NON-INVASIVE DENGUE SCREENING METHOD VIA OPTICAL SPECTROSCOPY: A MULTIVARIATE INVESTIGATION

ABDUL HALIM POH YUEN WU

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

NON-INVASIVE DENGUE SCREENING METHOD VIA OPTICAL SPECTROSCOPY: A MULTIVARIATE INVESTIGATION

ABDUL HALIM POH YUEN WU

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

ABSTRACT

More feared than understood, dengue fever often evokes emotions of morbidity in the public space, arguably due to a positive response to media campaigns. A symphony of clinical analysis of blood results (not necessarily dependent only on the NS1 detection), symptoms, and day of fever onset, among others, contributes to a successful diagnosis, often spanning many man-hours (or days) combined with costly overhead expenditure. For dengue-related management, the Government of Malaysia spent close to US$73.5 million in 2010, which was 0.03% of the country's GDP during that time.

With the aim of reduction of the time and monetary costs involved in this issue, a potentially non-invasive method was scrutinized in painstaking detail to circumvent at least a few hurdles in dengue patient care, especially in diagnostics. This method is namely Diffuse Optical Reflectance Skin Spectroscopy. With a probe and a light source combined with spectrometers, the forearm of a suspect patient in the University of Malaya Medical Centre (UMMC) is scanned. The spectroscopy data is then collected and saved via a software tailored to consolidate all information regarding the patient. The patient is then assigned to a diagnosis by UMMC physicians. Three groups were later formulated for the classification, namely confirmed dengue, probable dengue, and control patients.

Based on multivariate analysis on the spectroscopy data of 230 patients, we have come to at least two major findings. First, the modelling produced by the statistical algorithms predicted the accuracy of detection up to 98.65% on discriminating all three groups. Second, several feasible algorithmic models for classifying dengue patients was synthesized, ranging from sensitivities and specificities of 76.67%-89.29% and 94.85-100% respectively. From these findings, further clinical trials on non-invasive dengue screening are recommended.

ABSTRAK

Demam denggi merupakan suatu penyakit yang digeruni, walaupun mungkin kurang difahami dalam masyarakat umum. Ini adalah mungkin suatu sumbangan positif dari pihak media massa. Denggi dikesan melalui pertimbangan antara kajian darah di makmal, mengesan tanda-tanda gejala denggi dan tempoh bermulanya demam tersebut. Pada tahun 2010, Kerajaan Malaysia telah membelanjakan sejumlah US\$73.5 juta yang juga setara dengan 0.03% KDNK negara.

Berpandukan usaha untuk mengatasi kos masa dan wang dalam isu ini, suatu kaedah yang tidak menyakitkan, iaitu tidak memerlukan pengambilan darah dari pesakit telah diteliti dengan mendalam, terutamanya dalam ehwal mengesan denggi. Ianya dinamakan kaedah spektroskopi kulit optik. Menggunakan sebuah pengesan yang digabungkan dengan sumber cahaya dan spektrometer, lengan pesakit yang sedang diuruskan di Pusat Perubatan Universiti Malaya (PPUM) akan diimbas. Maklumat dari imbasan tersebut akan disimpan. Pesakit tersebut pula akan dikelaskan mengikut diagnosa oleh pakar klinikal di PPUM, sama ada dalam kumpulan yang disahkan denggi, disyaki denggi atau sebagai kawalan.

Berdasarkan analisa *multivariate* terhadap 230 pesakit yang diimbas, terdapat dua hasil dapatan yang amat memberangsangkan. Pertama adalah dari model statistik yang dijana dengan perisian yang digunakan berjaya meramalkan 90.43% kategori pesakit dengan tepat. Kedua, menggunakan kaedah jujukan statistik yang memilih pembolehubah secara berperingkat, pengasingan yang jelas dapat dihasilkan bagi tiga kumpulan tersebut. Berdasarkan skor yang terhasil dengan model yang dijanakan, kepekaan dan kekhususan antara 76.67%-89.29% and 94.85-100% dapat dihasilkan. Dengan hasil penemuan ini, ujian klinikal untuk kegunaan yang lebih meluas dengan kaedah ini perlu dipanjangkan dalam masa akan datang.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ACA	:	Antigen Capture Anti-DENV
AGE	:	Acute Gastroenteritis
ALT	:	Alanine Aminotransferase
APTT	:	Activated Partial Thromboplastin Time
AST	:	Aspartate Aminotransferase
BCC	:	Basal Cell Carcinoma
CAP	:	Community Acquired Pneumonia
CRP	:	C-Reactive Protein
CT	:	Computed Tomography
DA	:	Discriminant Analysis
DENV	:	Dengue Virus
DF	:	Dengue Fever
DHF	:	Dengue Haemorrhagic Fever
DRS	:	Diffuse Reflectance Spectroscopy
DSS	:	Dengue Shock Syndrome
DSS	:	Dengue Shock Syndrome
DWOS	:	Dengue Without Warning Signs
DWWS	:	Dengue With Warning Signs
ELISA	:	Enzyme-Linked Immunosorbent Assay
ESR	:	Erythrocyte Sedimentation Rate
FBC	:	Full Blood Count
GUI	:	Graphical User Interface
HCO_3	:	Bicarbonate Acid
HLH	:	Hemophagocytic Lymphohistiocytosis
ID	:	Infectious Diseases
IgG	:	Immunoglobulin G
IgM	:	Immunoglobulin M
LFT	:	Liver Function Test
LSPR	:	Localized Surface Plasmon Resonance
MAC-ELISA	:	M Antibody-Capture Enzyme-Linked Immunosorbent Assay
NIR	:	Near-Infrared
NIRS	:	Near-Infrared Spectroscopy
NS1	:	Non-Structural Protein 1
OLEDB	:	Object-Linked Environment Database
PCA	:	Principal Component Analysis
POC	:	Point Of Care
PT INR	:	Prothrombin Time And International Normalized Ratio
PTFE	:	Polytetrafluoroethylene
RDT	:	Rapid Diagnostic Test
RNA	:	Ribonucleic Acid
RT-LAMP	:	Reverse Transcription-Loop-Mediated Isothermal Amplification

RT-PCR	:	Reverse-Transcriptase Polymerase Chain Reaction
SC	:	Stratum Corneum
SD	:	Severe Dengue
SNR	:	Signal-To-Noise
SpO$_2$	:	Peripheral Capillary Oxygen Saturation
SPR	:	Surface Plasmon Resonance
SPT	:	Skin Phototype
SVW	:	Significant Variability Wavelengths
UFEME	:	Urine Full And Microscopic Examination
UMMC	:	University Of Malaya Medical Center
URTI	:	Upper Respiratory Tract Infection
UTI	:	Urinary Tract Infections
UV	:	Ultraviolet
VB.NET	:	Visual Basic .Net
VIS	:	Visible
VS2013	:	Visual Studio 2013
WBC	:	White Blood Count

CHAPTER 1: GENESIS, GENUS AND GEOGRAPHY

This chapter explores the motivation behind studying a possibly non-invasive method to screen dengue. This consists of a brief dengue history and epidemiology, the shortcomings of current diagnostic methods be it conventional or experimental, followed by the objectives and outlines.

1.1 Dengue, Delirium, and Demise

The words "Break bone fever" has been used to describe an emerging disease as early as in the year 1780 in perhaps an idyllic city in Philadelphia by Benjamin Rush, who happened to be one of the founding fathers of the United States of America. Later to be termed "Dengue fever", the name of the disease has no clear etymological origin. It has been speculated, among others, to be from the word "*Ka-dinga-pepo*", a Swahili term which means seizures attributed to the seeming possession suffered the victims, likely due to the debilitating pain. Others suggested that it came from the corruption of the English word *dandy* or Spanish *denguero*, or *dunga* in Cuba where the first two terms are similar in its context describing the subdued and stiff appearance of the victims when they walk (Barnett, 2017; Dick et al., 2012).

The deadly reputation of dengue is credited to its severe form, which is usually due to subsequent infections after the first, especially with different serotypes. There are four dengue serotypes, namely DENV-1 to DENV-4. Severe forms of dengue causes manifestations of plasma leakage with symptoms such as myalgia and arthralgia, with signs taking in the form of ascites or pleural effusion. This subsequently causes hypovolemic shock, causing organ failure, and ultimately, death (Jaenisch et al., 2016).

Albeit its acknowledgment by archaic Western literature, dengue remains to be a predominantly a South Asian disease, which is represented by India, Southeast Asia and

the Americas such as Brazil and Panama, with cases dating as early as the year 1635. This is due to the year-round availability of the main vector, *Aedes Aegypti*, a mosquito species attributed to most dengue infections (Dick et al., 2012).

The global burden of this disease has been acknowledged as a threat globally with climate change and cross-country travels changing the landscape of the epidemiology. The prevalent presence of these vectors are attributed to the temperate and tropical climate, ideal for a vicious cycle of perpetual breeding. Annually, 390 million people are infected globally (Jaenisch et al., 2016). In Malaysia, more than 100,000 dengue cases have been reported in 2016, with more than 200 dengue-related deaths recorded in the same year (Ahmad et al., 2018). There has been no known therapy for dengue-infected patients, and dengue vaccines have reportedly been mired in controversy (Aguiar, Stollenwerk, & Halstead, 2016; Flasche et al., 2016).

1.2 Diagnostics, Detection, and Dissonance

It takes many steps and procedures, and hence, time, to diagnose dengue. One of the vehicles of patient care begins with diagnostics for differentiating febrile patients suspected of dengue. Normally, this depends on the use of the MAC-ELISA method to detect the antigens resulting from the infection, namely the NS1, IgM or IgG antigens from the blood serum. While the method has a high sensitivity and specificity, it takes numerous resources including test reagents, trained personnel, and normally results can be obtained after several hours. However, as blood samples are run by the batch in healthcare centers, this often takes more than 24 hours. Golden-standard laboratory diagnostics such as the Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR) methods, which examines the RNA of the dengue takes longer time, with nontrivial setup and high maintenance costs (Lai et al., 2007).

Also, a clinical confirmation from the serological analysis of the patients would still be required, blood tests positive notwithstanding. This is performed by skilled physicians to evaluate clinical consistency for whittling out false positives due to the cross-reactivity on other antigen types via the MAC-ELISA methods. Combined with the serology time frame, and other human factors, a complete clinical diagnosis can take several days (Guzmán & Kourí, 2004).

The next logical step for many improvement-seeking medical technologists would be to pursue point-of-care (POC) and rapid diagnostic test (RDT) kits, which have made significant entry into scientific literature. From chemical immunoassays to optical detection via binding agents on various assays, most of these tests still require the blood sera, which necessitates procedures of blood drawing or phlebotomy, labelling and centrifugation. It can be deemed that however simple these schemes are, its disruptive element to the status quo (in this context, the MAC-ELISA method) is unsubstantial. Various reviews have been made to evaluate the economic impact of RDT schemes, which posed many challenges for replacement of already existing methods (Hunsperger, dos Santos, Vu, Yoksan, & Deubel, 2017; Muller, Depelsenaire, & Young, 2017).

It is worthy to note that in all cases of dengue diagnostic methods, whether standard or experimental, generate medical waste due to consumables used for serology or chemical assays. Specifically, Malaysia was projected to produce 33,000 tonnes of medical refuse in the year 2020 due to the increasing tourist traffic in the country (Ambali, Bakar, & Merican, 2013). Though no statistics exist to apportion the contribution of medical waste due to dengue diagnostics, it can be safely concluded that it contributes to the to-be-incinerated and later landfilled pile, at least in Malaysia (Lau, 2004).

Inevitably, the question to whether it is even remotely plausible to explore non-invasive avenues of detecting dengue comes into play in this research. A dissonant

perspective, which motivates establishing linkage between dengue with dermatological features detectable through optical means, tilts the disruptive scale. Such a method, would include skin spectroscopy as part of the large family of non-invasive techniques. Spectroscopic methods have reportedly shown capability for differentiating many diseases, such as diabetes, malignant skin such as basal cell carcinoma (BCC), and many others. These detection methods take place in numerous ways, including fluorescence, absorbance, diffuse reflectance, and Raman spectroscopy. The mechanism which allow spectral profiling from the skin practically gave birth to a distinctive branch of medical optics – namely skin or tissue optics. This discipline is a developing field which emerged arguably and partly when Sekelj (1954) established that saturated oxygenation of the blood can be optically monitored.

Since then, the field of tissue or skin optics has matured considerably. Modern instruments monitor numerous illnesses centralized in the non-invasive oxygenation measurements. This includes oximetry-based monitoring for cerebral, muscle, pulse and other aspects of human vitals (Hirasawa et al., 2016; Olesen et al., 2018). Diabetics were diagnosed by examining various fluorescence spectrometry peaks, attributed to the bonding of sugar molecules to the body fat (Ghazaryan, Hu, Chen, Tan, & Dong, 2012). Diagnostics of basal cell carcinoma, or other variants cancerous skin were credited to spectral patterns collectively exhibited by fluorophores in the skin, such as collagen, elastin, coenzymes, flavins, and proteins' cross-links. Analysis of spectral pattern exhibited by these skin components consequently discriminates healthy and cancerous subjects (Bodén et al., 2013).

However, non-invasive dengue diagnostics have never been considered, due to the adherence to rigorous disciplines of serological and clinical diagnosis. This has been the case despite reports on distinctive dermatological and histological features of dengue

patients have been covered in previous works, albeit with no spectroscopic profiling attempted for diagnostic value. For dengue patients, there has been several in-vitro and in-vivo measurements, reporting on various aspects of the patient, mostly represented by serological approach, such as the use of Raman spectroscopy for differentiating dengue and non-dengue infected blood sera (Shamaraz Firdous et al., 2012). Due to plasma leakage, the dilated capillaries of febrile dengue patients, which is prevalent through the body has been reported to reduce muscle oxygenation, a parameter measurable by near-infrared spectroscopy (Soller et al., 2014). It can be reasonably hypothesized, that due to the combinations of the physiological states, the tissue morphology of dengue patients may exhibit spectroscopic patterns that are statistically classifiable.

1.3 Objectives, Outlines, and Organization

The objectives of this research is to (a) establish a database of skin spectroscopy profiles from febrile human patients within the ultraviolet, visible, and near-infrared bandwidths of the light source, or between 200-2500 nm, (b) clinically confirm the recruitment of the patients, hence separating confirmed dengue cases and non-dengue diseases, and (c) statistically establish the feasibility of separating the classes of dengue and non-dengue illnesses, thus establishing a potentially ground-breaking dengue screening method, if rendered with high levels of confidence.

The thesis, as presented here, is outlined following this chapter. **Chapter 2** elucidates the literature on the epidemiology of dengue, diagnostic standards, and published experimental methods of rapid dengue diagnostics, its limitations and advantages, followed by the potentials of skin spectroscopy for dengue diagnostics.

Chapter 3 reveals an in-depth detailing of the methods used, including the spectroscopy setup, patient recruiting method, inclusion and exclusion criteria, data management, clinical diagnostics of the patients and analysis.

Chapter 4, in turn, details the results and discussion from patient demographics, raw spectroscopic data patterns, multivariate statistical analysis of the data, and normalization techniques for improving accuracies.

Chapter 5 summarizes Chapters 3 and 4 for a succinct review on the practical outcomes of the research activities, and what can be hoped to be derived from it.

Chapter 6 is practically the conclusion of this thesis, with a story to be told.

In an attempt to break away from the didacticisms so often found in technical writings, this whole thesis loosely adheres to a rhetorical linguistic structure from its titles and headings. Also, chapter maps are provided in all chapters for navigation of the narratives presented. All artwork in this thesis, unless specified, are rendered using a combination of scanned hand-drawn figures, Microsoft Paint, Microsoft Powerpoint and GIMP 2.0.

CHAPTER 2: DENGUE VIRUS, DIAGNOSTICS AND DISRUPTION OF CONVENTIONS

This chapter topology is provided as shown in **Figure 2.1**.
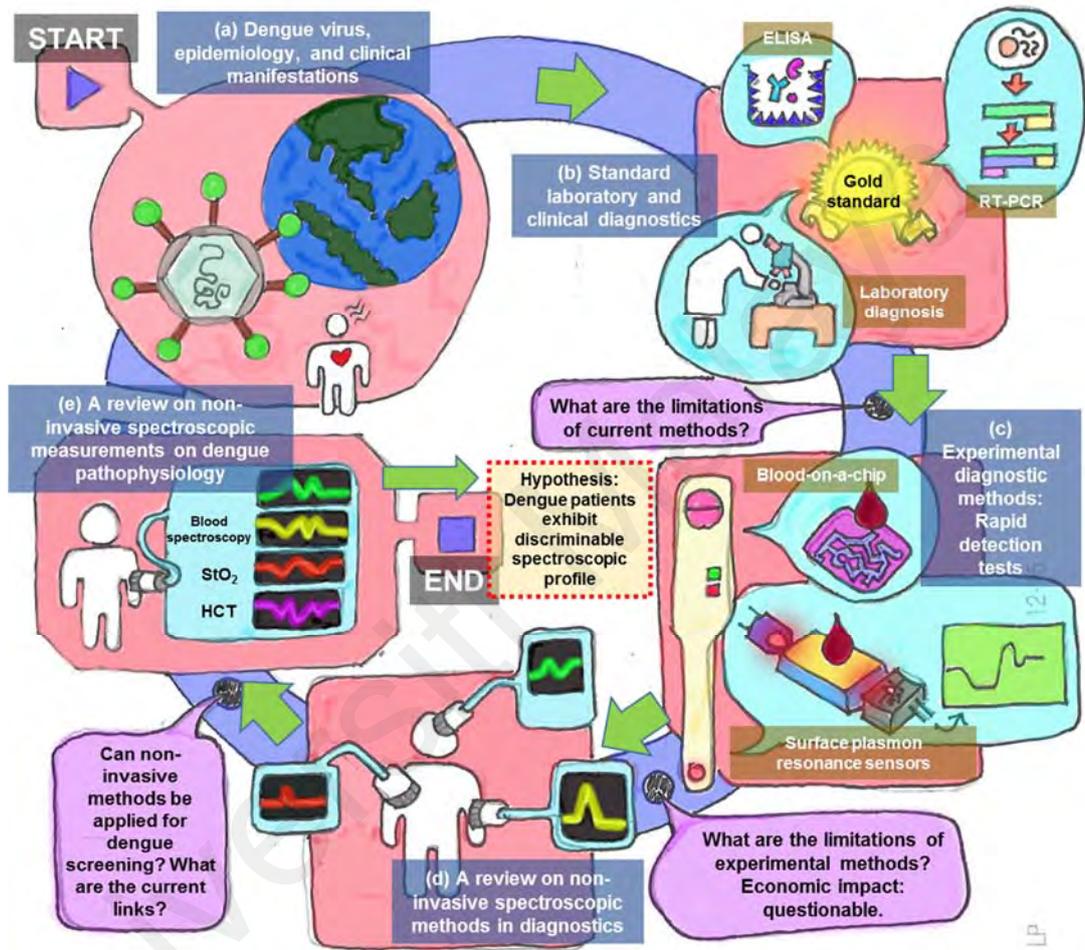
Figure 2.1: Chapter topology

As shown in **Figure 2.1(a)**, this chapter elaborates firstly on dengue as a virus, its vectors, and pathophysiology, which leads to its morbid outlook. Then, we discuss standard diagnostic methods as shown in **Figure 2.1(b)**, including its limitations. This is followed by attempts at breaking away from conventions, especially via rapid detection tests (RDTs) and point-of-care (POC) techniques as shown in **Figure 2.1(c)** and its

limitations. Next, we discuss reasons for consideration of non-invasive screening of dengue.

Figure 2.1(d) further reviews several non-invasive spectroscopic methods of diagnosing several diseases, and mechanisms of the diseases which allow detection. Finally, Figure 2.1(e) attempts to provide linkage between the pathophysiology of dengue patients with reflectance spectroscopy. This is done by highlighting several experimental and established publications, which indicate the nature of the patient in the febrile state, which may collectively contribute to in-vivo characteristics worthy of screening or diagnostic potential. The hypothesis is concluded based on this section.

2.1   The Dengue Trivium: Virus, Vectors and Valuations

Dengue is an arbovirus (or arthropod-borne virus) and a flavivirus, a genus under the Flaviviridae family (both flavi- prefixes derived from *flavus* in Latin, which means yellow), which bore similar characteristics to the Yellow Fever virus in terms of its size, symmetry and nucleic acid structure. It is a small single-stranded RNA virus, which consists of four separate serotypes, namely DEN-1 to DEN-4. Mature DENV particles are spherical with diameters of 50 nm, consisting of numerous three structural proteins, a host derived membrane bilayer, and a copy of the positive-sense, single-stranded RNA genome. Each serotype exhibits genetic diversity, where each show remarkable relativity in nucleotide sequence (Organization et al., 2009). The DEN-2 and 3 serotypes have been colloquially associated as the "Asian" genotypes, where both are attributed to the severe symptoms following secondary infections (Lanciotti, Lewis, Gubler, & Trent, 1994; Leitmeyer et al., 1999; Messer, Gubler, Harris, Sivananthan, & De Silva, 2003). A fifth serotype (DEN-5) was also discovered, to the chagrin of the health community (Mustafa, Rasotgi, Jain, & Gupta, 2015).

Mosquitoes are the main vectors of dengue infections. Besides the primary vector *Aedes Aegypti*, others also include *Aedes albopictus*, *Aedes polynesiensis* and certain species of the *Aedes scutellaris* complex. Outbreaks of dengue virus transmissions are attributed with the current diversity in domestic transportation systems and foreign travels, where latest first cases of outbreaks due to *A. Albopictus* have been reported from Havelock Islands, near Andaman Islands (Sivan et al., 2016). Although there are statistical disputes, there is an estimated 50 to 100 million cases of infection every year (Stanaway et al., 2016).

Upon a mosquito bite on a human victim, the virus makes its entry into the human host via infecting the nearby skin cells designated as keratinocytes, which is the commonest skin type. The virus also infects the immune cells in the skin, specifically a type of dendritic cell named Langerhans cell. The Langerhans cells in turn trigger the immune system by traveling to the lymph nodes, alerting of pathogen invasion by displaying the dengue viral antigen on its surface. The lymph nodes, in turn, produce white blood cells, namely monocytes and macrophages aimed at neutralizing the virus. However, the virus in turn infect both the white blood cells, spreading throughout the body by perpetually tricking the mechanism of the immunity (Diamond, 2003). **Figure 2.2** summarizes the process of infection to the viremic stage.
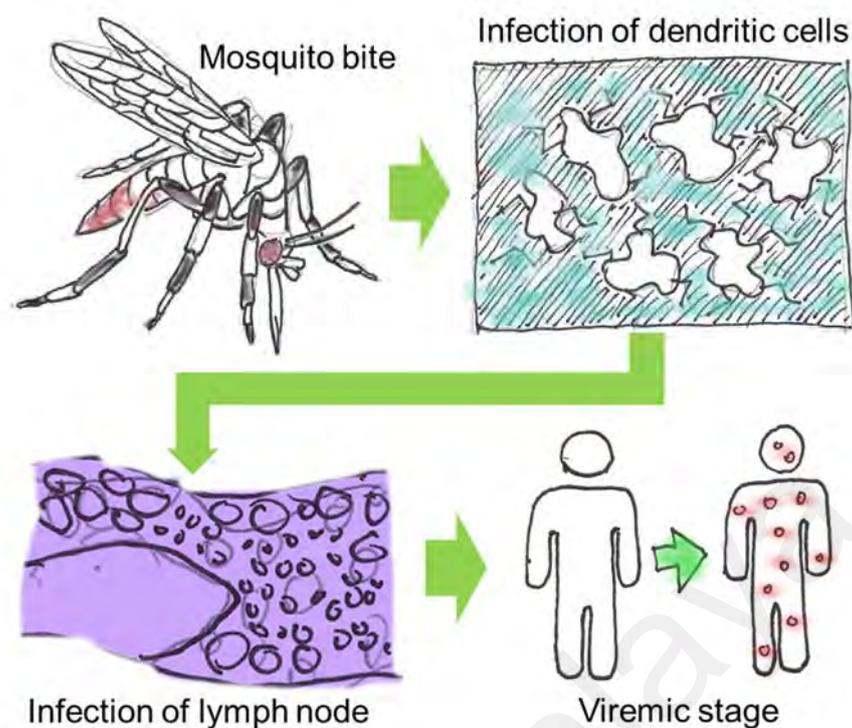
Figure 2.2: Process of dengue infection

Following an infected mosquito bite on a subject and an incubation period of 4-10 days, varied symptoms will give rise, although most cases exhibit subclinical and asymptomatic tendencies. The subjects are thought to have a lifelong immunity towards the specific serotype which invaded the body (Organization et al., 2009).

Although the immunity is maintained for other serotypes, this lasts for a typical 2-3 months after the primary infection. From here, cross-immunity against other serotypes are no longer available (Organization et al., 2009). Infants have been shown to acquire immunity from an infected mother, but unfortunately this acquisition acts as another precursor towards enhancing diseases (Clapham et al., 2015). Development of human vaccines, for which none are licensed yet globally, is based on providing immunity for all four serotypes remain elusive when assumed lifelong immunity is not totally guaranteed, although monkeys have shown full immunity (Forshey, Stoddard, & Morrison, 2016; Schwartz, Halloran, Durbin, & Longini, 2015; Valdés et al., 2016). This also projects the possibility of infection with the same serotype, which was proven when

homotypic re-infections was reported in Nicaraguan children (Waggoner et al., 2016). To date, dengue vaccines still face adversity in the health sector (Aguiar et al., 2016; Flasche et al., 2016).

The severity of the infections depend on a few factors including age, ethnicity, probable chronic diseases, and most severe cases are credited to secondary heterotypical infections, currently thought to be due to the human immunopathology (Screaton, Mongkolsapaya, Yacoub, & Roberts, 2015). In this case, severity also depends on the viral sequence and the period between infections (Organization et al., 2009). The extreme end of the variants of symptoms includes capillary/plasma leakage, coagulopathy, organ impairment, haemoconcentration and abnormalities in homeostasis (Organization et al., 2009; Yacoub, Mongkolsapaya, & Screaton, 2016). Several primary infections have also reported severe cases. Younger children are less capable to overcome capillary leakage and is subject to a range of fatalities, where for instance fulminant viral myocarditis was reported on a 5-year old boy following a dengue infection where the boy died within 24 hours (Abrar & Ansari, 2016). Diffuse cerebral hemorrhages, subdural hematoma and cranial diabetes insipidus was reported on a 24-year old healthy lady, where dengue infection was confirmed serologically, later died to the illness (Jayasinghe, Thalagala, Wattegama, & Thirumavalavan, 2016). **Figure 2.3** summarizes the normative body response towards infection.
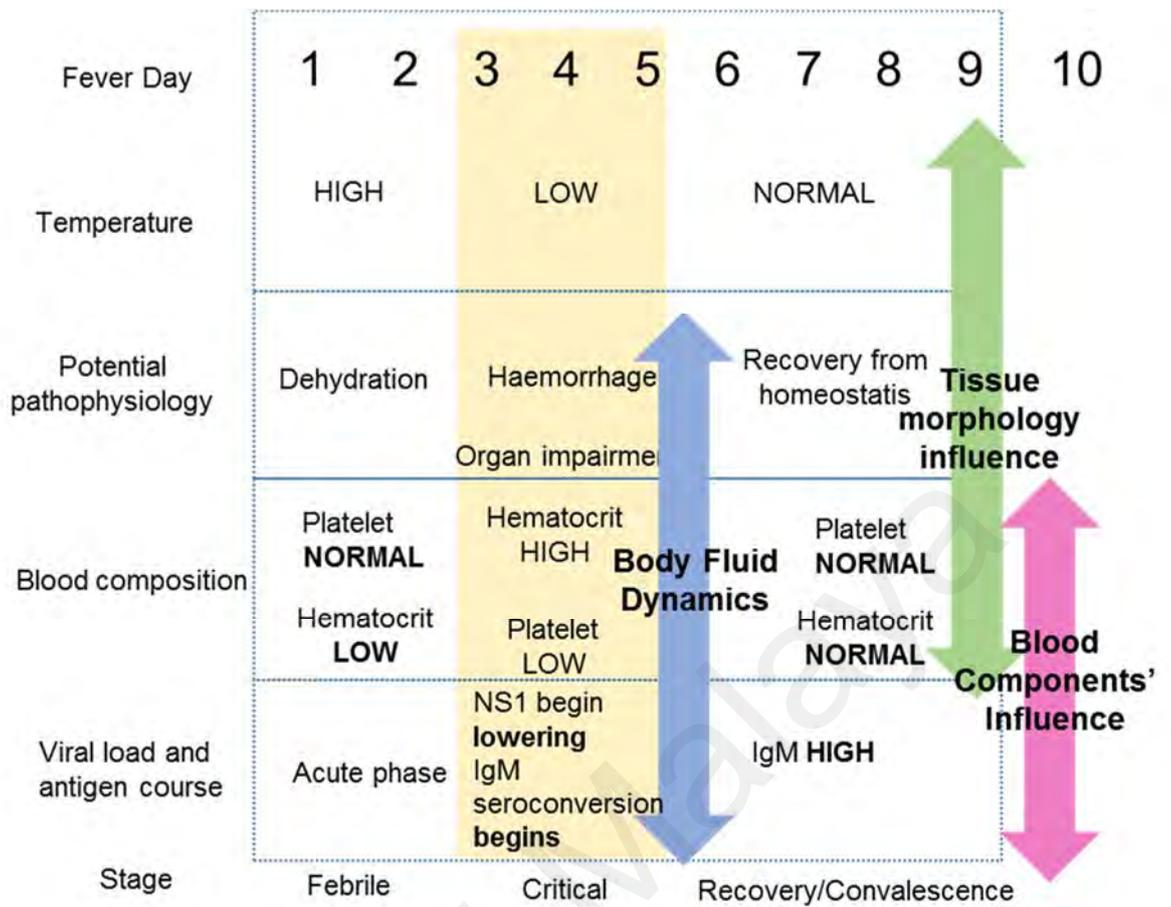
Figure 2.3 The course of dengue illness and relations to tissue morphology

From **Figure 2.3**, the critical stages of the infection is from 3-6 days, where the critical phase of the illness will commence from this period. A severe dengue symptom begins as the vascular permeability increases in the patient, where bouts of shock may occur due to hypovolemia. This usually occurs during defervescence, 3-7 days post-infection. The initial stage of the shock enters into the body when the compensatory mechanism in the body induces tachycardia and peripheral vasoconstriction with a reduction of the skin perfusion, which results in cold extremities in the patient and delayed capillary refill period (Organization et al., 2009).

Counterintuitively during this period, the diastolic pressure increases towards the systolic pressure, while the pulse pressure constricts as the vascular feedback resistance gains. In this critical period, the patient may feel normal and alert while dangerously teetering on the brink of severe shock. In this phase, less sensitive physicians monitoring the patient, may be registering normal systolic pressures despite the critical state of the patient. Following this period, decompensation occurs and abrupt disappearance of both diastolic and systolic pressures occur. Hypotensive shock and hypoxia, which follows, in prolonged periods ultimately causes multiple organ failure and leads to several extremely adverse clinical complexities, and ultimately death (Organization et al., 2009). Fatality rates differ from region to region between >1% to a record of 3.55%. **Figure 2.4** shows the distribution of dengue incidence rates in Malaysia per 100,000 population, rendered based on available data (Mohd-Zaki, Brett, Ismail, & L'Azou, 2014).
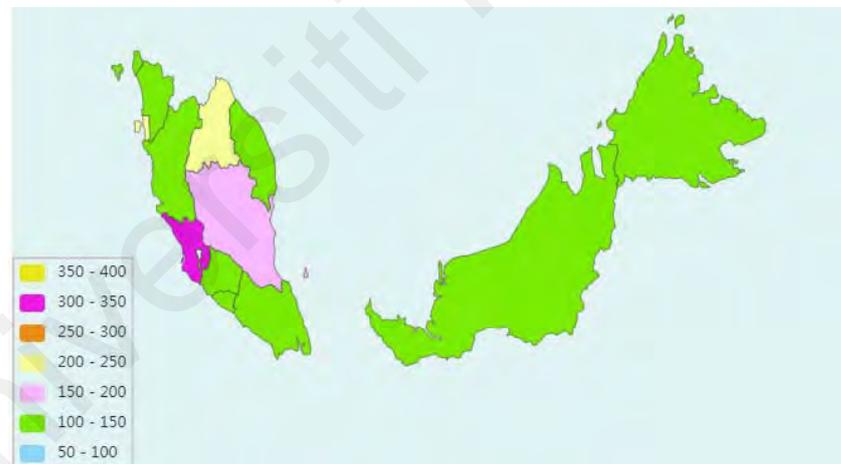
Figure 2.4: Dengue incidence rates in Malaysia

For severe cases, where potential cases of dengue shock syndrome (DSS) comes in, the temporal disadvantage of conventional diagnostics comes into question and a more important question arises: how can diagnostics become more efficient? Therefore, point-of-care diagnostics and rapid detection tests plays a game-changing importance. For this purpose, diagnostic methods will be discussed in the following section.

2.2   Standard Practices: Definitive Dengue Diagnostics

There are two general methods in diagnosing a DENV infection, either a direct or indirect approach. Direct methods use virus isolation, genome detection or antigen (NS1) detection methods. Indirect methods include serological methods including IgM or IgG detection. However, IgG/IgM detection is only available after a typical 6 days. Direct methods have a high specificity but lower opportunities due to the complicated procedures involved, where although several detection kits have been developed, none was validated (Peeling et al., 2010).

As part of clinical procedures in determining dengue virus infection, blood samples are taken from infected patients with suspected dengue infection. These samples are further separated to extract the serum from the blood and will be subjected to tests such as the Enzyme-Linked Immunosorbent Assay (ELISA) for IgM antibody detection. As the commonest practice in dengue detection, ELISA takes a long time to produce readings and depend on precise instruments to interpret the readings. Also, these equipment are costly. It takes a trained medical personnel to perform these tasks. This task also may take up to several hours. These equipment are not Point-of-Care (POC), will not be able to cater a speedy diagnostics test, may not be available in remote areas, and will not be portable for a trained clinician to handle (Jahanshahi, Zalnezhad, Sekaran, & Adikan, 2014). Other than ELISA, several other types of tests include reverse transcription-loop-mediated isothermal amplification (RT-LAMP) assay, quantitative RT- polymerase chain reaction (qRT-PCR). A recent breakthrough in clinical tests include a reverse transcription-recombinase polymerase amplification method where the DENV RNA is

detected within less than 20 minutes (Teoh et al., 2015). The summary of conventional

diagnostic methods are in **Table 2.1**.

Table 2.1: Summary of dengue diagnostics in standard setups

Methods/ Attributes	Viral isolation and identification	RNA detection	Antigen/Serological Tests		
			NS1	IgM	IgG
Confirmation strength	High (100%)	No	Yes	No	Yes
Serotype identification	Yes	Yes	No	No	No
Acute sample required? (0-5 days fever onset)	Yes	Yes	Yes	No	No
Expertise-dependent	Yes	Yes	Yes	Yes	Yes
Results <24 hours	No	Yes	Yes	Yes	Yes
Cost-effectivity	No	No	Yes	Yes	Yes
Requires blood sera?	Yes	Yes	Yes	Yes	Yes
Main disadvantage	Highly complex setup and personnel requirement Highly complex setup and personnel requirement (Teoh et al., 2015)		Prone to false positives due to cross-reactivity with other flaviviruses (Steinhagen et al., 2016)	IgM levels stay detectable after 60 days (Rigau-Pérez et al., 1998)	High seropositivity among Malaysians at 91.6% (Azami, Salleh, Neoh, Zakaria, & Jamal, 2011)

2.3   Experimental Methods: In-vivo, Rapid-Detection, Point-of-care, and Optical Methods

There were numerous attempts at breaking away from many of the complexities of

dengue diagnostics. In exploring more divergent techniques, a detection scheme which

translates into swift treatment is desirable, which is where rapid detection schemes/test

(RDT) kits boasts advantage. Other methods attempted at discounting the process of

blood withdrawal, by utilizing chemical assays for detection of other antigens in saliva

and urine samples. Perhaps, the most experimental techniques on the frontline revolve

around the use of optical methods, which have a reputation of obtaining serological

diagnosis nearly immediately. **Figure 2.5** non-exhaustively summarizes what can be expected for different diagnostics' path, with elaboration based on this graphic in the following subsections.
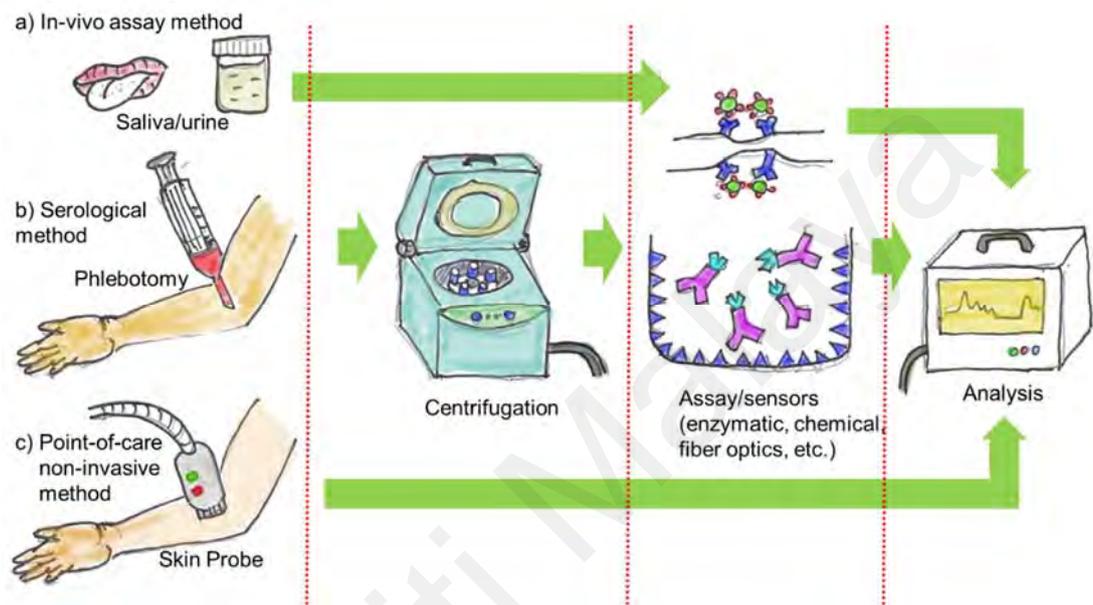
Figure 2.5: Summary of different paths taken for a diagnostics test

2.3.1 Saliva and Urine Assays – In-vivo Detection

Bloodless and in-vivo methods also have been reported for dengue diagnostics. A report on using antigen capture anti-DENV IgA (ACA) ELISA technique was also published, where the saliva of subjects was used with 36% sensitivity on primary infections and 100% for secondary infections (Yap, Sil, & Ng, 2011). Several non-invasive methods include using the saliva and urine via detection of NS1 antigen and RNA in the saliva and urine of travelers with dengue was also reported, where the sensitivity was reportedly 72% and 50% respectively. However, similarly to other clinical

tests, this test shall have to be conducted in-lab with exclusive instruments of precision (Korhonen, Huhtamo, Virtala, Kantele, & Vapalahti, 2014). A stacking flow immunoassay was developed which detected the presence of IgG in saliva. However, this method is not able to provide a definitive analysis for a currently infected patient (Y. Zhang, Bai, & Ying, 2015). However, these tests again require trained personnel, reagents and lab equipment for analysis, as shown in **Figure 2.4(a)** as per the path of diagnostic completion.

