi, M., Ferrari, A. and Signoroni, A., 2018. Automatic hemolysis identification on aligned dual-lighting images of cultured blood agar plates. *Computer Methods and Programs in Biomedicine*, 156, pp.13–24. https://doi.org/10.1016/j.cmpb.2017.12.017.

192. Savkare, S.S. and Narote, S.P., 2016. Blood cell segmentation from microscopic blood images. In: *IEEE International Conference on Information Processing, ICIP*. Pune, pp.502–505. 10.1109/INFOP.2015.7489435.

193. Scherer, D., Müller, A. and Behnke, S., 2010. Evaluation of pooling operations in convolutional architectures for object recognition. In: *20th International Conference on Artificial Neural Networks (ICANN)*. Thessaloniki, Greece.

194. Scott, I. and Mazhindu, D., 2014. *Statistics for healthcare professionals : An introduction*. 2nd ed. London: SAGE.

195. Sergyan, S., 2008. Color histogram features based image classification in content-based image retrieval systems. In: *6th International Symposium on Applied Machine Intelligence and Informatics*. Herlany, pp.221–224.

196. Serrano, J.M., Ruiz-lombera, R., Valdiande, J.J., Cobo, A., Lopez-higuera, J.M. and Member, S., 2016. Colorimetric analysis for on-line arc-welding diagnostics by means of plasma optical spectroscopy. *IEEE Sensors Journal*, 16(10), pp.3465–3471. 10.1109/ICSENS.2014.6985079.

197. Shabut, A.M., Tania, M.H., Lwin, K.T., Evans, B.A., Yusof, N.A., Abu-Hassan, K.J. and Hossain, M.A., 2018. An intelligent mobile-enabled expert system for Tuberculosis disease diagnosis in real time. *Expert Systems with Applications*, 114, pp. 65-77. https://doi.org/10.1016/j.eswa.2018.07.014.

198. Sharma, G., 2003. *Digital color imaging handbook*. New York: CRC Press.

199. Shen, L., Hagen, J. a. and Papautsky, I., 2012. Point-of-care colorimetric detection with a smartphone. *Lab on a Chip*, 12(21), p.4240-4243. 10.1039/C2LC40741H.

200. Shin, J., Choi, S., Yang, J.-S., Song, J., Choi, J.-S. and Jung, H.-I., 2017. Smart forensic phone: Colorimetric analysis of a bloodstain for age estimation using a smartphone. *Sensors and Actuators B: Chemical*, 243, pp.221–225. https://doi.org/10.1016/j.snb.2016.11.142

201. Sicard, C., Glen, C., Aubie, B., Wallace, D., Jahanshahi-Anbuhi, S., Pennings, K., Daigger,

G.T., Pelton, R., Brennan, J.D. and Filipe, C.D.M., 2015. Tools for water quality monitoring and mapping using paper-based sensors and cell phones. *Water Research*, 70, pp.360–369. https://doi.org/10.1016/j.watres.2014.12.005.

202. Sicasys Software GmbH, 2017a. *Spotxel® Reader 1.1 Smartphone-based colorimetric image reader quick start guide*. [pdf] Heidelberg: Sicasys Software GmbH. Available at: <https://www.sicasys.de/fileadmin/spotxel-reader/docs/SpotxelReader_QuickStartGuide.pdf> [Accessed 12 January 2018].

203. Sicasys Software GmbH, 2017b. *Spotxel® Reader*. [Google Play]. Available at: <https://play.google.com/store/apps/details?id=com.sicasys.spotxel&hl=en> [Accessed 12 January 2018].

204. Siemens Healthcare GmbH, 2018. *CLINITEK Status+ Analyzer*. [online] Available at: <https://www.healthcare.siemens.com/urinalysis-products/urinalysis-systems/clinitek-status-analyzer> [Accessed 19 May 2018].

205. Skandarajah, A., Reber, C.D., Switz, N.A. and Fletcher, D.A., 2014. Quantitative imaging with a mobile phone microscope. *PLoS ONE*, [online] 9(5), p.e96906. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4019540&tool=pmcentrez&rendertype=abstract> [Accessed 29 September 2015].

206. SMH17, 2016. *Image Analysis Toolset*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=tk.silviomarano.imageanalysistoolset&hl=en_GB> [Accessed 09 January 2017].

207. Smith, A.R., 1978. Color gamut transform pairs. In: *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*. New York, pp.12–19. 10.1145/965139.807361.

208. Smith, G.T., Dwork, N., Khan, S.A., Millet, M., Magar, K., Javanmard, M., Ellerbee Bowden, 2016. Robust dipstick urinalysis using a low-cost, micro-volume slipping manifold and mobile phone platform. *Lab on a Chip*, 58, pp.951–954. 10.1039/C6LC00340K.

209. Smith, J.E., Griffin, D.K., Leny, J.K., Hagen, J. a, Chávez, J.L. and Kelley-Loughnane, N., 2014. Colorimetric detection with aptamer-gold nanoparticle conjugates coupled to an android-based color analysis application for use in the field. *Talanta*, 121, pp.247–55. 10.1016/j.talanta.2013.12.062.

210. Soares, R.R.G., Santos, D.R., Chu, V., Azevedo, A.M., Aires-Barros, M.R. and Conde, J.P., 2017. A point-of-use microfluidic device with integrated photodetector array for immunoassay multiplexing: Detection of a panel of mycotoxins in multiple samples. *Biosensors and Bioelectronics*, 87, pp.823–831. https://doi.org/10.1016/j.bios.2016.09.041.

211. Soille, P., 2004. *Morphological image analysis: Principles and applications*. [e-book] Springer. Available through: Springer <https://www.springer.com/gb/book> [Accessed 31 August 2018].

212. Solmaz, M.E., Mutlu, A.Y., Alankus, G., Kılıç, V., Bayram, A. and Horzum, N., 2018. Quantifying colorimetric tests using a smartphone app based on machine learning classifiers. *Sensors and Actuators B: Chemical*, 255, pp.1967–1973. https://doi.org/10.1016/j.snb.2017.08.220.

213. Soni, A. and Jha, S.K., 2017. Smartphone based non-invasive salivary glucose biosensor. *Analytica Chimica Acta*, 996, pp.54–63. https://doi.org/10.1016/j.aca.2017.10.003.

214. Soria Olivas, E., 2010. *Handbook of research on machine learning applications and trends : Algorithms, methods and techniques*. Hershey, PA: Information Science Reference, IGI Global.

215. Statista, 2015. *Number of mobile phone users worldwide 2013-2019*. [online] The Statistics Portal. Available at: <https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/> [Accessed 24 April 2018].

216. Statista, 2017. *App Stores - Statistics & Facts*. [online] Available at: <https://www.statista.com/topics/1729/app-stores/> [Accessed 01 March 2017].

217. Sun, M., Li, Z., Xia, Y., Zhao, C. and Liu, H., 2018. Concentration cell-based potentiometric analysis for point-of-care testing with minimum background. *Analytica Chimica Acta*. https://doi.org/10.1016/j.aca.2018.09.029.

218. Suslick, K.S., Rakow, N.A. and Sen, A., 2004. Colorimetric sensor arrays for molecular recognition. *Tetrahedron*, 60(49), pp.11133–11138. doi:10.1016/j.tet.2004.09.007.

219. Sutherland, J.S., Mendy, J., Gindeh, A., Walzl, G., Togun, T., Owolabi, O., Donkor, S., Ota, M.O., Kon Fat, E.T., Ottenhoff, T.H.M., Geluk, A. and Corstjens, P.L.A.M., 2016. Use of lateral flow assays to determine IP-10 and CCL4 levels in pleural effusions and whole blood for TB diagnosis. *Tuberculosis*, 96, pp.31–36. https://doi.org/10.1016/j.tube.2015.10.011.

220. Sutton, R.S. and Barto, A.G., 2018. *Reinforcement learning : An introduction*. 2nd ed. London: MIT Press

221. Suykens, J.A.K, 2002. *Least squares support vector machines*. River Edge, NJ: World Scientific.

222. Suykens, J.A.K. and Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), pp.293–300. https://doi.org/10.1023/A:1018628609742.

223. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2015. Rethinking the Inception architecture for computer vision. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas. 10.1109/CVPR.2016.308.

224. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston. 10.1109/CVPR.2015.7298594.

225. Tania, M.H., Lwin, K. and Hossain, M.A., 2018. Advances in automated tongue diagnosis techniques. *Integrative Medicine Research*. https://doi.org/10.1016/j.imr.2018.03.001.

226. Tania, M.H., Lwin, K.T., Abuhassan, K. and Bakhori, N.M., 2017. An automated colourimetric test by computational chromaticity analysis: A case study of Tuberculosis test. In: *Fdez-Riverola F., Mohamad M., Rocha M., De Paz J., Pinto T. (eds) 11th International Conference on Practical Applications of Computational Biology & Bioinformatics. Advances in Intelligent Systems and Computing*, 616, pp.313–320. Cham: Springer. https://doi.org/10.1007/978-3-319-60816-7_38.

227. Tania, M.H., Lwin, K.T. and Hossain, M.A., 2016. Computational complexity of image processing algorithms for an intelligent mobile enabled tongue diagnosis scheme. In: *10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*. Chengdu, pp.29–36. 10.1109/SKIMA.2016.7916193.

228. TheDrApps, 2014. *Whitening At Home(WAH)*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=appinventor.ai_sacevedo704.WAH> [Accessed 10 January 2017].

229. Theodoridis, S.; Koutroumbas, K., 2009. *Pattern recognition*. 4th ed. London: Academic Press.

230. Thevenaz, P. and Unser, M., 2008. Snakuscules. *IEEE Transactions on Image Processing*, 17(4), pp.585–593. 10.1109/TIP.2007.914742.

231. TinEye, 2017. *TinEye Labs*. [online] Available at: <http://labs.tineye.com/> [Accessed 09 January 2017].

232. Tools, R.L., 2015. *Colorimeter*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.colorimeter&hl=en_GB> [Accessed 09 January 2017].

233. Tools, R.L., 2016. *Color Analysis*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.color.coloration&hl=en_GB> [Accessed 10 January 2017].

234. Torres, A., 2016. *Colorimetry Pro*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=uy.com.bigcheese.colorimetrypro> [Accessed 10 January 2017].

235. Tracey, B.H., Comina, G., Larson, S., Bravard, M., López, J.W. and Gilman, R.H., 2011. Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Boston, pp.6017–6020. 10.1109/IEMBS.2011.6091487.

236. Trumbo, T.A., Schultz, E., Borland, M.G. and Pugh, M.E., 2013. Applied spectrophotometry: Analysis of a biochemical mixture. *Biochemistry and Molecular Biology Education*, 41(4), pp.242–250. https://doi.org/10.1002/bmb.20694.

237. Tsai, T.-T., Shen, S.-W., Cheng, C.-M. and Chen, C.-F., 2013. Paper-based tuberculosis diagnostic devices with colorimetric gold nanoparticles. *Science and Technology of Advanced Materials*, 14(4), p.044404. 10.1088/1468-6996/14/4/044404.

238. Tukey, J.W., 1977. *Exploratory data analysis*. PA: Addison-Wesley.

239. U.S. Census Bureau, 2017. *U.S. and World Population Clock*. [online] Available at: <https://www.census.gov/popclock/> [Accessed 07 August 2017].

