A Quantitative Characterization of Tuberculin Skin Test Indurations using Hyperspectral Imaging to Enable Automated Latent Tuberculosis Screening

By

Ajibola Samson Oladokun

OLDAJI001

A dissertation submitted to the Division of Biomedical Engineering, Department of Human Biology, in fulfilment of the requirements for the degree of

Doctor of Philosophy

Supervisors: Dr. Bessie Malila

Prof. Tinashe Mutsvangwa

April 2025

DECLARATION

I, ***Ajibola Samson Oladokun***, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 22 April 2025

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, Dr. Bessie Malila and Prof. Tinashe Mutsvangwa, for their constant support and guidance throughout this PhD research. This research was produced with the financial assistance of the European Union (EU) (Grant no. DCI-PANAF/2020/420-028), through the African Research Initiative for Scientific Excellence (ARISE) pilot program, implemented by the African Academy of Sciences (AAS). The research was also supported by the University of Cape Town International Scholarship and the South African Research Chairs Initiative (SARChI) of the NRF and the Department of Science and Technology (grant no. 98788). I appreciate all the funders of this research.

I also appreciate my current and former lab mates—Alinafe Kaliwo, Abby Blocker, Maurine Chepkoech, Humphrey Otieno, Tadiwanashe Mabhudhu, Prosper Magara, Shannon Matiyashe, Hervé Nbonsou, Catherine Namayega, Cait Farrell, and Yvonne Karanja—from the telemedicine and connected care, and Mi2D2 research groups for their friendship and support.

I am deeply grateful to my parents, Festus and Foluke Oladokun, for their support throughout my academic journey from childhood to the present. From selling your cars to fund the secondary school education of me and my brothers (and remaining without a vehicle for years), to taking loans to fund my undergraduate degree, I could not have wished for better parents. I hope that you are proud of who your son is becoming. Your investments in me are yielding good fruits – may you continue to reap the rewards of your labour for many decades.

I appreciate my younger brothers, Olawale, Abimbola, Oluwaseun, and David Oladokun, for their friendship and moral support. I appreciate Olawale and Abimbola for occasionally supporting me financially from your side hustles. I am glad to be an elder brother to you all. You give me reasons to strive to be a positive example for you.

I appreciate the ministers and members of RCCG Latter House Church for their prayers and spiritual guidance throughout my research. I thank my church choir and RCF UCT society members for granting me the opportunity to serve as their coordinator. Every session of worship music, Bible study, and prayers with you consistently refreshed and renewed my mind whenever I was not actively working on my research. You are my family in Cape Town. Thank you for the invaluable experience.

Most importantly, I offer my utmost gratitude to God – the creator of my mind, consciousness, and identity. Thank you for setting up the initial conditions of my life – my wonderful family – and for fine-tuning me every step of the way to fulfil your purpose. Thank you for connecting me to all the people who have directly and indirectly contributed to the production of this research. You give me reasons

to keep moving forward by reminding me that "*those who trust in the Lord will find new strength. They will soar high on wings like eagles. They will run and not grow weary. They will walk and not faint*" – Isaiah 40:31.

ABSTRACT

An estimated two billion people globally are believed to harbour latent tuberculosis infection (LTBI), which is a precursor to active tuberculosis (TB). Immunocompromising conditions, such as human immunodeficiency virus (HIV) infection, diabetes, chronic renal failure, and chemotherapy, are associated with the risk of progression from LTBI to active TB. People with LTBI represent a major reservoir for new active TB cases. Thus, the detection and management of LTBI is key to the world health organisation (WHO) goal to reduce TB incidence globally by 90% by the year 2035.

The tuberculin skin test (TST) is the most widely used test for LTBI in low- and middle-income countries (LMICs). It involves the injection of tuberculin into the forearm skin of a participant. An induration – a subdermal lump – forms at the injection site 48 – 72 hours after the test. Using the Mantoux reading method, diagnosis is obtained using a pen and a ruler to measure the induration diameter and comparing it with consensus thresholds. Mantoux readings are subject to inter- and intra-observer subjective variations in readings. The manual nature of this method precludes any repeat analysis or validation after the 72-hour expiry of an induration. Furthermore, the Mantoux method does not provide any significant insight on the pathophysiology of indurations which could improve current understanding of LTBI. Recent applications of hyperspectral imaging (HSI) in other dermatological applications provides evidence of its viability in addressing the shortcomings of the TST. Thus, the aim of this research was to develop and validate an HSI-based approach for quantitative characterisation of TST indurations that can aid the diagnosis of LTBI. This aim was achieved through three objectives.

The first objective was to develop and validate an HSI protocol for reliable and reproducible acquisition of spectral signatures of TST injection sites. An enclosure was designed to implement and validate the developed protocol. The validation was done using a range of light source types and ten participants with a wide range of skin tones. The effect of light source type on skin spectral signature was investigated to reveal factors that impact spectral reproducibility.

The validated imaging protocol was utilized in the second objective to reliably capture the hyperspectral images of 70 participants (38 from South Africa and 32 from Vietnam). The objective was to develop and validate an HSI-model that can generate precise diameter estimations of TST indurations that are comparable to traditional Mantoux readings. Principal component analysis was utilized to transform the hyperspectral images (or hypercubes) into principal components and generate induration segmentation masks. Precise induration diameter estimations, which were comparable to corresponding Mantoux readings, were generated from the masks. An intraclass correlation coefficient score of 0.80, a median difference in reading of 0.29 mm, and a Pearson's

correlation of 0.664 ($p<0.001$), was achieved between the HSI-based diameter estimations and corresponding Mantoux readings. This metrics surpassed the performance of similar studies. This is evidence that HSI can mitigate the subjectivity of the Mantoux method by generating comparable but more precise diameter estimations of TST indurations.

The induration hypercubes and segmentation masks were utilized in the third objective to develop and validate an HSI framework capable of identifying biomarkers that accurately and consistently characterise TST indurations. The framework enabled the generation of chromophore maps from induration hypercubes. Radiomics was applied on the chromophore maps to generate a set of features. The features were used, in conjunction with a support vector machine model, to rank the ability of the generated chromophore maps to predict LTBI. Ferritin, water, and oxyhaemoglobin maps were identified as the most predictive chromophore maps for LTBI. The predictive accuracies for TST diagnosis using these chromophores were 90.0% across all participants, and 96.0% and 96.7% for the South African and Vietnamese participants. This is evidence that HSI can noninvasively facilitate the generation of chromophore biomarkers to characters TST indurations – a first of its kind in literature.

These findings present HSI as a potential modality for mitigating the subjectivity of the Mantoux method and identifying digital biomarkers that noninvasively offer insights into induration pathophysiology. The potential impact of this research is improved health outcome for at-risk TB groups via enhanced LTBI screening accuracy and precision.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BC	Beta Carotene
BCG	Bacillus Calmette-Guerin
BL	Bilirubin
CART	Classification and Regression Trees
CCDs	Charge-coupled Devices
CG	Collagen
DNA	Deoxyribonucleic Acids
DX	Deoxyhaemoglobin
EM	Eumelanin
ET	Elastin
FMC	Feature Map Channel
FPA	Focal Plane Array
FR	Ferritin
FT	Fat
FWHM	Full Width at Half Maximum
GLCM	Gray Level Co-occurrence Matrix
GLDM	Gray Level Dependency Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GMM	Gaussian Mixture Model
Hb	Deoxyhaemoglobin
HbO_2	Oxyhaemoglobin
HGLRE	High Gray Level Run Emphasis
HIV	Human Immunodeficiency Virus
HREC	Human Research Ethics Committee
HSI	Hyperspectral Imaging
ICC	Intraclass Correlation Coefficient
ID	Inverse Difference
IDM	Inverse Difference Moment
IFN-γ	Interferon-gamma
IGRA	Interferon Gamma Release Assay

IMC	Information Measure of Correlation
InGaAs	Indium Gallium Arsenide
LC	Lycopene
LED	Light-emitting Diode
LMICs	Low- and Middle-income Countries
LTBI	Latent Tuberculosis Infection
MAD	Mean Absolute Deviation
MDA	Model-based Discriminant Analysis
MIR	Mid-infrared
ML	Machine Learning
MOI	Medical Optical Imaging
Mtb	*Mycobacterium tuberculosis*
NGLDM	Neighbouring Gray Level Dependence Matrix
NGTDM	Neighbouring Gray Tone Difference Matrix
NIR	Near Infrared
OX	Oxyhaemoglobin
PCA	Principal Component Analysis
PLWH	People Living with HIV/AIDS
PM	Pheomelanin
PPD	Purified Protein Derivative
PVD	Peripheral Vascular Disease
RES	Residual Determinants
RGB	Red-Green-Blue
RNA	Ribonucleic Acids
ROI	Region of Interest
SVD	Singular Value Decomposition
SVM	Support Vector Machines
TB	Tuberculosis
TBST	Mtb Antigen-based Skin Tests
TNF-α	Tumour Necrosis Factor Alpha
TST	Tuberculin Skin Test
UCT	University of Cape Town
UV	Ultraviolet
VIS	Visible

VNIR	Visible and Near-infrared
WHO	World Health Organization
WT	Water

CHAPTER 1

Introduction

1.1. Motivation

About two billion of the global population is estimated to have latent tuberculosis infection (LTBI) (Sterling et al., 2020). This infection is a precursor to the onset of active tuberculosis (TB), which is a leading cause of mortality amongst people living with HIV/AIDS (PLWH) (Houben & Dodd, 2016). The causative agent of LTBI is *Mycobacterium tuberculosis* (*Mtb*) – a species of pathogenic bacteria. The *Mtb* bacteria is the leading cause of death globally from a single infectious agent. Over 10 million new TB cases and 1.4 million TB-related deaths were reported in the World Health Organization (WHO) Global Tuberculosis Report 2020 (Gong & Wu, 2021). Africa, Southeast Asia, and Western-Pacific are regions with the highest TB burden, and they account for about 80% of all *Mtb* infections globally (Gong & Wu, 2021). Particularly, South Africa has one of the highest TB burdens globally, with an incidence rate of 468 per 100,000 population reported in 2022 (Izrah, 2024). The effect of TB infection is compounded by biosocial risk factors such as HIV co-infection, alcohol use disorders, smoking, and diabetes. An estimated 57% of the 54,000 TB-related deaths in South Africa in 2022 occurred in HIV-infected persons (Izrah, 2024). This highlights the importance of early TB diagnosis in South Africa and other TB-endemic regions.

Two stages of *Mtb* infection are currently clinically recognized – LTBI and active TB. There is growing recognition that the infection could exist as a spectrum of states. While these intermediate states are still being investigated, LTBI and active TB are still the consensus broad classification of the infection states (Zaidi et al., 2023). The LTBI disease state is an asymptomatic clinical state that is not transmissible such that the bacillary replication of inhaled *Mtb* is impeded or kept below an undefined threshold by a human host's immune system. If the host's immune system becomes compromised, the contained *Mtb* bacteria can commence rapid replication and progression to active TB (Mir et al., 2022). Active TB is characterized by symptoms such as lung tissue damage, cough, haemoptysis, fever, night sweats, chills, fatigue, and weight loss (Shah & Dorman, 2021). Active TB is transmissible through *Mtb*-infected droplet nuclei which become airborne when an individual with active TB coughs. The lifetime risk of progression from LTBI to active TB in otherwise healthy individuals is between 5% and 15% (Shah & Dorman, 2021). People living with human immunodeficiency virus (HIV; PLHIV) who are co-infected with LTBI have the highest risk of progression from LTBI to active TB, more than a 100-fold compared with otherwise healthy members of the populace (Kiazyk & Ball, 2017). Thus, the importance of early diagnosis of LTBI is heightened by the vulnerability of immunocompromised PLWH

to suffer from TB. The main goal of the WHO End TB Strategy is to reduce TB incidence globally by 90% and TB-related deaths by 95% by the year 2035. People with LTBI represent a major reservoir for new active TB cases (Kiazyk & Ball, 2017). Thus, LTBI is one of the biggest obstacles to achieving the WHO End TB Strategy goal. The detection and management of LTBI is therefore a key component to the attainment of this goal (Gong & Wu, 2021).

Strategies for LTBI control are anchored in screening at-risk populations and offering preventive therapy to those at highest risk of developing active TB disease. Development and validation of accurate, affordable, and scalable diagnostic tests for LTBI infection remain a priority (Krutikov et al., 2022). The most widely used method for detecting LTBI in low- and middle-income countries (LMICs) is the Mantoux tuberculin skin test (TST) (Carranza et al., 2020). The TST is administered *in vivo* by injecting tuberculin (purified protein derivative) into the forearm skin dermis of a participant. Tuberculin is a cocktail of several antigens extracted from the *Mtb* bacteria. Macrophages, a type of white blood cell, engulf and process the injected tuberculin antigen as a foreign substance. The macrophages present processed fragments of the antigen to CD4+ T lymphocytes – another type of white blood cell (Flynn et al., 2011). The CD4+ T cells in LTBI positive individuals are sensitized to the *Mtb* proteins in the antigen fragments due to prior TB infection. The sensitized CD4+ T cells become activated and release cytokines. Cytokines are a group of cell-signalling proteins that increase permeability of local blood vessels to attract more macrophages, CD4+ T cells, other immune cells, and fluid to the reaction site (Arango Duque & Descoteaux, 2014; Vukmanovic-Stejic et al., 2006). This localised concentration of cells and molecules lead to the formation of an induration – a stiff region of dermal inflammation – 48-72 hours after the initial injection. Among the cytokines is interferon-gamma (IFN-γ) – a type of cytokine that activates more macrophages to process more antigens and repeat the cycle (Andersen et al., 2000). Using the Mantoux reading method, LTBI diagnosis is obtained by manually measuring the diameter of the induration, using a pen and a ruler, and comparing it with consensus thresholds (Kiazyk & Ball, 2017). The Mantoux method is a crude but simple approximation of the underlying physiological reactions of an induration.

The interferon gamma release assay (IGRA) is an alternative method for diagnosing LTBI with improved specificity compared to the TST method (Gualano et al., 2019). The IGRA is a whole blood-based *in vitro* test where the blood sample from an individual is exposed to *Mtb* antigens to assess cell-mediated immune response to the antigens. Using the IGRA test, LTBI diagnosis is obtained by measuring the levels of interferon-gamma produced by the immune system cells in response to the antigens. The IGRA test requires significant laboratory infrastructure and technical capacity which makes it more expensive to administer than the TST method (Kiazyk & Ball, 2017). A single IGRA test costs between 35 USD to 40 USD per patient while the TST costs as low as 3 USD per patient (Nijhawan

et al., 2016). Thus, the cost-effectiveness and ease of implementation of the Mantoux TST method makes it the most common test for LTBI in LMICs (Carranza et al., 2020).

The Mantoux method for reading TST indurations, which is the current clinical practice, is prone to inter- and intra- observer variations in readings (Goel et al., 2018; Moayedi-Nia et al., 2019). The inter- and intra- observer subjectivity of the method is due to its manual nature, which heavily relies on the reading technique and experience of a clinician. Moreover, the manual nature of the Mantoux method precludes the generation of a digital data format, thereby hindering repeat analysis, interpretation, visualisation, and validation of a patient's TST induration after the 48–72-hour validity window has elapsed. As a result, reassessment necessitates the administration of a new LTBI skin test. This introduces the need for a digital, standard, automated, precise, and reproducible method for capturing TST indurations and generating quantitative features that consistently characterise LTBI and give a diagnostic result.

The use of biopsy to take samples from a TST induration of patients for cellular and molecular analysis is an invasive procedure that may be impractical in large scale LTBI screening. The Mantoux reading method offers a simple, cheap, non-invasive, but imprecise way of approximating the underlying physiological processes of an induration. Medical optical imaging offers the ability to capture TST indurations in a digital format that facilitates post-induration analysis, validation, and generation of quantitative features that are precise and standardised. Medical optical imaging refers to the use of visible and near-infrared (VNIR) light in the range (380 nm to 1000 nm) of the electromagnetic spectrum to capture images of body tissue (Halicek et al., 2019). The use of the red-green-blue (RGB) sub modality of medical optical imaging for capturing TST indurations has been investigated in some studies [Dendere et al. (2017) and Naraghi et al. (2018)]. These studies propose leveraging the ubiquity of smartphones to capture RGB images of TST indurations and utilizing image processing techniques to generate TST readings from the images. These studies proposed the imaging of a TST induration from seven predefined smartphone orientations. This enabled the photogrammetric reconstruction of 3D image representations of the indurations from the seven RGB images. The rationale behind this was to generate a 3D digital representation of TST indurations such that induration diameters could be estimated from the 3D images. The approach was first tested on cosmetically generated mock indurations and promising results were obtained. However, the approach was unable to generate consistent results when tested on real TST indurations (Maclean, 2020).

Unlike simulated indurations, real TST indurations develop beneath the skin surface and may not consistently display distinct geometric contrast relative to the surrounding skin. In addition, real TST indurations are typically subsumed in erythema which makes it more difficult to identify the induration

boundaries (Toivgoogiin et al., 2005). Furthermore, the RGB imaging and 3D reconstruction approach used in the above studies does not natively offer non-geometric characterisation of TST indurations, such as the estimation of underlying pathophysiological processes, which could provide more insights on the formation of TST indurations. Thus, there is a need for an improved medical optical imaging approach that addresses the subjectivity of the Mantoux reading method by capturing TST indurations in a digital, standard, automatable, precise, and repeatable format. Furthermore, there is a need for the approach to be capable of generating quantitative non-geometric features that characterise TST indurations.

Hyperspectral imaging (HSI) is a non-invasive optical imaging technique with promising applications in medicine which include skin cancer, burn wounds, skin bruises and diabetic foot in dermatology (Vasefi et al., 2016). Hyperspectral imaging has the potential to address the limitations encountered by RGB imaging studies in the reading of TST indurations. It captures the spectral intensities of light reflected across VNIR wavelength range by skin tissue. Unlike in smartphone imaging where broad RGB bands are reliably captured, HSI enables the capturing of reflected light over a continuous range of wavelengths (Lu & Fei, 2014). This facilitates a richer acquisition of optical information. Skin tissue absorbs and reflects different wavelengths of light to varying degrees. This encodes quantitative subdermal information about the biological and structural properties of tissues, such as, inflammation, pigmentation, tissue oxygen saturation and blood perfusion, and vascular structure (Pardo et al., 2018). Hyperspectral images of skin tissue (also called hypercubes) are digital representations of spectral signatures of the skin captured in a standard format – a format that is consistent across most HSI studies (Vasefi et al., 2016). The digital and standard form of hypercubes enables the development of automated pipelines where the screening of a medical condition can be achieved using HSI with minimal human input as demonstrated in Hosking et al. (2019). This suggests that HSI can enable the capturing of TST indurations in a digital, standard, and automatable format which are features that the current Mantoux reading method is unable to provide.

Furthermore, HSI has shown remarkable precision, with consistent spectral signatures observed within body tissue types. It has also shown excellent reproducibility, with consistent spectral signatures observed for the same tissue under identical imaging conditions, in various medical applications (Abdlaty, Abbass, & Awadallah, 2021; Di Cecilia et al., 2018). This suggests that HSI could potentially facilitate the acquisition of precise and reproducible measurements from TST indurations in the form of spectral signatures. Additionally, HSI has been shown to generate quantitative non-geometric features from spectral signatures in the form of body tissue chromophore concentrations such oxyhaemoglobin and deoxyhaemoglobin concentrations (Giannoni et al., 2021).

Machine learning (ML) is a subfield of artificial intelligence (AI) techniques where models are developed to learn and recognize patterns in a dataset (Nozari & Sadeghi, 2021). Machine learning has been utilized to learn and recognize features in datasets of medical hyperspectral images that are indicative of underlying diseases (Cui et al., 2022). Thus, the ability of HSI, aided by ML, to generate quantitative non-geometric features in a digital, standard, automatable, precise, and reproducible form, suggests that it could be a viable modality to address the subjectivity of the Mantoux reading method of the TST which leads to variability of diagnostic information.

There is currently no significant optical imaging research that addresses the subjectivity of the Mantoux method of reading real TST indurations as a pathway to improved LTBI diagnosis. Thus, this doctoral research investigated the viability of quantitative characterization of TST indurations using HSI to enable automated and accurate LTBI screening.

1.2. Problem Statement

Active TB is a leading cause of mortality globally. People with LTBI represent a major reservoir for new active TB. Thus, LTBI is one of the biggest obstacles to achieving the goal of the WHO End TB Strategy. The screening of LTBI is a key component to the attainment of this goal. The Mantoux TST is the most common method for diagnosing LTBI in LMICs. The manual acquisition of Mantoux readings impedes automated screening of LTBI. Furthermore, the diameter readings obtained by the Mantoux method do not provide physiological insights on the formation of TST indurations or the underlying LTBI which are a result of biological responses to the tuberculin injection. Studies have shown the potential of optical imaging to address the shortcomings of the Mantoux method. Thus, there is a need for the development of an optical imaging solution that provides quantitative non-geometric characterisation of TST indurations based on physiological responses on the TST injection site in a digital, standard, automatable, precise, and reproducible form to enable accurate LTBI screening that is independent of inter-observer variability.

1.3. Research Scope

Besides the problem of the subjectivity of the Mantoux reading method, the TST suffers from the problem of the requirement for secondary clinical visits. Patients are required to return to the clinic 48-72 hours after the initial injection of tuberculin to read the induration that may have formed in the period. This requirement has resulted in the invalidation of several tests because many patients fail to return, often due to factors like distance, travel costs, and scheduling conflicts with other commitments. These rural healthcare challenges have led to a patient dropout rate of around 17% after initial clinical visits for the TST (Faujdar et al., 2022). From the perspective of a patient, the secondary visit requirement could be a more important problem than the subjectivity of the Mantoux

method. This problem is also of significant importance to the public health sector of governments, as low turnout for LTBI screening directly impacts their TB management initiatives. Studies such as Dendere et al. (2017) and Naraghi et al. (2018) proposed the utilisation of the wide accessibility of smartphone imaging to address this problem. The solution was proposed to enable patients to read their TST indurations at home using their smartphone camera which could eliminate the need to return to the clinic for reading. This would enable automated screening of LTBI which would also be important for public health initiatives for TB management. However, the inability of smartphone imaging to achieve consistent readings from real TST indurations, as reported in Maclean (2020), highlights the need to solve the subjectivity problem as a prelude to the secondary visit problem.

This study proposes an image-based pathway to achieving automated LTBI screening, as illustrated in Figure 1.1. It proposes the use of HSI to address the subjectivity of the Mantoux reading method and to provide a digital characterisation of TST indurations. The mitigation of subjectivity is of interest to clinicians as the ability to obtain precise and reproducible readings from TST indurations using HSI enables them to give consistent diagnosis. Researchers in the TB field rely on invasive skin biopsies to obtain insights on the induration pathophysiology with respect to normal skin physiology. The demonstrated ability of HSI to noninvasively obtain these insights is an outcome that will be of interest to the TB research community. The digital nature of the HSI provides researchers a snapshot of the reaction state of an induration which can be reanalysed for additional insights using a plethora of techniques. Thus, the scope of this thesis was to provide evidence – the first of its kind – that HSI can mitigate the subjectivity of the Mantoux reading method as well as to noninvasively provide insights into induration pathophysiology.

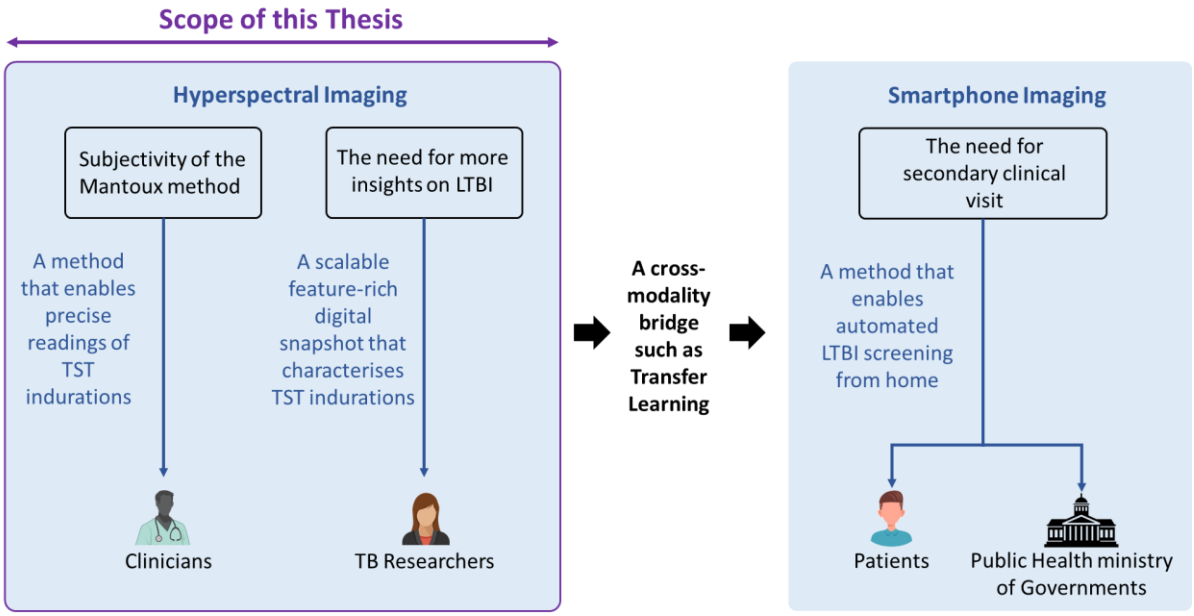


Figure 1.1: A possible image-based pathway to automated screening for latent tuberculosis infection. The scope of this thesis is an investigation of feasibility of the first part of this pathway.

The limited accessibility of HSI to patients limits its ability to directly enable automated LTBI screening. However, the existence of cross-modality techniques like transfer learning, which enable the transference of characteristics from one modality to another, means that there is a pathway through which HSI can facilitate automated LTBI screening. Transfer learning models can potentially be trained on a large dataset of matching pairs of HSI and smartphone images of TST indurations. This could enable the model to learn a mapping between the two modalities such that it is able to estimate hyperspectral features that predict LTBI when given only the smartphone image of an induration. With such an approach, it could be possible to achieve the superior HSI precision and characterisation of the TST using only smartphone images.

The scope of this thesis was limited to the investigation of the effectiveness of HSI for the characterisation of TST indurations. The output of this investigation is the evidence of effectiveness of HSI which can be a platform on which future studies can build on to achieve automated LTBI screening.

1.4. Research Questions

The research questions that encapsulate the research gap highlighted in the problem statement are as follows

1. Can spectral signatures of TST injection sites be reliably and reproducibly captured using a hyperspectral camera?
2. Can hyperspectral imaging of TST indurations enable the generation of precise diameter estimations that are comparable to traditional Mantoux readings?
3. Can hyperspectral imaging non-invasively capture features that are good representations of the underlying pathophysiological responses to tuberculin antigen at a TST injection site?

1.5. Research Hypothesis

The research hypothesis of this doctoral research is a testable statement that answers the research questions posed and it is as follows.

> H_1: Hyperspectral imaging, aided by machine learning, can capture clinically important subdermal features of TST indurations indicative of the underlying tuberculin-triggered immune response, which could address the inter-observer variability in the Mantoux TST method.

1.6. Aim and Objectives

The aim of this research was to develop and validate a hyperspectral imaging-based approach for quantitative characterisation of TST indurations that can aid the diagnosis of latent tuberculosis infection.

The aim was achieved through the following objectives:

I. Development and validation of a hyperspectral imaging protocol for reliable and reproducible acquisition of spectral signatures of TST injection sites.
II. Development and validation of a hyperspectral imaging model that can generate precise diameter estimations of TST indurations that are comparable to traditional Mantoux method readings.
III. Development and validation of a hyperspectral imaging framework capable of identifying biomarkers that accurately and consistently characterise TST indurations.

The validation of the characterisation framework in objective III complements the results of objectives I and II to fully test hypothesis H_1.

1.7. Contributions

The main contributions of this thesis are as follows:

1. A first of its kind investigation of the impact of various VNIR illumination sources, i.e., halogen, xenon, broadband light-emitting diode (LED), and white LED, on the reproducibility of skin spectral signatures across different skin tones (Objective 1). The study by He et al. (2020) is currently the only other study that does a similar investigation. However, that study only investigated the impact of various colour temperatures of visible spectrum LED on skin spectral signatures, compared to several VNIR illumination types on different skin tones investigated in our study. Furthermore, this thesis presents, for the first time in the literature, the use of flanking reflectors and a white reference panel within a constructed enclosure to minimise the impact of confounding ambient illumination. These contributions are published in the following paper:
    a. **Oladokun, A.S.,** Mutsvangwa, T., Malila, B. (2024). Understanding the Impact of Different Light Sources on the Reproducibility of Skin Spectral Signature for Different Skin Tones. In *2024 IST-Africa Conference (IST-Africa)* (pp. 1-12). IEEE. https://doi.org/10.23919/IST-Africa63983.2024.10569556

2. The generation of precise TST induration diameters using HSI images was demonstrated in this study for the first time in literature. Image-based generation of induration diameters have thus far involved RGB imaging of mock indurations (Naraghi et al., 2018) and thermal imaging (Fiz et al., 2015). The evidence of improved precision of induration diameter estimation presented in this study, presents HSI as the leading imaging modality for mitigating the subjectivity of Mantoux readings – a desirable outcome for clinicians. This contribution is published in the following paper:

    b. **Oladokun, A. S.,** Malila, B., Shey, M., & Mutsvangwa, T. (In Print). Hyperspectral Imaging for the Diagnosis of Latent Tuberculosis Infection. In A. Abraham & A. Bajaj (Eds.), *Computational Intelligence based Hyperspectral Image Analysis*. Springer.

3. The identification of chromophore biomarkers using HSI to accurately characterise TST indurations is a first of its kind in the literature. The thermal imaging study by Fiz et al. (2015) is the only research thus far that has proposed an image-based biomarker for TST indurations. It proposed temperature maps as digital representations of indurations. Conversely, the following research demonstrates the effectiveness of HSI-enabled chromophore maps generated as digital representations of TST indurations. The chromophore biomarkers identified here were related to the cellular and molecular characteristics of indurations described by biopsy studies such as Vukmanovic-Stejic et al. (2006). The identified chromophore biomarkers provide a new characterisation of TST indurations, different from thermal imaging, that may facilitate further TB investigations. This is a potentially significant outcome for TB researchers as this research presents evidence that HSI can facilitate scalable image-based biopsies of induration which could enrich the current understanding of LTBI. This contribution is published in the following paper:

    c. **Oladokun, A.S.,** Malila, B., Campello, V.M., Shey, M., Mutsvangwa, T.E.M. (2024). SpeChrOmics: A Biomarker Characterization Framework for Medical Hyperspectral Imaging. In: Linguraru, M.G., *et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. MICCAI 2024. Lecture Notes in Computer Science, vol 15003. Springer, Cham. https://doi.org/10.1007/978-3-031-72384-1_70

4. The identification of an image-based biomarker from TST indurations to predict IGRA diagnosis is studied in this study for the first time in literature. Biopsy studies such as Vukmanovic-Stejic et al. (2006) reveal that IFN-γ is among the cytokines present in a TST induration. This suggests that the gap in agreement between the TST and the IGRA, which is reported to be around 74.7% (Apriani et al., 2024), can in theory be bridged if IGRA readings

can be estimated directly from the IFN-γ levels in a TST induration. No current study has attempted to bridge this gap. Here, our research suggests a correlation between HSI-enabled chromophore maps and IGRA diagnosis implying that HSI may capture the IFN-γ levels of an induration. This serves as preliminary evidence that can be validated in future research.

To lay the groundwork for future research on the integration of transfer learning to merge HSI and RGB imaging for automated LTBI screening, an initial investigation was conducted on the utilization of telemedicine to transmit images to a remote server housing a diagnostic model. This preliminary investigation is detailed in the following papers.

i. **Oladokun, A.,** Malila, B., & Mutsvangwa, T. (2022). A Secure and Reliable Hyperspectral Imaging-Based Telemedicine Network Architecture. *Southern Africa Telecommunication Networks and Applications Conference (SATNAC) 2023* https://doi.org/10.25375/uct.26339719.v1

ii. **Oladokun, A.,** Mutsvangwa, T., & Malila, B. (2023). A Network Resource-Aware Hyperspectral Imaging-Based Telemedicine Architecture for Rural, Remote, and Underserved Communities. *Southern Africa Telecommunication Networks and Applications Conference (SATNAC) 2023* https://doi.org/10.25375/uct.26339734.v1

1.8. Thesis Outline

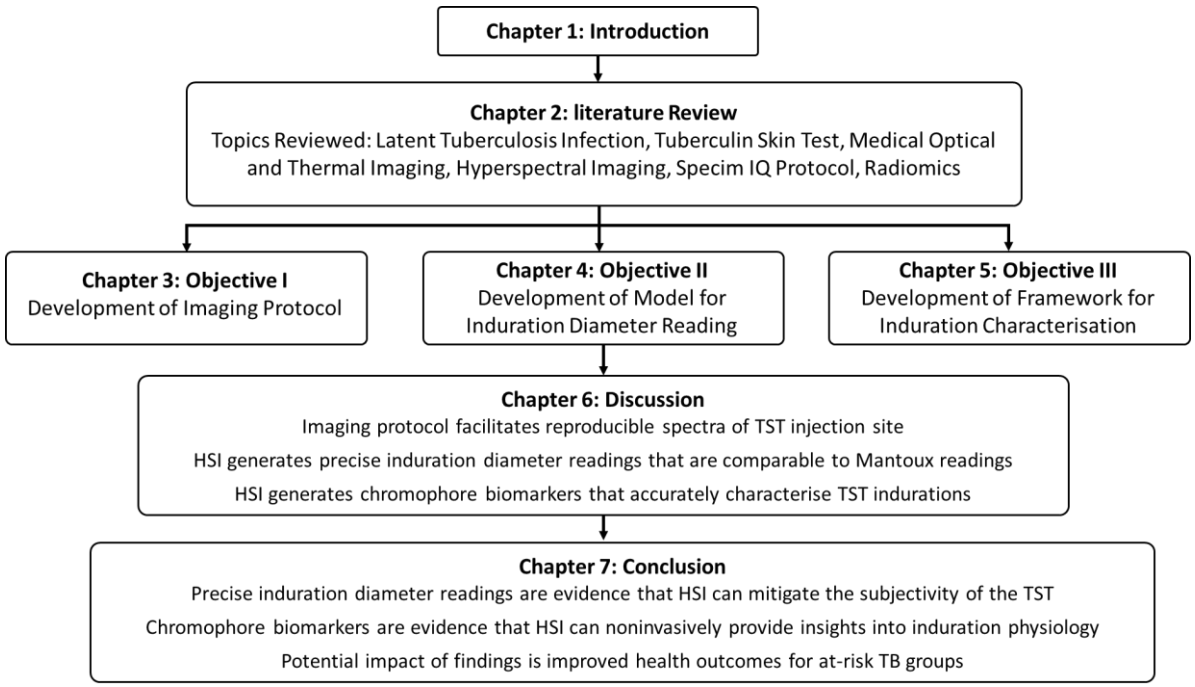


Figure 1.2: Thesis outline

This thesis is organised as follows. **Chapter 2** provides a review of literature on LTBI, TST, HSI and the modelling techniques utilized in the implementation of the research second and third objectives

outlined in Chapter 1. **Chapter 3** presents the development and validation of an imaging protocol for reproducible acquisition of spectral signatures of TST skin sites. **Chapter 4** introduces the development and validation of an HSI-based model that leverages the protocol outlined in Chapter 3 to capture hypercubes of real TST indurations. The chapter also describes the ability of the model to generate induration segmentation masks from the hypercubes. The process of utilizing the segmentation masks to generate precise diameter estimations of TST indurations that are comparable to traditional Mantoux readings, is further detailed in the chapter. **Chapter 5** describes the development and validation of an HSI-based framework, that utilises ML techniques and the masks from Chapter 4, to generate physiological biomarkers for accurate and consistent characterisation of TST indurations. **Chapter 6** presents an analysis of the results obtained in chapters 3 to 5. **Chapter 7** concludes the thesis, summarising the research findings, their implications for, and contributions to current scientific discourse. The chapter also outlines the limitations of this research and proposes recommendations for future research.

This research was motivated by the need for new measures in the detection and management of LTBI as people with this disease state represent a major reservoir for new active TB cases. Automated LTBI screening can improve the detection and management. Thus, this research proposed an image-based pathway to automated LTBI screening by presenting HSI as a modality to address the current problem of the subjectivity of the TST. This subjectivity problem informed the formulation of the hypothesis that HSI, aided by machine learning, can capture clinically important subdermal features of TST indurations indicative of the underlying tuberculin-triggered immune response, which could address the inter-observer variability in the Mantoux TST method. Evidence to support this hypothesis was generated using three objectives, first of which was to develop and validate an HSI protocol for reliable and reproducible acquisition of spectral signatures of TST injection sites. The developed protocol was utilized to capture hyperspectral images of TST indurations from 70 participants. The images served as input to the second objective which was to develop a model to generate precise induration diameter estimations from the images that are comparable to traditional Mantoux method readings. Induration masks were generated in the process of generating the induration readings. The hyperspectral images and the induration masks served as input to the third objective which was to develop an HSI framework that identified biomarkers that accurately and consistently characterise TST indurations. The key findings from this study include the identification of ferritin, water, and oxyhaemoglobin as chromophore biomarkers for LTBI diagnosis. These HSI-generated chromophore maps are presented as potential digital representations of induration pathophysiology. The relationship between the identified chromophore biomarkers and the induration pathophysiology described by skin biopsy is evidence of the viability of HSI as a potential imaging modality to

noninvasively facilitate largescale biopsies of indurations. Another key finding of this study is the identification of preliminary evidence to support the potential ability of HSI to bridge the gap between the TST and the IGRA tests by facilitating the identification of chromophores in indurations that predict IGRA results. This opens the possibility of new research area into the development of low cost HSI cameras to potentially estimate IGRA results from a TST induration, which could eliminate the laboratory infrastructure required for the IGRA while preserving the simplicity and affordability of the tuberculin injection process of the TST. This study also includes preliminary work to address the need for improving accessibility to healthcare for rural, remote and underserved communities by leveraging the ubiquitous nature of cellular networks and the capabilities of telemedicine systems. This opens the possibility of research into the use of transfer learning, supported by telemedicine, to combine the strengths of HSI and smartphone imaging to enable automated LTBI screening. Thus, the generation and visualisation of HSI-enabled features is presented in this study as a digital, standard, automatable, precise, and reproducible representation of TST indurations that potentially enables accurate LTBI screening that is independent of inter-observer variability.

CHAPTER 2

Literature Review

This chapter presents a background information on latent tuberculosis infection and the tuberculin skin test. It also reviews current literature on the application of medical optical imaging and thermal imaging for TST diagnosis. The protocol for the utilization of Specim IQ – the HSI camera of choice in this research – is also reviewed along with background information on the use of radiomics to extract features from medical images. The reviewed content presented here provides prerequisite knowledge required for subsequent chapters.

2.1 Latent Tuberculosis Infection and the Mantoux Tuberculin Skin Test

Mycobacterium tuberculosis – a species of pathogenic bacteria infects an estimated 2 billion people, about a quarter of the global population (Sterling et al., 2020); and is the causative agent of TB, one of the leading causes of death globally from a single infectious agent. Two stages of *Mtb* infection are currently clinically recognized – LTBI, an asymptomatic clinical state that is not transmissible, and active TB, which is characterized by symptoms such as lung tissue damage, cough, haemoptysis, fever, night sweats, chills, fatigue, and weight loss (Shah & Dorman, 2021). Exposure to *Mtb* via inhalation of infected airborne droplet nuclei leads to LTBI which is a state in which bacillary replication of *Mtb* is impeded or kept below some undefined threshold due to control mechanisms of the host's immune system. The LTBI can be clinically defined as a state of persistent immune response or sensitization to stimulation by *Mtb* antigens without evidence of clinically manifested active TB (Shah & Dorman, 2021).

The prevalence of the *Mtb* pathogen is partly due to its ability to persist within its host in an asymptomatic state of dormancy with the risk of reactivation into active TB at some point in the future (Chee et al., 2018). This state of dormancy is a reversible state of low bacterial metabolism where *Mtb*, as an immune-evading strategy, transitions from replicating to non-replicating bacilli and surviving for long periods without replication (Carranza et al., 2020). The lifetime risk of progression from LTBI to active TB in otherwise healthy individuals is between 5% to 15% (Shah & Dorman, 2021). People co-infected with HIV and LTBI are the group with the highest risk of progression from LTBI to active TB, with a risk of more than a 100-fold compared with otherwise healthy members of the populace (Kiazyk & Ball, 2017). There are other groups of individuals with varying risks, ranging from high to low risk, of progression from LTBI to active TB. High risk groups include chronic renal failure patients requiring haemodialysis, organ transplant patients who require the use of immune suppressants, cancer patients undergoing chemotherapy and radiotherapy, and silicosis patients (Jacobs et al., 2015; Kiazyk

& Ball, 2017). Groups with medium risk include patients who use tumour necrosis factor alpha (TNF-α) inhibitors due to autoimmune and inflammatory diseases, glucocorticoids patients, diabetics of all types (3-times higher risk than non-diabetics), COVID-19 patients, and children with LTBI and aged 4 years or less (Carranza et al., 2020; Gong & Wu, 2021; Kiazyk & Ball, 2017). Low risk group includes individuals with a history of alcohol abuse, cigarette smokers, and underweight or malnourished individuals (Kiazyk & Ball, 2017). A commonality across these groups is some level of suppression of the immune system.

The main goal of the WHO End TB Strategy is to reduce TB incidence globally by 90% and TB-related deaths by 95% by the year 2035. People with LTBI represent a major reservoir for new active TB cases (Kiazyk & Ball, 2017). Thus, LTBI is one of the biggest obstacles to achieving the goal of the WHO End TB Strategy. The detection and management of LTBI is therefore a key component to the attainment of this goal (Gong & Wu, 2021). Vaccination is one of the viable mechanisms to reduce TB prevalence globally. The Bacillus Calmette-Guerin (BCG) vaccine is currently the only vaccine against active TB. The BCG is most effective when given to infants just after birth or to children younger than 5 years old in regions with high TB burden. If infected with *Mtb*, young children are at very high risk of quickly progressing to TB meningitis which is an extrapulmonary form of active TB. The vaccination of infants immediately after birth protects them from developing active TB if they contract LTBI in childhood (Pooransingh & Sakhamuri, 2020). However, the effectiveness of the BCG vaccine wanes with time and offers protection of up to 10-15 years. The vaccine does not offer significant protection against progression to active TB in adults like it does in children. A study across 17 countries showed that the overall effectiveness of the BCG vaccination at preventing active TB in adults over 15 years was 18% (Martinez et al., 2022). The BCG vaccination of adults is not recommended. Revaccination of an individual who was BCG-vaccinated as a child is also not recommended due to significant reduction in effectiveness of a second dosage (Bannister et al., 2021). Therefore, BCG vaccination is a useful but insufficient mechanism in the impediment of the prevalence of TB in adults. This highlights the need for more measures to limit the prevalence of TB.

2.1.1 Detection and Treatment of LTBI

Current measures to control the prevalence of TB include testing for LTBI in regions with high *Mtb* infection rates and treating those who test positive for LTBI and belong to the group with high risk of progression to active TB. Treatment of LTBI involves a 6- to 9-month antibiotic therapy to clear the *Mtb* infection and reduce the risk of development to active TB by up to 90% (Kiazyk & Ball, 2017). The antibiotic therapy involves the use of medications such isoniazid, rifapentine, and rifampin; used individually or as a combination therapy (Sharma et al., 2014). Side effects due to these medications

include gastrointestinal problems, acneiform rashes, cognitive impairment, lethargy, headaches, sleep disturbance, blurred vision, numbness or tingling in the hands or feet, fever, yellowing of skin or eyes, and joint pain (Berg et al., 2004; Denholm et al., 2014). Having side effects over the lengthy 6 to 9 months treatment period results in the lack of completion of the therapy or dose skipping by patients; reducing the effectiveness of the therapy (Sandgren et al., 2016). Due to this, LTBI antibiotic therapy is typically only limited to LTBI positive individuals in groups with high risk or medium risk of progression to active TB. Thus, there is a need for a reliable test for diagnosing LTBI.

While active TB can be diagnosed via bacterial culture from the sputum to detect the presence of *Mtb*, this is not the case with LTBI (Kiazyk & Ball, 2017). The sputum of LTBI patients do not contain *Mtb* as the bacteria is in a dormant state in the lungs. Thus, current LTBI tests do not directly detect the presence of *Mtb*. Latent TB infection is inferred by stimulating the immune system of a patient with *Mtb* antigens and assessing the cell-mediated immune response of the patient. An individual with no prior exposure to *Mtb* will not develop an immune response to the antigens of *Mtb* (Shah & Dorman, 2021). The immune system of an individual with past exposure to *Mtb* will be sensitized to the antigens of the bacteria. Currently, the two most common tests for LTBI are the TST, which has been used for over a century, and the IGRA, which was introduced in 2005 (Chee et al., 2018).

The active substance in the Mantoux TST is tuberculin, a standardized extract of proteins derived from mycobacterium tuberculosis, also known as purified protein derivative (PPD). It consists of a cocktail of approximately 200 antigens extracted from the *Mtb* bacteria and other related bacteria. During the TST test, the Tuberculin is administered *in vivo* by injecting it into the forearm skin of an individual (Carranza et al., 2020; Zellweger et al., 2020). After injection, the cell-mediated immune response to the injected *Mtb* antigens manifests as a firm, raised and hardened area on the injection site, also known as an induration, that forms 48-72 hours after injection. The Mantoux reading method involves obtaining the LTBI diagnosis by measuring the diameter of the induration with a pen and a ruler/calliper, and comparing it with consensus thresholds (Kiazyk & Ball, 2017). The threshold for a positive LTBI diagnosis is ≥ 5 mm for HIV seropositive individuals and organ transplant patients. The threshold is set at ≥ 10 mm for individuals with other LTBI-related risk factors such as exposure to active TB patients, diabetes, and injection drug use; and increased to ≥ 15 mm for individuals with no TB-related risk factors (Al Jahdali et al., 2010).

The IGRA is a whole blood-based *in vitro* test where the blood sample from an individual is exposed to *Mtb* antigens to assess cell-mediated immune response to the antigens. Using the IGRA, LTBI diagnosis is obtained by measuring the levels of interferon-gamma produced by the immune system cells in response to the antigens. The IGRA requires significant laboratory infrastructure and technical

capacity which makes it more expensive to administer than the Mantoux TST (Kiazyk & Ball, 2017). Thus, the latter is the most common LTBI test in low- and middle-income countries due to its low cost and ease of implementation.

Tuberculin, which is the antigen used for the TST, contains some protein strains that are also found in the BCG vaccine (Shah & Dorman, 2021). Thus, the TST may trigger an induration in a BCG-vaccinated individual who is not LTBI positive even after the expiry of the window of effectiveness of BCG. This motivated the recent creation of *Mtb* antigen-based skin tests (TBST) which uses the ESAT-6 (early secretory antigenic target 6) and CFP-10 (culture filtrate protein 10) antigens instead of tuberculin (WHO, 2022). The ESAT-6 and CFP-10 antigens, also used in IGRA, are proteins produced by *Mtb* but are absent in the BCG vaccine and most non-tuberculous mycobacteria. As a result, TBST, which is commercially available as C-Tb, C-TST, and Diaskintest, triggers indurations that are not influenced by BCG vaccination (WHO, 2022). Like the TST, the TBST is read using the Mantoux reading method. The adoption of TBST is still in a very early phase, thus the TST is still the standard skin-based test for LTBI.