2.3.2   Rapid Detection Test

In field usage, there are increasing demands for rapid diagnostic test (RDT) kits. This alleviates clinicians in remote areas from the more complicated mechanisms of a hospital-run ELISA-based diagnostics. This, in result lowers the burden of the operation of larger and government-run health centers, and allows for patients to receive better healthcare in their own locality, while discounted from the costs of traveling and queuing period.

One of the latest techniques are reverse transcriptase loop-mediated isothermal amplification (RT-LAMP), used for detecting dengue viral RNA was a technique reportedly capable of differentiating the four serotypes as well (Lopez-Jimena et al., 2018). Using the similar RT-LAMP method, a recent smartphone-based diagnostic kit was also developed, by application of a pinpricked blood on a microfluidic platform and further image processing on the amplification reaction allows a diagnosis within 50 minutes (Ganguli et al., 2017). A commercial field-deployable device (POCKIT^{TM}, GeneReach Biotech, Taichung, Taiwan) reportedly boasts a variant of PCR method for detecting dengue RNA under 60 minutes. However, the setup is nontrivial, with multiple steps required before diagnostics are provided (Tsai et al., 2018). Some of the RDT schemes have been given a fairly poor review due to the dispute towards the claims of the

products. These include an RDT immunochromatographic test of a certain commercial brands which were given fairly poor reviews with varying sensitivities (Carter et al., 2015; Osorio et al., 2015).

In summary, these tests generally tread the path as shown in **Figure 2.4(b)**, where blood withdrawal, centrifugation, and assay processions are required before laboratory analysis and diagnosis is given. Based on comprehensive reviews, however, rapid detection schemes face many challenges, especially from the economic perspective (Hunsperger et al., 2017; Lim, Alexander, & Di Tanna, 2017; Muller et al., 2017). This is partly due to other complications which arise from the innovation which faces other cost-related issues, which is not found in conventional methods.

2.3.3 Optical and Photonics-based Detection

The use of optical methods for dengue diagnostics have recently gained traction. This is due to the rapid results characterized by optical detectors, such as spectrometers or photodetectors. These detection methods owe their fast response due to the real-time electrical properties of the detectors. This is in contrast to chemical or enzymatic assays which depend on temporal-based reaction for its dengue-specific characteristics to be detectable by any means available.

Optical methods to detect DENV have made several appearances, such as a chemiluminescent optical fiber was also developed to detect the presence of IgM antibody, where sensitivity and specificity of 98.1% and 87.0% was achieved respectively (Atias et al., 2009). Another type used a dengue immunoassay with a Localized Surface Plasmon Resonance (LSPR) fiber optic sensor. However, this study was still a proof-of-concept, and takes multiple processes within a long period and may be subjected to further

studies (Camara et al., 2013). A more recent and less time consuming setup used surface plasmon resonance (SPR) waveguides. However, this setup still requires separated blood for sera and a fully equipped lab to run tests completely (Jahanshahi et al., 2014; Wong, Krupin, Sekaran, Mahamd Adikan, & Berini, 2014).

The use of transmission and Raman spectroscopy on the blood of dengue-infected patients have shown diagnostic value. This was performed by a pattern analysis of the transmission and Raman spectra of sera of dengue-infected patients, compared to healthy control patients (S Firdous et al., 2012; Firdous & Anwar, 2015; Khan et al., 2016).

In the realm of optoelectronics, the use of fiber-based sensors also have been reported. This includes several modes of coated fiber, reacting to dengue NS1, IgG and E proteins in the blood sera, with its best detection period of under 120 seconds (Kamil et al., 2019; Mustapa, Bakar, Kamil, Hamzah, & Mahdi, 2018). Surface-plasmon resonance (SPR) sensors, recently gaining traction, have been reported for reaction with IgM antibodies, ranging from serological analysis to kinematics. This technique, however, depended on the binding effect of the antibodies on the surface of the sensors, rendering optical power fluctuations to generate the specific patterns for diagnostic analysis within 120 minutes (Jahanshahi et al., 2017; Wong, Berini, Fan, & Adikan, 2018).

2.4   Breaking Bad Blood: Is Non-Invasive Dengue Detection Possible?

Discounting saliva or urine assays, most diagnostic techniques in all methods depend on the blood serum, which necessitates blood withdrawal and centrifugation. This was a trend which can be seen in previous sections of this chapter. However, there has been virtually no feasibility studies on non-invasive dengue detection from the human body. Perhaps the only established linkage between non-invasive optical methods to dengue diagnostics was in the later stages of dengue. One study mentions the use of near-infrared spectroscopy (NIRS) to detect plasma leakage in dengue-infected children. In this study,

levels of muscle oxygenation in the thigh (SmO_2), measureable by means of NIRS was correlated with the manifestations of plasma leakage in the form of pleural effusion (Soller et al., 2014). However, no studies have reported potentials of detecting peculiarities of the tissue morphology in dengue patients via skin optics.

2.4.1 Skin Optics as a Tool for Diagnostics

The optics of the human skin has received numerous coverage for a relatively long time. This is due to the many applications and the insight it provides for detecting and management of diseases. Turning to skin spectroscopy, which has been performed for various reasons, including non-invasively diagnosing cancer (Broer, Liesenhoff, & Horch, 2004), diabetes (Ghazaryan et al., 2012), acne-based bacterial infections (Seo, Tseng, Cula, Bargo, & Kollias, 2009), and treatment of cutaneous disorders (Xiao, Wang, Luo, Li, & Wuerger, 2016). The mechanism allowing the diagnostic capabilities of this method range from the detection of specific fluorophores in the skin, to tissue morphology.

The spectrum of the human skin has been characterized in various studies between 250-2500 nm (Angelopoulou, 1999; Cooksey, Tsai, & Allen, 2014). A typical setup for measuring skin reflectance is as shown in **Figure 2.6**.
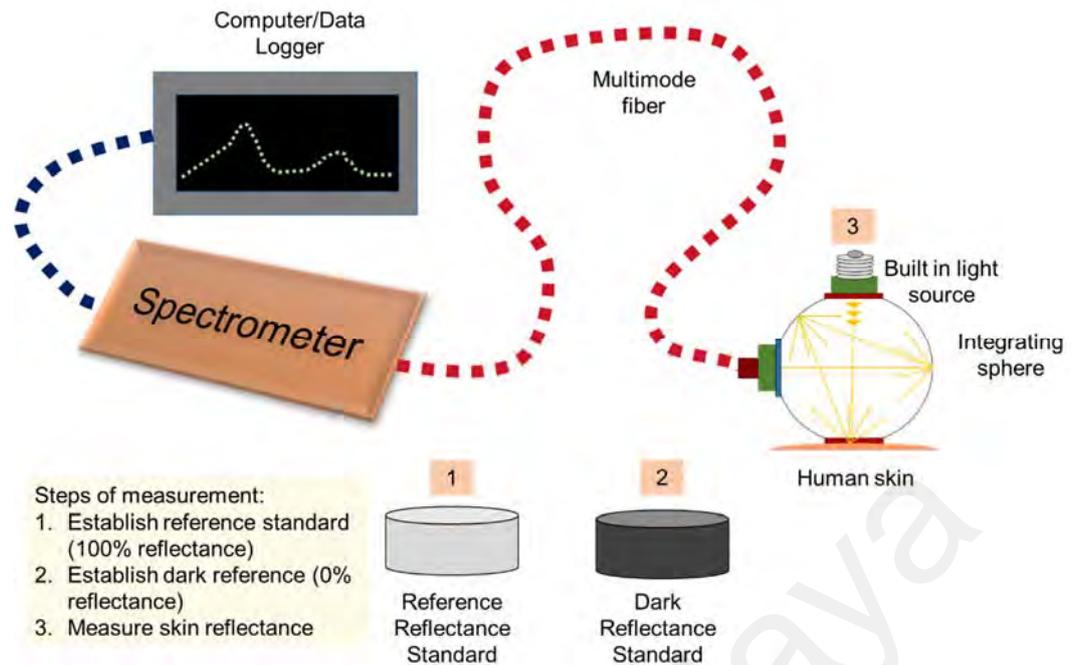
Figure 2.6: A setup for measuring skin reflectance

Normally, an integrating sphere is used to measure the directional-hemispherical reflectance measurements. The sphere is coated with a highly reflective material, such as sintered polytetrafluoroethylene (PTFE). The incident light from the illumination source is reflected from the human skin, and recollected using a bundle fiber. The reflected optical signal is further fed to a spectrometer, where further analysis will be made with the spectrum profiles (Cooksey et al., 2014; Yoshida & Nishidate, 2014).

There are various aspects of the human skin spectroscopy, which affects the profile. A spectral profile of the human skin will represent the collective effect of hair, the epidermis, dermis and the subcutaneous tissues. The epidermis consists of the stratum corneum (SC) and melanin or melanocytes which is a product of the skin against radiation of sun, where dark pigmentations form. The dermis comprises of veins, sweat glands and capillaries, nervous cell networks and hair follicles. The subcutaneous tissues comprises

of blood vessels and fat cells (Jacques, 1996). **Figure 2.7** shows the three dominant wavelength absorption ranges of different components of human skin:
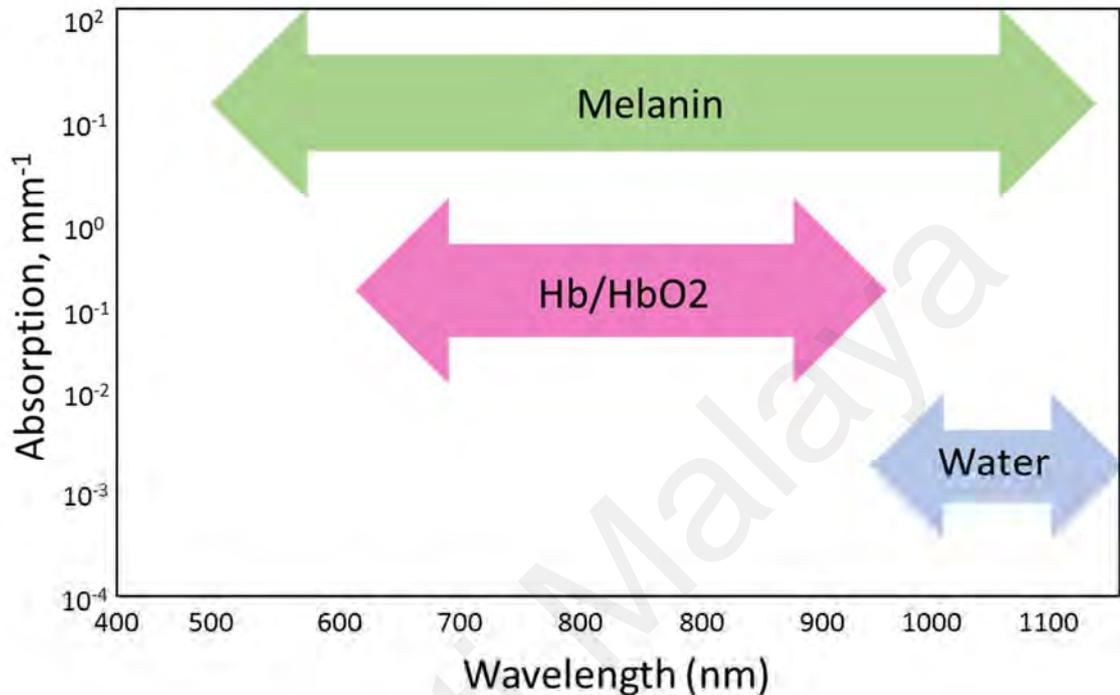
Figure 2.7: Absorption ranges of several skin components (mm^{-1}) vs wavelength (nm).

There are also variations of the optical properties for each of the components of the human skin across different ethnicities and regions where genetics and weather plays a role (Cooksey et al., 2014). Also, the optical properties vary with configurations of the profile, whether it is via scattering, absorption, or reflectance of the human skin. Also, in-vitro or in-vivo configurations will produce different results (Lister, Wright, & Chappell, 2012). Several wideband studies on human skin spectroscopy is non-exhaustively summarized in Table 2.2. It is to be noted that these studies mentioned in this section mainly reflect the diversity of the applications in skin spectroscopy.

Table 2.2: Summary of the optical properties of human skin in literature (SVW: Significant variability wavelengths)

No	Authors	Wavelength range	Type	Applications
1	(Angelopoulou, 1999)	400-700 nm	Reflectance	Differentiating mannequins and human skin
2	(Bashkatov, Genina, Kochubey, & Tuchin, 2005)	300-2000 nm	Absorption	Optical properties of human skin
3	(Kaxiras, Tsolakidis, Zonios, & Meng, 2006)	100-800 nm	Absorption	Eumelanins in skin
4	(Matts, Dykes, & Marks, 2007)	400-700 nm	Reflectance	Melanin content in skin
5	(Cooksey et al., 2014)	250-2500 nm	Reflectance	Skin reflectance variability
6	(Yoshida & Nishidate, 2014)	400-700 nm	Reflectance	Skin modelling
7	(Mendenhall, Nunez, & Martin, 2015)	450-1800 nm	Reflectance	Human skin detection

Based on the summary, we can derive several important characteristics of the human skin, where most of the variability as shown in the spectrum is limited to certain bandwidths. Significant Variability Wavelengths (SVW) is an observational measure introduced in this section for summarizing purposes, which is derived from the amount of diversity of the reflectance values corresponding to certain wavelength ranges. The SVW ranges are summarized in **Figure 2.8**.
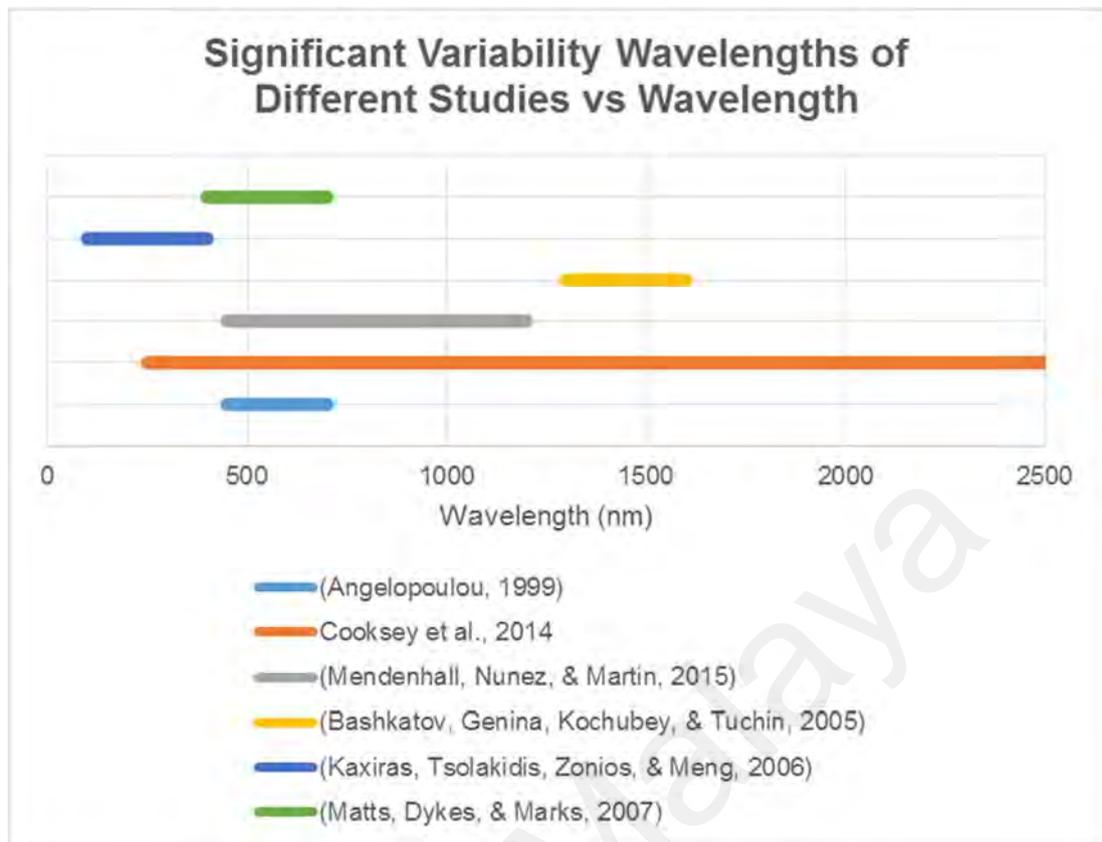
Figure 2.8: Significant Variability Wavelengths of different studies versus wavelength. The y-axis is merely to separate the different case studies.

By comparing all the studies in **Figure 2.8**, in can be projected that the most probably influential wavelengths of response range between 200-2500 nm for further experimental purposes. This is due to the absorption of the remaining source illumination beyond the infrared region by the presence of water in the human skin cells, rendering ranges beyond 2500 nm as redundant.

2.4.2   Physiology Features of Dengue Patients: Possible Link with Spectroscopy?

In line with the most efficient method projected in **Fig. 2.5(c)**, which bypasses at least two major steps in conventional diagnostics, optical spectroscopy offers a plausible solution. In conjunction with the previous section, we have demonstrated the wide range

of applications in optical spectroscopy for shedding light of the human condition, whether in illness or in health. Dermatological and physiological aspects of dengue patients, whether in acute or febrile to the convalescence stage have been receiving numerous coverage. These characteristics, combined with the penetration depth of a wideband light source which can penetrate up to 3,500 microns of human skin and tissue, as established by Anderson et al in 1981 (Anderson & Parrish, 1981), provides a clue on how to expect reflected spectral patterns which can be attributed to dengue patients.

A typical dengue patient, in the febrile phase are examined for several criteria. For probable dengue cases, this includes fever, nausea or vomiting, rashes, myalgia, leukopenia and capillary leakage, indicated with a tourniquet test. Warning signs include arthralgia, vomiting, fluid accumulation, mucosal bleeding, lethargy, enlarged liver and a rise in haemoconcentration (or hematocrit levels, HCT) accompanied by rapid decrease in platelets. In severe dengue cases, shock and fluid buildup accompanied by difficulty in breathing are the criteria (Organization et al., 2009).

The pathophysiological aspects of a dengue patient which may be detected by non-invasive optical means, in turn, can be divided in three categories, where the criteria for these different severity levels may partially contribute towards spectral uniqueness.

The first category is based on serological and viral component. Optical characteristics of dengue-infected patients' blood sera was analyzed by Firdous et al. without any enzymatic linkage or assay-based processes, yielding differentiable spectra between dengue and non-dengue patients. The range of absorption was reportedly between 300-900 nm, which covers part of the ultraviolet, visible and near-infrared regions (UV-VIS-NIR) (S Firdous et al., 2012; Firdous & Anwar, 2015; Khan et al., 2016).

The second component which may contribute to the spectral peculiarity of dengue patients is based on tissue morphology, which comprises of oxygenation and fluid components. Near-infrared spectroscopic (NIRS) measurements reportedly was capable of detecting muscle oxygenation fluctuations due to plasma leakage in dengue patients during the recovery phase (Soller et al., 2014). The device used in their measurements cover ranges in the NIR, albeit its actual range was not disclosed. However, measurements of tissue oxygenation in other publications indicate the ranges 400-1600 nm as the used bandwidth (Boushel & Piantadosi, 2000; Wollina, Liebold, Schmidt, Hartmann, & Fassler, 2002). Though this measurements occur within the recovery period, manifestations of plasma leakage can occur during the febrile phase, as confirmed in sonographic measurements (Srikiatkhachorn et al., 2007). Visible cutaneous features of dengue patients in the febrile phase also may register on the spectrum, where precursors of rashes, pruritus and petechial eruption occur (Azfar et al., 2012; Manoharan, Ramasamy, & Heinz, 2017; Tjandra, Ehrchen, Broekaert, & Sunderkötter, 2017). These features arise due to capillary dilatation, which facilitates the leakage of the blood components.

The third component relies on the composition of the blood or tissues, usually represented by Raman spectroscopy. In the acute phase of a dengue infection, thrombocytopenia and leukopenia is a common feature in patients. Though not reported for dengue, measurements of leucocytes from the blood has been reported via optical spectroscopy. Measurements between 350-1150 nm for plasma and platelet content was characterized and shows measurable differences (Meinke, Müller, Helfmann, & Friebel, 2007). Hematocrit levels have been reportedly detectable non-invasively within 575-1100 nm (S. Zhang, Soller, Kaur, Perras, & Vander Salm, 2000).

Table 2.3 summarizes the criteria, and potential contributions to a spectral regions.

Table 2.3: Potential contribution to spectral peculiarity in dengue patients

Dengue Patient Criteria	Measurable by spectroscopic methods?	Wavelength range (nm)	References
Fever (general criteria)	Yes, but irrelevant due to study design.	n/a	n/a
Serology			
NS1/IgM/IgG	Yes	300-900	(S Firdous et al., 2012; Firdous & Anwar, 2015; Khan et al., 2016)
Probable dengue			
Nausea/vomiting	No	n/a	n/a
Rashes	Yes (vasodilation)	780-830	(Takahashi et al., 2011)
General aches	No	n/a	n/a
Leukopenia	Yes	350-1150	(Meinke et al., 2007)
Capillary leakage	Yes (in terms of tissue oxygenation)	400-1600	(Boushel & Piantadosi, 2000; Soller et al., 2014; Wollina et al., 2002)
Warning/Other signs			
Myalgia	Probably (Tissue oxygenation and plasma reduction)	350-1600	(Boushel & Piantadosi, 2000; Flodgren, Crenshaw, Hellström, & Fahlström, 2010; Meinke et al., 2007; Soller et al., 2014; Wollina et al., 2002)
Vomiting	No	n/a	n/a
Clinical fluid accumulation	Yes (plasma or water)	1300-2200	(Arimoto, Egawa, & Yamada, 2005)
Mucosal bleeding	No	n/a	n/a
Lethargy	Probably (muscle oxygenation)	400-1600	(Boushel & Piantadosi, 2000; Soller et al., 2014; Wollina et al., 2002)
Enlarged liver	No	n/a	n/a
HCT levels	Yes	575-1100	(S. Zhang et al., 2000)
Platelet levels	Yes	350-1150	(Meinke et al., 2007)
Severe dengue			
Shock	Yes (plasma or water)	1300-2200	(Arimoto et al., 2005)
Fluid buildup with respiratory distress			

2.5 Chapter Conclusion

In this chapter, we have reviewed dengue virus, its vectors and the epidemiology, followed by the pathophysiology of the illness. This includes signs and symptoms and classes of dengue infections.

Following the issues surrounding the disease, several diagnostic techniques are also evaluated. Since there are numerous indicators of an infection, and serotypic identification plays an important factor as well, we have summarized two conventional and general methods, namely the direct and indirect methods. Direct methods isolate the virus or RNA of the dengue itself, while indirect methods focus on the aftermaths of an infection, either through the NS1 antigen detection, or immunoglobulin proteins (IgG or IgM). However, details of the biochemistry of the indirect methods are not discussed for the sake of brevity.

Consequently, we also discussed point-of-care (POC) schemes and several rapid detection tests or kits (RDTs) where although the achievement of these kits are revolutionary, none have been standardized yet. Several serological methods also employed the robustness of optical devices including long range and localized surface plasmon resonance (SPR) biosensors, and chemiluminescent fiber optics immunosensors were evaluated as well. However, all the methods above require sera from the blood, which is by itself not totally POC in principle. Several in-vivo methods were also covered, albeit its limitations due to the assay procession required to analyze saliva and urine.

Finally, potential precursors to non-invasive methods for dengue diagnostics are discussed in depth. This method excludes the need for invasive blood-drawing exercises, and can be practically implemented for monitoring patients' hematological compositions, if a definitive pattern can be observed linked to the serology. This was attempted within two phases, first by a fundamental study on skin reflectance, and attempting to link the studies with dengue diagnosis. As mentioned in **Section 2.4**, the assumption that this method may be studied is due to three confirmed optical characteristics of the human body. The first is the serological composition of the blood. Secondly, the cutaneous manifestations of dengue patients which may have unique optical characteristics. Thirdly,

dengue-infected blood sera of humans have unique optical features. Thus it is warranted that a further fundamental study is required to see if there are any optical peculiarities of the skin or tissue of humans infected with DENV.

With these overwhelming evidence presented, we hypothesize that by utilizing diffuse reflectance spectroscopy on febrile patients within the UV-VIS-NIR (200-2500 nm) range, dengue and non-dengue cases can be differentiated via spectrum pattern analysis.

CHAPTER 3: OPERATING PRINCIPLES, METHODOLOGY AND MANAGEMENT

This chapter discusses several aspects of the use of optical spectroscopy on patients for building a database of skin spectroscopy profiles. This chapter is divided into five sections. The first covers the spectrometry setup. The second elaborates on patient management. The third section elaborates on the data management. The fourth section covers the clinical discourse for data verification with physicians and dengue experts. The final section concludes the chapter. **Figure 3.1** is a map for navigating this chapter.
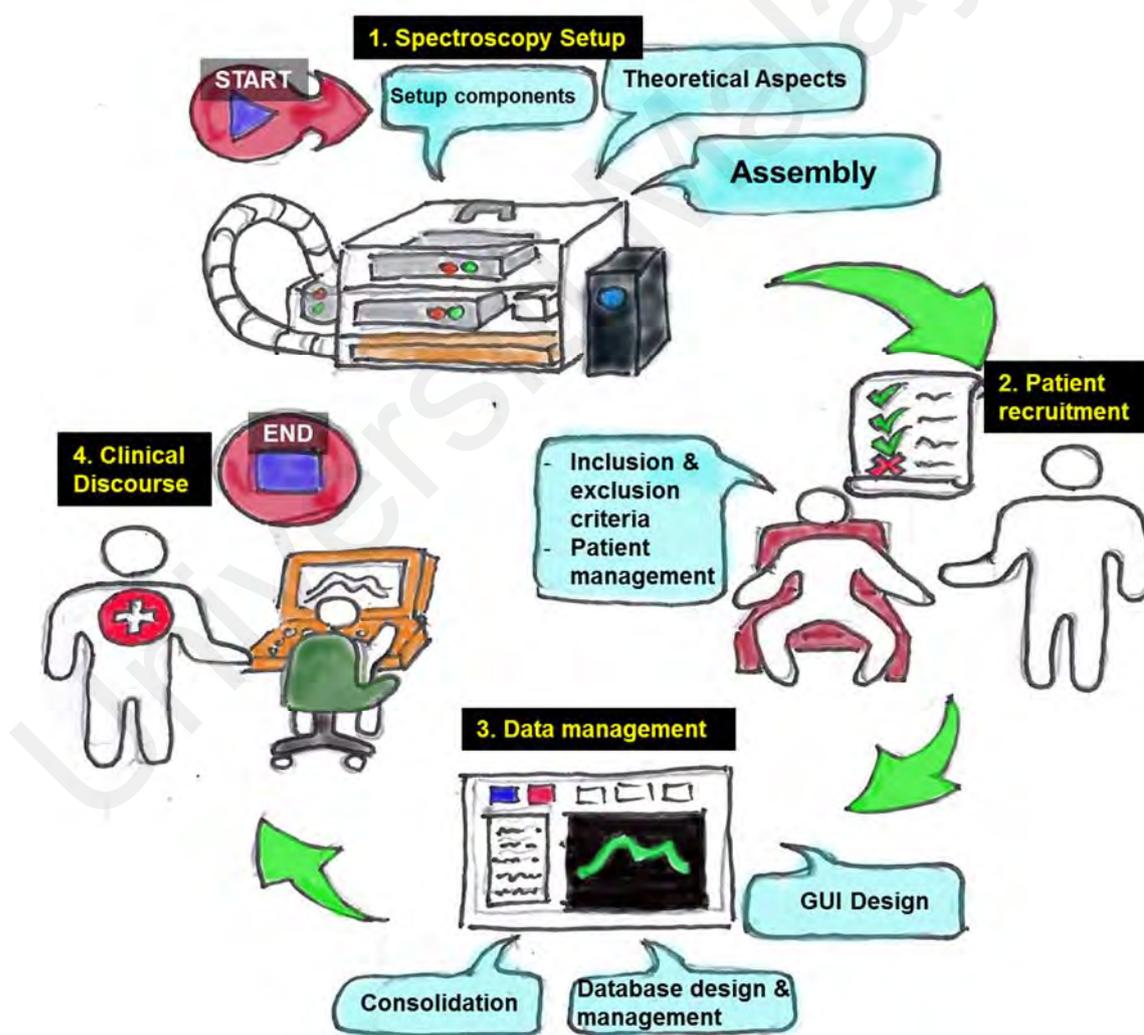
Figure 3.1: Chapter map

3.1 Theoretical Aspects: Operating principles

This section elaborates on the aspects which determine a spectral profile of a human skin, and technical aspects of how an integrating sphere is used to capture the reflected spectra. There are two types of reflectance, namely diffuse and specular. A diffuse reflectance spectroscopy (DRS) setup in this case relies on the principles of interaction of a light source with human skin or tissue, resulting in a reflection. In contrast to certain non-biological materials, such as a mirror, the reflection is mostly described as diffuse. In specular reflections, the beam of light reflected from a material such as a mirror is mainly concentrated in a wave propagation path where the reflection occurs at the surface of the material.

In most materials especially the human skin, however, both diffuse and specular reflection plays a role in determining the collective reflection spectra. Both the external and internal layers of a material of concern is involved in contributing to the specular and diffuse reflection respectively. While minor specular reflections occur on the skin surface due to different refractive indices of the stratum corneum (SC) surface and air, a preponderance of the light penetrating the skin causes diffuse reflections that occur mainly due to scattering within the SC and the components of the dermis, while absorbing part of the radiant energy as well.

Figure 3.2 in the following illustrates the structure of the human skin.
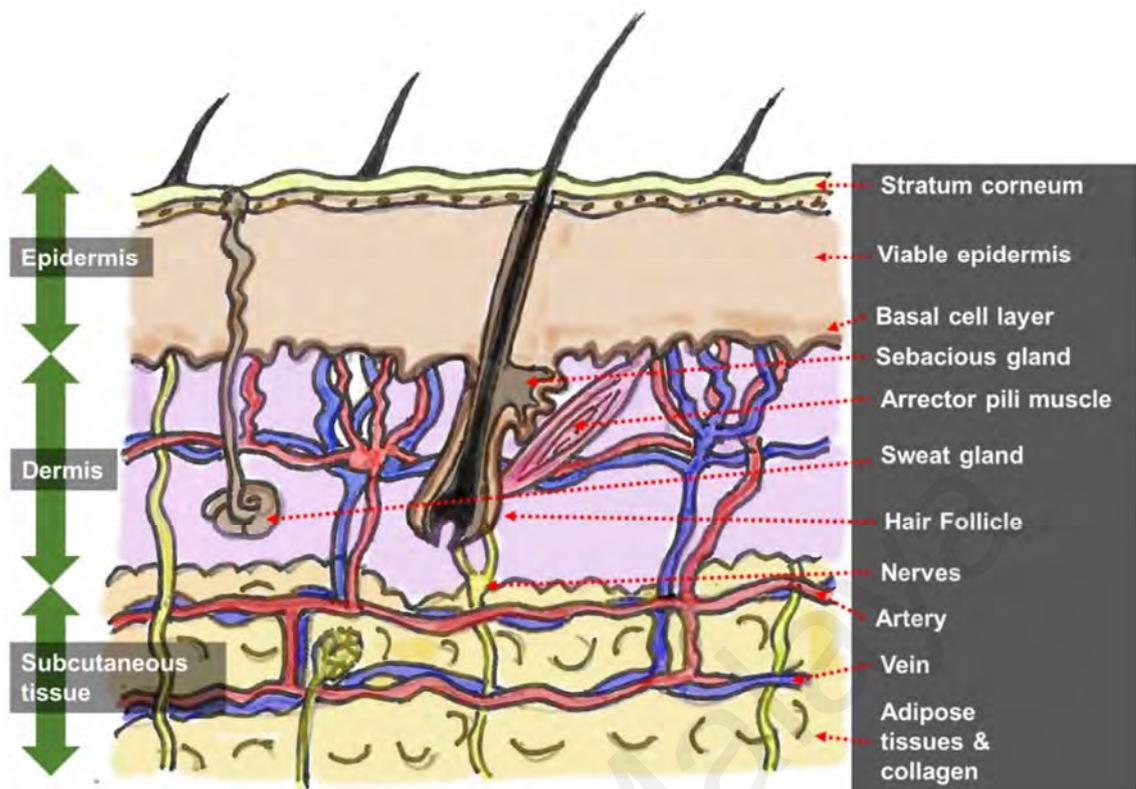
Figure 3.2: Skin structure

As shown in **Figure 3.2**, human skin, however, comprises of a turbid medium of non-uniform layers of components, each with its own optical properties. In most instances, the interaction of light with the skin structure is a complex combination of the absorption mainly of a few components. This includes melanin, produced and distributed around the layer of melanocytes. Another component is oxy- and deoxygenated hemoglobin, perpetually pulsating around the blood vessels under the epidermis. Yet another component is water, which is present throughout the skin structure non-homogenously (Cooksey, Tsai, & Allen, 2015). However, other components of the skin also contribute to the reflectance pattern as discussed in **Chapter 2**. It has been documented that separate optical properties of the skin do not necessarily equate a skin spectrum collectively (Gillies, Zonios, Anderson, & Kollias, 2000).

Despite the detailed attribution of several components to the spectral profile of human skin, it is noteworthy to bear in mind the utter complexity of the human skin structure. It

has been asserted that a precise treatise on the mathematical modelling in bridging radiation patterns to human skin is virtually impossible (Anderson & Parrish, 1981). Currently, this is still the case, but with improved mathematical modeling (C. Li, Brost, Benezeth, Marzani, & Yang, 2018; Masuda, Ogura, Inagaki, Yasui, & Aizu, 2018). The following sub-subsections discusses several theoretical aspects involved in quantifying skin reflectance.

3.1.1   Kubelka-Munk Theory of Reflectance

Before delving into the physical and mathematical foundations of the human skin, an approximate illustration of the relationship between a light source and human skin needs to be clarified. Following an illumination from a light source, the light propagates and reaches the surface of the skin, or the epidermis. Part of the light (~5%) is reflected due to the different refraction index between the air and the stratum corneum (SC) on the epidermis. Depending on the wavelengths, the non-reflected light waves penetrate the epidermis and reaches into the dermis and possibly a small part of the subcutaneous layer. In these layers, however, the interaction is nontrivial, with multiple dermal components scatter and absorb the energy from the photons of the light source. At the region 1090 nm, the light source reaches a maximum depth of 3.5-5.0 mm (Ash, Dubec, Donne, & Bashford, 2017; Cooksey et al., 2015), which approximately exceeds dermis depths as well (Anderson & Parrish, 1981). **Figure 3.3** illustrates the optical paths taken by a diffuse light source through the human skin.
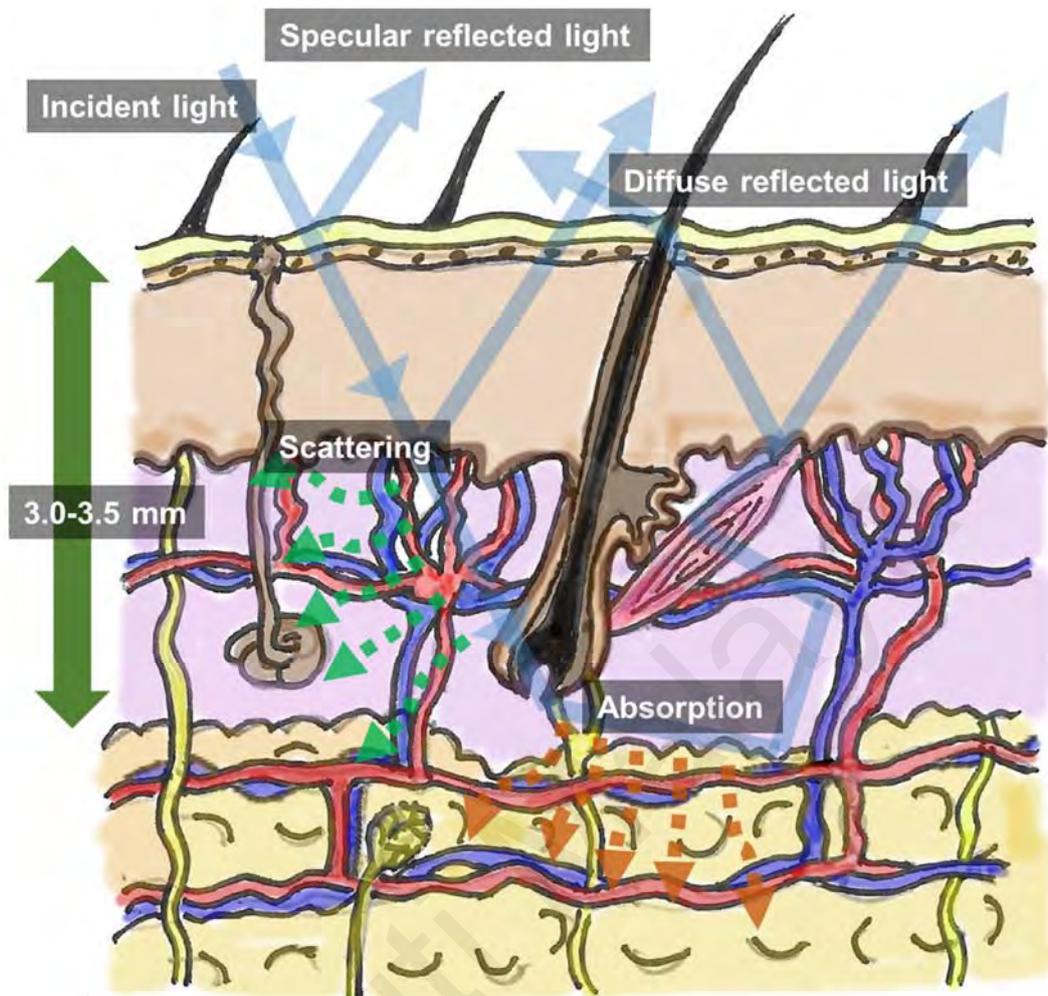
Figure 3.3: Optical pathways through the skin

Quantitatively, a generally accepted method to mathematically represent the interaction of light with the skin is by the values of scattering and absorption coefficients, pictured in **Figure 3.3**. These two processes generally describe the penetration of the incident radiation.

Generally in our case, scattering is defined as the deviation of a path of radiation (a diffuse light source in this instance), whereas absorption is defined as the dissipation of the radiation by the media in the skin, which transforms into other forms of energy, such as heat. A scattered light, in this medium, will continually propagate and undergo several scatterings before absorption of its energy dissipates completely. When a scattering is localized to certain spots in the skin, multiple scatterings will occur before absorption

occurs. Scattering generally occurs when the oncoming radiation in regions where inhomogeneity of the refractive index within the media is dominant. At any given wavelength, scattering is generally negligible for particles with dimensions less than 10% of the wavelength size (Anderson & Parrish, 1981).