240. UKVI, 2017. *Tuberculosis tests for visa applicants*. [online] Available at: <https://www.gov.uk/tb-test-visa> [Accessed 24 January 2017].

241. Ulep, T.-H. and Yoon, J.-Y., 2018. Challenges in paper-based fluorogenic optical sensing with smartphones. *Nano Convergence*, 5(14). https://doi.org/10.1186/s40580-018-0146-1.

242. Universal Medical, 2018. *BrandTech BRANDplates cellGrade Sterile 96-Well Plate with Lids*. [online] Universal Medical. Available at: <https://www.universalmedicalinc.com/brandtech-brandplates-96-well-plate-cellgrade-polystyrene-sterile-with-lids.html> [Accessed 22

September 2018].

243. University of Waikato, 2017. Weka. [computer program] University of Waikato. Available at: <https://www.cs.waikato.ac.nz/ml/weka/> [Accessed 18 September 2018].

244. Variable, I., 2017a. *CMG Color*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.variable.inspire.cmg&hl=en_GB> [Accessed 10 January 2017].

245. Variable, I., 2017b. *COLOR MUSE*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.variable.inspire&hl=en_GB> [Accessed 09 January 2017].

246. Variable, I., 2017c. *MeasureColor*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.variable.inspire.measurecolor&hl=en_GB > [Accessed 09 January 2017].

247. Vashist, S., Schneider, E. and Luong, J., 2014. Commercial smartphone-based devices and smart applications for personalized healthcare monitoring and management. *Diagnostics*, 4(3), pp.104–128. https://doi.org/10.3390/diagnostics4030104.

248. Vashist, S.K., van Oordt, T., Schneider, E.M., Zengerle, R., von Stetten, F. and Luong, J.H.T., 2015. A smartphone-based colorimetric reader for bioanalytical applications using the screen-based bottom illumination provided by gadgets. *Biosensors and Bioelectronics*, 67, pp.248–255. https://doi.org/10.1016/j.bios.2014.08.027.

249. Vistech.projects, 2014. *Color Meter Free - live colors*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.vistechprojects.colormeterfree&hl=en_G B> [Accessed 09 January 2017].

250. Vistech.projects, 2016. *Color Meter grab color live*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.vistechprojects.colormeter&hl=en_GB> [Accessed 09 January 2017].

251. Visual Studio, 2018. Xamarin App Development with Visual Studio. [computer program] Microsoft. Available at: <https://visualstudio.microsoft.com/xamarin/> [Accessed 24 September 2018].

252. VRprod, 2016. *Colorimeter*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.vrprod.colorimeter&hl=en_GB> [Accessed 09 January 2017].

253. Walpole, R.E., Myers, R.H., Myers, S.L. and Ye, K.E., 2011. *Probability and statistics for engineers and scientists*. 9th ed. Boston: Prentice Hall.

254. Wang, S., Xu, F. and Demirci, U., 2010. Advances in developing HIV-1 viral load assays for resource-limited settings. *Biotechnology Advances*, 28(6), pp.770–781. https://doi.org/10.1016/j.biotechadv.2010.06.004.

255. Wang, X.-Y., Wu, Z.-F., Chen, L., Zheng, H.-L. and Yang, H.-Y., 2016a. Pixel classification based color image segmentation using quaternion exponent moments. *Neural networks : The official journal of the International Neural Network Society*, 74, pp.1–13. https://doi.org/10.1016/j.neunet.2015.10.012.

256. Wang, X., 2013. *Digital tongue color imaging and analyzing for traditional Chinese medicine applications*. Ph.D. Hong Kong Polytechnic University. Available at: < http://hdl.handle.net/10397/6158> [Accessed 07 March 2017]

257. Wang, Y., Li, Y., Bao, X., Han, J., Xia, J., Tian, X. and Ni, L., 2016b. A smartphone-based colorimetric reader coupled with a remote server for rapid on-site catechols analysis. *Talanta*, 160, pp.194–204. https://doi.org/10.1016/j.talanta.2016.07.012.

258. Wei, Q., Nagi, R., Sadeghi, K., Feng, S., Yan, E., Ki, S.J., Caire, R., Tseng, D. and Ozcan, A., 2014. Detection and spatial mapping of mercury contamination in water samples using a smart-phone. *ACS Nano*, 8(2), pp.1121–1129. 10.1021/nn406571t.

259. Weiss, G.M., 2004. Mining with rarity: A unifying framework. *ACM SIGKDD Explorations Newsletter*, 6(1), pp.7–19. 10.1145/1007730.1007734.

260. Weiss, G.M. and Provost, F., 2003. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, 19, pp.315–354.

261. Welch, B.L., 1947. The generalization of `Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2), pp.28–35. 10.2307/2332510.

262. Wirth, M., Biswas, N., Ahmad, S., Nayak, H.S., Pugh, A., Gupta, T. and Mahmood, I., 2018. A prospective observational pilot study to test the feasibility of a smartphone enabled uChek© urinalysis device to detect biomarkers in urine indicative of preeclampsia/eclampsia. *Health and Technology*, pp.1–6. https://doi.org/10.1007/s12553-018-0248-0.

263. Wongsrichanalai, C., Barcus, M.J., Muth, S., Sutamihardja, A. and Wernsdorfer, W.H., 2007.

A review of malaria diagnostic tools: Microscopy and rapid diagnostic test (RDT). *American Journal of Tropical Medicine and Hygiene*, 77(Suppl. 6), pp.119–127..

264. World Bank Group, 2016. *World Development Report 2016: Digital Dividends*. [online] Washington DC. Available at: <http://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-0671-1> [Accessed 10 May 2018].

265. World Health Organization, 2017. *Global Health Observatory (GHO) data*. WHO. https://doi.org/10.1596/978-1-4648-0671-1.

266. Wu, Y., Boonloed, A., Sleszynski, N., Koesdjojo, M., Armstrong, C., Bracha, S. and Remcho, V.T., 2015. Clinical chemistry measurements with commercially available test slides on a smartphone platform: Colorimetric determination of glucose and urea. *Clinica Chimica Acta*, 448, pp.133–138. https://doi.org/10.1016/j.cca.2015.05.020.

267. X-Rite, Inc., 2015. *ColorTRUE*. [Google Play] Available at: <https://play.google.com/store/apps/details?id=com.xrite.mobiledisplaycalibration&hl=en_GB> [Accessed 09 January 2017].

268. Xiao, W., Huang, C., Xu, F., Yan, J., Bian, H., Fu, Q., Xie, K., Wang, L. and Tang, Y., 2018. A simple and compact smartphone-based device for the quantitative readout of colloidal gold lateral flow immunoassay strips. *Sensors and Actuators B: Chemical*, 266, pp.63–70. https://doi.org/10.1016/j.snb.2018.03.110.

269. Xu, D., Huang, X., Guo, J. and Ma, X., 2018. Automatic smartphone-based microfluidic biosensor system at the point of care. *Biosensors and Bioelectronics*, 110, pp.78–88. https://doi.org/10.1016/j.bios.2018.03.018.

270. Yan, S., Lai, X., Du, G. and Xiang, Y., 2018. Identification of aminoglycoside antibiotics in milk matrix with a colorimetric sensor array and pattern recognition methods. *Analytica Chimica Acta*. 1034, pp. 153-160. https://doi.org/10.1016/j.aca.2018.06.004.

271. Yang, R.-J., Tseng, C.-C., Ju, W.-J., Wang, H.-L. and Fu, L.-M., 2018. A rapid paper-based detection system for determination of human serum albumin concentration. *Chemical Engineering Journal*, 352, pp.241–246. https://doi.org/10.1016/j.cej.2018.07.022.

272. Yanikoglu, B., Aptoula, E. and Tirkaz, C., 2014. Automatic plant identification from photographs. *Machine Vision and Applications*, 25(6), pp.1369–1383.

273. Yetisen, A.K., 2014. *Holographic sensors*. [e-book]. Springer. Available through: Springer <https://www.springer.com/gb/book> [Accessed 07 August 2017]

274. Yetisen, A.K., Akram, M.S. and Lowe, C.R., 2013. Paper-based microfluidic point-of-care diagnostic devices. *Lab on a Chip*, 13(12), pp.2210–2251. 10.1039/C3LC50169H.

275. Yetisen, A.K., Martinez-Hurtado, J.L., Garcia-Melendrez, A., da Cruz Vasconcellos, F. and Lowe, C.R., 2014. A smartphone algorithm with inter-phone repeatability for the analysis of colorimetric tests. *Sensors and Actuators B: Chemical*, 196, pp.156–160. https://doi.org/10.1016/j.snb.2014.01.077.

276. You, D.J., Park, T.S. and Yoon, J.-Y., 2013. Cell-phone-based measurement of TSH using Mie scatter optimized lateral flow assays. *Biosensors and Bioelectronics*, 40(1), pp.180–185. 10.1016/j.bios.2012.07.014..

277. Yu, L., Shi, Z.Z., Fang, C., Zhang, Y.Y., Liu, Y.S. and Li, C.M., 2015. Disposable lateral flow-through strip for smartphone-camera to quantitatively detect alkaline phosphatase activity in milk. *Biosensors and Bioelectronics*, 69, pp.307–315. https://doi.org/10.1016/j.bios.2015.02.035.

Appendix A

Key Terminologies

In this section, a number of key definitions are clarified to provide a better understanding for readers from different disciplines.

Colourimetric test

This provides a decisive analysis of the present elements or concentration of the chemical compound facilitated by a colour agent.

Analyte

The point of interest, which could be a substance or chemical constituent that will go through an analytical process, is called an analyte.

Matrix

In computer science, a matrix refers to the mathematical matrix that represents a rectangular array consisting of numbers, which can be arranged in rows and columns. In a biochemical test, the components apart from the analyte are known as matrix.

Target

Although in machine learning, targets refer to the class label, this thesis also deals with biological targets. A biological target signifies an entity, to which another entity is directed. The target could be inside of a living organism, and this 'other' entity may bind or react with the target, which will cause a change in the behaviour or function of the target or sample being tested. For a colourimetric test, the change of behaviour will appear as a change of colour. The focus of this thesis is this colour change and how human visualisation of this colour can be transferred to the machine vision on the mobile platform, whereas the tristimulus colourimeter and spectrophotometry are prominent techniques used in clinical chemistry for colour analysis (Trumbo et al., 2013; ICC, 2004).

Assay

It is a mean to investigate the analyte. It is an analytical process in number of disciplines to determine the presence of a target entity or quantify the amount or functional activity of the analyte.

ELISA

It is a type of an assay involving a solid-phase enzyme immunoassay that can determine the presence of a ligand in a liquid sample with the assistance of antibodies. The technique is primarily used in diagnosis, plant pathology and industrial quality control.

LFA

LFA is a type of colourimetric test scheme and is commonly used either as a qualitative or a semi-quantitative assay. There are few other popular names for LFAs, such as lateral flow immune-chromatographic assays, rapid test, test strip and dipstick.

Colour Vision

Based on the wavelength of light, colour vision is the capability of human, animal and machine to differentiate objects. Recognising an object can be an easy task for a human, but it can be challenging for a machine. Regarding colour, the human vision of different individual can perceive the same colour in different ways.

Colour Perception

The physiology of colour perception involves pigments in retinal cells possessing particular spectral sensitivities, which are known as cone cells. Regarding colour perception, two things should be noted: i) even people with normal colour vision can have variation in the peak response of cone cells and ii) colour vision degrades at an older age (National Eye Institute, 2015).