2.1.2 Pathophysiology of TST Indurations

Intradermally injected PPD is recognised by the immune system as a foreign substance. Antigen-presenting cells, such as macrophages, dendritic cells, and Langerhans cells, encounter and engulf the injected antigen via phagocytosis. The macrophages present the antigens to T cells, specifically CD4+ T cells (helper T cells). If the CD4+ T cells have been previously sensitized to a *Mtb* infection, they become activated and differentiated into effector T cells (Flynn et al., 2011). The activated T cells secrete various cytokines (signalling molecules), such as IFN-γ, TNF-α, and interleukins. The IFN-γ molecules attract and activate more macrophages to enhance their phagocytic and antigen-presenting capabilities (Andersen et al., 2000). The TNF-α molecules play are significant role in induration formation by causing vasodilation and increased vascular permeability. This allows more immune cells, proteins, water, and red blood cells to migrate through vascular walls to the injection site. The accumulation of fluid (oedema), macrophages, other immune cells and molecules at the site, over the span of 48 to 72 hours post-injection, results in the formation of a firm and palpable induration (Arango Duque & Descoteaux, 2014). Induration diameter measurement, via the Mantoux method, is a manual approach to quantifying the volume of immune cells, molecules, and oedema at the TST site. Consequently, the volume of cells and molecules is regarded as proxy for the intensity of immune reaction to PPD. An induration diameter greater than or equal to consensus threshold is interpreted as proxy for significant immune response. The immune response to PPD indicates previous or current *Mtb* infection (Arango Duque & Descoteaux, 2014; Vukmanovic-Stejic et al., 2006).

The Mantoux method of reading TST indurations and estimating the intensity of immune reaction to *Mtb* antigens is susceptible to inter- and intra- observer variations in readings (Goel et al., 2018; Moayedi-Nia et al., 2019). Intra-observer variations in the readings of the same clinician can reach 3.0 mm, while inter-observer variation in readings can be as high as 3.7 mm (Pouchot et al., 1997). This subjectivity of the Mantoux method in the quantification of immune response leads to low precision in the diagnosis of LTBI compared to biopsy analysis. The invasiveness of biopsy makes its large-scale implementation for LTBI screening impractical. This has resulted in the persistence of the Mantoux method as a crude and imprecise but non-invasive and easily scalable way of reading indurations. The lack of digitization in the Mantoux method also prevents repeat analysis, validation, and automation. Thus, there is a need for a non-invasive method for quantifying the *Mtb*-based immune reaction embedded in TST indurations that is digital, standard, automatable, precise and reproducible.

2.2 Medical Optical Imaging and Thermography

Medical optical imaging (MOI) has the potential to address the shortcomings of the Mantoux method. It involves the utilization of light in the ultraviolet, visible and the near infrared (NIR) regions of the electromagnetic spectrum to acquire images of organs, tissues, cells, and molecules (Balas, 2009; Li & Wei, 2020). The visible, ultraviolet, and NIR bands are non-ionizing and exhibit a relatively stochastic behaviour in photon interaction and propagation in a heterogenous multi-layered biological tissue medium. Optical photons, unlike ionizing gamma rays and X-ray photons, do not penetrate deep into body tissue. However, they are useful for the capturing soft tissue properties.

The sub modalities of MOI include RGB imaging and HSI. In RGB imaging, the interaction between photons of visible light and body tissue are captured in three broad wavelengths of visible light, such that the wavelength bands 380 – 495 nm, 495 – 570 nm, and 620 – 750 nm are captured as the blue, green, and red bands, respectively (Bonnans et al., 2020; Mehraj et al., 2014). Thus, each pixel in an RGB image is composed of three intensities – a red, a green, and a blue band intensity value. On the other hand, in HSI the interaction between photons of light with body tissue are captured in at least 20 equally distributed contiguous wavelength bands anywhere from 200 nm to 25,000 nm (Lu & Fei, 2014).

2.2.1 Current Imaging Approaches for TST Indurations

Most smartphone cameras capture RGB images. Researchers have explored the use of smartphone-based RGB imaging to measure TST indurations and minimise the subjectivity of the Mantoux reading method. This was motivated by the relative ease of accessibility of smartphones today. One approach involves the use of structure-from-motion technique to capture RGB images of an induration at seven different angular orientations over a coverage of 120° along one angular axis (Dendere et al., 2017;

Naraghi et al., 2018). This facilitates the generation of a photogrammetric 3D representation of TST indurations. This approach was motivated by the hypothesis that the generation of RGB 3D reconstruction of a TST induration would facilitate a precise and repeatable measurement of induration diameters. The approach was proposed as an mHealth solution which is a subset of eHealth that leverages capabilities of mobile technologies to improve various aspects of healthcare (Twilt, 2023). The smartphone-based approach had a telemedicine configuration to transmit images from the frontend of an image-guidance app to a remote server hosting the photogrammetric model.

In the pilot study of Dendere et al. (2017), mock indurations of diameters ranging from 4 to 19 mm were cosmetically created on the forearm skin of 10 participants. Photogrammetric 3D reconstructions of the mock indurations were generated. Two clinicians were tasked with measuring the diameter of the mock indurations by manually placing landmarks on the induration boundaries in the 3D image. The clinicians also physically measured the mock indurations of participants using the Mantoux pen and ruler method. An intraclass correlation coefficient of 0.965 was obtained between the Mantoux diameter readings and 3D image-based readings. This showcased high level of agreement between the Mantoux reading method and the 3D RGB imaging method for mock indurations. The limitation of the manual placement of landmarks in the 3D images to obtain diameter estimations was addressed in Naraghi et al. (2018) where a sticker with a bounding box and fiducial markers was placed around the mock indurations of participants.

The acquisition of RGB images of mock indurations was done in Naraghi et al. (2018) using a custom-built smartphone application software. The sticker facilitated easier 3D reconstruction and identification of a rectangular region of interest around the mock indurations. To facilitate automated estimation of mock induration diameters, depth maps were generated from the 3D images and the indurations were segmented from the depth maps. The segmentation was achieved via the pipeline of image processing operations such as histogram equalisation, modified Otsu segmentation (Otsu, 1975), and elliptical estimation of induration boundaries. The diameter of the segmentation masks was regarded as the estimated induration diameter. The mean difference between the estimated readings and corresponding Mantoux readings was 1.1 mm with a standard deviation of 1.0 mm. This showcased the comparability of the approach to the Mantoux method for mock indurations. The application of this RGB 3D approach on real TST indurations, after the refinement of the custom-built application software in Farao et al. (2020), was investigated in Maclean (2020). The approach did not yield consistent segmentations or readings when applied on real indurations.

Recent studies by Parihar et al. (2021), Zang et al. (2024), and Akinola et al. (2024) explored a variety of approaches using RGB images of real indurations. Parihar et al. (2021) investigated the effectiveness

of four traditional thresholding algorithms – Otsu, region-growing, watershed, and region watershed – for the segmentation of real TST indurations of 26 participants. They found that watershed produced the lowest percentage error in area of 23.34%, with respect to the ground truth. Zang et al. (2024) introduced a new thresholding method to perform segmentation of real indurations for 66 participants. They achieved a mean DICE score of 0.845 across all participants with respect to the ground truth. Akinola et al. (2024) utilized a U-Net deep learning model to perform segmentation on a combination of 56 real and mock induration RGB images. They achieved an intersection-over-union score of 0.956 with respect to the ground truth. The ground truth utilized in these three studies was generated by manually annotating the RGB images of indurations. The assumption here that annotations on RGB images serve as a reliable proxy for real indurations is debatable, as clinicians traditionally rely on palpations or tactile cues to identify subdermal induration boundaries. This calls into question the clinical relevance of their approaches. There is currently no spatial ground-truth for induration segmentation. A spatial ground-truth for indurations will need to capture the boundaries of an induration in a repeatable non-subjective way. Manual annotation of RGB images introduces subjectivity as it is based on what regions of an induration a clinical can see in the images which are 3-channel abstractions of the physical induration. Hyperspectral imaging, with its subdermal imaging capability demonstrated in other medical applications, has greater potential as a spatial ground truth for indurations than manually annotated RGB images. Thus, there is a need for imaging approach, such as HSI, for real TST indurations that provide more clinically relevant readings.

The use of medical thermal imaging for estimating the real TST induration of 34 participants was investigated by Fiz et al. (2015) – currently the only study of its kind. This modality, also known as thermography, involves the use of long-wave infrared wavelength range (8µm to 14µm) of the electromagnetic spectrum to capture the heat radiation of body tissue (Shaikh et al., 2019). Fiz et al. (2015) performed thresholding on thermal images of indurations to segment perceived induration from surrounding tissue. The induration segmentation was not validated against image-based annotations, as it was in the preceding RGB studies. Instead, the diameter of the segmentation masks for 20 of the participants were compared to corresponding Mantoux readings. A Pearson's correlation score of 0.65 (with p-value<0.001) was obtained between the two sets of readings. The thermography-based diameter estimations have clinical relevance as they were validated against traditional clinician readings. In addition to image-based estimations, the study also identified temperature as a biomarker for characterising TST indurations – an insight that the Mantoux method does not provide. This is evidence of the effectiveness of thermography in the mitigation of the subjectivity of the TST. The long-wave infrared wavelength range is significantly absorbed by water and lipids in the skin. This leads to heating effect in skin tissue which is captured by a thermal camera. On the other hand, the

VNIR range of the electromagnetic spectrum is absorbed by a wider variety of molecules in the skin, including water and lipids. This gives the VNIR range a higher potential to generate more biomarkers of TST indurations in addition to facilitating clinically relevant diameter estimations. Thus, there is a need to investigate the ability of the VNIR range to characterise real TST indurations.

2.2.2 Hyperspectral Imaging

In Hyperspectral Imaging, a single pixel contains spectral information on the intensities of the various reflected wavelengths of light for the spatial point represented by the pixel. The plot of this spectral information is called a spectral signature (Aboughaleb et al., 2020). The spatial information of a scene captured in HSI is represented in two dimensions (also called the spatial axes) like a gray scale smartphone image. The spectral signature of each pixel location is layered on top of each point in the spatial dimension to form a third dimension – the spectral dimension (Holmer et al., 2018); resulting in a multidimensional data representation called the hypercube with the axes (x, y, λ). The HSI modality is a hybrid of optical spectroscopy, which involves the use of optical spectrometers to generate the spectral signature of a single point on a specimen, and digital imaging, which emphasises spatial information (Vasefi et al., 2016).

The hardware of a hyperspectral camera is typically composed of a front lens, an entrance slit, a collimator, a dispersive element, a focusing lens, and 2D detector arrays (Lu & Fei, 2014). The front lens focuses light from a specimen onto the entrance slit which allows only a line of light to pass to the collimator. The collimator transforms the diverging light from the entrance slit to rays of light which are separated into a series of narrow spectral bands by the dispersive device (Yang et al., 2021). The focusing lens focuses the decomposed spectral bands of light onto the 2D detector array. Hyperspectral cameras can be classified based on the image acquisition mode, detector array type, spectral range and resolution, measurement mode, and dispersive device type. The acquisition mode defines how the spatial and spectral information is captured from a scene. Conventional hyperspectral cameras utilize either spatial scanning or spectral scanning. In spatial scanning, a hypercube is generated by acquiring the complete spectrum for a single pixel or a line of pixel before moving to the next pixel or line of pixel (Vasefi et al., 2016).

Point-scanning HSI cameras that perform pixel-by-pixel scanning are also called whiskbroom cameras, and line-scanning ones are called pushbroom cameras. The whiskbroom and pushbroom cameras develop hypercubes by spatially scanning through a scene by moving the camera or the observed specimen (Paulus & Mahlein, 2020). In spectral scanning, which is also called staring or area-scanning, an entire scene is captured on the 2D detector array in a single exposure at a selected wavelength band. Multiple 2D images are then captured while stepping through multiple wavelength bands to

generate a complete hypercube (Li et al., 2023). Both spatial and spectral scanning methods have their merits. The spatial scanning method offers higher accuracy of spectral information but compromises spatial information. With this method, the spectral signatures of pixels (as in whiskbroom) or lines of pixels (as in pushbroom) are stitched together. Thus, there could be misalignments of pixels due to motion of the camera or the target object. On the other hand, in spectral scanning, spatial information is more accurate as an array of pixel positions are captured in single exposures. However, some spectral information could be lost during the transitions between wavelength exposures. This scanning method also allows for live preview of scene during capturing. Conversely, in spatial scanning, images are typically viewed after the entire capture process (Lu & Fei, 2014).

Hyperspectral cameras can also be classified based on their spectral range and resolution. The spectral range defines the span of wavelengths an HSI camera can capture. The commonly defined wavelength ranges are ultraviolet (UV) (100 to 400 nm), visible (VIS) (400 to 750nm), NIR (750 to 2500nm), and mid-infrared (MIR) (2500 to 10000nm) (Agiomavriti et al., 2024). The wavelength ranges an HSI camera would be sensitive to depends on the type of its detector focal plane array (FPA). The detector FPA is composed of 2D semiconductor arrays that convert radiation energy of dispersed light into electrical signals. Charge-coupled devices (CCDs) are the most widely used detector arrays due to their low dark current and high quantum yield. Silicon-based CCDs are sensitive to light in the VIS and early NIR range, while indium gallium arsenide (InGaAs) CCD detectors are sensitive to 900 to 1700 nm range (Vollmer et al., 2015). Mercury cadmium telluride (HgCdTe) detectors are sensitive to the NIR and MIR wavelengths (Huck, 2014). In dermatological HSI studies, VIS and NIR are the most used spectral ranges. The hybrid VNIR range (400 to 1000 nm) has been created to enable the development of hypercubes with VIS light and the frontend of the NIR (Lu & Fei, 2014). Thus, the HSI cameras used in dermatology typically have silicon-based detector FPAs.

The spectral resolution of an HSI camera defines the smallest interval between two distinguishable points in its spectral signature. The higher the spectral resolution of an HSI camera, the closer the captured spectral signature is to the true spectral profile of the observed specimen (Vasefi et al., 2016). The number of spectral bands determines the number of stacked greyscale images in the developed hypercube. Thus, the higher the number of spectral bands, the larger the size of the hypercube in the spectral axis. The measurement mode of an HSI camera defines what optical information is captured based on the interaction of light with biological tissue. This includes reflectance, fluorescence, and transmission (Lu & Fei, 2014). Reflectance is the preferred mode in dermatology as it allows *in vivo* non-invasive tissue examination by simply capturing light reflected by the skin. Conversely, in transmission and fluorescence, the visual information of interest is how light is transmitted through and within biological tissues and cells, typically involving *in vitro* analysis (Lu &

Fei, 2014). The Specim IQ (specim.com/iq) is an example of a VNIR HSI camera that has a high number of spectral bands (204) and operates in reflectance mode, which makes it a viable choice for dermatology applications such as in the LTBI screening.

2.3 Physiology and Optical Characteristics of the Human Skin

The human skin is composed of the epidermis (the outermost layer of the skin), the dermis (middle layer), and the hypodermis (deepest layer). The thickness of the human epidermis is typically between 77 and 267 µm (Oltulu et al., 2018). The different layers of the skin possess different chromophores, i.e., molecules that absorb light at specific wavelengths to give the appearance of colour. Light is absorbed in the epidermis by chromophores such as eumelanin, pheomelanin, lycopene, and beta-carotene (Darvin et al., 2011; Vasefi et al., 2016). Eumelanin is a pigment that has the appearance of dark brown to black colouration. Pheomelanin is a reddish-yellow pigment. Eumelanin offers protection to skin from the damaging effects of UV radiation. Pheomelanin offers protection to a lesser degree. Lycopene and beta-carotene belong to the family of carotenoids – a type of antioxidant (Darvin et al., 2011).

The thickness of the dermis typically varies from 2.1 mm to 5.9 mm and is composed of connective tissues, blood vessels, and nerves (Oltulu et al., 2018; Weber et al., 2006). The blood concentration in the dermis is relatively low between 0.2% and 7% (Vasefi et al., 2016). The chromophores in the dermis include oxyhaemoglobin (HbO_2), deoxyhaemoglobin (Hb), fat, water, collagen, elastin, and bilirubin (Vasefi et al., 2016; Villaret et al., 2019). The haemoglobin (both HbO_2 and Hb) content in blood tends to absorb light in the blue and green wavelengths. Thus, it reflects more of red wavelengths which gives blood its reddish colour. The function of haemoglobin is to transport oxygen. The HbO_2 molecule is haemoglobin with oxygen molecules bound to it and is responsible for the bright red appearance of oxygenated arterial blood. Conversely, Hb is haemoglobin that has released is oxygen molecules. It is responsible for the dark red appearance of deoxygenated venous blood (Alkawaz et al., 2015). Water and fat have relatively low absorption in the VIS range and higher absorption in the NIR range. Collagen and elastin also have higher absorption in the NIR range (Konugolu Venkata Sekar, Bargigia, et al., 2017). Collagen is fibrous protein in the dermis that provides strength and structural support to the skin. Conversely, elastin is a flexible fibrous protein that provides elasticity to the skin (Baumann et al., 2021). Bilirubin is a by-product of the breakdown of red blood cells. It is typically generated in the liver. When bilirubin levels are high, it circulates in the bloodstream and diffuses into the skin (Singla et al., 2023).

The thickness of the hypodermis (also called subcutaneous tissue) typically varies from 1.9 mm to 7.1 mm (Hwang et al., 2016). It is primarily composed of fat and connective tissue. It is not as rich in

chromophores as the dermis or epidermis. The main chromophores in this layer include fat, water, HbO_2 and Hb.

The penetration of VNIR light through the skin increases with wavelength. As light travels farther, the various constituents of the skin absorb some of the photons various wavelengths, transmits some, and scatters the rest (Lu & Fei, 2014). The intensity of the various reflected wavelengths of light depends on how much of each wavelength is absorbed. This is the visual information captured by an HSI camera. The high eumelanin content of darker skin tones makes it absorb more and reflect less light below 600 nm (Vasefi et al., 2016). This means a camera would capture less visual information from darker skin than lighter skin. Normal RGB imaging suffers more from this problem due to approximating the intensities of the several wavelengths of reflected visible light by only three intensities. This low precision of captured light spectra makes it difficult to observe subtle differences in the intensities of reflected wavelengths. Conversely, HSI cameras capture the intensities of hundreds of reflected wavelengths – making it possible to observe subtle differences in intensities. Biochemical and morphological changes due to skin lesions modify the absorption, scattering, and fluorescence properties of skin tissue when they interact with photons of light. Acquisition and observation of the changes in these optical characteristics provide valuable diagnostic information (Lu et al., 2020). The unique ability of HSI to obtain optical information at hundreds of wavelengths has motivated several *in vivo* studies in dermatology.

2.4 Application of Hyperspectral Imaging and Artificial Intelligence in Dermatology

2.4.1 Hyperspectral Imaging in Dermatology

Skin cancer research is among the most common application of HSI in dermatology. In addition to this, HSI has been employed in the assessment of wound extent and rate of healing. Skin wound research that has benefited from HSI include chronic wound (Daeschlein et al., 2017), burns (Nunez et al., 2022), bruises (Randeberg & Hernandez-Palacios, 2012), diabetic foot ulcers (Kounas et al., 2023), peripheral vascular disease (PVD) (Chiang et al., 2017), erythema (Abdlaty, Doerwald-Munoz, et al., 2021), and flap transplantation (Kohler et al., 2021). Skin infection research has also benefited from the use of HSI, for example eczema (Hald et al., 2019), wound infection (Ramirez-GarciaLuna et al., 2023), cellulitis (Chen et al., 2018), cutaneous leishmaniasis (Botina-Monsalve et al., 2020), and shingles (herpes zoster) (Stamatas & Kollias, 2006).

2.4.1 Artificial Intelligence in Hyperspectral Imaging

Recent studies also demonstrate the effectiveness of HSI in conjunction with AI in dermatology. One of the most common diagnosis guideline in dermatology is the ABCDE rule (Nachbar et al., 1994),

which emphasizes the asymmetry, border, colour, differential structures, and evolution of skin lesions (Abbasi et al., 2004). In skin cancer research a score is assigned for each of the five features, obtained by visual inspection, and the total score determines the level of potential malignancy (Johansen et al., 2020). These features are not as pronounced in TST indurations, which contributes to the subjectiveness of the Mantoux reading method. The use of AI and feature extraction models in conjunction with hyperspectral images can facilitate the extraction of more lesion features than the visually observable ones (Johansen et al., 2020).

The ML algorithms commonly used on HSI studies in dermatology include singular value decomposition (SVD), as seen in the skin cancer study by Pardo et al. (2018), and principal component analysis (PCA), as seen in Rey-Barroso et al. (2018). Chen et al. (2020) implemented unsupervised ML algorithms such as k-means, Gaussian mixture model (GMM), and hierarchical clustering. The unsupervised algorithms were used to reduce the dimensionality of the hypercubes before applying supervised learning algorithms like support vector machines (SVM). Kazianka et al. (2008) developed a classification model based on SVM and model-based discriminant analysis (MDA) to classify skin cancer regions in 300 hypercubes. Torti et al. (2020) also used a SVM for the classification of skin cancer images. Zheludev et al. (2015) developed a HSI-based pipeline for malignant skin tumor delineation using diffusion maps dimensionality reduction. In that study, the average nearest neighbor classifier, and classification and regression trees (CART) were used for malignant/healthy pixel classification.

Data augmentation techniques have been employed to improve the performance of ML models by generating derivative subsets of original HSI datasets. Kato et al. (2020) employed data augmentation (horizontal and vertical inversion) on 619 hypercubes of skin lesions to improve the sensitivity and specificity of their deep learning classification model from 77.4% and 75.6% to 79.9% and 82.4%, respectively. Similarly, Lihacova et al. (2022) perfomed random horizontal/vertical flipping, and random $\pm 90°$ rotation to generate 19,560 augmented spectral images from 1,304 original hyperspectral images of skin lesions for a deep learning classification model. Though ML techniques have been used to automate diagnosis in HSI skin cancer studies, the characterisation of TST indurations using HSI has not been reported.

2.5 The Imaging Protocol of the Specim IQ Hyperspectral Camera

The Specim IQ (specim.com/iq) HSI camera was the device used in this research to capture hyperspectral imaging. Its selection was motivated by the portability of the device, and the spectral range of 400 nm to 1000 nm (Behmann et al., 2018). The default imaging protocol for the acquisition of images using the camera is as illustrated in Figure 2.1. The protocol comprises five steps. Step 1

involves the positioning the target object at least 150 mm in front the camera lens and defining the primary illumination type for the imaging scene. Halogen or sunlight are the illumination types recommended by Specim. The viewfinder of the camera can be used to verify the field of view of the imaging scene.

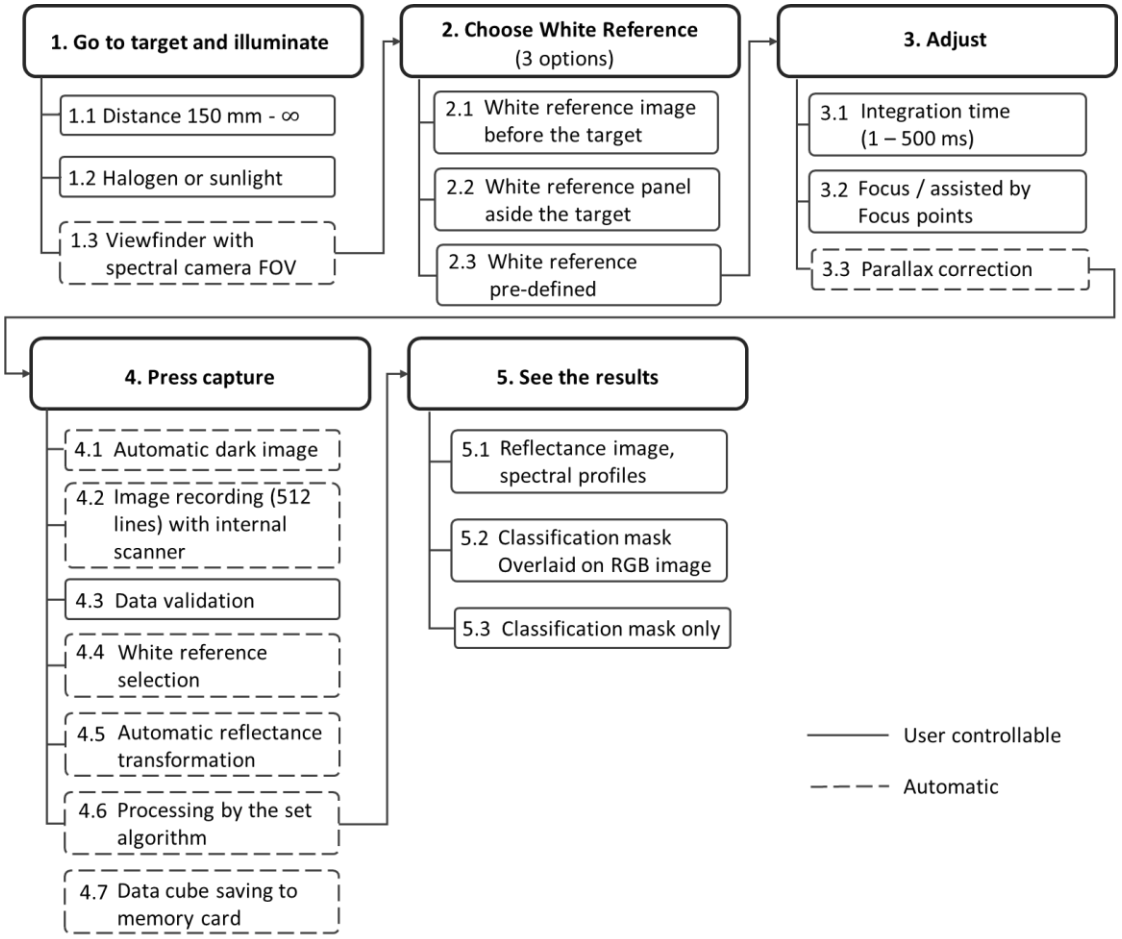


Figure 2.1: Steps 1–5 of the imaging protocol for spectral signature acquisition using the Specim IQ HSI camera (Behmann et al., 2018).

Protocol Step 2 describes the choice and placement of a white reference in the imaging scene before the capturing of a hypercube. A white reference panel, placed in the imaging scene, reflects the incident illumination to the Specim IQ camera. The white reference should be made of Spectralon, which is currently the most widely used standard reflectance surface in HSI. It is a fluoropolymer which has the highest diffuse reflectance of any known material in the VNIR region (Lu & Fei, 2014). With a diffuse reflectance of 99%, the Spectralon white reference panel can reflect 99% of the spectrum of the incident illumination to the camera. This enables the camera to capture the illumination condition of the imaging scene which in turn facilitates the normalisation of captured spectral signatures.

Protocol item 2.1 describes the placement of the white reference panel in the imaging scene to capture illumination condition of the scene before the introduction of the target object for the

acquisition of spectral signatures. This approach introduces the risk of temporal illumination condition changes between when the white reference panel is removed from the scene and when the target object is introduced. Item 2.2 is an alternative to 2.1 and describes the placement of the white reference panel in close proximity to the target for simultaneous acquisition of the ambient illumination condition and the target spectral signatures. This mitigates the risk of error in spectral signatures due to temporal changes in illumination conditions. Item 2.3 outlines an alternative to Items 2.1 and 2.2, proposing the use of a pre-defined illumination condition. However, this condition may not accurately reflect the actual illumination present in the current imaging scene.

Step 3 describes a set of task items that follows after the placement of the white reference panel in the imaging scene. Item 3.1 describes the definition of the integration time to be used by the Specim IQ camera in the acquisition of each line of spectral signatures. The Specim camera uses a line scan acquisition mode where a line of spectral signatures is captured from the imaging scene per sensor exposure. The time duration to capture a line of spectral signatures is the integration time of the camera. The range of allowable integration time for the Specim camera is 1 ms to 500 ms. The Specim IQ captures 512 lines of spectral signatures per imaging scene. Thus, the range acquisition time for a single hyperspectral image is (1 to 500) ms x 512 = (512 ms to 4.3 minutes) (Behmann et al., 2018). Selecting an integration time that is too short may result in an underexposed hyperspectral image as the sensor aperture is not open long enough for sufficient photons of light from the scene to be captured. On the other hand, selecting an integration time that is too long may result in an overexposed hyperspectral image as the camera sensor becomes saturated from gathering too many photons of light from the scene. The viewfinder of the Specim IQ device assesses the average illumination intensity of the current scene and calculates the optimal integration time required to produce a well-exposed hyperspectral image. The user is given the option to choose the estimated optimal integration time or to adjust it to their preference.

Item 3.2 describes the adjustment of the focus lens of the hyperspectral camera by the user to ensure that the target is in-focus. Focus points (fiducial markers) can be placed in the imaging scene to facilitate the determination of optimal focus. Due to the physical distance between the RGB viewfinder and the HSI lens, there is a parallax error between the viewfinder feed and the scene coverage to be captured through the HSI lens. The Specim device attempts to automatically align these two feeds, via image registration, before the acquisition of a hyperspectral image. The user is provided a visualisation of the alignment with the option to modify the alignment for better parallax correction.

Step 4 describes the acquisition of a hyperspectral image from an imaging scene after illumination, white reference, lens focus, and parallax correction tasks have been completed. The inherent sensor

noise of the Specim HSI camera is utilised in the normalisation of acquired spectral signatures. The sensor noise is estimated via the acquisition of a dark reference image. The Specim device acquires this by automatically capturing an image with the HSI camera shutter closed. Item 4.2 describes the line-by-line recording of the 512 lines of spectral signatures that makeup a hypercube from the camera. Item 4.3 describes the user-based data validation of the acquired raw hypercube. The user is presented with the raw hypercube to visualise if there are regions of the hypercube that are underexposed or overexposed. The user can discard an ill-exposed hypercube at this stage. If item 2.2 was utilised during the white reference task block, then the device, in item 3.4, automatically extracts regions of the raw hypercube that captured the white reference panel in the imaging scene. The camera transforms the reflectance data in item 3.5, and normalises the acquired hypercube in item 3.6, using the white reference and dark reference, before saving the normalised hypercube to memory in item 3.7. Step 5 describes the presentation of the acquired reflectance spectral signatures to the user in item 5.1, and downstream tasks like classification of objects in the imaging scene, as described in items 5.2 and 5.3.

2.6 Radiomics for Feature Extraction

With the inception of human genome mapping and sequencing, advancements in technology have enabled the acquisition of a vast array of molecular measurements within tissues and cells. Scientific disciplines related to high-throughput measurement of biological molecules are collectively referred to as "omics". Examples of omics domains include proteomics, transcriptomics, genomics, epigenomics, metabolomics, and lipidomics (Micheel et al., 2012). These correspond to the analyses of proteins, ribonucleic acids (RNA), genes, methylated deoxyribonucleic acids (DNA), metabolites, and lipids, respectively (Micheel et al., 2012).

In recent times, radiological imaging has experienced a rise in usage for medical applications. This has led to the development of a specialized omics category known as radiomics. Radiomics involves the extraction and analysis of quantitative data from radiological images. It is predicated on the hypothesis that medical images capture underlying pathophysiologic characteristics. Hence, qualitative analysis of these images can provide useful insights that may elude human perception and are not accessible through traditional visual inspection of medical images (van Timmeren et al., 2020). Radiomics aims to enhance the existing data available to clinicians by means of advanced mathematical and data-driven analysis of medical images. It quantifies the spatial relationship between pixels of a medical image as large array features which can be used to describe tissue and lesion properties such as shape and heterogeneity (Mayerhoefer et al., 2020).

The level of discretization of each image in a set of medical images impacts the reproducibility of the radiomics features extracted from across all the images. This level can be defined in terms of three parameters: range of intensities, bin count or the number of bins (the number of discrete intensity levels), and bin width (the range of intensity values grouped into each discrete bin) [reference]. Variations in the values of these parameters among images in a dataset can introduce bias towards certain images when extracting features. Standardizing these parameters across all images removes this bias and promotes the reproducibility of the generated radiomics features (van Timmeren et al., 2020). The standardization process includes fixing the bin count, intensity range and bin width. Setting a fixed bin count is beneficial in data with arbitrary intensity units (like the pixel intensities of medical images) where contrast is important. Bin count and bin width are related. When a fixed intensity range is established, choosing a fixed bin width determines the resulting bin count, and vice versa. Choosing a bin width that is too wide can result in the averaging-out of spatial features. Conversely, when the bins are two narrow, features can become indistinguishable from noise. The optimal bin width for each dataset is typically estimated empirically (van Timmeren et al., 2020).

The discretization parameters of a medical image as well as the generation radiomics features from the images can be achieved using an open-source library such as Pyradiomics which was developed by van Griethuysen et al. (2017). It is one of the most comprehensive open-source libraries for extracting radiomics features from medical images. The library contains a comprehensive set of functions that can be applied to an image to generate radiomics features. The features can be termed deterministic or handcrafted radiomics features as they are generated using predefined functions. Pyradiomics comprises of six classes of intensity-based radiomics features. These include the first order statistics-based, the gray level co-occurrence matrix (GLCM)-, the gray level run length matrix (GLRLM)-, the gray level size zone matrix (GLSZM)-, the neighbouring gray tone difference matrix (NGTDM)-, and the neighbouring gray level dependence matrix (NGLDM)-based features (Zwanenburg et al., 2016).

The first order statistics-based functions used to generate intensity-based statistical features provide basic descriptions of the distribution of pixel intensities in an image. The histogram of a given image $\boldsymbol{C}$, with number of pixels N_p, can be denoted as P^{fo}. The normalised histogram can be denoted as $p^{fo} = P^{fo}/N_p$ (Zwanenburg et al., 2016). The number of gray levels can be denoted as N_g. First order statistics-based features defined using these variables are as shown in Table 2.1 (Zwanenburg et al., 2016). The energy function is used to emphasize pixels with high intensities. Applying the area of a pixel in an image as a scaling factor to the energy feature, results in the generation of the total energy feature. Entropy quantifies the uncertainty or randomness of the intensities in the image. The mean

absolute deviation quantifies the mean distance of all intensities from the mean value of $\boldsymbol{C}$. The robust mean absolute deviation quantifies the mean distance of all intensities from the mean of $\boldsymbol{C}$ calculated on a subset of intensities within the 10th and 90th percentile of $\boldsymbol{C}$. The root mean squared feature quantifies the average value of the intensities of $\boldsymbol{C}$ while being more sensitive to higher intensities. The standard deviation feature measures the amount of variation of the intensities from the mean intensity (Wu et al., 2019). Skewness quantifies the asymmetry of the distribution of pixel intensities about the mean intensity. Furthermore, the kurtosis feature quantifies the concentration of intensity distribution with respect to the mean. A high kurtosis implies that intensity values are concentrated towards the tails of the histogram rather than the mean. Variance measures the spread of intensities about the mean of $\boldsymbol{C}$. The uniformity feature quantifies the homogeneity of the pixel intensities (van Timmeren et al., 2020; Zwanenburg et al., 2016).

Table 2.1: First Order Deterministic Radiomics Feature (Zwanenburg et al., 2016)

$energy = \sum_{i=1}^{N_p} C(i)^2$ (2.1)	$total\ energy = pixel\ area \sum_{i=1}^{N_p} C(i)^2$ (2.2)
$entropy = -\sum_{i=1}^{N_g} p^{fo}(i) log_2(p^{fo}(i) + 2.2 \times 10^{-6})$ (2.3)	
$minimum = \min(C)$ (2.4)	$10^{th}\ percentile\ of\ intensities\ in\ C$ (2.5)
$90^{th}\ percentile\ of\ intensities\ in\ C$ (2.6)	$maximum = \max(C)$ (2.7)
$mean = \frac{1}{N_p} \sum_{i=1}^{N_p} C(i)$ (2.8)	$median\ intensity\ in\ C$ (2.9)
$interquartile\ range = difference\ between\ 75^{th}\ and\ 25^{th}\ percentile\ in\ C$ (2.10)	
$range = \max(C) - \min(C)$ (2.11)	$root\ mean\ squared\ (RMS) = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} C(i)^2}$ (2.12)
$mean\ absolute\ deviation\ (MAD) = \frac{1}{N_p} \sum_{i=1}^{N_p} \lvert C(i) - \bar{C} \rvert$ (2.13)	
$robust\ mean\ absolute\ deviation\ (rMAD) = \frac{1}{N_{10-90}} \sum_{i=1}^{N_{10-90}} \lvert C_{10-90}(i) - \bar{C}_{10-90} \rvert$ (2.14)	
$standard\ deviation = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (C(i) - \bar{C})^2}$ (2.15)	$variance = \frac{1}{N_p} \sum_{i=1}^{N_p} (C(i) - \bar{C})^2$ (2.16)
$skewness = \frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (C(i) - \bar{C})^3}{\left(\sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (C(i) - \bar{C})^2} \right)^3}$ (2.17)	$kurtosis = \frac{\frac{1}{N_p} \sum_{i=1}^{N_p} (C(i) - \bar{C})^4}{\left(\frac{1}{N_p} \sum_{i=1}^{N_p} (C(i) - \bar{C})^2 \right)^2}$ (2.18)
$variance = \frac{1}{N_p} \sum_{i=1}^{N_p} (C(i) - \bar{C})^2$ (2.19)	$uniformity = \sum_{i=1}^{N_g} p^{fo}(i)^2$ (2.20)

The GLCM-based features are generated from GLCM matrix $\boldsymbol{P^{cm}}$ of an image $\boldsymbol{C}$. The $(j,k)^{th}$ element in $\boldsymbol{P^{cm}}$ corresponds to the number of times the combination of gray levels j and k occur in two neighbouring pixels in image $\boldsymbol{C}$ (Nanni et al., 2013). The generation of matrix $\boldsymbol{P^{cm}}$ from a given image $\boldsymbol{C}$ for a neighbourhood distance of one pixel in the horizontal direction is as illustrated in Figure 2.2. As shown in the figure, the number of occurrences of a pair of neighbouring pixel intensities in image $\boldsymbol{C}$, in the horizontal direction, is captured in the matrix $\boldsymbol{P^{cm}}$ (Nanni et al., 2013).

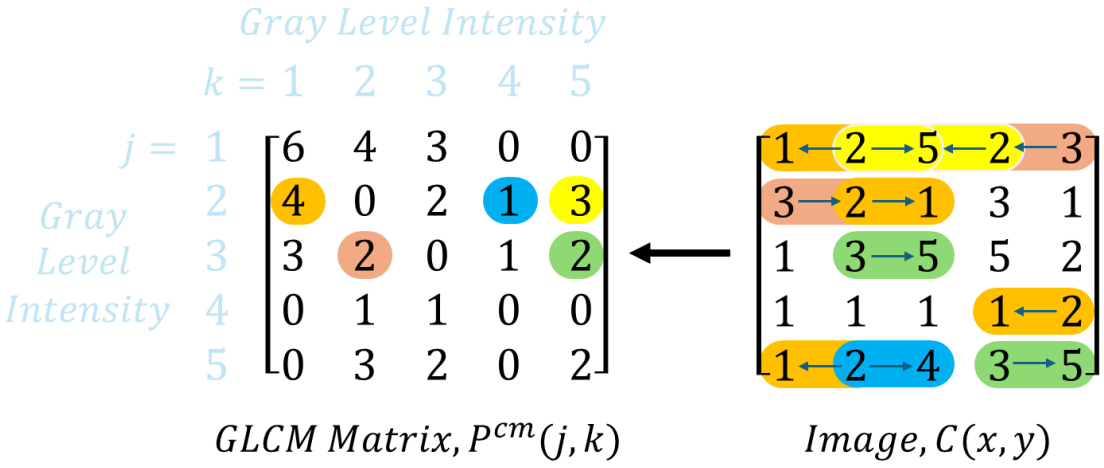


Figure 2.2: Illustration of the generation of the GLCM matrix from a given image.

The variables of $\boldsymbol{P^{cm}}$ that facilitate the generation of GLCM-based features are as outlined in Table 2.2. The resulting GLCM-based functions are outlined in Table 2.3. A measure of the fineness and coarseness of the texture of image $\boldsymbol{C}$ is captured by the autocorrelation feature. Joint average represents the average value from the distribution of j gray levels. The cluster prominence feature captures the skewness and asymmetry of $\boldsymbol{P^{cm}}$. Conversely, the cluster shade measures the skewness and uniformity of $\boldsymbol{P^{cm}}$. Cluster tendency is a measure of groupings of pixels with similar gray level intensities. The contrast feature captures local intensity variation, favouring intensities away from the diagonal direction (Nanni et al., 2013). The linear dependency of gray level values to their respective pixels in $\boldsymbol{P^{cm}}$ is captured by the correlation variable. Difference average is a measure of the relationship between the occurrences of similar intensity pairs and the occurrences of dissimilar intensity pairs. On the other hand, difference variance captures heterogeneity but places higher weights on differing intensity pairs that deviate more from the mean. The joint energy feature captures the frequency of the occurrence of intensity pairs that are neighbours in an image. The randomness of a neighbourhood of intensity values is captured by the joint entropy feature (Nanni et al., 2013). The information measure of correlation (IMC) 1 and IMC2 features quantify the correlation between the probability distributions of j and k. The inverse difference moment (IDM) quantifies the local homogeneity of the image $\boldsymbol{C}$. The IDMN, which is the normalised form of the IDM, captures the normalises the square of the difference between neighbouring intensity values. The inverse difference

(ID) is another feature that captures a measure of local homogeneity of image $\boldsymbol{C}$. The IDN, which is the normalised form of the ID, normalises the difference between neighbouring intensity values. The inverse variance is a measure of the variance of $\boldsymbol{P^{cm}}$. The rate of occurrence of the most predominant pair of neighbouring intensity values is quantified by the maximum probability feature. On the other hand, the sum average feature quantifies the relationship between occurrences of low intensity pairs and high intensity pairs. The sum of neighbourhood intensity value differences is captured by the sum entropy feature. The sum of squares is a measure of the distribution of neighbouring intensity pairs that are about the mean intensity level in $\boldsymbol{P^{cm}}$ (Nanni et al., 2013; Zwanenburg et al., 2016). The formulation of the other radiomics features is described in appendix C.

Table 2.2: Outline of GLCM Variables (Zwanenburg et al., 2016)

$normalised\ GLCM, p^{cm} = \frac{P^{cm}(j,k)}{\sum P^{cm}(j,k)}$ (2.21)	$marginal\ row\ probability,$ $p_x^{cm}(j) = \sum_{k=1}^{N_g} p^{cm}(j,k)$ (2.22)
$marginal\ column\ probability,$ $p_y^{cm}(k) = \sum_{j=1}^{N_g} p^{cm}(j,k)$ (2.23)	$mean\ gray\ level\ intensity\ of\ p_x^{cm},$ $\mu_x(k) = \sum_{j=1}^{N_g} p_x^{cm}(j) \cdot j$ (2.24)
$mean\ gray\ level\ intensity\ of\ p_y^{cm},$ $\mu_y = \sum_{k=1}^{N_g} p_y^{cm}(k) \cdot k$ (2.25)	$\sigma_x = standard\ deviation\ of\ \mu_x$ (2.26)
$\sigma_y = standard\ deviation\ of\ \mu_y$ (2.27)	$p_{x+y}^{cm}(l) = \sum_{j=1}^{N_g} \sum_{k=1}^{N_g} p^{cm}(j,k)$ (2.28) $, where\ l = j + k$
$p_{x-y}^{cm}(l) = \sum_{j=1}^{N_g} \sum_{k=1}^{N_g} p^{cm}(j,k)$ (2.29) $, where\ l = j - k$	$entropy\ of\ p_x^{cm},$ $HX = -\sum_{j=1}^{N_g} p_x^{cm}(j) \cdot \log_2(p_x^{cm}(j) + \epsilon)$ (2.30) $, where\ \epsilon \approx 2.2 \times 10^{-16}$
$entropy\ of\ p_y^{cm},$ $HY = -\sum_{k=1}^{N_g} p_y^{cm}(k) \cdot \log_2\left(p_y^{cm}(k) + \epsilon\right)$ (2.31)	
$entropy\ of\ p^{cm}(j,k),$ $HXY = -\sum_{j=1}^{N_g} \sum_{k=1}^{N_g} p^{cm}(j,k) \cdot \log_2(p^{cm}(j,k) + \epsilon)$ (2.32)	
$HXY1 = -\sum_{j=1}^{N_g} \sum_{k=1}^{N_g} p^{cm}(j,k) \cdot \log_2\left(p_x^{cm}(j) p_y^{cm}(k) + \epsilon\right)$ (2.33)	
$HXY2 = -\sum_{j=1}^{N_g} \sum_{k=1}^{N_g} p_x^{cm}(j) p_y^{cm}(k) \cdot \log_2\left(p_x^{cm}(j) p_y^{cm}(k) + \epsilon\right)$ (2.34)	

Table 2.3: GLCM-based Deterministic Radiomics Features (Zwanenburg et al., 2016)

$autocorrelation = \sum_{j=1}^{N_g} \sum_{k=1}^{N_g} p^{cm}(j,k) \cdot j \cdot k$ (2.35)	$joint\ average = \sum_{j=1}^{N_g} \sum_{k=1}^{N_g} p^{cm}(j,k) \cdot j$ (2.36)
$cluster\ prominence =$ $\sum_{j=1}^{N_g} \sum_{k=1}^{N_g} \left(j + k - \mu_x - \mu_y\right)^4 \cdot p^{cm}(j,k)$ (2.37)	$cluster\ shade =$ $\sum_{j=1}^{N_g} \sum_{k=1}^{N_g} \left(j + k - \mu_x - \mu_y\right)^3 \cdot p^{cm}(j,k)$ (2.38)
$cluster\ tendency =$	$contrast =$

$\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}\left(j+k-\mu_x-\mu_y\right)^2\cdot p^{cm}(j,k)$ (2.39)	$\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}(j-k)^2\cdot p^{cm}(j,k)$ (2.40)
$correlation=\frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}p^{cm}(j,k)\cdot j\cdot k-\mu_x\mu_y}{\sigma_x(j)\sigma_y(k)}$ (2.41)	
$difference\ average=\sum_{l=0}^{N_g-1}l\cdot p_{x-y}^{cm}(l)$ (2.42)	$difference\ entropy=\sum_{l=0}^{N_g-1}l\cdot p_{x-y}^{cm}(l)\log_2(p_{x-y}^{cm}(l)+\epsilon)$ (2.43)
$difference\ variance=\sum_{l=0}^{N_g-1}(l-difference\ average)^2\cdot p_{x-y}^{cm}(l)$ (2.44)	
$dissimilarity=\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}\lvert j-k\rvert\cdot p^{cm}(j,k)$ (2.45)	$joint\ energy=\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}(p^{cm}(j,k))^2$ (2.46)
$joint\ entropy=\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}p^{cm}(j,k)\log_2(p^{cm}(j,k)+\epsilon)$ (2.47)	$information\ measure\ of\ correlation\ (IMC)=\frac{HXY-HXY1}{\max(HX,HY)}$ (2.48)
$IMC2=\sqrt{1-e^{-2(HXY2-HXY)}}$ (2.49)	$inverse\ difference\ moment\ (IDM)=\sum_{l=0}^{N_g}\frac{p_{x-y}^{cm}(l)}{1+l^2}$ (2.50)
$maximal\ correlation\ coefficient\ (MCC)=\sum_{l=0}^{N_g}\frac{p^{cm}(j,l)\cdot p^{cm}(k,l)}{p_x^{cm}(j)\cdot p_y^{cm}(l)}$ (2.51)	$inverse\ difference\ moment\ normalized,\ (IDMN)=\sum_{l=0}^{N_g-1}\frac{p_{x-y}^{cm}(l)}{1+\left(\frac{l^2}{N_g^2}\right)}$ (2.52)
$inverse\ difference\ normalised\ (IDN)=\sum_{l=0}^{N_g-1}\frac{p_{x-y}^{cm}(l)}{1+\left(\frac{l}{N_g}\right)}$ (2.53)	$inverse\ variance=\sum_{l=1}^{N_g-1}\frac{p_{x-y}^{cm}(l)}{l^2}$ (2.54)
$maximum\ probability=\max(p(j,k))$ (2.55)	$sum\ average=\sum_{l=2}^{2N_g}p_{x+y}^{cm}(l)\cdot l$ (2.56)
$sum\ entropy=\sum_{l=2}^{2N_g}p_{x+y}^{cm}(l)\cdot\log_2\left(p_{x+y}^{cm}(l)+\epsilon\right)$ (2.57)	$sum\ of\ squares=\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}(i-\mu_x)^2\cdot p^{cm}(j,k)$ (2.58)

The ability for radiomics to extract spatial features from medical images was important in the identification of biomarkers that characterize TST indurations. The descriptions of each radiomics feature, presented in this chapter, aided in understanding which pixel regions or spatial attributes of the downstream hyperspectral image transformations effectively predict LTBI diagnosis.