To describe a spectroscopic pattern in simpler terms, a generally accepted model used for mathematically describing the human skin in terms of its scattering and absorption components is the Kubelka-Munk theory of reflectance. We consider a simplified version of a human skin cross-section as shown in **Figure 3.4**. Note that to qualify the use of Kubelka-Munk method, we approximate the human skin as a relatively homogenous medium compared to the thickness of the object measured. Also, we assert the incident radiation as diffuse (not collimated or concentrated) and incidental reflection occurring from the skin surface is negligible (Anderson & Parrish, 1981).
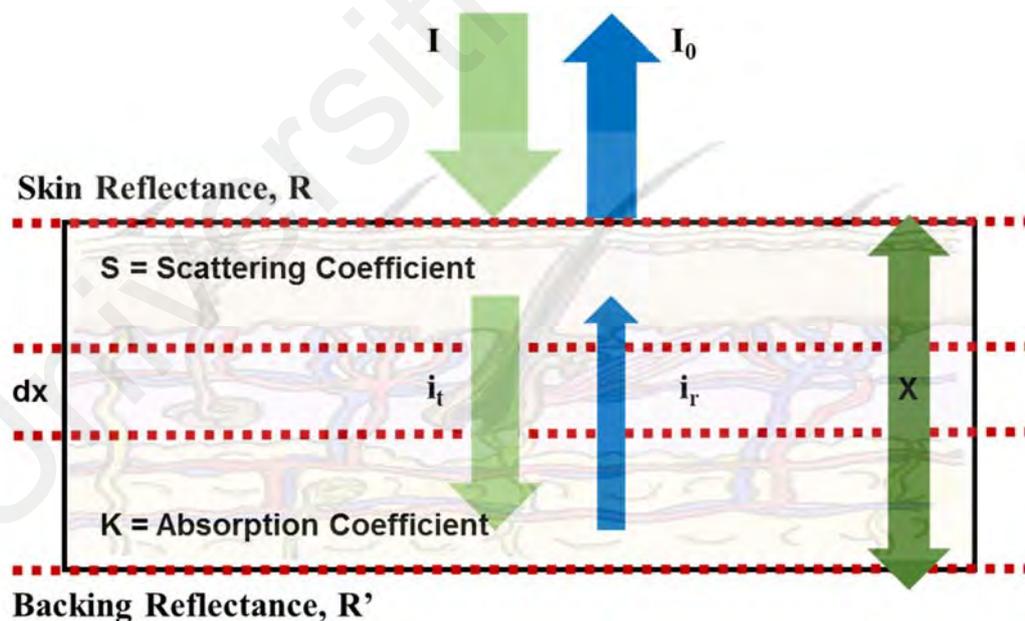
Figure 3.4: Approximate model of the human skin

We elaborate based on combined works of previous authors (Anderson & Parrish, 1981; Kubelka, 1931). This model has undergone several modifications and still is currently used in recent works (C. Li et al., 2018; Masuda et al., 2018). Based on **Figure 3.4**, we denote $\mathbf{I_T}$ and $\mathbf{I_R}$ (with i_t and i_r as the infinitesimal components) as opposing fluxes representing radiation. Reflectance (*R*), in turn, is defined as the ratio I/I0 and transmittance (*T*) is defined as I0/I. The parameters scattering (S) and absorption (K) coefficients is denoted as such. Both of these parameters, which determine the nature of the reflected spectra, are expressed with the following differential equations:

$$-di_t = -(S + K)i_t dx + i_r S dx \tag{3.1}$$

$$di_r = -(S + K)i_r dx + i_t S dx \tag{3.2}$$

By dividing Equation (3.1) with i_t and (3.2) with i_r, we arrive at the following:

$$-\frac{di_t}{i_t} = -\frac{(S + K)i_t dx}{i_t} + \frac{i_r S dx}{i_t} \tag{3.3}$$

$$\frac{di_r}{i_r} = -\frac{(S + K)i_r dx}{i_r} + \frac{i_t S dx}{i_r} \tag{3.4}$$

Adding (3.3) and (3.4) results in the following:

$$\frac{di_r}{i_t} - \frac{di_t}{i_t} = -\frac{(S + K)\cancel{i_t} dx}{\cancel{i_t}} + \frac{i_r S dx}{i_t} - \frac{(S + K)\cancel{i_r} dx}{\cancel{i_r}} + \frac{i_t S dx}{i_r} \tag{3.5}$$

Where it is further simplified as follows:

$$\frac{di_r}{i_t} - \frac{di_t}{i_t} = -2(S + K)dx + S\left(\frac{i_t}{i_r} + \frac{i_r}{i_t}\right) \tag{3.6}$$

Bearing in mind of the identity of the following:

$$\frac{d\left[\ln\left(\frac{i_r}{i_t}\right)\right]}{dx} = \frac{d[\ln(i_r) - \ln(i_t)]}{dx}$$

$$= \frac{\delta\ln(i_r)}{di_r}\frac{di_r}{dx} - \frac{\delta\ln(i_t)}{di_t}\frac{di_t}{dx}$$

$$= \frac{\delta\ln(i_r)}{di_r}\frac{di_r}{dx} - \frac{\delta\ln(i_t)}{di_t}\frac{di_t}{dx}$$

$$= \frac{1}{i_r}\frac{di_r}{dx} - \frac{1}{i_t}\frac{di_t}{dx}$$

Therefore,

$$\frac{d\left[\ln\left(\frac{i_r}{i_t}\right)\right]}{dx} = \frac{di_r}{i_r} - \frac{di_t}{i_t} \tag{3.7}$$

If we define reflectance $R = I/I_0$, incremental reflectance $r = i_r/i_t$ and $d(\ln r) = \frac{dr}{r}$, we

arrive to the following (referring to equation 3.6):

$$\frac{dr}{r} = -2(S + K) + S\left(\frac{1}{r} + r\right) \tag{3.8}$$

After rearranging and integrating with respect to dx:

$$\int_{R'}^{R} \frac{dr}{r\left\{\left(\frac{dr}{r}\right) - 2\left(\frac{K}{S} + 1\right)\right\}} = S\int_{0}^{x} dx$$

$$\int_{R'}^{R} \frac{dr}{1 + r^2 - 2\left(\frac{K}{S} + 1\right)r} = S\int_{0}^{x} dx \tag{3.9}$$

And this allows us to express R in terms of S, K and R' as per the boundary conditions

shown in **Figure 3.4.**

By arbitrarily simplifying $\frac{S+K}{S} = 1 + \frac{K}{S} = a$, equation 3.9 turns into:

$$\int_{R'}^{R} \frac{dr}{1 + r^2 - 2ar} = S \int_{0}^{x} dx \tag{3.10}$$

By employing partial fractions, one can solve the denominator $1 + r^2 - 2ar$ using the quadratic expression:

If $r^2 + 1 - 2ar = 0$

$$r = \frac{2a \pm \sqrt{4a^2 - 4}}{2} = a \pm \sqrt{a^2 - 1}$$

Therefore $r^2 + 1 - 2ar = \left(r - a\sqrt{a^2 - 1}\right)\left(r + a\sqrt{a^2 - 1}\right)$

By putting the right hand side of the equation over a common denominator results in:

$$A\left(r - a\sqrt{a^2 - 1}\right) + B\left(r + a\sqrt{a^2 - 1}\right) = 1$$

By comparing coefficients of r, one can arrive to the fact that

A+B =0 \hfill (3.11)

And by comparing constants between Equation 3.11 and the following:

$$A\left(-a + \sqrt{a^2 - 1}\right) + B\left(-a - a\sqrt{a^2 - 1}\right) = 1$$

One can derive that

$$A = \frac{1}{2\sqrt{a^2 - 1}} \quad ,$$

$$B = \frac{1}{2\sqrt{a^2 - 1}}$$

In turn, Equation 3.10 turns into:

$$\int_{R'}^{R} \frac{1}{\left(r - a - \sqrt{a^2 - 1}\right)} - \frac{1}{\left(r - a + \sqrt{a^2 - 1}\right)} dr = 2\sqrt{a^2 - 1} * S \int_{0}^{x} dx$$

and $\left[\ln\left\{\frac{r-a-\sqrt{a^2-1}}{r-a+\sqrt{a^2-1}}\right\}\right]_{R'}^{R} = 2\sqrt{a^2 - 1} * Sx$

Following substitution of R and R' into r:

$$\ln\left\{\frac{R - a - \sqrt{a^2 - 1}}{R - a + \sqrt{a^2 - 1}} * \frac{R' - a + \sqrt{a^2 - 1}}{R' - a - \sqrt{a^2 - 1}}\right\} = 2\sqrt{a^2 - 1} * Sx$$

And by substituting $b = \sqrt{a^2 - 1}$,

$$\ln\left\{\frac{R - a - b}{R - a + b} * \frac{R' - a + b}{R' - a - b}\right\} = 2bSx \tag{3.12}$$

When we limit the condition for the thickness X and reflectance R to be infinite or large, it will be assumed that X and R= ∞. Since no light is assumed to have penetrated completely the material or skin, we may assume that R' = 0. If the left hand side of Equation 3.12 is equal to infinity, this means the denominator of the said equation is close or equal to 0. Therefore:

$$R_\infty = a - \sqrt{a^2 - 1}$$

$$= 1 + \frac{K}{S} - \sqrt{\frac{K^2}{S^2} + \frac{2K}{s}} \tag{3.13}$$

This can be approximated further into the common expression, for infinite skin thickness:

$$R = 1 - \sqrt{\frac{2K}{s}} \tag{3.14}$$

As shown in Equation (3.14), however, the expression in terms of R alleviates the necessity for obtaining two measurements for obtaining K/S. One can further deduce that for thick samples, which indicates the transmission R' values rendered negligible due to $I_T \ll I_0$, Equation 3.12 becomes:

$$\frac{K}{S} = \frac{1 + R^2}{2R} - 1 \tag{3.15}$$

Rearranging the quadratic terms and factorizing, one arrives to the final and common Kubelka-Munk expression as shown in Equation (5):

$$\frac{K}{S} = \frac{(R - 1)^2}{2R} \tag{3.16}$$

Therefore, we can express remittance solely based on the ratio of the absorbance and scattering coefficients. The following subsections elaborate on the equipment used which manifest the collected spectral patterns of human skin. It is to be noted that in diffuse reflectance measurements, Beer-Lambert's law, though well-known in the use of spectroscopic methods, is not a favorable approach in mathematically expressing the human skin. This is due to the inhomogeneity of the skin's multiple layers, while Kubelka-Munk is generally more inclusive, though less elegant (Anderson & Parrish, 1981).

3.1.2   Integrating Sphere Technique

Since the collection of diffuse spectra in our case employs the use of an integrating sphere technique, a brief treatise on the latter is provided in this section. An integrating

sphere is mainly used to integrate radiant flux which originates from diffuse lighting in our case. Consider a hollow sphere with highly reflective material coating on the inside, with entrance and exit ports, for interaction with the human skin and the coupling fiber respectively. **Figure 3.5** illustrates an integrating sphere diagram.
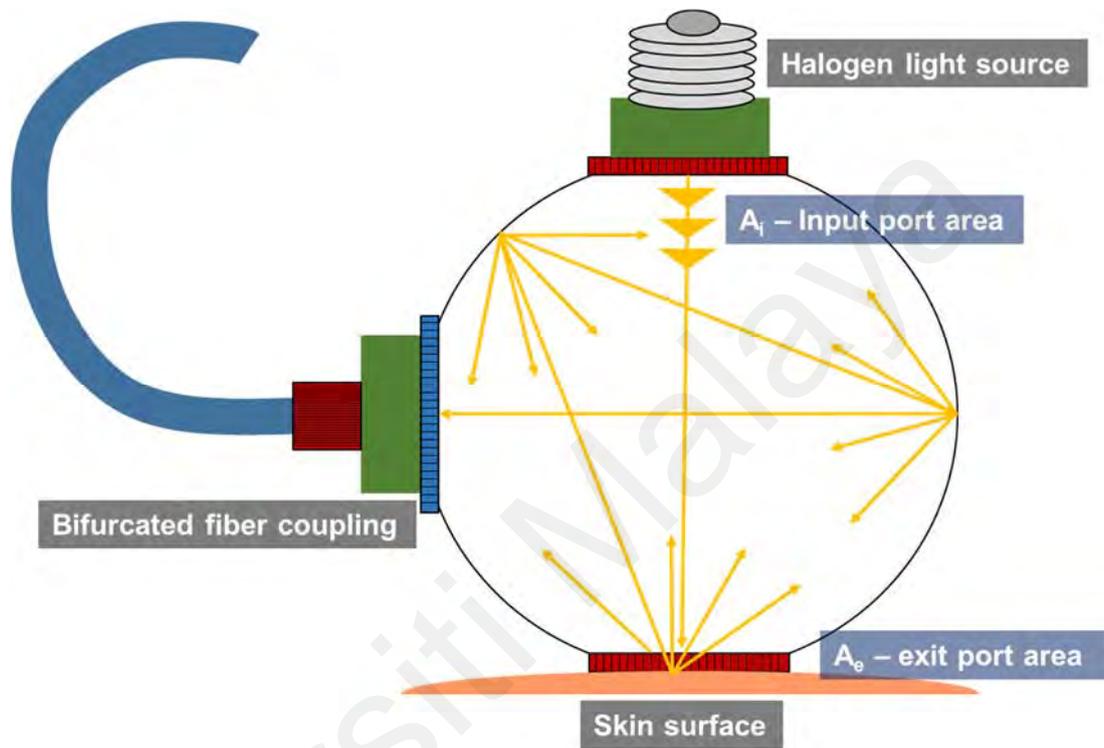
Figure 3.5: Integrating sphere diagram

The theory on integrating spheres is derived from theories revolving exchange of radiation between diffuse surfaces within an enclosed space. Consider a radiant flux from the light source, reaching all surface in the sphere. Since the reflective material in the integrating sphere is diffuse, light is reflected in all angles, including the sample port as well. This perpetual process of reflection, after very brief moment, is integrated over all directions into the sample (in this case, human skin), which receives light over all possible directions. Therefore, it can be surmised that an integrating sphere facilitates a spatially uniform light source to the sample. This form of diffuse light is best expressed in the term *radiance*.

Radiance, L (SI unit Watt per steradian square meter or W. sr/m^2) is mathematically expressed as follows:

$$L = \frac{\Phi_i \rho}{\pi A} \text{ (W. sr/m}^2) \tag{3.17}$$

Where Φ_i is the input flux, ρ the reflectance, and A the illuminated surface area.

To express the total radiance in an integrating sphere with a total area A_s, with an input port area A_i and exit port area A_e as shown in **Figure 3.5**, the input flux radiance is expressed as follows:

$$L = \Phi_i \rho \left(\frac{A_s - A_i - A_e}{A_s} \right) \tag{3.18}$$

Equation 3.18 expresses the flux radiated on the surface which is not absorbed by the port openings. For simplification, the parenthesized part in **Equation 3.18** is expressed as (1-f), where f represents the fraction of radiant flux by omission of the contribution of the port areas devoid of reflective material in the integrating sphere. After a second successive reflection, this results in **Equation 3.19**, the input flux transforms into the following expression:

$$= \Phi_i \rho^2 (1 - f)^2 \tag{3.19}$$

Similarly, the following occurs for a third reflection:

$$= \Phi_i \rho^3 (1 - f)^3 \tag{3.20}$$

After infinite cycles of the reflection, we obtain the following series:

$$= \Phi_i \rho (1 - f)\{1 + \rho(1 - f) + \cdots + \rho^{n-1}(1 - f)^{n-1}\} \tag{3.21}$$

By employing the power series, and by extension since r(1-f) < 1, the final radiant flux expression above can be simplified as follows:

$$= \Phi_i \rho \left(\frac{1-f}{1-\rho(1-f)} \right) \qquad \text{(in Watts)} \tag{3.22}$$

The equation above shows that the total flux which permeates all around the sphere is higher than the input flux as shown in **Equation 3.22**, a result of perpetual reflections occurring inside. This results in the expression of the sphere surface radiation as follows:

$$L_s = \frac{\Phi_i \rho}{\pi A_s (1-f)} * \left(\frac{1-f}{1-\rho(1-f)} \right) \tag{3.23}$$

And finally in its well-known form as shown in the following:

$$L_s = \frac{\Phi_i \rho}{\pi A_s} * \left(\frac{\rho}{1-\rho(1-f)} \right) \tag{3.24}$$

After the diffuse light reaches the skin, the reflected light from the skin in turn is reflected again in the sphere, forming a spatially uniform reflection which then reaches the detector port, in our case the common side of a bifurcated fiber connected to spectrometers. The collected spectra in turn encompasses both specular and diffuse reflection components.

Calibration of the reflectance scale is performed by comparing the incident flux with a diffuse reflection standard, made from highly reflective material such as Spectralon$^{\text{TM}}$ or Polytetrafluoroethylene (PTFE). The mathematical relationship between the reflection

standard and the sample (human skin) is then established as shown in **Equation 3.22**, that is the ratio of both radiance and reflectance of the sample and reflectance standard.

$$\frac{L_s}{L_r} = \frac{\rho_s}{\rho_r}$$

(3.25)

Subscripts r and s corresponds to sample and reflectance standard respectively. In substitution schemes used in our measurements, when the reflectance standard shifts to the human skin sample, the true expression of the radiance ratio is shown in a more accurate model:

$$\frac{L_s}{L_r} = \frac{\rho_s}{\rho_r} * \frac{1 - \boldsymbol{\rho_s}}{1 - \boldsymbol{\rho_r}}$$

(3.26)

Where both $\boldsymbol{\rho_s}$ and $\boldsymbol{\rho_s}$ (bolded) is the average wall reflectance with sample and reference respectively. In our case, averaged readings are automatically configured in the software.

3.2 Spectrometry Setup and Initialization

The setup in the following combines methods from several setups in literature (Cooksey et al., 2015; Fang, Fu, & He, 2016). **Figure 3.5** shows the setup diagram.
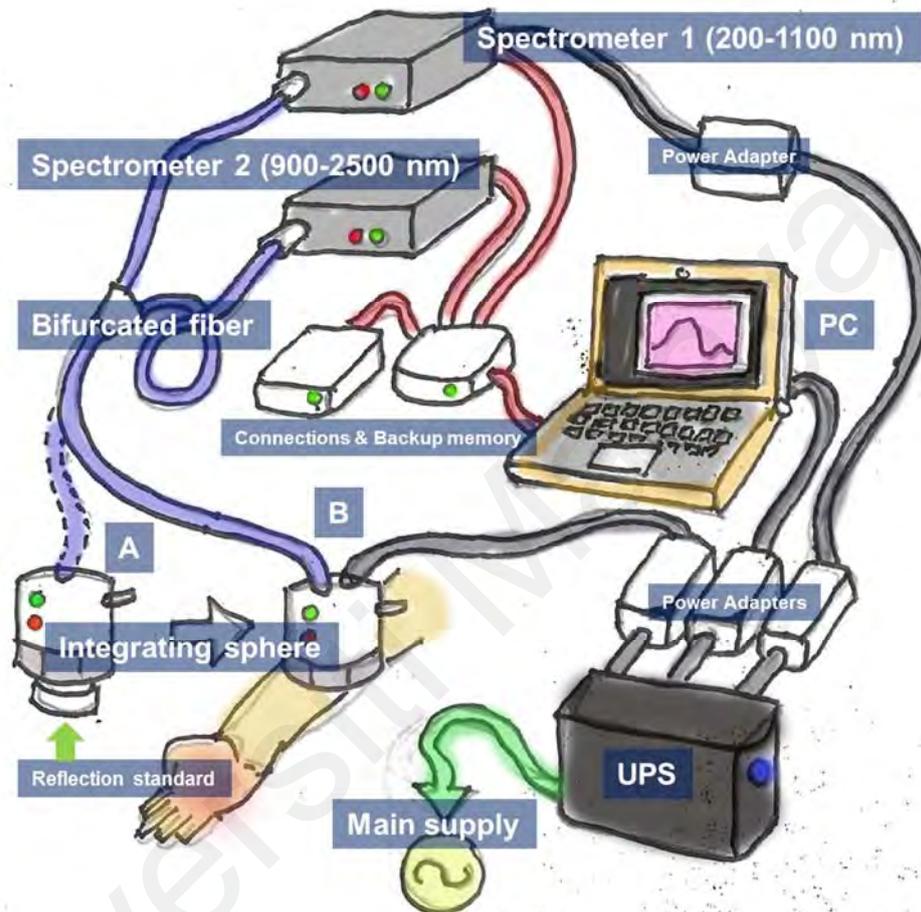
Figure 3.6: Diffuse reflectance spectroscopy setup

As shown in **Figure 3.6**, the setup is described as follows. The setup is perhaps best described by beginning with the integrating sphere (AvaSphere-50-LS-HAL-12V, Avantes), that acts, in our case, as a patient skin probe. The probe has an opening port with a 10mm diameter for contact with the human skin, an internal diameter of 50 mm, and 0.1% light source drift. The integrating sphere's halogen light source (bandwidth 360-2500 nm) is turned on, and left to stabilize for a few minutes. A reference spectrum which establishes an approximate 100% reflectance is obtained with a reflection standard (WS-

1-SL, Ocean Optics), as shown by the transition *A* in **Figure 3.6**. The reflection standard has a reflectivity of 96-99% for wavelengths between 250-2500nm. Next, dark measurements (0% reflection baseline) are taken where no light enters the integrating sphere port by turning off the light source, and closing the port with a black surface[1].

This is promptly followed by a measurement on the forearm of the subject as shown in transition *B* in **Figure 3.6**. Upon the interaction of light source on the skin of the forearm, the diffuse reflected light propagates through a Y-split hex-nut sub-miniature assembly (SMA) bifurcated fiber and split according the wavelength region accorded to both spectrometers. Both spectrometers are connected via the bifurcated fiber cable. The reflected spectra is collected by both spectrometers, where spectrometer 1 (*Avaspec-3648*, Avantes) collects the spectrum between 173-1100 nm (~0.3 nm resolution), and spectrometer 2 (*NIRQuest 256-2.5*, Ocean Optics) covers 900-2000 nm (~6.5 nm resolution).

For the use in a health center, a rugged setup is required. This necessitates placing of all the equipment in a casing which allows only the integrating sphere to extend from the setup. **Figure 3.7** shows the setup assembly in an acrylic casing, enabling mobility for the data collection to commence efficiently.

---

[1] There are configurations which stresses upon turning the light source on in a perfectly dark room for dark measurements. This is not always the case, as a dark room is not available in hospital conditions. The next best configuration is to settle with turning off the light source, as performed by some researchers as cited.

Figure 3.7: Setup in secondary triage area. The inset is the reflection standard.

As shown in **Figure 3.7**, a netbook computer is used to operate the three core software systems. This setup is also designed to be able to move around different areas in University of Malaya Medical Center (UMMC), Kuala Lumpur, where patients residing in other areas may be recruited. The trolley serves as the docking area for all items required for the measurements to be carried out, including the setup, patient information forms and consent forms as well. The setup is stationed mainly in the Secondary Triage area in UMMC for patient recruitment during the data collection sessions.

3.3 Patient Management

Ethical clearance was obtained from UMMC for data collection within the center's premises, approved by the UMMC Medical Ethics Committee (reference number 201511-1904). The subjects recruited in this study mainly consists of triaged patients in the Secondary Triage Area in UMMC. The reason of selecting this area is to maximize the recruitment numbers, due to the higher traffic of patients in this area, where the

hospital requires their vitals to be recorded before assigned to a medical officer for further check-up. During this time, temperature, blood pressure, blood oxygenation (SpO_2), pulse rates, respiratory rates are mainly obtained. During this time, the medical assistants (MA) who were responsible for the triaging process will alert the data collector if the patient fits the recruitment criteria, which will be elaborated in the following.

3.3.1 Inclusion/Exclusion Criteria

Patients above the age 12 are considered eligible for the study, if presented with the following signs or symptoms. Patients with a chief complain of fever, although with no documented temperature ($=>37.5°C$) or with the documented temperature will be included. Also, patients with no fever, but presented other dengue signs such as fatigue, nausea and others will be included. The reason for these criteria is due to fever being one of the most common symptom for both primary and secondary dengue infections. Patients who require immediate medical attention upon triaging process will be excluded.

3.3.2 Patient Recruitment Process

Figure 3.8 shows a typical scenario of the data collection activity from the patient recruitment to the end of the dataset consolidation.
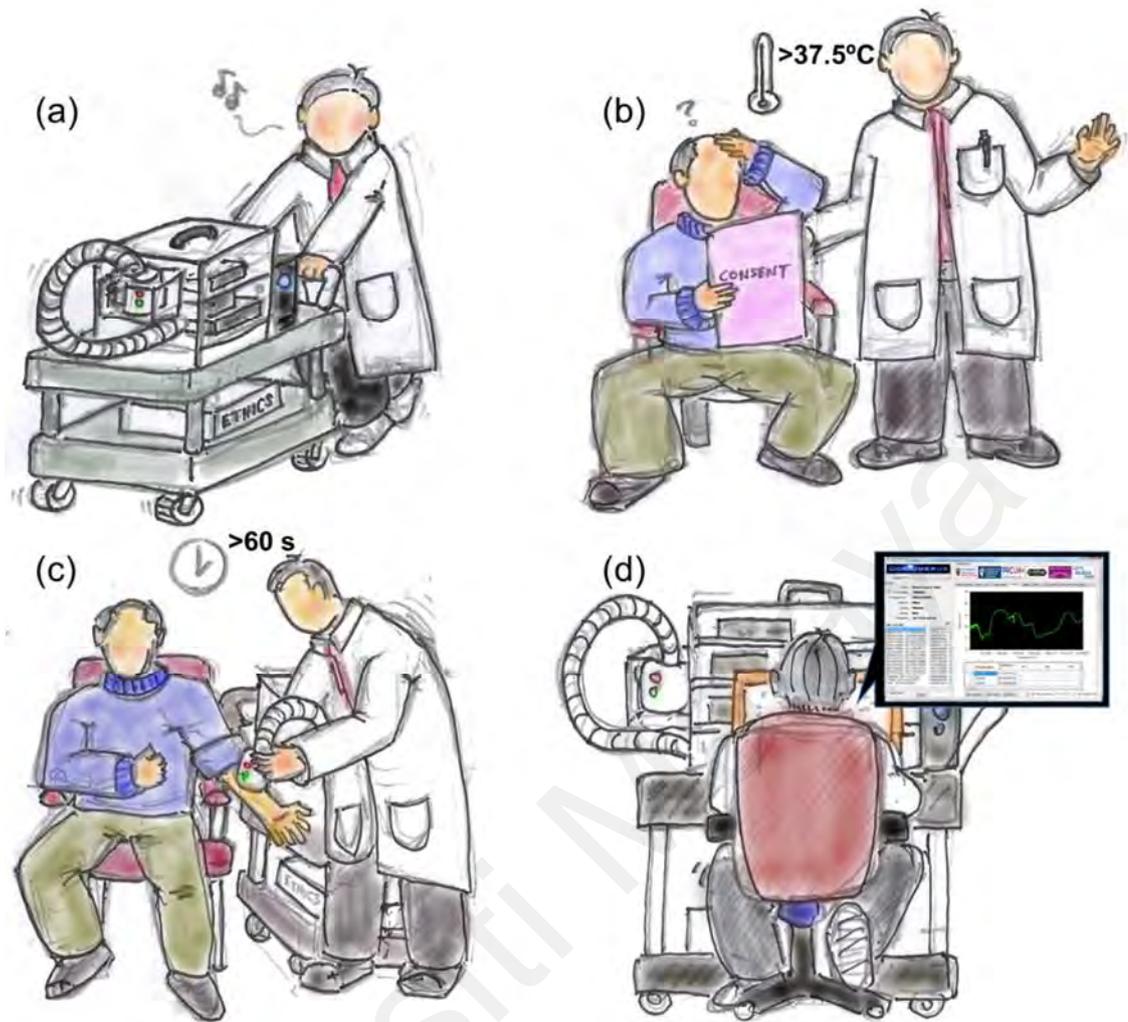
Figure 3.8: Patient recruitment and data consolidation

As shown in **Figure 3.8(a-b)**, after commuting the setup to the recruitment area, a triaged patient typically with fever or dengue symptoms are directed to the vicinity. Next, written consent is obtained either from the patient or guardian, if the patient/guardian agrees for the data collection. The patient will be seated and requested to bare the forearm with the ventral side facing upwards. **Figure 3.8(c)** shows the use of the integrating sphere on the subject's forearm for data collection. This phase takes about 60 seconds to complete. Next, completion of the clinical data is performed, including the vitals, dengue assessment, and blood results as shown in **Figure 3.8(d)**. The data aspects will be covered in the next section.

3.4   Software and Data Management

There are three core software involved, of which two are related to the spectrometers and one is custom-made for data key-in. The details of these software such as parametric settings and formatting will be discussed. Details of data analysis, however, will be discussed in the following chapters due to the foundational nature of this study that requires the data to be observed collectively before an analytical method is decided upon.

3.4.1   Data Key-in and Synchronization

Spectra from both spectrometers are collected in reflectance mode via software provided by respective manufacturers. For spectrometer 1's software, (*Avasoft 8*, Avantes) is set to an integration time of 4.18 ms. Spectrometer 2 (*OceanView*, Ocean Optics) in turn is set to an integrating time of 100 ms. It is to be noted that these differing values are not of any concern for the measurements since they are kept constant throughout the operation, and these values are set for the least noise spectrum. Therefore we maintain the consistency of the data.

During measurements upon recruited patients, upon generating the reflectance values versus wavelengths, three datasets are generated for each spectrometer, totaling six data files. This is done for precautionary reasons, although averaging may also be beneficial but not necessary due to the readings from the spectrometer are already averaged. *Avasoft 8* generates Excel (**.xlsx**) files, while *OceanView* generates text (**.txt**) files. All original data files are saved during the data collection before consolidation with the main patient dataset. The consolidation method is elaborated in the next subsection.

3.4.2 VB.NET Development

For data consolidation, a graphical user interface (GUI) was developed using VB.NET via Visual Studio 2013 (VS2013). The design and deployment of this application was due to the vast amount of information which requires at least the patient ID to create a profile. **Figure 3.9** shows the GUI design. For the sake of brevity, we discuss only salient features of GUI development, as in-depth algorithms for analytics developed are discussed in following chapters.
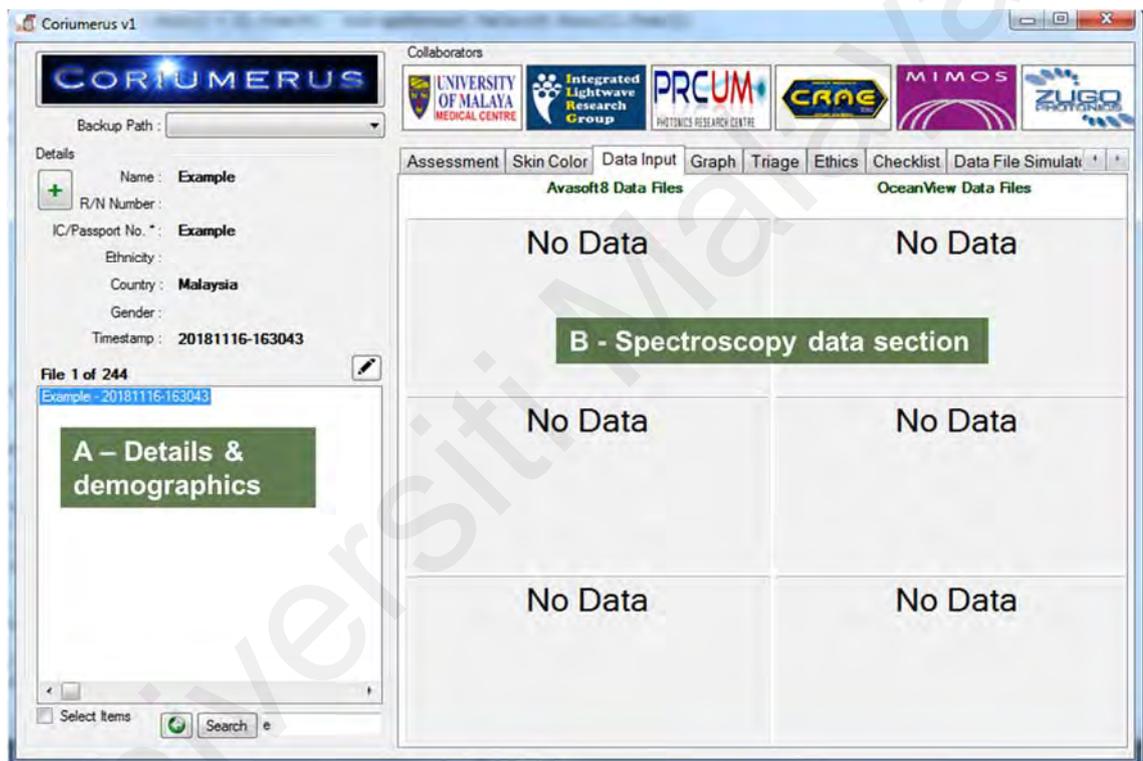
Figure 3.9: Graphical user interface (GUI) for data collection

Figure 3.9 is briefly described as follows. The GUI is described as a version 4.5 .NET framework that enables connectivity with all features on Microsoft platform including Microsoft Office features such as Excel, Word and others. The GUI consists of two main sections, with demographics and a list to navigate the data as shown in section *A* in **Figure 3.9**. Section *B* comprises of several tabs. The main tabs in this design is centered on the "data input" tab as shown in Section *B*, with other tabs such as Graph, Checklist, and Data File Simulator. However, these tabs are intended for auxiliary purposes in the dataset,

enabling data verification and confirmation later. This is due to the manual keying-in of each parameter, which may result in errors or data loss.

During recruitment, a profile of the patient for the patient database is created using VB.NET's Microsoft Office Interop function that converts all the data generated from the measurement sessions into a single Excel (.xlsx) file. After saving the data from *Avasoft 8* and *OceanView*, the data files are then dragged into each box representing both spectrometers' data, which prompts the saving of each files into the patient folder. Following a successful backup, the respective box shows a tick (✓) sign. Error-keying in is programmed for precautionary measures by detecting file type in each section.

3.4.3   Data Topology

As the dataset requires a clear and rigid structure to ease future scripting, the topology of the latter needs to be defined with clarity. We begin with the spectroscopic data. The values from the reflectance spectroscopy from the Excel (from Spectrometer 1) and text files (from Spectrometer 2) are read with the Object-Linked Environment Database (OLEDB) routine in VS2013, where a graphical representation can be obtained from the OLEDB paired with VS2013's Graph Control immediately. A satisfactory measurement is based on the normative appearance from the Graph Control in VB.NET. **Figure 3.10** shows a typical spectra prior to synchronization with the main dataset.
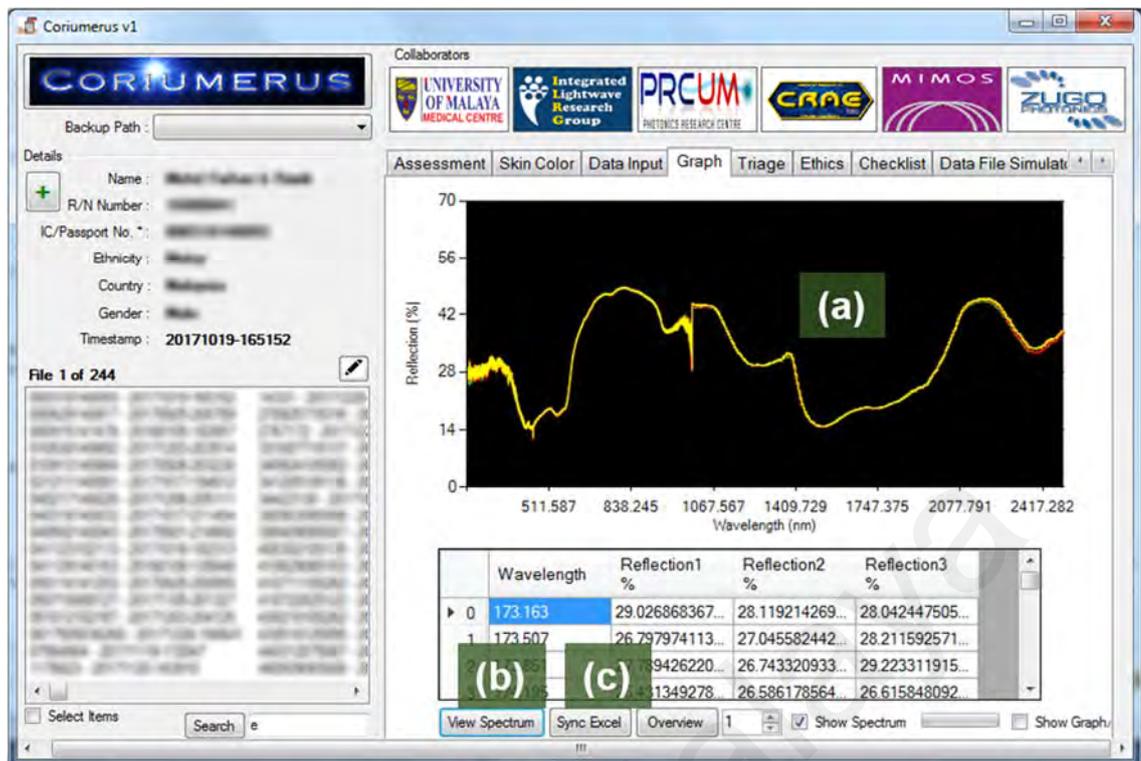
Figure 3.10: Graph control and data synchronization.

A satisfactory spectra example is as shown in the graph area in **Figure 3.10(a)** after triggering the graph function (*"View Spectrum"*) from the control in **Figure 3.10(b)**. A synchronization with the main data profile is then commenced by the "*Sync Excel*" button in **Figure 3.10(c)**. This, in turn, transfers the reflectance percentages versus wavelength values from both spectrometers into the dataset, spanning 2,974 rows of data for each profile.

After spectroscopic measurements were obtained, other details pertaining to patient demographics, vitals, dengue assessment, and blood results are keyed–in. **Table 3.1** shows the breakdown of the items recorded to form the dataset.

Table 3.1: Data structure

Classification	Details
Demographics/Essentials	- Name - ID - Ethnicity - Gender - Nationality - Timestamp - Spectroscopy data
Vitals	- Temperature - Blood Pressure - Respiratory rate - Blood oxygenation (SpO_2) - Pulse Rate

Dengue Assessment	- **Date of fever onset**
	- **Symptoms**:
	   i. Generalized body ache, myalgia, arthralgia, sore throat, conjunctival injection, anorexia, nausea and vomiting.
	 - **Risk factors**:
	  ii. Infant (<1 year old), hypertension, obesity, pregnancy, diabetes mellitus, heart disease, renal or liver failure, >60 years of age.
	 - **5-1 Magic Touch**:
	   iii. Colour of extremities, capillary refill time, temperature of extremities, temperature of extremities and pulse volume
	- **Three golden questions**:
	(a) Could drink 3-4 glasses of water in the last 12 hours,
	(b) Passed urine at least twice in the last 12 hours
	(c) Able to walk around in the last 6 hours.
	- **Danger signs**:
	(a) Severe abdominal pain
	(b) Vomiting >3 times
	(c) Weakness/Lethargy/Confusion
	(d) Mucosal bleeding, including heavy menstruation
	(e) Cold or pale hands/feet
	(f) Breathing difficulties or chest pain
	(g) Dizziness or fainting

Blood/serology data	- **Full blood count** (a) Haemoglobin (b) Haematocrit (c) White Blood Cell Count (d) Platelets - **Liver Function Test** (a) Aspartate Aminotransferase (AST) (b) Alanine Aminotransferase (ALT) - **Coagulation** (a) Activated Partial Thromboplastin Time (APTT)

	(b) Prothrombin time and International Normalized Ratio (PT INR)
	- **Renal Function Test**
	(a) Creatinine
	(b) Urea
	(c) CO_2
	- **Blood Gas Analysis**
	(a) Blood pH
	(b) Bicarbonate Acid (HCO_3)
	(c) Lactate
	- **Virology and Serology**
	(a) NS1
	(b) IgM
	(c) IgG
Final diagnosis	- Statistical Category
	- Diagnosis
	- Briefing on diagnosis

All triage and clinical data are obtained via the UMMC patient system database under strict regulations. Also, particulars of the patient such as name and ID are further encrypted and redacted in the compiled dataset to preserve the patients' privacy. Although each parameter as shown in **Table 3.1** is a viable parameter to be used for the analytics in this study, not all patients are expected to undergo dengue assessment or specific blood tests due to lack of clinical presentation of dengue. However, these parameters are recorded for the assessment by clinicians.