Colour Space

The world is full of an infinite combination of colours. The human colour perception was quantified for the first time using CIE (International Commission on Illumination) 1931 XYZ colour space. It is the foundation of most colour spaces. According to CIE 1931, the response of human eyes can be portrayed as a horseshoe shape diagram. However, there are commercial colour spaces which were developed to improve the image quality.

The abstract mathematical model of colours in forms of triples (or quadruples) of numbers is referred to the colour model. A colour space signifies a precise combination of the colour model and the mapping function. Therefore, denoting a colour space involuntarily specifies the colour model. In this thesis as well as in the literature, these two terms are often used as interchangeable terminologies.

Perceptual Linearity

It is desirable that the perceived visual sensation be linearly correlated with the original form of the data (Cheng and Badano, 2018). This attribute is expressed as perceptual linearity. For

instance, using luminance of a monochromatic display, such as X-Ray, stimulates the user to prompt a visual sensation of brightness.

Perceptual Uniformity

A two-dimensional colour wheel represents hue variables and lightness. If chroma or colourfulness is added to this concept, then it results in a three-dimensional (3D) representation of a colour model. The purpose of this added dimension is to describe colour variation. This 3D structure is expressed as colour solid.

The logical expectation would be to perceive any colour difference within the colour solid equally. In this way, every possible 'object colour percept' would be arranged uniformly. This phenomenon is expressed as perceptual uniformity. In reality, it is not possible to develop such colour solid due to geometric constraints (Kuehni, 2010).

Data Pre-processing

This process involves converting raw data into an understandable format. In this thesis, the extent of data pre-processing includes image pre-processing, segmentation, post-processing, feature extraction and feature selection.

Machine Learning

"The field of Machine Learning seeks to answer the question "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?" (Michalski, Carbonell and Mitchell, 1983). It is a subset of the wider domain of AI which concentrates on 'teaching' computers regarding how to 'learn' without requiring to be programmed for explicit tasks (Pal and Gulli, 2017).

Deep Learning

Deep learning is a subset of machine learning. Therefore, rest of the models can be thought of as the traditional machine learning techniques. Deep learning possesses higher power and greater flexibility (as compared to traditional machine learning) to perceive as well as to represent a problem in the form of a nested hierarchy of concepts.

Appendix B

Computer-aided Colourimetric Test

This work categorises the computer-aided colourimetric test into the following four groups: systems with additional hardware attachments, proof-of-concept provided as a desktop application, mobile-enabled systems, and server-based approach. Some of the systems described in this section incorporated more than one of these approaches. For example, an interesting research was conducted by Feng et al. (2014) involving a Google glass (Glass Partners, 2017) for the qualitative detection of HIV and quantitative determination PSA from RDTs. The system at the same time falls under the category of a hardware and a server-based system. The Google glass is consist of a camera, prism display, and inertia, proximity and ambient light sensors along with many other features. Therefore, the system can be considered as an alternative to the smartphone-based approach as well. Due to the advanced features provided by Google Glass and the server connection, it can be also be thought of as a complete reader. Therefore, one may argue whether the system is using any additional attachment or not.

B.1 Hardware Based Systems

Substantial amount of work conducted in the literature are hardware based, which are often presented as lab-on-chip assays (Akraa et al., 2018; Roda et al., 2016; Vashist, Schneider and Luong, 2014; Lin and Scott, 2012a; Coskun et al., 2013; Coskun, Ahmet F.Wong et al., 2012; Zhu et al., 2011). This section studies the functionality of the hardware-based colourimetric tests to explore the possible replacement with an intelligent system.

The influence of optomechanical attachments on mobile-enabled microfluidic biosensor systems in the recent articles is clearly visible from the survey conducted by Xu et al. (2018). Xu et al. admitted that expense, size and ease of use are the constraints of traditional biosensor systems to be implemented on POC platform. While reviewing the system architecture, analysis procedure and sensing mechanism of the biosensors suitable for POC platforms, the study showed the potential and application of the image-based systems.

The additional hardware as an optomechanical attachment has been utilised for various purposes in the literature. For example, Akraa et al. (2018) presented a system to eliminate device agnosticism to perform urinalysis for CKD. The mobile devices with various size will have an impact on the exposure plane while capturing an image of the sample, which can effectively alter the colour intensity and size of the sample. Therefore, Akraa et al. designed

and developed an image housing attachment with adjustable viewports and clamp controls for self-monitoring of CKD.

Aymerich et al. (2018) developed an electrochemical instrument to be attached with a smartphone to measure alcohol in the blood. Yang et al. (2018) developed an analytic box for measuring Human Serum Albumin (HSA) concentration using a smartphone. The assay type used in the study was a quantitative one, which was validated on 36 real samples. Their assay benefited from built-in heating chamber and battery placed in an analytic box; these hardware are not the point of interest in this thesis. However, the box also contains a camera connector, lens and light-guide; these are the hardware attachments under scrutiny in this section.

Xiao et al. (2018) used a colloidal gold lateral flow immunoassay and extended its utilisation from qualitative to quantitative using additional hardware attachments with a smartphone. The optomechanical attachment includes LED, battery, resistor, switch, holes and holder to be attached with a mobile phone. Calabria et al. (2017) developed a mobile-enable oral fluid L-lactate detection scheme that can provide the result taking one minute. The hardware attachment includes a small dark box and a smartphone adapter. Kim et al. (2017) used a small dark box of 25mm x 12mm x 12mm dimension containing additional lens, and mirror and diffuser to guide the LED light to maintain a consistency in imaging and lighting condition while capturing the image of paper strips with a mobile phone camera to measure the alcohol concentration in saliva. Lee et al. (2014) used a mobile-enabled system with a small hardware attachment containing an LED and battery to determine vitamin-D deficiency on the POC platform. Considering the conventional laboratory scale detection method, the presented system by Lee et al. was undoubtedly a useful method.

The reported articles (Kim et al., 2017b; Coskun, Ahmet F.Wong et al., 2012) often used hardware attachments that acted as the sample holder as well. Kim et al. (2017b) developed a compartment to place the sample in a tube to analyse the luminescence using a smartphone. Coskun, Ahmet F.Wong et al. (2012) presented a system with hardware attachments for smartphones with 22mm x 67mm x 75mm dimension which has a platform for both sample and control, and the optical arrangement includes additional lenses, LEDs, light diffusers and aperture to control the imaging field-of-view. Their study claimed that the developed system is suitable for home-settings to perform allergen test. Later, Coskun et al. (2013) demonstrated a mobile-enabled system with an optomechanical attachment using laser technology for fluorescent assays to detect albumin. Earlier, Zhu et al. (2011) presented a similar system for fluorescent microscopy. These studies (Coskun, Ahmet F.Wong et al., 2012; Coskun et al., 2013; Zhu et al., 2011) can be considered as an extended scope of the current context.

As compared to the similar recent researches mentioned above, Konnaiyan et al. (2017) utilised a relatively bulky attachment that can contain the assay, i.e. urine dip-stick, which can aid in the detection of early signs of pregnancy complications such as preeclampsia and gestational diabetes. The 3D-printed sample holder was utilised to improve the sensitivity of the assay and control the illumination condition. Previously, Ozkan and Kayhan (2016) discussed in their study on the development of automatic RDT platform that 3D-printed attachments can function as both advantageous and disadvantageous way.

The reported articles more often failed to scale-down of the POC system due to their involvement with various tabletop boxes. Chen et al. (2018) presented a mobile-enabled hardware-based image acquisition scheme for microfluidic assay at POC platform utilising two case studies, which are colourimetric analysis of bicinchoninic acid (BCA) based protein detection assay and ABO blood type determination. The hardware attachment is consist of an internal backlighting and magnification system placed in 37mm x 26mm x 23mm box. Purpose of this attachment was to convert the smartphone light source into a backlight illumination source, guide the light waves and magnify the image of a sample. Later, Chen et al. (2018b) improved the box, however, failed to make it box-free.

Barnes et al. (2018) designed a mobile-enabled real-time diagnostic system for microbial infections. The assay is required to be placed in a cardboard box, probably for heat insulation. The 96-well assay plate requires the aid of DC powered 96 LEDs and a bandpass filter to be processed by the mobile application.

Li et al. (2018) placed plastic micro-pit array chips inside of a six cm long paper box with a hole on top for the smartphone camera lens and an astral lamp at the bottom to provide the quantification of HIV p24 antigen. Kim et al. (2017c) placed a colourimetric assay for blood haematocrit analysis inside of an acrylic box to capture the image using a mobile phone camera and 1 cm poly-dimethylsiloxane light diffuser. Their system showed only a 0.1% limit of detection to analyse the blood. Wang et al. (2016b) presented a mobile-enable server-based system for rapid catechol analysis. The system was able to differentiate between thirteen catechols from images of a 96-well Perspex plate placed in a light-tight box containing white LED lights as well.

Masawat, Harfield and Namwong (2015) used a plywood based photography light-box of 10xm x 16.5cm x 25cm dimension to capture the image with an iOS device and ColorCon mobile application. The image was captured from outside of the light-tight box using a drill hole to trace tetracycline in milk.

Vashist et al. (2015) developed a mobile-enabled colourimetric reader and tested the system for 96 or 24 well microtitre plate of different assays such as direct and sandwich ELISA, and BCA. Similar to the systems above, the assay requires to be placed inside a box. The system described by Vashist et al. (2015) not only involved additional hardware but also necessitated a virtual plate to capture the image adequately. However, the study discussed the potential of more small, compact and cost-effective systems in the future.

Comparing the impact of the additional hardware attachments, Soni and Jha (2017) concluded that the lighting condition has to be dealt with either using an algorithm or using a hardware attachment. Their study included three smartphones (Samsung Galaxy SIII, HTC Desire 526G Plus and Gionee P5 mini) and the assay was placed both inside and outside of a dark box to analyse saliva glucose. The conclusion was drawn from the observation of the impact of camera positioning and analysis of flashlight intensity on the biosensor.

Using both flashlight and no-flashlight, and additional hardware attachment and without hardware attachment, Mutlu et al. (2017) showed that it is possible to provide a colourimetric classification with high accuracy without requiring such attachments.

B.2 Desktop Applications

A proof-of-concept for a computer-aided colourimetric test is often provided as a desktop application rather on a mobile platform or any standalone device. However, in most cases, these applications utilised additional hardware attachments, which are already discussed in the previous section. As the colourimetric test is a point of interest for many disciplines such as biochemistry, agriculture, nanotechnology, optoelectronics, a detailed description of the computational system is often not provided with the literature. Another concern to analyse these articles is lack of a consistent format to present the result such as regarding accuracy, specificity, sensitivity, error rate, convergence and computation time.

Bu et al. (2018) captured that images of pregnancy test strips with a smartphone to determine the pathogenic bacteria in milk. The images were analysed on a desktop platform using ImageJ software (Rasband, 2016). The image was processed by separating the background after grayscale conversion, and the colourimetric decision was produced from the relative pixel intensity of the test and control band. Using a calibration curve[29], the recovery rate was within

---

[29] In order to understand a matrix effect, the colourimetric tests often utilise a calibration curve using a standard sample, where the concentration of the analyte is known.

80.7 to 141.7%. The study precluded all the detail of the image processing, colourimetric decision and computation time.