2.7 Conclusion

This chapter highlights the current measures in the management and diagnosis of LTBI and the research gap to be addressed in the next chapters. The pathophysiology of TST indurations was reviewed to understand the cellular and biological processes that underpin the formation of

indurations. This knowledge was useful in the downstream validation of the TST biomarkers generated from HSI. Current research on the use medical optical imaging and thermal imaging for reading TST indurations was reviewed. This enabled the understanding of related works and their limitations. The studies also served as benchmark for the demonstration of the effectiveness of HSI for TST reading and characterisation. Furthermore, background knowledge on the physiology and optical characteristics of the human skin was reviewed to obtain understanding of relationship between spectral signatures and the constituents of the human skin. The background of HSI was presented as a justification for its selection as the preferred modality over RGB imaging and thermography, for addressing the identified research gap. Recent dermatological applications of HSI, used in conjunction with AI, were reviewed to provide justification that HSI is applicable for the TST, which is likewise a dermatological application. Furthermore, the application of HSI in other skin studies was reviewed to highlight its increasing popularity and the need for the generation of evidence of its applicability in TST characterisation. The imaging protocol of the HSI camera used in this research for the acquisition of spectral signatures of TST injection sites was also reviewed. Also presented was background information on radiomics and the different classes of mathematical functions which will facilitate the extraction of features for TST characterisation. The next chapter describes the development and validation of an HSI protocol, which is based on the reviewed Specim IQ protocol, for the reproducible acquisition of spectral signatures of TST injection sites.

CHAPTER 3

An Imaging Protocol for Reproducible Acquisition of Spectral Signatures of TST Injection Sites

---

Disclaimer: The contents of this chapter are partly published in the following paper:

I. Oladokun, A.S., Mutsvangwa, T., Malila, B. (2024). Understanding the Impact of Different Light Sources on the Reproducibility of Skin Spectral Signature for Different Skin Tones. In *2024 IST-Africa Conference (IST-Africa)* (pp. 1-12). IEEE. https://doi.org/10.23919/IST-Africa63983.2024.10569556

This chapter addresses the first research question of the thesis: *Can spectral signatures of TST injection sites be reliably and reproducibly captured using a hyperspectral camera?* To answer this, the chapter details the development and validation of an imaging protocol designed to enable reproducible acquisition of spectral signatures from TST skin sites. The imaging protocol contributes to the research aim by enabling the reliable acquisition of hyperspectral images for quantitative characterisation of TST indurations to aid the diagnosis of latent tuberculosis infection. The protocol was derived by modifying the Specim IQ imaging protocol—reviewed in Chapter 2—to account for specific illumination conditions, namely the type of primary light source and the presence of secondary, confounding illumination commonly found in clinical settings.

To minimise the impact of these illumination factors on spectral reproducibility, the modified protocol included targeted adjustments and was validated using the ventral forearm of ten participants with varying skin tones. Validation involved the construction of a controlled imaging enclosure to standardise conditions across all participants. The findings from this validation study identified key illumination-related variables affecting spectral consistency, which informed further fine-tuning of the protocol. The background, methodology, validation procedures, and results of this refined imaging protocol are presented in the sections that follow.

3.1 Background and Related Works

Illumination conditions are among the most critical camera-independent factors influencing the quality and reproducibility of HSI, particularly in dermatological applications where spectral precision is essential (He et al., 2020). Skin spectral signatures are sensitive to variations in illumination, and this is especially relevant in studies such as TST imaging, where subtle spectral features may inform

diagnostic decisions. While parameters such as intensity, angle, and distance can also influence outcomes, this chapter focuses specifically on the impact of **illumination source type**, assuming its spectral characteristics have a more pronounced effect on skin reflectance than other lighting variables.

The **Specim IQ hyperspectral camera** (400–1000 nm) was selected for this study due to its spectral coverage of both VNIR regions, enabling characterisation of the epidermal, dermal, and subcutaneous skin layers (Huang et al., 2023; Johansen et al., 2019). This makes it well suited for dermatological research (Lu & Fei, 2014). Previous studies, including those by Behmann et al. (2018), Wawrzyk-Bochenek et al. (2023), and Zdrada-Nowak et al. (2023), have demonstrated the effectiveness of the Specim IQ for imaging skin. Accordingly, the imaging protocol established by Behmann et al. (2018)—outlined in Figure 3.1—served as the foundation for this study.

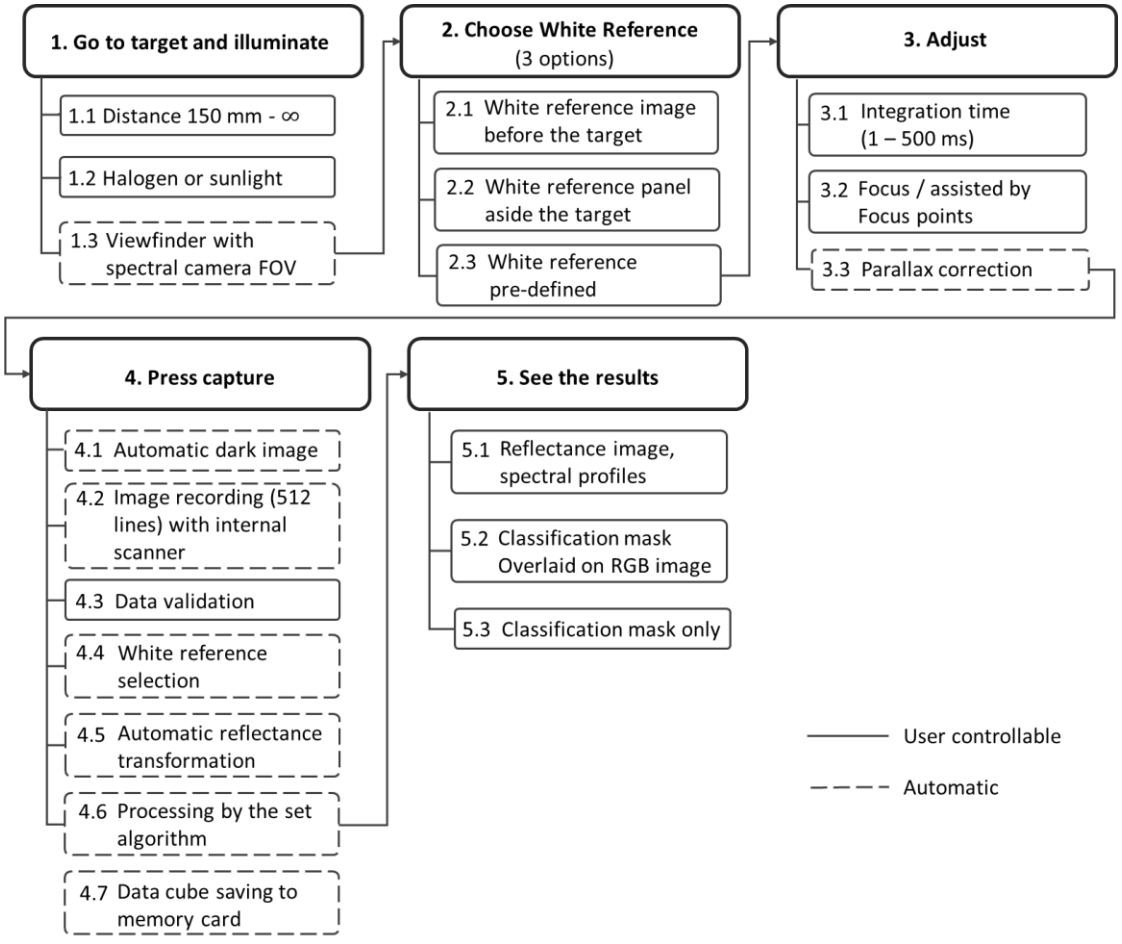


Figure 3.1: Revisiting the imaging protocol for spectral signature acquisition using the Specim IQ HSI camera (Behmann et al., 2018).

To investigate the reproducibility of spectral signatures from TST sites, **item 1.0 of the protocol (illumination conditions)** was refined. This decision was informed by the understanding that all subsequent imaging steps are contingent on reliable and uniform lighting at the target site.

At the time of writing, He et al. (2020) remains the only study to directly assess how different illumination sources affect the reproducibility of skin spectral signatures across varying skin tones. Their use of a Thouslite LED array, tuned to six wavelengths in the 400–700 nm range, allowed controlled comparisons across the visible spectrum. However, this setup failed to simulate clinically relevant light sources such as Halogen and Xenon and omitted the near-infrared (NIR) region—limiting its relevance to medical HSI applications. This highlights a critical research gap: the need to evaluate the impact of **real-world VNIR illumination sources**—Halogen, Xenon, and broadband LED—on the reproducibility of skin spectral signatures.

In addition to light source selection, **ambient illumination** in clinical environments introduces confounding variability. While some portable HSI systems (e.g., Fabelo et al. (2019); Hetz et al. (2024)) incorporate lenses with short focal distances to minimise ambient light interference, high-performance systems like the Specim IQ and Tivita Tissue have longer minimum focus distances (150 mm and 300 mm, respectively), which increase their susceptibility to ambient light (Barberio et al., 2018). Some studies address this by imaging in rooms completely sealed from ambient light (e.g., Deda et al. (2024); Grambow et al. (2019)), but this is impractical in most clinical settings.

To address this limitation, this study introduces **flanking reflectors** into the imaging scene. These reflectors, used alongside a white reference panel, helped to distribute both primary and ambient light more uniformly, thereby reducing the impact of ambient illumination on spectral measurements. This enabled realistic imaging conditions while improving spectral reproducibility across a range of skin tones.

Taken together, this chapter addresses two key gaps in the literature:

1. The limited evaluation of reproducibility under **clinically relevant illumination types**, and
2. The lack of practical solutions to **mitigate ambient light interference** in clinical HSI.

Both issues are resolved through the refinement of the Behmann et al. (2018) protocol and the introduction of novel scene design elements. The methodology and validation of these contributions are described in the following sections.

3.2 Methodology

A physical enclosure was constructed to accommodate the unmodified Specim IQ imaging protocol for capturing spectral data. The enclosure was further modified to address limitations in the illumination procedure 1.0. This was done by removing sunlight as an illumination source and

incorporating artificial alternatives, such as a broadband LED, white LED, and Xenon, in addition to the baseline Halogen.

The design and the dimensions of the constructed enclosure are as shown in Figure 3.2. The enclosure was designed using the SOLIDWORKS 2020 (solidworks.com) computer-aided design software enabling the enclosure to be 3D printed. The enclosure was designed to be modular for easy assembly and disassembly; and transportation between study sites. The modularity also enables the enclosure to be printed in parts rather than as a single object. The proposed design of the enclosure enables easy replication through 3D printing by other researchers who might use the proposed protocol.

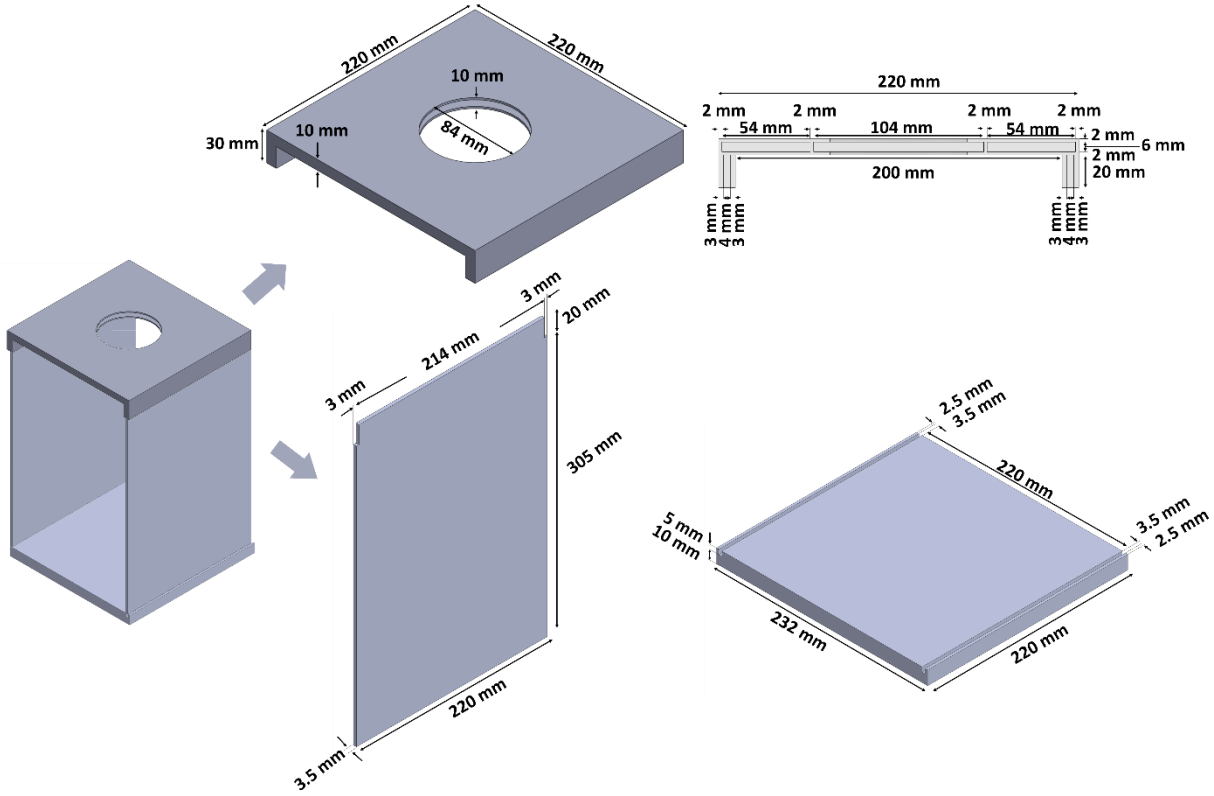


Figure 3.2: Design and dimensions of the 3D-printed enclosure for hyperspectral imaging of TST injection site

The top section of the enclosure features a circular slot to accommodate the Specim camera lens, enabling it to capture images within the enclosure. The design facilitates the positioning of a participant's left forearm—specifically the ventral side, which is the typical site for a TST injection—directly beneath the camera lens and light sources. The flanking sections of the enclosure allow vertical adjustment of the camera, enabling it to be positioned up to approximately 300 mm above the participant's arm. The choice of 300 mm was motivated by the minimum focus distance of Specim IQ which is 150 mm.

Both the front and rear of the enclosure are intentionally left open to simplify the placement of the participant's arm. However, this open design also permits ambient light to enter the imaging scene, potentially introducing confounding illumination. To address this, aluminium reflectors were installed on the inner flanks of the enclosure to assess their effectiveness in reducing ambient light interference.

The base section was engineered to provide structural stability for the enclosure, camera, and various light sources. It is also equipped with a cushion to enhance comfort for participants as they rest their forearm during imaging.

The resulting multi-illumination enclosure represents the physical implementation of the modified imaging protocol. This system was validated across a diverse range of participant skin tones to assess the impact of varying illumination conditions on the reproducibility of spectral signatures.

3.1.1. Description of the Imaging Protocol for the Acquisition of Spectral Signatures of TST Skin Sites

Each hyperspectral image, captured by an HSI camera, also called a hypercube, is denoted as $\boldsymbol{H(x, y, \lambda)}$. It is described here as an array of spectral signatures $\boldsymbol{s_{x,y,\lambda}}$, and as defined in equation (3.1). Each spectral signature $\boldsymbol{s_{x,y,\lambda}}$ contains diagnostic information about a fractional area on the skin or diseased tissue area of a participant. The spatial axes $\boldsymbol{x}$ and $\boldsymbol{y}$, and a spectral axis $\boldsymbol{\lambda}$, parameterise the hypercube, as shown in Figure 3.3. Each hypercube generated by the Specim IQ camera consisted of 512×512 spatial pixels (262,114 total), spanning wavelengths from $\lambda_0 = 397\ nm$ to $\lambda_{N-1} = 1003\ nm$, (where N=204). The spectral resolution of each wavelength band is 7 nm full width at half maximum (FWHM).

$$H(x, y, \lambda) = \{s_{x,y,\lambda} | s_{x,y,\lambda} \in \mathbb{R}, 0 \leq x \leq X-1, 0 \leq y \leq Y-1, 0 \leq \lambda \leq N-1\} \quad (3.1)$$

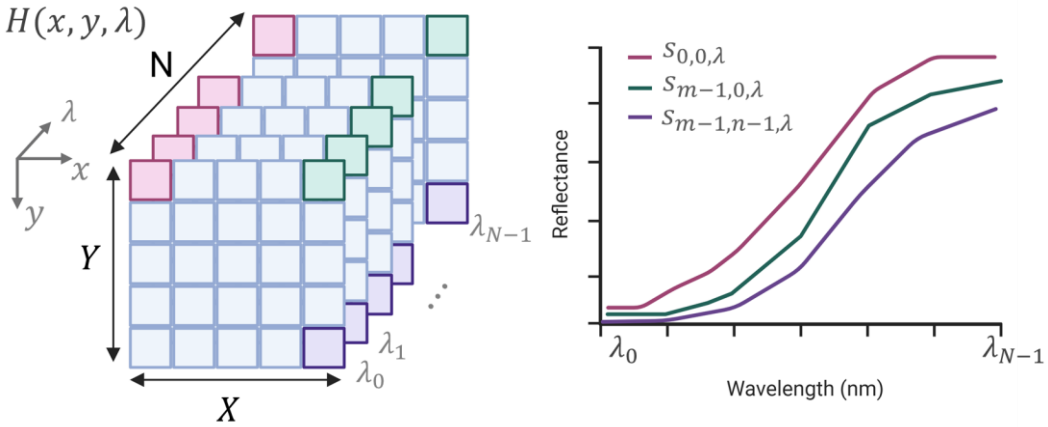


Figure 3.3: Description of a hypercube and constituent spectral signatures

The enclosure was designed to enable the installation of multiple illumination source types. They were positioned to illuminate a participant's arm placed on the enclosure base. The enclosure allows positioning the camera directly above the forearm, with illumination sources flanking the camera on both sides, as shown in Figure 3.4. The dispersive effect of the introduced flanking reflectors is as illustrated in Figure 3.5. The physical distance between the camera lens and the base of the enclosure is set to 300 mm, assuming the forearm thickness does not exceed 150 mm.

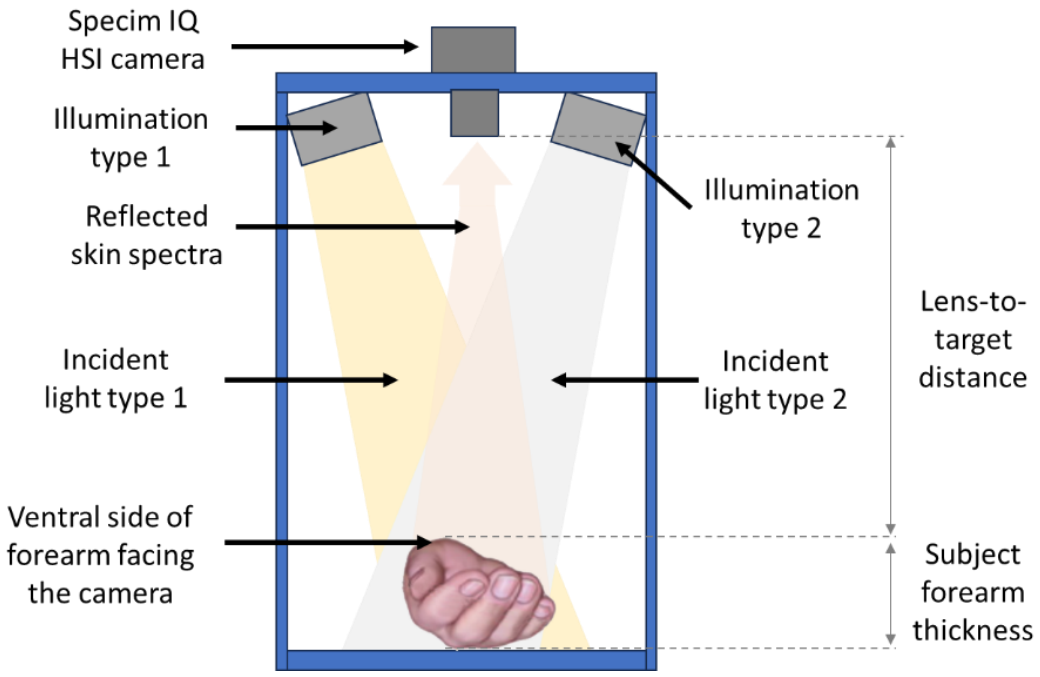


Figure 3.4: Enclosure design for the modification of the Specim IQ imaging protocol

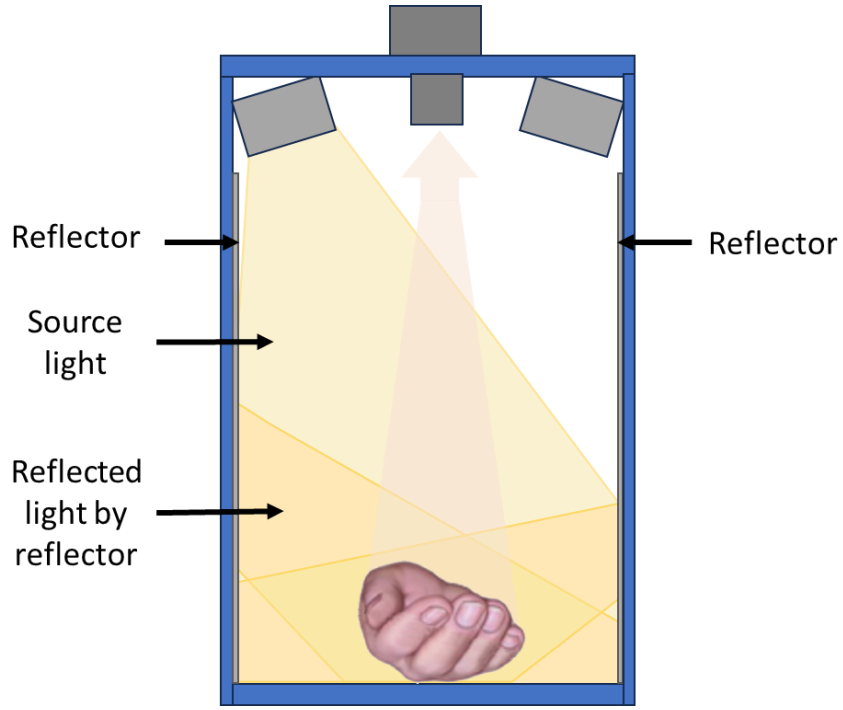


Figure 3.5: Inclusion of reflectors in enclosure design for uniform illumination of the target

Variations in forearm thickness among participants led to differences in the lens-to-target distance, resulting in variations in apparent spatial dimensions captured by the camera. Ink-stamped fiducial markers of known dimensions were applied to the surface of the forearms to standardise the spatial dimensions of the hyperspectral images across the participants. This approach accounted for differences in lens-to-target distance, as illustrated in Figure 3.6. The CAD design of the 3D printed stamp for the fiducial markers is also shown in the figure.

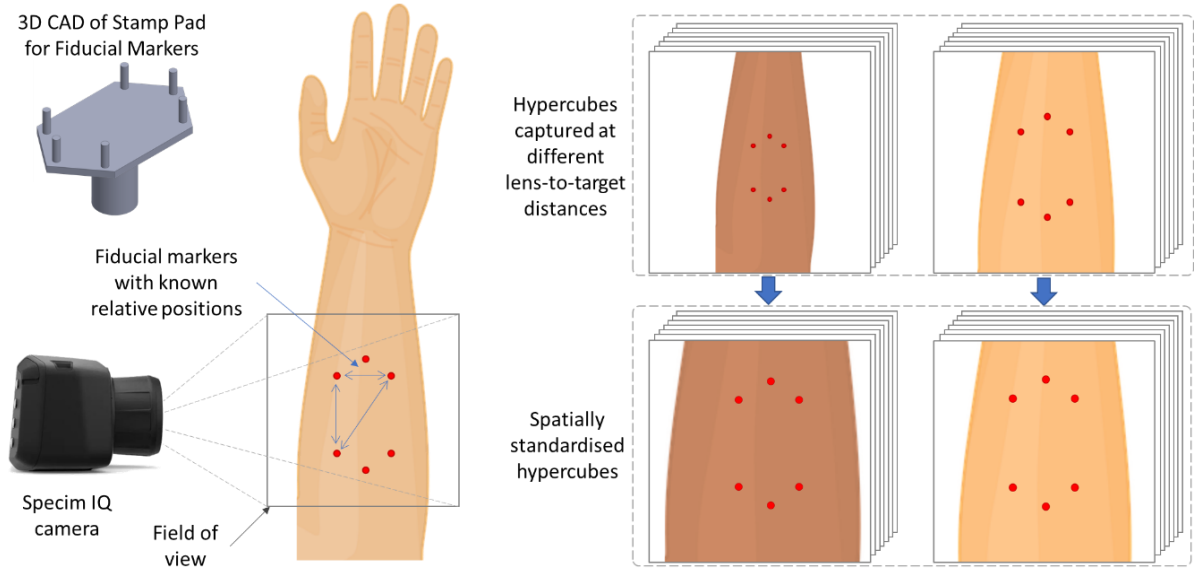


Figure 3.6: Standardisation of spatial dimension of hypercubes using ink-stamped fiducial markers (red dots) of known dimensions

The RGB viewfinder of the Specim camera was utilized to visualise the imaging scene in real-time. It enabled the verification of the position of participant forearms before capturing each hypercube. A Spectralon white reference panel (with 99% diffuse reflection) was placed in the imaging scene to reflect incident illumination to the camera. This enabled the camera to capture the illumination condition of the scene. This information was used to normalise the captured spectral signatures.

During the acquisition of each hypercube, the integration time of the camera was set to approximately 117 ms per line of spectral signatures. This resulted in an average acquisition of 60 seconds (117 ms x 512 lines) per hypercube, as illustrated in Figure 3.7. The physical distance between the viewfinder and the HSI camera lens led to a parallax error between what was visualised in the scene before imaging and the spatial data of the captured hypercube. Image registration was performed on each captured hypercube to correct this error by aligning its spatial axis to the viewfinder image, as illustrated in Figure 3.8. The fiducial markers were utilized as anchor points to perform the registration.

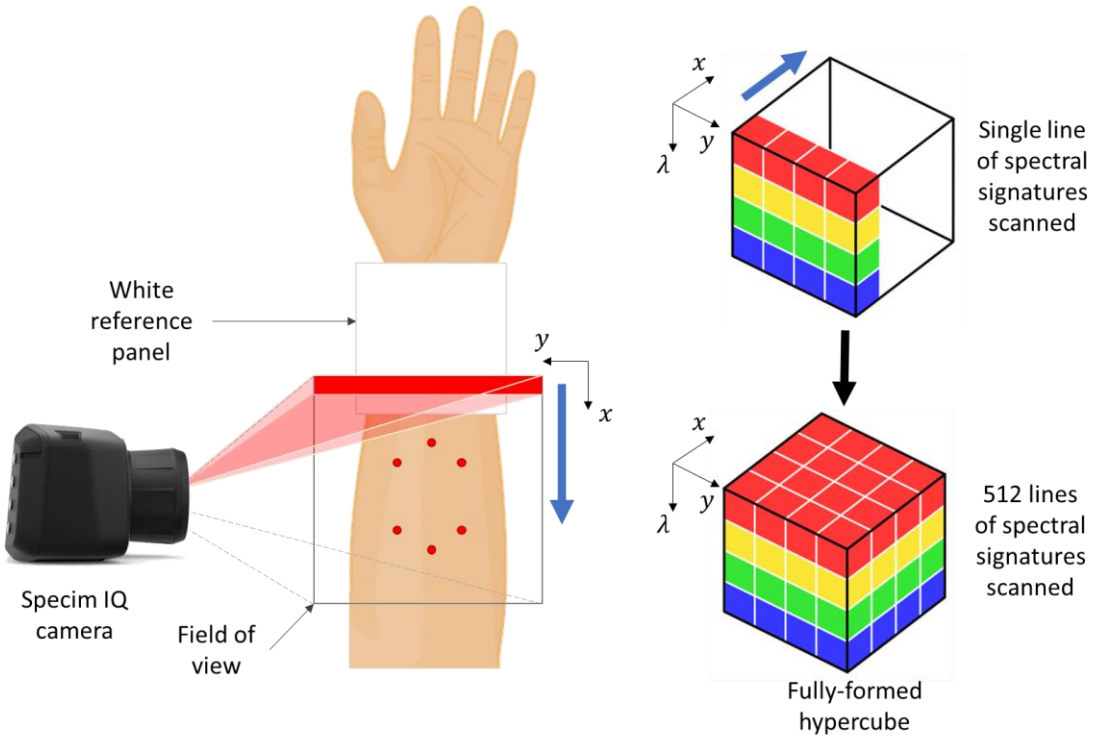


Figure 3.7: Description of line scan acquisition of spectral signatures from TST skin site

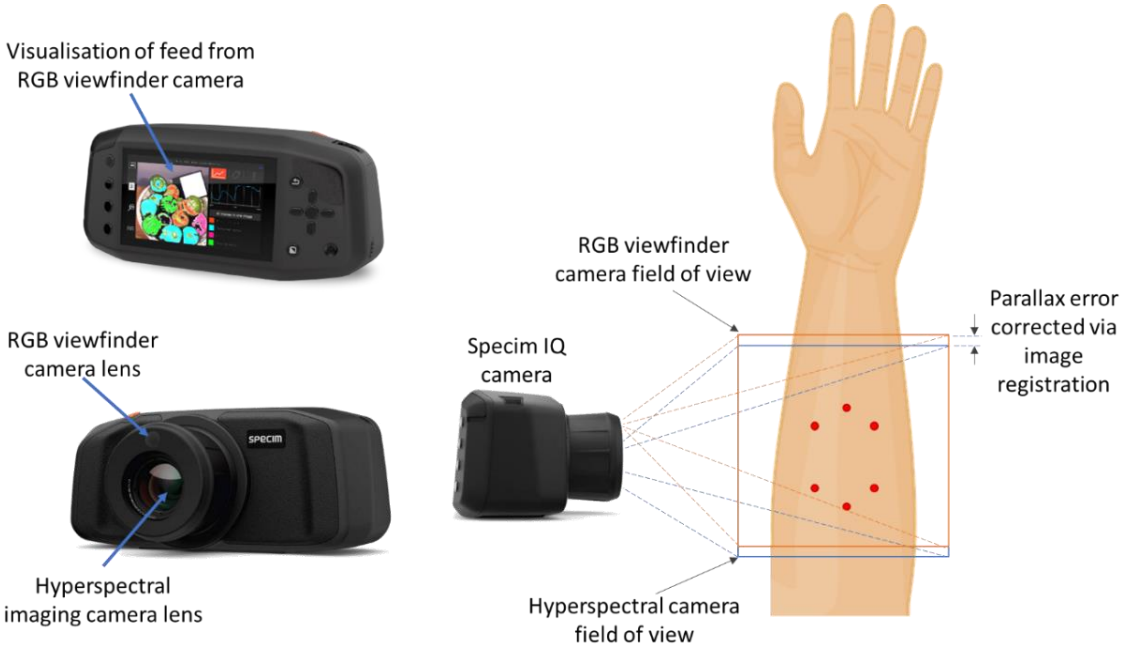


Figure 3.8: Description of the hyperspectral and RGB viewfinder field of views

3.1.2. Evaluation of the Impact of Different Light Sources on the Reproducibility of Skin Spectral Signature for Different Skin Tones

Ten human participants with varying skin tones were recruited to evaluate the impact of the specified illumination conditions on spectral reproducibility for the TST injection site. The skin tone classification was based on the Fitzpatrick scale (Oliveira et al., 2023). The range of tones was from dark brown

(Fitzpatrick type VI) to pale white (Fitzpatrick type I), as shown by the RGB images in Figure 3.9 and Table 3.1. Human research ethics approval was obtained from the University of Cape Town (UCT) human research ethics committee (HREC) (with ethics number HREC 319/2018) before the participants were recruited and imaged.

Multiple illumination conditions were simulated within the constructed enclosure per participant. A hypercube was captured for each simulated illumination condition. Spectral signatures of the same skin region per participant were captured for each illumination condition. Six illumination conditions were simulated:

- The use of a single primary light source type such as:
  - **Condition I:** Halogen
  - **Condition II:** Broadband LED
  - **Condition III:** White LED
  - **Condition IV:** Xenon
- **Condition V:** A use of a primary light source (Halogen) and a single confounding secondary illumination (broadband LED). Flanking reflectors were not installed in the simulation of this condition.
- **Condition VI:** The effect of introducing flanking reflectors on the minimisation of confounding secondary illumination. This condition was simulated using Halogen as the primary source and the combination of the other three sources (broadband LED, white LED, and Xenon) as a proxy for a complex confounding illumination.

The new steps introduced into the Specim protocol to investigate the reproducibility of TST site spectra under specified illumination conditions is as shown in Figure 3.10.

Figure 3.9: RGB images showing range of participant skin tones

Table 3.1: Metadata of Participants for Spectral Reproducibility Test

PARTICIPANT ID	SEX	RACE/NATIONALITY	SKIN TONE
PID1	Male	Black/Togolese	Dark Brown (Type VI)
PID2	Female	Black/South African	Dark Brown (Type VI)
PID3	Female	Black/Zimbabwean	Dark Brown (Type VI)
PID4	Female	Black/Zimbabwean	Brown (Type V)
PID5	Male	Black/Nigerian	Brown (Type V)
PID6	Male	South Asian/Pakistani	Moderate Brown (Type IV)
PID7	Female	White/Italian	Medium White (Type III)
PID8	Male	White/South African	White (Type II)
PID9	Male	White/South African	Pale White (Type I)
PID10	Female	White/American	Pale White (Type I)

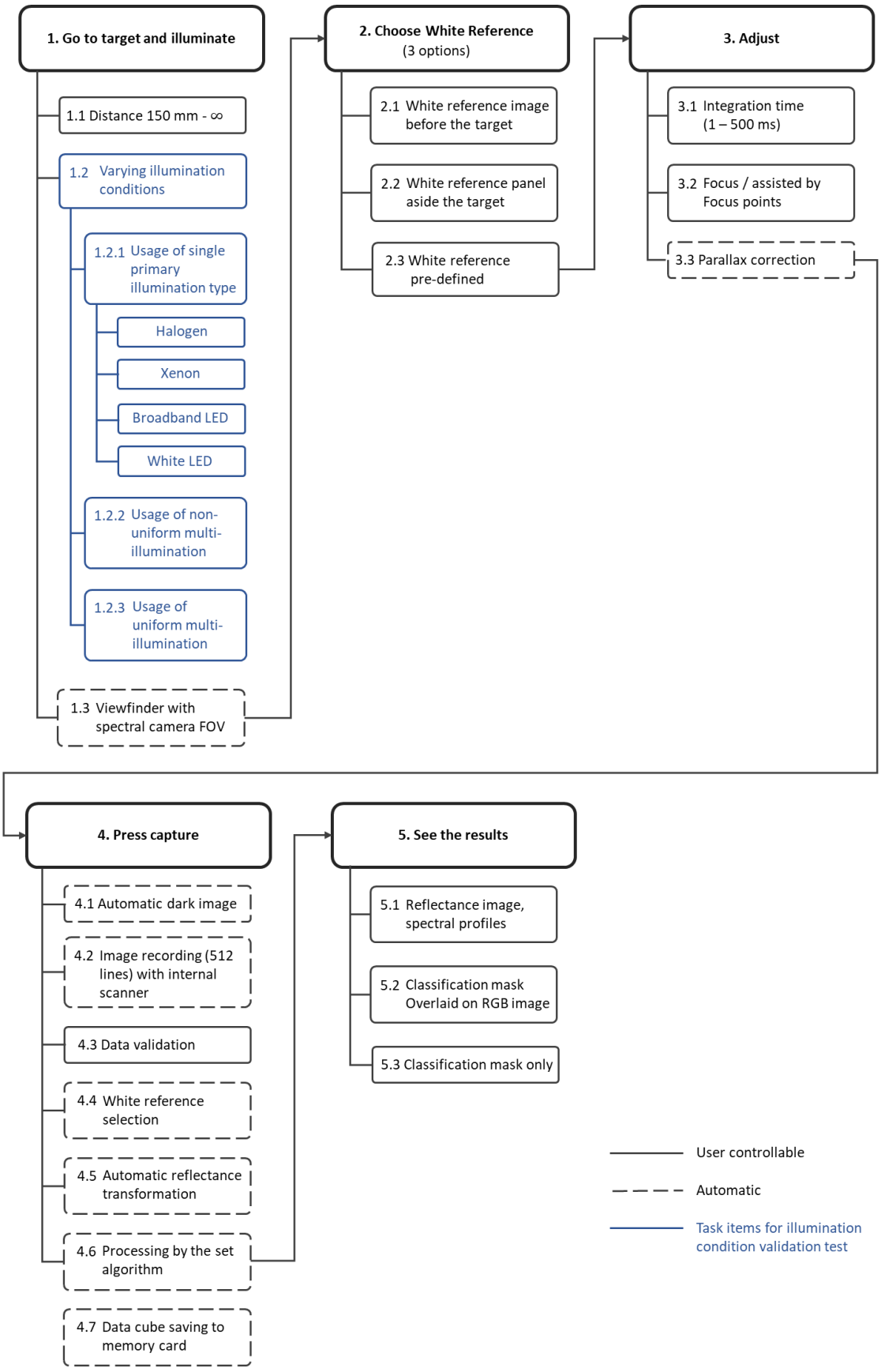


Figure 3.10: Imaging protocol of Specim IQ modified to perform validation tests with varying illumination conditions

The reproducibility of skin spectral signatures across participants was assessed by evaluating the correlations and spectral angles under the simulated illumination conditions. The enclosure housed the camera, four 10-watt halogen lamps (OZB-A4804, kern-sohn.com), a 25-watt Effilux broadband LED array (EFFI-Flex-HSI, effilux.com), a 2-watt white LED array, and a 45-watt Xenon arc lamp (H4EW-2, hids4u.co.uk), as shown in figure 3.11. For the single illumination imaging, the Specim camera, using its viewfinder to scan the illumination intensity of the scene, automatically adjusted the exposure time. This adjustment enabled the standardisation of the image brightness to account for differences in the illumination intensities of the light sources. This was the measure employed for controlling the effect of relative differences in illumination intensities. The acquisition time per hypercube was 60 seconds on average across all imaging tests.

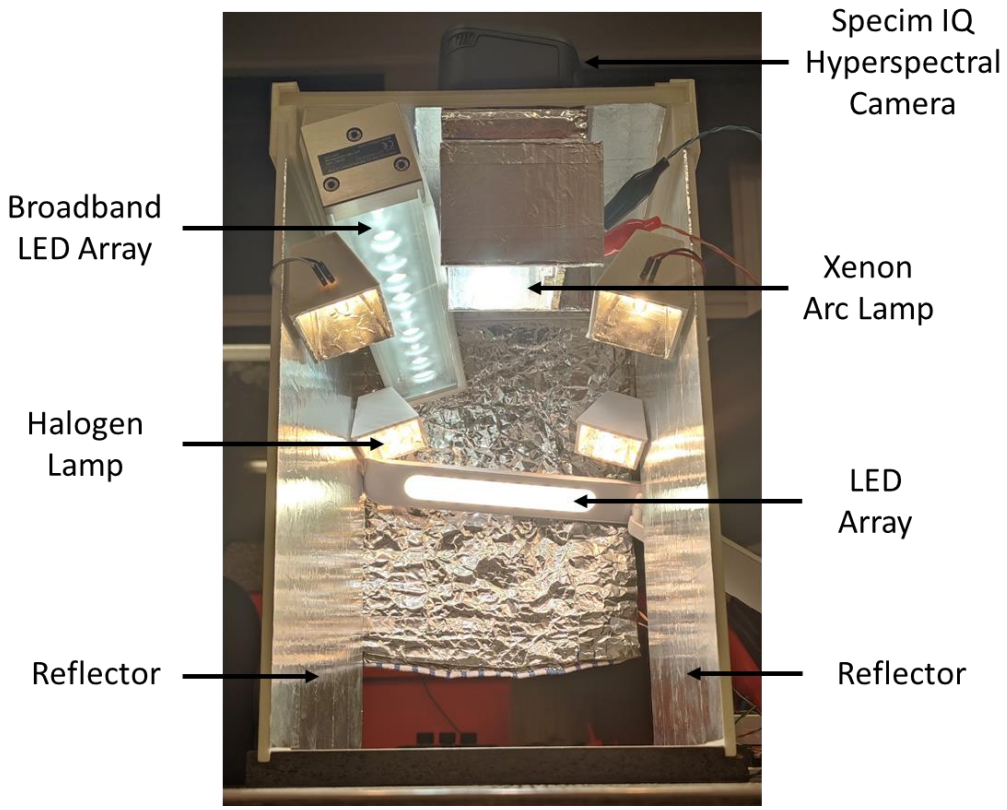


Figure 3.11: Implementation of enclosure constructed with multiple illumination source types and the Specim IQ hyperspectral camera for capturing spectral signatures

The setup for the imaging is as shown in Figure 3.12. The participants were made to grasp a gripping platform with their left hand. This was done to minimise the involuntary motion of the forearm during imaging. The platform had suction points under its base to affix it to the table on which the enclosure rested. The light sources did not expose the participants to any discomfort. The sources were at least 150 mm away from the forearm of participants. They did not significantly raise the skin surface temperature during the duration of imaging.

Figure 3.12: Pictures of the setup for imaging protocol validation tests

Each single pixel raw reflectance spectrum captured by the camera was denoted by $\boldsymbol{I_{raw}(\lambda)}$. It was estimated to be composed of single pixel skin spectral signature $\boldsymbol{I_{skin}(\lambda)}$ and an illumination source spectra $\boldsymbol{I_{white}(\lambda)}$, as illustrated in Figure 3.13(a). The inherent dark current spectra $\boldsymbol{I_{dark}(\lambda)}$ of the camera was extracted from the device. The per pixel skin spectral signature was extracted from $\boldsymbol{I_{raw}(\lambda)}$, $\boldsymbol{I_{white}(\lambda)}$, and $\boldsymbol{I_{dark}(\lambda)}$, using equation (3.2) (Liu et al., 2011). The illumination source spectra $\boldsymbol{I_{white}(\lambda)}$ was captured by placing the white reference panel in the imaging scene, next to the skin region of interest (ROI).

$$\boldsymbol{I_{skin}(\lambda)} = \frac{\boldsymbol{I_{raw}(\lambda)} - \boldsymbol{I_{dark}(\lambda)}}{\boldsymbol{I_{white}(\lambda)} - \boldsymbol{I_{dark}(\lambda)}} \qquad 3.2$$

The imaging of participants was conducted in a laboratory sealed off from sunlight and artificial illumination other than the selected sources in the enclosure. This minimised the risk of unintended confounding illumination in the imaging scene. Illumination conditions I to IV were simulated by capturing four hypercubes per participant using each of the selected light sources independently, as illustrated in Figure 3.13(a). Illumination condition V was simulated by capturing a single hypercube per participant in the presence of the Halogen primary source and broadband LED as the confounding illumination, as illustrated in Figure 3.13(b). The flanking reflectors were not included for this simulation to investigate the effect of their absence. Condition VI was simulated by capturing a single hypercube per participant with the introduction of two other confounding illuminations (white LED and Xenon) and flanking reflectors, as illustrated in Figure 3.13 (c). This resulted in the acquisition of a total of six hypercubes per participant (one hypercube per illumination condition).

During the simulation of conditions I to IV, the illumination spectra captured by the white reference panel for the Halogen, broadband LED, white LED, and Xenon primary sources were denoted by $\boldsymbol{I_{white(Hg)}}$, $\boldsymbol{I_{white(BL)}}$, $\boldsymbol{I_{white(WL)}}$, and $\boldsymbol{I_{white(Xn)}}$, respectively. The white reference illumination spectra captured during the simulation of condition V was denoted by $\boldsymbol{I_{white(cond\_V)}}$. As illustrated in Figure 3.13(b), $\boldsymbol{I_{white(cond\_V)}}$ was composed of only the primary illumination while $\boldsymbol{I_{raw}(\lambda)}$ included the spectra contribution of the confounding illumination and $\boldsymbol{I_{skin}(\lambda)}$. As illustrated in Figure 3.13(c), the introduction of the flanking reflectors in condition VI evenly distributed all illumination in the scene such that the white reference, denoted by $\boldsymbol{I_{white(cond\_VI)}}$, captured the primary illumination and all the confounding illumination. Similarly, $\boldsymbol{I_{raw}(\lambda)}$ comprised all illumination in the scene in addition to $\boldsymbol{I_{skin}(\lambda)}$.