A data key-in in this stage is considered complete after the blood or serology data has been released for viewing in the system. This normally takes under 72 hours. Meticulous attention will be given for patients with dengue serological readings, which include NS1, IgM or IgG. After this period, if no blood data is released that traces back to the time when the spectroscopy data is taken (with a time window of <3 hours), the data is archived into cloud storage to depopulate the list of patients in the GUI. This enables the data collector to focus his/her attention on the remaining patients with pending clinical data.

3.5   Clinical Discourse

In this section, the final diagnosis for the patients recruited will be decided, where dengue patients will be classified either as confirmed dengue or probable dengue, in line with clinical practice guidelines. Other patients recruited in our study who does not fall into the two latter categories will be classified as control non-dengue patients.

It is to be noted that the conventions of the 1997 diagnostic categories, including dengue fever (DF), dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) (Srikiatkhachorn et al., 2011) is not employed for statistical analysis, though used for diagnosis. Also, the more recent 2009 classification, consisting of dengue with warning signs (DWWS), dengue without warning signs (DWOS) and severe dengue (SD) follows the same fashion (Srikiatkhachorn et al., 2011). This is due to the nature of the objectives in this research, which seeks to confirm a dengue infection rather than classify the nuances. These conventions may change over time due to various technical reasons as demonstrated by the two aforementioned classification style.

Rather than relying on golden standard laboratory tests such as the RT-PCR method or HI assays for dengue confirmation, we defer to the standard clinical diagnosis. This is due to lack of resources required for such a scale for establishing a dengue patient database.

The following subsection elaborates this aspect in an in-depth perspective.

3.5.1   Classification of Dengue Patients

Clinical confirmation of dengue cases is performed by an Infectious Diseases (ID) ward physician[2]. This is carried out by examining the clinical notes of each of the patients that were accessed from the patient folders under strict access by the physician.

A patient is considered confirmed dengue if the NS1 antigen results returned positive, with clinical consistency such as day of fever onset <6 days, and presenting symptoms of acute dengue infection (Organization et al., 2009). NS1 seropositivity is unlikely after day 7 of illness. However, if the NS1 readings are negative between days 1-3 of the illness, it is unlikely to be dengue and will be diagnosed according to other criterions causing the illnesses.

A patient is also considered confirmed dengue if the IgM antigen readings returned positive, provided the day of fever onset is more than 5 days, combined with serological features such as leukopenia (low white blood cell count) and thrombocytopenia (low platelet count) (Organization et al., 2009). This is due to the possibility of the presence of IgM antibody within 60 days after convalescence from a dengue infection (Rigau-Pérez et al., 1998).

Another scenario where dengue is confirmed is if IgM readings are negative, and consequently becomes positive after repeated serology tests (Organization et al., 2009).

Patients with IgG-positive results, however, will be approached with caution. This is due to the high percentage of IgG seropositivity among Malaysians, at 91.6%, indicating that the majority of Malaysians have had past dengue infections, with lifelong availability of the IgG antibody (Azami et al., 2011). However, examination of clinical notes may

---

[2] The clinical evaluation was performed by Assoc. Prof. Dr. Syarifah Faridah Syed Omar, specialist in the Infectious Diseases and Medicine department, University of Malaya Medical Centre, KL

allow a confirmed or probable dengue diagnosis if serological data is unavailable from UMMC database, clinical referrals may report on NS1 or IgM positivity from other health institutions. However, IgG-positive patients will be classified as confirmed dengue if the IgG seropositivity is differentiated from primary or a previous infection, with a positive NS1 or IgM combined with clinical consistencies as stipulated before.

A patient is considered to have probable dengue if the NS1 is negative, and IgM positive but with a below day 5 fever onset, with clinical consistencies (Organization et al., 2009). This is due to the seroconversion phase for the production of IgM which normally takes 5 days to be at detectable titer levels.

A negative IgM reading after day 7 or 8 of the illness is considered to be a non-dengue case (Organization et al., 2009), and will be diagnosed according to non-dengue diseases as described in the following.

3.5.2   Classification of Non-dengue Patients

Patients which undergo other clinical examinations, and ultimately diagnosed as non-dengue illnesses are considered as control cases. This includes all forms of illnesses that results in the febrile state. For these classifications, a physician from the Emergency Department, UMMC examined the clinical notes for a final diagnosis[3].

For tabulating the data, the original diagnoses are maintained, with a section on the justifications on each diagnosis given.

---

[3] Confirmation of non-dengue diagnoses was performed by Dr Khadijah Poh, a physician from Emergency Department (ED), University Malaya Medical Centre, Kuala Lumpur.

3.6 Chapter Conclusion

As per the chapter map in **Figure 3.1**, this chapter is concluded with the following. The process began with the operating principles of diffuse reflectance spectroscopy (DRS) by highlighting a well-known theory by Kubelka-Munk. The mentioning of this theory is to ground the experimental aspects of the interaction of light with human skin in a mathematical form. This can be expressed in the scattering and absorption coefficients governing the spectral pattern.

However, the approximations are not wholly accurate, nor they are elegant. This is due to the nature of the human skin, comprised of complex layers with different optical properties. However, Kubelka-Munk theory approximates the mathematical model reasonably well. The next theoretical aspect covered is the integrating sphere, which is the main conduit of obtaining highest accuracy of diffuse skin reflectance possible. Mathematically, the diffuse radiant flux shows higher combined radiant flux compared to the input radiance. This demonstrates the advantages of the technique. Next, we covered the spectroscopy setup, comprising mainly of an integrating sphere coupled with the two spectrometers, which are housed with other electronics components in a rugged acrylic casing.

Following elaboration of the setup, we discuss the methods of recruitment, from ethics, inclusion and exclusion criteria, and data collection flow. The setup is used to collect diffuse reflectance spectra from triaged patients, where other data pertaining to the patients are collected. This includes the patient's particulars (though to be encrypted later), vitals, blood results, dengue assessment, and dengue serology, if performed.

Next, we cover data management schemes where data is organized and consolidated with a custom-built GUI for a rigid database for ease of scripting and manipulation. This includes the use of a .NET framework setting for creating patient profiles, Microsoft

Excel data files, and a built in verification system via graphical visualization. All these data are compiled from both spectrometers into the main patient profile folder.

Finally, the clinical discourse is covered, for validation of the patients diagnosis, covering three categories, namely for confirmed and probable dengue, and non-dengue control. All these tasks are performed by trained physicians in UMMC. All these data are later incorporated into the dataset for further statistical analysis.

The next chapter discusses the results which are structured around the methods outlined in this chapter.

CHAPTER 4: RESULTS - PATIENT DEMOGRAPHY, PROCURED SPECTROSCOPY, AND CLINICAL SUMMARY

This chapter mirrors the previous one by presenting the *prima facie* (first face) of three main outputs from the data collection exercise, as outlined. These three result sections are (a) the demographics of recruited patients including gender, ethnicity, and others, (b) reflectance spectroscopy data obtained, based on demographic categories, and finally (c) the clinical conclusions on each patient recruited. **Figure 4.1** illustrates the flow as mentioned.
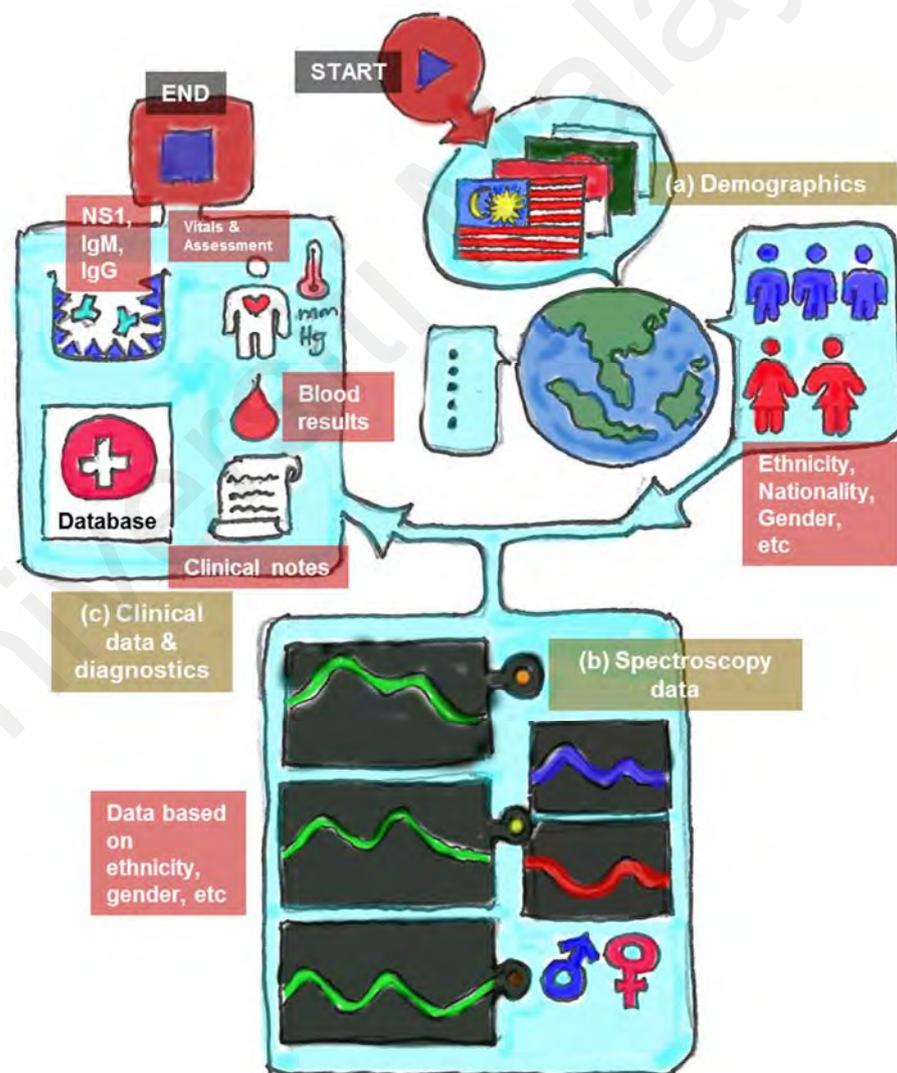
Figure 4.1: Chapter structure

4.1 Data Collection: Sojourn to Adjourn

Data collection sessions were carried out from 27 April 2017 until 16 January 2018, within approximately 8 months. 257 patients were recruited within that period. Recruitment in the Secondary Triage Area (STA), UMMC vary between 10-0 patients per day, mostly during the evening between 3.00 pm to 12.00 am, though peak hours usually are 9.00 to 11.00 pm. Other factors also at play in many days viable for data collection, where the operation of the STA may be shifted to other departments due to technical reasons, causing cessation of the data collection in that day. The breakdown of the demographics is provided in the following section.

4.1.1 Demographics

Out of 257 patients, 240 patients were deemed fit for inclusion into the data analysis, after exclusion of 17 patients with faulty spectroscopic data.

We compare the data with current Malaysian demographics (Malaysia, 2016). 217 (90.4%) of the recruited patients are Malaysians while 23 (9.5%) are non-Malaysians. Of all the patients included, males are represented quite accurately at 124 (51.7%), compared to 114 (48.3%) females.

Patients were of 12-90 years of age. 8 (3.3%) are aged between 12-14 years old, 202 (84.2%) are aged between 15-64 years old, and the remainder 30 (12.5%) are aged above 65 years old. The median age is 32 years old.

114 (47.5%) are Malays, 59 (24.6%) are Indians and 39 (16.3%) are Chinese. Other subjects comprised of 3 (1.3%) East Malaysians (Sabah and Sarawak) and 2 (0.8%) Punjabis. The Malays, defined as per the identified ethnicity based on available documents, are well-represented based on the demographics of Malaysia. The Chinese,

however are underrepresented based on current Malaysian demographics at 24.6%, and Indians are overrepresented with the latest census at 6.9%.

The remainder 23 patients include 7 (2.9%) Bangladeshis, 6 (2.5%) Indonesians, and 10 (4.1%) from China, Algeria, Nepal and other Asian countries. **Figure 4.2** illustrates the breakdown of the demographics.
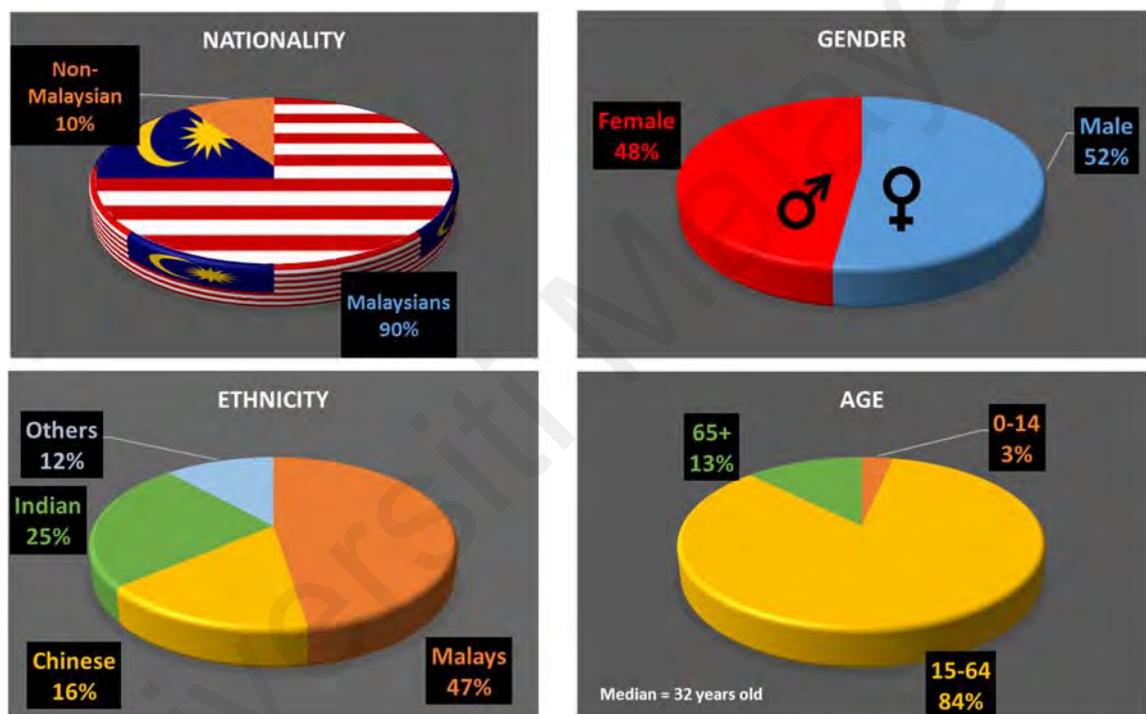
Figure 4.2: Demographics

4.2 Raw Spectrum – Reflectance Spectroscopic Data: Based on Demographics

This section presents the collected diffuse reflectance spectroscopy data from different Malaysian ethnicities, and the two genders. Salient points will be mentioned from the spectra. The main reason for presenting the data as such is to ground it in terms of primary observation (what can be noticed from a simple visual analysis), before in-depth statistical analysis. Also, these two themes (ethnicity and gender) serve as normalization factors, as

will be discussed. Most of the data presented here, however, are foundational for Malaysians within 200-2500 nm. Therefore, comparison of the data with published literature will be performed moderately to avoid potentially obfuscating the discussion.

It is to be noted that conventional in-depth tissue optical properties are not computed from the spectral patterns. This also includes the use of values of scattering or absorption coefficients (μ_s/μ_a or S/K) to describe the optical properties of the skin. This is due to the conceptualism behind the objectives for the sake of brevity. It is in these lines that we pivot the observation of the spectrum obtained in terms of raw reflectance percentages[4], versus what influences the patterns in three main regions, namely the ultraviolet (UV) between 173-380 nm, visible (VIS) between approximately 380-700 nm and near infra-red (NIR) between 700-2500 nm.

All reflection spectra in the following sections are reflection values (in percentages, %) versus wavelengths (nm). All graphs are rendered using Microsoft Graph Control in VB.NET via VS2013.

---

[4] This was done under consultation with Dr Catherine Cooksey, National Institute of Standards and Technology (NIST), Maryland, US.

4.2.1 Raw Spectrum – Gender

In another configuration, **Figure 4.3** shows the gender-based spectroscopic data, which serves as a normalization factor.
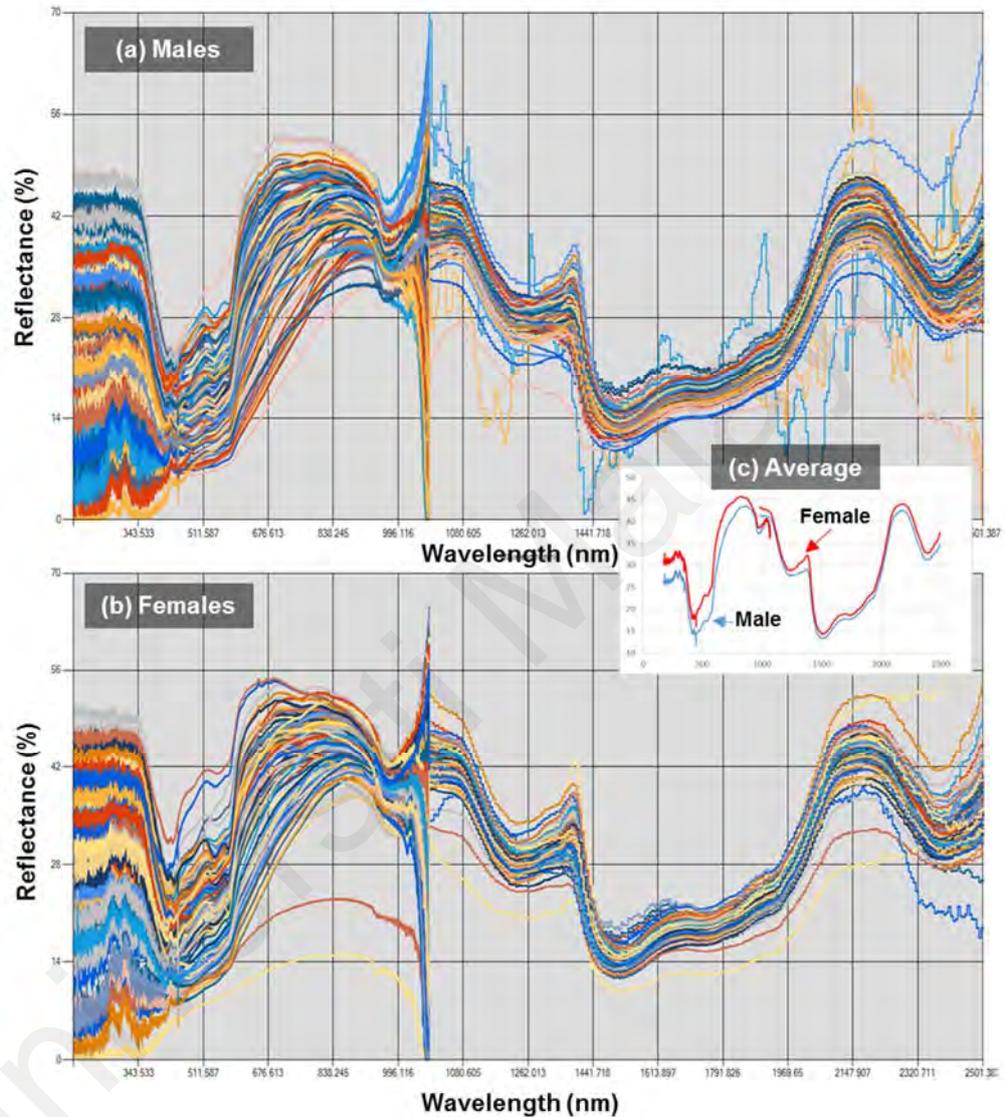
Figure 4.3: Spectroscopic data of (a) males, (b) females and (c) average

As shown in **Figure 4.3**, it can generally be seen that although the distribution of reflection percentages across skin types appear to be uniform, subtle patterns can be observed from the averaged values. The main differences from the spectrum which can be observed mainly occurs in several segments in the UV-VIS region, mainly between 170-350 nm and 400-900 nm. It also appears to be significant in the NIR region between

1200-1400 nm. Females seem to collectively have higher reflectance values, but which depends on the ethnicity.

The differences between the gender spectrums, in our case, is dependent on a pattern analysis. The significance of these initial findings will be elaborated in the discussion section.

4.2.2   Raw Spectrum – Ethnicity

Due to the lack of literature on reflectance spectroscopy of Malaysians in general, we extrapolate the correlation between the spectra and the skin phototype (SPT) or more known as the Fitzpatrick skin types (covering type I-VI) of the major ethnicities in Malaysia in literature. This includes the Malays, Chinese and Indians, including other less represented groups in the data pool as well. Skin type I (generally in Caucasians/Anglo Europeans) results in highest reflectance readings, and skin type VI (generally in dark-skinned people) results in the lowest reflection levels.

4.2.2.1 Raw Spectrum – Malays

Figure 4.4 in the following compiles the Malay spectra.
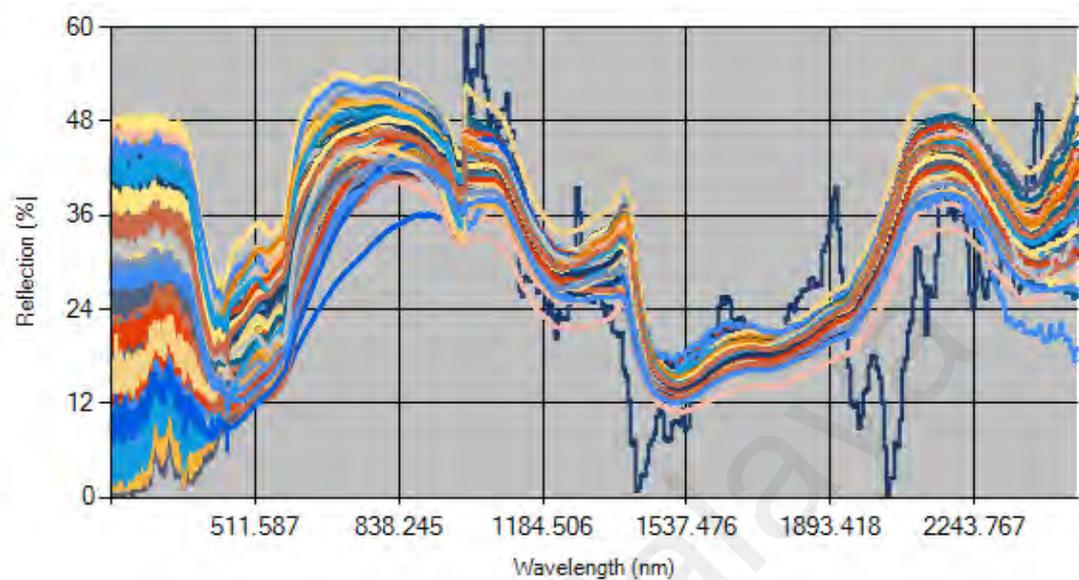
Figure 4.4: Spectroscopic data of Malaysian Malays

As shown in **Figure 4.4**, the spectra presents several points. The first is the apparently high variance in the UV-VIS and a small part of the NIR region between 173-1100 nm. The scattering and absorption which occurs in part of this region has been attributed to the melanisation of the skin (Pena, Strupler, Boulesteix, & Schanne-Klein, 2005), a product mainly from the reaction from the sunlight's UV, while physiological makeup also determines the pervasion (Lerner, Shizume, & Bunding, 1954).

The second noticeable feature of the spectra is the seeming cutoff pattern between 996-1000 nm. This is due to the concatenation and slight disparity of two different spectrometers' readings, which is to be expected. However, the equipment uncertainty in this region is offset by readings of Spectrometer 2, which compensates for loss of the noise region. This accounts for all spectroscopic patterns discussed in this work.

Reports on melanin content in Malay skin is scarce. However, since Malays are generally classified as Fitzpatrick skin types IV and V (Bin Yap, 2010), these readings

are consistent for percentages in the UV region ranging between 3-45%. Subjects with fairer skin such as the Japanese and Caucasians (Type II or III) has reportedly averaged >60% reflection readings between 500-800 nm (Zonios, Bykowski, & Kollias, 2001).

Although one of the spectral profiles showed some aberration, this is probably due to the heat in the surrounding experimental setup which distorted the spectrum in the NIR. However, this is deemed negligible due to the low pattern variance in this region, and also with the more significant patterns from 173-1070 nm, accounting ~90% of the data points which are unaffected due to the lower resolution of the second spectrometer.

4.2.2.2 Raw Spectrum – Chinese

Figure 4.5 in the following shows the compendium of Chinese spectra.
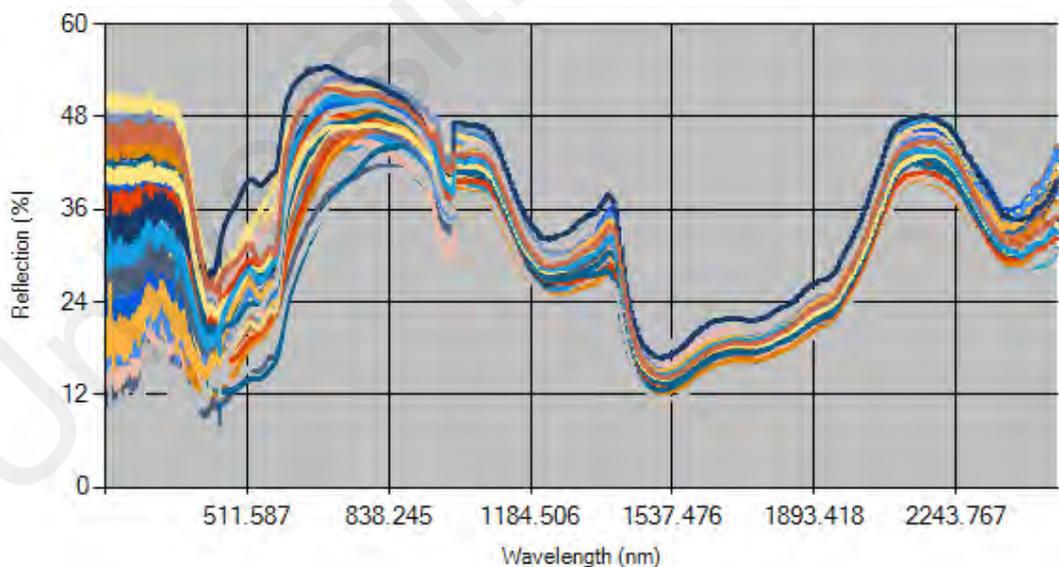
Figure 4.5: Spectroscopic data of the Malaysian Chinese ethnicity

Figure 4.5 illustrates a slightly higher collective reflection percentages. This is intuitively understood based on the slight conflation between Chinese and Malays' skin types, where the latter is generally classified as Type III and IV (Bin Yap, 2010). This is

similarly prevalent along the UV-VIS-NIR spectrum, until 1100nm. These findings are expected due to the generally fairer complexion of the Chinese, which have lower skin melanin content (Alaluf et al., 2002).

Similarly, the Japanese also are generally Type III, but registering higher reflectance readings compared to Type IV and V (Zonios et al., 2001).

4.2.2.3 Raw Spectrum – Indians

Figure 4.6 in the following shows the collection of spectra from Indians in Malaysia.
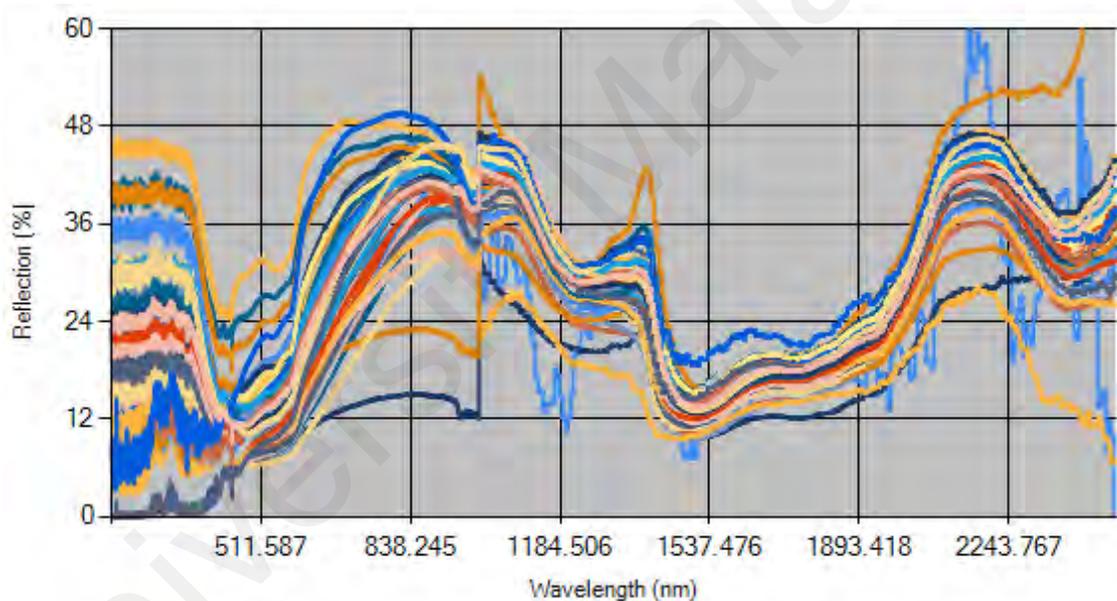
Figure 4.6: Spectroscopic data of Malaysian Indians

Intuitively from **Figure 4.6**, the reflection spectra of Indians, which generally are of Type VI (Koh et al., 2003), registered a higher density in the lower percentage tiers in the UV range, between 0-28%. The reflectance values, as shown in **Figure 4.5** are relatively consistent compared to **Figures 4.3** and **4.6**, where skin Types III-V as represented by the Malays and Chinese have higher reflectance percentages. Skin melanin of Indians have generally been attributed to its lower reflectance readings throughout the spectrum until 1200 nm.

In the VIS region, however, is only slightly lower than its previous counterparts. This is due to the steadily lessening gap between the skin types' reflectance from 300-800 nm (Zonios et al., 2001). This trend can be observed similarly until the pre-NIR range.

Again, the aberration in the spectral pattern in one subject is most likely due to the heating of the spectrometer. This is again deemed acceptable with the main structure of the spectrum still intact.

4.2.2.4 Raw Spectrum – Others: Non-Malaysian nationals
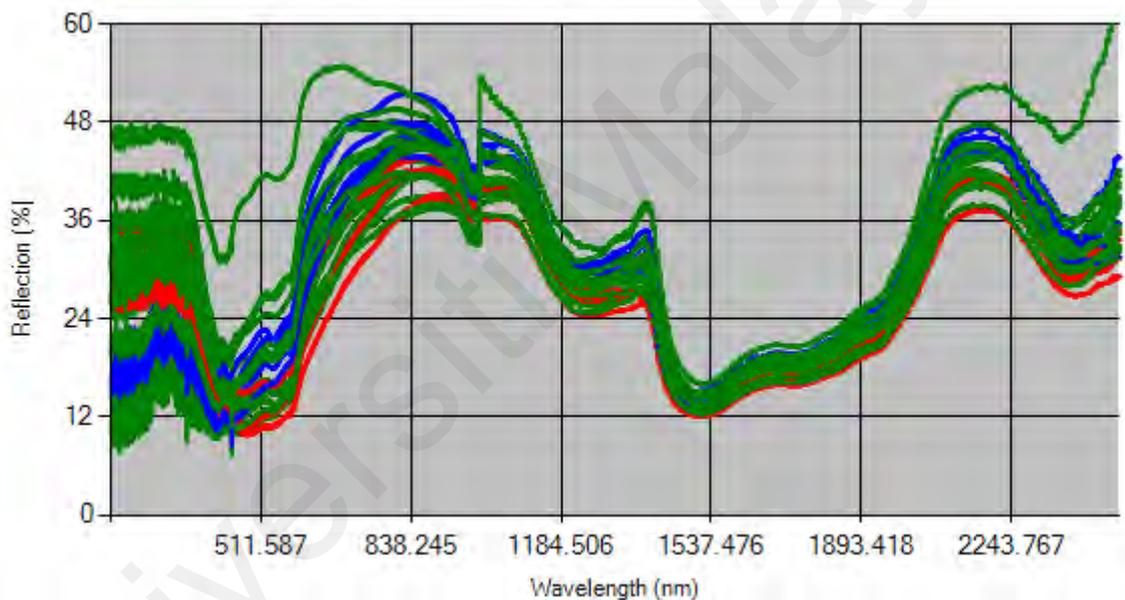
Figure 4.7: Spectroscopic data of other nationalities. Blue represents Indonesians while red represents Bangladeshis. Others are colored green.

Other than Malays, Chinese and Indians, other relatively major populations in our data pool are Indonesians and Bangladeshis. **Figure 4.7** compiles the remaining subjects recruited in our study. Since this group consists of various Asian nationalities with generally different skin types, it can be evidently seen that there is more spectral pattern differences.

Indonesians generally share the same skin Type IV or V as Malays (Halder & Nootheti, 2003), as shown in **Figure 4.2** with a blue series, which returns higher reflection spectra compared to Bangladeshis in red. This is also intuitive since Bangladeshis generally have Type V skin (Huggins et al., 2012; Zambelis et al., 2017), which explains the similarity of the spectrum in terms of the reflection peaks with Indians as shown in **Figure 4.6**.

As for the other subjects, it can be observed that the skin types are all generally in agreement with most Asian skin types.

4.2.3 Raw Spectrum – Age Groups

Figure 4.8 in the following highlights the age groups as discussed in **Figure 4.2**.
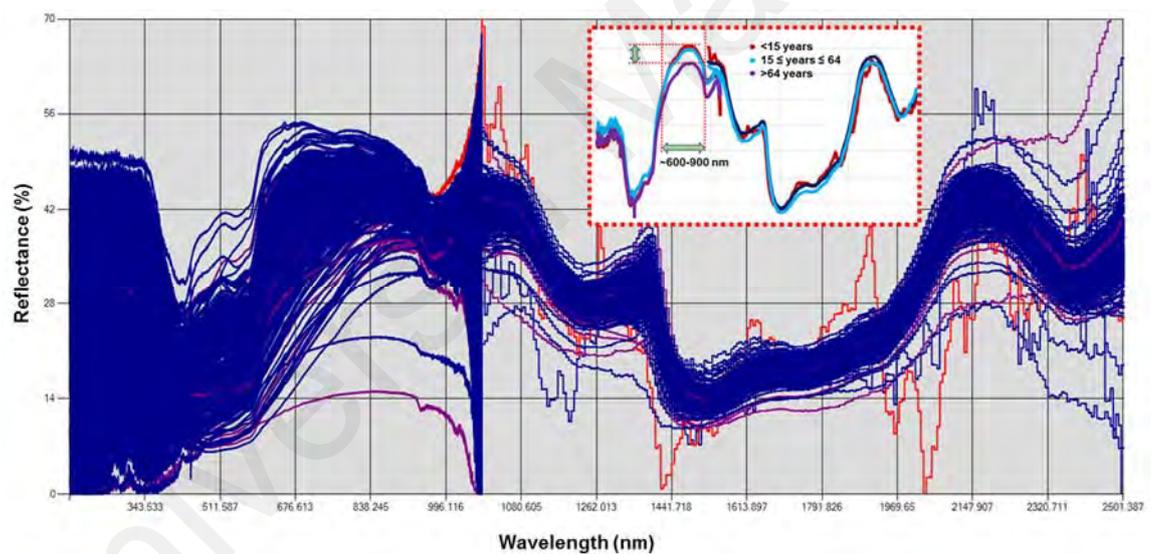
Figure 4.8: Spectroscopic data of age groups. Inset shows average spectrum of each group. The red series represent subjects (A) aged below 15, (B) blue; 15-64 years old, and (C) purple; more than 64 years old.

As shown in **Figure 4.8**, due to the overwhelming majority of the data belonging to group **B**, which comprises of 202 (84.2%) patients, the spectra minimally represents group **A** and **C**. However, by examining the inset which show the average age group spectra as shown in **Figure 4.8**, while group **A** and **B** exhibit very similar properties, group **C**, comprising of older subjects aged 65 and above, have mild pattern differences

ranging 2-5% reflection levels in the 600-900 nm. These pattern subtleties will be discussed in the next chapter.

4.3 Clinical Diagnosis

This segment presents the clinical diagnosis of each patient. This section is much more nuanced, therefore necessitating the data to be presented as part of the text in this section. This is due to the nature of a clinical diagnosis, which synthesizes various resources at hand which is processed by highly skilled physicians. As per the methodology outlined, the final diagnosis, or non-diagnosis (uncertain diagnosis leading to exclusion – in part only for the diagnostic analytics, but still maintained for other classifiable categories such as gender and ethnicity) is compiled in this section. The conclusions arrived for each subject recruited in this study are scrutinized in detail.

From 240 patients included in this study (as elaborated in **Section 4.1.1**), 122 (50.8%) have been subjected to clinical dengue assessment. 28 (11.7%) have been classified as confirmed dengue. 8 (3.3%) patients are diagnosed with probable dengue. The remainder 194 (80.8%) are diagnosed with mostly febrile illnesses, serving as a control in our study.

Amongst the patients diagnosed with non-dengue diseases, the majority of the illnesses causing fever are upper respiratory tract infection (URTI) at 69 (28.8%), 22 (9.2%) acute gastroenteritis (AGE), 24 (10.0%) undifferentiated acute febrile illness, 14 (5.8%) pneumonia, 9 (3.8%) urinary tract infections (UTI), 10 (4.2%) viral/acute gastritis, and 46 (19.2%) others.

Finally, 10 (4.2%) are excluded from the data analysis due to insufficient information for a final diagnosis from the clinical notes. The reasons includes the patient leaving UMMC before further assessment can be performed and insufficient clinical notes.

The breakdown of the diagnosis is illustrated in **Figure 4.9**.
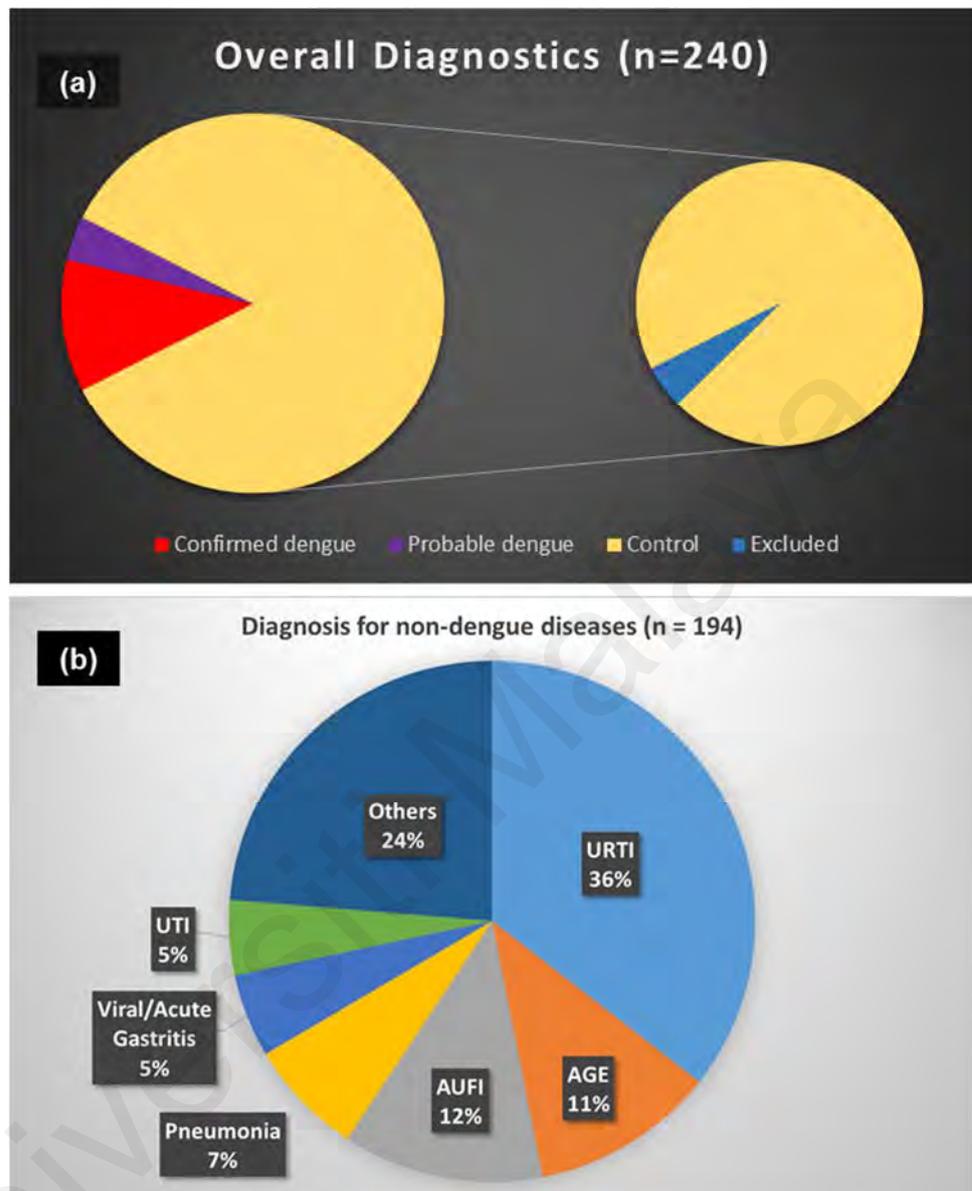
Figure 4.9: Diagnostics' composition

Table 4.1 describes the diagnosis of each patient with a clinical note summary. Confirmed dengue cases are highlighted in red, while probable dengue is highlighted in purple. Excluded patients are highlighted in blue, while the remainder control patients are highlighted in green as the default table background.