Garg et al. (2014) presented a system to analyse blood glucose from a paper-based assay. The assay requires to be attached with the computational device using additional hardware. In this way, the sample position is fixed with respect to the camera, which can aid while processing the image. Their study included an automatic image processing algorithm to determine the ROI and provided a quantification of the analytes using linear least-squares. The result achieved by the system can have 3-5% error. Garg et al. claimed that the system can be deployed as a mobile application and can provide quantification of any analyte.

Mutlu et al. (2017) performed colourimetric classification of pH-test strips using least-squares support vector machine (LS-SVM) with and without additional hardware and attained 100% accuracy for both cases. Although a smartphone captured the images, the proof-of-concept was provided on the desktop platform.

Yu et al. (2015) also used an LFA to measure the activity of alkaline phosphatase milk by capturing images of the assay using a smartphone. The AuNP based ALP testing zone was analysed in the MATLAB platform by quantifying the optical intensity. The proposed system traced $0.1–150$ U L^{-1} ALP within 10 minutes.

The colour-based analysis is well utilised in the literature beyond colourimetric tests and healthcare applications such as plant identification (Yanikoglu, Aptoula and Tirkaz, 2014) and ripeness of fruit (Phothisonothai, Tantisatirapong and Aurasopon, 2016). In order to maintain the focus of the study concentrated toward the colourimetric test, such articles were not reviewed in this chapter. Nevertheless, conceptually, if an algorithm is efficient enough, it can be developed as a mobile application. Our research suggests that, in such cases, the algorithm would have to comply with the limited capacity of the mobile devices.

B.3 Mobile-enabled Systems

The mobile application market is booming and exponentially increasing every year. In 2016, 224.8 billion mobile applications were downloaded from different platforms (Statista, 2017). From these application downloads, Apple Inc. reported revenue of £15.95 billion for developers. The mobile applications are estimated to generate £150.67 billion revenues from app-stores and in-app advertising in 2020. Among different studies, considering the one predicted lowest revenue is illustrated in Fig. B. 1.
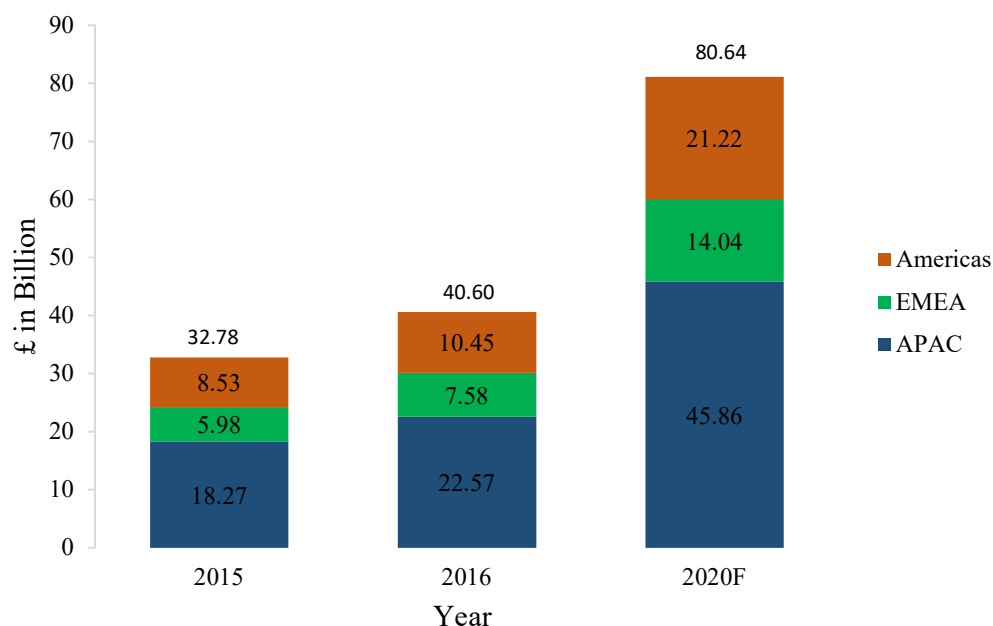
Fig. B. 1: Region-based forecast of mobile applications' annual revenue (App Annie, 2016)

Medical application market for healthcare professionals is expected to reach £11.67 billion by 2020; which implies, 14.47% of the revenue in mobile application will come from the applications designed for health professionals (Markets and Markets, 2015, 2016). The forecast is shown in Fig. B. 2.
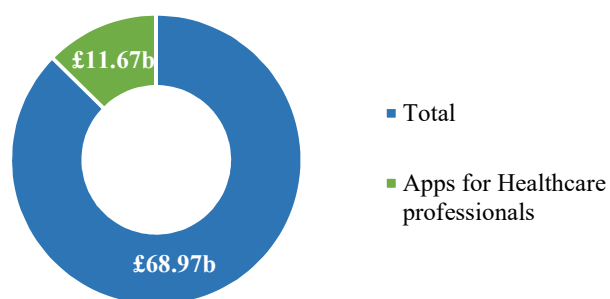
Fig. B. 2: Forecast of mobile applications' revenue for health professional in 2020. (Unit is in Billion GBP)

The market for mobile health (mHealth) solutions connected with external devices (e.g. glucose meter, BP monitor) is estimated to reach £47.14 billion by 2020. The global immunoassay market, considering both product and service, is expected to reach £20.30 billion by 2021 from £13.69 billion in 2016 (Markets and Markets, 2016). These statistics indicate the financial aspect and potential of intelligent colourimetric tests.

However, currently, there are only a few mobile applications available in the commercial and public app-stores to perform colourimetric tests on a 96-well plate, e.g. Spotxel® Reader (Sicasys Software GmbH, 2017b), Enzo ELISA Plate Reader (Enzo Life Sciences inc., 2015) and AssayColor (Alidans srl, 2015). Enzo ELISA Plate Reader and AssayColor neither provide any automatic complete analysis nor include any decision support system to interpret the colourimetric results. The Spotxel® Reader (Sicasys Software GmbH, 2017b) comprising plate annotation and alignment, uses powerful noise processing and signal detection techniques. Based on our research, this is the best mobile application till date to perform wet-chemical-based colourimetric quantification on cross-platform.

Instead of intelligent sensing, the Spotxel® Reader (Sicasys Software GmbH, 2017b) uses a virtual plate which can be laid over the plate image. The application expects the wells to be aligned with the virtual plate. The user is required to match the corner and centre wells with the grid. The virtual plate or grid can be scaled and rotated. However, aligning the wells with the grid requires some image capturing skills, which reduces the ease of use. The developers also acknowledged the limitations in the image processing scheme (Sicasys Software GmbH, 2017b). The application is capable of performing statistical analysis to quantify the result. The accuracy of such quantification is yet to be revealed.

Recently, for the first time, a computer-aided, to be precise a mobile-enabled urinalysis system (Healthy.io, 2018) has achieved clinical approval to be commercialised in EU. Although the application can be downloaded from the app-store, access would require an endorsement from a general practitioner.

It is not uncommon for the literary works to use a commercially available mobile applications to provide a proof-of-concept. By utilising a commercially available mobile application[30] (FTLapps, 2012), Wu et al. (2015) performed colourimetric analysis for Vitros® glucose and urea colourimetric assays using iPhone 4. The performance of the system was compared with the gold standard instruments and was found to be equal or more accurate than the conventional methods.

Due to the parallel advancement in bio-sensing as well as handheld technology, a number of critical surveys are conducted in this context (Hernández-Neuta et al., 2018; Ulep and Yoon, 2018; Roda et al., 2016; Vashist, Schneider and Luong, 2014; Yetisen, Akram and Lowe, 2013). While exploring paper-based microfluidic assays, Yetisen, Akram and Lowe (2013) anticipated

---

[30] Full list of such applications is provided in Appendix C.

such systems to be free from additional hardware attachment. Exploring smartphone based biosensors and bioelectronics, Zhang and Liu (2016) favoured for a mobile-enabled platform for colour analysis such as DNA imaging, serum diagnosis, and environmental monitoring. Zhang and Liu stressed on the issue with computational efficiency and ease of use. Hernández-Neuta et al. (2018) anticipated these mobile-enabled systems to contemplate the cost and robustness factors.

Vashist, Schneider and Luong (2014) studied mobile-enabled systems for personalised healthcare and discussed its capability to facilitate better management of healthcare systems to compensate for the imbalance between patient and health professionals. While critically surveying the recent biosensors, Roda et al. (2016) also favoured for the mobile-enabled POC systems due to its ubiquitous distribution and connectivity. The survey indicated a widespread research focus on colourimetry, reflectance and luminescence to utilise smartphone camera as a detector or instrumental interface. Although most of these research works are just a proof-of-concept, Roda et al. (2016) stressed on its potential of real-time self-measurement of target analytes. However, in literature, it is difficult to find a mobile-enabled system for the colourimetric test without using any additional hardware. A list of such systems without hardware attachments are provided in Table B. 1.

Table B. 1: Mobile-enabled colourimetric test in different platforms

Reference	Android	iOS	Others
Solmaz et al. (2018)	Yes	x	x
Alankus et al. (2018)	Yes	x	x
Karlsen and Dong (2017)	Yes	Yes	Yes
Jonas et al. (2016)	x	Yes	x
Wu et al. (2015)	x	Yes	x
Sicard et al. (2015)	x	Yes	x
Smith et al. (2014)	Yes	x	x
Yetisen et al. (2014)	Yes	Yes	x
Cooper et al. (2012)	Yes	Yes	x
Cooper (2012)	Yes	Yes	x

Solmaz et al. (2018) developed a mobile application named ChemTrainer to perform semi-quantitative colourimetric test on peroxide paper strip supported in the Android platform of 4.0.3 or higher. The application supported the facility to capture new images as well as to use an image stored from the gallery. The user would require to crop the image to separate the ROI and then mean R, G and B values were measured before sending the information to a message-

oriented middleware service. A progress bar was included in the application to inform the user about the status of the analysis. The result of the test was displayed to the user after obtaining it from the server. The application was tested on six Android smartphones.

Sun et al. (2018) showed the scope of a potentiometric method based colourimetric test using a mobile phone. The system developed by Yang et al. (2018) included a hardware attachment, mobile application and server integration. However, no discussion is provided in the article regarding the mobile application. Similarly, Konnaiyan et al. (2017) excluded the detail description of the mobile phone application.

Karlsen and Dong (2017), a part of the doctoral research of Karlsen, (2018), presented a mobile-enabled smart-diaper to detect urinary biomarkers. The system utilised a commercially available mobile application (Flavio Gonzalez App-Entwicklung, 2016). The study performed colourimetric classification of three pH levels, i.e. 4, 7 and 10 from the colour-references of ten pH levels (1-10) without any optomechanical attachment and tested on Samsung Galaxy S6 Edge, S7 Edge, ZTE Nubia V7 mini and iPhone 6. Earlier Chen, Wu and Dong (2014), in their work on smart-diaper, discussed the potential of a mobile application in this context.