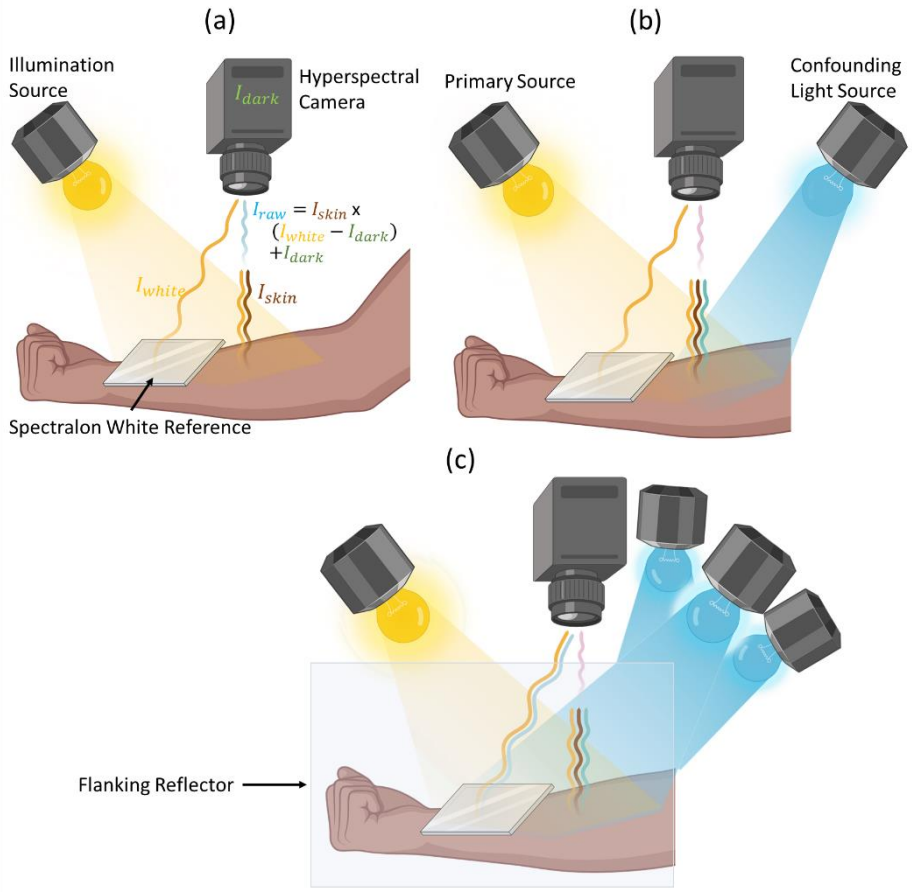


Figure 3.13: Illustration showing raw spectra acquisition with (a) one primary light source, (b) inclusion of a confounding illumination, and (c) introduction of flanking reflectors to evenly distribute all illumination in the scene

The position of the forearm in all six hypercubes per participant was not perfectly identical due to some minor involuntary motion of participants between image scans. Image registration was performed on the six images per participant to align skin tissue ROI to the same pixels. The first hypercube per participant was selected as the reference image. The other five images were aligned to this image using the fiducial markers in each image as control points. The registration was

implemented using the homography estimation via linear least-squares minimization approach (Hartley & Zisserman, 2003). The mean spectral signature of a 20 x 20-pixel ROI was extracted for each of the six hypercubes per participant. The mean spectra for each hypercube were denoted by $\boldsymbol{I_{skin(Hg)}}$, $\boldsymbol{I_{skin(BL)}}$, $\boldsymbol{I_{skin(WL)}}$, $\boldsymbol{I_{skin(Xn)}}$, $\boldsymbol{I_{skin(cond\_V)}}$, and $\boldsymbol{I_{skin(cond\_VI)}}$, respectively. This is as shown in Figure 3.14.

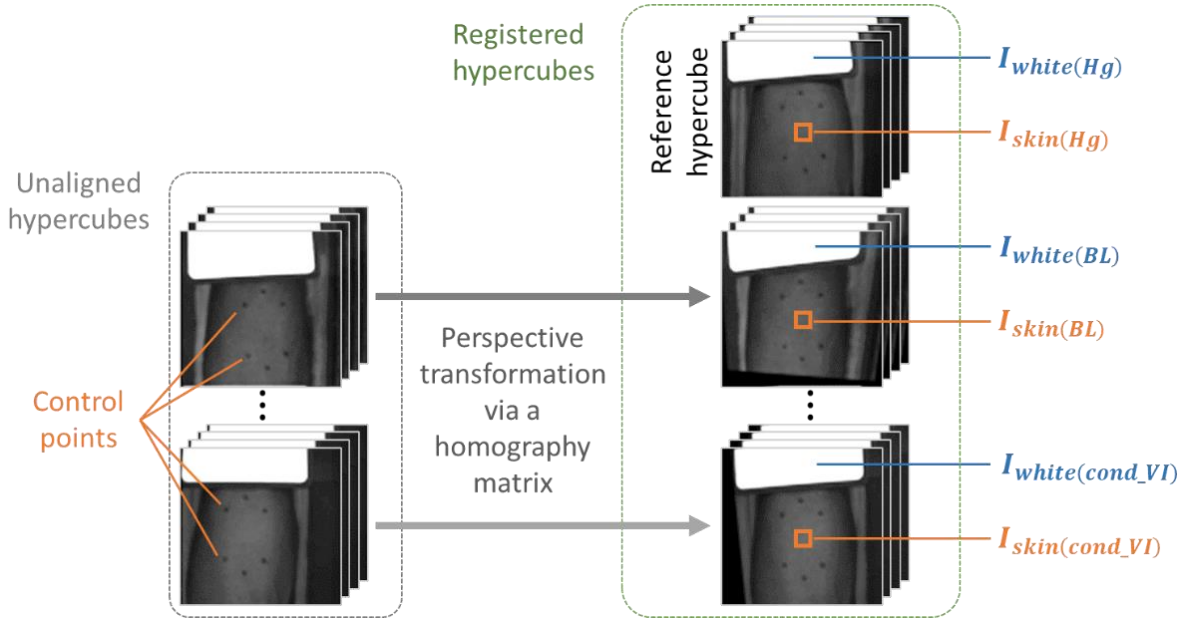


Figure 3.14: Registration of hypercubes captured for a single participant

The consistency between the mean skin spectral signature across the various illumination conditions per participant was estimated. This served as proxy for quantifying reproducibility of the spectral signature of a ROI under the six conditions. Pearson's correlation coefficient and spectral angle were the metrics employed as measures for estimating this spectral reproducibility. The correlation coefficient σ_{ab} between the mean ROI skin spectral signatures $\boldsymbol{I_{skin(a)}}$ and $\boldsymbol{I_{skin(b)}}$ was estimated using equation (3.3) (Liu et al., 2016).

$$\sigma_{ab} = \frac{\sum_{\lambda=0}^{N-1}\left(\boldsymbol{I_{skin(a)}}(\lambda) - \overline{\boldsymbol{I_{skin(a)}}(\lambda)}\right)\left(\boldsymbol{I_{skin(b)}}(\lambda) - \overline{\boldsymbol{I_{skin(b)}}(\lambda)}\right)}{\sqrt{\sum_{\lambda=0}^{N-1}\left(\boldsymbol{I_{skin(a)}}(\lambda) - \overline{\boldsymbol{I_{skin(a)}}(\lambda)}\right)^2\left(\boldsymbol{I_{skin(b)}}(\lambda) - \overline{\boldsymbol{I_{skin(b)}}(\lambda)}\right)^2}} \quad 3.3$$

The spectral angle α_{ab} between the mean ROI skin spectral signatures $\boldsymbol{I_{skin(a)}}$ and $\boldsymbol{I_{skin(b)}}$ was estimated using equation (3.4) (Cheng et al., 2018).

$$\alpha_{ab} = \cos^{-1}\left(\frac{\sum_{\lambda=0}^{N-1}\boldsymbol{I_{skin(a)}}(\lambda)\boldsymbol{I_{skin(b)}}(\lambda)}{\sqrt{\left(\sum_{\lambda=0}^{N-1}\boldsymbol{I_{skin(a)}}(\lambda)\right)^2}\sqrt{\left(\sum_{\lambda=0}^{N-1}\boldsymbol{I_{skin(b)}}(\lambda)\right)^2}}\right) \quad 3.4$$

3.3 Results

Using the modified imaging protocol, 60 hypercubes, consisting of six hypercubes per participant for 10 participants, were captured. Figure 3.15(a) shows the mean raw and calibrated spectral signatures $\boldsymbol{I_{raw(Hg)}}$ and $\boldsymbol{I_{skin(Hg)}}$, respectively, for a ROI of a single participant under Halogen illumination. The Halogen white reference and the dark reference spectra, $\boldsymbol{I_{white(Hg)}}$ and $\boldsymbol{I_{dark(Hg)}}$, are also shown in the figure. The white reference spectra $\boldsymbol{I_{white(Hg)}}$, $\boldsymbol{I_{white(BL)}}$, $\boldsymbol{I_{white(WL)}}$, $\boldsymbol{I_{white(Xn)}}$, $\boldsymbol{I_{white(cond\_V)}}$, and $\boldsymbol{I_{white(cond\_VI)}}$ for each of the six simulated illumination conditions are shown in Figure 3.15(b). The average raw and calibrated spectral signatures for a single participant under the six conditions are shown in Figure 3.15(c) and Figure 3.15(d) respectively.

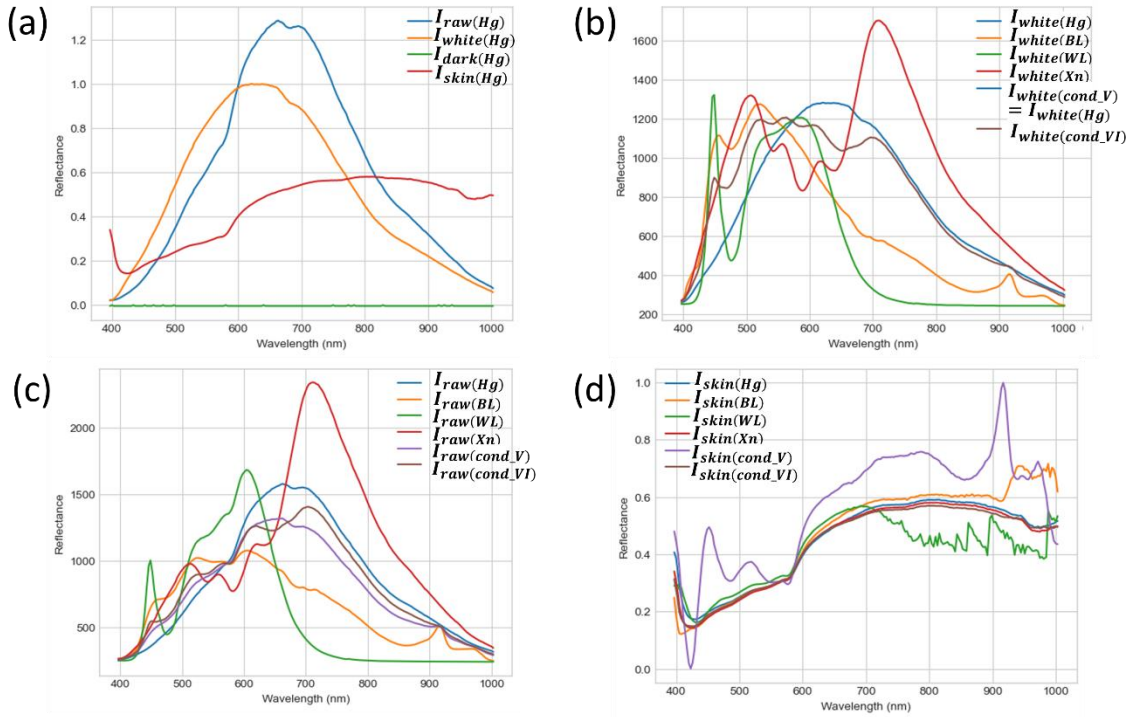


Figure 3.15: Spectral signatures under different illumination conditions for participant PID4: (a) The halogen-illuminated spectra from the participant, (b) white reflectance spectra for each of the six simulated illumination conditions, (c) corresponding raw spectra for participant, (d) the corresponding calibrated mean ROI skin spectral signatures

Figure 3.16(a) shows the mean correlation matrix for all 10 participants across the six simulated illumination conditions. Each cell in the matrix shows the average correlation, calculated across all 10 participants, between the calibrated ROI skin spectral signatures $\boldsymbol{I_{skin(a)}}$ and $\boldsymbol{I_{skin(b)}}$ captured under the respective illumination conditions a and b. The p-value of each mean pairwise correlation was < 0.01 which shows the statistical significance of computed correlations. The mean spectral angle matrix for all 10 participants across the six simulated illumination conditions is shown in Figure 3.16(b). Each cell in the matrix shows the average spectral angle, calculated across all 10 participants, between the

calibrated ROI skin spectral signatures $\boldsymbol{I_{skin(a)}}$ and $\boldsymbol{I_{skin(b)}}$ captured under the respective illumination conditions a and b.

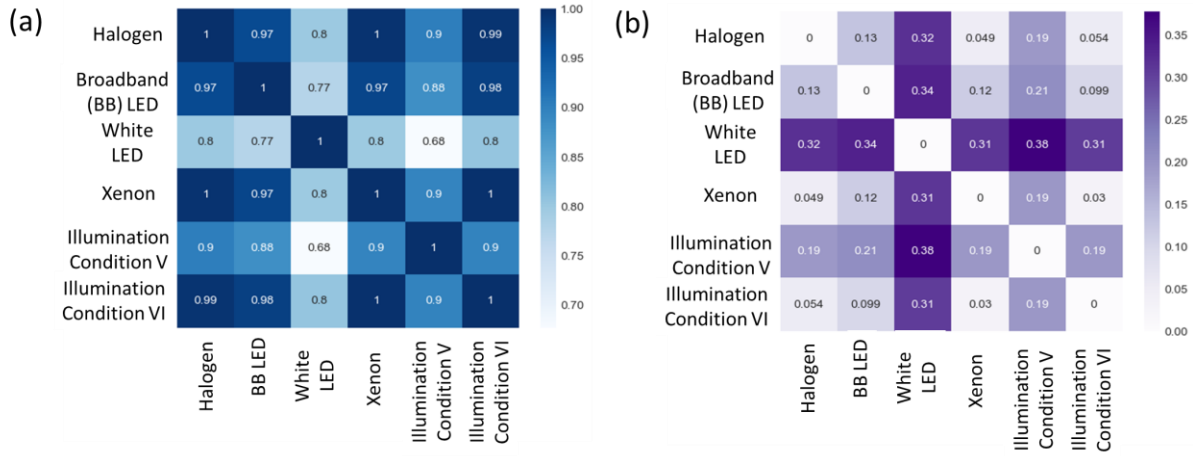


Figure 3.16: Pairwise estimation of consistency between skin spectra obtained under different illumination conditions using (a) mean correlation matrix, (b) mean spectral angle matrix.

Figure 3.17(a) shows the mean spectra of $\boldsymbol{I_{skin(Hg)}}$, $\boldsymbol{I_{skin(BL)}}$, $\boldsymbol{I_{skin(WL)}}$, $\boldsymbol{I_{skin(Xn)}}$, $\boldsymbol{I_{skin(cond\_V)}}$, and $\boldsymbol{I_{skin(cond\_VI)}}$ for three representative participants. It also shows the margin of error of the mean skin signatures for a representative participant with dark brown skin tone (Fitzpatrick type VI), with moderate brown (Fitzpatrick type IV), and one with pale white skin tone (Fitzpatrick type I). Figure 3.15(b) shows the mean spectra for the same three participants but now excluding the $\boldsymbol{I_{skin(WL)}}$ and $\boldsymbol{I_{skin(cond\_V)}}$. Based, on these results, a list of illumination factors that influence the reproducibility of skin spectral signatures, are described in table 3.1. Based on these factors, procedures in the modified protocol of Figure 3.10 which produced low spectral reproducibility values were excluded from the protocol. This information was used to fine-tune the final imaging protocol to the version shown in Figure 3.18, which produced spectral signatures per participant that were consistent under four illumination conditions.

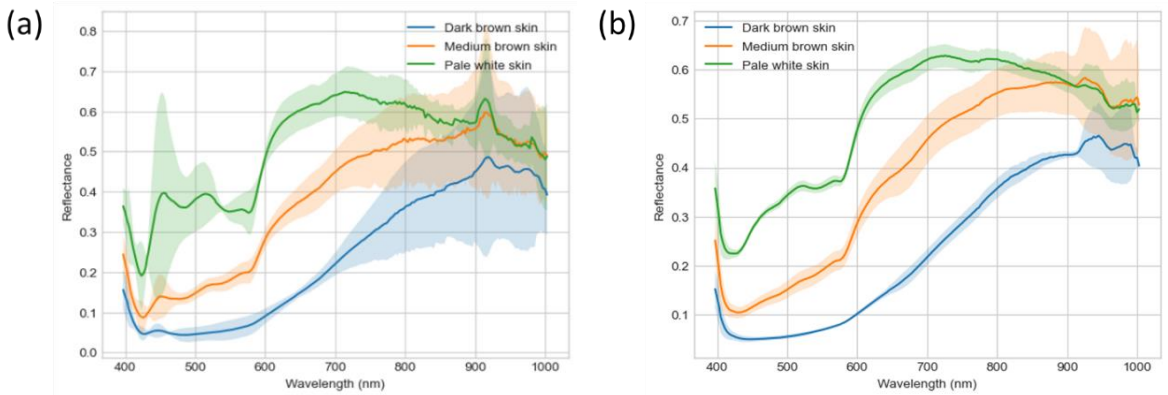


Figure 3.17: Mean skin spectral signature across all illumination conditions for three representative participants, (b) mean skin spectral signature excluding the white LED and uneven illumination conditions.

Table 3.2: Illumination Factors that Influence the Reproducibility of Skin Spectral Signatures

	Illumination Factors
1	Selection of a primary artificial light source type from the options [Halogen, Xenon, Broadband LED, White LED] – Sorted from most to least preferred
2	Uniformity of primary and ambient lighting, facilitated by flanking reflectors, across the skin region of interest and white reference panel area.
3	The use of white LED as a primary light source for TST HSI and other skin applications at wavelengths beyond 700 nm.
4	The use of broadband LED as a primary light source for TST HSI and other skin applications at wavelengths beyond 900 nm.

Illumination factor 1 was informed by the pairwise correlations between skin spectra generated by [Xenon, broadband LED, White LED] and skin spectra generated by halogen exhibit a decreasing order, while the corresponding pairwise spectral angles show an increasing order. This is as shown in figures 3.16(a) and 3.16 (b) respectively. Illumination factor 2 was informed by the correlations between skin spectra generated under illumination condition V other illumination conditions are relatively lower, while the spectral angles are higher. In contrast, the correlations between skin spectra generated under illumination condition VI and other illumination conditions are higher, with lower spectral angles. This is as shown in figures 3.16(a) and 3.16(b). Furthermore, the margin of error decreased from figure 3.17(a) to 3.17(b), partially due to the exclusion of skin spectra from condition V. Illumination factor 3 was informed by the fact that there were relatively low intensities in the white LED spectra and resulting white LED-based raw skin spectra beyond 700 nm, as shown in figures 3.15(b) and 3.15(c). There is significant divergence in the spectral characteristics of white LED-based calibrated skin spectra beyond 700 nm, as shown in figure 3.15(d). Furthermore, the margin of error decreased from figure 3.17(a) to 3.17(b), partially due to the exclusion of white LED-based skin spectra. Illumination factor 4 was informed by the fact that there were characteristic spikes in the spectral intensities of broadband LED spectra and resulting raw skin spectra between 900 and 1000 nm, as shown in figures 3.15(b) and 3.15(c). There is significant divergence in the spectral characteristics of broadband LED-based calibrated skin spectra beyond 900 nm, as shown in figure 3.15(d).

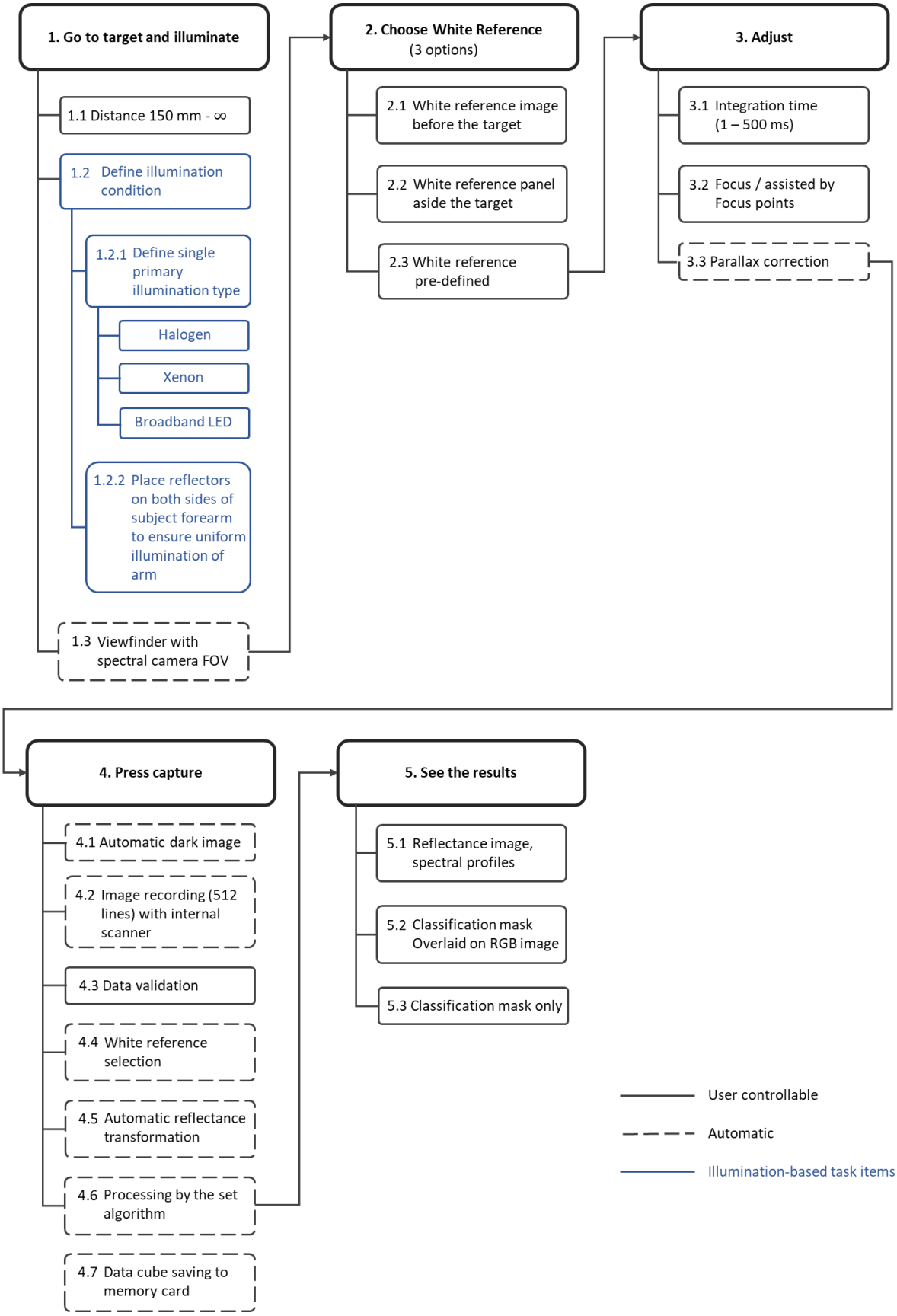


Figure 3.18: Validated imaging protocol informed by prescription factors for the acquisition of reproducible spectral signatures from TST skin sites

3.4 Discussion

This chapter set out to address the question on whether spectral signatures of TST injection sites can be reliably and reproducibly captured using an HSI camera. Thus, the existing Specim IQ imaging protocol was modified, and a new imaging protocol was proposed for capturing spectral signatures of TST sites using the Specim IQ camera. The protocol was validated using 10 participants to provide evidence of the ability to achieve reproducible TST site spectra using an HSI camera. The goal was to present an HSI protocol which also describes factors that positively and negatively spectral reproducibility for the TST sites.

The validation test was done under six simulated illumination conditions. The variation in the white reference spectra characteristics, shown in Figure 3.15(b), highlights the diversity of the illumination conditions. It also highlights the inability of white LED to provide illumination of TST site and other applications in the NIR range. This meant that white LED had poor reproducibility in the NIR range. On the other hand, the Xenon, Halogen, and broadband light sources, in the order of highest to least NIR performance, produced significant NIR intensities. This suggests that they can illicit spectral response in the NIR range from a TST site and are viable full-range VNIR sources for the TST.

The large variation in the characteristics of the raw skin spectra for the same ROI, under the six conditions, shown in Figure 3.15(b), highlights the impact of illumination type on spectral reproducibility. This is further emphasized by similarity in characteristics between the raw spectra in Figure 3.15(b) and corresponding white reference spectra in Figure 3.15(c). The transformation of the raw spectra to a more similar spectral shape across the six conditions, as shown in Figure 3.15(d), highlights the importance of correct calibration to spectral reproducibility.

The mean correlation and spectral angle matrices computed across all 10 participants, as shown in figures 3.16(a) and 3.16(b) suggest a consistent relationship between the skin spectral signatures captured under the different simulated illumination conditions. A correlation score of 1.0 indicates perfect linear correlation between a pair of mean skin spectral signatures, while a spectral angle score of 0 suggests identical direction in the wavelength vector space. Both figures 3.16(a) and 3.16(b) suggest that among the single source type illumination, the Halogen, and the Xenon light sources, produced the most identical skin spectral signatures across all 10 participants with different skin tones. Thus, both sources are interchangeable as a primary illumination source for acquiring reproducible spectral signatures of TST skin sites.

The broadband and the white LED produced less identical skin spectral signatures to the signatures produced by Halogen and Xenon. The broadband LED produced a more similar spectra than the white LED. The high correlation (≥0.98) and low spectral angle (≤0.054) in figures 3.16(a) and 3.16(b),

respectively, show that the skin spectral signatures produced under illumination condition VI was almost identical to the spectra produced by using only Halogen or Xenon light. This is evidence of the effectiveness of the introduction of the flanking reflector in minimisation of the impact of three simultaneous confounding illumination sources. This is also corroborated by the similarity of the spectral shape of the condition VI spectra in figure 3.15(d) to conditions I and IV. Conversely, figures 3.16(a) and 3.16(b) show a lower correlation of 0.9 and a higher spectral angle of 0.19 between skin signatures of condition V and the skin spectra of conditions I and IV, across all participants. This highlights the detrimental impact of a single confounding illumination, without the presence of the flanking reflectors, on spectral reproducibility for TST sites. The confounding ambient illumination in a real clinical setting is bound to be more complex and unpredictable, which could lead to further decrease in spectral reproducibility. The high spectral consistency achieved under the combination of three confounding source types in condition VI suggests that the introduction of flanking reflectors can enable spectral reproducibility in real-world clinical settings.

Visualising the average skin spectra across all conditions for each of the three representative participants, as shown in Figure 3.17(a) highlights the wide margin of error introduced by conditions III and V. This is validated by the significant reduction in the width of the margin of error for all three representatives after the exclusion of conditions III and V, as shown Figure 3.17(b). The relatively higher margin of error from 900 - 1000 nm across the three representative participants in figure 3.17(b) is likely due to the influence of the spectral spike in the 900 - 1000 nm spectral range of broadband LED in figure 3.17(b). This means that caution should be exercised in the interpretation of spectral signatures in this range if this source type is utilised for TST induration imaging.

The results from the simulation of the six illumination conditions informed the generation of the spectral reproducibility factors in Table 3.1 and the validated HSI protocol in Figure 3.18. The finetuned protocol presents Halogen and Xenon light as the ideal primary light types for reproducible spectral from TST sites, with broadband LED as a good alternative. It also presents the introduction of flanking reflectors as a new procedure to minimize the impact of confounding illumination.

A limitation of the protocol could be that other illumination and non-illumination conditions were not simulated during its validation tests. Furthermore, the test was not done with a larger representation of each skin one and a larger cohort size. Nonetheless, the result of the controlled validation tests presented here is evidence that the finetuned imaging protocol can enable the acquisition of reproducible spectral signatures from TST injection sites.

3.5 Conclusion

The output of this chapter is an imaging protocol for the acquisition of reproducible spectral signatures of TST skin sites. The 3D printability of the enclosure of the protocol can facilitates reproducibility of results by other researchers for the TST or other medical HSI applications involving limbs. This imaging protocol will be utilized in the next chapters for the acquisition of the spectral signatures of real-world TST indurations towards the goal of image-based diagnosis of LTBI.

CHAPTER 4

Hyperspectral Imaging for Induration Diameter Estimation in Tuberculin Skin Tests

---

Disclaimer: The contents of this chapter are partly published in the following paper:

II. Oladokun, A. S., Malila, B., Shey, M., & Mutsvangwa, T. (In Print). Hyperspectral Imaging for the Diagnosis of Latent Tuberculosis Infection. In A. Abraham & A. Bajaj (Eds.), *Computational Intelligence based Hyperspectral Image Analysis*. Springer.

This chapter addresses the second research question of this thesis: *Can hyperspectral imaging of TST indurations enable the generation of precise diameter estimations that are comparable to traditional Mantoux readings?* To answer this, the chapter details a model was developed by the primary researcher of this study to generate precise diameter estimations of TST indurations that are comparable to traditional Mantoux readings. The model contributes to the research aim by providing evidence that hyperspectral imaging can generate comparable estimations to Mantoux readings. The model's segmentation masks for induration diameter estimations contribute to the aim of induration characterisation by providing a region of interest for the reaction site of each participant. The model was developed using principal component analysis to project the hypercubes of TST indurations onto principal components that clusters spectral signatures based on variance. The model was validated using the principal components generated from the induration hypercubes of 70 participants captured using the protocol developed in Chapter 3. A support vector machine classification algorithm was utilized to validate the ability of the principal components to predict TST diagnosis of participants. A threshold was applied to the most predictive principal component to segment the induration from the surrounding skin tissue per participant. The diameter of the segmentation mask per participant was estimated and presented as HSI-based induration estimations that are comparable to Mantoux readings. The rest of the chapter is organised as follows. Section 4.1 presents the background and related works. Sections 4.2, 4.3, 4.4, and 4.5 present the methodology, validation procedure, and the results obtained from the TST diameter-generating model.

4.1 Background and Related Works

The findings of the previous chapter demonstrated that, under controlled illumination conditions, a hyperspectral camera can generate consistent spectral signatures from the TST skin site of a participant. This suggests that HSI, when facilitated by the developed imaging protocol, can reliably capture spectral signatures of real TST indurations. Consequently, any downstream readings or extracted features from these spectral signatures should also be reproducible. To confirm this, it was necessary to acquire hyperspectral images of real TST indurations using the previously established protocol.

However, the spectral signatures captured using a hyperspectral camera are not directly comparable to the TST results obtained using the traditional Mantoux method. Therefore, a transformation was required to convert the spectral signatures from a high-dimensional wavelength-based representation into a diameter estimation that aligns with the pen-and-ruler measurement of the Mantoux method. The objective was to demonstrate that HSI is a viable alternative for TST induration measurement, which can produce results that are comparable to those obtained using the Mantoux method, while eliminating the method's inherent limitations.

One of the primary limitations of the Mantoux method is its manual measurement process, which introduces inter- and intra-observer variability. Intra-observer variation in Mantoux readings can reach 3.0 mm, while inter-observer variation can be as high as 3.7 mm (Pouchot et al., 1997). These variations indicate that the Mantoux method suffers from both accuracy and precision issues. By transforming hyperspectral images of TST indurations into a form that allows diameter estimation of the indurations, the precision of the HSI-based method can be assessed and compared to the Mantoux readings.

The structured ENVI data format (Hook et al., 2010; Vibhute et al., 2015) is used to capture and store hyperspectral hypercubes, ensuring standardized digital storage with associated sensor and image metadata. This interoperability across computational systems allows hypercubes from various HSI cameras to be processed consistently, which could support automated LTBI screening.

Given that Mantoux readings lack a universally accepted gold standard, investigating the accuracy of LTBI diagnoses using the proposed HSI method may be impractical. Instead, this study focuses on addressing the precision problem inherent in the Mantoux method. The reproducibility of spectral signatures observed in the previous chapter suggests that the diameter estimations derived from them would also be repeatable, ensuring higher precision. This opens up an opportunity to assess

whether HSI can generate diameter estimations that match Mantoux readings in magnitude but with superior precision.

The concept of generating a diameter estimation of TST indurations using optical imaging modalities has been explored in previous studies. Research has investigated the use of RGB imaging to produce a diameter estimation comparable to the Mantoux method. Specifically, past studies have examined RGB imaging for segmenting TST indurations and measuring their diameters to aid in LTBI diagnosis.

One approach utilized a structure-from-motion technique, capturing RGB images of an induration from seven different angular orientations within a 120° coverage along a single angular axis (Dendere et al., 2017; Naraghi et al., 2018). This approach facilitated the generation of a photogrammetrically derived 3D representation of TST indurations. It was motivated by the hypothesis that such a 3D reconstruction of a TST induration could enable precise and repeatable size measurements comparable to Mantoux readings. In the pilot study by Dendere et al. (2017), mock indurations ranging from 4 to 19 mm in diameter were cosmetically created on the forearm skin of 10 participants. Three-dimensional reconstructions of these mock indurations were generated from RGB images. Two clinicians were tasked with physically measuring the mock indurations as they would in a typical Mantoux test. Additionally, they were required to measure the size of the corresponding 3D reconstructions by manually placing digital landmarks on the perceived induration boundaries and estimating the point-to-point distance between these landmarks. This digitally measured distance between manually placed landmarks on the RGB 3D reconstruction was the diameter estimation proposed by Dendere et al. (2017) as a comparable alternative to the Mantoux reading method. For the first clinician, the intraclass correlation coefficient (ICC) between the Mantoux method readings and the landmark-based point-to-point measurements from the 3D images was 0.965 (95% CI: 0.865–0.991). Similarly, for the second clinician, the ICC was 0.954 (95% CI: 0.830–0.988). These results indicated a high level of agreement between the virtual estimations on the 3D reconstructions and the physical Mantoux readings. Additionally, the ICC between the RGB photogrammetry-based diameter estimations of the two clinicians was 0.990 (95% CI: 0.938–0.998), demonstrating strong interobserver reliability of this measurement approach. However, the requirement for manual landmark placement on perceived induration boundaries in the 3D reconstructed indurations introduces an element of observer subjectivity, potentially affecting the consistency of readings.

This limitation of manual placement of landmarks in 3D images to obtain induration measurements was addressed in (Naraghi et al., 2018). In this study, a sticker with a bounding box was placed around a mock induration before capturing RGB images for the 3D image generation. The bounding box facilitated automated extraction of a ROI of point clouds from a 3D image of a mock induration. Depth

maps were generated from the point clouds to generate a 2D representation of the mock induration and surrounding skin regions. Image processing operations such as histogram equalisation, Otsu binarization (Otsu, 1975), and elliptical estimation of induration boundaries were utilized to segment mock induration from surrounding skin in the depth maps. The width of the segmented mock indurations in the depth maps constituted the virtual estimation comparable to physical Mantoux readings. The mean difference between the two readings for seven mock indurations was estimated as 1.1 mm with a standard deviation of 1.0 mm. This suggested that segmentation-based virtual measurement of mock indurations, was within the interobserver error margin of 3.7 mm for real TST indurations reported in (Pouchot et al., 1997).

Maclean (2020) investigated the effectiveness of this approach for real TST indurations and found that it failed to generate consistent segmentations from photogrammetry-enabled depth maps. Unlike cosmetically generated mock indurations (Dendere et al., 2017; Naraghi et al., 2018), the real TST indurations lacked well-defined boundaries and did not exhibit significant height elevation, making segmentation challenging. The cosmetically applied mock indurations in Dendere et al. (2017) and Naraghi et al. (2018) exhibited significant colour and texture contrasts between the induration sites and surrounding skin. These mock indurations also had prominent height elevations, well-defined boundaries, and lacked encapsulating erythema (skin redness). In contrast, the real TST induration RGB images captured by Maclean (2020) displayed less pronounced colour and texture contrasts, lower induration height, and weaker boundary definitions, particularly in darker skin tones. This highlighted the limitation of RGB imaging as a viable non-subjective alternative to the traditional Mantoux readings.

The presence of erythema within and around real TST indurations is a major challenge for induration segmentation using RGB imaging. The injection of tuberculin during TST administration often induces erythema at the injection site, which persists regardless of whether an induration develops (Toivgoogiin et al., 2005). Erythema—characterized by redness caused by capillary dilation, irritation, and increased blood flow—typically results from the trauma of subdermal injection. This erythema contributes to the subjectivity of the Mantoux reading method, as indurations often blend into encapsulating erythema, making boundary identification difficult for clinicians (Pahal & Sharma, 2020). As a result, clinicians frequently rely on tactile cues during Mantoux readings to estimate the subdermal boundaries of indurations. For any optical imaging solution to effectively measure TST indurations, it must be capable of distinguishing indurations from surrounding erythema to produce a diameter estimation that reduces the subjectivity of Mantoux readings. HSI's ability to capture rich subdermal spectral features, compared to the limited three-channel data of RGB imaging, suggests

that it can precisely segment TST indurations from erythema, enabling a non-subjective estimation of real indurations.

Recent studies, including Zang et al. (2024), Parihar et al. (2021), and Akinola et al. (2024), have explored RGB image-based analysis of 66, 26, and 56 TST indurations, respectively—with Akinola et al. (2024) analysing a mix of real and mock indurations. However, these studies did not aim to generate an image-based diameter estimation of real TST indurations comparable to traditional Mantoux readings. Instead, their objective was to assess the effectiveness of:

- A new thresholding method (Zang et al., 2024),
- A set of traditional thresholding methods (Parihar et al., 2021), and
- A deep learning model (Akinola et al., 2024),

in predicting clinician annotations of perceived induration boundaries in RGB images. However, the assumption that clinician annotations on RGB images serve as a reliable proxy for *in vivo* Mantoux readings is debatable, as clinicians traditionally rely on palpation or tactile cues to identify subdermal TST induration boundaries. Thus, in these RGB image annotation-based studies, the proposed induration segmentations were based on the portion of indurations that are visible on the surface of the skin, neglecting the subdermal induration boundaries that may be out-of-reach for RGB imaging. The ability of HSI to capture subdermal features suggests that it may be capable of capturing subdermal boundaries of real TST indurations which could facilitate the generation of a diameter measurement comparable to the Mantoux reading obtained by identifying induration boundaries *in vivo* via palpations or tactile cues.

The use of medical thermal imaging, an alternative to medical optical imaging, for TST indurations has been investigated by Fiz et al. (2015). Medical thermal imaging (also called thermography) refers to the use of long-wave infrared wavelength range (8μm to 14μm) of the electromagnetic spectrum to capture the heat radiation of body tissue (Shaikh et al., 2019). The research by Fiz et al. (2015) is currently the only significant thermography study for LTBI testing. The authors segmented regions believed to be induration from regions believed to be erythema based on the hypothesis that indurations have higher temperatures compared to erythema. The segmentation was also based on the hypothesis that TST indurations are subsumed within a skin region of erythema. The heat signature-based segmentations by the authors present a viable subdermal visualization of TST indurations that could mitigate the subjectivity of the Mantoux method. Long-wave infrared is majorly absorbed by water and lipids in the skin which leads to molecular vibrations in the fluids to produce heat (Cho et al., 2009; Horton et al., 2023). The resulting increase in skin temperature is captured by a thermal camera. Conversely, light in the VNIR range is absorbed by a wide range of molecules in the

skin (including water and lipids) (Vasefi et al., 2016). This suggests that VNIR light reflected by indurations may contain more information than the induration heat signatures. Therefore, medical optical imaging, particularly HSI, has the potential to capture more precise subdermal information and facilitate precise segmentations towards the goal of generating induration diameter estimations.

In the RGB images of real TST indurations in Maclean (2020), each pixel was a vector of three components captured from three wavelength bands – the red, green, and blue. On the other hand, each pixel of the hyperspectral images of real TST indurations captured by the Specim IQ HSI camera (Behmann et al., 2018)is a vector of 204 components generated from 204 wavelength . Such images have been captured using the imaging protocol presented in the preceding chapter. However, the use of HSI for TST induration introduces the challenge of high dataset dimensionality, a problem that is not present in the RGB imaging or thermography approaches of previous studies. Thus, there is a need for a segmentation approach that incorporates dimensionality reduction when segmenting the induration from erythema and normal skin tissue in TST induration images. The subjective nature of the Mantoux reading method prevents clinicians from generating definitive pixel-wise ground truth labels. Furthermore, the lack of objective ground truth labels for TST images transforms image-based segmentation of real TST indurations into a clustering problem. Consequently, the segmentation task involves grouping spectral signatures within a given hypercube into clusters based on their similarities while ensuring that these similarities generalize across all hypercube samples.

Traditional clustering techniques like K-means clustering or agglomerative hierarchical clustering (Karthikeyan et al., 2020) can be utilized for this segmentation task. However, their lack of implicit dimensionality reduction suggest that they may struggle with the high dimensionality of the HSI hypercubes. Principal component analysis is a dimensionality reduction technique that was used by the authors to perform pixelwise clustering across all hypercubes. With PCA, the 204-dimensionality of each pixel of a hypercube was reduced by transforming each pixel into a feature space where only relevant features across all the HSI spectral signatures were preserved. In the PCA feature space, there is a formation of natural clusters of pixels with similar characteristics. Autoencoders are deep learning-based models that can perform the task of dimensionality reduction to facilitate downstream pixelwise clustering of TST hypercubes. However, they require significant large dataset of hypercubes, in the hundreds, in order to generate a feature space that captures salient features. On the other hand, since the goal of this chapter is to delineate induration from erythema and generate a diameter measurement from delineated indurations, it is not apparent that the latent space of an autoencoder would directly cluster indurations in a way that delineates them from erythema while preserving the spatial relationship between induration pixels to facilitate induration diameter measurement. Each spectral signatures from a TST hypercube were considered in this study as a sample, and 40,000

signatures were acquired per participant. This resulted in a sample size of 40,000 signatures multiplied by the number of participants, a, which prevents a high-dimensional, low sample size scenario (Jung & Marron, 2009). Thus, the ability of PCA to capture salient features in high dimensional data that potentially delineate induration from erythema, while preserving spatial relationship between induration pixels to facilitate induration diameter measurement, motivated its selection as the clustering model of choice in this chapter. Furthermore, the option of a few-labelled-sample deep learning technique for HS image classification described in (Jia et al., 2021) is not desired as the unsupervised clustering capability offered by PCA addresses the subjectivity problem of the TST better than a black box classification of hypercubes that deep learning offers.

4.2 Methodology

The dataset of hyperspectral images of real TST indurations utilized in this study was captured from participants recruited in two parallel TST screening studies that were conducted in South Africa and Vietnam. The parallel study in South Africa, with ethics number 031/2019, was conducted by researchers at the University of Cape Town. The parallel study in Vietnam was conducted by the Department of health of Ca Mau Province under the ethics number 31/QD-BVLBP. The TST was administered in the parallel studies to 38 healthcare professionals at the TB ward of Tygerberg and Brewelskloof hospitals, Western Cape, South Africa, and to 32 healthcare professionals at the TB and Lung Disease Hospital of Ca Mau Province, Ca Mau, Vietnam. This resulted in the recruitment of a total of 70 healthcare professionals from the parallel studies into this study. The skin tones of the participants were dark brown (type VI), brown (type V), moderate brown (type IV), and moderate white skin tone (type III), as shown in the metadata presented by table 4.1. The assigned Fitzpatrick skin tone classifications were based on estimations from the RGB images of the participants. The HSI images of the TST reaction sites of the participants were captured by the primary researcher between 48 and 72 hours after the TST was administered in the parallel studies. The setup for the acquisition of TST hypercubes from participants, using previously developed enclosure and Halogen lights, is as shown in Figure 4.1. Human research ethics approvals were obtained by the primary researcher from the UCT HREC (Ethics number 319/2018 and 073/2023) before the participants from both countries were recruited and imaged. The 031/2019 number was the ethics approval for clinicians to administer the TST to the South African participants. The 31/QD-BVLBP was the ethics approval for clinicians to administer TST to the Vietnamese participants. The 319/2018 was the ethics approval to capture hyperspectral images of the TST reaction sites for both sets of participants. The 073/2023 was the ethics approval to utilize the captured images in this thesis. The persistent exposure of the healthcare

professionals to active TB patients increases their likelihood of being LTBI positive. This informed the choice of this cohort from both countries. The consensus threshold (10 mm) for healthcare professionals is the same for many other groups such as recent immigrants from high TB prevalent countries, children under 5 years, residents or employees of high-risk congregate settings (i.e. prisons, homeless shelters), persons with high-risk clinical conditions (e.g. silicosis, diabetes mellitus, chronic renal failure), malnourished individuals, and injection drug users (Al Jahdali et al., 2010). This suggests that the findings of this research should generalise to these groups which constitute a significant proportion of the target population for LTBI screening. Future studies will be required to investigate the generalisability of findings in groups whose TST positive threshold is 5mm. These groups include persons infected with HIV, persons in recent contact with an active TB patient, persons with indications of prior active TB infection, organ transplant patients, and patients on immunosuppressants (Al Jahdali et al., 2010).

Trained clinicians administered Tuberculin to the forearm dermis of the participants. Hyperspectral images of the TST indurations resulting from the administration of the tuberculin were captured within a 48-to-72-hour window after application of the tuberculin. Visual observation of the spectral signatures captured in this chapter revealed that the flanking reflector of the Chapter 3 imaging protocol minimised the effects of the different ambient illumination encountered during image acquisition at the different sites.

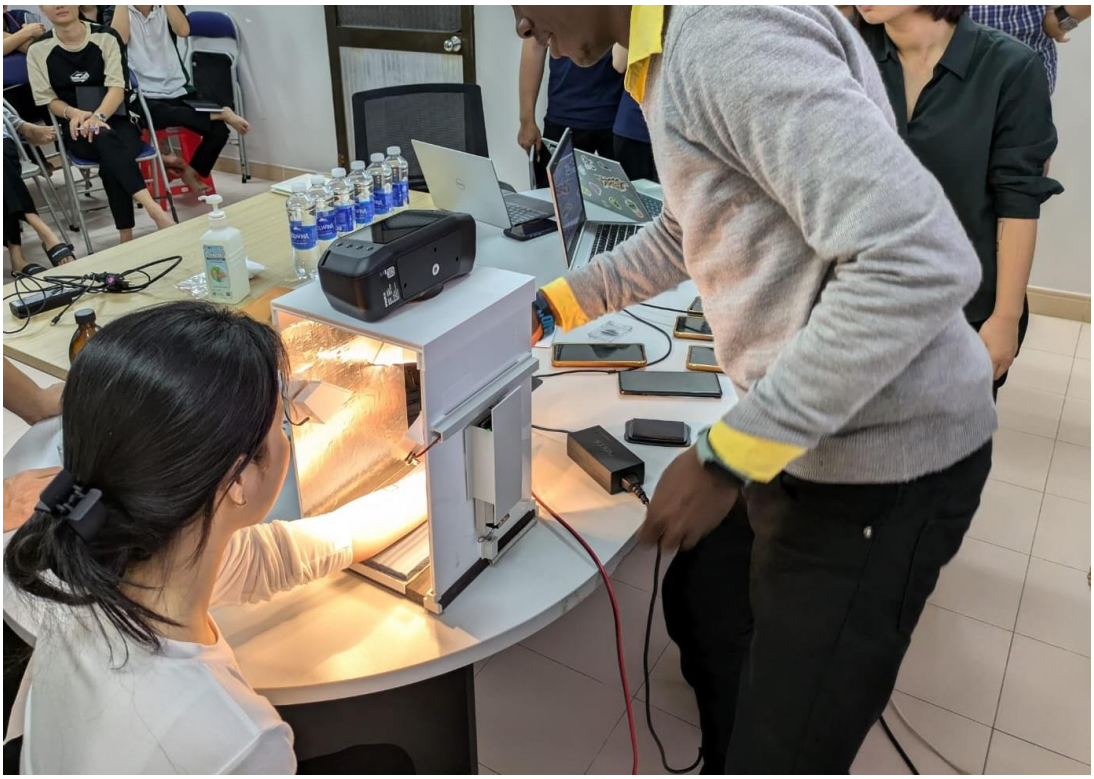


Figure 4.1: Picture showing the acquisition of hypercubes of TST indurations at the study site in Vietnam.