Table 4.1: Breakdown of clinical diagnosis

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0001	Undifferentiated acute febrile illness	FBC normal on D4	n/a
S0002	Viral gastritis	FBC normal on D2 and D3	n/a
S0003	Viral gastritis	FBC normal on D1, vomiting diarrhoea D3	n/a
S0004	Leptospirosis	Laboratory confirmation	n/a
S0005	URTI	FBC normal on D5 and D6, Dengue diagnostics all negative at D5	n/a
S0006	Viral gastritis	FBC normal D3, Dengue diagnostics all negative at D3	n/a
S0007	Community acquired pneumonia	Raised WBC, Chest X-Ray findings	n/a
S0008	Confirmed dengue	NS1 Positive at day 4 from Private GP	NS1 positive IgM negative IgG positive
S0009	Viral Gastritis	FBC normal on D2	n/a
S0010	URTI	FBC normal on D5	n/a
S0011	Confirmed dengue	NS1 & IgM positive at GP at D6	NS1 positive IgM positive
S0012	URTI	FBC normal on D4	n/a
S0013	Acute Pharyngitis/URTI	No FBC done	n/a
S0014	Undifferentiated acute febrile illness	FBC normal on D3	n/a
S0015	UTI	WBC high, Platelet normal at D3	n/a
S0016	Confirmed dengue	NS1 detected on D5, clinical consistency	NS1 positive
S0017	Appendicitis	Perforated gastric ulcer	n/a
S0018	Excluded	Call no action	n/a
S0019	UTI	WBC high, UFEME suggestive	n/a
S0020	URTI	FBC normal D2	n/a
S0021	URTI	FBC normal D7	n/a
S0022	Hospital acquired pneumonia	2 weeks hospital admission	n/a
S0023	Community acquired pneumonia	Normal FBC at D5	n/a
S0024	Confirmed dengue	IgM detected on D7, clinical consistency	NS1 negative

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
			IgM positive
S0025	Lower limb Cellulitis	WBC high, Ultrasound findings	
S0026	Urosepsis	WBC high, UFEME suggestive, blood culture E.Coli	n/a
S0027	Urosepsis	WBC high, UFEME suggestive	n/a
S0028	Undifferentiated acute febrile illness	Only one FBC which is suggestive of dengue	n/a
S0029	Community acquired pneumonia	WBC high at D7	n/a
S0030	Aspiration pneumonia due to left Cerebrovascular accident (stroke)	CT scan of brain	n/a
S0031	AGE and Community acquired pneumonia	Blood culture E.Coli	n/a
S0032	Confirmed dengue	NS1 positive without warning signs	NS1 positive
S0033	URTI and AGE	WBC high	n/a
S0034	Undifferentiated acute febrile illness	FBC and LFT trend in keeping with dengue however no leucopenia and NS1 IgG IgM not detected on D7	NS1 negative IgM negative IgG negative
S0035	Head of pancreas carcinoma	CT scan findings, FBC normal	n/a
S0036	CAP (Community acquired Pneumonia)	Chest X-Ray suggestive	n/a
S0037	URTI	WBC normal, low platelet at D3	n/a
S0038	Acute pharyngitis	WBC high	n/a
S0039	CAP (Community acquired Pneumonia)	WBC high	n/a
S0040	CAP (Community acquired Pneumonia)	FBC normal	n/a
S0041	URTI	Blood not available	n/a
S0042	URTI	FBC normal at D3	n/a
S0043	URTI	FBC Normal at D2	n/a
S0044	Viral gastritis	WBC normal, low platelet at D2	n/a
S0045	UTI	WBC high, UFEME suggestive	n/a
S0046	Urosepsis	WBC high, UFEME suggestive, blood culture Citrobacter koseri	n/a
S0047	AGE	FBC normal at D4	n/a
S0048	URTI	NS1 on D3 negative	NS1 negative

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0049	Liver Abcess	WBC high, Ultrasound findings from Private Hospital	n/a
S0050	Acute pharyngitis	No FBC done, clinical diagnosis	n/a
S0051	Right foot cellulitis	No FBC done, clinical diagnosis	n/a
S0052	URTI	Normal FBC at D3	n/a
S0053	Acute gastritis	Normal FBC, clinical presentation	n/a
S0054	URTI	WBC high at D1	n/a
S0055	URTI	Normal FBC at D1	n/a
S0056	Liver Abcess	WBC and platelets very high, ultrasound findings	n/a
S0057	URTI	WBC high at 2 weeks of fever	n/a
S0058	Surgical site infection (Inguinal hernia)	WBC high, discharge inflamed wound	n/a
S0059	URTI	FBC normal at D2 and D4	n/a
S0060	Confirmed dengue	NS1 positive on D3, clinical consistency	NS1 positive
S0061	CAP (Community acquired Pneumonia)	Chest X-Ray suggestive	n/a
S0062	CAP (Community acquired Pneumonia)	Clinical diagnosis, FBC normal	n/a
S0063	URTI	FBC normal at D7	n/a
S0064	Undifferentiated acute febrile illness	Platelets high at D10	n/a
S0065	Undifferentiated acute febrile illness	WBC high	n/a
S0066	AGE	WBC high, clinical diagnosis	n/a
S0067	URTI	FBC normal at D4	n/a
S0068	Acute pharyngitis	No FBC done, clinical diagnosis	n/a
S0069	URTI	Raised WBC and platelets, Chest X-Ray normal, AFB and mantoux test negative	n/a
S0070	Left foot cellulitis	Clinical diagnosis, WBC and Platelet elevated	n/a
S0071	CAP (Community acquired Pneumonia)	Clinical diagnosis - WBC normal, Platelet low, persistent thrombocytopenia, X-ray normal	n/a
S0072	AGE	WBC elevated, normal platelet	n/a
S0073	Pneumonia	Chest X-Ray suggestive	n/a
S0074	AGE	NS1 on D2 negative	NS1 negative
S0075	AGE	Clinical diagnosis, FBC normal	n/a
S0076	Urosepsis	WBC normal, platelet low, UFEME suggestive, Urine culture E. Coli	n/a

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0077	URTI	WBC elevated, normal platelet	n/a
S0078	Undifferentiated acute febrile illness	FBC normal at D6	n/a
S0079	Symptomatic anemia due to uterine fibroid	Hb low, WBC and platelet raised	n/a
S0080	Probable dengue	Rapid test positive KK, query NS1/IgM/IgG. Since taken on D12 most likely IgM/IgG positive therefore can only be said to be probable.	IgM positive/ IgG positive
S0081	Acute Pharyngitis	WBC elevated, normal platelet	n/a
S0082	URTI	FBC not done, Clinically suggestive	n/a
S0083	Undifferentiated acute febrile illness	NS1 on D4 negative, Platelet and WBC normal	NS1 negative
S0084	URTI	FBC normal at D7	n/a
S0085	Acute tonsillitis	WBC elevated, normal platelet, tonsils exudates	n/a
S0086	URTI	Clinical Diagnosis	n/a
S0087	Hypopituitarism	WBC normal, MRI findings of bilateral pituitary microadenoma	n/a
S0088	UTI	FBC normal, UFEME suggestive	n/a
S0089	URTI	FBC normal at D7	n/a
S0090	URTI	FBC not done, Clinical diagnosis	n/a
S0091	Brain metastasis secondary to brain carcinoma	WBC normal, platelet high, Blood culture no growth, UFEME negative, CT brain: Brain & possible skull vault metastases, Right front calcified meningioma	n/a
S0092	URTI	WBC and platelets high	n/a
S0093	Undecided 1	WBC normal, Fibrin clot seen	n/a
S0094	URTI	FBC not done, Clinical diagnosis	n/a
S0095	Symptomatic anemia due to uterine fibroid	Hb low, WBC and platelet normal	n/a
S0096	Sacral ulcer & UTI	WBC and platelet high, UFEME suggestive, swab culture proteus mirabilis	n/a
S0097	Infected left DFU	WBC and platelets elevated, clinically suggestive	n/a
S0098	URTI	FBC normal	n/a
S0099	Musculoskeletal pain	Clinically suggestive	n/a
S0100	Confirmed dengue	NS1 positive on D3, clinical consistency	NS1 positive
S0101	CRPSI	WBC elevated, Platelet normal, blood culture Corynebacterium Striatum	n/a

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0102	URTI	FBC normal at D2	n/a
S0103	Confirmed dengue	NS1 positive on D6, measurement done in recovery phase	NS1 positive
S0104	Confirmed dengue	NS1 also positive from GP & minimal PV bleeding	NS1 positive
S0105	Undifferentiated acute febrile illness	FBC normal at D3, clinical diagnosis	n/a
S0106	Undifferentiated acute febrile illness	WBC low, platelet normal, subsequently NS1, IgM, IgG negative	NS1 negative IgM negative IgG negative
S0107	URTI	FBC normal at D4, clinically suggestive	n/a
S0108	Confirmed dengue	IgM positive at D7	NS1 negative IgM positive
S0109	Confirmed dengue	IgM positive at D7	NS1 negative IgM positive
S0110	URTI	WBC elevated, Platelet normal	n/a
S0111	Acute pharyngitis	FBC normal at D7	n/a
S0112	Excluded	Call no Action	n/a
S0113	Right lower limb cellulitis	WBC elevated, platelet normal, Clinical diagnosis	n/a
S0114	Acute tonsillitis	WBC high, platelet normal at D4, tonsils exudate	n/a
S0115	URTI	WBC high, platelet normal at D2	n/a
S0116	AGE	No blood taken, clinically suggestive	n/a
S0117	Confirmed dengue	NS1 positive at D6	NS1 positive
S0118	Acute pharyngitis	FBC normal at D5	n/a
S0119	CAP (Community acquired Pneumonia)	Chest X-Ray suggestive	n/a
S0120	Confirmed dengue	NS1 positive at D3	NS1 positive
S0121	AGE	WBC elevated, platelet normal, clinically suggestive	n/a
S0122	Probable dengue	IgM positive at D3, probable dengue	NS1 negative IgM positive
S0123	URTI	WBC normal, platelet low	n/a
S0124	Close fracture right distal radius	X-Ray findings and clinical diagnosis	n/a
S0125	Probable dengue	IgM positive at D4, probable dengue	NS1 negative IgM positive
S0126	URTI	No blood taken, clinical diagnosis	n/a
S0127	Orchitis	NS1, IgM, IgG negative at GP, red swollen tender right testis	NS1 negative IgM negative IgG negative
S0128	Undifferentiated acute febrile illness	FBC, UFEME and chest X-ray normal at D30	n/a
S0129	URTI	WBC elevated, platelets normal at D3	n/a

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0130	Viral gastritis	No blood taken, clinically suggestive	n/a
S0131	UTI	WBC high, platelet normal, UFEME suggestive	n/a
S0132	UTI	FBC normal, UFEME suggestive	n/a
S0133	Confirmed dengue	NS1 positive but FBC normal	NS1 positive
S0134	URTI	FBC normal, clinically suggestive	n/a
S0135	Confirmed dengue	NS1 positive at D2 with HLH	NS1 positive
S0136	URTI	FBC normal at D7, clinically suggestive, chest x-ray normal	n/a
S0137	Varicella Zoster infection	No blood taken, clinical diagnosis	n/a
S0138	URTI	WBC elevated, platelet normal at D2, clinically suggestive	n/a
S0139	Tonsillitis	FBC normal at D3, clinical diagnosis	n/a
S0140	Undifferentiated acute febrile illness	FBC normal at D2, clinical diagnosis	n/a
S0141	Tonsillitis	FBC normal at D5, clinical diagnosis	n/a
S0142	AGE	FBC normal at D7, clinical diagnosis	n/a
S0143	URTI	FBC normal at D2, clinical diagnosis	n/a
S0144	URTI	WBC high, platelet normal	n/a
S0145	URTI	No blood taken, clinically suggestive	n/a
S0146	AGE	FBC normal at D5, had dengue 10 years ago	n/a
S0147	Tonsillitis	FBC normal, clinically suggestive	n/a
S0148	Excluded	Cannot be accessed for cohort	n/a
S0149	Confirmed dengue	Clinically suggestive: Low WBC and platelets, went to Private Hospital, NS1 positive at D3	NS1 positive
S0150	Right peritonsillar abcess	WBC elevated, platelet normal, abcess Strepcoccus mitis	n/a
S0151	URTI	FBC normal at D3, clinical diagnosis	n/a
S0152	Undifferentiated acute febrile illness	FBC normal at D7, clinical diagnosis	n/a
S0153	AGE	FBC normal at D2, clinical diagnosis	n/a
S0154	URTI	WBC elevated, platelet normal, clinically suggestive	n/a
S0155	URTI	FBC normal at D5, clinically suggestive	n/a
S0156	URTI	Pregnant at 21/52, FBC normal, clinically suggestive diagnosis	n/a
S0157	URTI	FBC normal at D1 and D3, clinical diagnosis	n/a

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0158	URTI	WBC and platelets normal, NS1, IgM, IgG negative at private at D4, WBC low, platelet normal but receding at D6, other details: recent fogging, myalgia, athralgia	NS1 negative IgM negative IgG negative
S0159	Undifferentiated acute febrile illness	WBC elevated, platelet normal, Clinical diagnosis	n/a
S0160	Acute pharyngitis	FBC normal at D3	n/a
S0161	AGE	FBC normal at D1, UFEME negative, clinical diagnosis	n/a
S0162	Parotitis	WBC low, platelet normal, bilateral parotic gland enlarged and tender, tonsils enlarged	n/a
S0163	URTI	No blood taken, clinical diagnosis	n/a
S0164	Viral gastritis	Normal FBC at D2, clinical diagnosis	n/a
S0165	Acute pharyngitis	Normal FBC at D3, clinically suggestive	n/a
S0166	AGE	WBC elevated, Platelet normal at D4, clinically suggestive	n/a
S0167	Probable dengue	IgG positive, clinical consistency	NS1 negative IgM negative IgG positive
S0168	Undecided 2	Undifferentiated acute febrile illness (Though epidemiologically and clinically could be dengue though serology not sent. FBC WBC low, and platelet normal	n/a
S0169	Acute pharyngitis	FBC normal, clinically suggestive	n/a
S0170	Acute pharyngitis	No blood taken, clinically suggestive	n/a
S0171	Periodic hypokalaemic paralysis	Potassium readings low, WBC and platelet elevated, included due to lethargy	n/a
S0172	Left epididimo-orchitis	WBC very high, platelets normal, ultrasound scrotum suggestive	n/a
S0173	AGE	FBC normal at D4	n/a
S0174	Rheumatic fever	Jones criteria: fulfil one major (polyathritis), two minor (raised ESR, CRP, leucocytosis)	n/a
S0175	Mesenteric adenitis	WBC elevated, platelet normal, Clinical diagnosis	n/a
S0176	URTI with hyperemesis gravidarum	Pregnant 12/52, FBC normal at D2, clinically suggestive	n/a
S0177	Viral gastritis	FBC normal, clinically suggestive	n/a
S0178	URTI	FBC at Private Hospital normal at D2, clinical diagnosis	n/a

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0179	Excluded	Call no action	n/a
S0180	Confirmed dengue	NS1 positive day D3	NS1 positive
S0181	Excluded	Call no Action	n/a
S0182	Confirmed dengue	IgM positive day D7	NS1 positive IgM positive
S0183	AGE	FBC normal at D2, clinically suggestive	n/a
S0184	URTI	FBC not taken, clinical diagnosis	n/a
S0185	Confirmed dengue	NS1 positive at D5, clinical consistency	NS1 positive
S0186	Acute pharyngitis	FBC not taken, clinical diagnosis	n/a
S0187	Undecided (default: AGE)	Dengue suspect though no documented temperature, FBC normal, clinically suggestive	n/a
S0188	URTI	FBC normal at D2, clinically suggestive	n/a
S0189	CSF leak post Foramen Magnum Decompression	WBC and platelet elevated at D1, clinical diagnosis	n/a
S0190	URTI	WBC and platelet elevated at D2, clinical diagnosis	n/a
S0191	URTI	FBC normal at D5, clinical diagnosis	n/a
S0192	URTI	FBC normal at D7, clinical diagnosis	n/a
S0193	Undifferentiated acute febrile illness	WBC normal, platelet elevated, NS1, IgM, IgG negative	NS1 negative IgM negative IgG negative
S0194	Undifferentiated acute febrile illness	FBC normal at D4, clinically suggestive	n/a
S0195	Excluded	Call no Action	n/a
S0196	Tonsillitis	FBC normal at D5, clinical diagnosis	n/a
S0197	Confirmed dengue	NS1 Positive at D3	NS1 positive
S0198	Undifferentiated acute febrile illness	No blood taken, clinical diagnosis	n/a
S0199	URTI	FBC normal at D5, Chest X-ray normal.	n/a
S0200	AGE	FBC normal at D2, clinical diagnosis	n/a
S0201	UTI	WBC elevated, UFEME positive	n/a
S0202	URTI	FBC normal at D5, clinically suggestive	n/a
S0203	Acute pharyngitis	FBC normal at D4, clinically suggestive	n/a
S0204	Undifferentiated acute febrile illness	FBC normal at D4, clinically suggestive	n/a
S0205	Undifferentiated acute febrile illness	FBC normal at D4, clinically suggestive	n/a
S0206	AEBA secondary to URTI	WBC elevated, UFEME suggestive	n/a
S0207	Confirmed dengue	NS1 Positive at D5	NS1 positive

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0208	Post tonsillectomy D3	WBC elevated, clinically suggestive	n/a
S0209	Acute pharyngitis	FBC normal, clinically suggestive	n/a
S0210	AGE	No blood taken, vomiting and diarrhea at D1, clinically suggestive	n/a
S0211	Confirmed dengue	NS1 Positive at day 3 from Private GP	NS1 positive IgM positive
S0212	Excluded	Not enough data (MA assessment only)	n/a
S0213	Confirmed dengue	NS1 positive on D5, Measurements on D7	NS1 positive
S0214	URTI	No blood taken, clinically suggestive	n/a
S0215	Undifferentiated acute febrile illness	No blood taken, clinically suggestive	n/a
S0216	UTI	FBC normal, UFEME suggestive	n/a
S0217	URTI	FBC normal, clinically suggestive	n/a
S0218	Undifferentiated acute febrile illness	FBC normal, clinically suggestive	n/a
S0219	URTI	WBC elevated, platelet normal	n/a
S0220	URTI	No blood taken, fever too long for dengue. Symptoms suggestive of URTI	n/a
S0221	AGE	FBC normal, UFEME negative	n/a
S0222	URTI	FBC normal, clinically suggestive	n/a
S0223	Confirmed dengue	NS1 positive day D6	NS1 positive
S0224	Probable dengue	Verified from Clinical Notes: Had FBC in GP showing reducing trend in Platelet and WBC	NS1 negative IgM negative IgG negative
S0225	Probable dengue	IgM positive day D4	NS1 negative IgM positive
S0226	Confirmed dengue	NS1 positive at day 3 from Private GP	NS1 positive IgM positive
S0227	AGE	WBC high, Platelet normal	n/a
S0228	URTI	FBC normal, clinically suggestive	n/a
S0229	Confirmed dengue	NS1 positive at D4	NS1 positive
S0230	URTI	No blood taken, clinically suggestive	n/a
S0231	Probable dengue	At D8 during recruiting, towards recovery	IgM positive
S0232	Confirmed dengue	NS1 detected at GP at D4, petechia, showing reducing WBC and platelet trends	NS1 positive
S0233	Probable dengue	IgG positive at D5, clinical consistency with dengue	NS1 negative IgM negative IgG positive
S0234	Confirmed dengue	NS1 positive D6, compensated shock	NS1 positive
S0235	Undifferentiated acute febrile illness	Pregnant at 2/52, HgB low, WBC and platelets normal, UFEME normal	n/a

ID	Final Diagnosis	Clinical Note Summary	Dengue Serology
S0236	URTI	FBC normal, clinically suggestive	n/a
S0237	Acute Appendicitis	WBC elevated, clinically suggestive	n/a
S0238	Undifferentiated acute febrile illness	WBC elevated, UFEME negative, clinical diagnosis	n/a
S0239	URTI	FBC normal, clinically suggestive	n/a
S0240	AGE	WBC elevated, clinically suggestive	n/a

All the patients in this case have recorded fever day onsets, except **P0171** which were finally diagnosed with hypokalemia though it was linked with dengue due to lethargy (Jha & Ansari, 2010) but ultimately categorized as control. All of the patients' reflection spectra are tied to the designation for each patient in **Table 4.1** (S0001-S0240) which is further analyzed for statistical correlation. This will be discussed in the next chapter.

4.4   Chapter Conclusion

As per the chapter navigation illustrated in **Figure 4.1**, we have presented the external impressions on three components of the data collected. This includes the demographics, diffuse reflectance spectroscopy data and finally the clinical diagnosis.

The first part, which briefly summarizes the demographics such as nationality, age, gender and ethnicity outlines the population dynamics. Most of the patients (~90%) who were recruited are Malaysians.

As the reflectance spectroscopy data is dependent on mainly melanin content of the skin, among others, the data is partly presented as a function of the ethnicities in **Figures 4.2-4.6**. This is where categorical melanisation may be projected from the ethnicities, in the form of the Fitzpatrick skin phototype. However, it is to be noted that

the skin type is not a measure used in our study, though mentioned. Its mention is basically to ground the data obtained with the literature on the reflectance levels from subjects with demographically similar skin.

The final output from this study is the clinical diagnosis, which is derived from the clinical notes examined by authoritative personnel from UMMC. These clinical notes are summarized and the final diagnosis is provided for all the patients, excluding 10 patients, as exhausted in **Table 4.1**. Overall, there are 28 confirmed dengue, 8 probable dengue and 194 control cases to be linked with the spectroscopy data, and finally given an in-depth statistical analysis.

The statistical and pattern analysis of these three components mentioned in this section will be integrated and concluded in the next chapter.

CHAPTER 5: DISCUSSION - A SYNTHESIS BETWEEN ANALYTICS AND DIAGNOSTICS – PROVING SCREENING POTENTIAL

This chapter is a four-part compendium to analyze and conclude the link between the diffuse reflectance spectroscopy patterns versus patients with a either a confirmed or probable dengue, or non-dengue diagnosis. The chapter structure is illustrated in **Figure 5.1**.
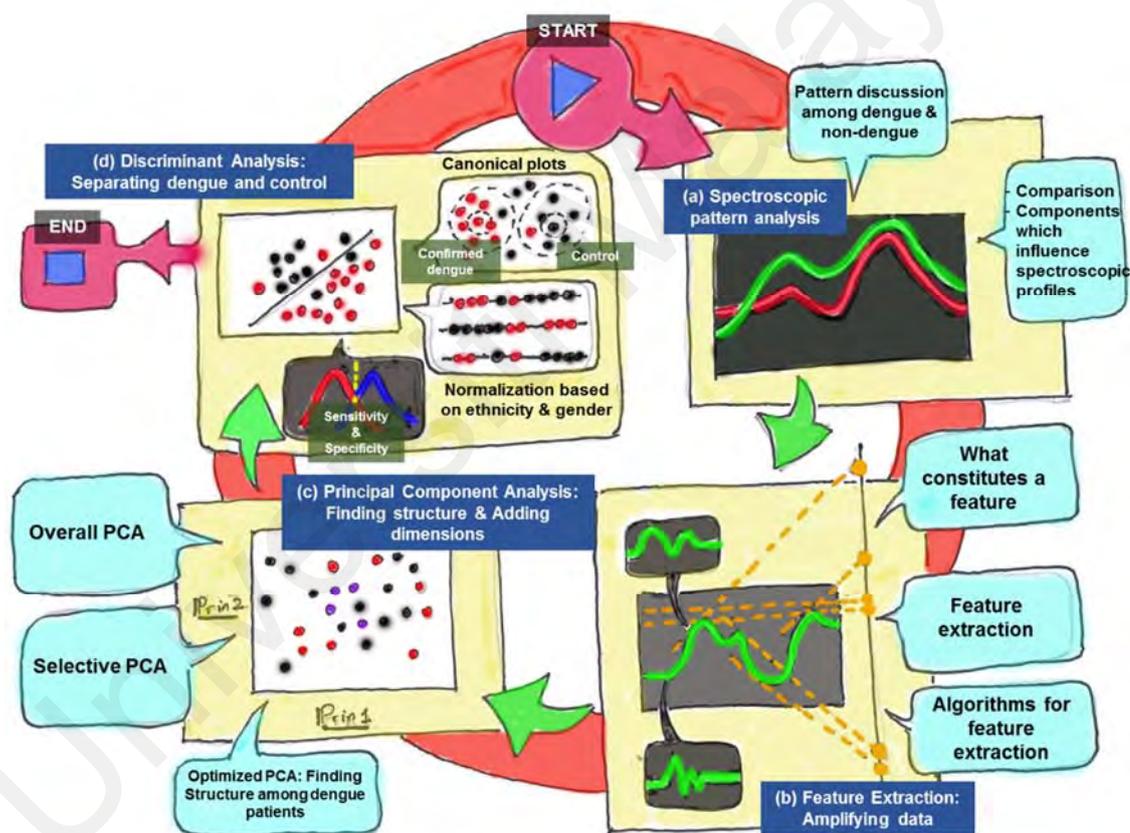
Figure 5.1: Chapter structure

As shown in the chapter structure in **Figure 5.1(a)**, this chapter begins with a raw spectroscopic data analysis of the three categories of patients: confirmed/probable dengue, and control patients. We will then prove that the spectroscopic patterns, other than shedding light on the ethnicity of the patients, are non-trivial (or convoluted), and

closely related profiles, rendering manual observation impossible. This, in result, necessitates the use of multivariate methods.

The second section in **Figure 5.1(b)** is used to augment the first. In order to increase the variance of the data, features are extracted by geometric amplification of the patterns in the Cartesian space. This increases the number of covariates for a further multivariate analysis. This is performed by identifying key observable patterns in the spectrum, and classifying the types. Then, the computation of these features are carried out.

The third section in **Figure 5.1(c)** is a platform to transcend from human observation to a multidimensional observation. This is performed by integrating both the raw spectroscopy data points and also the feature-extracted parameters. The covariates from the feature-extracted are proven for utility in finding structure to find clustering tendencies in order to reduce the dimensionality of the data. The feature-extraction method is used primarily to prove that patients with dengue exhibit a common feature using dimensionality reduction methods.

Finally in **Figure 5.1(d)**, we conclude in the fourth chapter with the use of discriminant analysis to classify confirmed dengue versus control. This subsection, in turn, is divided into four major sections. The first is using discriminant analysis for classifying ethnicities. The second focuses on gender. The third is where both latter sections are used as normalization factors for the reflectance spectroscopy pattern. This includes imputation of ethnicities for skin category for subjects which lacks a normalization reference, such as Bangladeshis and Punjabis. The fourth discusses the classification accuracies based on the diagnosis provided on each patient, with computations of sensitivity and specificity. An optimized model for the best dengue screening algorithm will be presented as a conclusion.

Finally, we discuss the aspects of the data which render the dengue infection detectable with the use diffuse reflectance spectroscopy. This chapter is finally concluded with a graphical representation on the different analytical methods from **Section 5.1** to **Section 5.4**.

5.1 Brief Analysis of Spectroscopy Data

As outlined previously, we present the decomposition of the confounding factors in objectively analysing the spectroscopic patterns. However, observable patterns which can be gleaned from the data is noted as an amplification factor in the next section. **Figure 5.2** summarizes the spectroscopic data in graphical representation.
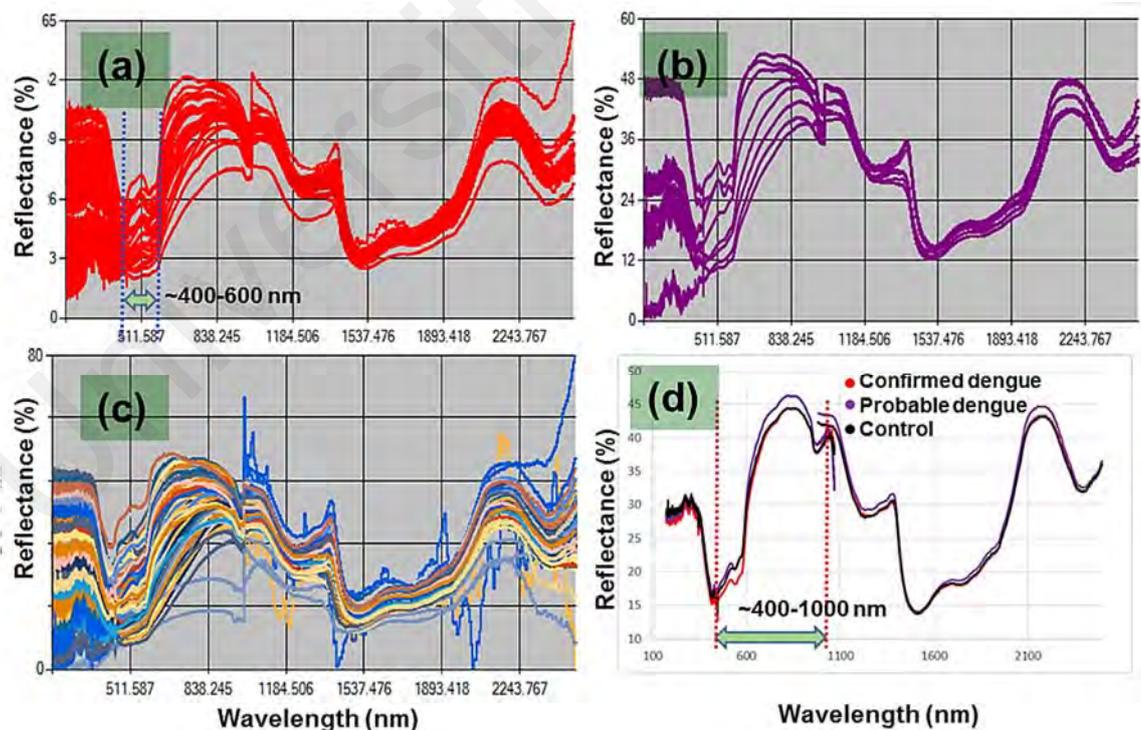
Figure 5.2: Spectroscopic data of (a) confirmed dengue patients, (b) probable dengue, (c) control patients and (d) average patterns from each class in (a) to (c).

Figure 5.2(a) shows the compiled 28 dengue patients' reflectance profiles, with 10 (35.7%) out of 28 patients female and 18 (64.3%) males. The composition of ethnicity is varied between 13 (46.4%) Malays, 3 (10.7%) Chinese, 4 (14.3%) Indians, 3 (10.7%) Bangladeshis, and 4 (14.3%) others. In this aspect, other than the observable reflection percentages due to the skin types prevalent in each ethnicity, visual cues which indicate any pattern unique to this class is absent. It can be safely concluded in this group, we have sufficient gender and ethnicity variability to avoid any issue with biased samples.

Figure 5.2(b) comprises of 8 probable dengue patients with 3 (37.5%) males and 5 (62.5%) females. In terms of ethnicity, there are no Malays, 3 (37.5%) Chinese, 2 (25.0%) Indians, 1 (12.5%) Bangladeshi and 2 (25.0%) others. Similarly, one may conclude the same in terms of the pattern commonalities for this class. It can be also maintained that the variability in both gender and ethnicity categories this group relatively offsets sampling bias problems. This is despite with no Malay subjects, it can be expected that the Malays in the control group offsets this issue.

Figure 5.2(c) shows a preponderance of the 194 control patients previously diagnosed with various febrile illnesses other than dengue. Equally, 97 (50.0%) male and female are categorised in this class. 95 (49.0%) are Malays, 33 (17.0%) are Chinese, 48 (24.7%) Indians, 2 (1.00%) Punjabis, 3 (1.5%) Bangladeshis, and 13 (6.7%) others. The control group is perhaps the strongest statistical anchor to separate commonalities between dengue and non-dengue patients.

Figure 5.2(d) allows us to look at each average class patterns collectively (excluding the "noisy" band between 1070-1100 nm from Spectrometer 1), with a relatively small variance in the range shown between 400-1000. Though one may be tempted to allude that the "double peak" pattern is prevalent in **Figure 5.2(a)** and **Figure 5.2(b)** for both dengue classes, which can be seen in the highlighted band between 400-600 nm, this

pattern is certainly not unique only to dengue groups, though it is slightly more prominent in both dengue categories, as shown in **Figure 5.2(d)**.

It can be concluded in this case, that the data classically displays subtleties which require more sophisticated processing. The next logical step is to attempt amplifying the data to increase the salience of the patterns. The main reason for the failure of a visual analysis lies in the fact that the majority of the spectroscopic patterns are determined by four major components in the skin, namely melanin, oxygenated and deoxygenated haemoglobin, and water (Cooksey et al., 2014). Especially for melanin, the range of absorption pervades within practically a majority of the UV-VIS-NIR range, up to 1300 nm.

However, subtleties which exist in the patterns are due to other components of the human skin and tissue. These aspects will be discussed in-depth in the following subsections.

5.2   Feature Extraction: A Closer Inspection of Pattern Subtlety

There are three main reasons to advance our discussion into data features as follows: (a) dimensionality reduction, (b) useful wavelength range observation and (c) pattern observation.

Feature extraction, as the name implies, is a method of which the dimensions of the data is reduced by computing outliers from the pattern. In many cases, it is heuristically driven, meaning a set of rules applied on the pattern variations constitute a feature. Throughout a pattern comprised of thousands of data points in closely-related patterns, often it can be reduced to a representation of only a few data points.

For example, the use of facial recognition in computer learning uses a digitized image of the face, and is then turned into a binary grayscale image. The features of the 2-D

image is then extracted and reduced to an array of scalar values which then can be statistically computed to match the actual owner of the face. This is generally the case though in reality it is much more nuanced and often a nontrivial computational conundrum.

A multivariate method such as Principal Component Analysis (PCA) depends on the use of variables selectively in order for the analyst to find structure of the data, especially in a 2-dimensional space. However, to selectively perform this task for spectroscopic data, spanning nearly 3,000 data points, is a daunting task. Therefore, to reduce the spectrum to a set of more manageable variables is very much desired.

In our case, the main reference on how the features will be determined will be based on mainly the spectrum of confirmed dengue patients as discussed in the previous section. Then, the rules are expanded for commonalities in the spectroscopic data of control patients. All computations of the feature extraction are performed by close scrutiny of the spectra with Microsoft Visual Studio Graph Control via VS2013, combined with VB.NET on producing the transformed datasets.

5.2.1 What Constitutes a Feature?

It is imperative that a few assumptions are made to define the heuristics employed for a feature to be considered. This is due to the nature of the spectroscopic patterns observed which has both subtle and non-subtle patterns. The grading of these patterns in turn are determined by the noticeability, in our case subjective. Graphical representations are used to demonstrate how these features appear.

There are generally three types of features: (a) scalar or absolute values from certain pattern peaks/troughs, (b) gradient and intersection values, and (c) relative values of

peaks/troughs from the baseline surrounding it. These features are in turn divided into four categories, which are **major** and **minor** features for both **absolute** and **gradient/projected** values. Regions with less signal-to-noise (SNR) ratio, or with apparently smoother texture are prioritized. This, as we shall see, is exhibited in regions between 400-1000 nm from Spectrometer 1 and the rest from Spectrometer 2. Instrument uncertainty that exists between 1000-1100 nm are excluded from the pattern analysis, but are compensated from the readings of Spectrometer 2, though with much less spectral resolution.

Absolute values (denoted by P_n, Q_n and VIS_n) later) are derived directly from the dataset without any computation, where relative values are computed with respect to the baseline surrounding the coordinate of interest (or $P_{relative} = P_{absolute} - P_{baseline}$). Gradients, in turn, allow for two parameters, namely the value of the gradient m_n and also projection values c_n. The intersection values reflect a projection of a certain point of the data on a much larger axis.

The following elaborates on each sections.

5.2.1.1 Major features – Absolute values

Figure 5.3 in the following outlines several major features, based on confirmed dengue patients.
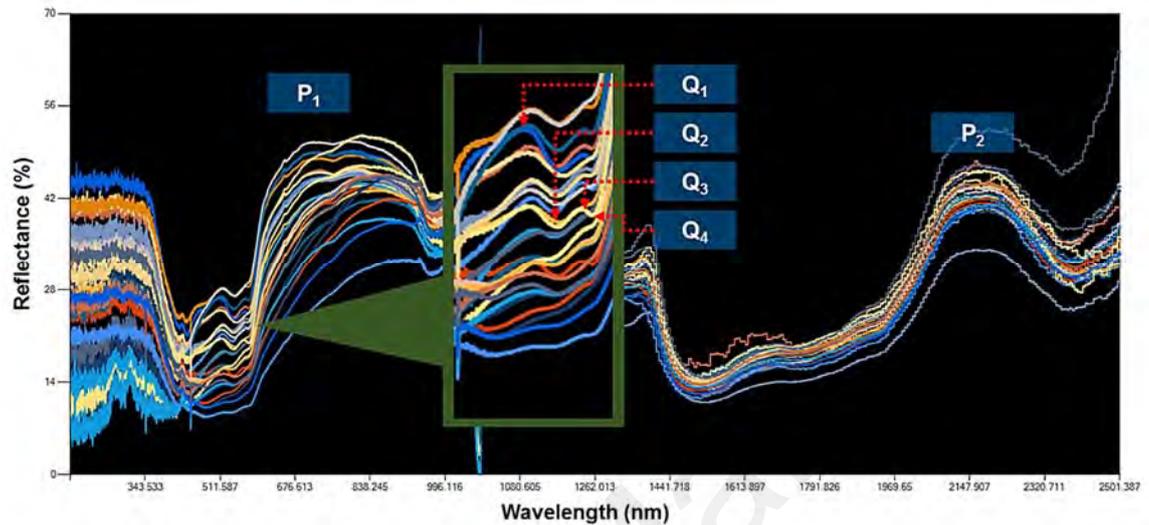
Figure 5.3: Scalar major features

There are two main peaks noticeable from the spectrum, denoted as P_1 and P_2. P_1 occurs in the end visible and part of the NIR spectrum (700-850 nm), signifying a combination of the optical properties of melanin in the skin, and blood or tissue oxygenation (Matas, Sowa, Taylor, & Mantsch, 2002; Sprigle, Linden, & Riordan, 2002). The next features which are more prevalent among confirmed dengue patients are highlighted as Q_1 to Q_4, occurring between 515-575 nm, with the same components affecting the spectrum for P_1 and P_2. These patterns were classically attributed to oxygenated and deoxygenated blood (Zonios et al., 2001).