Another relevant doctoral research was conducted by Yetisen (2014). For quantification of the colourimetric test, Yetisen et al. (2014) developed a cross-platform smartphone application that can measure the concentration of glucose, protein and pH from a multiplex assay. The capturing process started with calibration at the given light conditions, i.e. colour, intensity, tone, followed by the user's inputs of sensor type, target analyte, unit of concentration and number of data points. The image of the calibration point captured by the user was stored. The sample image of the corresponding test zone was processed utilising electromagnetic radiation from the coloured zone; the concentration of the analyte and corresponding value then was returned on the screen. Once the image was transformed, and the result was produced from the measured value versus the calibration curve, the information was required to be synchronised. The study asserted the scope of such system to replace the requirement of a spectrometer and microplate reader. The system featured interphone repeatability for urine analysis. The study claimed that the system could be implemented for different assays, e.g. colloidal gold, latex labels, solution based assay and microfluidic diagnostic device. However, the study did not provide any indication of how it is going to be implemented for different assays.

Jonas et al. (2016) performed a CRD test for diagnosis and prognosis of preeclampsia. The system intentionally used an older version of iPhone to make it cost effective. The developed system was tested on 300 samples. In addition to achieving a promising result, the system

demonstrated that a computer-aided system can make a colourimetric detection faster. They minimised the wet-chemical testing time for up to 2 minutes.

Shen, Hagen and Papautsky (2012) presented the potential of smartphone-based colourimetric tests, not to eradicate the conventional method, but to provide portable, transferable, immediate, low-cost diagnosis to a huge population with limited access to resources. They compensated the ambient lighting environment and formed the calibration curve of concentration from the chromaticity value to measure the pH. They envisioned their colour conversion analysis techniques to be useful to any POC diagnosis with the colourimetric response, even for fluorescence data.

Cooper et al. (2012) and Cooper (2012) used mobile image ratiometry to perform cocaine and benzoylecgonine analysis and on-sight fungus detection, respectively. Capturing the luminosity by Adobe Photoshop CS3, the binary class was drawn from the relative pixel densities. Although the detailed analysis was performed in the server, it is worth mentioning that these studies were conducted using only 3.2 megapixels camera. Use of such low-resolution mobile phone camera can be noticed in the early stage research works on mobile-enabled colourimetric tests such as Lin and Scott (2012b), where the mobile application utilised MATLAB to analyse the ROI. The result (AUROC) attained on the mobile platform was around 4% less than the biosensor platform. Lin and Scott probably compensated the low resolution of the camera by utilising backlighting and placed the assay in a box before taking images. Hence their research can also be categorised under the system with optomechanical attachments.

Smith et al. (2014) performed mobile-enabled aptamer AuNP based on-sight cocaine detection. The system was not free from hardware. Smith et al. (2014) explained that an LED light source placed beneath the assay could aid in combating the battle with ambient lighting condition, specifically the reflection of the ceiling light. Their assay plate should contain the known concentration to perform a qualitative test on an unknown sample. The colour of the known samples would act as the control or ground truth for the test samples. The performance was compared with the quantitative analysis performed by a conventional plate reader. The study indicated that a mobile-enabled system could perform even better than these plate readers. From the video of an experiment provided in the article (Smith et al., 2014), the following observations are made in this section: i) the system provides a flexibility to the users to correct the size of the image; ii) the control sample has to be provided in two specific wells, which is true for conventional plate readers as well and iii) user will require to manually select the sample to know the result, a better interface may solve this issue.

To monitor water quality by detecting organophosphate pesticides, Sicard et al. (2015) developed a geo-tagging enabled smartphone application utilising paper-based sensors. The images of test and control strips were captured in varying lighting condition by using a Flip-pal 100c mobile scanner, followed by colour intensity quantification using ImageJ software. The Objective-C based algorithm for the iOS 6 platform provided the processed image with RGB components and pixel counts. The system supported users' comments and manual editing like cropping feature, but the result had to be analysed from the Dose-response curve. Even with the inadequacy of the linear models consisting of the relative magnitude of the stressor, e.g. chemicals in response to the receptor called the Dose-response curve, the model is utilised well in water quality monitoring (Wei et al., 2014; Sicard et al., 2015). Our research suggests that, inclusion of an intelligent system can enhance such systems.

Alankus et al. (2018) analysed water quality using a custom-made Android application. The detection accuracy for chromium, phosphate, nitrite, and phenol were 100, 98.7, 100, and 76%, respectively. In order to make the system rapid on the mobile platform, the system calculated the result from the colour difference (ΔE), instead using any heavy algorithm. Undoubtedly the technique showed good performance; our observation suggests that the dependency on the user-interactions might not be suitable for non-technical personnel and could induce human-error.

In order to reduce the dependency on hardware and to produce the result of a colourimetric test using the limited capacity of the mobile phones, server-based approaches are often utilised. Therefore, our next section will provide an overview of server-based approaches for the colourimetric test.

B.4 Server-based Approach

The reported articles often utilise a server to process a colourimetric test result and send the result to others. Feng et al. (2014) used server-connected Google Glass to perform HIV and PSA test using RDTs. Google Glass envisions to bring ubiquitous computing into our daily life (Glass Partners, 2017). It has a built-in camera, wireless networking and voice-command facility. Using the system proposed by Feng et al. (2014), the user would require to define the test in the website, produce a QR code to identify the test and mark this QR code with the corresponding test before capturing an image of the RDT. The image was enhanced within the native device, i.e. Google Glass, however, required to be processed in the server due to the limited capacity of this personalised device. If there were no wireless connection available in time of the image capturing, the information could be preserved in the Glass, which would be uploaded to the server whenever any connection was available. In order to maintain a consistency in the camera to assay position to ensure the quality of the image and to instruct

the user to hold the assay in the correct orientation, a virtual guideline or overlay was utilised in the system. Once the result was processed in the server, it was sent back to the Google Glass. Each test took 8 seconds to produce and display the result. The test result was also obtainable via an internet browser. Geo-tagging was another added feature of the system.

Geo-tagging can aid in preventing and managing an epidemic, assist in health tracking among different demographics and provide environmental assessments. Sicard et al. (2015) used geo-tagging to monitor water quality by testing the pesticides level using a paper-based colourimetric test.

Wang et al. (2016b) used a remote server to measure catechols utilising LED and a light-tight box. The result was provided on a mobile platform. The connection between the server and a smartphone was established using a user datagram protocol. The server was comprised of database, processing and communication modules. The processing module utilised MATLAB whereas the communication module utilised Eclipse Integrated Development Environment, which served the purpose to receive the colour related features and sent the result to a smartphone after processing the result using machine learning algorithms. The study claimed that the result could be produced in real-time.

Similar to Wang et al. (2016b), Akraa et al. (2018) used additional hardware as well as a server, to perform a urinalysis. The purpose of the server was to reduce the computational burden from the mobile devices. The web-based system would require access to the smartphone camera of the user. Therefore, the mobile client was required to be developed separately using the Xamarin platform (Visual Studio, 2018). The system discussed the potential of such tests for telemedicine facilities. However, the system did not include any arrangement to connect with the health professionals; future researchers may explore such opportunities.

Recently, Solmaz et al. (2018) demonstrated a server-connected mobile application to provide a colourimetric classification of peroxide test strips. When a user provides a cropped image of the test strip, the mobile device client sends this information as well as a temporary unique queue identifier to RabbitMQ using cloud-based advanced message queuing protocol. Solmaz et al. (2018) used a desktop computer to act as a server and to perform the computation in MATLAB and presented the result in Android operated mobile devices. The system could handle multiple users, however the study lack in quantifying the computation time for a single as well as multiple users. The result obtained on mobile platform was only 90.3% to classify six classes of test strips possessing single object per sample. The accuracy obtained in the desktop environment was 95%. Our observation suggests that, as the image was cropped, instead using any advanced segmentation algorithm, the performance might have improved by

utilising a better classifier or fine-tuning the feature-set. One also needs to investigate if the test image by the users resulted in an improper cropping of the ROI. Moreover, the performance of the system described in Solmaz et al. degraded when the number of classes was extended to 11. No detailed explanation is provided in the article regarding the performance degradation for the change of platform and the extention of the classes.

Appendix C

Relevant Mobile Applications

C.1 Colourimeters in the App-store

There are number of colourimeters on the web (TinEye, 2017; Color Thesaurus, 2017) and often integrated to a mobile-enabled platform to match live colours as well as from previously stored images. Using word search 'colorimeter' in Google Play 58 mobile phone applications was found on January 9, 2017. The list is provided in Table C. 1.

Table C. 1: Colourimeters on android platform

Input	Camera/ Mobile phone	Web	Mobile	Free	OS (Android)	Other requirements	Features/ Remarks	Data share	Ref.
Live	back-camera with D65 white balance and manual exposure controls	x	√	√	≥ 5.0	Bluetooth	-Colour metering - Can work with wide range of camera - Take too long to 'warm up'	Facebook	(Camera Colorimeter, 2016, Calibrated Photo Viewer Demo, 2016)
Stored image	back-camera of Nexus 5, 6, 7, Samsung Galaxy S4, Tab S 10.5, Nook HD Plus	x	√	x	≥ 4.0.3		-Automated, comprehensive colour and gamma calibration (sRGB image). -Colour balance - No rooting -Correct colour space is automatically selected based on image's EXIF metadata. - Light exposure correction. -Works on JPEG format only		(Auralisoft, 2016a)

Stored Image	Versatile	x	√	√	≥ 4.0	x	- Colour Percentage - Colour Name - RGB & Hex Code - Sometimes inaccurate	Email, Cloud	(Leizer, 2016)
	Versatile (Illuminant = D65 Observer = (1931))	x	√	x	≥ 2.3	x	-Colour metering (CIE L * a * b RGB, XYZ and HEX, Chromaticity, HUE angle)	email , Bluetooth , USB	(Tools, 2016)
Live, Stored Image	Versatile	x	√	√/ x	≥ 2.2	x	- Colour metering - white balance - hex code - use colour as WALLPAPER - colour palette - Flashlight, Autofocus, Zoom View	√	(vistech.projects, 2014; Vistech.projects, 2016)
Live, Stored Image	Versatile	x	√	x	≥ 2.3	x	- Colour metering (CIE LAB, Chroma, Hueº, RGB, delta E) - colour names - real time visible spectra - Suitable for interior designing	x	(Tools, 2015)

n/m	n/m	x	√	√	≥ 4.4	Additional device	-Transmittance, Absorbance - Intended for experimental science e.g. chemistry, biology	x	(Mystrica, 2016)
Live, Stored Image	Versatile	x	√	√	≥ 5.0	COLOR MUSE scanning device for increased functionality	-Colour metering -Colour palette - Colour recommendation - Suitable for interior designing	√	(Variable, 2017b; c; a)
Live, Stored Image	Versatile	x	√	√	≥ 4.0.3		-Colour metering (RGB value, hex code) -Object Identifier -Logo Identifier - Censorship Risk Meter	Email, Cloud, text message and various messengers	(SMH17, 2016)
x	Not required	x	√	√	≥ 4.1	x	-Colour referencing (RGB, hex code) -No colour matching facility is included	x	(VRprod, 2016)
Stored Image	Samsung Galaxy, Nexus, HTC One, Moto X, LG, Xiaomi	x	√	√	≥ 4.0.3	x	-Checks the displayed colour quality of the mobile device - Not a colorimeter in the conventional sense. Suitable for photography	x	(X-Rite, 2015)