Table 4.1: Metadata of TST Participants

Participant ID	South African Clinician Readings (mm)	Vietnamese Clinician Readings (mm)			Median Readings (mm)	Skin Tone	LTBI Diagnosis
		Clinician 1 (mm)	Clinician 2 (mm)	Clinician 3 (mm)			
SA1	0	-	-	-	0	V	Negative
SA2	0	-	-	-	0	V	Negative
SA3	0	-	-	-	0	V	Negative
SA4	0	-	-	-	0	V	Negative
SA5	0	-	-	-	0	V	Negative
SA6	0	-	-	-	0	V	Negative
SA7	0	-	-	-	0	VI	Negative
SA8	10	-	-	-	10	VI	Positive
SA9	0	-	-	-	0	VI	Negative
SA10	0	-	-	-	0	VI	Negative
SA11	0	-	-	-	0	VI	Negative
SA12	0	-	-	-	0	VI	Negative
SA13	0	-	-	-	0	V	Negative
SA14	0	-	-	-	0	V	Negative
SA15	0	-	-	-	0	VI	Negative
SA16	0	-	-	-	0	V	Negative
SA17	10	-	-	-	10	V	Positive
SA18	0	-	-	-	0	VI	Negative
SA19	8	-	-	-	8	VI	Negative
SA20	0	-	-	-	0	VI	Negative
SA21	0	-	-	-	0	V	Negative
SA22	5	-	-	-	5	VI	Negative
SA23	0	-	-	-	0	VI	Negative
SA24	21	-	-	-	21	VI	Positive
SA25	0	-	-	-	0	VI	Negative
SA26	14	-	-	-	14	VI	Positive
SA27	0	-	-	-	0	VI	Negative
SA28	12	-	-	-	12	VI	Positive
SA29	0	-	-	-	0	VI	Negative
SA30	8	-	-	-	8	VI	Negative
SA31	16	-	-	-	16	VI	Positive
SA32	0	-	-	-	0	VI	Negative
SA33	15	-	-	-	15	V	Positive
SA34	9	-	-	-	9	V	Negative
SA35	12	-	-	-	12	VI	Positive
SA36	11	-	-	-	11	VI	Positive
SA37	0	-	-	-	0	VI	Negative
SA38	15	-	-	-	15	VI	Positive
VTN1	-	12	15	16	15	III	Positive
VTN2	-	23	23	24	23	III	Positive
VTN3	-	20	15	20	20	III	Positive
VTN4	-	18	23	16	18	IV	Positive
VTN5	-	14	12	16	14	III	Positive
VTN6	-	9	9	11	9	III	Negative
VTN7	-	9	7	6	7	III	Negative
VTN8	-	6	6	6	6	III	Negative
VTN9	-	10	9	10	10	III	Positive
VTN10	-	9	10	11	10	III	Positive
VTN11	-	10	10	9	10	III	Positive

VTN12	-	12	10	10	10	III	Positive
VTN13	-	5	6	7	6	IV	Negative
VTN14	-	15	12	12	12	IV	Positive
VTN15	-	6	8	6	6	III	Negative
VTN16	-	5	7	7	7	III	Negative
VTN17	-	4	7	8	7	III	Negative
VTN18	-	12	14	14	12	IV	Positive
VTN19	-	7	8	6	7	IV	Negative
VTN20	-	16	21	20	20	III	Positive
VTN21	-	15	18	21	18	IV	Positive
VTN22	-	4	8	7	7	IV	Negative
VTN23	-	11	17	16	16	III	Positive
VTN24	-	6	7	5	6	III	Negative
VTN25	-	5	8	6	6	IV	Negative
VTN26	-	0	0	2	0	III	Negative
VTN27	-	0	5	2	2	III	Negative
VTN28	-	0	4	3	3	III	Negative
VTN29	-	0	6	6	6	III	Negative
VTN30	-	10	16	12	12	III	Positive
VTN31	-	0	0	0	0	IV	Negative
VTN32	-	0	5	7	5	III	Negative

4.2.1 Hyperspectral Image Capturing Procedure

Prior to imaging, a clinician made markings encircling an area they believed encapsulated the induration region, erythema region, and significant portions of normal skin for each participant. A 200 x 200-pixel square region of interest was cropped around the marked area across all spectral bands for each hypercube, as shown in figure 4.2. Thus, the cropped hypercube of all participants encapsulated the entirety of the TST perimeter as well as significant regions of the skin believed to be outside the test perimeter. The cropped hypercube was assumed to be composed of spectral signatures of skin tissue within the TST perimeter $\boldsymbol{P_{TST}}$ and spectral signatures of skin tissue outside the TST perimeter $\boldsymbol{P_{skin}}$, as shown in figure 4.2.

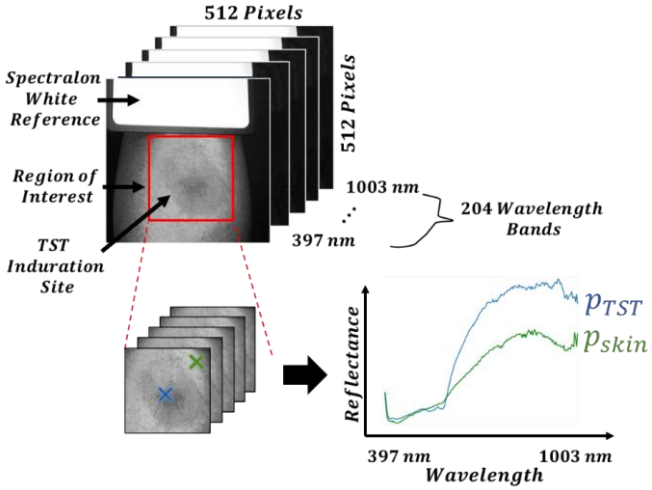


Figure 4.2: Description of the region of interest for TST site

For some of the captured hypercubes, the long axis angle of the participant arm and the marked bounding box did not align with the vertical spatial axis of the hypercube, as shown in figure 4.3(a). In some hypercubes there was a shearing deformation of the bounding box due to the elasticity of the skin, as shown in figure 4.3(b). In both sets of cases, homography transformation, as described in chapter 3, was utilized to rotate each wavelength band image of a hypercube such that the long axis of the participant arm in the image and the bounding box coincide with the vertical axis of the hypercube. This facilitated the extraction 200 x 200-pixel regions of interest from the bounding boxes as shown in figure 4.3.

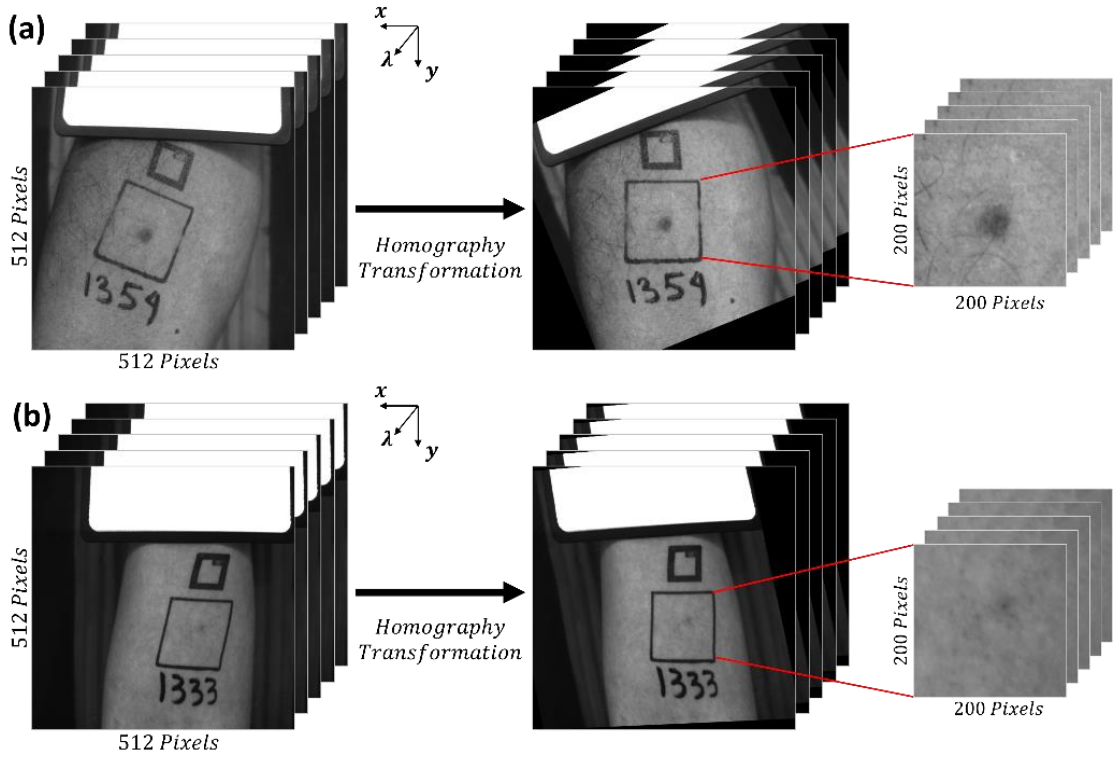


Figure 4.3: Homography transformation applied on hypercubes with (a) rotational misalignment of bounding box and (b) shear deformation of bounding box

The wavelength band images of the cropped hypercube for the TST site of each participant are shown in figure 4.4. The band images around 397 nm and 1003 nm contained less visual information than other bands as they were captured at the tail ends of the camera's sensitivity range. As observed in the figure, similar skin tissue characteristics were captured at each wavelength band across the human participants. An RGB estimate of the imaged scene for each participant was obtained from the hypercubes. The RGB estimates were indicative of what the clinician saw during the TST reading as well as approximation of what an RGB camera would have captured. The Specim IQ camera generated RGB estimates from hypercubes by selecting the 598 nm, 548 nm, and 449 nm wavelength band images as the red, green, and blue channel images respectively, as shown in figure 4.4.

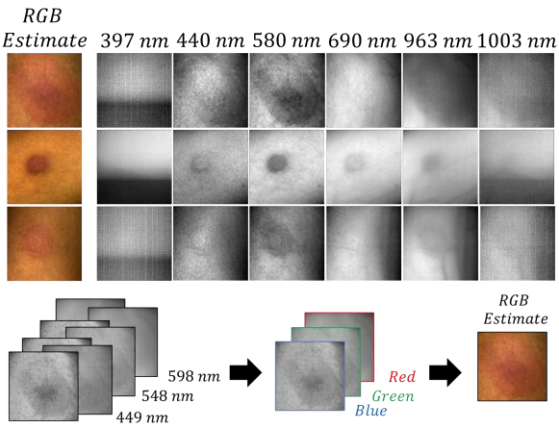


Figure 4.4: Wavelength band images for TST site

A set of Mantoux pen-and-ruler readings, $\boldsymbol{G_{Rx}}$, of 38 participants (from South African sites) were obtained by a clinician. The Mantoux readings of the remaining 32 participants (Vietnam site) were obtained by three different clinicians, as shown in figure 4.5. The three clinicians conducted the Mantoux measurements $\boldsymbol{G_{R1}}$, $\boldsymbol{G_{R2}}$, $\boldsymbol{G_{R3}}$ on a significant subset (32) of the participants in order to evaluate interobserver variation in the readings. A set of median values $\boldsymbol{G_{RM}}$ of the Mantoux measurements was then evaluated for the subset of 32 participants. It was assumed that the median of three TST readings per participant accounts more for interobserver variation than a single clinician reading. The order of the clinician reading was randomised and each of the three clinicians was blind to the reading of the other, to avoid bias. Hyperspectral images of the TST induration sites were captured prior to the clinician readings. The set of single clinician readings, $\boldsymbol{G_{Rx}}$, from the subset of 38 participants and the set of median TST readings, $\boldsymbol{G_{RM}}$, from the subset of 32 participants were combined to form a set, $\boldsymbol{G}$, of 70 Mantoux readings of real TST indurations. The 70 Mantoux readings served as labels for the corresponding hypercubes. Thus, the dataset was comprised of 70 hypercubes of real TST indurations and corresponding hypercube-level labels, $\boldsymbol{G}$.

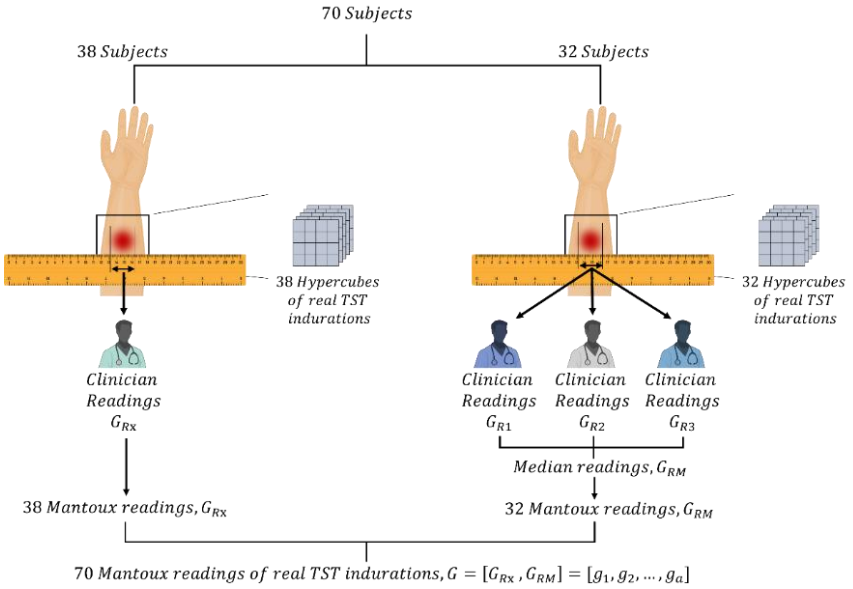


Figure 4.5: Description of the generation of hypercube-level labels for the dataset

Though there were labels for each hypercube, the goal was to obtain labels for each spectral signature (at discrete pixel locations) contained in a hypercube. The problem of subjectiveness of the Mantoux reading was addressed by setting the target labels for each spectral signature per hypercube as "induration", "erythema only", or "normal skin". This enabled a precise definition of the boundaries of an induration using HSI which aimed to give a unique visualisation of TST indurations that neither eye inspection nor RGB imaging can offer. The goal was to accurately generate the labels for each spectral signature in the hypercube and use the hypercube-level label of the participant to validate the accuracy of the generated spectral signature labels. This is an unsupervised learning problem since, as mentioned before, there are no spectral signature (pixelwise) labels to use as the ground truth for the label prediction. The labels offered by the Mantoux reading specifies the presence of a significant induration somewhere within the test perimeter. Indurations with measured diameters of ≥ 10 mm were considered to be significantly sized indurations which would be interpreted as positive LTBI diagnoses for the cohort of the participants.

Principal component analysis was applied to the images of TST sites, which resulted in projecting the hypercubes into a coordinate space that maximized the variance between the constituent spectral signatures to form natural clusters of spectral signatures. These natural clusters included a cluster of spectral signatures that corresponded to skin tissue that contained an induration. The spectral signatures within this natural cluster were ascribed the "induration" pixelwise label, and the boundaries of the cluster were defined as the boundaries of an induration. The unsupervised pixelwise clusters of samples generated by the PCA were validated against the hypercube-level labels. Extracting cluster features and classifying them using the hypercube-level labels were considered as a viable method of estimating the validity of a cluster of induration samples.

The hypercube for each human participant i was defined as $\boldsymbol{H_i}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the spatial axes, and $\boldsymbol{\lambda}$ is the spectral axis of the hypercube. Each hypercube was defined as a $k \times k \times n$ matrix, such that $\boldsymbol{H_i} \in \mathbb{R}^{k \times k \times n}$, and $k = 200$ pixels and $n = 204$ wavelength bands, as shown in figure 4.6. The hypercube-level label for participant i was defined as g_i. Intensity values at pixel position $l = 1, 2, \ldots, k \times k$ across the wavelength bands formed a vector $\vec{\boldsymbol{P}}_{\boldsymbol{l}}$ which was defined as a sample of the spectral signature of the skin tissue at point l on the forearm of participant i, as shown in figure 4.7.

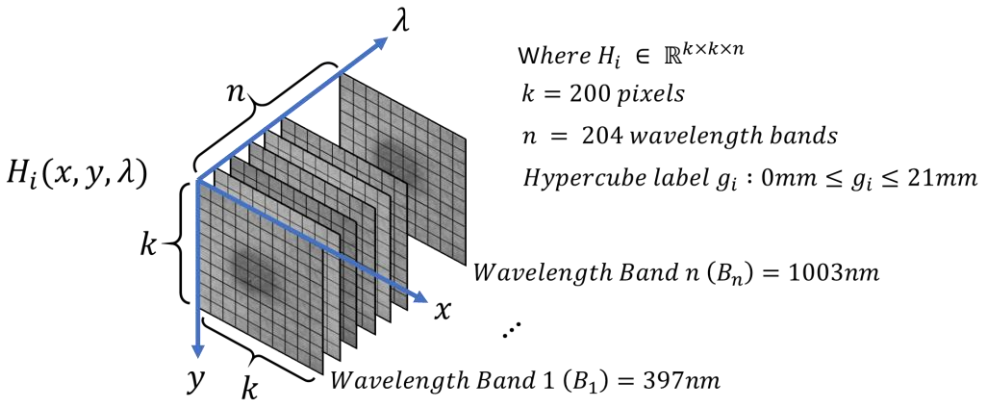


Figure 4.6: Hypercube description

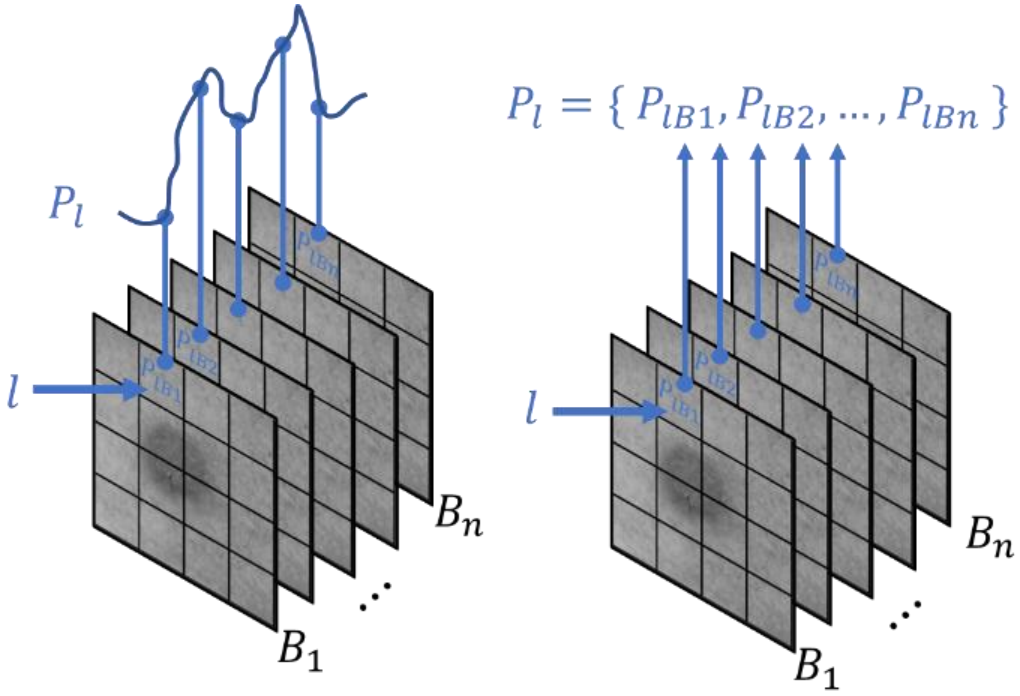


Figure 4.7: Spectral signature description

An 80:20 split of the 70 hypercubes was done to generate a training dataset of 56 hypercubes and a test dataset of 14 hypercubes. The training dataset $\boldsymbol{D}$ contains all skin spectral signature samples obtained from the hypercubes of the 56 training participants across wavelength bands $\boldsymbol{B_1, B_2, \ldots, B_n}$, such that $\boldsymbol{D} = \{ \boldsymbol{H_1} \cup \boldsymbol{H_2} \cup \cdots \cup \boldsymbol{H_{56}} \}$, as shown in figure 4.8.

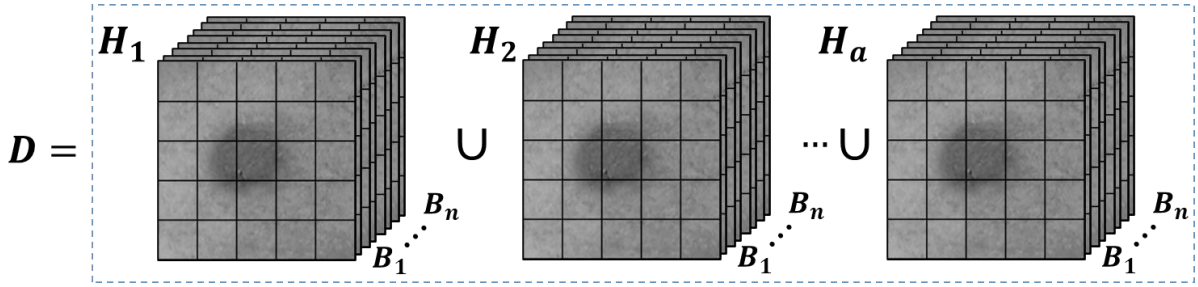


Figure 4.8: Concatenation of hypercubes

Each hypercube consisted of $k \times k$ skin spectral signature samples $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{k}\times\boldsymbol{k}}$. Thus, the training dataset $\boldsymbol{D}$ comprised m samples of skin spectral signatures $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$ across all training participants, where $m = (k \times k) \times a$ and $\boldsymbol{G}$ was the set of hypercube-level labels $\{g_1, g_2, \ldots, g_a\}$. For each hypercube $\boldsymbol{H_i(x, y, \lambda)}$, pixel-wise labels $\{q_1, q_2, \ldots, q_{k\times k}\}$ were derived for the corresponding $(k \times k)$ skin spectral signature samples $\{\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{k}\times\boldsymbol{k}}\}$, as shown in figure 4.9.

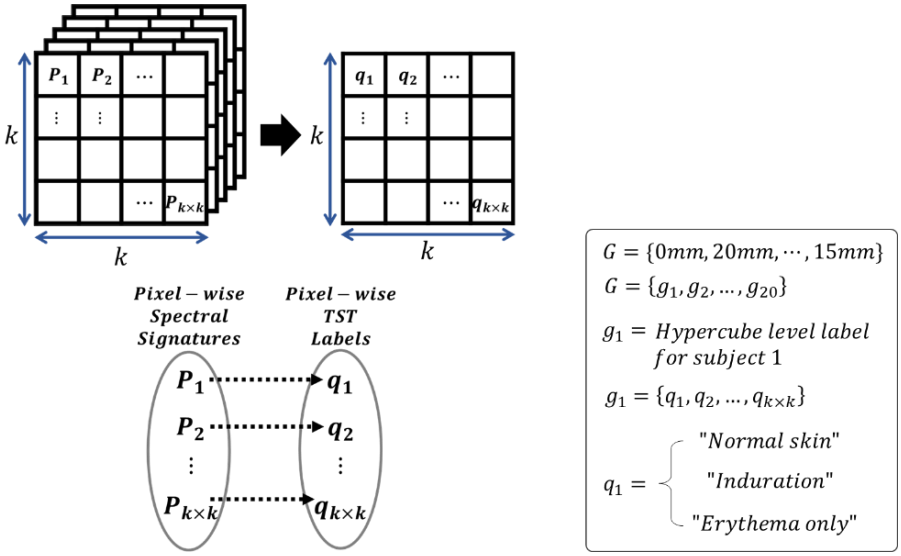


Figure 4.9: Hypercube-level and pixel-wise label definitions

In the training dataset $\boldsymbol{D}$, the nomenclatures of the samples were modified to track which samples belonged to which participant. This was implemented such that spectral signature samples that belong to training participant $i = 1$ were denoted as $\{\vec{\boldsymbol{P}}_{\boldsymbol{1,1}}, \vec{\boldsymbol{P}}_{\boldsymbol{1,2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{1,k}\times\boldsymbol{k}}\}$, and the samples for training participant $i = a$ were denoted as $\{\vec{\boldsymbol{P}}_{\boldsymbol{a,1}}, \vec{\boldsymbol{P}}_{\boldsymbol{a,2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{a,k}\times\boldsymbol{k}}\}$, as shown in figure 4.10, where $a = 56$ for the training set and $a = 70$ for the full dataset. This was linked to corresponding TST labels $\boldsymbol{Q}$. The PCA-transformation done on the dataset is as illustrated in Figure 4.11. To avoid bias of the PCA to wavelength bands in $\boldsymbol{D}$ with higher intensity range, the dataset $\boldsymbol{D}$ was standardised to generate $\widetilde{\boldsymbol{D}}$. All wavelength bands in $\widetilde{\boldsymbol{D}}$, across all participants, had the same intensity range while maintaining their original intra band spatial relationships. The standardisation approach is as described in appendix A. A SVM classification was performed on the principal components to validate the presence of a consistent cluster of induration samples. Further information on the methodology employed for the PCA transformation and SVM classification is provided in appendices A and B, respectively.

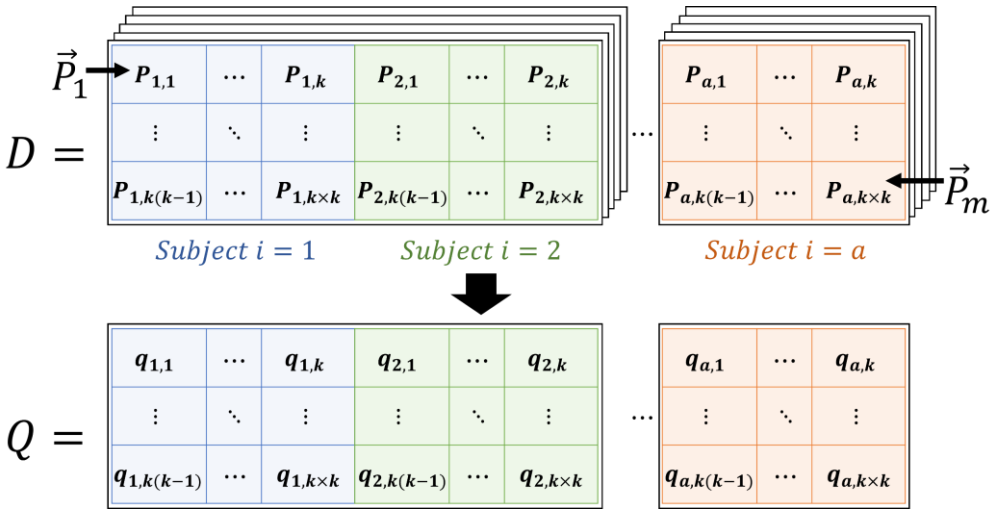


Figure 4.10: Generation of corresponding TST labels for spectral signature samples

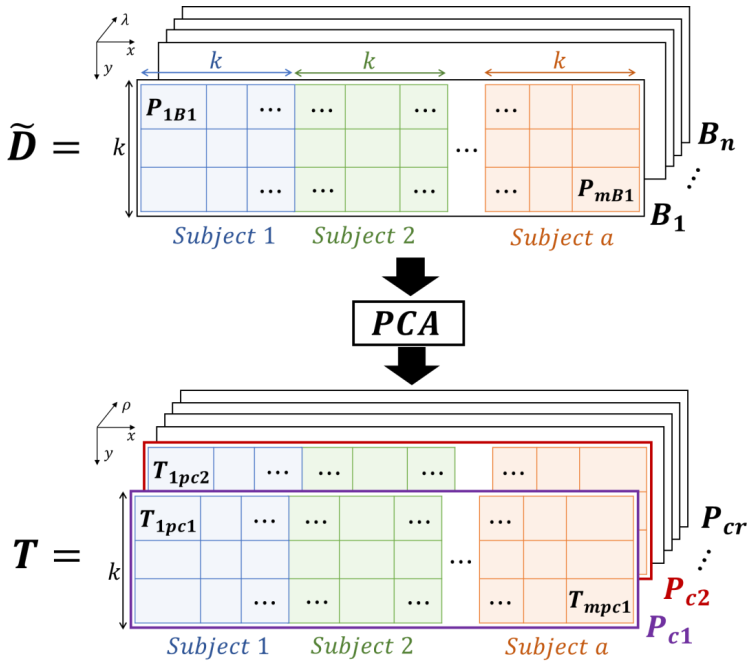


Figure 4.11: Transformation of training dataset from wavelength-based hypercube to an array of PCA-based projections

4.2.2 PCA-Based Erythema Segmentation

Based on the hypothesis that indurations are always subsumed within erythema, the model proposed in this study was configured to segment erythema samples from normal skin samples. As seen in Hossain et al. (2022), erythema, which is characterized by oxyhaemoglobin, has two absorption peaks within the 500nm and 600nm wavelength window of visible light. This coincides with the green band of visible light. It was assumed that some of the natural clusters produced in $\boldsymbol{T}$ were erythema related. Since the loadings $\boldsymbol{L} = \boldsymbol{L_1}, \boldsymbol{L_2}, \dots, \boldsymbol{L_n}$ of the principal component axes $\boldsymbol{P_{c1}}, \boldsymbol{P_{c2}}, \dots, \boldsymbol{P_{cn}}$ respectively ($n = 204$ wavelength bands), were projections from the wavelength band space, the presence of erythema clusters in $\boldsymbol{T}$ was validated by correlating $\boldsymbol{L}$ with the oxyhaemoglobin wavelength characteristic defined in Hossain et al. (2022). To increase the likelihood of accurate segmentation of

erythema from normal skin samples, the training dataset $\widetilde{\boldsymbol{D}}$ was cropped along the wavelength band axis from $\boldsymbol{\lambda} = 500nm$ to $\boldsymbol{\lambda} = 600nm$ to form the cropped hypercube $\widetilde{\boldsymbol{D}}_{\boldsymbol{500nm,600nm}}$, as shown in figure 4.12. The PCA was applied to $\widetilde{\boldsymbol{D}}_{\boldsymbol{500nm,600nm}}$ to generate the PCA transformation $\boldsymbol{T}_{\boldsymbol{500nm,600nm}}$ which facilitated a clustering and segmentation of erythema samples from normal skin samples.

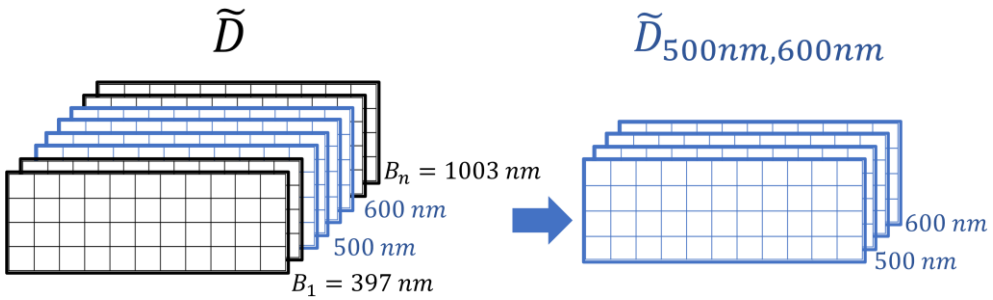


Figure 4.12: Wavelength cropping of HSI dataset to erythema wavelength window

Based on the assumption that the erythema region in the TST skin site captured for each participant is surrounded by normal skin tissue, a bivariate Gaussian kernel-based vignette $\boldsymbol{f}_{\boldsymbol{v}}$, defined in equation (4.1), and also described in (Hamedani & Tata, 1975), was applied to adjacent $(k \times k)$ blocks in each principal component axes of $\boldsymbol{T}_{\boldsymbol{500nm,600nm}}$. Each adjacent $(k \times k)$ block in $\boldsymbol{T}_{\boldsymbol{500nm,600nm}}$ represented a 2D shaped projection in the PCA space for a participant. The means μ_x, μ_y of the bivariate Gaussian kernel mapped the centre of the kernel to the centre of the 2D projection for each participant. The standard deviation σ_v of the kernel was identical for the $\boldsymbol{x}$ and $\boldsymbol{y}$ spatial axes to avoid bias in either direction during the erythema segmentation process. The Gaussian function was passed through a median filter $\boldsymbol{f}_{\boldsymbol{med}}$, to reduce noise, and an intensity thresholding function $\boldsymbol{f}_{\boldsymbol{th}}$ enabled the generation of an array of erythema masks $\boldsymbol{M}$, as shown in figure 4.13. The mask facilitated the segmentation of erythema samples from normal skin samples.

$$f_v(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma^2}} \tag{4.1}$$

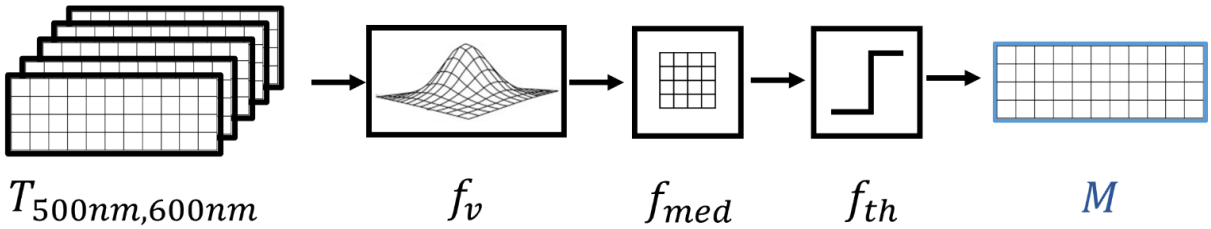


Figure 4.13: Pathway to the generation of erythema mask array

4.2.3 PCA-Based Induration Segmentation from Erythema Region

After segmentation of the erythema samples from the normal skin samples was done, the next goal was to segment the induration samples presumably subsumed amongst the erythema samples.

Though wavelength cropping of the training dataset $\widetilde{\boldsymbol{D}}$ in the $500nm - 600nm$ wavelength window to generate $\widetilde{\boldsymbol{D}}_{\boldsymbol{500nm,600nm}}$ facilitated clustering of erythema samples using PCA, a similar window with significant wavelength characteristics unique to induration samples is not currently known in literature. Thus, the full wavelength range of the training dataset $\widetilde{\boldsymbol{D}}$ was utilized for induration segmentation. Normal skin samples were removed from the training dataset $\widetilde{\boldsymbol{D}}$ to minimise their impact on the likelihood of induration samples having a unique cluster in the principal component space. This was achieved by applying the erythema mask $\boldsymbol{M}$ to $\widetilde{\boldsymbol{D}}$ and performing Hadamard (elementwise) product to generate $\breve{\boldsymbol{D}}_{\boldsymbol{e}}$ – the erythema-masked training dataset. Furthermore, $\breve{\boldsymbol{D}}_{\boldsymbol{e}}$ was projected to the principal component space to generate the PCA-transformed dataset $\breve{\boldsymbol{T}}$, such that the samples formed the "erythema only" and "induration" clusters, and there was maximal delineation of the subsumed induration from surrounding erythema. The complete architecture of the described induration model is as shown in figure 4.14.

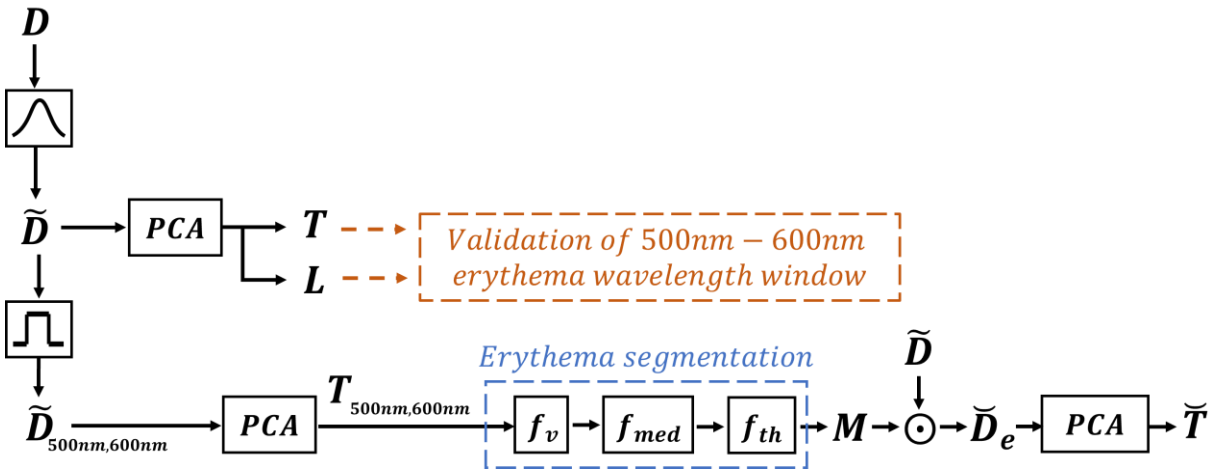


Figure 4.14: Overview of induration segmentation model

4.2.4 Metrics for Benchmarking the HSI Segmentation-based Diameter Estimations for the TST

In this study, the HSI-based estimation for real TST induration was defined as the estimated diameter of the segmentation mask from the PCA transformation. This diameter was obtained by estimating the median value of all diameters, 15 degrees on either side of the axis along which the clinicians take their Mantoux readings, as shown in figure 4.15. The South African clinician in this research obtained Mantoux measurement readings along the long axis of the participant arm while the Vietnamese clinicians measured transversely to the long axis. Studies, such as Nayak and Acharjya (2012), suggests that measuring transversely is potentially the preferred clinical approach. However, this disparity in measurement direction should not impact the results here as a like-for-like comparison was made with whichever direction a clinician used. The centre of the segmentation mask was identified by

calculating the median pixel location of only the white pixels in the binary mask image. Multiple lines whose midpoint coincided with the centre of the mask were generated within a 30-degree span of the clinicians' axis of choice for each participant. The diameter lines terminated at the boundary of the mask and the length of each line was estimated using the Euclidean distance between its end points. The length of each line was translated from Euclidean distance to real-world millimetres by comparing it to the Euclidean distance of fiducial markers with a known real-word dimension in millimetres placed in the imaging scene. The median value among the line lengths, in millimetres, was estimated as the HSI-based diameter estimation of a real TST induration that is comparable to the traditional Mantoux reading.

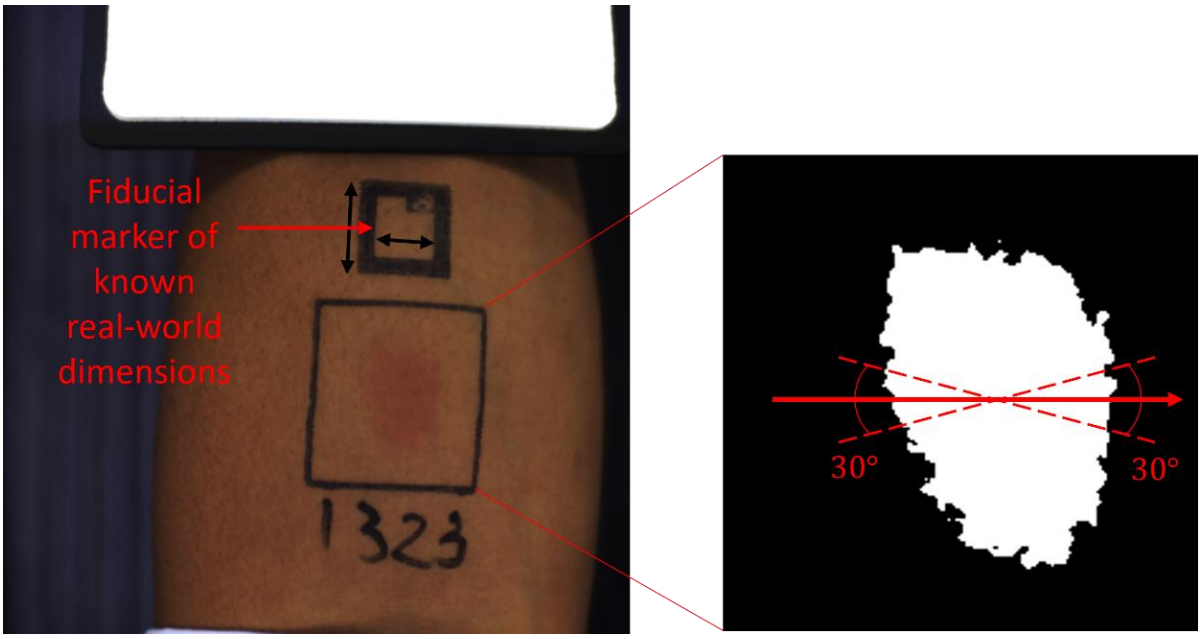


Figure 4.15: Induration mask diameter estimation

Three major studies so far, Dendere et al. (2017), Naraghi et al. (2018), and Fiz et al. (2015), have generated image based diameter estimations for mock or real TST indurations that are comparable to the traditional Mantoux reading. These studies utilized different metrics to evaluate the validity of their diameter estimations as a viable alternative to the Mantoux reading method. It was essential that the HSI-based diameter estimation proposed in this chapter be evaluated using the same metrics. This ensured a fair comparison and facilitated benchmarking against the diameter estimation used in previous studies. The diameter estimation proposed in Dendere et al. (2017) was generated from the manually obtained landmark distances from RGB-based 3D images from the mock indurations of 10 participants. The evaluation metric utilized was the intraclass correlation coefficient calculated between the proposed estimations and the corresponding Mantoux readings. The estimation proposed in Naraghi et al. (2018) was generated from the diameter of the segmentation masks from RGB-based depth maps from the mock indurations of 7 participants. The evaluation metric utilized

was the difference in mean value of the proposed estimations and the mean value of the corresponding Mantoux readings. The estimation proposed in Fiz et al. (2015) was generated from the diameter of the segmentation masks generated from the infrared thermal images of the real indurations of 34 participants. It is the only major study so far that proposed an image-based diameter estimation for real TST indurations that is comparable to the Mantoux method. The evaluation metric utilized was the Pearson's linear correlation between the proposed estimations and the corresponding Mantoux readings.

Studies such as Parihar et al. (2021), Zang et al. (2024), and Akinola et al. (2024), investigated the segmentation of RGB images of 26 real TST indurations, 66 real TST indurations, and 56 mixed TST images (real and mock induration) respectively. All three studies generated ground truths by annotating RGB images of real/mock indurations and used utilized deep learning models or traditional thresholding models to generate segmentation masks that are compared to the annotations. The validity of the annotating RGB images of indurations can be called into question as it is expected that the subdermal boundaries may not be visible in an RGB image, which necessitated the use of palpations or tactile cues by clinicians to identify induration boundaries before performing the Mantoux reading. The evaluation metric utilized by the studies is the DICE similarity coefficient (Zou et al., 2004) which is a spatial overlap index that estimates the level of pixel-wise agreement between a predicted segmentation and its corresponding ground truth. This metric, as utilized in these three studies, compared the generated segmentation masks to annotations that are questionable as ground truths, rather than comparing to the Mantoux reading of a clinician which is currently clinically valid despite its subjectivity.

The HSI-based approach proposed in this study generated diameter estimations that are directly comparable to traditional Mantoux readings. Thus, the metrics utilized for the evaluation and benchmarking were the intraclass correlation coefficient, Pearson's linear correlation coefficient, and the mean/median difference in readings. The ICC quantified the strength of inter-rater agreement. It was used to measure the level of agreement (or reliability) between the Mantoux readings $\boldsymbol{G}$ measured by clinicians and the HSI-based geometric estimations for 70 corresponding real TST indurations. The ICC score was also used to measure the level of agreement between the readings $\boldsymbol{G_{R1}}$, $\boldsymbol{G_{R2}}$, $\boldsymbol{G_{R3}}$ by three clinicians, respectively, for the subset of 32 participants to ascertain the validity of the set of median readings $\boldsymbol{G_{RM}}$. The Pearson correlation was utilized to evaluate the level and direction of linear correlation between the clinician Mantoux readings $\boldsymbol{G}$ and the HSI-based diameter estimations for 70 corresponding real TST indurations. The mean/median difference in readings between the two sets of readings was the third metric reported here. It estimated the deviation between the two sets of readings on average. The median difference was preferred over the

mean difference due to the robustness of the median of a set to outliers, unlike the mean which is more susceptible.

4.3 Results

The PCA-transformed samples for some of the participants in the training dataset, along the first eight principal components $\boldsymbol{P_{c1}}, \boldsymbol{P_{c2}}, \ldots, \boldsymbol{P_{cr}}$, where $r = 8$, are as shown in figure 4.16. The PCA transformation for all participants is presented in the Appendix D. The RGB estimates for the participants, along with the corresponding clinician Mantoux readings, are also visualised in figure 4.16.