P_2 is within the NIR region, and is not influenced by melanin or blood components, but mainly water content in the skin as elaborated on the role played in the NIR region (Arimoto et al., 2005; Attas et al., 2002; De Rigal et al., 1993; Martin, 1993; Matas et al., 2002; Sprigle et al., 2002; Walling & Dabney, 1989).

5.2.1.2 Major features – Gradient values
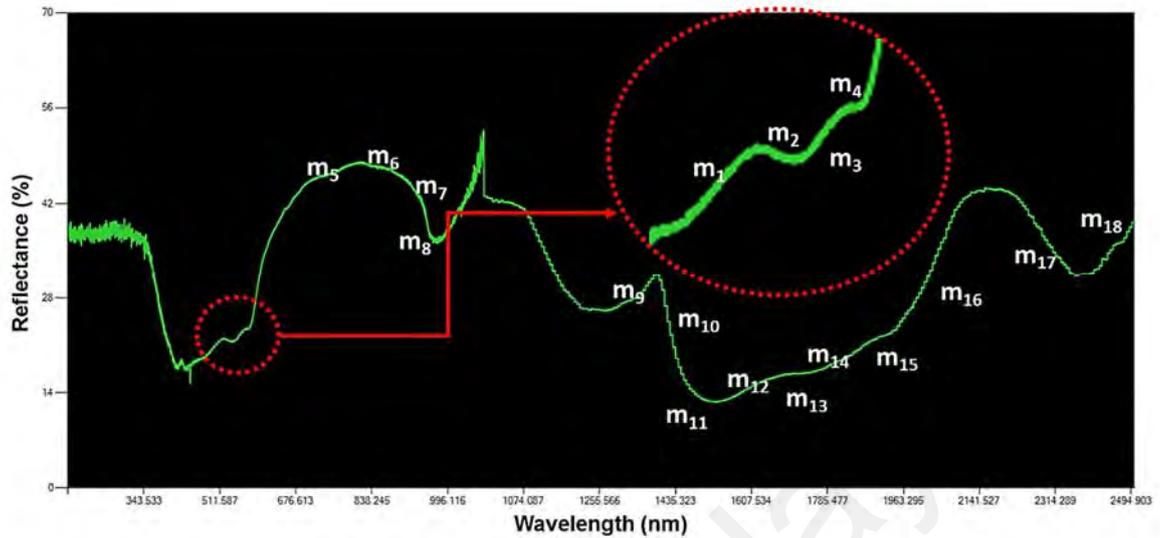
Figure 5.4: Gradient and intersection features

Figure 5.4 describes the gradient patterns gleaned from the profile throughout a typical spectroscopic graph. These values are denoted as m_1 to m_{18} whereas linear intersection values, augmenting the gradient values, are denoted as c_1 to c_{18}. These values cover both VIS and NIR ranges, signifying the role played by melanin, blood or tissue oxygenation for m_1-m_8 and water in the dermis for m_9-m_{18} (Arimoto et al., 2005; Attas et al., 2002; De Rigal et al., 1993; Martin, 1993; Matas et al., 2002; Sprigle et al., 2002; Walling & Dabney, 1989).

The NIR region (800-2500 nm), spanning between m_9 to m_{18}, is considerably higher in its reflection percentage compared to literature (Cooksey et al., 2015) only due to its settings in the spectrometer (Spectrometer 2) with higher values of the integration time. In previous studies, it is often <10% in this region. Though lower settings of the integrating time allows for more accurate readings, noise peaks render the region unusable for pattern analysis. This setting was chosen to maximize the smoothness of the output spectra, allowing features from the data to be extracted with less occlusion from the noise.

5.2.1.3 Minor features – Absolute values

Features discussed here are a combination of a bird's-eye-view of the spectrum, and a close-up scrutiny which allows for certain peaks or troughs to be observed.

While these features often span a range of 1-3% reflectance percentage, the presence of these features are important to study since these patterns are poorly understood in literature. Since these patterns are not coincidental and occur in most subjects, it is included as a feature. **Figure 5.5** elaborates these features:
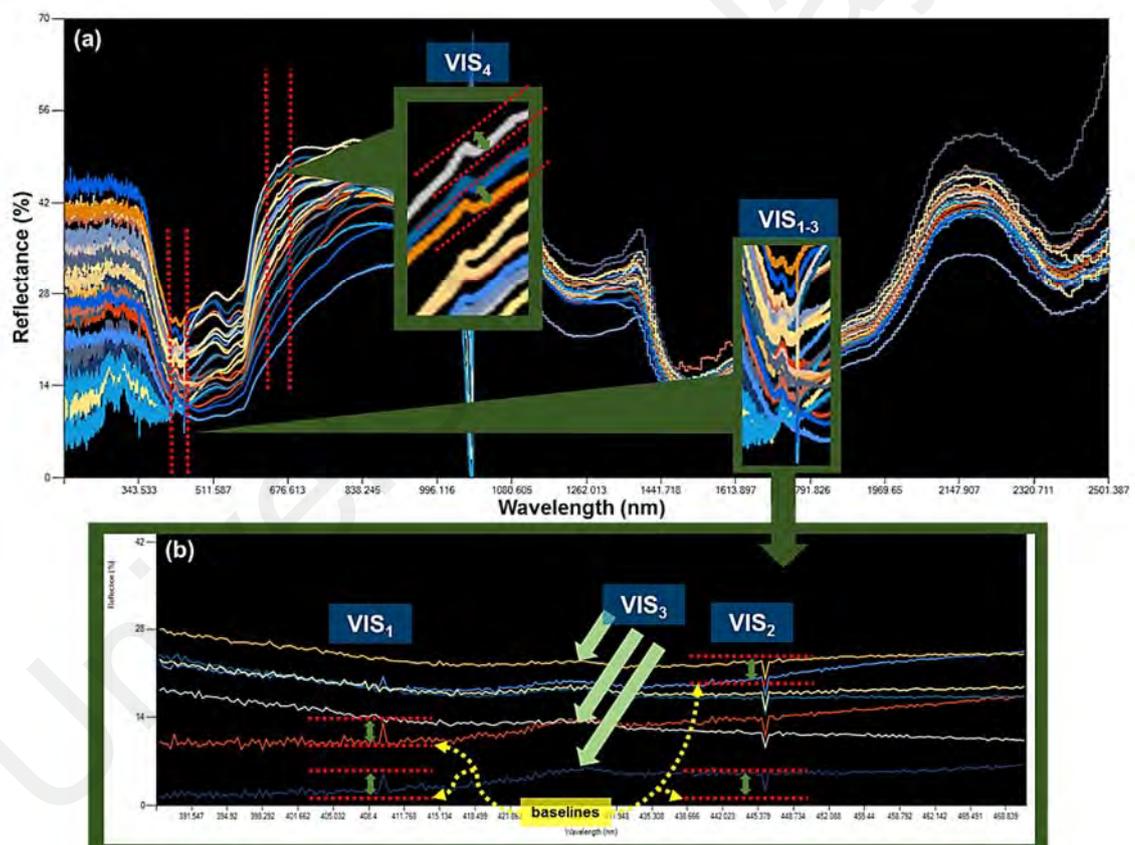
Figure 5.5: Minor scalar features.

Figure 5.5 shows the minor scalar features. These features signifies that subtleties of the pattern are more prevalent in the VIS region, denoted as *VIS*-features. *VIS₁* to *VIS₄* occurs at 409.41 nm, 446.05 nm, 426.235 nm and 652.394 nm respectively. As shown in **Figure 5.5(a)**, there are two regions which exhibit these minor features. The inset for

VIS_{1-3} as shown in **Figure 5.5(a)** is decomposed into **Figure 5.5(b)**, which bore three distinct features. VIS_4, in turn, constitutes a lone peak as shown in the inset for VIS_4 in **Figure 5.5(a)**. It is highly likely that these features are influenced by fluorescence emissions.

Fluorescence profiles, although different than reflectance, is in many cases positively correlated by the principle of superposition of both components, as discussed in several studies on separation of these aspects in spectroscopy (Fuchs, 2001; Gillies et al., 2000) and is a nontrivial computation issue (Ji, Miao, Zhang, Lin, & Dai, 2017). For this reason, we shall not attempt in decomposing these patterns further. In correlation with VIS_{1-3} features, similar profiles occurring within 400-460 nm shows considerate peak formation via fluorescence spectroscopy studies with different excitation sources. These properties were linked with both of the properties of collagen in the subcutaneous tissue and the skin dermis, though the contribution of the dermis to the peak was more significant (J. Chen et al., 2009; Gillies et al., 2000).

Pattern-wise, the general appearance of the spectra is feature-wise similar with previous studies (Glennie, Hayward, & Farrell, 2015; Hennessy, Markey, & Tunnell, 2015), though with differing baseline features. Note that as shown in **Figure 5.5(b)**, both VIS_1 and VIS_2 are more noticeable than VIS_3, hence the order of the variables' denotation.

5.2.1.4 Minor features – Gradient values

Figure 5.6 shows the remaining minor gradient features from the spectral pattern.
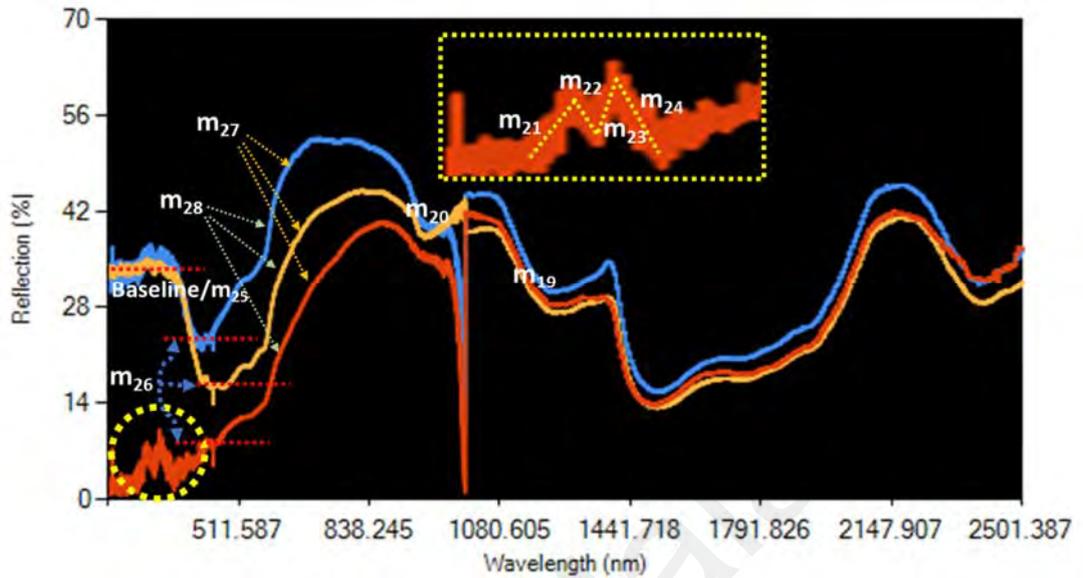
Figure 5.6: Minor gradient features. The yellow circle is magnified by the yellow-outlined inset

In the least visually noticeable features, three common spectra representing different major ethnicities in Malaysia (Malay, Chinese and Indian) is shown. As shown in **Figure 5.6**, m_{19} and m_{20} resides in the NIR region, with less pronounced features compared to previous counterparts in **Figure 5.3** to **Figure 5.5**. These features are still partly influenced by melanin content, but mostly water as discussed in the previous sections.

The patterns between m_{21-24} occur within the UV between 250-350 nm, and signify response toward the epidermis, consisting of aromatic amino acids, urocanic acids, nucleic acids and melanin which respond to the short wavelengths (Na, Stender, Henriksen, & Wulf, 2001). They are better exhibited with subjects with darker skin (such as Type V and VI skin). Two different subjects may have the similar baseline as shown by m_{25}, but the "double peak" pattern as shown in the inset in **Figure 5.6**, perhaps can be attributed to optical properties of stratum corneum (Cohen, 1977; Van Gemert, Jacques,

Sterenborg, & Star, 1989), may not necessarily be exhibited. This qualifies as a feature due to these aspects.

The following section summarizes the findings of these features, and provides a fresh lead towards multivariate analyses as will be discussed later.

5.2.2   Algorithms for Feature Extraction

To summarize the four aspects of feature extraction as highlighted in **Section 5.2.1**, the quantitative elements are further computed in this section. To recapitulate, there are (a) major scalar peak values of reflectance percentages, (b) major gradient and intersection values from the spectral pattern, (c) minor scalar values of reflectance percentage in reference to the respective baselines, and (d) minor gradients and intersection of non-obvious patterns.

While items (a) and (c) only requires direct access to its value from the database, the major and minor gradient and intersection features are computed as follows:

$$m_n = \frac{P_{\lambda_2} - P_{\lambda_1}}{\lambda_2 - \lambda_1} \tag{5.1}$$

m_n symbolizes the corresponding gradient, graded as per order of distinctiveness. $P_{\lambda 2}$ and $P_{\lambda 1}$ corresponds to the extremes of the range which forms the gradient, while λ_2 and λ_1 are the wavelengths within the mentioned range.

The intersection values, c_n are projection values to amplify the magnitude of the reflection percentages, computed as follows in Equation 5.2:

$$c_n = P_{\lambda_2} - \frac{P_{\lambda_2} - P_{\lambda_1}}{\lambda_2 - \lambda_1} * \lambda_2 \tag{5.2}$$

c_n is simplified as per the linear expression in Equation 5.3:

$$c_n = P_{\lambda_2} - \ m_n * \lambda_2 \qquad\qquad (5.3)$$

In simpler terms, c_n denotes the y-intersection values for the 2-D data topology in Cartesian geometry. **Table 5.1** itemizes all 67 features, with corresponding wavelengths.

Table 5.1: Feature Extracted Parameters

Feature Parameters	Wavelengths (nm)	Definition
a) Scalar Values		
P_1	650.00-850.00	Peak values in both VIS and NIR
P_2	2100.00-2250.00	
Q_1	515.92	Values of the double/ troughs. If these peaks/troughs do
Q_2	545.51	not exist, scalar value is taken.
Q_3	566.73	
Q_4	575.00	
Baseline	175.00-285.00	Average value along this band
b) Gradient Values		
m_1, c_1	485.23 - 505.25	Slopes in VIS and NIR. Note that m_9 is predominantly an
m_2, c_2	520.25 - 531.89	important factor in isolation in ethnicity.
m_3, c_3	551.82 - 561.76	
m_4, c_4	566.73 - 571.70	All in the form of a range.
m_5, c_5	735.54 - 752.05	
m_6, c_6	767.89 - 784.34	
m_7, c_7	961.41 - 970.50	
m_8, c_8	976.76 - 989.25	
m_9, c_9	1332.79 - 1371.29	
m_{10}, c_{10}	1390.52 - 1435.32	
m_{11}, c_{11}	1435.32 - 1480.06	
m_{12}, c_{12}	1531.10 - 1588.44	
m_{13}, c_{13}	1690.20 - 1734.67	
m_{14}, c_{14}	1766.43 - 1861.67	
m_{15}, c_{15}	1956.94 - 2020.51	
m_{16}, c_{16}	2052.33 - 2084.16	
m_{17}, c_{17}	2243.77 - 2275.80	
m_{18}, c_{18}	2417.28 - 2481.94	
m_{19}, c_{19}	1093.63 - 1158.59	
m_{20}, c_{20}	995.65-1041.458	
m_{21}, c_{21}	242.09 - 275.89	Minor gradients in UV region. All in the form of
m_{22}, c_{22}	275.89 - 295.31	wavelength range.
m_{23}, c_{23}	295.31 - 310.27	
m_{24}, c_{24}	310.27 - 327.25	

m_{25}, c_{25}	644.20 - 676.94	Minor gradients in VIS and NIR.
m_{26}, c_{26}	594.83 - 611.32	
m_{27}, c_{27}	360.80 - 394.58	
m_{28}, c_{28}	841.76 - 870.45	
Minor Features		
VIS_1	409.41	A minor peak relative to the surrounding baseline i.e minus the value at 409.1 nm or 409.7 nm
VIS_2	446.05	Values relative to a baseline which is defined around 445.04 nm
VIS_3	426.235	
VIS_4	652.394	Relative to the baseline value at 649.12 nm

The final form of the feature extraction is shown in **Figure 5.9** in the following.
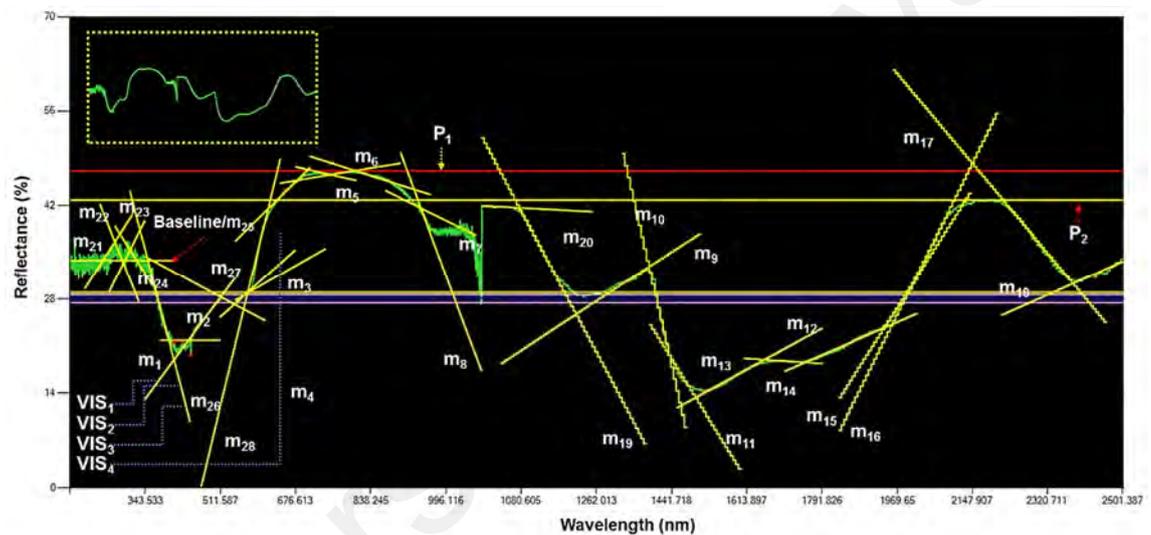
Figure 5.7: Compiled feature-extracted parameters from an example spectrum. Q_1-Q_4 are signified as the four closely-spaced banded lines. The inset shows the original spectrum.

There are a total of 67 features generated from the spectrum as shown in **Figure 5.7**, which significantly reduces the aim to find structure within the data. This is performed in the following section using multivariate analysis of the parameters to further reduce the dimensions. The following section elaborates on the use of these features.

5.3   Principal Component Analysis (PCA): Decomposing Data Structure and Confounding Factors and Cross-Validation of Feature Extraction

There are three main reasons to analyze the data in principal components, which is for (a) finding structure amongst different groups (ethnicity, gender, age and clinical diagnosis), (b) if structure is found among ethnicity, gender or/and age, it qualifies as a normalization (or confounding) factor, and (c) cross-validate the use of feature-extracted data in the principal component space, where if data structure can be gained from the features, additional accuracy can be obtained for classification.

As outlined in the chapter structure in **Figure 5.1(c)**, this section attempts to transcend the data obtained into a more useful observation. Principal component analysis (PCA) is briefly described as follows. In basic representation of data with two or three parameters, this often is performed with relative ease using graphical representation. The plots which result in such exercise allows one to draw conclusions on whether patterns exhibited fit a certain pattern.

However, most scientific data covers a broad range of variables. In this instance, a 2-D or 3-D representation of a data with more than three variables is impossible. This gave rise to the use of PCA, which reduces the parameters to a 2-D or 3-D representation by transforming the variables, or covariates from its original matrix into a covariance matrix.

All PCA computations and graphs are produced using *JMP Pro 12.2.0* (SAS, 2015). When structure is found from the data by means of a 2-D biplot, we can conclude that the parameters involved in the analysis correlate positively with the classification. As demonstrated in the previous section, the 97.7% reduction from 2,974 parameters from the spectroscopic data to 67 parameters allows us to examine the data structure. Also, we elaborate the rationale of representing the data in the form of ethnicity and gender.

We demonstrate the primary structure of the data by inclusion of all ranges of the raw spectroscopy data (without feature-extracted data), spanning 2,974 parameters from the two spectrometers as discussed in previous sections. This allows us to put the first two sections in this chapter into perspective on the advantages or disadvantages of the modes of data. All of the plots in this section were rendered using the Wide method[5], since the number of parameters p significantly outweigh the number of subjects n ($p>>n$).

In a parallel analysis, we attempt to improve the clustering of the different groups by comparing a fully-utilized feature-extracted covariates versus selected covariates. A satisfactory structure is construed as having visually discernible groups in the PCA biplot. As discussed in **Section 5.3.1**, PCA on raw data was performed on all 2,974 data rows, using the Wide method, which is a complex and inefficient analysis method. By reducing the number of variables to a set of more manageable dataset, we can perform the analysis more efficiently.

However, achieving maximum group separation in a biplot is dependent on selecting variables. In this section, only optimal combinations of the 67 parameters as discussed in **Section 5.2** (with covariates P_{1-2}, Q_{1-4}, m_{1-28}, c_{1-28}, *Baseline* and *VIS_{1-4}*) generated from feature extraction is presented. For the optimal permutation achieved, the covariates used in a certain category (gender, ethnicity and age group) can be deduced as influential to the category. Other permutations of the covariates are not covered in this section for the sake of brevity. Notable works using PCA by selective methods due to ethnicity and gender have used for facial recognition and (C. Chen, Chang, Ricanek, & Wang, 2010; Givens, Beveridge, Draper, & Bolme, 2003). It is also noteworthy that features as

---

[5] According to John Sall of SAS, Wide method is basically required for efficient computations with limited memory as an alternative from the computation of the covariance and inverse covariance matrix, as per the case in many instances for $p>>n$. His articulation was recorded in the following: https://community.jmp.com/t5/JMP-Blog/Wide-data-discriminant-analysis/ba-p/30585

mentioned in these studies are also somewhat subjective, which sheds light into the subjective nature of our methods as well.

5.3.1   PCA – Gender

Figure 5.8 in the following highlights the first demographic group as first highlighted in Chapter 4, specifically gender, on both fully raw and feature-extracted data.
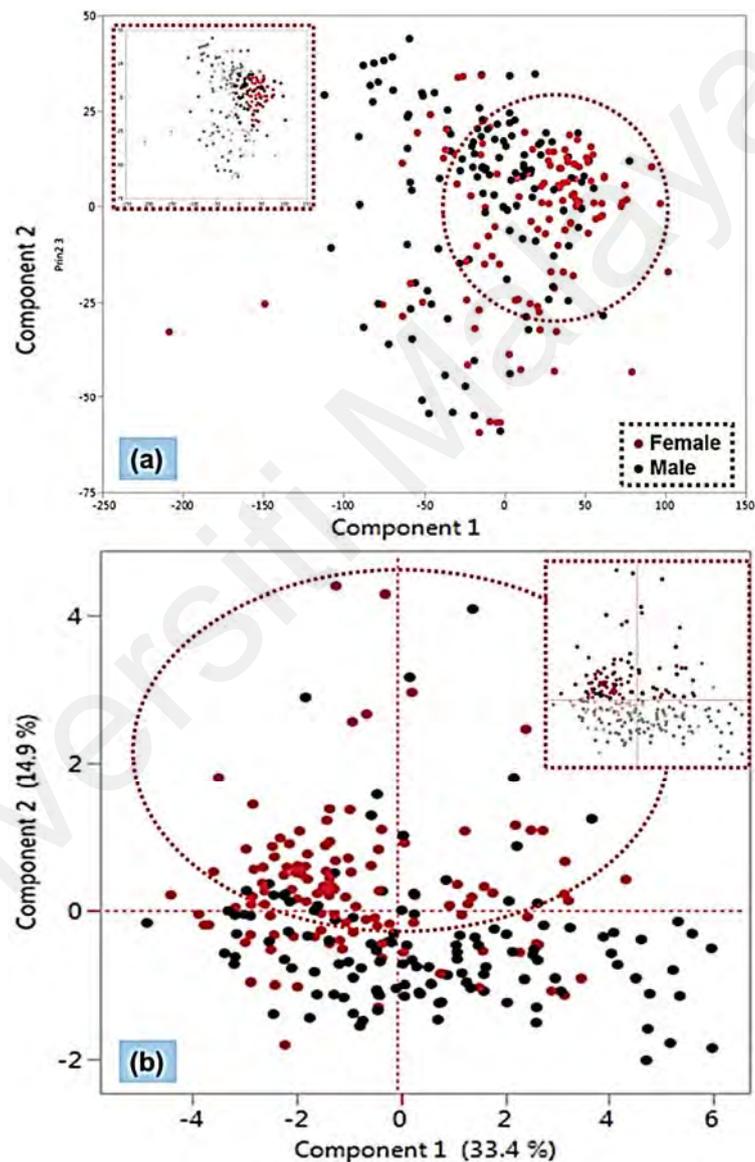
Figure 5.8: PCA biplot on (a) whole raw spectroscopy data, and (b) selected feature data, both gender highlighted. Inset in both figures shows selected patients approximately corresponding to the red dashed circle.

As shown in **Figure 5.8(a)**, females tend to precipitate along the positive side of the *Component 1* axis. Out of the highlighted region, totaling 116 patients, 85 (73.3%) of the

highlighted region are females (indicated by the red dots), versus 31 males (26.7%). This suggests that gender plays a partial role in the determining the spectrum pattern. This can also be observed in the collective spectra on genders as highlighted in **Figure 4.7(c)**, where the spectrum patterns differ mostly in the UV and VIS range (173-800 nm). These differences also have been observed within the same range in previous research (Prabhu et al., 2014).

With feature-extracted data, **Figure 5.8(b)** shows the augmentation of gender precipitation in the biplot especially along the *Prin2* axis. The corresponding PCA plot for feature-extracted data, based on information on fluorescence patterns in mice (Prabhu et al., 2014) (in relation to gender), we included the m_1-m_5 features as discussed in **Section 5.2** and additionally m_6-m_{15} for potential autofluorescence in these regions. This results in an improved disparity between males and females. The circled region covers 93 subjects, where 72 (77.4%) females reside in the region in contrast to 21 (22.5%) males in the region, demonstrating both the validity of feature-extracted data, and the influence of gender within these bands. However, since *Component 2* axis corresponds to a lower correlative strength compared to *Component 1*, it can be deduced that at least another confounding factor resides in the spectrum. In this case, we have completely omitted the use of the scalar amplification corresponding to c_{1-15} features, which removes the reflection percentage level as a factor, and dealing only with gradient pattern analysis.

In literature, however, the correlation between reflectance patterns versus gender is scarce. However, several works have pointed out the skin collagen content, which is well-established fluorophore (i.e. emitting light upon an excitation light source) (Zonios et al., 2001) between males and females have notable differences (Z. Li et al., 2017), also as demonstrated in fluorescence spectroscopy in mice within 400-750nm (Prabhu et al., 2014). Similarly, reflectance spectroscopy profiles as shown in **Figure 4.7(c)** in Chapter

4 also show noticeably different reflectance percentage and pattern of males compared to females, as also attested in independent works (Calabro, Curtis, Galarneau, Krucker, & Bigio, 2011; Prabhu et al., 2014).

5.3.2 PCA – Ethnicity

Figure 5.9 in the following shows the PCA plot with an emphasis on ethnicity.
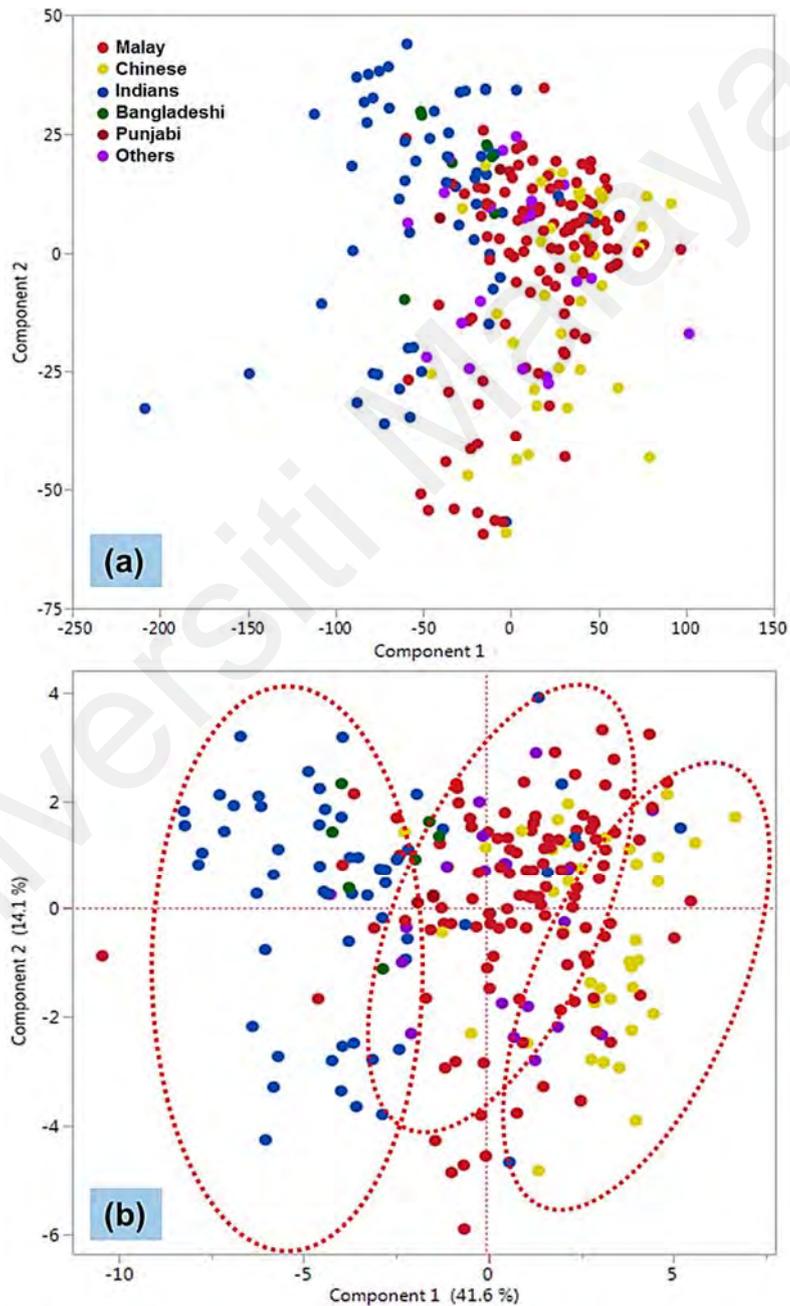
Figure 5.9: PCA biplot on (a) whole raw spectroscopy data, and (b) selected feature data, ethnicity highlighted.

Figure 5.9 shows the combination of the first two components illustrating the aspect of ethnicity or skin type when all parameters are used for the analysis. It is to be noted that this included the "noisy" band occurring in the right side of the bandwidth for Spectrometer 1 between ~950-1070 nm. This is partially compensated by readings of Spectrometer 2 which spans ~980-2500 nm. The inclusion or exclusion of this band, however, has proven to be negligible to the final outcome of the PCA biplot.

Table 5.2 shows the first four significant components.

Table 5.2: Eigenvalues from PCA

Number	Eigenvalue	Percent	Cumulative Percent	Visual	Singular Value
1	2022.6	68.034	68.034		695.28
2	487.5	16.397	84.430		341.33
3	287.3	9.6649	94.095		262.06
4	90.8	3.0556	97.151		147.35
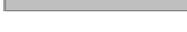
Based on **Table 5.2**, in this rendition, there are four significant components, out of 46 components from the spectrum (21 excluding the "noisy" band). It is to be noted that after the fourth component, the percentage of variabilities are redundant. This signifies at least four distinct groups from the data, in this case most likely due to the skin types, or indirectly ethnicity. By color-coding the series in **Figure 5.9**, it can be seen that the ethnicities play an important role in the data structure, where some Malays, Indians and Bangladeshis correlate positively on the left side of the *Component 1* axis due to generally having a Type VI or V skin (Alaluf et al., 2002; Bin Yap, 2010; Huggins et al., 2012; Ong, Tan, & Cheng, 2018; Zambelis et al., 2017). *Component 2* axis, however, which ranks below *Component 1* in its variability plays a lesser role in showing the density of the ethnicities on the biplot, though both components allow better visualization.

A majority of Malays and Chinese are classically similar in their skin type, divided between Type III and VI (Alaluf et al., 2002; Bin Yap, 2010) as shown in **Figure 5.9**. In

further analysis, we will demonstrate how this plays an important role in normalizing the data.

As for selective feature-extracted data, the parameters m_{1-12}, c_{1-12} and the *Baseline* were included after several permutation iterations for the best separation. As shown in **Figure 5.9(b)**, cluster appeal is more noticeable in this region. This is intuitively the case, due to melanin content of different skin types collectively contribute to the reflectance percentages between 1730-1200 nm, which is the range covered by the selected feature covariates in our case. It can be seen that *Component 1* is primarily the main factor for the ethnicity scatter.

5.3.3   PCA - Age

Figure 5.10 shows the highlighted plots emphasizing three age groups corresponding to the demographics as mentioned in **Figure 4.2** (Chapter 4), namely (**A**) less than 15 years old (*<15 years*), (**B**) between 15 years old to 64 years old (*15 ≤ years ≤ 64*), and (**C**) more than 64 years old (*>64 years*).
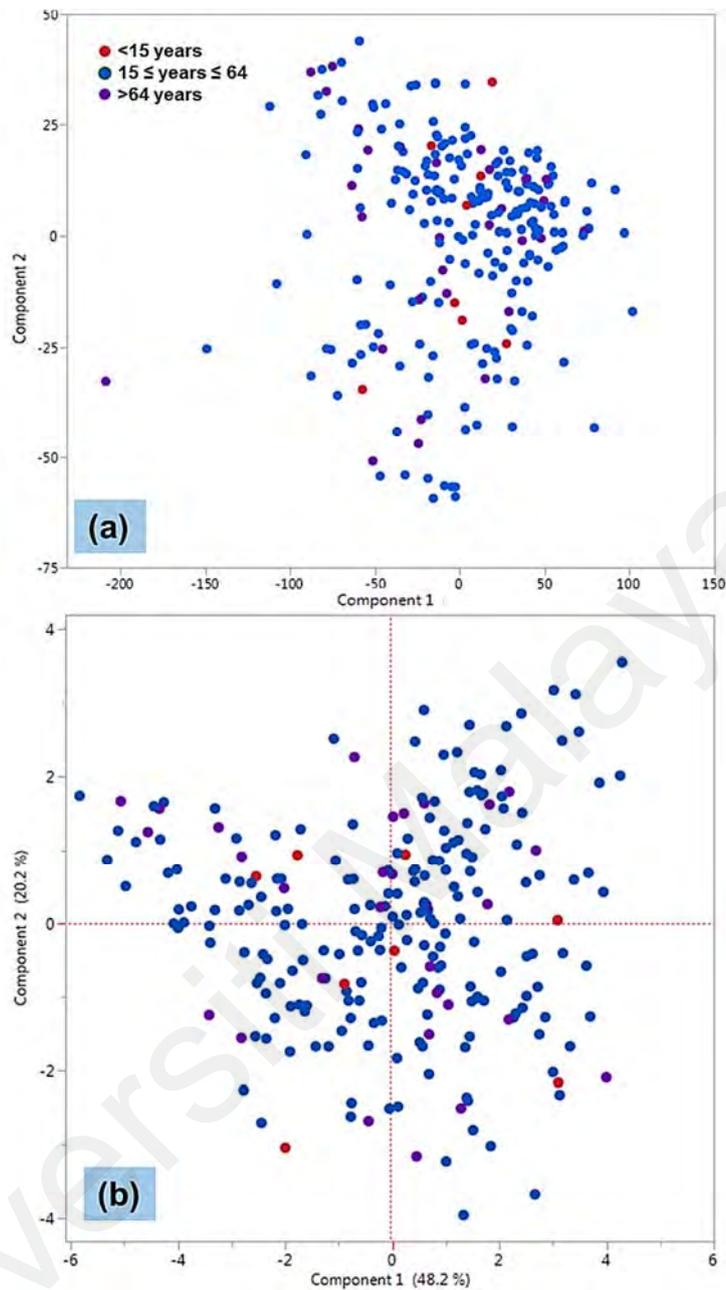
Figure 5.10: PCA biplot on whole raw spectroscopy data; age highlighted.

Although age has predominantly been established to influence the spectral patterns in fluorescence spectroscopy, it can be seen that there is virtually an absence of structure of the scatter in **Figure 5.10**. Based on the several works, the reflectance percentage in fluorescence spectroscopy has reportedly shown notable disparity between age groups. These differences were attributed to higher levels of pigmentation, or melanin due to higher exposure against UV from sunlight as subjects' age progresses (Lock-Andersen, Knudstorp, & Wulf, 1998; Na et al., 2001). However, in our case, the PCA plots have

shown that it has negligible effect on the clustering tendencies of the plot. As also shown in **Figure 4.8** (Chapter 4) there are negligible differences in the spectral pattern between the three age groups. This was shown with small reflectance percentage differing between **A** (*<15 years*) combined with **B** (*15 ≤ years ≤ 64*) compared to **C** (*>64 years*) in the region between 600-900 nm, whereas in literature, the regions of fluorescence spectroscopy corresponding to age can be observed within 555-660 nm (Na et al., 2001). It can be deduced that since fluorescence is a substantially less significant component of the reflectance spectrum, the difference between the age groups are not noticeable in this case.

As for **Figure 5.10(b)**, since it was shown that the variability between group **A**, **B** and **C** was within 600-900nm, and 550-660 nm in fluorescence mode (Na et al., 2001), an educated estimate employs the use of m_{3-7} and c_{3-7} features which covers these bands. However, similar to **Figure 5.10(a)**, no utility was derived from the biplot, suggesting that at least in reflectance mode, age plays a negligible role as a confounding factor.