Live	Versatile	x	√	√	≥ 4.0	USB On The Go (OTG) capable Android tablet or phone. Bluetooth.	-Colour metering (XYZ, LAB, CCT) -Colour names -Spot reflectance	x	(Computer Graphics Technology P.L., 2015; P.L., 2015)
Live	Limited versatility	x	√	√	≥ 2.1	colorStriker colorimeter Bluetooth	-Colour metering including delta E	Zeroconf, Bonjour	(Fheldt, 2014)
Stored Image	Versatile	x	√	√	≥ 4.0.3	x	-Colour metering including delta E - Movable colour aiming cursor - User has to manually click on image and name the colour	x	(Nicksdroid, 2016)
n/m	n/m	x	√	√	≥ 3.1	RS232 to USB Gender (FTDI Chip Only), OTG Cable	Colour metering	x	(CM-H505 App, 2015)

n/m	Limited versatility	x	√	√	≥ 4.0.3	Open CV manager	-Colour metering -Suitable for field of chemistry	x	(Lab4U, 2016)
Stored Image	Versatile	x	√	√	≥ 3.0	Spyder4™ colorimeter	-Colour correction -Suitable for field of photograph	x	(Datacolor, 2014)
Stored Image	Versatile	x	√	√	≥ 1.5	x	-Colour comparison for teeth -Has lot of features. Suitable for field of dentistry	x	(TheDrApps, 2014)
Stored Image	Versatile	x	√	x	≥ 4.0	√/x	- Colour metering (RGB value, Hex code) -Suitable for the field of analytical chemistry, biochemical assays, chemical assays, medical analyses, pharmaceutical , analyses Food, analyses diagnostic tests	Email, Cloud, text message and various messengers	(Alidans srl, 2015)
Live	Versatile	x	√	x	≥ 2.2	White paper to place as background	-Colour metering (Red, yellow, Blue) -Based on colour subtraction	x	(Folkstedt, 2013)

Stored Image	Versatile	x	√	x	≥ 4.1		-Colour palate -Colour recommendation - Suitable for field of Lifestyle e.g. clothes, accessories, hair colour	Email	(Torres, 2016)
Stored Image	Versatile	x	√	x	≥ 4.0.3	Colorcatch Nano (not mandatory)	-Colour metering -Colour palate -Colour recommendation - Suitable for field of interior design	Email	(Colorix, 2016)
Live	Limited versatility	x	√	√	≥ 4.0	x	-Colour metering (RGB, HSV, hex code) - Colour palate - Does not work properly	Social media	(Chromaflo Technologies, 2015)
Live, Stored Image	Versatile	√	√	x	≥ 4.0	x	- Colour metering (RGB) - TV colour fidelity checker -Requires calibration. -Also available in iOS platform	x	(Falcon Solutions Co, 2016)

N/m: Not mentioned. Some of the applications have in-app purchase facility (√/x). EXIF= Exchangeable Image File Format. The table excluded the following applications: used only for service or promotional service, without any other functionality; not available in English; mobile game etc.

In Auralisoft (Camera Colorimeter, 2016), two cameras are needed for colour calibration and calibrated colour viewing. Although it can work with a wide range of cameras, it takes 20 minutes to warm up after connecting both devices via Bluetooth. There are free applications with 'in-app purchase' facility to enable more features (vistech.projects, 2014). Many applications do not even require the image to be saved to analyse the colours and provide the result such as RGB value, colour histogram, hexadecimal code and the name of the colour.

C.2 TB Related Mobile Applications in the App-store

There are commercial and endorsed mobile applications for TB in the popular application stores, e.g. Google Play (Table C. 2, searched on 21-09-2017) and Apple app store.

Table C. 2: TB related mobile applications on the Android platform

User	Region	Aspect	Questionnaire	Intelligent Systems	Ref.
Department of Health	South Africa	Management; TB and HIV diagnostic data	√	X	(Interactive Health Solutions, 2016a)
Specific Users	Bangladesh	Management	√	X	(Interactive Health Solutions, 2017)
Mine community	South Africa	TB screening	√	X	(Interactive Health Solutions, 2016b)
Patients	Pakistan	Control TB and drug-resistance	√	X	(Interactive Health Solutions, 2016c)
Clinicians	Global	Decision on rapid diagnosis of TB and resistance	X	X	(Open Medicine Project, 2014)
Clinicians and Patients	Cambodia	Track lab test result	X	X	(Operation Asha, 2017)

When it comes to diagnosis, the applications are for screening purpose only (Interactive Health Solutions, 2016b; c, 2017). These applications store the screening data via the OpenMRS server. Either the user needs to insert the answers to a series of questions, or the lab test results have to be manually inserted by the user or clinician. The available applications can ensure the data portability (Table C. 2), and in some cases diagnostic decision (Open Medicine Project, 2014), however, they lack automation to produce a diagnostic result from the specimen. Thus,

there is a need for a system that does not require any additional hardware, e.g. a plate reader and can produce laboratory scale test results.

Appendix D

Instrumentations, Materials and Datasets

D.1 Choice of Platforms

This work has been simultaneously conducted using three computers with Windows 10. Two of them possess the same specification with Intel® Core™ i7-4470 CPU at 3.40GHz, 16GB and 64-bit OS; the other one possesses Intel® Core™ i7-7700HQ CPU at 2.80GHz.

This work explored a number of platforms to perform image processing and classification including MATLAB (MathWorks, 2017), ImageJ (Rasband, 2016), Weka (University of Waikato, 2017) and TensorFlow (Google Brain Team, 2018). Regarding software, this work is performed on the MATLAB[31] platform due to the convenience in research and convenience in analysing and presenting the research. Most of the relevant research work in the literature utilised MATLAB as well (Mutlu et al., 2017; Kim et al., 2017a; Solmaz et al., 2018). In the literature, among the other software, this research found widespread use of ImageJ (Yu et al., 2015; Tsai et al., 2013; Soares et al., 2017), OpenCV (Savardi, Ferrari and Signoroni, 2018; Oliveira, Pereira and Tavares, 2017; Lopez-Ruiz et al., 2014) and Weka (Oliveira, Pereira and Tavares, 2017; Tracey et al., 2011).

D.2 Sample Preparation of TB-test

D.2.1 Materials Preparation

The experiments on plasmonic ELISA were mainly conducted in Universiti Putra Malaysia. However, the TB patient sputum samples were provided by School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Malaysia through their University's hospital. The

---

[31] This research has been carried out for three years. Within this timeframe, all relevant software have been updated several times. In order to maintain a consistency, the result is presented in Windows 10 machines. In case of MATLAB, most of the result is obtained using MATLAB 2017a. In order to benefit from newly added advanced features in MATLAB platform for Deep learning such as data augmentation, visualisation of deep layers, this work utilises MATLAB 2018a.

fresh sputum sample was delivered to their lab, and smear microscope analysis was carried out before culture method. The ELISA analysis was carried out simultaneously in the same lab.

For the detection of CFP-10, a ten kDa secreted antigen from Mycobacterium tuberculosis, we first coated the ELISA plate with 100 μL of CFP-10 in carbonate buffer and then incubated for 1.5 h. Following the period, the plate was washed three times with PBS pH 7.6 and 0.05% Tween-20 (PBST) by tapping it against a clean paper towel. Now the plate was blocked with 370 μL of PBS containing BSA (PBSA) (1 mg/mL) for 1.5 h. All the antibodies and enzyme conjugates were diluted in diluent antibody containing PBST and 1% BSA. The plate was washed with PBST three times and kept the plate (invert) at 4°C for two hours. Now, 100 μL of the monoclonal anti-CFP-10 antibody as a primary antibody was added to the plate at 4°C for 1.5 h. After 1.5 h, the plate was washed with PBST for three times and the plate was added with 100 μL of biotinylated polyclonal secondary antibody and incubated for another 1.5 h at 4°C. The plate then washed three times, and 100 μL of the catalase-streptavidin conjugate (v/v 1:20) was pipetted into the plates and left for 1.5 h at 4°C. After the period, the wells in the plate were washed three times with PBST, two times with PBS, one time with deionized water and then dried. Now, 100 μL of hydrogen peroxide (in 1 mM MES, pH 6.5) buffer was pipetted into the wells. Immediately, 100 μL of gold ion solution freshly prepared in 1 mM MES buffer was added to the wells prepared in 1 mM MES buffer was added to the wells at room temperature. At this stage, the GNPs formation in the form of a coloured solution can be seen, and this can be read with a microplate reader at an absorbance of 550 nm. For the analysis of real samples, the sputum from positive and negative TB patients was diluted in 4% sodium hydroxide first and then proceeded to the same coating process as mentioned above.

The plasmonic ELISA links the colour of plasmonic nanoparticles to the presence or absence of the analyte (target protein). *Mycobacterium tuberculosis* ESAT-6-like protein esxB (CFP-10) was used as a target protein biomarker for the TB detection Plasmonic is accomplished by linking the growth of gold nanoparticles with the biocatalytic cycle of the enzyme label. The protocol adapts a conventional ELISA procedure with catalase-labelled antibodies. The enzyme consumes hydrogen peroxide (H_2O_2), and then gold (III) ions are added to generate gold nanoparticles. The concentration of hydrogen peroxide dictates the state of aggregation of gold nanoparticles. This allows for the naked-eye detection of analytes by observing the generation of blue- or red-coloured gold nanoparticle solution.

D.2.2 Data Collection

The dataset, generated as stated above and in Section 5.4.1, contains 252 images; 106 of them, captured with an iPhone 8-megapixel camera without mobile phone holder, were initially

considered for our preliminary study (Abuhassan et al., 2017). The dataset also contains few images taken from the video frame (Section 5.3.1), from a phase where it can be considered as an end point assay.

Blurry images, images with inadequate camera exposure, observations intended for biosensor optimisation and the initial experiments where the colour widely varied from the final representative colours were removed. Finally, 27 images were selected from 22 independent observations. Among these images, 13, 3 and 2 images were captured by Samsung Galaxy J5 Prime (13-MP), iPhone 7 plus (12-MP) and iPhone 6 (8-MP) cameras respectively. The remaining images were captured with an iPhone 4S (8-MP camera). A mobile phone holder (NJS Telescopic Music Record Mobile Phone iPad iPhone Stand Inc G Clamp Mount 68G) was used while capturing the images. However, the images varied regarding 35mm focal length, which means the physical distance between camera to sample were same, varying the effective camera exposure and size of the wells.

D.3 Dataset and Control Colours of the LFA

D.3.1 Preparation of Dataset

It was discussed in Chapter 6 (Section 6.2.1), that some of the test strips were allowed to dry on tissue paper, while some of them were purposefully not allowed to dry the residue on tissue paper. The impact of such preference is shown in Fig. D. 1 .
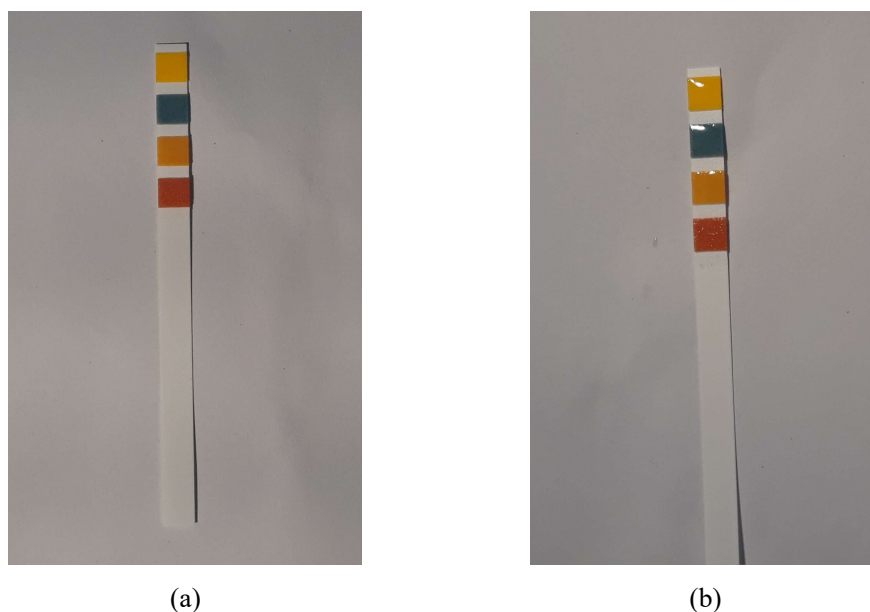

(a)   (b)

Fig. D. 1: Universal pH indicator paper showing pH level 9. (a) Sample dried on tissue paper and (b) sample not dried

When the sample in Fig. D. 1(a) were allowed to be dried on a tissue paper, there is no microfluid on the surface of the pH test strip. On the other hand, the microfluid in Fig. D. 1(b) suffered from slight light reflection problem. The purpose of the undried test strips was to create a robust dataset which contains such a noise.