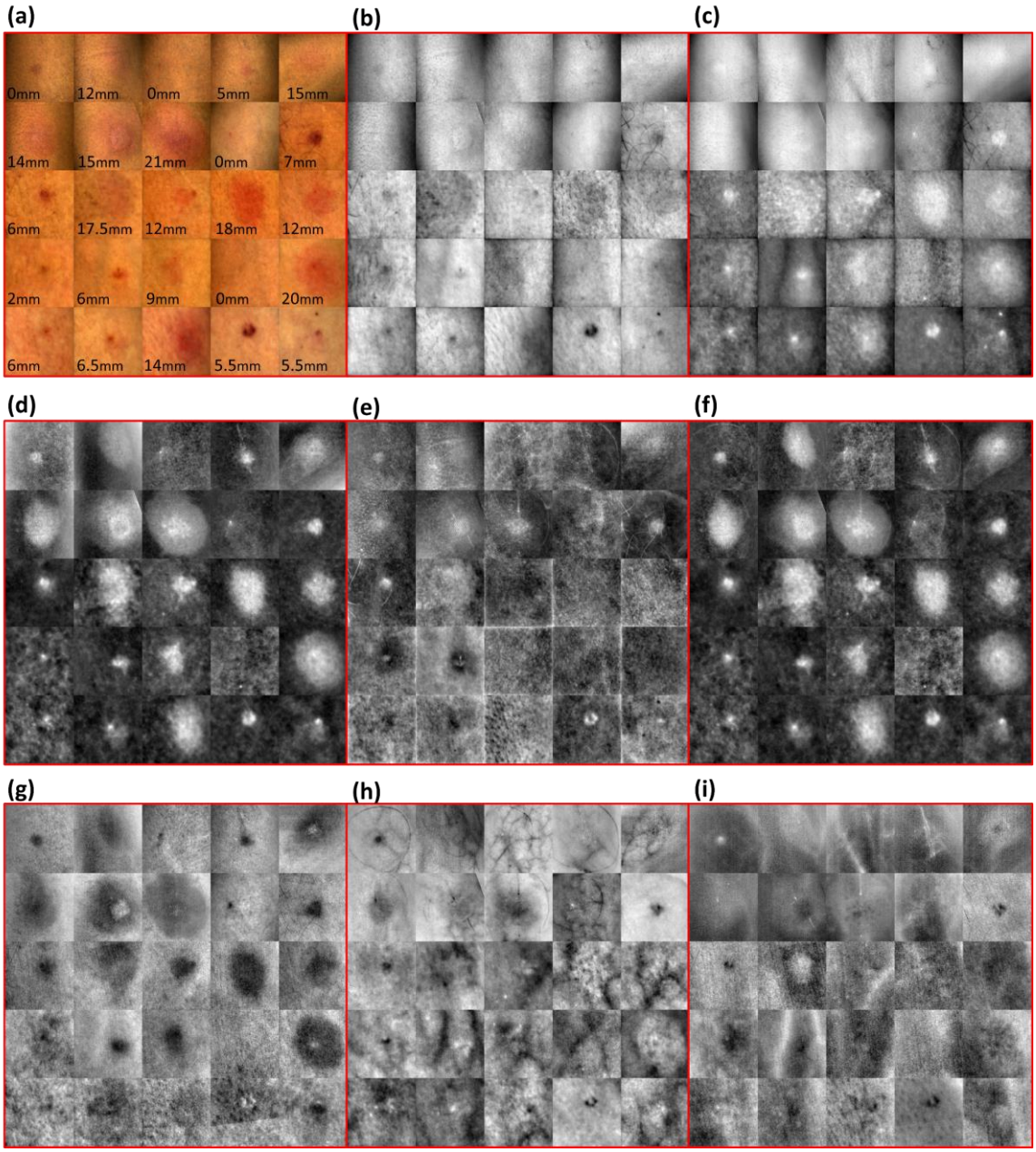


Figure 4.16: Training dataset outputs showing: (a) the RGB images estimated from the hypercubes, and the projection of the corresponding hypercubes to the (b) $\boldsymbol{P_{c1}}$, (c) $\boldsymbol{P_{c2}}$, (d) $\boldsymbol{P_{c3}}$, (e) $\boldsymbol{P_{c4}}$, (f) $\boldsymbol{P_{c5}}$, (g) $\boldsymbol{P_{c6}}$, (h) $\boldsymbol{P_{c7}}$, and (i) $\boldsymbol{P_{c8}}$ coordinates

Based on the ratio of the eigenvalue for $\boldsymbol{P_{c1}}$ to the sum of all eigenvalues, the $\boldsymbol{P_{c1}}$ axis captured 93.6% of the explained variance between the samples in the training dataset. The percentage of explained variances for $\boldsymbol{P_{c2}}, \boldsymbol{P_{c3}}, \ldots, \boldsymbol{P_{c8}}$ range in decreasing order from 5.6% to 0.1%. The sum of the percentages of explained variance for $\boldsymbol{P_{c1}}, \boldsymbol{P_{c2}}, \ldots, \boldsymbol{P_{c8}}$ was 99.9%. This means that the first eight principal components captured 99.9% of all the variations that exists between the training samples. This informed the choice of $r = 8$ and facilitated the dimension reduction of the samples from a 204 −dimensional wavelength band space to an 8 −dimensional PCA space while preserving 99.9% of inter-sample variations. The colourmap representation of the PCA-transformed samples projected to the PC1-PC8 coordinates, for a subset of the training dataset, are as shown in figure 4.17.

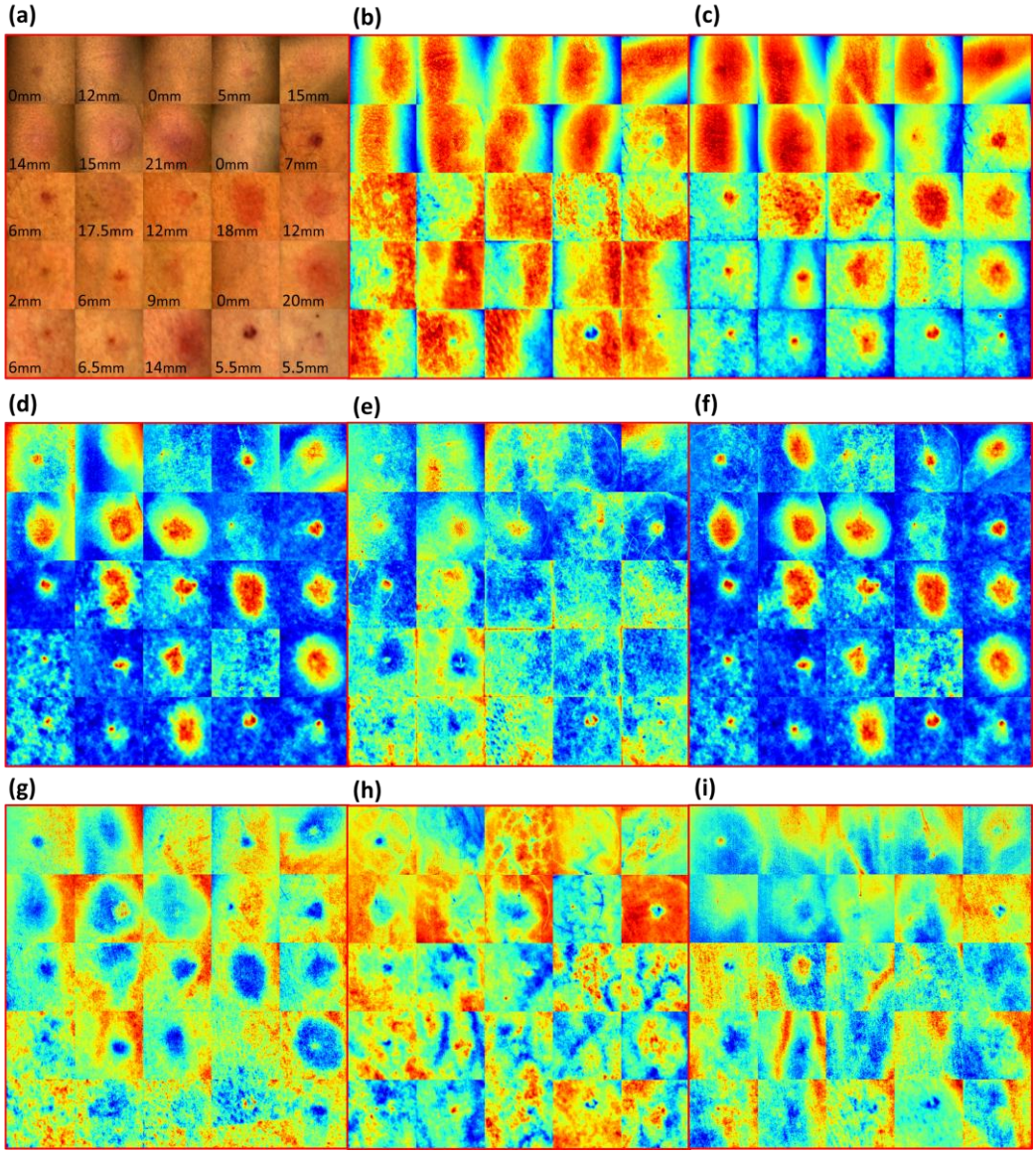


Figure 4.17: Training dataset colourmap outputs showing: (a) RGB images estimated from the hypercubes in the training dataset, and the colourmap visualisations of projection of the training hypercubes to the (b) $\boldsymbol{P_{c1}}$, (c) $\boldsymbol{P_{c2}}$, (d) $\boldsymbol{P_{c3}}$, (e) $\boldsymbol{P_{c4}}$, (f) $\boldsymbol{P_{c5}}$, (g) $\boldsymbol{P_{c6}}$, (h) $\boldsymbol{P_{c7}}$, and (i) $\boldsymbol{P_{c8}}$ coordinates

The colourmap enables easier visualisation of the natural clusters of samples, per participant, along each principal component axis. A visual inspection of the projections in the $\boldsymbol{P_{c1}, P_{c2}, \ldots, P_{c8}}$ axes shows that the samples are clustered uniquely along each principal component axes. In other words, each principal component axes captured a distinct relationship between the samples across all training participants. A visual inspection of $\boldsymbol{P_{c1}}$ and $\boldsymbol{P_{c2}}$,which capture the most significant inter-sample variations, suggests that skin surface features such as skin texture and superficial light reflection (glare) were the most significant variation between the training participants. Visual inspection of $\boldsymbol{P_{c3}}$, and $\boldsymbol{P_{c5}}$ suggests that they cluster the spectral signature samples for the training participants in a way that emphasizes the erythema observable in the RGB estimates across all participants. Visual inspection of $\boldsymbol{P_{c7}}$ and $\boldsymbol{P_{c8}}$ suggests that they capture some variations in the blood vessel patterns between the training participants. A plot of the absolute loadings $\boldsymbol{L = L_1, L_2, \ldots, L_n}$ ($n = 204$ wavelength bands) for $\boldsymbol{P_{c3}}$, and $\boldsymbol{P_{c5}}$ reveal peaks at around 545 nm and 579 nm wavelength bands, as shown in figure 4.18. These peaks coincide with the oxyhaemoglobin absorption peaks defined in Hossain et al. (2022). This validates the observation that $\boldsymbol{P_{c3}}$, and $\boldsymbol{P_{c5}}$ captures the erythema variations between the training participants. The principal component $\boldsymbol{P_{c5}}$ offers a unique clustering of spectral signature samples per training participant, as shown in the colourmap of figure 4.19, that seems to suggests the natural formation of a normal skin cluster (dark blue samples), a "erythema-only" cluster (light green samples), an induration boundary cluster (yellow samples), and an induration peak cluster (red samples).

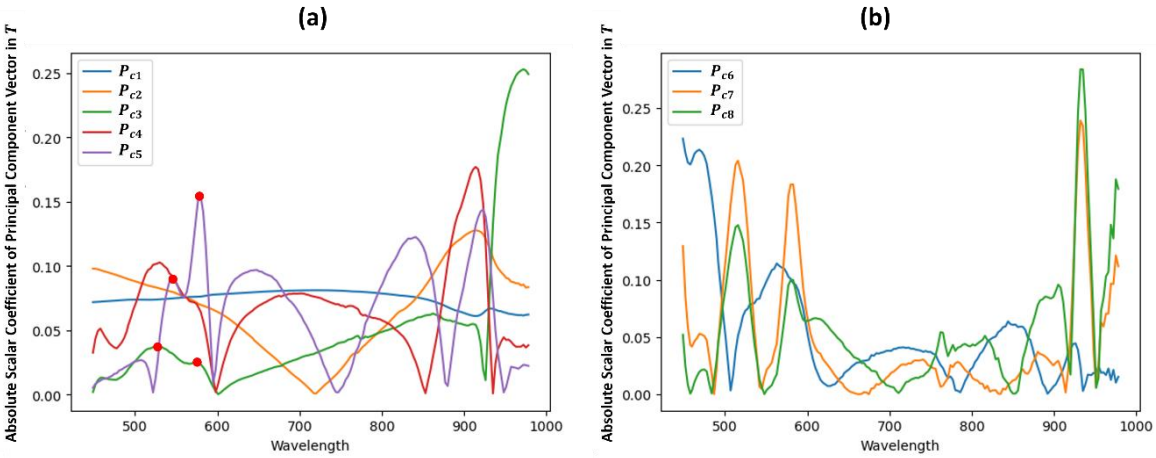


Figure 4.18: Absolute loadings for principal component (a) $\boldsymbol{P_{c1}}$-$\boldsymbol{P_{c5}}$, and (b) $\boldsymbol{P_{c6}}$-$\boldsymbol{P_{c8}}$

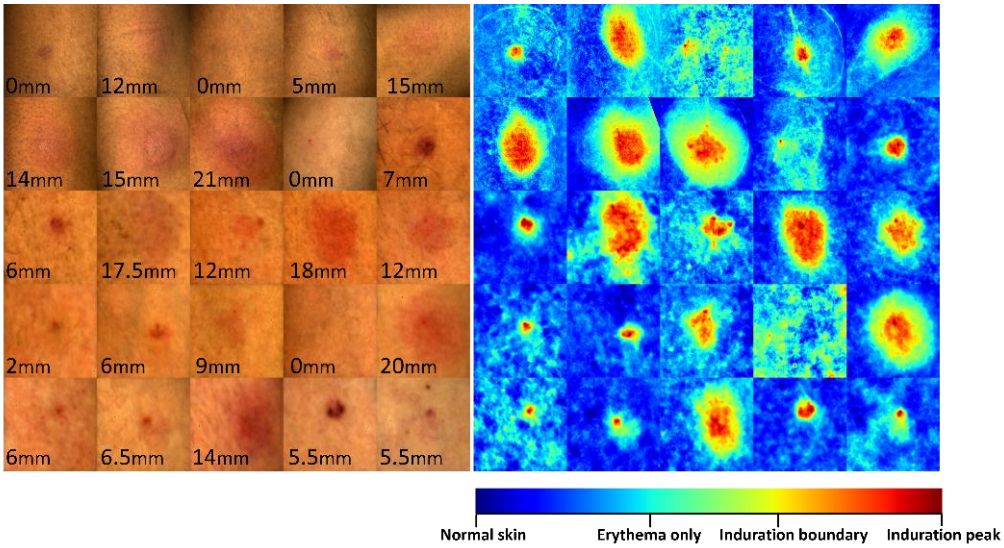


Figure 4.19: Colourmap of $\boldsymbol{P_{c5}}$ for a subset of the training dataset which offers natural clusters of samples that can facilitate erythema-induration separation

The encapsulating region of green-coloured samples in the $\boldsymbol{P_{c5}}$ colourmap of figure 4.19 supports the hypothesis that TST indurations are subsumed within erythema. This facilitated the delineation of the encapsulating erythema, and its constituent sample mixture, from normal skin samples across all training participants to generate the set of erythema masks $\boldsymbol{M}$ shown in figure 4.20.

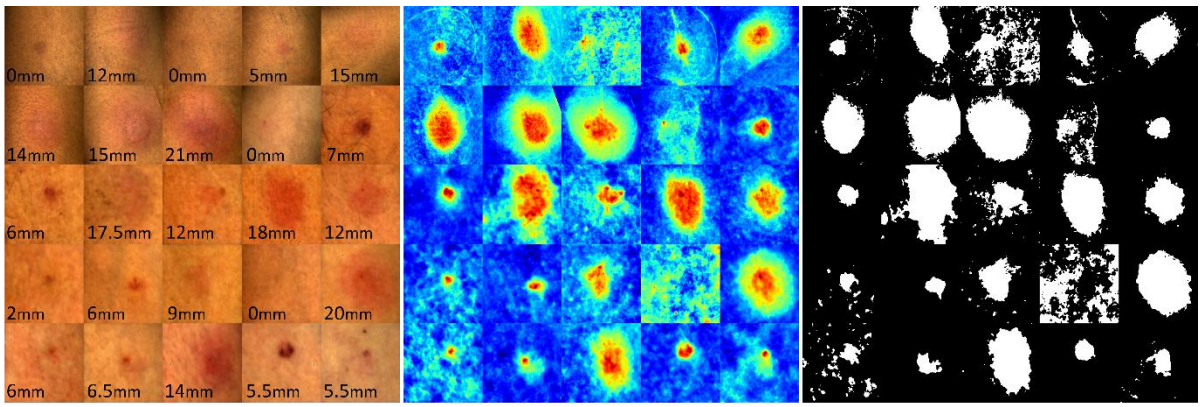


Figure 4.20: Segmentation of erythema from unaffected skin region in a subset of the training dataset

The assumption that the samples in the $\boldsymbol{P_{c5}}$ of $\tilde{\boldsymbol{T}}$ were "erythema-only" samples and induration samples was validated by extracting features from all samples within the erythema region for each principal component projection $\boldsymbol{P_{c1}}$ to $\boldsymbol{P_{c8}}$. The mean and standard deviation of the samples in each of the principal components were extracted as features. The rationale behind the choice of these features was that participants with mostly "erythema-only" samples would have different sample means/standard deviations compared to participants with a significant cluster of induration samples in addition to "erythema-only" samples. These features were validated, with the aid of a soft margin SVM classifier model, to predict LTBI diagnosis informed by the Mantoux readings of clinicians. Mantoux readings ≥10mm were considered LTBI positive. The 5-fold cross validation accuracy of the

samples in $\boldsymbol{P_{c1}}$ to $\boldsymbol{P_{c8}}$ of $\tilde{\boldsymbol{T}}$, as shown in table 4.2, validates the assumption that $\boldsymbol{P_{c5}}$, with the highest cross validation accuracy of 82%, possess the best delineation of erythema and induration samples amongst all the principal components in $\tilde{\boldsymbol{T}}$.

Table 4.2: The SVM-based Cross Validation Accuracies for the Prediction of LTBI using Features Generated from the Principal Components in $\tilde{\boldsymbol{T}}$

	$\boldsymbol{P_{c1}}$	$\boldsymbol{P_{c2}}$	$\boldsymbol{P_{c3}}$	$\boldsymbol{P_{c4}}$	$\boldsymbol{P_{c5}}$	$\boldsymbol{P_{c6}}$	$\boldsymbol{P_{c7}}$	$\boldsymbol{P_{c8}}$
5-Fold Cross-Validation Accuracy	61%	70%	75%	62%	**82%**	68%	73%	70%

The feature plot in figure 4.21 is a visualisation of the PCA-based feature representation of each participant. The PCA-based features for participants with clinically measured indurations ≥10mm formed a recognisable cluster in the feature plot which resulted in the cross-validation accuracy of 82% when the features were fed to a SVM classifier.

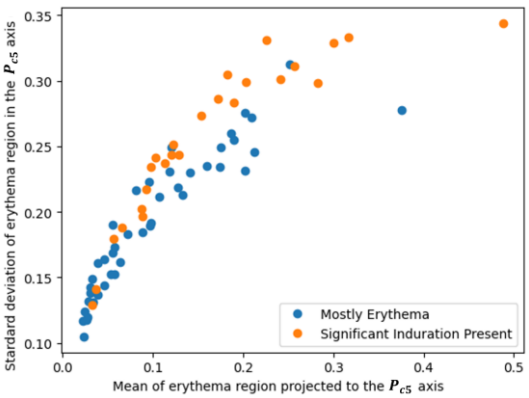


Figure 4.21: PCA-based feature plot of the participants

The cross-validation accuracy validated the assumption that the erythema-masked region in $\tilde{\boldsymbol{T}}$, across all the training participants, encapsulate clusters of induration samples. Thresholding the samples in $\tilde{\boldsymbol{T}}$ resulted in the segmentation of a cluster within the erythema that is believed to be a cluster of induration samples. A single threshold was applied simultaneously to the samples of all participants based on the assumption that induration samples, across all participants, belong to the same cluster in $\tilde{\boldsymbol{T}}$. The segmentation of induration samples from encapsulating erythema samples for the training participants is as shown in figure 4.22. The PCA-informed HSI-based diameter estimation for real TST indurations was generated from the induration mask via the automated measurement of the median value of the diameters of the mask 30 degrees about the clinician's axis of measurement per participant. The automated diameter lines for the induration mask of each participant in the training dataset are as shown in figure 4.23.

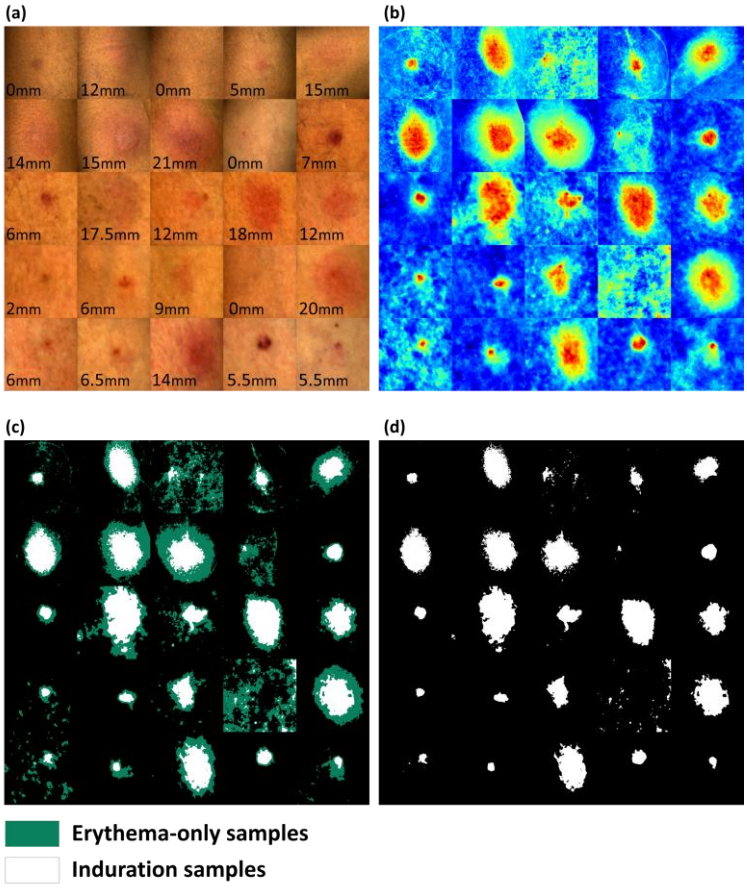


Figure 4.22: Visualisation of the segmentation of induration from erythema for the training dataset showing (a) the RBG estimates, (b) transformation of samples in $\boldsymbol{P_{c5}}$, (c) thresholding of induration samples from erythema, and (d) resulting induration masks

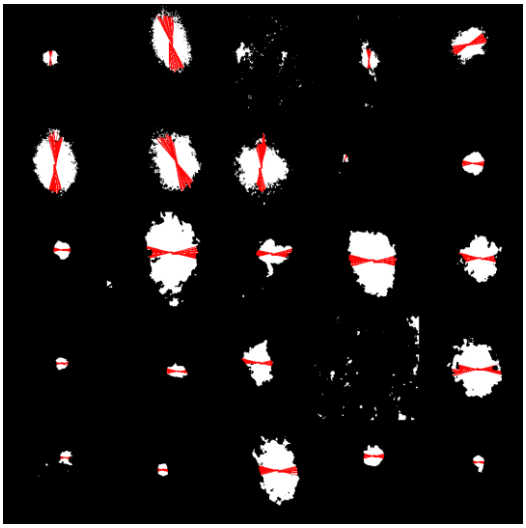


Figure 4.23: Induration masks for the training subset with induration diameter lines 30 degrees about the clinician's axis of measurement per participant

4.3.1 Test Dataset Results

The PCA transformation was applied to the standardised test dataset of 14 participants to generate $\boldsymbol{S}$, which is analogous to $\boldsymbol{T}$. The principal component transformations $\boldsymbol{P_{c1}}$ to $\boldsymbol{P_{c8}}$ in $\boldsymbol{S}$ are as shown in figure 4.24 and figure 4.25.

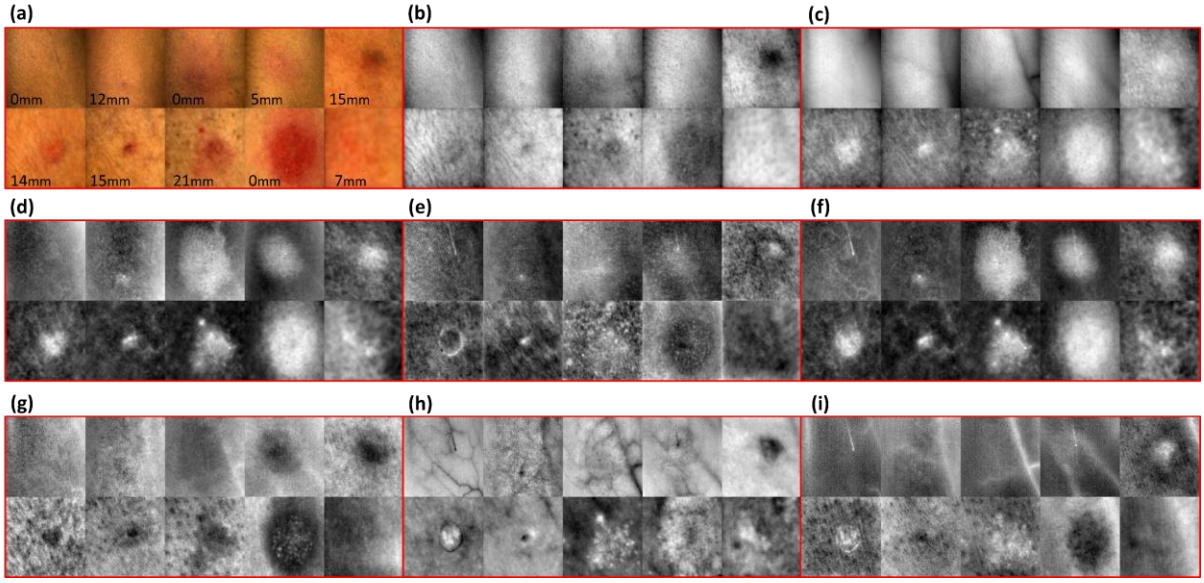


Figure 4.24: Test dataset outputs showing: (a) RGB images estimated from the hypercubes, and the projection of the corresponding test hypercubes to the (b) $\boldsymbol{P_{c1}}$, (c) $\boldsymbol{P_{c2}}$, (d) $\boldsymbol{P_{c3}}$, (e) $\boldsymbol{P_{c4}}$, (f) $\boldsymbol{P_{c5}}$, (g) $\boldsymbol{P_{c6}}$, (h) $\boldsymbol{P_{c7}}$, and (i) $\boldsymbol{P_{c8}}$ coordinates in $\boldsymbol{S}$

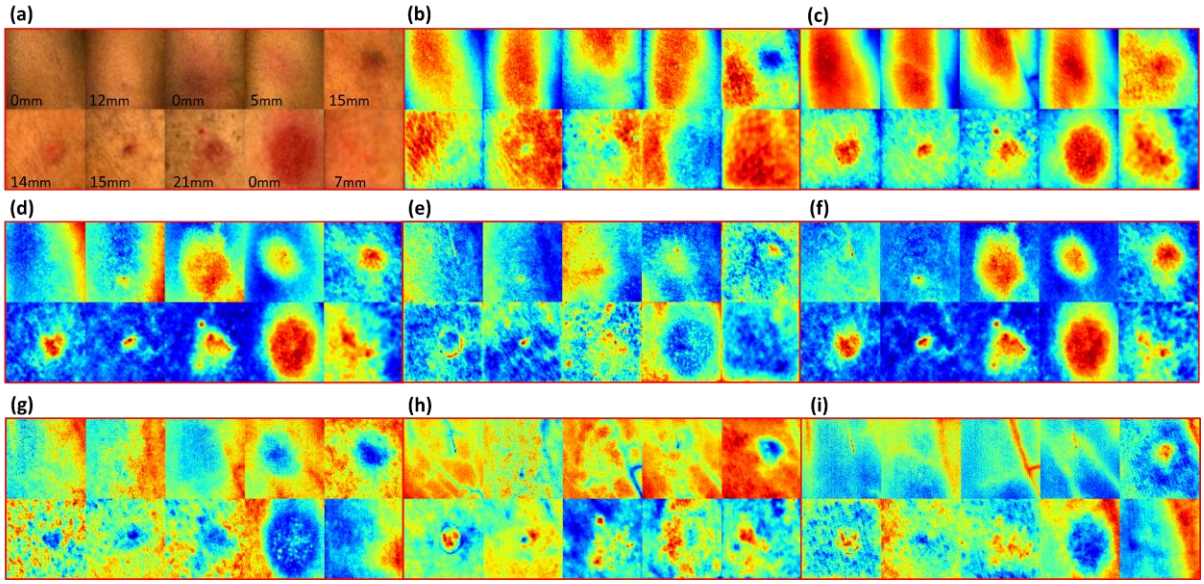


Figure 4.25: Test dataset colourmap outputs showing: (a) RGB images estimated from the hypercubes in the test dataset, and the colourmap visualisations of projection of the test hypercubes to the (b) $\boldsymbol{P_{c1}}$, (c) $\boldsymbol{P_{c2}}$, (d) $\boldsymbol{P_{c3}}$, (e) $\boldsymbol{P_{c4}}$, (f) $\boldsymbol{P_{c5}}$, (g) $\boldsymbol{P_{c6}}$, (h) $\boldsymbol{P_{c7}}$, and (i) $\boldsymbol{P_{c8}}$ coordinates in $\boldsymbol{S}$

These figures show that the characteristics of the sample clusters in $\boldsymbol{S}$ are similar to the characteristics in $\boldsymbol{T}$. The clusters of samples in $\boldsymbol{P_{c5}}$ of $\boldsymbol{S}$ are similar to the $\boldsymbol{P_{c5}}$ clusters in $\boldsymbol{T}$ showing erythema-only, and induration (boundary and peak) samples, as shown in figure 4.26. The generation of $\boldsymbol{S_{500nm,600nm}}$, analogous to $\boldsymbol{T_{500nm,600nm}}$, facilitates the delineation of the encapsulating erythema in the test dataset which results in the erythema masks shown in figure 4.27.

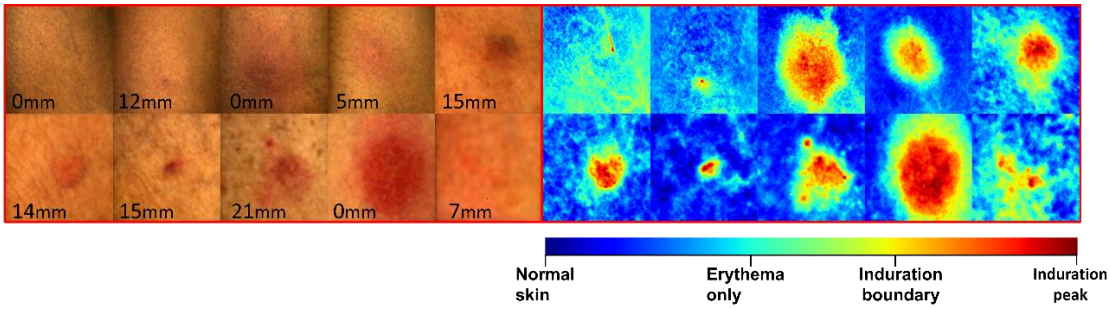


Figure 4.26: Colourmap of $\boldsymbol{P_{c5}}$ in $\boldsymbol{S}$ which offers natural clusters of samples that can facilitate erythema-induration separation

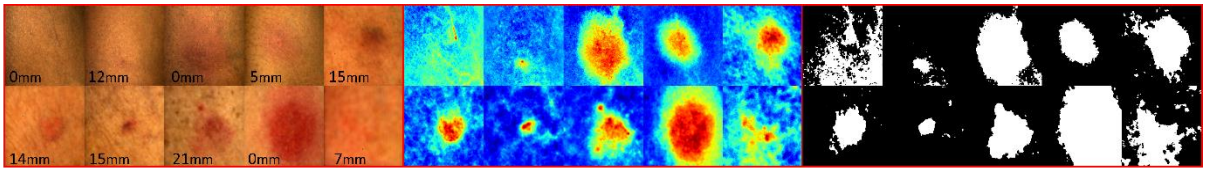


Figure 4.27: Segmentation of erythema from unaffected skin region in the test dataset

To validate the generalisability of the $\boldsymbol{P_{c5}}$ features to predict the presence of a significant clusters of real TST indurations, the SVM classifier models which were cross validated on $\boldsymbol{\breve{T}}$, were tested on $\boldsymbol{\breve{S}}$, to generate the test accuracies shown in table 4.3. The PCA-informed HSI-based diameter estimations for the TST indurations of the test dataset are derived from the induration mask diameters, as shown in figure 4.28 and figure 4.29.

Table 4.3: The SVM-based Test Dataset Predictive Accuracies for the Prediction of LTBI using Features Generated from the Principal Components in $\boldsymbol{\breve{S}}$

	$\boldsymbol{P_{c1}}$	$\boldsymbol{P_{c2}}$	$\boldsymbol{P_{c3}}$	$\boldsymbol{P_{c4}}$	$\boldsymbol{P_{c5}}$	$\boldsymbol{P_{c6}}$	$\boldsymbol{P_{c7}}$	$\boldsymbol{P_{c8}}$
Predictive Accuracies on Test Dataset	57%	71%	71%	71%	**93%**	86%	64%	64%

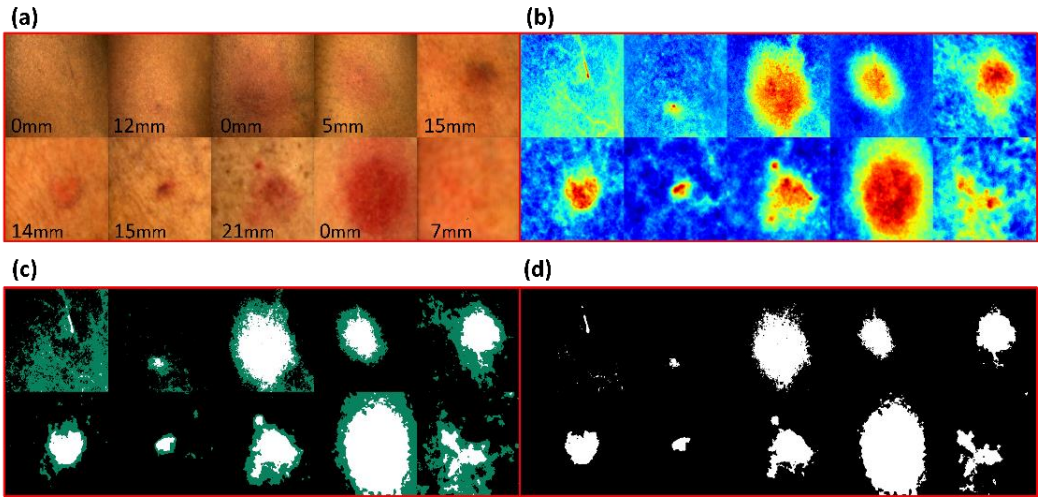


Figure 4.28: Visualisation of the segmentation of induration from erythema for the test dataset showing (a) the RBG estimates, (b) transformation of samples in $\boldsymbol{P_{c5}}$, (c) thresholding of induration samples from erythema, and (d) resulting induration masks

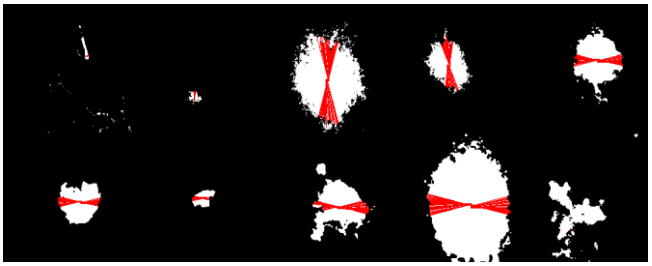
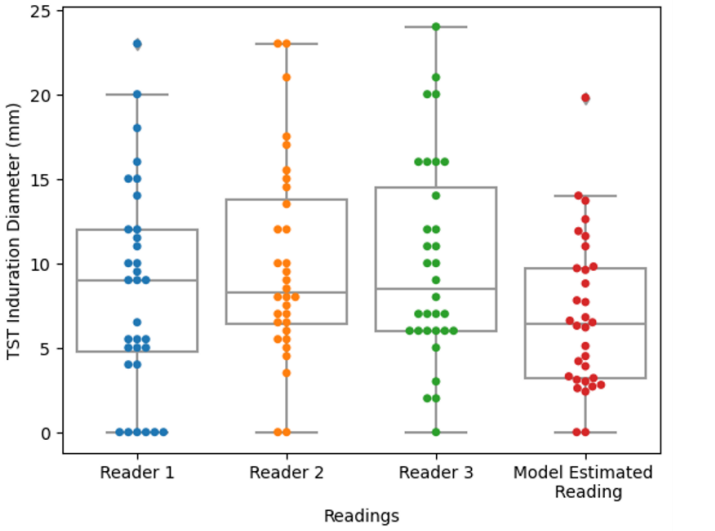


Figure 4.29: Induration masks for the test subset with induration diameter lines 30 degrees about the clinician's axis of measurement per participant

As shown in Figure 4.30, the medians of clinician Mantoux readings $\boldsymbol{G_{R1}}$, $\boldsymbol{G_{R2}}$, $\boldsymbol{G_{R3}}$ for the subset of 32 participants were 9.00 mm, 8.25 mm, and 8.50 respectively. The ICC score was calculated for the three sets of readings to be 0.96 (with a 95% confidence interval of 0.92-0.98). This shows excellent agreement between the Mantoux TST readings $\boldsymbol{G_{R1}}$, $\boldsymbol{G_{R2}}$, $\boldsymbol{G_{R3}}$ of the three clinicians for the subset of 32 participants. This excellent level of agreement validates the reliability of the use of $\boldsymbol{G_{RM}}$, which is the set of median readings from $\boldsymbol{G_{R1}}$, $\boldsymbol{G_{R2}}$, $\boldsymbol{G_{R3}}$, as the representative Mantoux readings for the subset of 32 participants. Figure 4.30 also shows that the HSI-based diameter estimations for the subset of 32 participants is comparable to the clinician Mantoux readings $\boldsymbol{G_{R1}}$, $\boldsymbol{G_{R2}}$, $\boldsymbol{G_{R3}}$.

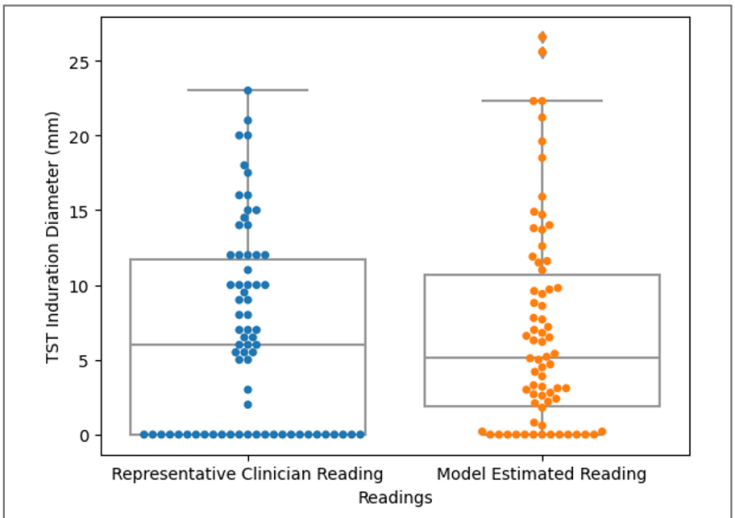


Median Reading:	9.00 mm (Reader 1)	8.25 mm (Reader 2)	8.50 mm (Reader 3)	6.40 mm (HSI Model)
ICC (95% CI):	0.95 (0.89-0.98)			

Figure 4.30: Box plot showing that the HSI-based induration diameter estimations for a subset of 32 participants is comparable to the clinician Mantoux readings $\boldsymbol{G_{R1}}$, $\boldsymbol{G_{R2}}$, $\boldsymbol{G_{R3}}$

The ICC score of 0.95 (with a 95% confidence interval of 0.88-0.98) shows excellent agreement between the four sets of readings. This implies that the HSI model estimations are reliably comparable to the readings by the three clinicians. It also implies that the HSI-model estimations have similar orders of magnitude as the Mantoux readings of the three clinicians. The median value for the HSI-based estimations was 6.40 mm for the subset of 32 participants and the maximum median reading

among the three clinicians was 9.00. Thus, the maximum difference in median TST induration reading for the subset of 32 participants was 2.6 mm which is within the interobserver TST variation of 3.7 mm reported in (Pouchot et al., 1997). This provides more evidence that the proposed HSI-based diameter estimation for real TST indurations is comparable to the traditional Mantoux TST reading. The median representative clinician Mantoux readings $\boldsymbol{G_{RM}}$ for the subset of 32 participants was combined with the single clinician readings $\boldsymbol{G_{Rx}}$ for the subset of the remaining 38 participants to generate $\boldsymbol{G}$, which is the set of single clinician Mantoux readings for the full dataset of 70 participants, such that $\boldsymbol{G} = [\boldsymbol{G_{Rx}}, \boldsymbol{G_{RM}}] = [\boldsymbol{g_1}, \boldsymbol{g_2}, \ldots, \boldsymbol{g_{70}}]$. The box plot in figure 4.31 shows the level of agreement between the HSI-based diameter estimations and the representative clinician Mantoux readings for the real TST indurations of all 70 participants in the dataset. The difference in the median readings was 0.85 mm and in the mean reading was 0.29 mm which are both within the 3.7 mm interobserver variation reported in literature. The Pearson's correlation coefficient between the two readings was 0.664 which implies a positive correlation between the readings. The ICC between the two sets of readings was 0.80 (with a 95% confidence interval of 0.68-0.88) which implies good agreement between the HSI-estimations and the clinician Mantoux readings for the real TST indurations of 70 participants. The agreement between the model estimations and the clinician readings is also reflected in the Bland-Altman plot shown in figure 4.32.



Median Reading:	6.00 mm (Reader 1)	5.15 mm (HSI Model)	0.85 mm (Difference in median)
Mean Reading:	6.74 mm (Reader 1)	7.03 mm (HSI Model)	0.29 mm (Difference in mean)
Pearson's Correlation	0.664 ($p < 0.001$)		
ICC (95% CI):	0.80 (0.68-0.88)		

Figure 4.31: Box plot showing level of agreement between the HSI-based induration diameter estimations and the representative clinician Mantoux readings for the real TST indurations of 70 participants

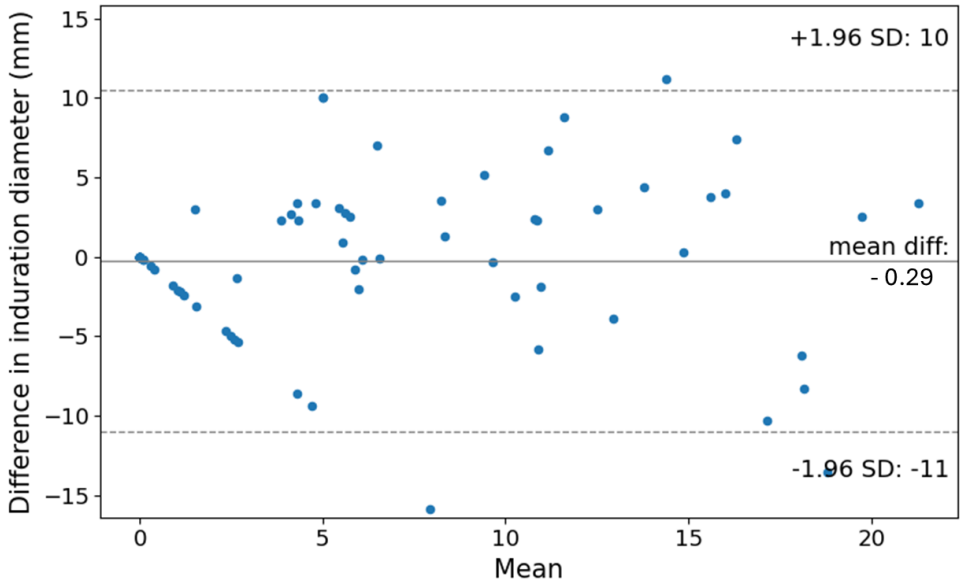


Figure 4.32: Bland-Altman plot showing level of agreement between the HSI-based and the clinician readings

4.4 Discussion

This chapter set out to address the question on whether HSI-based imaging of TST indurations can enable the generation of precise diameter estimations that are comparable to traditional Mantoux readings. Thus, an HSI model, based on PCA, was developed to project the hypercubes of 70 participants – captured using the protocol in Chapter 3 – to a set of principal components. The principal components were validated against clinician Mantoux readings using an SVM classifier. A pixel intensity threshold was applied to the most predictive principal component and the diameter of the resulting segmentation mask was presented as the HSI-based diameter estimation that is comparable to Mantoux readings for real TST indurations.

The proposed HSI model was outlined using the PCA-based architecture in figure 4.14 and implemented on a training dataset of 56 hypercubes of real TST indurations and a test dataset of 16 hypercubes. The implemented model generated an HSI-based diameter estimation of real TST indurations in form of the median value of the 30-degree span diameters of PCA-enabled induration segmentation masks, as shown in figure 4.23 and figure 4.29. The viability of the cluster of the PCA-transformed samples that formed the induration masks was validated against the LTBI clinician diagnoses using an SVM classifier with a 5-fold cross validation accuracy of 82% and test accuracy of 93%, as shown in table 4.1 and table 4.2. The cross-validation accuracy of 82% and predictive accuracy of 93% are comparable to the 5-fold cross-validation accuracy of 77.2% achieved in the HSI melanoma study in Hirano et al. (2020), and the estimated accuracy of 87.43% in Neittaanmäki et al. (2017), which suggests the usefulness of the novel application of HSI for TST to the field of HSI in dermatology.

The viability of the PCA-enabled HSI-based diameter estimation of real TST indurations as a comparable alternative to the traditional Mantoux reading method of clinicians was validated by the ICC, Pearson correlation, and mean/median difference tests reported in figure 4.30 and figure 4.31. The ICC score of 0.80 (with a 95% confidence interval of 0.68-0.88) estimated between the HSI-model estimations and representative clinician readings for 70 participants showed good agreement between the two readings. This suggests that real TST indurations can be reliably measured, with respect to the traditional Mantoux reading, using HSI. Currently, the only TST study that utilized the ICC metric is the study by Dendere et al. (2017). The study achieved an ICC score of up to 0.965 (with a 95% confidence interval of 0.865-0.991) for 3D RGB image-based diameter estimations from 10 mock indurations in reference to corresponding Mantoux readings. Benchmarking the ICC score of 0.80 achieved in this study using HSI, with the 0.965 achieved in Dendere et al. (2017), showcases the usefulness of the proposed HSI-based method. The ICC score of 0.965 achieved (Dendere et al., 2017) was based on mock indurations. There is currently no evidence that a similar ICC score is achievable in real TST indurations using the method proposed in the study. Also, the sample size of 10 participants, compared to the sample size of 70 in this chapter, suggests that the ICC score of 0.965 may be less replicable with a larger sample size. Thus, the ICC score of 0.80 achieved in this chapter for real TST indurations is a first of its kind and marks a good benchmark for future image-based diameter estimations for real TST indurations. The downward trend of the HSI-based estimations compared to the manual readings in figures 4.30 and 4.31 suggests that there may be additional regions of normal skin that clinicians unintentionally include in their readings which contribute the subjectivity of manual readings. Conversely, HSI-based estimations have more precise boundaries which are devoid of additional normal skin regions. The agreement between the model estimations and the clinician readings is also reflected in the Bland-Altman plot shown in figure 4.32. The plot reveals a bias of 0.29 mm which means that on average the HSI-model estimations were 0.29 mm less than the clinician readings. The plot also shows the 95% confidence interval, with most of the reading differences clustered near the limits of agreement.