5.3.4   PCA - Confirmed/Probable Dengue versus Non-dengue

Finally, **Figure 5.11** highlights the confirmed and probable dengue cases, versus other control cases while excluding the patients due to reasons discussed in Chapter 4, **Section 4.3**.
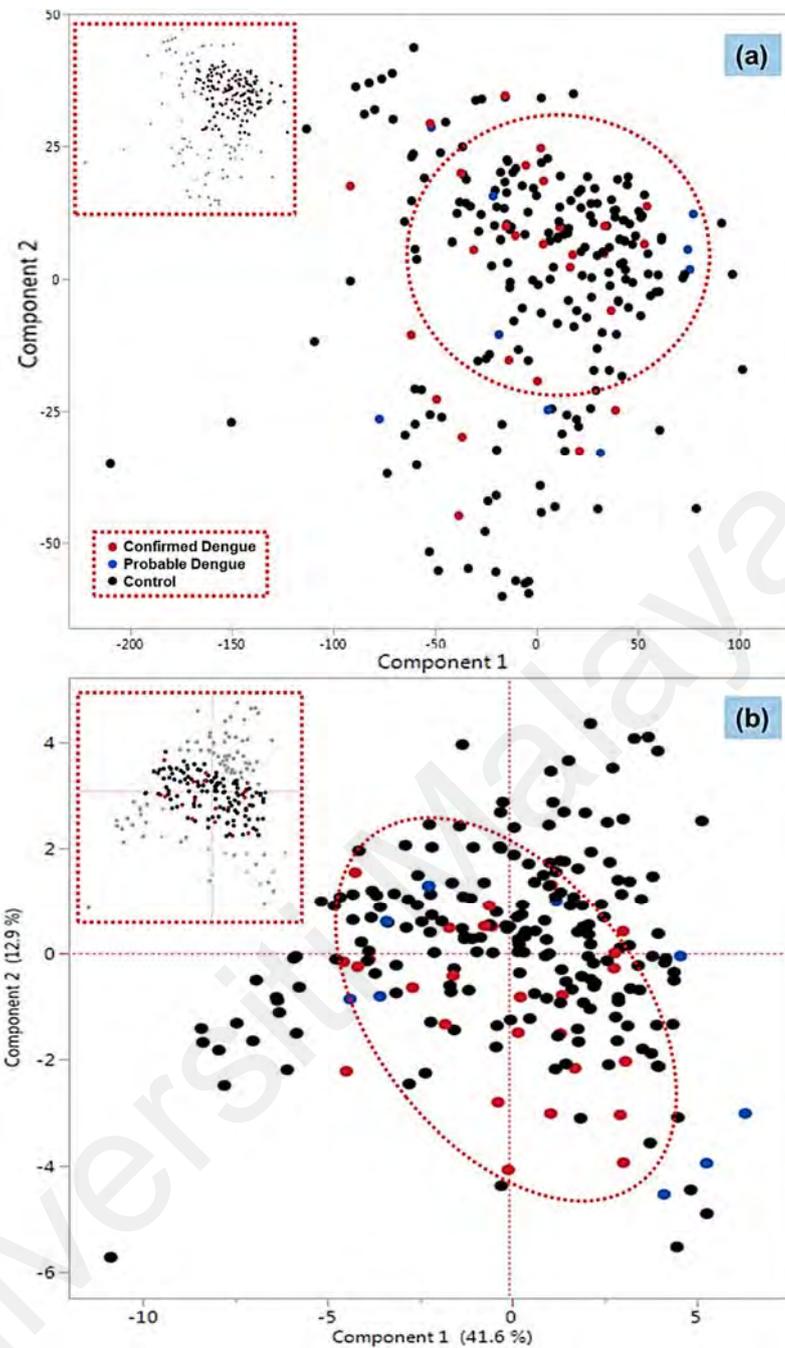
Figure 5.11: PCA biplot on (a) whole raw spectroscopy data, and (b) selected feature data, confirmed/probable dengue highlighted. Inset in both figures shows selected patients approximately corresponding to the red dashed circle.

As shown in **Figure 5.11(a)**, it appears that there is virtually no structure exhibited by confirmed and probable dengue cases amongst the control population. In short, confirmed and probable dengue groups are uniformly scattered throughout the biplot. It is to be noted although there seems to be tendency for convergence of both 18 (64.3%) of confirmed

and 4 (50.0%) probable dengue groups[6] as highlighted in **Figure 5.11**, it does not qualify as a data structure due to the significantly higher concentration of control subjects in the same area. This is understandably the case due to the subtleties of the reflectance pattern on both groups in contrast to the control as also shown in the raw spectrum graphs in **Figure 5.2**.

Figure 5.11(b) highlights the feature-extracted population, which shows an improved structural tendency by utilizing m_{1-12} and c_{1-12}. As shown by the ellipse constructed in **Figure 5.11(b)**, it can be shown that both *Component 1* and *Component 2* axes indicates multiple correlation of several parameters. 12 eigenvalues were generated from the scores, with 9 can be considered of significant influence, indicating multiple factors in the data structure. 20 (74.1%) confirmed dengue, and 4 (44.4%) probable cases reside in the boundary as shown in the inset. However, both confirmed and probable dengue only occupy 16.9% of the region's population of 142 subjects, indicating a complex interaction with the covariates. From here, we have demonstrated that a commonality in dengue patients from their spectroscopic data can be derived. Therefore, this exercise necessitates the use of discriminant analysis, which assigns the weightage of each parameter to its class, in this case, the three groups of confirmed or probable dengue, and control groups. The conclusion of this finding will be elaborated in the next final section.

---

[6] In this case, the percentages refer to the class population. Confirmed dengue percentage is 18 (64.3%) of 28 cases, while probable dengue is 4 (50.0%) of 8.

5.4　　Discriminant Analysis – Non-normalized Data

As discussed in the chapter structure shown in **Figure 5.1(d)**, this section concludes the analytics by further grounding the multivariate nature of the data via classification algorithms, known as discriminant analysis (DA).

To describe briefly[7], discriminant analysis is a method of separation of class objects by a set of continuous independent variables and a category variable. Similar to PCA, it uses a linear combination of the variables, but with the added weightage of the class categories which assists in maximizing class separation. This method was used for classifying closely-related patterns, where visually subtle outliers in a spectroscopic data finds it applications.

There are three subsections of DA that will be used in this part. The first section is to classify the data based on (a) raw and (b) feature-extracted data of the three demographic measures (gender, ethnicity, age group) and on the clinical diagnosis (discriminating dengue patients from control). This part marks the end of proving the influence (or non-influence) of confounding (or normalization) factors, namely gender, ethnicity and age groups. Also, in this subsection, feature-extracted data is combined with the raw data to maximize the algorithms' accuracy[8].

The next subsection focuses purely on normalized data based on the demographics as discussed in the previous sections. This includes solitary normalized data (gender and

---

[7] Discriminant Analysis (DA) was largely credited to Sir Ronald Fisher in 1936 who introduced the famous "Iris flower dataset". This dataset has been used to test multiple multivariate data analysis, and is the cornerstone of modern analytics for data with multiple parameters.

[8] The combination of different data types for a multivariate environment is considered cascading, which is while not commonly performed due to its experimental nature, is a developing technique (Katajamaa & Orešič, 2007). This allows increasing chances of classification accuracy with a high number of covariates.

ethnicity) or a combination of different classes of normalization by combining both modes with different operands.

The third subsection finally summarizes all the previous permutations into a binary representation, with only confirmed dengue and control as the outcome. This is concluded with grounding the analysis with measurements of sensitivity versus specificity, to demonstrate the screening potential of a diffuse reflectance spectroscopy for dengue.

How and why dengue patients can be screened with this method will be briefly discussed as a closure to this chapter. All DA biplots and analysis in this subsection are performed with *JMP Pro 12.2.0* (SAS, 2015).

5.4.1 Discriminant Analysis – Raw and Raw Combined with Feature Extracted Data

Figure 5.12 in the following shows four discriminant canonical plots showing the response of the raw spectra versus the demographic categories and clinical diagnoses.
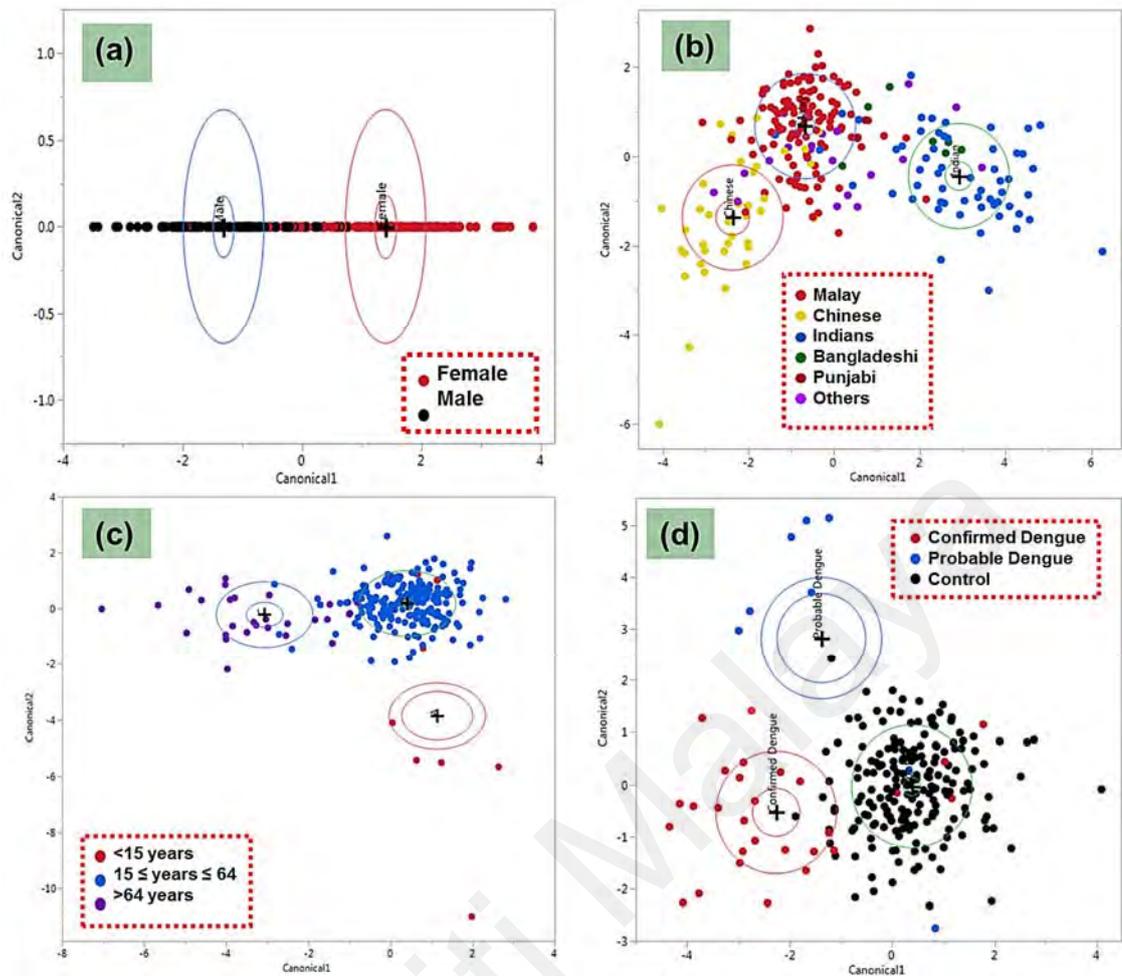
Figure 5.12: DA on raw data – demographics and clinical diagnoses

Table 5.3 shows the breakdown of the classifications for each category as per the discussion. All of the classifications were rendered under JMP software as mentioned earlier.

Table 5.3: Breakdown on classification accuracy for raw spectra

Class	Count	Misclassified	% Misclassified	Breakdown			
Gender	240	22	9.16667	**Actual**	**Predicted Count**		
				Gender	**Female**	**Male**	
				Female	104	12	
				Male	10	114	
Ethnicity	211	22	10.4265	**Actual**	**Predicted Count**		
				Ethnicity	**Chinese**	**Indian**	**Malay**
				Chinese	31	0	8
				Indian	0	53	5
				Malay	7	2	105
Age Group	240	14	5.83333	**Actual**	**Predicted Count**		
				Age Group	**A**	**B**	**C**
				A	5	3	0
				B	0	196	6
				C	0	5	25
Clinical Diagnoses	230	15	6.52174	**Actual**	**Predicted Count**		
				Category	**Confirmed Dengue**	**Control**	**Probable Dengue**
				Confirmed Dengue	23	4	1
				Control	10	180	4
				Probable Dengue	0	2	6

Figure 5.12 shows the implementation of all raw spectra (173-2500 nm) spanning 2,974 data points into the Discriminant Analysis, without any normalization. Beginning with **Figure 5.12(a)**, we can visually see a very well-defined separation between genders, with a small population of overlapping. This confirms the influence of gender in the spectrum, at a 90.83% accuracy in classification with 12 (5%) and 10 (4.1%) misclassified females and males.

Figure 5.12(b) shows the ethnicity limited to the three major Malaysian groups, namely Malays, Chinese and Indian. As for the other groups, classification is performed by the existing model in JMP by training the pre-classified data. Bangladeshis, who generally had Type V skin type (Huggins et al., 2012; Zambelis et al., 2017), predictably fell within Indian cluster with similar type IV and V skin types (Koh et al., 2003). Also, similar to PCA clustering tendencies observed in **Figure 5.9** between Malays and Chinese

classes, the centroids of Malays and Chinese groups are less distant compared to Indians and Bangladeshis. This again, demonstrates the higher overlap in skin types between Malays and Chinese, although Malaysian Chinese subjects intuitively have a tendency for higher-tiered skin type (Type III) due to lower melanin content in the skin (Alaluf et al., 2002; Bin Yap, 2010).

Figure 5.12(c) shows the age classification. Though clustering was considered wholesome for **Group B** (*15 ≤ years ≤ 64*), other subjects have high misclassification rates, with only 5 (62.5%) out of 8 were correctly classified for **Group A** (*<15 years*), though **Group C** (*>64 years*) had higher accuracy with 25 (83.3%) out of 30 compared to **Group A**.

Finally, **Figure 5.12(d)** shows the classification of both probable and confirmed dengue cases against control subjects. In this case, confirmed dengue cases returned 23 (82.1%) out of 28, while probable dengue returned 6 (75.0%) of 8 cases, in contrast to a 177 (91.2%) of 194 control cases. This excluded 10 patients due to reasons stated in **Section 4.3** in Chapter 4. In this instance, we insist that the accuracy is not yet ideal due to the significantly lower number of dengue patients ($n_{confirmed} = 28$) compared to control ($n_{control} = 194$), which may allude sampling bias problems. This argument will be refined in the next section.

In contrast to **Figure 5.12, Figure 5.13** in the following shows four discriminant canonical plots showing the response of the raw spectra plus feature-extracted data versus the demographic categories and clinical diagnoses. With increased number of covariates, a higher classification accuracy is expected.
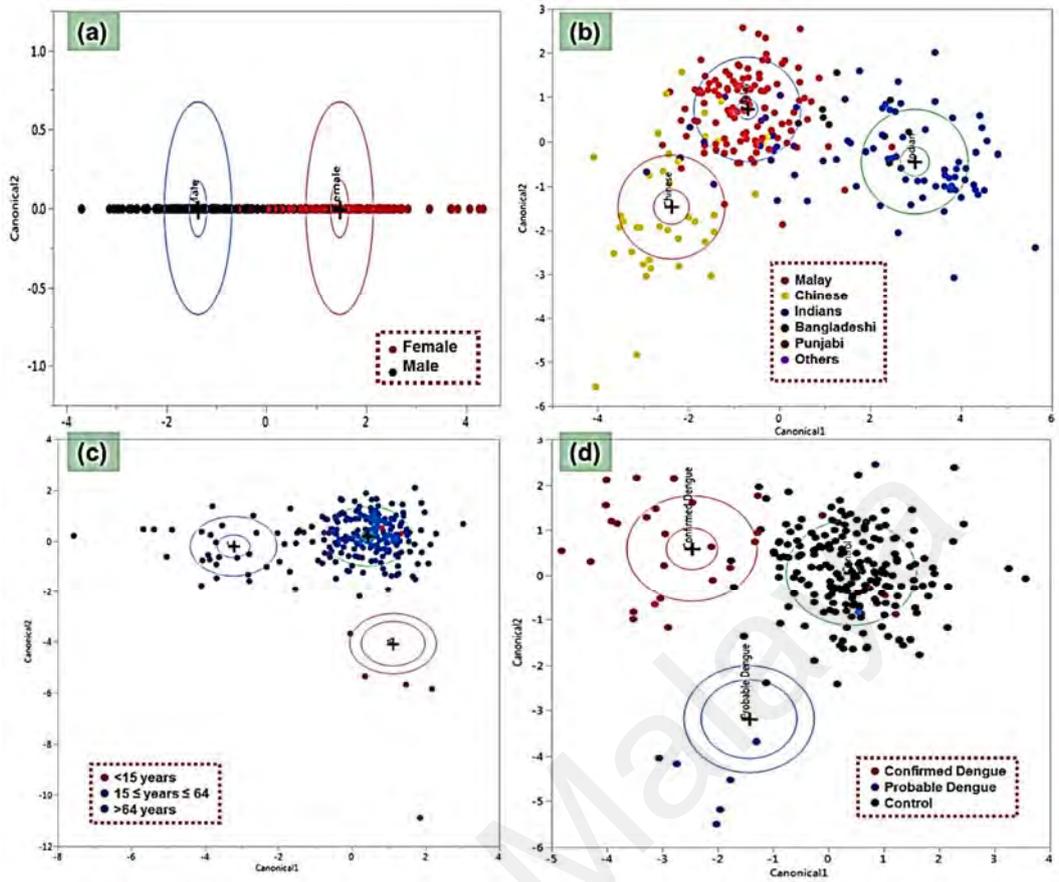
Figure 5.13: DA on raw plus feature-extracted data – demographics and clinical diagnoses

The analytics of this section is broken down into **Table 5.4.**

Table 5.4: Breakdown on classification accuracy for raw plus feature extracted data

Class	Count	Misclassified	% Misclassified	Breakdown		
Gender	240	18	7.50000	**Actual**	**Predicted Count**	
				Gender	**Female**	**Male**
				Female	106	10
				Male	8	116
Ethnicity	211	20	9.47867	**Actual**	**Predicted Count**	
				Ethnicity	**Chinese**	**Indian** / **Malay**
				Chinese	31	0 / 8
				Indian	0	53 / 5
				Malay	4	3 / 107
Age Group	240	15	6.25000	**Actual**	**Predicted Count**	
				Age Group	**A**	**B** / **C**
				A	6	2 / 0
				B	1	194 / 7
				C	0	5 / 25
Clinical Diagnoses	230	15	6.52174	**Actual**	**Predicted Count**	
				Category	**Confirmed Dengue**	**Control** / **Probable Dengue**
				Confirmed Dengue	24	4 / 0
				Control	5	185 / 4
				Probable Dengue	0	2 / 6

As shown in **Figure 5.13**, visually if compared to **Figure 5.12**, all plots are similar despite small differences. However, based on **Table 5.4**, the classification based on the statistical model produced in this case show higher accuracy. An increase of 1.66% accuracy for gender classification was achieved (represented by **Figure 5.13(a)**), with improved male categorization.

For ethnicity as shown in **Figure 5.13(b)**, a slight increase of 0.95% was achieved. It is to be noted as well that in both **Figure 5.12(b)** and **5.13(b)**, the classification of ethnicity was performed by maintaining only the categories of three major Malaysian ethnicities (Malay, Chinese and Indian). This was done to avoid complicating the

classification algorithm[9]. The utility of this step, in turn, will be elaborated as part of an objective normalization procedure.

For the age groups, the use of feature extracted data as shown in **Figure 5.13(c)** returns lower accuracy levels by a margin of 0.42%. As for clinical diagnoses covered in **Figure 5.13(d)**, the accuracy increased to a slight 0.30%, though a lower accuracy for confirmed dengue cases, while the overall increase of the accuracy is attributed to the control group, with subjects algorithmically reclassified correctly as indicated in the breakdown column in **Table 5.4** and **Table 5.3** by comparing the numbers indicated by Clinical Diagnoses. One common feature observed in both **Figure 5.12(d)** and **5.13(d)** is that both probable and confirmed dengue cases exhibit highly common properties by mirroring plane against the control group, suggesting a tendency in dengue patients of both categories to converge in the primary *Canonical1* axis.

Considering a five-fold demonstration from (1) visual inspection of the graphical reflectance spectrum, (2) PCA and (3) DA of both the raw and feature-extracted data, this section concludes the following. These conclusions are (a) gender is a normalization factor, (b) ethnicity is also a normalization factor, while (c) age is a negligible factor in determining the classification, and (d) classification of confirmed dengue requires normalization to mitigate sampling bias issues. We address this in the following section.

---

[9] This is due to the fact that the values from both *Canonical1* and *Canonical2* scores explain 100% of the data variability (as per elaboration by exchanges with SAS personnel), and more categories induce confusion in the apparent "clusters" in the canonical plots.

5.4.2   Discriminant Analysis of Normalized spectra

To provide a perspective on how the algorithms might be field-deployed for clinical use, we deal exclusively with clinical diagnoses for confirmed or probable dengue, versus control subjects in this section. This is by providing a dichotomous analysis, with probable dengue cases reclassified either into confirmed dengue or control cases.

By normalizing the data, we attempt to mold the spectra into a form which removes its redundant features, and at the same time augmenting salient features in the pattern. This increases classification accuracy models. Generally, normalization techniques in literature concerning skin spectroscopy have been based on regions of well-known responses, such as oxygenated blood (HbO_2) and reactive hyperemia (RH) are normalized based in regions of 542 nm and 574 nm (Sprigle et al., 2002), while full-range normalization within 600-1200 nm was performed for studying collagen fiber orientation in human skin (Nickell et al., 2000). The studies covering normalization techniques are scarce due to most studies generally involving small number of patients, without specifying the studies.

For the sake of simplicity, we normalize the spectrum of each subjects based on ethnicity and gender separately, while analyzing separate and combined cases of the data modes in a non-exhaustive manner. All normalized spectrum are computed using VB.NET's Microsoft Excel Interop to produce the compendium, which spans 5,000+ parameters after combining with the original dataset, and feature-extracted data.

Though conventions on normalization vary according to usage[10], we use a subtractive operand for normalization, commonly known as a mean-centered operation, also used in

---

[10] There were arguments on distinctions between centering, standardizing and normalizing in the field of data transformation, which are all computationally different. However, for the sake of keeping the semantics consistent in this work, centering is construed as normalization.

several other research in biological data analytics (van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006), and specifically in skin spectroscopy (Vrabie et al., 2007). This method is computationally efficient, and allows minimal data processing, as shown in **Equation 5.4**.

$$R_{normalized} = R_{\lambda_{raw}} - R_{\lambda_{average}} \qquad\qquad (5.4)$$

R_{normalized} is the value used to either be utilized directly in discriminant analysis, or to be convoluted later for the analysis. R_{λraw} is the original value of the spectrum, and R_{λaverage} is the average value corresponding to either gender or ethnicity. For example, normalized data of a Chinese female would have two separate datasets, with one set centered to the average reflectance spectrum of females, and the other set centered on the average reflectance spectrum of Chinese subjects. It is to be noted that this equation is applied to every 2,974 rows of the data with its corresponding normalizer.

Due to limited number of subjects from ethnicities other than Malay, Chinese or Indian to establish a proper average reference, these subjects are ethnically imputed based on the classification algorithms obtained in **Figure 5.13(b)**. For instance, Indonesians were intuitively either classified as Chinese or Malays based on the Canonical plots in **Figure 5.13(b)**, depending on the nearest cluster and resulting classification. This is to say, a male Indonesian, if classified as a Malay, will be centered to males and Malay, producing two separate normalized datasets, alongside with all subjects as well. This is well supported by the fact that Asians generally exhibit skin types within the spectrum of the three major Malaysian ethnicities.

It is to be noted the problematic issue of objectively defining ethnicity in several circles of the scientific community. Generally in our case, Chinese and Indian subjects fall within the category provided by a widely used glossary (Bhopal, 2004). Incidentally, this is the case for Malaysian Government which retain these data as per identification cards.

Malays, on the other hand, are classified only by means accessible by the identification documents under the purview of the Malaysian Government.

In this section, we substitute the probable dengue group with a blank as a test to see the algorithmic categorization. This will return a dichotomous result; either confirmed dengue or non-dengue control, while probable cases will register as either. For further classification augmentation, a final data component of the covariates are included, which are the patient vitals, namely the temperature, systolic and diastolic rates, respiratory rate, SpO_2 and pulse rate. 18 patients in the dataset, however, were imputed for data loss as per the general guideline of imputations (Sterne et al., 2009). The use of this variant of data is deemed useful for observation purposes as will be discussed later.

The crux of the analysis in this section is finally represented by convolutions of the normalized data. This combines the augmentation of the normalization modes, and enables better discriminant classification.

5.4.2.1 Gender- and Ethnicity- Normalized Data: Separate Analysis

Figure 5.14 presents the discriminant analysis on convolution between the two modes of normalization by different operations. In following parts, when the term *data residue* is mentioned, this encompasses all original spectra, feature-extracted data, and vitals.
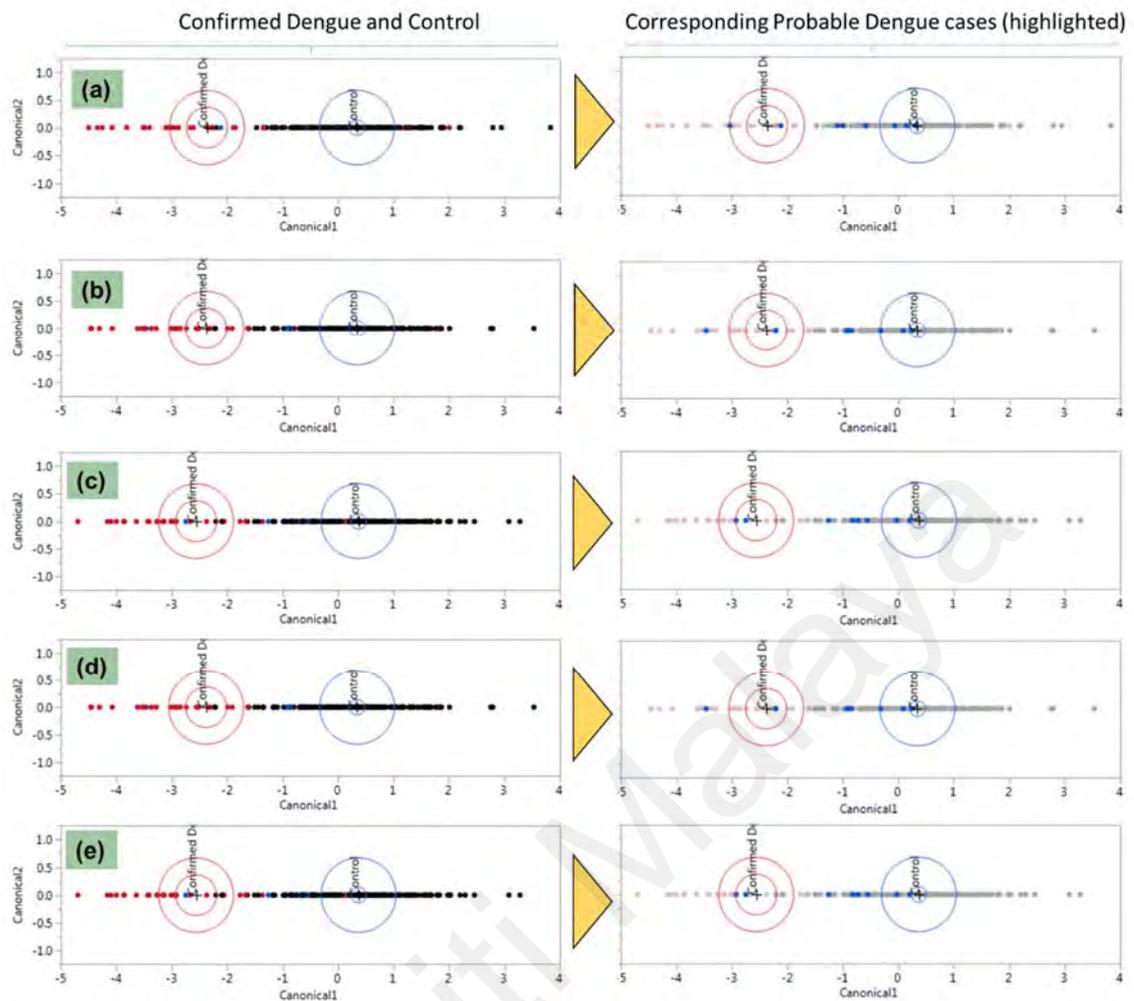
Figure 5.14: Dichotomous discriminant analysis on (a) raw data spectrum, (b) Normalized gender-centered spectrum, (c) Normalized gender-centered spectrum combined with data residue, (d) Normalized ethnicity-centered spectrum, and (e) Normalized ethnicity-centered spectrum combined with data residue. Red indicates confirmed dengue, black control, and purple for probable dengue. The right side highlights the corresponding probable dengue cases.

The summary of the analysis obtained in **Figure 5.14** is shown in Table 5.4[11].

---

[11] To facilitate the understanding of this figure, we take Figure 5.14(a) as an example. The figure on the left shows confirmed dengue cases on the left in red. The right side shows the scores of control patients in black. Probable cases are left unclassified in the software unlike the confirmed dengue and control counterparts, to see where it would land on based on the classification algorithms. Part of the probable dengue cases (highlighted in blue) can be seen on the left. However, the corresponding figure on the right side of the figure is exactly the same as the left, but only with probable cases made obvious to be seen, while both confirmed dengue and control cases are dimmed.

Table 5.5: Summary analysis for normalized data. Each row corresponds to the
designated items in **Figure 5.14**

Data Type	Count	Misclassified	% Misclassified	Breakdown		
Raw data spectrum	222	14	6.30631	**Actual Category**	**Predicted Count**	
					Confirmed Dengue	**Control**
				Confirmed Dengue	24	4
				Control	10	184
Normalized gender-centered spectrum	222	17	7.65766	**Actual Category**	**Predicted Count**	
					Confirmed Dengue	**Control**
				Confirmed Dengue	23	5
				Control	12	182
Normalized gender-centered spectrum combined with data residue	222	12	5.40541	**Actual Category**	**Predicted Count**	
					Confirmed Dengue	**Control**
				Confirmed Dengue	24	4
				Control	8	186
Normalized ethnicity-centered spectrum	222	17	7.65766	**Actual Category**	**Predicted Count**	
					Confirmed Dengue	**Control**
				Confirmed Dengue	23	5
				Control	12	182
Normalized ethnicity-centered spectrum combined with data residue	222	12	5.40541	**Actual Category**	**Predicted Count**	
					Confirmed Dengue	**Control**
				Confirmed Dengue	24	4
				Control	8	186

As shown in **Figure 5.14** which corresponds to each element in **Table 5.5**, after substituting probable cases as a test data in this section, a binary output is produced. **Figure 5.14(a)** shows the discriminant analysis on the unprocessed data, returning a 6.31% misclassification. **Figure 5.14(b)** shows the discriminant analysis on only the gender-centered spectrum, while **5.14(c)** shows the combination of the gender-centered spectrum with the residual data. The accuracy can be further augmented by inclusion of the residual parameters, as shown in **Figure 5.14(c)**, with a 5.41% error.

On ethnicity-centered normalization shown in **Figure 5.14(d)**, the accuracy is slightly lesser than the original analysis as shown in **Figure 5.14(a)**. However, when combined with residual parameters, the accuracy improved considerably at 4.51% error, similar to **Figure 5.14(c)**. This part thus shows that normalizing the data yields marginally better results.

5.4.2.2 Summary: Convolution Analysis with Sensitivity and Specificity

In this section, we finalize the analysis by convoluting the gender- and ethnicity-centered normalized data with different computations, with the same steps used in **Section 5.4.2.1**. By convoluting the ethnicity- and gender-centered data, the outliers which were amplified with each normalization is combined.

For the sake of brevity, the analysis in this section is directly attributed to the accuracy models from the prediction based on discriminant analysis in JMP. The canonical plots, in this case, are not shown as extensively discussed in the previous section for this purpose, due to the similar structure exhibited in the dichotomous biplots as shown in **Figure 5.14**.

Each functions and the formula of convolution describes how resulting spectra is used for classification. The combination of this analysis is summarized in **Table 5.6**, with all previous data modes included. All gender- and ethnicity-centered normalized data is denoted as X_{gender} and X_{eth} respectively.

Table 5.6: Summary of discriminant analysis of different data modes with sensitivity and specificity

Function ID	Formula for Normalized Data	% Misclassified	% Accuracy	TP	FP	TN	FN	Sensitivity (%)	Specificity (%)	Test: Probable Cases (n=8)		Data Types			
										Confirmed Dengue	Control	Raw	Feature	Vitals	Normalized
F1	n/a (Raw Data)	6.3	93.7	24	10	184	4	85.71	94.85	3	5	✓	✗	✗	✗
F2	n/a (Raw + Feature)	4.95	95.05	24	7	187	4	85.71	96.39	3	5	✓	✓	✗	✗
F3	n/a (Raw + Feature + Vitals)	5.4	94.6	24	8	186	4	85.71	95.88	3	5	✓	✓	✓	✗
F4	$X_{gender} = X_{raw} - X_{gender}$	6.3	93.7	21	7	187	4	84.00	96.39	2	6	✓	✓	✓	✓
F5	$X_{eth} = X_{raw} - X_{eth}$	5.41	94.59	22	6	188	4	84.62	96.91	1	7	✓	✓	✓	✓
F6	$(X_{eth} * X_{gender}) / (X_{eth} + X_{gender})$	1.35	98.65	25	0	194	3	89.29	100.00	1	7	✗	✗	✗	✓
F7	$(X_{eth} * X_{gender}) / (X_{eth} - X_{gender})$	5.85	94.15	23	8	186	7	76.67	95.88	2	6	✓	✓	✓	✓
F8	X_{eth} / X_{gender}	5.4	94.6	24	8	186	6	80.00	95.88	3	5	✓	✓	✓	✓
F9	$Sqrt(X_{eth}^2 + X_{gen}^2)$	5.4	94.6	24	8	186	5	82.76	95.88	3	5	✓	✓	✓	✓
F10	$X_{eth} * X_{gender}$	6.31	93.69	24	10	184	4	85.71	94.85	3	5	✓	✓	✓	✓

Confirmed Dengue: 28
Control: 194

Note:

TP: True Positive, TN: True Negative, FP: False Positive. Sensitivity = (TP/(TP + FN))%, Specificity = (TN/TN+FP)%

125

Table 5.5 summarizes different combinations of the discriminant analysis outputs, where each functions as denoted from **F1** to **F10** have their own subjective elements of merit. These elements which determine their merits are as follows (a) computation complexity, (b) sensitivity (c) specificity, (d) hypothetically capable to differentiating probable dengue into confirmed or control cases, and (e) possessing the least combinations of data types. These elements will be discussed as follows.

As shown in **F1-F3**, without the normalized data, the classification accuracies range between 93.7-95.05%, with best sensitivity and specificity values of 85.71% and 96.39% respectively. **F1**, while computationally simpler than **F2** and **F3**, had the least classification accuracy. **F2**, with raw and feature data, showed improved accuracy and performs even better than **F3**, which included three data landscapes, as shown in the "Data Types" column. The use of vitals (temperature, respiratory rates, etc) as covariates for the classification not only proved to be ineffective, it negates the accuracy marginally as proven in **F3**. This is due to the variability of the vitals as parameters in this case does not necessarily contribute to dengue diagnosis. 3 probable dengue patients were classified as confirmed in this region in each case.

F4 and **F5** show the separate use of the gender- and ethnicity-centered data, exhibiting better algorithmic performance in ethnicity-centered data. This is intuitively the case where PCA plots showed that ethnicity played a significant role in determining data structure as shown in **Figure 5.11**. However, **F5** does not show any output difference compared to only a raw and feature combo in **F2**. This is most likely due to the small variability produced in gender-based normalization, which has a binary outcome compared to a triple category ethnicity as a normalization factor. Also, probable cases were consistently reclassified as 1-2 confirmed dengue and 6-7 control.

F6-F10 shows the resulting output from convolutions of both gender- and ethnicity-centered data. All except **F6** (by only using normalized data) utilizes the residual data as shown in the "Data Types" column in **Table 5.5**. **F6** showed the highest accuracy of 98.65%, with both sensitivity and specificity of 89.29% and 100% respectively. Within these functions, only one or two probable dengue patients are reclassified as confirmed dengue.

In **F7-F10**, while demonstrating the convolution of both the ethnicity- and gender-centered data, showed less accuracy than the next best function **F4**, which is purely an ethnicity-normalized data plus residue. However, these convolutions have mixed results in reclassifying probable dengue. **F7-F10** reclassified 3 probable cases into the confirmed group respectively.

If we project that heuristically, 10-30% of probable dengue cases should be able to be reclassified as confirmed dengue, then it can be said that **F6** has the best output, with a reasonable sensitivity and specificity of 89.29% and 100%, while registering 1 out of 8 probable cases as confirmed dengue. We maintain this claim for further consideration, since there are virtually no publications to support nor to debunk this heuristically derived rule. If the least data combinations are construed as the main determinant in selecting the data processing technique, **F6** is also perhaps the best due to the use of purely normalized data, and excluding all other residues. We maintain the plausibility of this model for further consideration in future.

5.4.3   Chapter Conclusion

This chapter is concluded from the first analysis in **Section 5.1** where the skin reflectance from various patients of confirmed or probable dengue, and control groups. The analysis concludes that the patterns exhibited subtleties not observable to the human eye to discern. It was concluded that while there are collective differences in the average spectral patterns of each categories, it is too subtle to be simply rule-driven for classification.

Section 5.2 discusses several aspects of the data observed from afar and up-close to explain the occurrences of the patterns, and to reduce the parameters required for the following analyses which is aimed at removing confounding factors. In the case for discriminant analysis, this section provided a new data landscape for the classification formula that increases accuracy.

Section 5.3 allows a much simplified view of the data by using principal component analysis (PCA), further demonstrating the gender and ethnicity as a main factor in the data structure.

Finally, **Section 5.4** confirmed gender and ethnicity as the normalization factors using discriminant analysis (DA). These findings transformed the landscape of the data, which combined both the salient features of a gender- and ethnicity-centred data for a more accurate analysis. In a best-fit model, an accuracy of 98.65% was reported for separating confirmed dengue and control cases.

The factors that allow dengue patients to be detected by this method would most likely be a superposition of the spectral patterns due to tissue morphology, dengue pathophysiology and blood composition of the patient. To confirm each of these components as a factor in determining a diagnosis, however, numerous cross-validation

on the use of diffuse reflectance spectroscopy is required for histological scans of dengue patients. Also, in-depth studies on the tissue morphology with optical spectroscopy for this purpose is also warranted.