D.3.2 Control Colours

The colour chart provided by the manufacturers are shown in Fig. D. 2. This chart can be used as the reference colours for the naked-eye measurement of the pH levels using the universal pH indicator paper. This is one of the conventional methods.

Fig. D. 2: Colour chart of the pH indicator paper

Alternatively, there are analysers for colour measurements. An example is shown for urinalysis (Fig. D. 3).
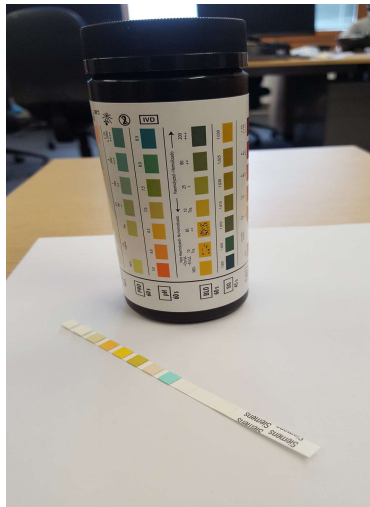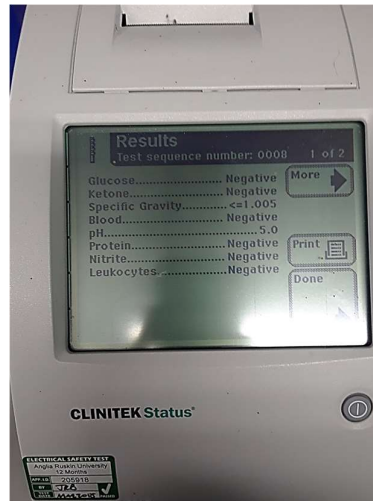
(a)   (b)

Fig. D. 3: Colour chart and analyser for urinalysis. (a) Colour chart provided by the manufacturer, (b) CLINITEK Status® + Analyzer

The colour chart can be thought of as the ground truth colours or the controls.

Appendix E

Major Steps of the Algorithm

E.1 Reaction Phase and Time-dependent Approach

Out first approach was to analyse the dynamic changes in the images of an end-point assay. This is part of the first specific objective. Major steps of the algorithms is given below.

Algorithmic Steps 1: Image Segmentation for Reaction phase and time dependent approach

---

Input: 1. I= initial image at start point;

    2. F= final image at end point

Output: Mask

Steps:

5.1 READ Input 1

     READ Input 2

5.2 RESIZE INPUT 1 AND INPUT 2 WITH SAME SCALING FACTOR

\\ L_P distance; $1 \leq p \leq \infty$

5.3 Z= |Input 1 - Input 2|

\\ Z in green and blue channel

5.4 $Z_g = I(:,:,2)$

     $Z_b = I(:,:,3)$

\\ Make the wells more prominent

5.5 ADD Z_g and Z_b

\\ Smooth the image to clear the empty

5.6 GAUSSIAN 2D FILTERING

5.7 GB → Grey

5.8 Grey → Binary

MORPHOLOGICAL OPERATION

E.2 Reaction Phase and Time-independent Approach

The system architecture of the end point assay considering the final static stage of the reaction was described in the Chapter 3. After experimentation and trial and error the framework was finalised. Before describing the key components of the algorithm, the workflow of the finalised system is outlined in Fig. E. 1.
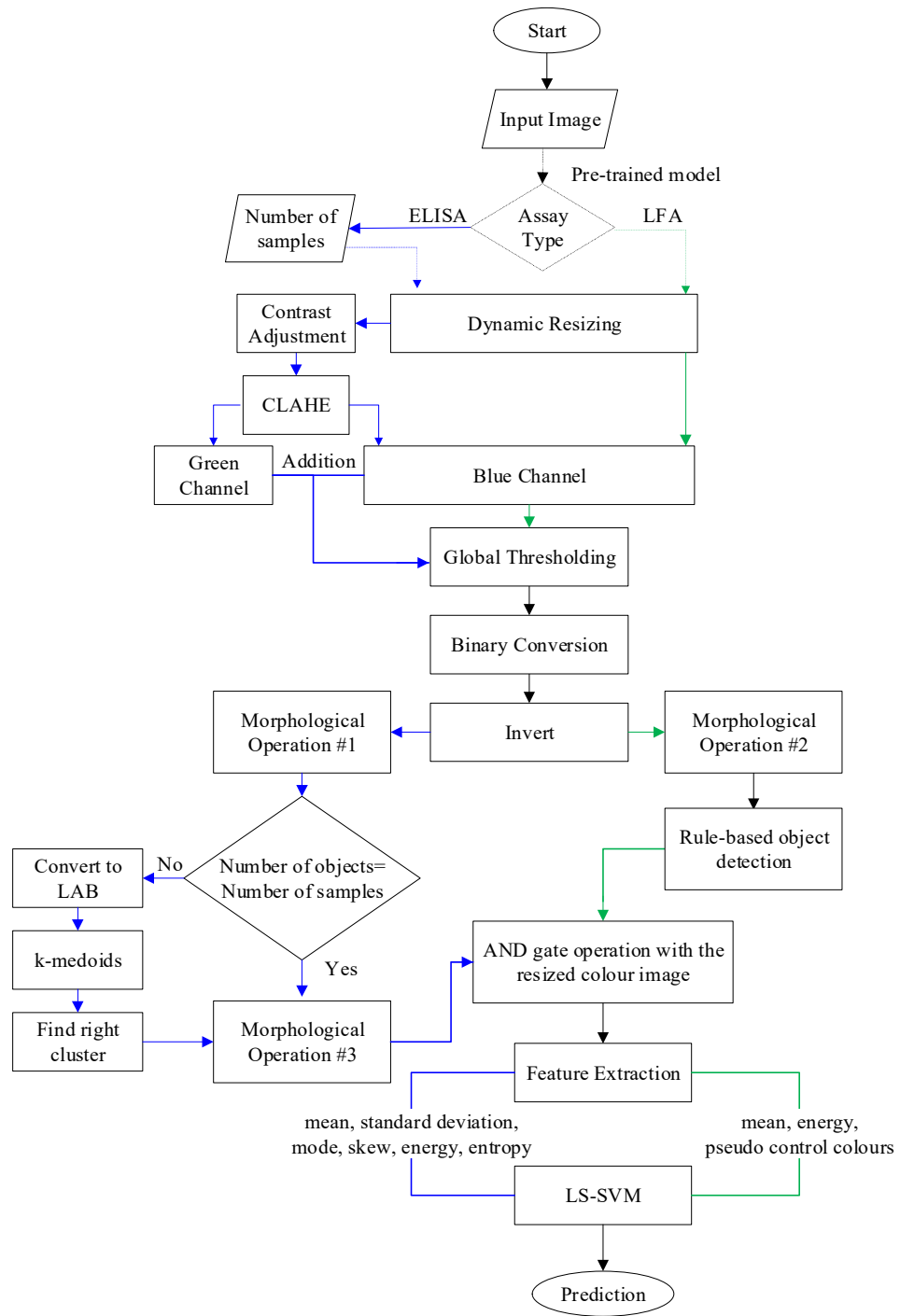
Fig. E. 1: Flowchart of the overall system

E.2.1 Dynamic Scaling

The dynamic scaling on Android platform is conducted by Android Developers (2018). In this work, the scaling mechanism is provided below.

Algorithmic Steps 2: Dynamic Scaling

Input: I= final image at end point;

Output: Scaled image

Steps:

2.1 READ I

\\ Dimension= length x width

2.2 ESTIMATE the dimension of I

    L=length of Input 1

    W=Width of Input 1

    Ch= number of colour channels

5.9 Dynamically SCALE the image

    Scale=300/ maximum (L,W)

\\ B< Input

2.3 B= Resize (Input, Scale)

E.2.2 Image processing of Case Study 1

The major steps to process TB-test images are provided below. The steps include pre-processing, segmentation and post-processing methods used in Chapter 5.

Algorithmic Steps 3: Image processing for TB-test

Input: 1. I= final image at end point;

      2. NS= number of samples;

Output: Segmented wells

Steps:

3.1 READ I

3.2 Algorithmic Steps 2

3.3 CONTRAST ENHANCEMENT of B

3.4 SMOOTHENING

3.5 ENHANCEMENT of smoothened image

3.6 x_r = 8-bit unsigned integer array of filtered image in red channel= B(:, :, 1)

    x_g = 8-bit unsigned integer array of filtered image in green channel= B(:, :, 2)

    x_b = 8-bit unsigned integer array of filtered image in blue channel= B(:, :, 3)

3.7 XBG = x_g + x_b

\\ Binary conversion by Global Thresholding

3.8 BW= Binary image of XBG by Otsu (1979)

\\ Inverse the binary image

3.9 I= BW'

3.10   Morphological Operation #2 (Appendix E3)

3.11   **IF** Number of objects = NS

3.12   Morphological Operation #3 (Appendix E3)

3.13 I_c= 8-bit unsigned integer array from logical array containing correct number of objects

\\ Segmented image

3.14 $Y = \begin{cases} x_r \times I_c \\ x_g \times I_c \\ x_b \times I_c \end{cases}$

3.15   **ELSE** Conversion of the processed image into the CIELAB colour space

3.16   Use colours in the ab space to measure the Euclidean distance for clustering. Select k seeds by implementing the k-means++ algorithm (Arthur & Vassilvitskii, 2007) for cluster medoid initialization. k=2.

3.17   Dynamically repeat step to avoid local minima

\\ k images

3.18   For clusters 1 to k, separate the objects using the index clustering

3.19   Convert k images to binary images

3.20   FILTERING

\\ Find the right cluster

3.21   **WHILE cl**uster ≠ empty

3.22   **IF** Number of objects = NS,

3.23   Repeat Step 3.12 to 3.14

3.24   **ELSE** Calculate the distance among the objects in the cluster

3.25   **IF** distribution of distance ≈ threshold

3.26   Repeat Step 3.12 to 3.14

3.27   **ELSE** ESTIMATE average sizes of the objects in the cluster

3.28   **IF** object to object size ≈ average size

3.29   Repeat Step 3.12 to 3.14

3.30   **ELSE** COUNT objects in Cluster 1 and Cluster 2

3.31   **IF** objects in Cluster 1 + Cluster 2=NS

3.32   Repeat Step 3.12 to 3.14

As described in Section 5.5 (Chapter 5), the image processing started with the pre-processing steps. To summarise the outcome of the image processing framework, once again for the convenience of the readers, a brief discussion is provided in this section.