The study by Naraghi et al. (2018) achieved a mean difference of 1.0 mm in estimations between their proposed 3D RGB depth map-based diameter estimations for the mock induration of 7 participants and corresponding Mantoux readings. Conversely, the HSI-based diameter estimations proposed in this study achieved a mean difference of 0.85 mm and a median difference of 0.29 mm in readings with respect to corresponding clinician Mantoux readings for the real TST induration of 70 participants. This shows that the HSI-based estimations achieves superior precision with real indurations and a larger sample size compared to the study by Naraghi et al. (2018). This positions the

proposed HSI-based induration diameter estimation as a new benchmark in precision for future image-based diameter estimations for real TST indurations.

The study by Fiz et al. (2015) is currently the only study on real TST indurations that is comparable to the HSI-based diameter estimations introduced in this chapter in terms of its applicability with real TST indurations rather than mock indurations, and in terms of validation with respect to Mantoux readings by clinicians. Fiz et al. (2015) achieved a Pearson's correlation coefficient score of 0.65 (with p-value<0.001) between their proposed thermography-based diameter estimation for the real TST indurations of 34 participants in Spain and corresponding Mantoux readings of a clinician. Similarly, the HSI-based diameter estimation proposed in this study achieves a slightly higher Pearson's correlation score of 0.664 (with p-value<0.001) with respect to corresponding Mantoux readings of a clinician for 70 participants. The ability of the proposed HSI-based diameter estimations to slightly outperform the thermography-based estimations in terms of linear correlation, with a larger dataset (70 versus 34 participants), and with a more skin-tone-diverse dataset of real TST indurations from two countries/continents ((South Africa/Africa and Vietnam/Asia) versus Spain/Europe), positions the HSI-based method as a new approach for image based diameter estimations of real TST indurations that is a viable and precise alternative to the traditional Mantoux method.

There have been more recent studies such as Zang et al. (2024), Parihar et al. (2021), and Akinola et al. (2024) which have attempted to perform an RGB image based analysis of 66 real TST indurations, 26 real TST indurations, and 56 real plus mock indurations, respectively. The research objective of these studies was to investigate the ability of a new thresholding method, a set of traditional thresholding methods, and a deep learning model, respectively, to replicate the annotations that a clinician makes on the perceived induration boundaries in RGB images. On the other hand, the objective of this chapter was to develop an image-based diameter estimation of real TST indurations that is comparable to traditional Mantoux readings while being more precise, repeatable, and automatable. These RGB annotation-based studies were not aiming to generate induration diameter estimations that are directly comparable to Mantoux readings. This is corroborated by the lack of the demonstration of agreement or reliability, by any of the three studies, with respect to in *vivo* Mantoux readings using standard reliability metrics such as ICC or Person's correlation coefficient. The assumption that annotations made by clinicians in these studies on RGB images of TST indurations is a good proxy for in *vivo* Mantoux readings is significantly debatable as clinicians traditionally rely on palpations or tactile cues to identify induration boundaries which tend to be subdermal.

Thus, in these RGB image annotation-based studies, the proposed induration segmentations were based on the portion of indurations that are visible on the surface of the skin, neglecting the

subdermal induration boundaries that might be out-of-reach for RGB imaging. On the other hand, this chapter has demonstrated the ability of HSI to capture subdermal boundaries of real TST indurations across a wide range of skin tones from different nationalities. This positions this chapter's novel PCA-enabled HSI-based diameter estimation for real TST induration as a unique benchmark that potentially surpasses similar studies as a viable image-based alternative to the traditional in *vivo* Mantoux method, while offering the digitization, standardisation, precision, repeatability, and automation that Mantoux method currently lacks.

4.5 Conclusion

The results of this chapter, including an ICC score of 0.80, a median difference in reading of 0.29 mm with respect to the Mantoux method, and a Pearson's correlation score of 0.664, demonstrates that HSI can aid the diagnosis of LTBI by providing comparable diameter estimations to the Mantoux method while introducing the digitization, standardisation, precision, repeatability, and automation that Mantoux method currently lacks. This is a significant contribution to the aim of the overall thesis in terms of how it aids the diagnosis of LTBI. However, the aim also includes the characterisation of TST indurations. The HSI-based induration diameter estimation introduced in here addresses the subjectivity problem of the Mantoux method in that the readings are repeatable and do not exhibit the intra-observer variations experienced in the Mantoux method. However, diameter-based readings from TST indurations, whether HSI-based or Mantoux-based, offer a limited representation of real TST indurations which are products complex biological subdermal processes at molecular and cellular levels. The diameter readings of the Mantoux method do not give any significant insight into any of these processes towards a better understanding of underlying LTBI. A skin biopsy is the gold standard method to obtain such insights, but it is an invasive procedure. Hyperspectral imaging has the potential to non-invasively provide insights into some of the biological processes underlying real TST indurations. This is investigated in the next chapter using the 70 hypercubes captured in this chapter and the resulting induration masks. The induration masks generated in here would contribute to next chapter's investigation of the ability of HSI to characterise TST indurations by providing spatial localisation of spectral signatures that correlate more to the presence of induration rather than normal skin tissue.

CHAPTER 5

Hyperspectral Imaging for the Characterisation of Tuberculin Skin Test Indurations

---

Disclaimer: The contents of this chapter are partly published in the following paper:

I. Oladokun, A.S., Malila, B., Campello, V.M., Shey, M., Mutsvangwa, T.E.M. (2024). SpeChrOmics: A Biomarker Characterization Framework for Medical Hyperspectral Imaging. In: Linguraru, M.G., *et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. MICCAI 2024. Lecture Notes in Computer Science, vol 15003. Springer, Cham. https://doi.org/10.1007/978-3-031-72384-1_70

This chapter addresses the third research question of this thesis: *Can hyperspectral imaging non-invasively capture features that are good representations of the underlying physiological responses to tuberculin antigen at a TST injection site?* To address this question, the chapter details the development and validation of a hyperspectral imaging framework capable of identifying biomarkers for accurate and consistent characterization of TST indurations. The framework was developed to generate candidate chromophore maps from the hypercubes and induration masks of 70 participants that were generated in chapter 4. The framework utilizes TST hypercubes, captured using the imaging protocol developed in chapter 3, and the induration masks from chapter 4, to generate chromophore maps that characterise TST indurations – which is the research aim. The unsupervised identification of oxyhaemoglobin, from HSI dataset used in chapter 4, as a potential LTBI biomarker, suggests that the dataset may contain other potential chromophore biomarkers. This makes the dataset also appropriate for the objective of this chapter which is to characterise TST indurations via the identification of chromophore biomarkers. Quantitative features were extracted from the chromophore maps using radiomics techniques to quantify spatial characteristics of the maps. The extracted features were validated against the diagnostic results of TST and IGRA tests of the participants which were obtained by a clinician. This was done to identify chromophores that were good predictors of LTBI. Contribution maps and heatmaps of the radiomics features were generated to provide a visualisation of the most predictive chromophore maps. These visualisations emphasize regions of interests that contribute to high predictive accuracy for LTBI. The most-predictive chromophores were proposed as a good digital representation and characterization of induration pathophysiology. The rest of the chapter is organised as follows. Section 5.1 presents the background

and related works. Section 5.2 presents the methodology and validation procedures, and section 5.3 presents the results obtained from the developed characterisation frameworks.

5.1. Background and Related Works

The previous chapter achieved the objective of developing and validating a hyperspectral imaging-based model for precise diameter estimations for TST indurations that are comparable to traditional Mantoux readings. The ability of HSI to capture subdermal information that facilitates the segmentation of TST indurations from surrounding tissue (erythema and normal skin) was demonstrated. This suggests that HSI may offer insights into the subdermal physiological processes of indurations, providing a method for the characterisation of indurations that is more informative than the Mantoux distance measurement method.

By ensuring digital, standardized, precise, and reproducible characterisation of induration images—similar to the HSI-based distance estimations in Chapter 4—HSI could address the subjectivity problem of the Mantoux TST reading method while noninvasively providing additional physiological insights. These insights could potentially contribute to future understanding of latent tuberculosis infection.

At the time of writing, there are no reported studies that have investigate the use of medical optical imaging to generate non-geometric biomarkers for TST indurations. The 3D reconstruction of TST indurations by Dendere et al. (2017) and Naraghi et al. (2018) did not result in any insights on the pathophysiology of indurations beyond their diameters. Similarly, the RGB imaging approaches of Zang et al. (2024), Parihar et al. (2021), and Akinola et al. (2024) do not offer any additional induration insights beyond segmentation masks.

The medical thermal imaging approach for TST indurations by Fiz et al. (2015) offer additional physiological insights beyond diameter estimations. The study reported the mean temperature of the TST site as $1℃$ higher for positive participants. The total number of participants in the study was 34. The average injection site temperature for the positive group was ($36.2 \pm 1.1℃$) and ($35.1 \pm 1.6℃$) for the negative group. The study stated that the observed temperature increase was a biomarker indicative of elevated blood flow, vascular dilation, oedema, blood proteins, metabolic activity, and cellular extravasations. It postulated that, during induration formation, vasodilation facilitates cellular extravasation – the leakage of red and white blood cells through blood vessel walls to surrounding skin tissue to combat the perceived threat of tuberculin. The combination of vasodilation and cellular extravasations contribute to the development of oedema – accumulation of fluid in the interstitial spaces of tissues. The study postulated that these physiological processes in the formation of TST indurations produce heat as a by-product which can be captured by thermal imaging.

The study demonstrated that imaging in the long-wave infrared range (8,000 – 14,000 nm) of the electromagnetic spectrum can facilitate a characterisation of TST indurations in addition to diameter estimations. There is a need to investigate the potential of the visible and VNIR range (400 – 1000 nm) to generate similar or new characterisation of indurations. Thus, this chapter presents an HSI framework to generate potential VNIR-based biomarkers for TST indurations.

5.2. Methodology

The dataset of hyperspectral images of real TST indurations that were captured in chapter 4 was utilized in the development of the characterisation framework in this chapter. It was assumed that each pixel of the 70 TST hypercubes captured in chapter 4 contained a mixture of φ number of pure spectral components. Each TST site spectral signature was defined as a combination of the pure components (also called endmembers). Spectral unmixing was used to identify the endmembers of the TST site HSI images and estimate their abundances (or concentrations). The endmembers of the TST site were considered as chromophores, which are molecules that absorb light at specific wavelengths to give the appearance of colour in skin tissue (Shukla et al., 2012). The Beer-Lambert law, which defines a proportional relationship between each chromophore and its concentration and optical path length in medium, was utilized to perform spectral unmixing. The law is as described in Calin et al. (2023) and given in equation 5.1.

The product of the raw concentration $\boldsymbol{C_{raw}^{i}(x,y)}$ and the optical path length $\boldsymbol{\ell}$ was defined as the apparent chromophore concentration $\boldsymbol{C^{i}(x,y)}$. The extinction coefficient, which is a unique constant measure that defines the absorptivity of a chromophore at each specified wavelength was denoted as $\boldsymbol{\varepsilon}$. The absorbance representation utilized here was obtained from Calin et al. (2023) and is defined in equation 5.2. The reflectance TST hypercube $\boldsymbol{H_i(x,y,\lambda)}$ for participant i, generated from chapter 4, was transformed into an absorbance hypercube $\boldsymbol{\widehat{H}_i(x,y,\lambda)}$ using equation 5.3 obtained from Calin et al. (2023). Figure 5.1 presents a visualisation of this transformation.

$$\text{Absorbance, } \boldsymbol{\widehat{H}_i(x,y,\lambda) \propto C_{raw}^{i}(x,y) \cdot \ell} \tag{5.1}$$

$$\boldsymbol{\widehat{H}_i(x,y,\lambda) = \varepsilon \cdot C_{raw}^{i}(x,y) \cdot \ell = \varepsilon \cdot C^{i}(x,y)} \tag{5.2}$$

$$\boldsymbol{\widehat{H}_i(x,y,\lambda) = \log\left({}^{1}/_{H_i(x,y,\lambda)}\right)} \tag{5.3}$$

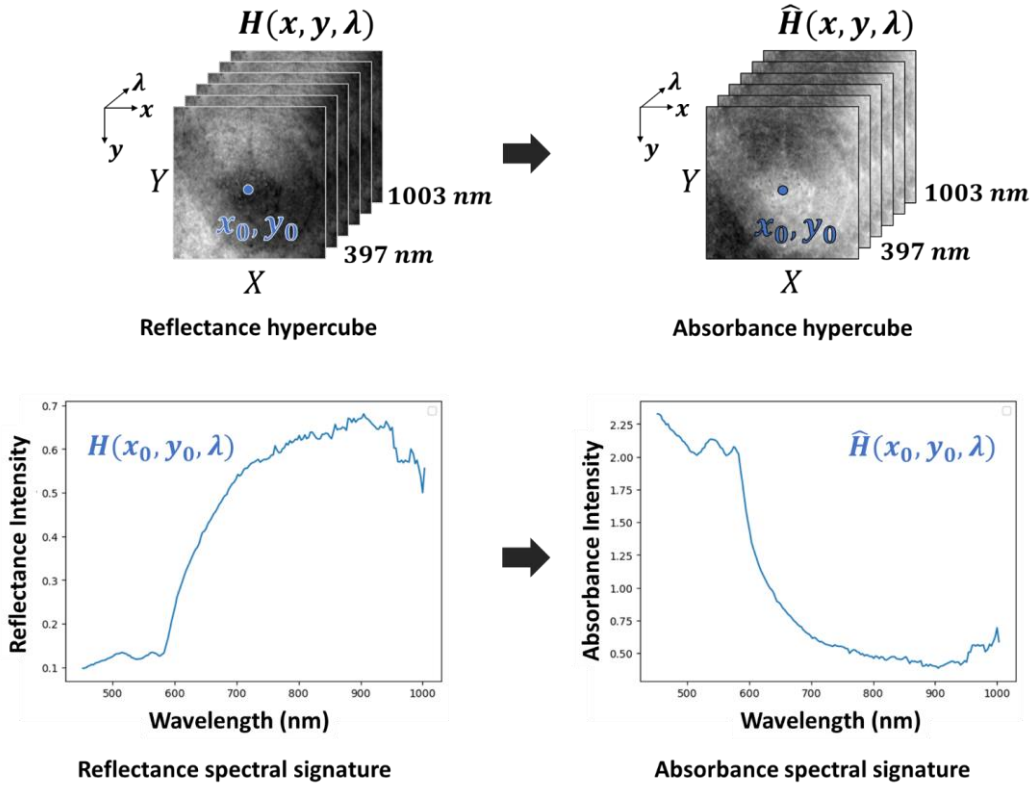


Figure 5.1: Transformation of a reflectance hypercube to an absorbance hypercube

The φ candidate chromophores present on a TST site were regarded as a subset of a library of chromophores $\boldsymbol{C_{lib}}$ which are likely to be present on the site. The array of candidate apparent chromophore maps $\boldsymbol{C_{\varphi}^{i}} = [\boldsymbol{C_{e}^{i}}, \boldsymbol{C_{res}^{i}}] = [\boldsymbol{C_{1}^{i}}, \boldsymbol{C_{2}^{i}}, \dots, \boldsymbol{C_{\phi}^{i}}, \boldsymbol{C_{res}^{i}}]$ were estimated using the Beer-Lambert model. The model was defined here in terms of a non-linear regression function, $\acute{\boldsymbol{g}}$, which estimates the relationship between the independent spatial and spectral variables $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{\lambda}$, and the dependent chromophore map variables in $\boldsymbol{C_{\varphi}^{i}}$, as shown in equation 5.4.

$$\hat{\boldsymbol{H}}_{\boldsymbol{i}}^{\boldsymbol{pred}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = \acute{\boldsymbol{g}}\left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}, \boldsymbol{C_{e}^{i}}\right) + \boldsymbol{C_{res}^{i}} \qquad 5.4$$

The array of ϕ candidate chromophores was composed of an array of specific chromophores $\boldsymbol{C_{e}^{i}}$ and unmodelled residuals $\boldsymbol{C_{res}^{i}}$. The model $\hat{\boldsymbol{H}}_{\boldsymbol{i}}^{\boldsymbol{pred}}$ was considered here as a sum of the products of extinction coefficients $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_{\phi}}]$ and corresponding chromophore maps $\boldsymbol{C_{e}^{i}} = [\boldsymbol{C_{1}^{i}}, \boldsymbol{C_{2}^{i}}, \dots, \boldsymbol{C_{\phi}^{i}}]$, as shown in equation 5.5.

$$\hat{\boldsymbol{H}}_{\boldsymbol{i}}^{\boldsymbol{pred}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = \boldsymbol{\varepsilon_1}(\boldsymbol{\lambda}) \cdot \boldsymbol{C_{1}^{i}}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{\varepsilon_2}(\boldsymbol{\lambda}) \cdot \boldsymbol{C_{2}^{i}}(\boldsymbol{x}, \boldsymbol{y}) + \cdots + \boldsymbol{\varepsilon_{\phi}}(\boldsymbol{\lambda}) \cdot \boldsymbol{C_{\phi}^{i}}(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{C_{res}^{i}}(\boldsymbol{x}, \boldsymbol{y}) \qquad 5.5$$

The unmodeled determinants in the model were captured by $\boldsymbol{C_{res}^{i}}$. A spectral unmixing function $\mathcal{L}$ was defined to transform each absorbance hypercube $\hat{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$ to an array of apparent chromophore maps $\boldsymbol{C_{\varphi}^{i}}(\boldsymbol{x}, \boldsymbol{y})$ with φ channels. The function estimated the candidate chromophore map array that

minimized the squared error between the measured absorbance hypercube $\widehat{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})$ and the modelled absorbance hypercube $\boldsymbol{H}_{\boldsymbol{i}}^{\boldsymbol{pred}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})$, as shown in equation 5.6.

$$\mathcal{L}\left(\widehat{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})\right) = \underset{\boldsymbol{C}_{\boldsymbol{\varphi}}^{\boldsymbol{i}}}{\arg\min}\left(\widehat{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) - \widehat{\boldsymbol{H}}_{\boldsymbol{i}}^{\boldsymbol{pred}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})\right)^{2} \quad 5.6$$

The chromophores that were considered to likely be present at a TST site include oxyhaemoglobin (OX), deoxyhaemoglobin (DX), eumelanin (EM), pheomelanin (PM), bilirubin (BL), water (WT), fat (FT), lycopene (LC), beta carotene (BC), collagen (CG), elastin (ET), and ferritin (FR). These chromophores were considered as the elements of the library of chromophores $\boldsymbol{C}_{\boldsymbol{lib}}$. The selection of OX and DX was informed by the presence of erythema at TST sites. The choice of EM and PM was informed by their influence on the skin tone of participants. The selection of bilirubin was informed by its presence in skin bruises, as reported in Singla et al. (2023). The choice of WT, FT, CG and ET was informed by their presence in the dermis region of the skin, as reported in Vasefi et al. (2016). The selection of LC and BC was informed by their presence in the epidermis region of the skin, as reported in Darvin et al. (2011). The selection of the molecule FR was informed by its property of storing and releasing iron which is vital for haemoglobin synthesis, as reported in Abbaspour et al. (2014). The extinction coefficients of the chromophores in the library $\boldsymbol{C}_{\boldsymbol{lib}}$ are as shown in Figure 5.2.

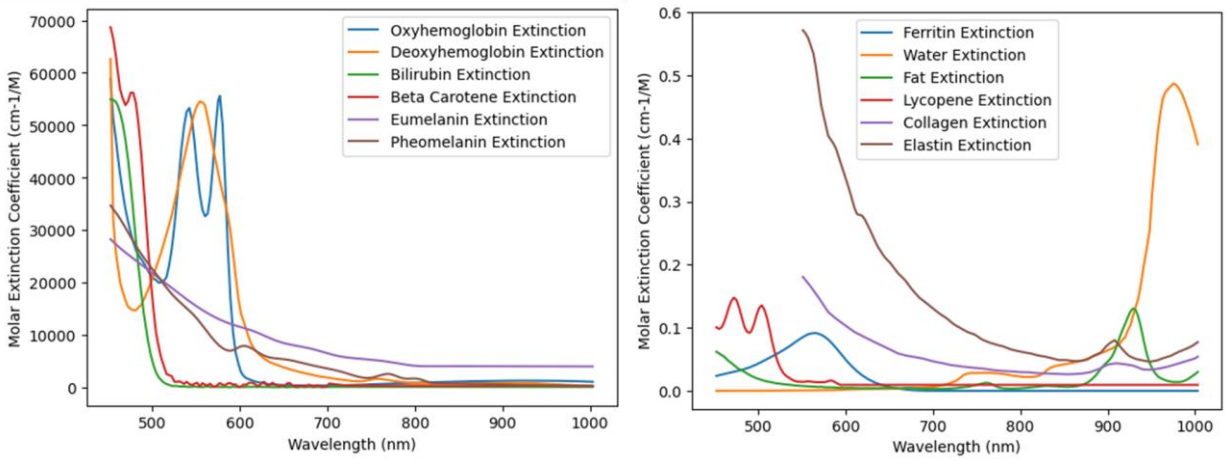


Figure 5.2: Extinction coefficients of the chromophore library

The extinction coefficients were obtained from Prahl (2018) (OX, DX, EM, PM, WT and FT), Prahl (2017b) (BL), Cámara et al. (2012) (LC), Prahl (2017a) (BC), Konugolu Venkata Sekar, Beh, et al. (2017) (CG and ET), and Lai et al. (2015) (FR). The spectrometer devices used in the acquisition of these coefficients measure intensities at wavelengths points different to the wavelengths of the Specim IQ device used for $\widehat{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda})$. Linear interpolation was performed on the raw coefficients to generate intensities at the Specim wavelengths. This is illustrated in Figure 5.3.

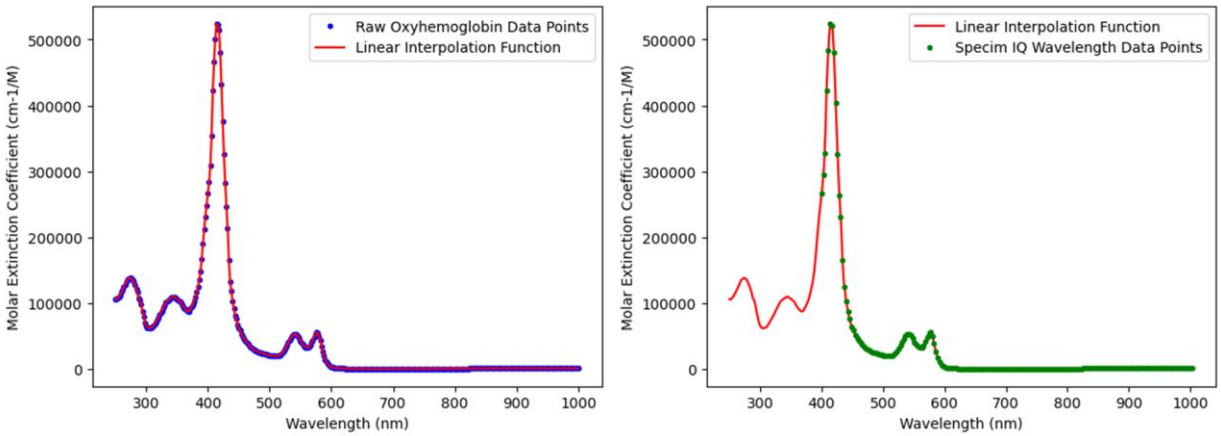


Figure 5.3: Interpolation of Specim IQ wavelengths from raw OX extinction coefficient

For this project, a set of candidate chromophore $\boldsymbol{C_{cand}}$ was generated from the chromophore library as a subset. The candidate chromophores were OX, DX, EM, PM, WT, and FR. Absorbance spectra were randomly sampled from the TST induration region per participant. The sampling was done within the boundaries of the PCA-enabled TST induration masks generated in chapter 4. This is illustrated in Figure 5.4.

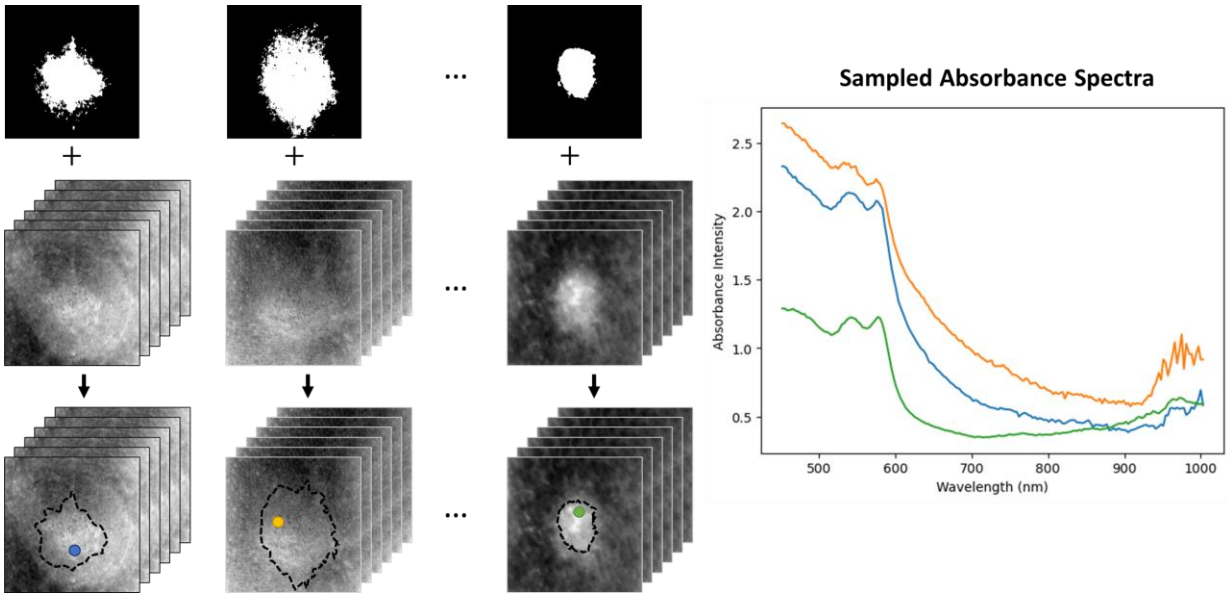


Figure 5.4: Sampling of absorbance spectra for candidate chromophore selection

The sampled spectral signatures were utilized in the selection of candidate chromophores that had significant presence at the induration site across participants. The extinction coefficients of the chromophore library were iteratively included in the model for a single signature in $\widehat{\boldsymbol{H}}_{\boldsymbol{i}}^{\boldsymbol{pred}}$. The selection was informed by the change in median squared error across all participants following the inclusion of each chromophore. Only chromophores that resulted in significant reduction in the MSE were candidate chromophores. The wavelength range of 450 – 1003 nm was utilized since it was the

widest range that minimized the regression error. The change in the error between $\widehat{\boldsymbol{H}}_{\boldsymbol{i}}^{\boldsymbol{pred}}(\boldsymbol{x_0}, \boldsymbol{y_0})$ and $\widehat{\boldsymbol{H}}_{\boldsymbol{i}}(\boldsymbol{x_0}, \boldsymbol{y_0})$ following the inclusion of candidate chromophores is illustrated in Figure 5.5.

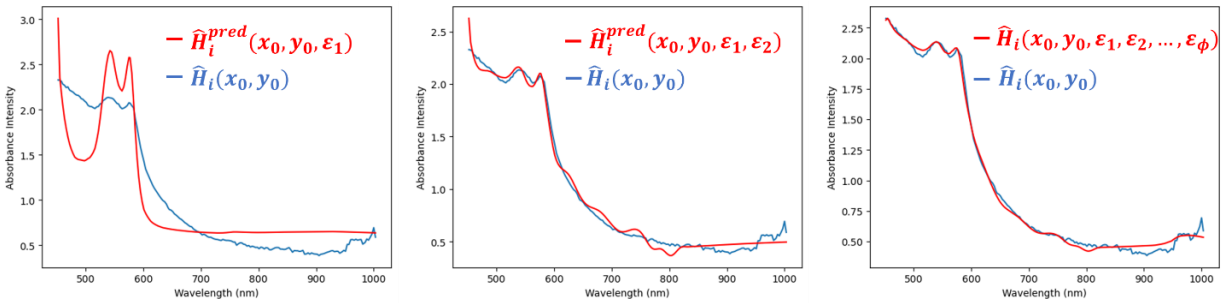


Figure 5.5: Illustration of change in absorbance model following iterative inclusion of candidate chromophores

The extinction coefficients of the candidate chromophores were applied to the absorbance hypercubes per participant to generate the chromophore maps. The $\phi + 1$ number of chromophore maps (including residual determinant) was less than the 185 wavelength bands from 450 – 1003 nm range. This portrays a dimensionality reduction as illustrated in Figure 5.6.

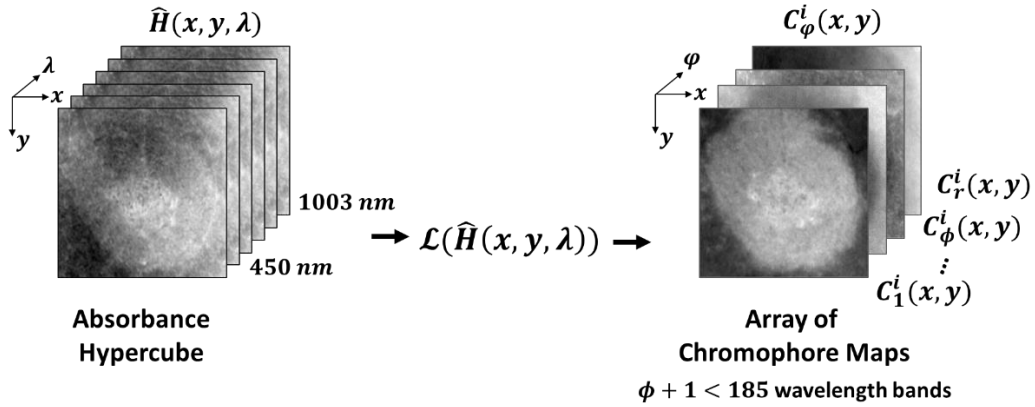


Figure 5.6: Illustration of the generation of an array chromophore maps

5.2.1. Deterministic Radiomics for Spatial Feature Extraction

The goal of the spectral unmixing here was not to compare the estimated intensities between chromophores or between participants. This would have involved calibrating against materials of known concentration per chromophore. Rather, the goal was to extract spatial features per chromophore map that characterise the relationship between normal skin and induration tissue regions within each map. The extracted spatial features per chromophore map were then compared between participants instead of absolute chromophore intensities. Hence, the intensity range of each chromophore map $[\boldsymbol{C_1^i}, \boldsymbol{C_2^i}, \ldots, \boldsymbol{C_\phi^i}, \boldsymbol{C_{res}^i}]$ was normalized to $0.0 - 1.0$. This brought all chromophore maps to the same intensity window and made the extracted spatial features comparable between participants. The normalized intensities for each chromophore map were clustered into bins. The bin

width was 0.0167 which approximately corresponds to 60 bins per chromophore map. This improved the comparability of the spatial features per chromophore map.

The generation of the array of chromophore maps from an absorbance hypercube represented spectral feature extraction. Spatial feature extraction was layered on the spectral feature extraction to distil features from the chromophore maps – an approach introduced for the first time in literature by the primary researcher of this study. The approach was given the pseudonym SpeChrOmics and published in Oladokun et al. (2024). The resulting features amalgamated spatial information and chromophore features. Deterministic (predefined) radiomics functions $\boldsymbol{M^{det}} = \boldsymbol{M_1^{det}}, \boldsymbol{M_2^{det}}, \ldots, \boldsymbol{M_\alpha^{det}}$ were applied to the dataset $\boldsymbol{I}$ of all chromophore maps to generate composite features $[\boldsymbol{F_{m1,\varphi}^{i,det}}, \boldsymbol{F_{m2,\varphi}^{i,det}}, \ldots, \boldsymbol{F_{m\alpha,\varphi}^{i,det}}]$. This is illustrated in equation 5.7. A number of functions, ($\alpha = 94$), were obtained from the Pyradiomics library (van Griethuysen et al., 2017). Table 5.1 illustrates the classification of some of the functions. The full description of all the functions was given in the Chapter 2.

$$\boldsymbol{D^{det}} = \boldsymbol{F_{m,\varphi}^{i,det}} = \left[\boldsymbol{M_1^{det}}(\boldsymbol{I}), \boldsymbol{M_2^{det}}(\boldsymbol{I}), \ldots, \boldsymbol{M_\alpha^{det}}(\boldsymbol{I})\right] = [\boldsymbol{F_{m1,\varphi}^{i,det}}, \boldsymbol{F_{m2,\varphi}^{i,det}}, \ldots, \boldsymbol{F_{m\alpha,\varphi}^{i,det}}] \quad 5.7$$

Table 5.1: Summary of Deterministic Radiomics Functions

Feature Type	**Functions,** $\boldsymbol{M^{det}} = \boldsymbol{M_1^{det}}, \boldsymbol{M_2^{det}}, \ldots, \boldsymbol{M_\alpha^{det}}$
First Order Functions	$\boldsymbol{M_1^{det}} = energy, \boldsymbol{M_2^{det}} = total\ energy, \boldsymbol{M_3^{det}} = \ entropy, \ldots$
Gray Level Co-occurrence Matrix (GLCM) Functions	$\boldsymbol{M_{20}^{det}} = autocorrelation, \boldsymbol{M_{21}^{det}} = joint\ average, \ldots$
Gray Level Size Zone Matrix (GLSZM) Functions	$\boldsymbol{M_{44}^{det}} = small\ area\ emphasis, \boldsymbol{M_{45}^{det}} = large\ area\ emphasis, \ldots$
Gray Level Run Length Matrix (GLRLM) Functions	$\boldsymbol{M_{60}^{det}} = short\ run\ emphasis, \boldsymbol{M_{61}^{det}} = long\ run\ emphasis, \ldots$
Neighbouring Gray Tone Difference Matrix (NGTDM) Functions	$\boldsymbol{M_{76}^{det}} = coarseness, \boldsymbol{M_{77}^{det}} = contrast, \ldots$
Gray Level Dependence Matrix (GLDM) Functions	$\boldsymbol{M_{81}^{det}} = small\ dependence\ emphasis, \ldots,$ $\boldsymbol{M_\alpha^{det}} = large\ dependence\ high\ gray\ level\ emphasis$

5.2.2. Deep Radiomics for Spatial Feature Extraction

Deep learning models were also utilized to extract spatial features from chromophore maps. The deterministic features $\boldsymbol{D^{det}}$ were more explainable as they are based on known functions. On the other hand, the deep radiomics features $\boldsymbol{D^{deep}}$ extracted spatial features that captured more exhaustive pixel relationships. The backbones of pre-trained models were utilized rather than training naïve deep learning models on the limited dataset of 70 participants. Utilizing backbones pre-trained on large datasets as a spatial feature extractor for the 70 arrays of chromophore maps reduced the risk of model overfitting. The pre-trained model layers were all frozen during the feature extraction. The models used were ResNet-50 (Koonce, 2021), Inception-V3 (Szegedy et al., 2016), and DenseNet-121 (Huang et al., 2017). They were pretrained on RadImageNet (Mei et al., 2022) – a dataset of 1.35 million medical images across three radiological modalities. The models learned to identify tissue texture features from RadImageNet. The learned knowledge was utilized to extract spatial features in the chromophore maps. The assumption was that there were similarities between the tissue texture information in grayscale radiological images and in chromophore maps (Oladokun et al., 2024). Each chromophore map was fed to the model backbones to generate β feature maps $\left[\boldsymbol{F_{m1,\varphi}^{i,deep}}, \boldsymbol{F_{m2,\varphi}^{i,deep}}, \dots, \boldsymbol{F_{m\beta,\varphi}^{i,deep}}\right]$, as shown in equation 5.8. The feature map channel count β was 2048, 2048, and 1024, for ResNet-50, Inception-V3, and DenseNet-121, respectively. Each feature map was considered as the output of virtual functions $\boldsymbol{M^{deep}} = [\boldsymbol{M_1^{deep}}, \boldsymbol{M_2^{deep}}, \dots, \boldsymbol{M_\beta^{deep}}]$.

$$\boldsymbol{D^{deep}} = \boldsymbol{F_{m,\varphi}^{i,deep}} = \left[\boldsymbol{M_1^{deep}}(\boldsymbol{I}), \boldsymbol{M_2^{deep}}(\boldsymbol{I}), \dots, \boldsymbol{M_\beta^{deep}}(\boldsymbol{I})\right] = \left[\boldsymbol{F_{m1,\varphi}^{i,deep}}, \boldsymbol{F_{m2,\varphi}^{i,deep}}, \dots, \boldsymbol{F_{m\beta,\varphi}^{i,deep}}\right] \quad 5.8$$

The (200×200) dimension for each chromophore map resulted in feature map arrays with dimensions $(7 \times 7 \times 2048)$, $(4 \times 4 \times 2048)$, and $(6 \times 6 \times 1024)$ for ResNet-50, Inception-V3, and DenseNet-121, respectively. This is as illustrated in Figure 5.7.

The choice of handcrafted radiomics feature extraction was motivated by their reproducibility and explainability. This is augmented by deep radiomics feature extraction which may identify texture features in the chromophore maps that were not captured by the static radiomics functions. Furthermore, performing feature extraction in the induration region of interest – generated by the PCA model in chapter 4 – enabled the localisation of induration characterisation.

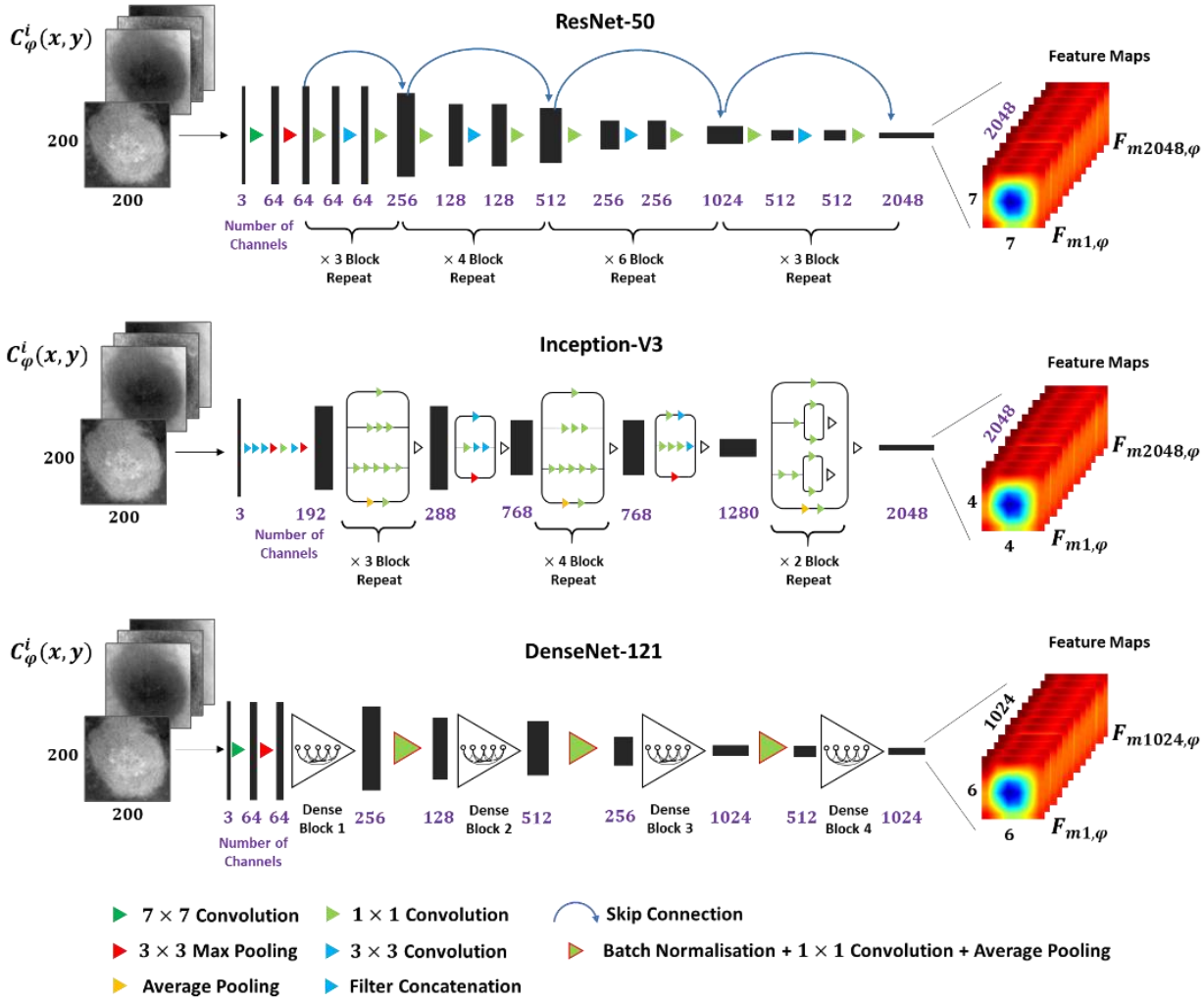


Figure 5.7: Illustration of the generation of deep radiomics features

5.2.3. Feature Ranking and Selection

The deterministic and deep radiomics features converged at $\boldsymbol{D^{rad}}$ which selects between the two sets of features. It allowed for feature selection to be done on either set. A 5-batch nested cross-validation was performed on the features for all participants. A random 80:20 training-test split was done in each batch. A SVM model was fitted to the training subset and evaluated on the test subset per batch to generate the score $\boldsymbol{S^{rad}_{b,\text{test}}}$. The labels for the classification were the TST-based LTBI diagnoses for all participants. The IGRA based diagnoses for a subset of 21 of the South African participants were acquired from a clinical laboratory in Cape Town, South Africa. Classification scores were also computed for the IGRA-based labels. A 5-fold cross-validation was performed on the training subset per batch to generate the training score $\boldsymbol{S^{rad}_{b,\text{valid}}}$. The metric of evaluation used was the balanced accuracy (Brodersen et al., 2010) for $\boldsymbol{S^{rad}_{b,\text{valid}}}$ and $\boldsymbol{S^{rad}_{b,\text{test}}}$. The mean scores $\boldsymbol{S^{rad}_{\text{mean,valid}}}$ and $\boldsymbol{S^{rad}_{\text{mean,test}}}$ across the batches were computed for each candidate feature $\boldsymbol{f_{\text{cand}}}$ in $\boldsymbol{D^{rad}}$. The features were ranked by the scores and the set of best-scoring single features $\boldsymbol{F^{best}_{solo}}$ were selected. A constrained

sequential feature selection was performed to repeat the selection process with every pair of features in $\boldsymbol{F_{solo}^{best}}$. The top-ranking feature and feature-pair were identified as potential biomarkers for LTBI. This is illustrated in equation 5.9. The schematic of the characterization framework for LTBI biomarker identification is illustrated in Figure 5.8.

$$\boldsymbol{F^{best} = (f_{cand}, S_b^{rad} | S_b^{rad} = CV(SVM(f_{cand}, y), k); S_b^{rad} > \tau, f_{cand} \epsilon F_{m,\varphi}^{i,rad})} \quad 5.9$$

5.2.4. Feature Visualisation

The best deterministic and deep radiomics features were considered as potential biomarkers for LTBI. The deterministic radiomics feature per chromophore map was a numerical value. To improve the interpretability of the features, a function $\boldsymbol{\mathcal{B}^{det}}$ was defined to generate a contribution map $\boldsymbol{\mathcal{B}^{det}(C_{\phi}^{i}, F_{m\alpha,\phi}^{i,det})}$ of a deterministic feature $\boldsymbol{F_{m\alpha,\phi}^{i,det}}$ extracted from chromophore map $\boldsymbol{C_{\phi}^{i}}$. The contribution map estimates the contribution of each pixel in the map $\boldsymbol{C_{\phi}^{i}}$ to the generation of the feature $\boldsymbol{F_{m\alpha,\phi}^{i,det}}$. This provided a visualization of regions in $\boldsymbol{C_{\phi}^{i}}$ that significantly influenced $\boldsymbol{F_{m\alpha,\phi}^{i,det}}$. This presented a qualitative interpretation of deterministic radiomics features. The contribution map for a first order feature like $total\ energy$ was defined in this study as illustrated in Figure 5.9. The contribution maps for the other deterministic features which, unlike the first order features, involve the generation of a feature matrix, are defined here and illustrated in Figure 5.10. The figure illustrates how the contribution map of the GLRLM $gray\ level\ nonuniformity$ feature was generated, as an exemplar of other features. The feature involved the estimation of the feature matrix $\mathbf{P}^{\boldsymbol{rlm}}(\boldsymbol{j}, \boldsymbol{k})$ which quantifies the length of consecutive horizontal pixels with the same binned intensity.