This chapter ends on the note of cautious optimism: separating confirmed dengue from non-dengue patients by means of discriminant analysis on diffuse reflectance spectroscopic patterns of a febrile human patient is possible to a high degree. This was shown from the data-driven analytics in this chapter. Finally, the interactions between all the analytical methods outlined in this chapter is summarized in **Figure 5.15**.
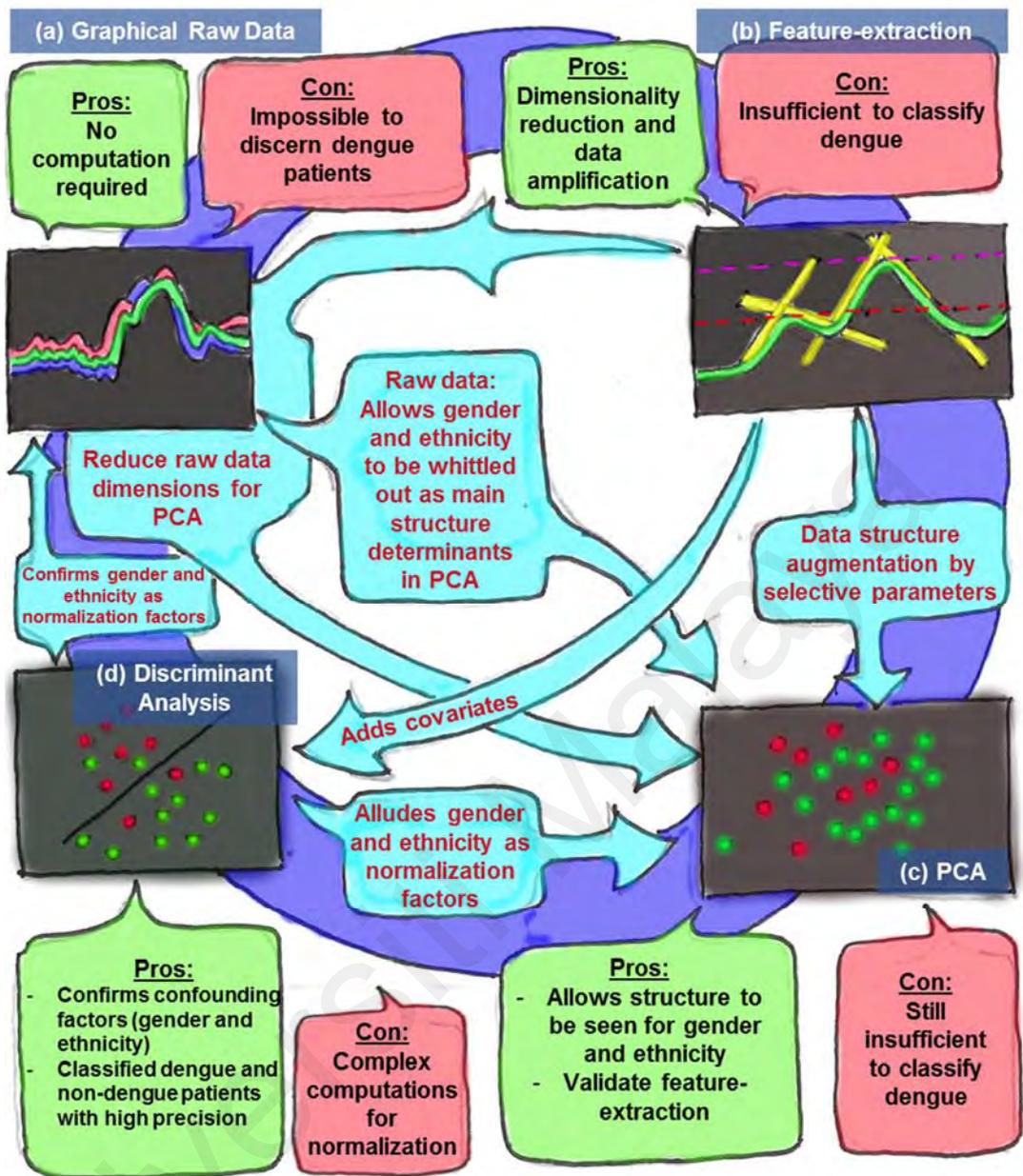
Figure 5.15: Chapter summary

CHAPTER 6: CONCLUSION – CLARITY UPON CLOSURE

This chapter begins with the following account:

"*An infant, not more than a year old, screamed in agony when blood was drawn. She had fever two days ago, but it has subsided. Mysteriously, she remains listless and weak, and still refused her bottled formula milk. Her cheeks and lips were slightly puffed and rashes speckled her tender stomach. The specialist of whom was entrusted the child ushered us to wait outside the triage room for she knew it was not a scene parents would bear. It took three rounds of piercing because blood from the small arms were not sufficient; the infant was dehydrated thus the unyielding veins. The parents should have been much further; the screams still pierced the pediatric area.*"

This is an unpublished but genuine account, rephrased to illustrate a father's witnessing of what needs to be done for a dengue suspect, a child notwithstanding, if blood is what is required. This, among others, triggered our pivotal research question: is it possible to non-invasively determine if a person is likely or unlikely to have dengue? The simplest way to obtain signatures from the human body was likely to be in the form of electromagnetic waves. The safest was perhaps the use of light, in mimicry of the sun. This requirement enabled us to narrow down on the use of a relatively simple and harmless technique called optical diffuse reflectance spectroscopy.

This led to unscrambling the intricate mesh on how dengue patients and photonics-based research cross paths, perhaps best articulated in **Chapter 2**, though for the most part, none of the correlations made were trivial and mostly were projections. The hypothesis led by the research question above was predicated on various conditions (or pathophysiology) of dengue patients linked with optical spectroscopy. This was not only confined to serological components of dengue patients, but also tissue morphology and

131

symptoms not only confined to dengue. All these were hypothesized to collectively contribute to the non-invasive signature profile of dengue patients.

This, in turn, led to fomentation of the objectives, as outlined in **Chapter 1**, which are (a) to build a database of diffuse skin reflectance spectroscopy of febrile patients, (b) confirm the diagnosis for each patient to determine the inclusivity of spectroscopic data, and finally (c) to statistically prove that from an in-depth analysis of the data, dengue and non-dengue illnesses can be classified to a high degree.

Chapter 3 describes our methods briefly on each major undertakings which needed to tessellate with each other. This includes the spectroscopic setup, which is a standard diffuse reflectance spectroscopy using the integrating sphere technique as the patient skin probe. Recruitment of patients (after ethical clearance) were devised to maximize both control and dengue patients for analytics in practical conditions. A custom-made software was devised to navigate and organize the data obtained. Finally, diagnostics of each recruited patient were later concluded by reviewing the clinical notes by UMMC physicians.

The results of our task were outlined in simple tones in **Chapter 4** to provide a first impression of the data. The spectroscopic data was divided into major demographic categories, namely gender, ethnicity and age groups. Diagnostics of each patient were exhausted piecewise in **Table 4.1**. 257 patients were recruited, which was later narrowed down to 240 after spectroscopic data error considerations, fulfilling the first two objectives.

We discussed each categories with depth on the factors which were predominant in the spectrum of the patients in **Chapter 5**. Four modes of discussing the data was established. The first is by briefly analyzing the graphical reflectance spectra by class categories. On

average, it seemed that gender and ethnicity played a role in the spectrum pattern. The second was by the use of feature extraction, in part as an attempt to provide an explanation of the patterns in the spectra, though mainly to simplify the use of principal component analysis (PCA) by selecting desired variables. Third, PCA was then used to identify the structure of the data, and to establish factors of redundancy. This was later proven with growing evidence that gender and ethnicity were indeed confounding factors.

The fourth and strongest argument of this thesis was presented by the use of discriminant analysis (DA) on the said factors. Also, with the use of the discriminant analysis, mathematically consistent normalization procedures were made possible by reclassifying ethnicities of patients other than the major Malaysian groups, namely Malays, Chinese and Indians. The spectrum of the subjects were flattened by mean-centering (or normalizing) the spectrum against gender and ethnicity, which later was proven to increase the classification accuracy. The most plausible model argued in this chapter is a convolution of both gender- and ethnicity-centered spectrum, **F6**, with classification accuracy of 98.65% while having a reasonable sensitivity and specificity of 89.29% and 100% respectively.

Rather than verifying which part of the condition in a dengue patient contributes to its peculiarity by a non-invasive reflectance spectroscopy, we'd rather treat the patient as the black box in this study. With the data-driven outcomes presented as per objectives, we cautiously conclude our third and final objective: that there is an optimistic outlook on the use of diffuse reflectance spectroscopy to screen dengue patients. In all, the findings of this work are hoped to be a landmark study for further trials and celebrate the self-righting mechanism of science.

On optimism, this chapter ends with an epilogue of the aforementioned account.

"Her mother slept beside her in the ward on a thin mattress. She was told to write down how much milk was taken by the child, and dutifully she did so. After the third day, as the child suckled the Tommy-Tippee ravenously, the doctor signed the discharge papers."

REFERENCES

Abrar, S., & Ansari, M. J. (2016). Acute fulminant myocarditis in a case of dengue fever: A case report. *Asian Pacific Journal of Tropical Disease, 6*(4), 328-329.

Aguiar, M., Stollenwerk, N., & Halstead, S. B. (2016). The risks behind Dengvaxia recommendation. *The Lancet infectious diseases, 16*(8), 882-883.

Ahmad, R., Suzilah, I., Najdah, W. M. A. W., Topek, O., Mustafakamal, I., & Lee, H. L. (2018). Factors determining dengue outbreak in Malaysia. *PloS one, 13*(2), e0193326.

Alaluf, S., Atkins, D., Barrett, K., Blount, M., Carter, N., & Heath, A. (2002). Ethnic variation in melanin content and composition in photoexposed and photoprotected human skin. *Pigment Cell Research, 15*(2), 112-118.

Ambali, A. R., Bakar, A. N., & Merican, F. M. (2013). Environmental policy in Malaysia: biomedical waste, strategies and issues. *Journal of Administrative Science, 10*(1), 1-17.

Anderson, R. R., & Parrish, J. A. (1981). The optics of human skin. *Journal of investigative dermatology, 77*(1), 13-19.

Angelopoulou, E. (1999). The reflectance spectrum of human skin. *Technical Reports (CIS)*, 584.

Arimoto, H., Egawa, M., & Yamada, Y. (2005). Depth profile of diffuse reflectance near-infrared spectroscopy for measurement of water content in skin. *Skin Research and Technology, 11*(1), 27-35.

Ash, C., Dubec, M., Donne, K., & Bashford, T. (2017). Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers in medical science, 32*(8), 1909-1918.

Atias, D., Liebes, Y., Chalifa-Caspi, V., Bremand, L., Lobel, L., Marks, R. S., & Dussart, P. (2009). Chemiluminescent optical fiber immunosensor for the detection of IgM antibody to dengue virus in humans. *Sensors and Actuators B: Chemical, 140*(1), 206-215.

Attas, M., Posthumus, T., Schattka, B., Sowa, M., Mantsch, H., & Zhang, S. (2002). Long-wavelength near-infrared spectroscopic imaging for in-vivo skin hydration measurements. *Vibrational spectroscopy, 28*(1), 37-43.

Azami, N. A. M., Salleh, S. A., Neoh, H.-m., Zakaria, S. Z. S., & Jamal, R. (2011). Dengue epidemic in Malaysia: Not a predominantly urban disease anymore. *BMC research notes, 4*(1), 216.

Azfar, N. A., Malik, L. M., Jamil, A., Jahangir, M., Tirmizi, N., Majid, A., . . . Malik, M. (2012). Cutaneous manifestations in patients of dengue fever. *Journal of Pakistan Association of Dermatologists, 22*(4), 320-324.

Barnett, R. (2017). Dengue. *The Lancet, 390*(10106), 1941.

Bashkatov, A., Genina, E., Kochubey, V., & Tuchin, V. (2005). Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *Journal of Physics D: Applied Physics, 38*(15), 2543.

Bhopal, R. (2004). Glossary of terms relating to ethnicity and race: for reflection and debate. *Journal of Epidemiology & Community Health, 58*(6), 441-445.

Bin Yap, F. B. (2010). Clinical characteristics of basal cell carcinoma in a tertiary hospital in Sarawak, Malaysia. *International journal of dermatology, 49*(2), 176-179.

Bodén, I., Nyström, J., Lundskog, B., Zazo, V., Geladi, P., Lindholm-Sethson, B., & Naredi, P. (2013). Non-invasive identification of melanoma with near-infrared and skin impedance spectroscopy. *Skin Research and Technology, 19*(1), e473-e478.

Boushel, R., & Piantadosi, C. (2000). Near-infrared spectroscopy for monitoring muscle oxygenation. *Acta Physiologica, 168*(4), 615-622.

Broer, N. M., Liesenhoff, T., & Horch, H.-H. (2004). Laserinduced fluorescence spectroscopy for real-time tissue differentiation. *Medical Laser Application, 19*(1), 45-53.

Calabro, K., Curtis, A., Galarneau, J.-R., Krucker, T., & Bigio, I. J. (2011). Gender variations in the optical properties of skin in murine animal models. *Journal of biomedical optics, 16*(1), 011008.

Camara, A. R., Gouvêa, P. M., Dias, A. C., Braga, A., Dutra, R. F., de Araujo, R. E., & Carvalho, I. (2013). Dengue immunoassay with an LSPR fiber optic sensor. *Optics express, 21*(22), 27023-27031.

Carter, M. J., Emary, K. R., Moore, C. E., Parry, C. M., Sona, S., Putchhat, H., . . . Dobson, A. D. (2015). Rapid Diagnostic Tests for Dengue Virus Infection in Febrile Cambodian Children: Diagnostic Accuracy and Incorporation into Diagnostic Algorithms. *PLoS neglected tropical diseases, 9*(2), e0003424-e0003424.

Chen, C., Chang, Y., Ricanek, K., & Wang, Y. (2010). *Face age estimation using model selection.* Paper presented at the Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on.

Chen, J., Lee, A., Zhao, J., Wang, H., Lui, H., McLean, D. I., & Zeng, H. (2009). Spectroscopic characterization and microscopic imaging of extracted and in situ cutaneous collagen and elastic tissue components under two-photon excitation. *Skin Research and Technology, 15*(4), 418-426.

Clapham, H., Cummings, D. A., Nisalak, A., Kalayanarooj, S., Thaisomboonsuk, B., Klungthong, C., . . . Lessler, J. (2015). Epidemiology of Infant Dengue Cases Illuminates Serotype-Specificity in the Interaction between Immunity and Disease, and Changes in Transmission Dynamics. *PLOS Negl Trop Dis, 9*(12), e0004262.

Cohen, M. L. (1977). Measurement of the thermal properties of human skin. A review. *Journal of investigative dermatology, 69*(3), 333-338.

Cooksey, C. C., Tsai, B. K., & Allen, D. W. (2014). *A collection and statistical analysis of skin reflectance signatures for inherent variability over the 250 nm to 2500 nm spectral range.* Paper presented at the SPIE Defense+ Security.

Cooksey, C. C., Tsai, B. K., & Allen, D. W. (2015). *Spectral reflectance variability of skin and attributing factors.* Paper presented at the Radar Sensor Technology XIX; and Active and Passive Signatures VI.

De Rigal, J., Losch, M., Bazin, R., Camus, C., Sturelle, C., Descamps, V., & Leveque, J. (1993). Near infrared spectroscopy: a new approach to the characterization of dry skin. *Journal-society of cosmetic chemists, 44*, 197-197.

Diamond, M. S. (2003). Evasion of innate and adaptive immunity by flaviviruses. *Immunology and cell biology, 81*(3), 196-206.

Dick, O. B., San Martín, J. L., Montoya, R. H., del Diego, J., Zambrano, B., & Dayan, G. H. (2012). The history of dengue outbreaks in the Americas. *The American journal of tropical medicine and hygiene, 87*(4), 584-593.

Fang, Z.-h., Fu, X.-p., & He, X.-m. (2016). Investigation of absorption and scattering characteristics of kiwifruit tissue using a single integrating sphere system. *Journal of Zhejiang University-SCIENCE B, 17*(6), 484-492.

Firdous, S., Ahmed, M., Rehman, A., Nawaz, M., Anwar, S., & Murtaza, S. (2012). Transmission spectroscopy of dengue viral infection. *Laser Physics Letters, 9*(4), 317.

Firdous, S., Ahmed, M., Rehman, A., Nawaz, M., Anwar, S., & Murtaza, S. (2012). Transmission spectroscopy of dengue viral infection. *Laser Physics Letters, 9*(4), 317-321.

Firdous, S., & Anwar, S. (2015). Dengue viral infection monitoring from diagnostic to recovery using Raman spectroscopy. *Laser Physics Letters, 12*(8), 085601.

Flasche, S., Jit, M., Rodríguez-Barraquer, I., Coudeville, L., Recker, M., Koelle, K., . . . Cummings, D. A. (2016). The long-term safety, public health impact, and cost-effectiveness of routine vaccination with a recombinant, live-attenuated dengue vaccine (Dengvaxia): a model comparison study. *PLoS medicine, 13*(11), e1002181.

Flodgren, G. M., Crenshaw, A. G., Hellström, F., & Fahlström, M. (2010). Combining microdialysis and near-infrared spectroscopy for studying effects of low-load repetitive work on the intramuscular chemistry in trapezius myalgia. *BioMed Research International, 2010*.

Forshey, B. M., Stoddard, S. T., & Morrison, A. C. (2016). Dengue Viruses and Lifelong Immunity: Reevaluating the Conventional Wisdom. *Journal of Infectious Diseases*, jiw102.

Fuchs, E. (2001). Separating the fluorescence and reflectance components of coral spectra. *Applied optics, 40*(21), 3614-3621.

Ganguli, A., Ornob, A., Yu, H., Damhorst, G., Chen, W., Sun, F., . . . Bashir, R. (2017). Hands-free smartphone-based diagnostics for simultaneous detection of Zika, Chikungunya, and Dengue at point-of-care. *Biomedical microdevices, 19*(4), 73.

Ghazaryan, A. A., Hu, P.-S., Chen, S.-J., Tan, H.-Y., & Dong, C.-Y. (2012). Spatial and temporal analysis of skin glycation by the use of multiphoton microscopy and spectroscopy. *Journal of dermatological science, 65*(3), 189-195.

Gillies, R., Zonios, G., Anderson, R. R., & Kollias, N. (2000). Fluorescence excitation spectroscopy provides information about human skin in vivo. *Journal of investigative dermatology, 115*(4), 704-707.

Givens, G., Beveridge, J. R., Draper, B. A., & Bolme, D. (2003). *A statistical assessment of subject factors in the pca recognition of human faces.* Paper presented at the Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on.

Glennie, D. L., Hayward, J. E., & Farrell, T. J. (2015). Modeling changes in the hemoglobin concentration of skin with total diffuse reflectance spectroscopy. *Journal of biomedical optics, 20*(3), 035002.

Guzmán, M. G., & Kourí, G. (2004). Dengue diagnosis, advances and challenges. *International journal of infectious diseases, 8*(2), 69-80.

Halder, R. M., & Nootheti, P. K. (2003). Ethnic skin disorders overview. *Journal of the American Academy of Dermatology, 48*(6), S143-S148.

Hennessy, R., Markey, M. K., & Tunnell, J. W. (2015). Impact of one-layer assumption on diffuse reflectance spectroscopy of skin. *Journal of biomedical optics, 20*(2), 027001.

Hirasawa, A., Kaneko, T., Tanaka, N., Funane, T., Kiguchi, M., Sørensen, H., . . . Ogoh, S. (2016). Near-infrared spectroscopy determined cerebral oxygenation with eliminated skin blood flow in young males. *Journal of clinical monitoring and computing, 30*(2), 243-250.

Huggins, R. H., Henderson, M. D., Mulekar, S. V., Ozog, D. M., Kerr, H. A., Jabobsen, G., . . . Hamzavi, I. H. (2012). Melanocyte-keratinocyte transplantation procedure in the treatment of vitiligo: the experience of an academic medical center in the United States. *Journal of the American Academy of Dermatology, 66*(5), 785-793.

Hunsperger, E. A., dos Santos, C. N. D., Vu, H. T. Q., Yoksan, S., & Deubel, V. (2017). Rapid and accurate interpretation of dengue diagnostics in the context of dengue vaccination implementation: Viewpoints and guidelines issued from an experts group consultation. *PLoS neglected tropical diseases, 11*(9), e0005719.

Jacques, S. L. (1996). Origins of tissue optical properties in the UVA, visible, and NIR regions. *OSA TOPS on advances in optical imaging and photon migration, 2*, 364-369.

Jaenisch, T., Tam, D. T. H., Kieu, N. T. T., Ngoc, T., Nam, N. T., Van Kinh, N., . . . See, L. L. C. (2016). Clinical evaluation of dengue and identification of risk factors for severe disease: protocol for a multicentre study in 8 countries. *BMC infectious diseases, 16*(1), 120.

Jahanshahi, P., Wei, Q., Jie, Z., Ghomeishi, M., Sekaran, S. D., & Mahamd Adikan, F. R. (2017). Kinetic analysis of IgM monoclonal antibodies for determination of dengue sample concentration using SPR technique. *Bioengineered, 8*(3), 239-247.

Jahanshahi, P., Zalnezhad, E., Sekaran, S. D., & Adikan, F. R. M. (2014). Rapid immunoglobulin M-based dengue diagnostic test using surface plasmon resonance biosensor. *Scientific reports, 4*.

Jayasinghe, N. S., Thalagala, E., Wattegama, M., & Thirumavalavan, K. (2016). Dengue fever with diffuse cerebral hemorrhages, subdural hematoma and cranial diabetes insipidus. *BMC research notes, 9*(1), 1.

Jha, S., & Ansari, M. (2010). Dengue infection causing acute hypokalemic quadriparesis. *Neurology India, 58*(4), 592.

Ji, X., Miao, C., Zhang, Y., Lin, X., & Dai, Q. (2017). Separating reflective and fluorescent components for dynamic scenes. *Optics Communications, 404*, 11-17.

Kamil, Y. M., Bakar, M. H. A., Yaacob, M. H., Syahir, A., Lim, H. N., & Mahdi, M. A. (2019). Dengue E Protein Detection Using a Graphene Oxide Integrated Tapered Optical Fiber Sensor. *IEEE Journal of Selected Topics in Quantum Electronics, 25*(1), 1-8.

Katajamaa, M., & Orešič, M. (2007). Data processing for mass spectrometry-based metabolomics. *Journal of chromatography A, 1158*(1-2), 318-328.

Kaxiras, E., Tsolakidis, A., Zonios, G., & Meng, S. (2006). Structural model of eumelanin. *Physical review letters, 97*(21), 218102.

Khan, S., Ullah, R., Saleem, M., Bilal, M., Rashid, R., Khan, I., . . . Nawaz, M. (2016). Raman spectroscopic analysis of dengue virus infection in human blood sera. *Optik-International Journal for Light and Electron Optics, 127*(4), 2086-2088.

Koh, D., Wang, H., Lee, J., Chia, K., Lee, H., & Goh, C. (2003). Basal cell carcinoma, squamous cell carcinoma and melanoma of the skin: analysis of the Singapore Cancer Registry data 1968–97. *British Journal of Dermatology, 148*(6), 1161-1166.

Korhonen, E. M., Huhtamo, E., Virtala, A.-M. K., Kantele, A., & Vapalahti, O. (2014). Approach to non-invasive sampling in dengue diagnostics: Exploring virus and NS1 antigen detection in saliva and urine of travelers with dengue. *Journal of Clinical Virology, 61*(3), 353-358.

Kubelka, P. (1931). The kubelka-munk theory of reflectance. *Z Tech Phys, 12*, 539.

Lai, Y.-L., Chung, Y.-K., Tan, H.-C., Yap, H.-F., Yap, G., Ooi, E.-E., & Ng, L.-C. (2007). Cost-effective real-time reverse transcriptase PCR (RT-PCR) to screen for

Dengue virus followed by rapid single-tube multiplex RT-PCR for serotyping of the virus. *Journal of clinical microbiology, 45*(3), 935-941.

Lanciotti, R. S., Lewis, J. G., Gubler, D. J., & Trent, D. W. (1994). Molecular evolution and epidemiology of dengue-3 viruses. *Journal of General Virology, 75*(1), 65-75.

Lau, V. L. (2004). *Case study on the management of waste materials in Malaysia.* Paper presented at the Forum Geookol.

Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., Villalobos, I., Ramos, C., & Rico-Hesse, R. (1999). Dengue virus structural differences that correlate with pathogenesis. *Journal of virology, 73*(6), 4738-4747.

Lerner, A. B., Shizume, K., & Bunding, I. (1954). The mechanism of endocrine control of melanin pigmentation. *The Journal of Clinical Endocrinology & Metabolism, 14*(12), 1463-1490.

Li, C., Brost, V., Benezeth, Y., Marzani, F., & Yang, F. (2018). Design and evaluation of a parallel and optimized light–tissue interaction-based method for fast skin lesion assessment. *Journal of Real-Time Image Processing, 15*(2), 407-420.

Li, Z., Wang, Z., Yin, Z., Zhang, Y., Xue, X., Han, J., . . . Wang, H. (2017). Gender differences in fibrosis remodeling in patients with long-standing persistent atrial fibrillation. *Oncotarget, 8*(32), 53714.

Lim, J. K., Alexander, N., & Di Tanna, G. L. (2017). A systematic review of the economic impact of rapid diagnostic tests for dengue. *BMC health services research, 17*(1), 850.

Lister, T., Wright, P. A., & Chappell, P. H. (2012). Optical properties of human skin. *Journal of Biomedical Optics, 17*(9), 0909011-09090115.

Lock-Andersen, J., Knudstorp, N. D., & Wulf, H. C. (1998). Facultative skin pigmentation in caucasians: an objective biological indicator of lifetime exposure to ultraviolet radiation? *The British journal of dermatology, 138*(5), 826-832.

Lopez-Jimena, B., Bekaert, M., Bakheit, M., Frischmann, S., Patel, P., Simon-Loriere, E., . . . Harold, G. (2018). Development and validation of four one-step real-time RT-LAMP assays for specific detection of each dengue virus serotype. *PLoS neglected tropical diseases, 12*(5), e0006381.

Malaysia, D. o. S. (2016). Current population estimates, Malaysia, 2014–2016.

Manoharan, K. S., Ramasamy, K., & Heinz, P. (2017). Rash with fever in children: a clinical approach. *Paediatrics and Child Health, 27*(4), 196-201.

Martin, K. A. (1993). Direct measurement of moisture in skin by NIR spectroscopy. *Journal-society of cosmetic chemists, 44*, 249-249.

Masuda, Y., Ogura, Y., Inagaki, Y., Yasui, T., & Aizu, Y. (2018). Analysis of the influence of collagen fibres in the dermis on skin optical reflectance by Monte

Carlo simulation in a nine-layered skin model. *Skin Research and Technology, 24*(2), 248-255.

Matas, A., Sowa, M. G., Taylor, G., & Mantsch, H. H. (2002). Melanin as a confounding factor in near infrared spectroscopy of skin. *Vibrational spectroscopy, 28*(1), 45-52.

Matts, P., Dykes, P., & Marks, R. (2007). The distribution of melanin in skin determined in vivo. *British Journal of Dermatology, 156*(4), 620-628.

Meinke, M. C., Müller, G. J., Helfmann, J., & Friebel, M. (2007). Optical properties of platelets and blood plasma and their influence on the optical behavior of whole blood in the visible to near infrared wavelength range. *Journal of biomedical optics, 12*(1), 014024.

Mendenhall, M. J., Nunez, A. S., & Martin, R. K. (2015). Human skin detection in the visible and near infrared. *Applied optics, 54*(35), 10559-10570.

Messer, W. B., Gubler, D. J., Harris, E., Sivananthan, K., & De Silva, A. M. (2003). Emergence and global spread of a dengue serotype 3, subtype III virus. *Emergence*.

Mohd-Zaki, A. H., Brett, J., Ismail, E., & L'Azou, M. (2014). Epidemiology of dengue disease in Malaysia (2000–2012): a systematic literature review. *PLoS neglected tropical diseases, 8*(11), e3159.

Muller, D. A., Depelsenaire, A. C., & Young, P. R. (2017). Clinical and laboratory diagnosis of dengue virus infection. *The Journal of infectious diseases, 215*(suppl_2), S89-S95.

Mustafa, M., Rasotgi, V., Jain, S., & Gupta, V. (2015). Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control. *Medical Journal Armed Forces India, 71*(1), 67-70.

Mustapa, M., Bakar, M. A., Kamil, Y. M., Hamzah, A. S. A., & Mahdi, M. (2018). Bio-functionalized Tapered Multimode Fiber Coated with Dengue Virus NS1 Glycoprotein for Label Free Detection of Anti-Dengue Virus NS1 IgG Antibody. *IEEE Sensors Journal*.

Na, R., Stender, I.-M., Henriksen, M., & Wulf, H. C. (2001). Autofluorescence of human skin is age-related after correction for skin pigmentation and redness. *Journal of investigative dermatology, 116*(4), 536-540.

Nickell, S., Hermann, M., Essenpreis, M., Farrell, T. J., Krämer, U., & Patterson, M. S. (2000). Anisotropy of light propagation in human skin. *Physics in Medicine & Biology, 45*(10), 2873.

Olesen, N. D., Sørensen, H., Ambrus, R., Svendsen, L. B., Lund, A., & Secher, N. H. (2018). A mesenteric traction syndrome affects near-infrared spectroscopy evaluated cerebral oxygenation because skin blood flow increases. *Journal of clinical monitoring and computing, 32*(2), 261-268.

Ong, M. W., Tan, C. H., & Cheng, A. K. S. (2018). Prevalence and Determinants of Vitamin D Deficiency Among the Overweight and Obese Singaporeans Seeking Weight Management Including Bariatric Surgery: a Relationship with Bone Health. *Obesity surgery, 28*(8), 2305-2312.

Organization, W. H., Research, S. P. f., Diseases, T. i. T., Diseases, W. H. O. D. o. C. o. N. T., Epidemic, W. H. O., & Alert, P. (2009). *Dengue: guidelines for diagnosis, treatment, prevention and control*: World Health Organization.

Osorio, L., Uribe, M., Ardila, G. I., Orejuela, Y., Velasco, M., Bonelo, A., & Parra, B. (2015). The use of rapid dengue diagnostic tests in a routine clinical setting in a dengue-endemic area of Colombia. *Memórias do Instituto Oswaldo Cruz*(AHEAD), 00-00.

Peeling, R. W., Artsob, H., Pelegrino, J. L., Buchy, P., Cardosa, M. J., Devi, S., . . . Guzman, M. G. (2010). Evaluation of diagnostic tests: dengue. *Nature Reviews Microbiology, 8*, S30-S37.

Pena, A.-M., Strupler, M., Boulesteix, T., & Schanne-Klein, M.-C. (2005). Spectroscopic analysis of keratin endogenous signal for skin multiphoton microscopy. *Optics express, 13*(16), 6268-6274.

Prabhu, V., Rao, S. B., Fernandes, E. M., Rao, A. C., Prasad, K., & Mahato, K. K. (2014). Objective assessment of endogenous collagen in vivo during tissue repair by laser induced fluorescence. *PloS one, 9*(5), e98609.

Rigau-Pérez, J. G., Clark, G. G., Gubler, D. J., Reiter, P., Sanders, E. J., & Vorndam, A. V. (1998). Dengue and dengue haemorrhagic fever. *The Lancet, 352*(9132), 971-977.

Schwartz, L. M., Halloran, M. E., Durbin, A. P., & Longini, I. M. (2015). The dengue vaccine pipeline: Implications for the future of dengue control. *Vaccine, 33*(29), 3293-3298.

Screaton, G., Mongkolsapaya, J., Yacoub, S., & Roberts, C. (2015). New insights into the immunopathology and control of dengue virus infection. *Nature Reviews Immunology, 15*(12), 745-759.

Seo, I., Tseng, S., Cula, G., Bargo, P., & Kollias, N. (2009). *Fluorescence spectroscopy for endogenous porphyrins in human facial skin.* Paper presented at the Proc. SPIE.

Sivan, A., Shriram, A., Sugunan, A., Anwesh, M., Muruganandam, N., Kartik, C., & Vijayachari, P. (2016). Natural transmission of dengue virus serotype 3 by Aedes albopictus (Skuse) during an outbreak in Havelock Island: Entomological characteristics. *Acta tropica*.

Soller, B., Srikiatkachorn, A., Zou, F., Rothman, A. L., Yoon, I.-K., Gibbons, R. V., . . . Green, S. (2014). Preliminary evaluation of near infrared spectroscopy as a method to detect plasma leakage in children with dengue hemorrhagic fever. *BMC infectious diseases, 14*(1), 396.

Sprigle, S., Linden, M., & Riordan, B. (2002). Characterizing reactive hyperemia via tissue reflectance spectroscopy in response to an ischemic load across gender, age, skin pigmentation and diabetes. *Medical engineering & physics, 24*(10), 651-661.

Srikiatkhachorn, A., Krautrachue, A., Ratanaprakarn, W., Wongtapradit, L., Nithipanya, N., Kalayanarooj, S., . . . Mammen Jr, M. P. (2007). Natural history of plasma leakage in dengue hemorrhagic fever: a serial ultrasonographic study. *The Pediatric infectious disease journal, 26*(4), 283-290.

Srikiatkhachorn, A., Rothman, A. L., Gibbons, R. V., Sittisombut, N., Malasit, P., Ennis, F. A., . . . Kalayanarooj, S. (2011). Dengue—how best to classify it. *Clinical Infectious Diseases, 53*(6), 563-567.

Stanaway, J. D., Shepard, D. S., Undurraga, E. A., Halasa, Y. A., Coffeng, L. E., Brady, O. J., . . . Castañeda-Orjuela, C. A. (2016). The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *The Lancet Infectious Diseases*.

Steinhagen, K., Probst, C., Radzimski, C., Schmidt-Chanasit, J., Emmerich, P., van Esbroeck, M., . . . Warnecke, J. M. (2016). Serodiagnosis of Zika virus (ZIKV) infections by a novel NS1-based ELISA devoid of cross-reactivity with dengue virus antibodies: a multicohort study of assay performance, 2015 to 2016. *Eurosurveillance, 21*(50).

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., . . . Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj, 338*, b2393.

Takahashi, T., Takikawa, Y., Kawagoe, R., Shibuya, S., Iwano, T., & Kitazawa, S. (2011). Influence of skin blood flow on near-infrared spectroscopy signals measured on the forehead during a verbal fluency task. *Neuroimage, 57*(3), 991-1002.

Teoh, B.-T., Sam, S.-S., Tan, K.-K., Danlami, M. B., Shu, M.-H., Johari, J., . . . Nentwich, O. (2015). Early Detection of Dengue Virus by Use of Reverse Transcription-Recombinase Polymerase Amplification. *Journal of clinical microbiology, 53*(3), 830-837.

Tjandra, E. T., Ehrchen, J., Broekaert, S., & Sunderkötter, C. (2017). Dengue fever and the differential diagnoses of rash, fever, and headache following travel to the tropics. *JDDG: Journal der Deutschen Dermatologischen Gesellschaft, 15*(1), 82-85.

Tsai, J.-J., Liu, L.-T., Lin, P.-C., Tsai, C.-Y., Chou, P.-H., Tsai, Y.-L., . . . Lee, P.-Y. A. (2018). Validation of POCKIT™ Dengue Virus Reagent Set for Rapid Detection of Dengue Virus in Human Serum on a Field-Deployable PCR System. *Journal of clinical microbiology*, JCM. 01865-01817.

Valdés, I., Gil, L., Lazo, L., Marcos, E., Martín, J., Suzarte, E., . . . Hermida, L. (2016). Could an experimental dengue virus infection fail to induce solid immunity against homologous viral challenge in non-human primates? *Archives of virology, 161*(2), 465-470.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics, 7*(1), 142.

Van Gemert, M., Jacques, S. L., Sterenborg, H., & Star, W. (1989). Skin optics. *IEEE Transactions on Biomedical Engineering, 36*(12), 1146-1154.

Vrabie, V., Gobinet, C., Piot, O., Tfayli, A., Bernard, P., Huez, R., & Manfait, M. (2007). Independent component analysis of Raman spectra: application on paraffin-embedded skin biopsies. *Biomedical Signal Processing and Control, 2*(1), 40-50.

Waggoner, J. J., Balmaseda, A., Gresh, L., Sahoo, M. K., Montoya, M., Wang, C., . . . Harris, E. (2016). Homotypic Dengue Virus Re-infections in Nicaraguan Children. *Journal of Infectious Diseases*, jiw099.

Walling, P. L., & Dabney, J. M. (1989). Moisture in skin by near-infrared reflectance spectroscopy. *J. Soc. Cosmet. Chem, 40*(3), 151-171.

Wollina, U., Liebold, K., Schmidt, W. D., Hartmann, M., & Fassler, D. (2002). Biosurgery supports granulation and debridement in chronic wounds–clinical data and remittance spectroscopy measurement. *International journal of dermatology, 41*(10), 635-639.

Wong, W. R., Berini, P., Fan, H., & Adikan, F. R. M. (2018). Multichannel Long-range Surface Plasmon Waveguides for Parallel Biosensing. *Journal of Lightwave Technology*.

Wong, W. R., Krupin, O., Sekaran, S. D., Mahamd Adikan, F. R., & Berini, P. (2014). Serological diagnosis of dengue infection in blood plasma using long-range surface plasmon waveguides. *Analytical chemistry, 86*(3), 1735-1743.

Xiao, K., Wang, M., Luo, R., Li, C., & Wuerger, S. (2016). Characterisation of skin spectra in a Caucasian and Oriental sample. *Electronic Imaging, 2016*(20), 1-4.

Yacoub, S., Mongkolsapaya, J., & Screaton, G. (2016). Recent advances in understanding dengue. *F1000Research, 5*.

Yap, G., Sil, B. K., & Ng, L.-C. (2011). Use of saliva for early dengue diagnosis. *PLOS Negl Trop Dis, 5*(5), e1046.

Yoshida, K., & Nishidate, I. (2014). Rapid calculation of diffuse reflectance from a multilayered model by combination of the white Monte Carlo and adding-doubling methods. *Biomedical optics express, 5*(11), 3901-3920.

Zambelis, T., Papadakis, G., Kokotis, P., Villiotou, V., Dogkas, N., & Karandreas, N. (2017). Lack of Definite Association of Vitamin D Deficiency with Diabetic Neuropathy. Investigation in Greek and in Bangladeshi Patients. *in vivo, 31*(2), 259-261.

Zhang, S., Soller, B. R., Kaur, S., Perras, K., & Vander Salm, T. J. (2000). Investigation of noninvasive in vivo blood hematocrit measurement using NIR reflectance

spectroscopy and partial least-squares regression. *Applied spectroscopy, 54*(2), 294-299.

Zhang, Y., Bai, J., & Ying, J. Y. (2015). A stacking flow immunoassay for the detection of dengue-specific immunoglobulins in salivary fluid. *Lab on a Chip, 15*(6), 1465-1471.

Zonios, G., Bykowski, J., & Kollias, N. (2001). Skin melanin, hemoglobin, and light scattering properties can be quantitatively assessed in vivo using diffuse reflectance spectroscopy. *Journal of investigative dermatology, 117*(6), 1452-1457.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

1. **Published:** Poh, A. H., Moghavvemi, M., Leong, C. S., Lau, Y. L., Ghandari, A. S., Apau, A., & Adikan, F. R. M. (2017). Collective behavior quantification on human odor effects against female Aedes aegypti mosquitoes—Open source development. *PLoS ONE*, 12(2), e0171555.

2. **Published:** Poh, A. H., Moghavvemi, M., Shafiei, M. M., Leong, C. S., Lau, Y. L., Adikan, F. R. M., ... & Hassan, M. A. A. (2017). Effects of low-powered RF sweep between 0.01-20 GHz on female Aedes Aegypti mosquitoes: A collective behaviour analysis. *PLoS ONE*, 12(6), e0178766.

3. **Published:** Poh, A. H., Jamaludin, M. F., Fadzallah, I. A., Nik Ibrahim, N. M. J., Yusof, F., Adikan, F. R. M., & Moghavvemi, M. (2019). Diffuse reflectance spectroscopic analysis of barium sulfate as a reflection standard within 173–2500 nm: From pure to sintered form. *Journal of Near Infrared Spectroscopy*, 0967033519868241.

4. Poh, A. H., Adikan, F. R. M, Moghavvemi, M, (2019). The Past, Present and Prospective on UV-VIS-NIR Skin Photonics and Spectroscopy - A Wavelength Guide. *Medical, Biological Engineering and Computing*, MBEC-D-18-00029R2. (Conditional acceptance, third revision).

5. Poh, A. H., Adikan, F. R. M, Moghavvemi, M, Sharifah Faridah Syed Omar, Poh, K., Mahyuddin, M.B.H., Yan, G., Azizah Ariffin, M.A., Wadi Harun, S. (2019). Precursors to Non- Invasive Clinical Dengue Screening: Multivariate Signature Analysis of In-Vivo Diffuse Skin Reflectance Spectroscopy on Febrile Patients in Malaysia. *PLoS ONE* (PONE-D-19-15612)

6. Poh, A. H., Adikan, F. R. M, Moghavvemi, M, Poh, K., Mahyuddin, M.B.H., Yan, G., (2019). Twofold Multivariate Analytics-Based Method for Auto-detection of Spectroscopic Data Inadmissibility: A Case in Medical Applications. *Spectroscopy Letters* (MS_SL_178_19)