After the pre-processing stage, the enhanced image was taken into the green and blue channel. The global thresholding by Otsu's method was applied on the collective image of the green and blue channel. The resultant image was transformed into a binary form to reverse the 0 and 1 values. Due to extensive pre-processing and colour channel manipulation before global thresholding, most (~78%) of the images can be corrected segmented within this stage as shown in Example A (Fig. E. 2).
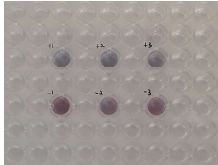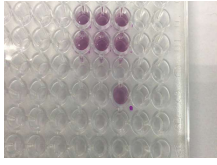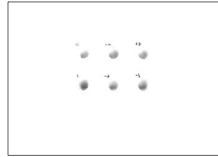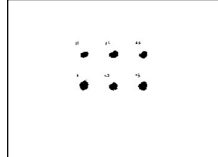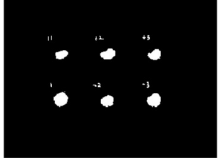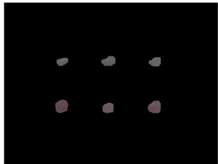

Fig. E. 2: Key output steps of the Hybrid algorithm

In order to present a robust and reliable image processing scheme that provides more autonomy to the user, k-medoids clustering followed by rules-based best cluster selection were integrated to the algorithm to separate the colours of background and foreground using only two clusters as suggested by the silhouette method. A demonstration is provided using Example B (Fig. E. 2). The observations from Section 5.4.2 governed the image processing framework to select

appropriate clustering techniques. The output image for both examples (A and B) were attained by post-processing using morphological operation

E.2.3 Feature Extraction Process of Case Study 1

After separating the ROI, the features can be extracted as the following.

Algorithmic Steps 4: Feature Extraction of TB-test

Input: Segmented Image

Output: Extracted Features

Steps:

4.1 Convert the data into the desired colour spaces

> RGB → LAB

4.2 Use a counter to record the features

\\ For each colour channel record the colour moments. The features are extracted for individual objects

4.3 Calculate mean of the elements of 4.1 along the first array dimension whose size is not equal to 1.

$$\overline{g} = \sum_{g=0}^{L-1} g\, P(g) = \sum_{r}\sum_{c} \frac{I(r,c)}{M}$$

4.4 Calculate the sample standard variation using Bessel's correction

$$\sigma_g = \sqrt{\frac{1}{1-n} \sum_{g=0}^{L-1} (g - \overline{g})^2 P(g)}$$

4.5 Find the most frequently occurring value in the colour histogram

$mode$

4.6 Calculate skewness using 1st skewness co-efficient i.e. Pearson mode skewness

$$SKEW = \frac{\overline{g} - mode}{\sigma_g}$$

4.7 Calculate the amplitude using the mean in Step 4.3

$$E = \sum_{g=0}^{L-1} [P(g)]^2$$

4.8 Calculate the entropy, which is a statistical measure of randomness that can be used to characterize the texture of the input image

$$ENTROPY = -\sum_{g=0}^{L-1} P(g)\, log_2\, [P(g)]$$

E.2.4 Image Processing and Feature Extraction Algorithm of Case Study 2

The image processing algorithm followed by the feature extraction algorithm performed for the Case Study 2 is provided below. The following algorithm is presented with the help of the algorithmic steps mentioned above to highlight the similarities for both of the case studies.

Algorithmic Steps 5: Image processing and Feature Extraction Algorithm for Case Study 2

Input: 1. Image of pH test strips in RGB;

Output: Extracted features of the segmented ROI in LAB

Steps:

\\ Image Processing

5.1 READ Input 1

5.2 Algorithmic Steps 2

5.3 Perform Step 3.6

\\ Eliminate Red and Green channel

5.4 x_b = B(:, :, 3)

\\ Take only Blue channel of filtered image

\\ Binary conversion by Global Thresholding (Step 3.8)

5.5 BW= Binary image of x_b by Otsu (1979)

5.6 Perform Step 3.9

5.7 Morphological Operation #1 (Appendix E3)

    C = Detected objects

\\ Noise filtering

5.8 P$_c$ = Pixels in C

      IF P$_c$ < 50

      ELIMINATE P$_c$

5.9 COUNT C

5.10 **IF** C= number of colour blocks in sample

5.11 Perform Step 3.13 to 3.14

\\ Feature Extraction

5.12 Perform Step 4.1 to 4.3

5.13 Perform Step 4.7

\\ Pseudo-control Colour

5.14 $\Delta E^*_{b_p} = \sqrt{\{(mean\ L^*_{b_p} - mean\ L^*_{b_{p-1}})^2 + (mean\ a^*_{b_p} - mean\ a^*_{b_{p-1}})^2 + (mean\ b^*_{b_p} - mean\ b^*_{b_{p-1}})^2\}}$

5.15   REARRANGE the matrix

E.3 Morphological Operation

Key post-processing techniques utilised in this work are listed in below. These operations are conducted on the binary image.

Area opening

This technique is used to remove all connected objects lower than a threshold pixel value. In several cases, this operation is performed by specifying the value for the connectivity as well.

Skeleton

While performing the morphological operation, with $n = \infty$, removes pixels on the boundaries of objects but does not allow objects to break apart. The pixels remaining convert the image into a skeleton of the image. This option preserves the Euler number. Then, additionally, the endpoints of the skeleton are outlined.

Dilatation

In this case, dilation is mainly translation-invariant operator, mathematically equivalent to Minkowski addition. This work utilises the dilation presented in Gonzalez, Woods and Eddins (2009).

Morphological reconstruction

As part of morphological reconstruction sometimes it is required to fill the holes in the image, where a hole is a set of background pixels that cannot be reached by filling in the background from the edge of the image (Soille, 2004).

Sometimes, it is also required to suppress the light structures connected to image border as part of the reconstruction. The input of this step is taken as the mask image. The marker image is considered to be zero everywhere except along the border, where it equals the mask image. The connectivity is also specified.

Erosion

Erosion is one of the major elements of morphology. This work mainly uses erosion by specifying the structural element neighbourhood. This work utilises the erosion as presented in Gonzalez, Woods and Eddins (2009).

Boundary tracing

Using Moore-Neighbor tracing algorithm, modified by Jacob's stopping criteria, the boundary of the objects are traced by searching for parent and child, i.e. object boundaries (Gonzalez,

Woods and Eddins, 2009). The outermost objects are called parents and objects entirely enclosed by the parents are called children.

The morphological operations of Fig. E. 1 is sequentially listed below. The Morphological operation #1 was part of the Case Study 2.

Algorithmic Steps 6: Morphological operation #1

Input: Inverted binary image of TB-test after Global thresholding

Output: Segmented, separated and clean ROI

Steps:

1. Perform AREA OPENING
2. Take the END POINTS of the SKELETON of step 1.
3. DILATION
4. MORPHOLOGICAL RECONSTRUCTION
5. EROSION
6. EROSION of the image of step 5
7. Perform AREA OPENING
8. BOUNDARY TRACING

The Morphological operation #2 was part of the Case Study 1.

Algorithmic Steps 7: Morphological operation #2

Input: Inverted binary image of LFA after Global thresholding

Output: Segmented, separated and clean ROI

Steps:

1. SEARCH Euler Number Property =1
2. FILTER input using step 1 \\ Output= Regions of original image without holes
3. Perform AREA OPENING

The Morphological operation #3 was also part of the Case Study 2.

Algorithmic Steps 8: Morphological operation #3

Input: Right cluster of TB-test

Output: Segmented, separated and clean ROI

Steps:

1. RGB → GRAY
2. Global thresholding
3. CONVERT to binary image
4. Perform AREA OPENING

---

5. Take the END POINTS of the SKELETON of step 4.
6. DILATION
7. MORPHOLOGICAL RECONSTRUCTION
8. EROSION
9. EROSION of the image of step 8
10. Perform AREA OPENING

E.4 Computation Time

The key definitions related to computational time is provided below.

Elapse time

It is the amount of time that passes from the start of an event to its finish.

Epoch

The presentation of the set of training, which could be input and/or target vectors, to a network and the calculation of new weights and biases. In the case of batch training in deep learning, all of the training samples pass through the learning algorithm simultaneously in one epoch before weights are updated.

Epochs can stop the training in case the solution of the training algorithm does not converge, to prevent infinitely running the training.

Appendix F

Statistical Analysis Method

The output of an effectively designed system should be reproducible. Despite the use of random number generator, shuffled as well as uniformly distributed random numbers, there can be minor variation in the output of the system due to random initial seed selection and random partitioning during cross-validation. Therefore, each of the performance or execution should be repeated several times and statistically validated. In this thesis, the statistical analysis is conducted in SPSS.

F.1 T-test

In this research, we compare the performance of the best performing classifiers by using Student's t-test (Walpole et al., 2011). Due to the bi-directional nature of the assessment, this work considers two-tailed t-test. The analysis method is depicted from Lwin (2015) and Hatem Al-jamal (2017). The z-tests are more commonly used in the case of a higher number of observations (samples>30). However, for the most computer packages t-test results are equivalent (Scott and Mazhindu, 2014). On this occasion, it should also be noted that the number of observations refers to the number of execution of the algorithm.

F.1.1 Consideration of Bias

This work carefully considers the issue of a continuous variable, normal distribution and equal variance. In order to perform a t-test, care should be provided to prevent bias, which can be induced through outliers or violation of assumptions.

Outliers

This can be thought of as an anomaly in the result. Ideally, the output of the classifiers used in this study, obtained in MATLAB, should not contain outliers. Otherwise, the outliers can be identified using the descriptive statistics option in SPSS.

Assumptions

   i)   Independence

The algorithms are not dependent on each other; hence independence can be assumed.

ii) Normality

The t-test assumes a normal distribution of the samples. The assumptions for t-test include: the parent population has to be normally distributed, and the samples possess the t-distribution. If the same size is 'large enough', then the central limit theorem can be considered, which adopt normality regardless of the shape of the data (Lumley et al., 2002). It is a common practice to consider 30 observations as the 'large enough' sample size to exercise the central limit theorem and to assume the normality (Field, 2013). In this thesis, both of the case studies include 30 observations for the statistical analysis to be performed. Therefore, the central limit theorem will be applicable, and normality test is not necessary.

iii) Homogeneity of variance

For the same set of input, if a system always provides the same output, then the method is deterministic. For the same set of input, if a system provides a different result at different execution, then the algorithm is non-deterministic. Due to initial seed selection, splitting and randomisation, an algorithm may show non-determinism, e.g. Random Forrest.

While comparing the accuracy of different algorithms, this work assumes the variance in one sample, i.e. the accuracy of the algorithm is not larger than twice of the size of the other. This work utilises Levene's test (Levene, 1960) to determine the homogeneity of variance. If the list of top performing classifiers includes both deterministic and non-deterministic algorithms, then clearly this assumption is going to be violated. In such cases, the violation of the assumption for equal variance needs to be compensated by utilising Welch-Satterthwaite method (Welch, 1947; Satterthwaite, 1946).