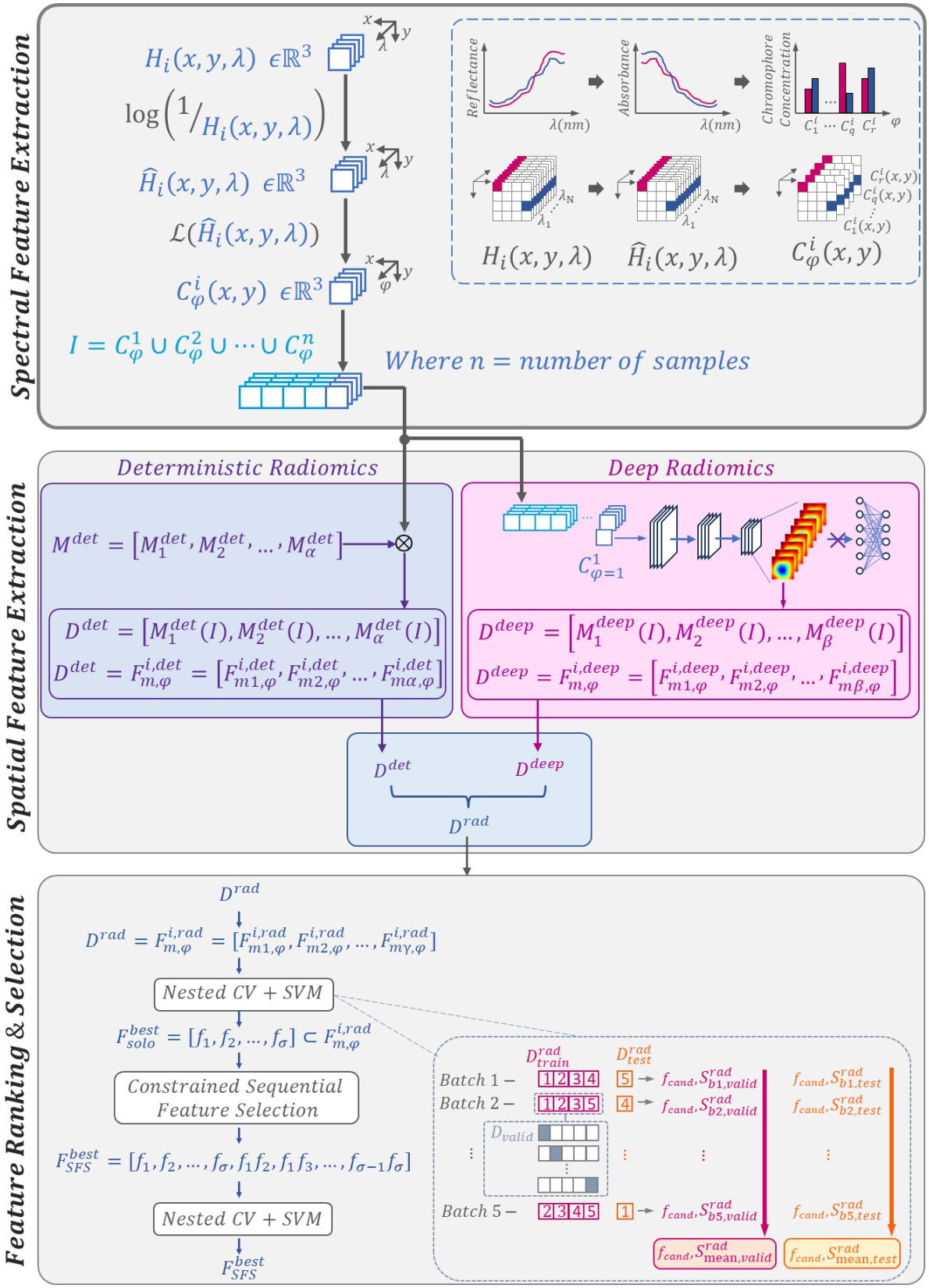


Figure 5.8: Schematic of the characterization framework for LTBI biomarker identification

$$F_{m\alpha,\phi}^{i,det} = total\ energy = pixel\ area \sum_{j=1}^{N_p} (C_\phi(j))^2 \quad \text{Where, } N_p \text{ = number of pixels in } C_\phi$$

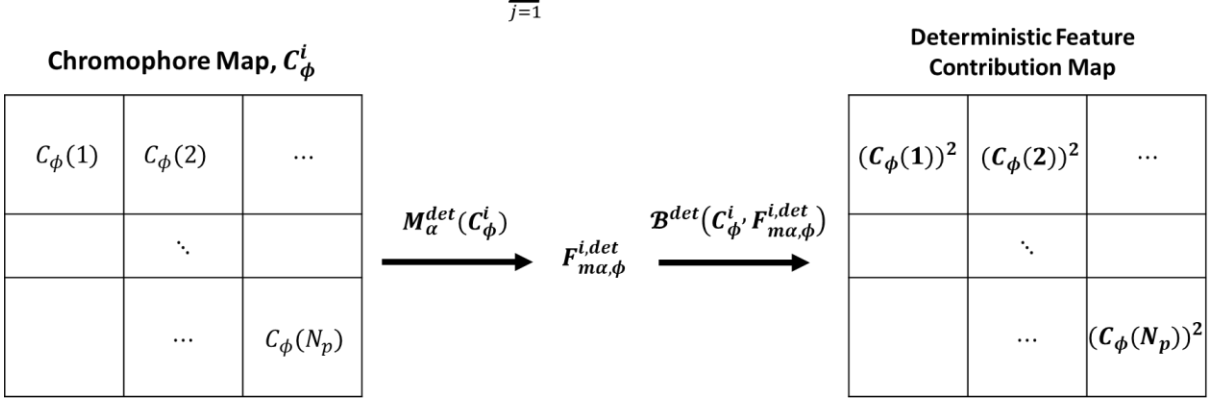
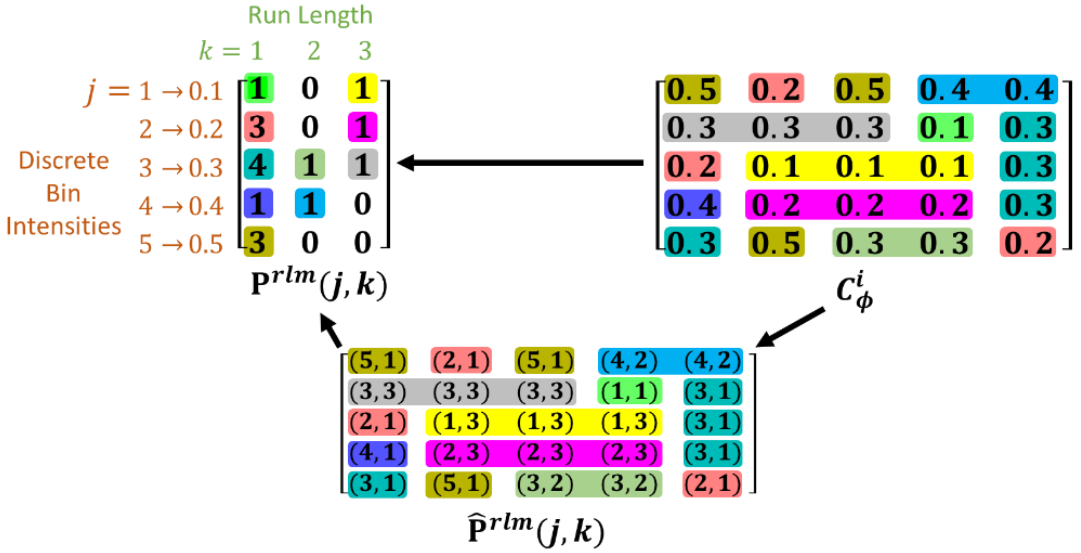


Figure 5.9: Definition of feature contribution map for $total\ energy$

$$F_{m\alpha,\phi}^{i,det} = gray\ level\ nonuniformity = \frac{\sum_{j=1}^{N_p}\left(\sum_{k=1}^{N_r} \mathbf{P}^{rlm}(j,k)\right)^2}{N_r}$$

Where, N_g = Number of discrete bin intensities ; N_r = Number of discrete run lengths

N_p = number of pixels in C_ϕ



$$\mathbf{P}_g^{rlm}(j) = \frac{\left(\sum_{k=1}^{N_r} \mathbf{P}^{rlm}(j,k)\right)^2}{N_r}$$

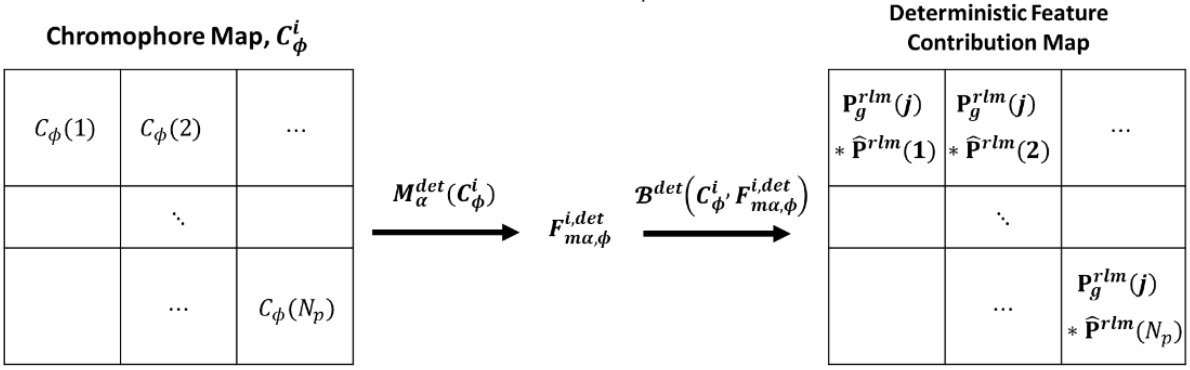


Figure 5.10: Definition of feature contribution map for GLRLM $gray\ level\ nonuniformity$

The deep radiomics feature per chromophore map was a $\upsilon \times \upsilon$ feature map, where υ is 7, 4, and 6 for ResNet-50, Inception-V3, and DenseNet-121, respectively. The feature maps, generated by the last layer of the model backbones, captured high-level features from the input chromophore maps. Thus, each feature map was treated as an attention map. Overlaying each feature map as a heatmap over its input chromophore map provided a visualisation that highlighted regions in the chromophore maps that significantly contribute to the feature. This is as illustrated in Figure 5.11.

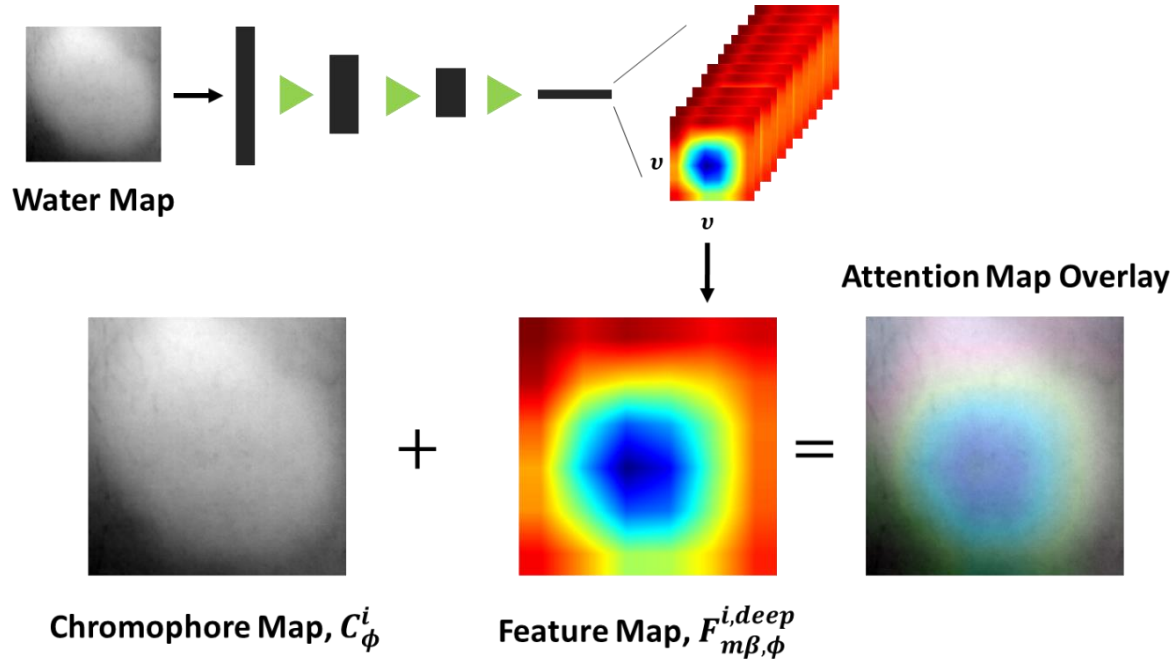


Figure 5.11: Attention map overlay for visualisation of deep radiomics features

5.3. Results

The median squared-error between the absorbance model $\widehat{\boldsymbol{H}}_{\boldsymbol{i}}^{\boldsymbol{pred}}$ and the spectral signatures sampled from the induration region of all participants is as shown in Figure 5.12 (a). The figure shows the error after the iterative inclusion of each chromophore from the library of chromophores into the regression model. The changes in the error following the inclusion of each chromophore are as shown in Figure 5.12(b). The changes were measured after the inclusion of OX and DX as they were considered as baseline chromophore verified by the presence of erythema. The chromophores that resulted in significant change in regression error were EM, PM, WT, and FR. Thus, the candidate chromophores were OX, DX, EM, PM, WT, FR, and residual determinants (RES).

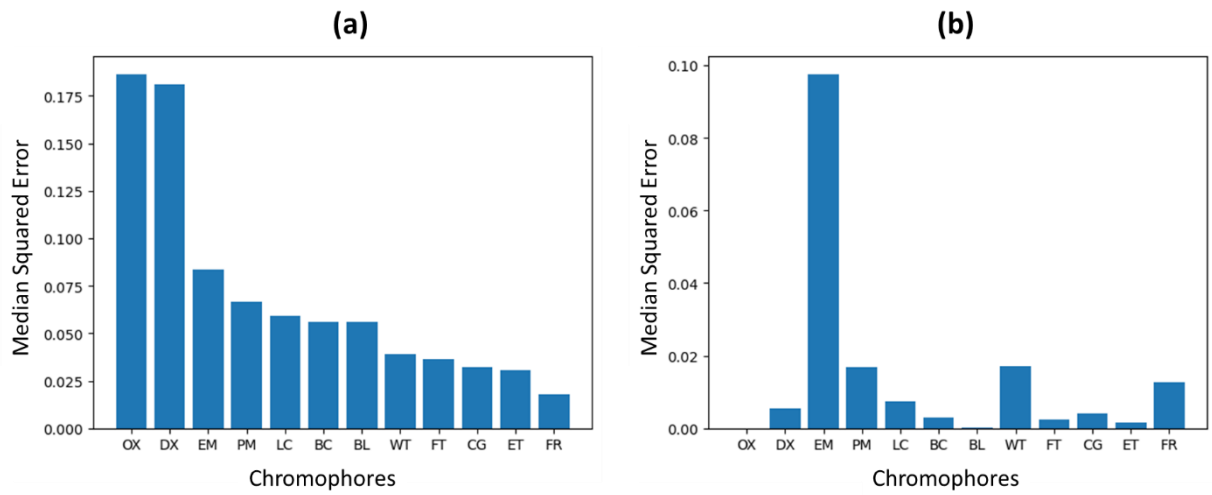


Figure 5.12: The output of candidate chromophore selection process showing: (a) median squared error and (b) change in error, following the inclusion of each chromophore from the library of chromophores

The change in the absorbance model with the inclusion of the candidate chromophores for a single sampled spectral signature is as shown in Figure 5.13. The fit of the model to the spectral signature significantly improved with each chromophore inclusion.

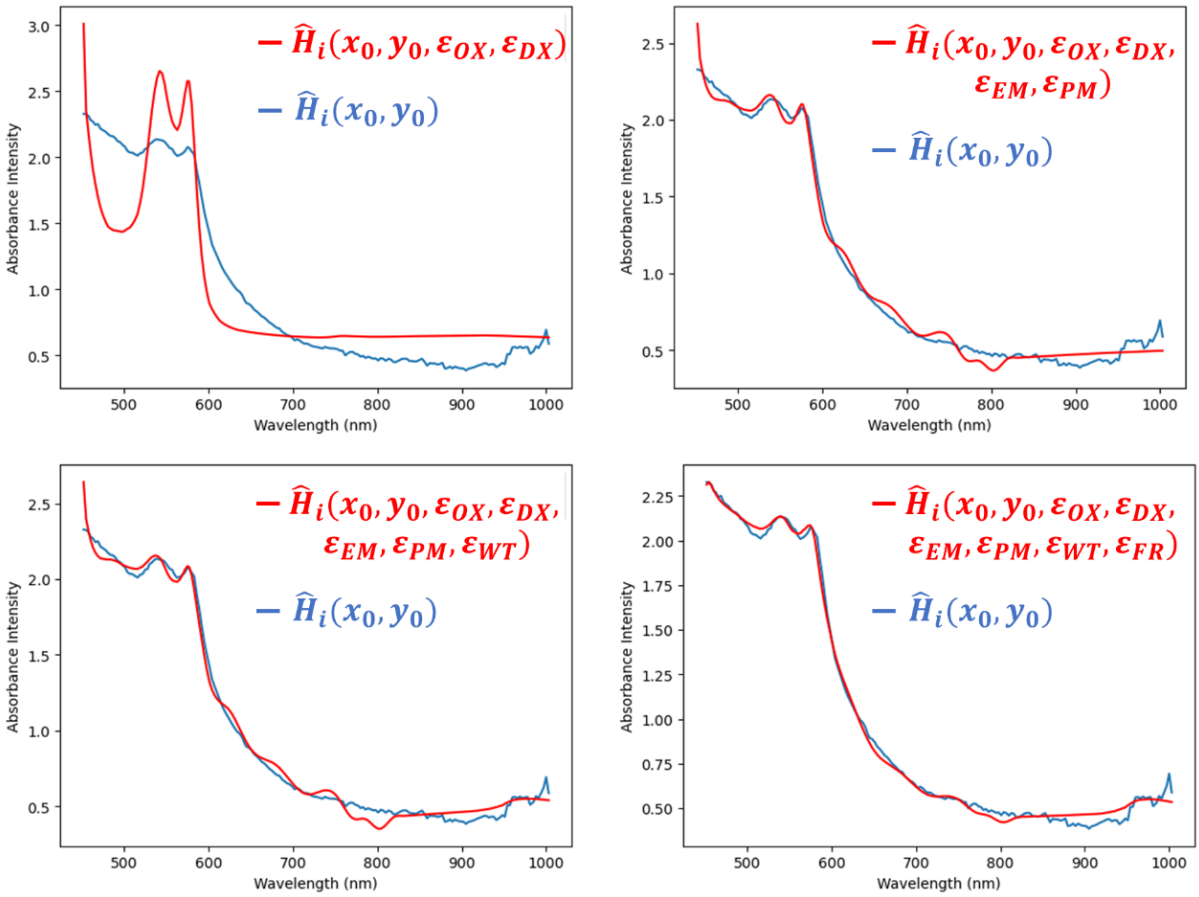


Figure 5.13: Improvement in the fit of absorbance model to a single sampled spectral signature with successive inclusion of candidate chromophores

Applying spectral unmixing to the hypercube of all participants using only the candidate chromophores resulted in the grayscale images in Figure 5.14. The grayscale maps were normalised per participant to compare the chromophore distribution between participants rather than absolute concentrations. The first and second rows of images in the RGB estimates are of five representative South African and Vietnamese participants, respectively. The corresponding chromophore maps for the representative participants are also shown in Figure 5.14. The chromophore maps of all participants are presented in Appendix D.

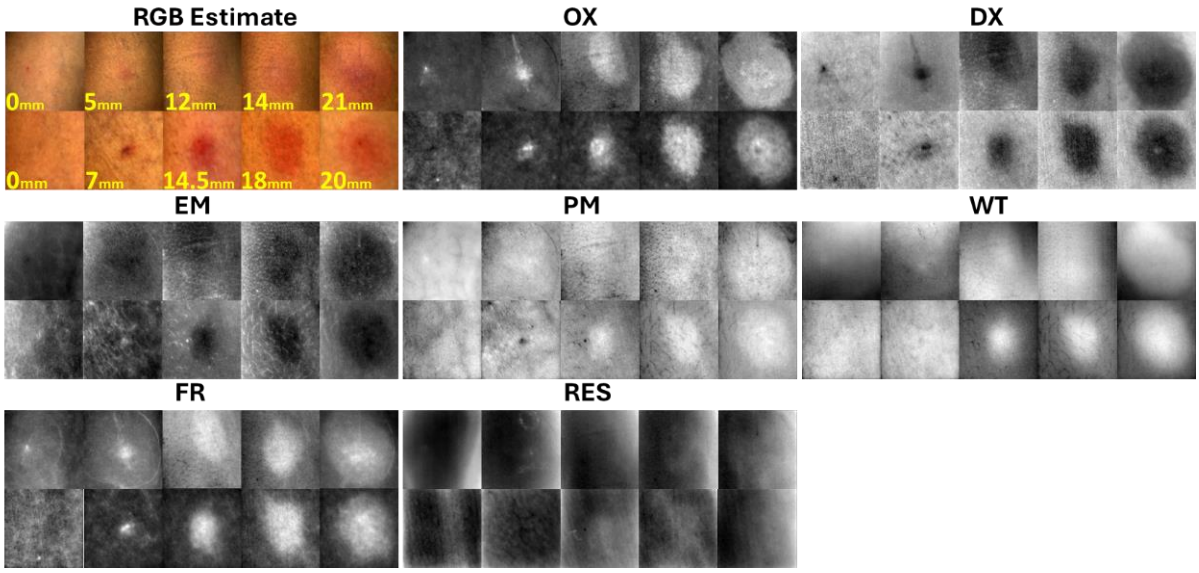


Figure 5.14: RGB estimates, and grayscale chromophore maps for a subset of all participants

The colour map visualisation of the chromophores is as shown in Figure 5.15. The figure enables easier visualisation of regions of high chromophore concentrations relative to baseline normal skin per participant. The OX maps for the participants possess high concentration in regions that correspond to regions of skin redness (erythema) in the RGB images. This suggests that OX, characterised by the bright red colour of arterial blood, has higher levels at the reaction site compared to the surrounding baseline skin tissue per participant. Conversely, the DX maps possess darker pixels in regions that correspond to erythema in the RGB images. This suggests that DX, which is characterised by the dark reddish-purple colour of venous blood, has relatively lower concentrations at the reaction site compared to baseline skin per participant. In the RGB image of the participant exhibiting an 18 mm induration, hair strands are visible. However, these strands are not appreciably discernible in the OX or DX maps, indicating that they do not meaningfully influence the observed haemoglobin signal. This finding is consistent with the well-established fact that hair does not contain significant haemoglobin. The size of the high OX cluster and low DX cluster per participant appear to correlate with the size of the Mantoux reading. This suggests that haemoglobin concentration is related to the pathophysiology of TST indurations.

Similarly, the relatively low concentration regions in the EM maps suggests that the reaction sites have lower EM levels than baseline skin per participant. The EM chromophore is a dark brown or black pigment that is found in black coloured hair. It is also responsible for the brown tone of people of African ancestry. The Black hair strands at the test site for the participant with 18mm induration correspond to bright (red) pixels of high concentration in the 18mm EM map, as expected. The reverse is the case for the PM maps. The hair strands appear as low (blue) concentration regions in the 18mm PM map. The PM chromophore is a yellowish-to-reddish pigment. Thus, as expected, regions in the RGB images with reddish colouration correspond to bright pixels in the PM maps. The contrast between these bright pixels and the baseline surrounding skin in the PM maps is not as pronounced as it was in the OX maps. This suggests that there is not a large variation in the PM levels across the imaged region per participant. It also indicates that OX contributes more to the bright red regions in the RGB images compared to PM and that the redness is more due to subdermal bleeding than localised increase in PM levels.

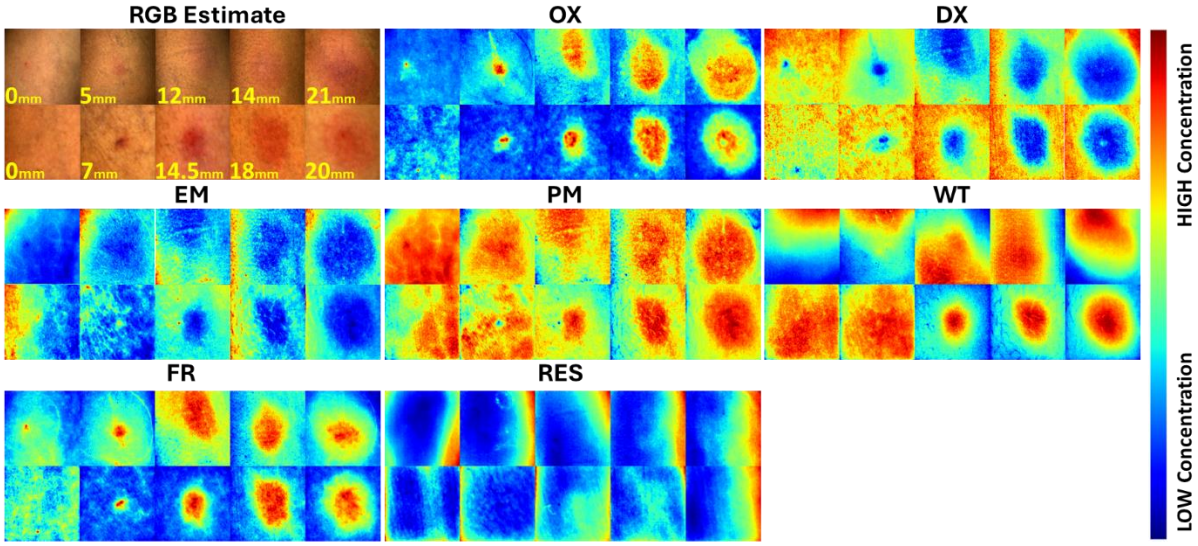


Figure 5.15: RGB estimates and chromophore colourmaps for a subset of all participants

The regions in the RGB images that correspond to erythema also produce bright pixels in the WT maps. However, unlike in the OX maps, the boundaries of the bright regions in WT have lower contrast against surrounding baseline skin. This implies that WT is a more subtle chromophore at the reaction site than OX. Hair strands appear as dark pixels in the 18mm induration WT map which suggests that they have significantly lower WT concentration compared to the reaction site and baseline skin, as expected. Vietnamese participants with <10mm induration appear to have a monotonous distribution of WT. Conversely, the Vietnamese participants with ≥10mm induration appear to have a significant contrast in WT concentration at the reaction site compared to baseline skin. The relationship between WT concentration clusters and Mantoux readings for South African participants does not have a similarly consistent pattern.

Similar to the OX maps, the FR maps have bright pixels in regions that correspond to erythema in the RGB images. This indicates that significant FR levels are present the TST reaction site. There also appears to be a sharp contrast between the bright regions in FR and baseline surrounding skin. This indicates a localised pool of FR that is specific to the TST. The bright regions in FR appear to have a different texture compared to the bright regions in OX. This implies that there are significant differences in the distribution of these chromophores at the reaction site despite their similarities.

The RES maps are characterised by predominantly homogenous clusters of dark pixels, suggesting that they capture a characteristic that is common to the reaction site and baseline skin. This also implies that most of the major chromophore characteristics across the imaged region have been captured by the candidate chromophores. The outline of some large blood vessels (probably veins) is shown as bright pixels in the RES maps. This implies that there is possibly an unmodelled determinant related to these vessels. There is no visually apparent relationship between the distribution of concentrations in RES and the TST readings.

Results of the application of deterministic radiomics feature extraction to the chromophore maps of all participants and ranking the features are illustrated in Table 5.2. The table presents the performance of the best deterministic radiomics features in $\boldsymbol{D^{det}}$ for the unmasked chromophore map of all participants. A FR-based feature produced the best balanced-accuracy scores across all participants. The combination of the FR-based and a WT-based feature produced the overall best mean cross-validation accuracy of 88.4% and mean test accuracy of 85.0% across all participants. This suggests that FR and WT are potential biomarkers for the TST-based LTBI diagnosis across South African and Vietnamese participants.

However, repeating this chapter's methodology separately on the South African and Vietnamese participants resulted in differences in the best deterministic features. As indicated in the table, the best performing feature for the South African participants is an OX-based feature. The best pair of features for the cohort was an OX-based and a RES-based feature. This suggests that the distribution of OX at a TST site, in conjunction with a baseline residual characteristic, is a good predictor for TST-based LTBI diagnosis. Conversely, WT-based features were the best predictor for the TST diagnosis in the Vietnamese participants. The mean cross validation accuracy obtained for the Vietnamese participants is higher (97.3% versus 91.2%) than the accuracy for the South African participants. The differences in the best chromophores between the two cohorts suggest that there are unique characteristics of each cohort that influence their TST reactions. Further investigation, using larger sample sizes of each cohort, may be necessary in future research to ascertain these characteristics.

Table 5.2: Accuracy Scores of the Best Deterministic Features for Chromophore Maps
(Where BSF = best single feature, BFP = best feature pair, BMCA = balanced mean cross-validation accuracy, MTA = mean test accuracy, GLNU = *gray level nonuniformity*, GLV = *gray level variance*, MAD = *mean absolute deviation*, HGLRE = *high gray level run emphasis*)

	South African and Vietnamese Participants			**South African Participants Only**			**Vietnamese Participants Only**		
	Feature	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**
BSF	FR (GLRLM GLNU)	78.2%	84.0%	OX (GLSZM GLV)	86.7%	93.8%	WT (GLRLM HGLRE)	94.0%	93.3%
BFP	FR (GLRLM GLNU) + WT (total energy)	88.4%	85.0%	OX (MAD) + RES (Variance)	91.2%	96.0%	WT (GLRLM HGLRE) + WT (Mean)	97.3%	96.7%

A visual interpretation of the best combination of deterministic features for TST prediction across all participants is presented by the feature contribution maps in Figure 5.16. As seen in Figure 5.16 (a), the contribution maps for GLRLM *gray level nonuniformity* of FR show that texture difference between TST reaction site of FR maps and normal skin regions is a good predictor for TST-based diagnosis across cohorts. As shown in Figure 5.16 (b), the contribution maps for *total energy* of WT indicate that the contrast between regions of high and low intensity in WT are also good predictors for the TST across cohorts.

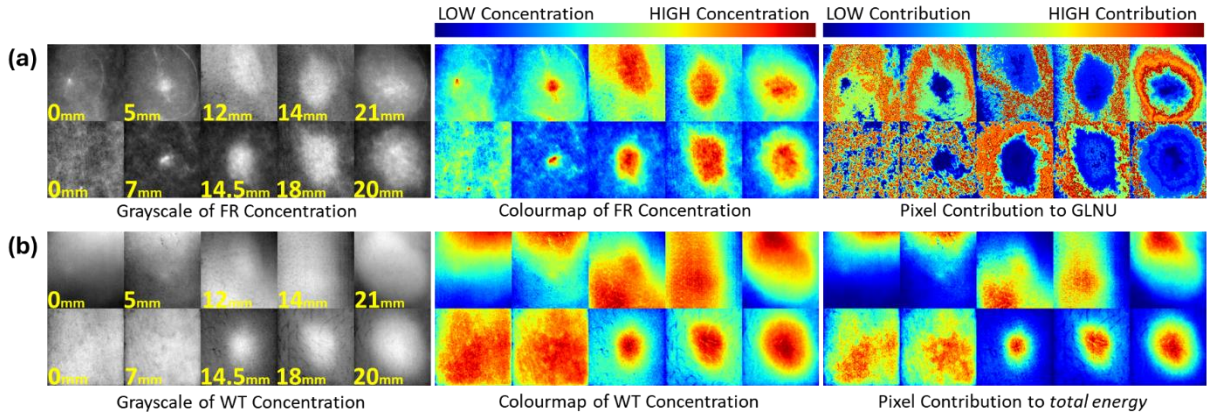


Figure 5.16: Visualisation of best pair of deterministic features for five representative South African and Vietnamese participants showing: (a) the feature contribution map for the GLRLM *gray level nonuniformity* of FR, and (b) the feature contribution map for the *total energy* of WT

Normal skin pixels are shown to have a high contribution to the GLNU feature while pixels at the reaction site have a low contribution. The total energy feature is the square of each pixel intensity in WT which increases the contrast between high and low intensities. Increasing this contrast in WT

increases its predictive ability for TST. The combination of the GLNU feature in FR and the total energy of WT produce the most generalizable characterization across cohorts using deterministic radiomics. This can be described as the combination of texture contrast in FR and intensity contrast in WT.

The contribution maps for the best deterministic feature combination for only South African participants are shown in Figure 17. The combination of the MAD feature of OX and variance of RES is a good predictor for the TST in South African participants. The MAD feature quantified how distinct each pixel in OX is compared to the average intensity per participant. As seen in Figure 17(a), participants with induration ≥10mm had more pixels with high contribution to MAD than participants with <10mm induration. This suggests that higher levels of OX at the TST site compared to the average level is a good predictor for TST. The contribution map for the variance of RES, as shown in Figure 17(b), emphasizes the homogeneity of the pixel contribution. There is no visible distinction between the variance contribution maps between the participants. This suggests that by itself, the RES-based feature may not be a good predictor of the TST, but it augments the predictive ability of the MAD of OX. Thus, the combination of the mean deviation of OX intensities and the homogeneity of RES is a good predictor for the TST in South African participants. The visualisation of these contribution maps for all South African participants is presented in Appendix E.

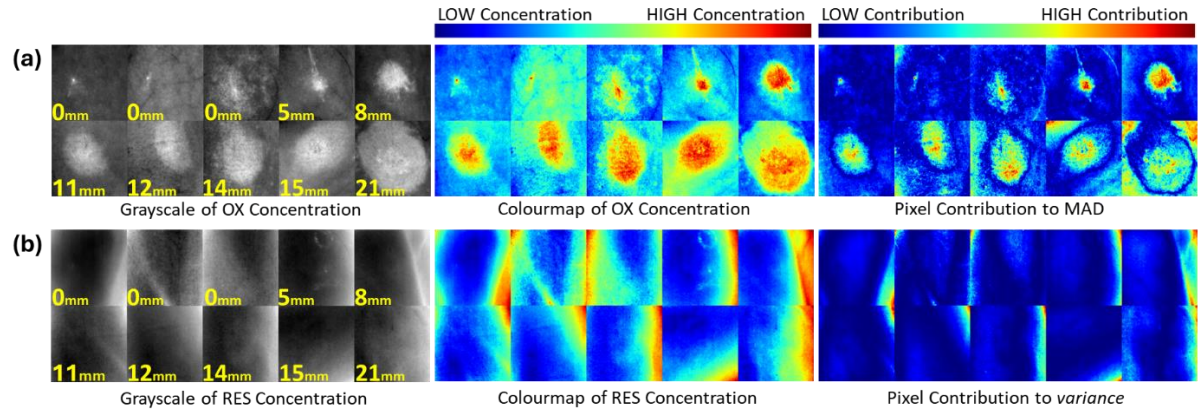


Figure 5.17: Visualisation of best pair of deterministic features for ten representative South African participants showing: (a) the feature contribution map for the $mean\ absolute\ deviation$ (MAD) of OX, and (b) the feature contribution map for the $variance$ of RES.

The contribution maps for the best deterministic feature combination for only Vietnamese participants are shown in Figure 18. The combination of the HGLRE feature and mean of WT was the best predictor of the TST for this cohort. The HGLRE feature quantifies the proliferation of pixel clusters with similar intensity in WT. Most of the pixels in participants with induration <10mm had high contribution to HGLRE. Conversely, pixels of the reaction site in participants ≥10mm had low contributions. This suggests that the induration region in WT is more homogenous than normal skin region. This texture contrast in WT is presented as the best TST predictor for the Vietnamese

participants. The mean of WT is the second feature that augments the predictive accuracy of the HGLRE of WT. The contribution map for the mean of WT is identical to the colourmap of WT in Figure 5.18 since each pixel's contribution to the mean is directly proportional to pixel's intensity. The contribution maps for all Vietnamese participants are presented in Appendix E.

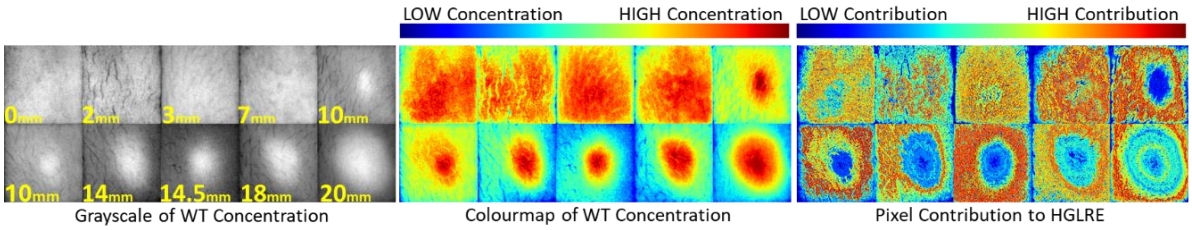


Figure 5.18: Visualisation of the best deterministic feature for ten representative Vietnamese participants showing the feature contribution map for the $high\ gray\ level\ run\ emphasis$ (HGLRE) of WT

Feature plots to visualize the values for the best feature pair for all, South African, and Vietnamese participants are shown in Figure 5.19. Positive participants tended to have low values for the FR and WT-based features and high values for the OX-based feature.

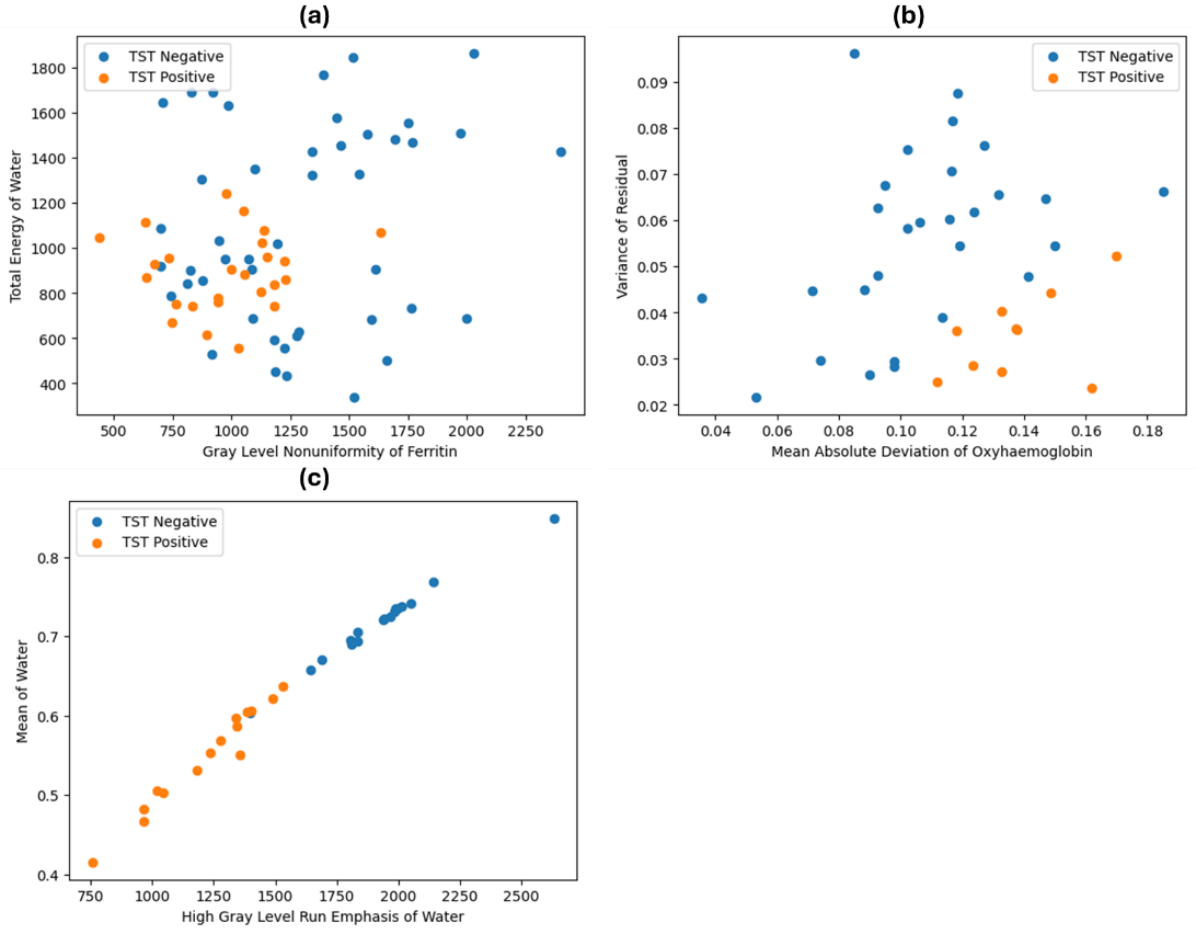


Figure 5.19: The scatterplots of the best pair of deterministic radiomics features for (a) all participants, (b) only South African participants, and (c) only Vietnamese participants

The results of the application of the PCA-enabled induration masks from chapter 4 to each chromophore map before performing deterministic feature extraction are given, in Table 5.3. The accuracy of the best features for all participants, and for each of the cohorts, are significantly lower than the corresponding accuracies for the unmasked chromophore maps. This suggests that the spatial relationship between the normal skin region and reaction site in the chromophore maps significantly contributes to the prediction of LTBI diagnosis. However, despite the lower accuracies compared to Table 5.2, FR and WT consistently appear to produce the best features across both cohorts. This insinuates that applying an induration mask significantly minimizes the impact of unique cohort-specific characteristics. Generally, the performance of the deterministic features suggests that the chromophore biomarkers for TST-based LTBI diagnosis are FR, WT, and OX.

Table 5.3: Accuracy Scores of the Best Deterministic Features for Induration-Masked Chromophore Maps
(Where RLNUN = run length nonuniformity normalized,
LAHGLE = $large\ area\ high\ gray\ level\ emphasis$, SAE = $small\ area\ emphasis$)

	South African and Vietnamese Participants			**South African Participants Only**			**Vietnamese Participants Only**		
	Feature	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**
BSF	FR (GLRLM RLNUN)	84.8%	83.3%	FR (GLSZM LAHGLE)	79.2%	87.5%	FR (GLRLM RLNUN)	92.0%	90.0%
BFP	FR (GLRLM RLNUN) + WT (NGTDM coarseness)	85.2%	83.3%	FR (GLSZM LAHGLE) + WT (GLSZM SAE)	86.3%	81.3%	FR (GLRLM RLNUN) + FR (GLRLM run percentage)	92.0%	90.0%

The results of the application of deep radiomics feature extraction on the chromophores of all participants using the pretrained ResNet-50, Inception-V3, and DenseNet-121, are given in Table 5.4. The table shows that DenseNet-121 had the best overall performance as a feature extractor. The features extracted by DenseNet-121 generated a mean cross validation accuracy of 86.5% and a mean test accuracy of 90% across all participants. This test accuracy is significantly greater than the test accuracy of the best pair of features identified by ResNet-50 and Inception-V3.

Table 5.4: Performance of Pre-trained Model Backbones for Feature Extraction on Chromophore Maps of all Participants (Where FMC-β = feature map channel – index)

	ResNet-50			Inception-V3			DenseNet-121		
	Feature	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**
BSF	DX (FMC-41)	86.5%	70.0%	PM (FMC-1297)	84.0%	80.0%	FR (FMC-534)	**82.0%**	**90.0%**
BFP	DX (FMC-41) + FR (FMC-524)	86.5%	70.0%	PM (FMC-1297) + WT (FMC-169)	84.0%	80.0%	FR (FMC-534) + WT (FMC-566)	**86.5%**	**90.0%**

The accuracy scores of the best deep radiomics features from DenseNet-121 for each cohort is shown in Table 5.5. Like the deterministic radiomics approach, DenseNet identified FR and WT as the best predictive chromophores for all participants. The test accuracy achieved by DenseNet-121 across all participants is higher (90% vs 85%) than the accuracy of the deterministic model. DenseNet identified DX as a good predictor at cohort-level but identified FR and WT as the best across both cohorts.

Table 5.5: Accuracy Scores of the Best DenseNet-121 Deep Radiomics Features for Chromophore Maps

	South African and Vietnamese Participants			South African Participants Only			Vietnamese Participants Only		
	Feature	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**
BSF	FR (FMC-523)	82.0%	90.0%	DX (FMC-692)	90.0%	75.0%	DX (FMC-28)	91.7%	96.0%
BFP	FR (FMC-523) + WT (FMC-566)	86.5%	90.0%	DX (FMC-692) + OX (FMC-188)	90.0%	95.0%	DX (FMC-28) + WT (FMC-698)	93.3%	96.7%

The difference between the most predictive chromophores at cohort level and across cohort highlights the cohort-specific characteristics also observed in Table 5.2. Generally, the performance of this

backbone indicates that a deep learning model pretrained on a large dataset of radiological images can be a good feature extractor for chromophore maps of the TST without finetuning.

A visual interpretation of the best combination of DenseNet-121 features for TST prediction across all participants is presented by the heatmaps in Figure 5.20. The heatmaps of feature channel-523 for FR suggest that pixel regions where the model paid low attention correspond to induration region. The differences in the polarity of the attention paid by the model to the normal skin region and to the reaction site in FR is a good predictor for the TST across cohorts. The feature channel-566 for WT improves the accuracy of the FR feature. Its heatmap suggests that differences in the polarity of the attention paid by the model to the top and bottom half of WT, across cohorts, is a good contributor to the accuracy of the FR feature

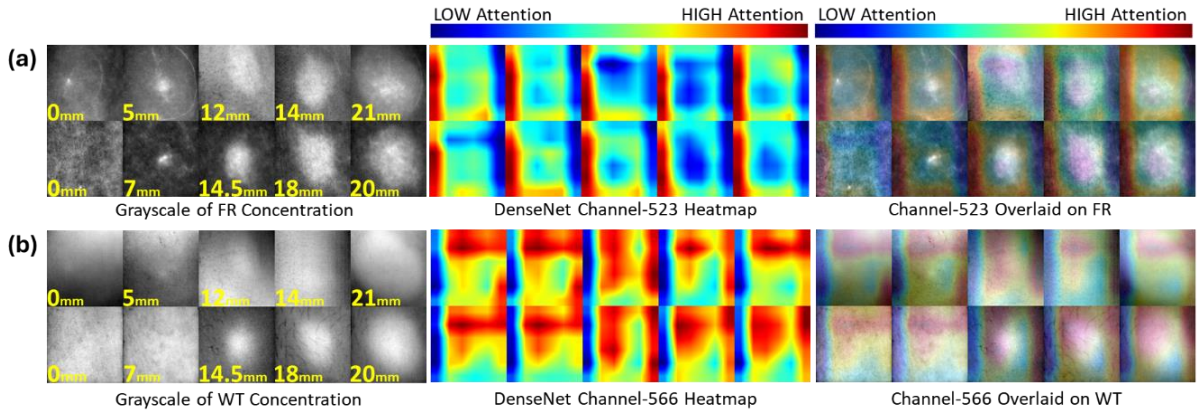


Figure 5.20: Visualisation of best DenseNet-121 deep feature pair for five representative South African and five Vietnamese participants showing: (a) heatmap of feature channel-523 for FR, and (b) heatmap of feature channel-566 for WT

The best pair of DenseNet channels for only the South African participants are visualised by the heatmaps in Figure 5.20. The channel-692 for DX and channel-188 for OX were the best feature pair for the cohort. The heatmap of channel-692 shows that the model tended to pay positive attention to regions in DX that have lower concentration of DX than baseline normal skin regions. Conversely, the heatmap of channel-188 shows that the model tended to pay low attention to regions of high concentration of OX. The best pair of channels for only the Vietnamese participants are visualised by the heatmaps in Figure 5.21. The channel-28 for DX and channel-698 for WT were the best feature pair for the Vietnamese cohort. The heatmap of channel-28 shows that the model tended to pay positive attention to regions of relatively low DX concentration. The heatmap of channel-698 does not show any consistent relationship between the attention regions and the intensity clusters in WT.

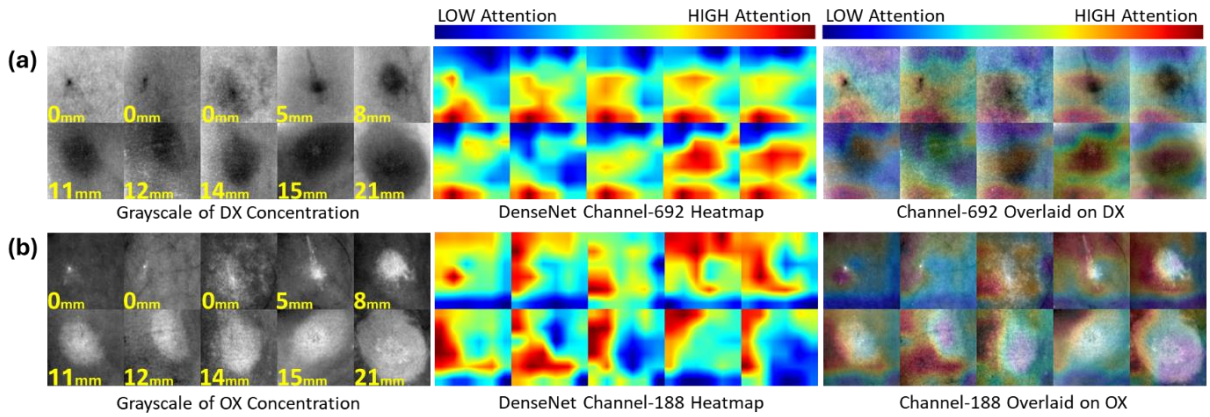


Figure 5.21: Visualisation of best DenseNet-121 deep feature pair for ten representative South African participants showing: (a) heatmap of feature channel-692 for DX, and (b) heatmap of feature channel-188 for OX

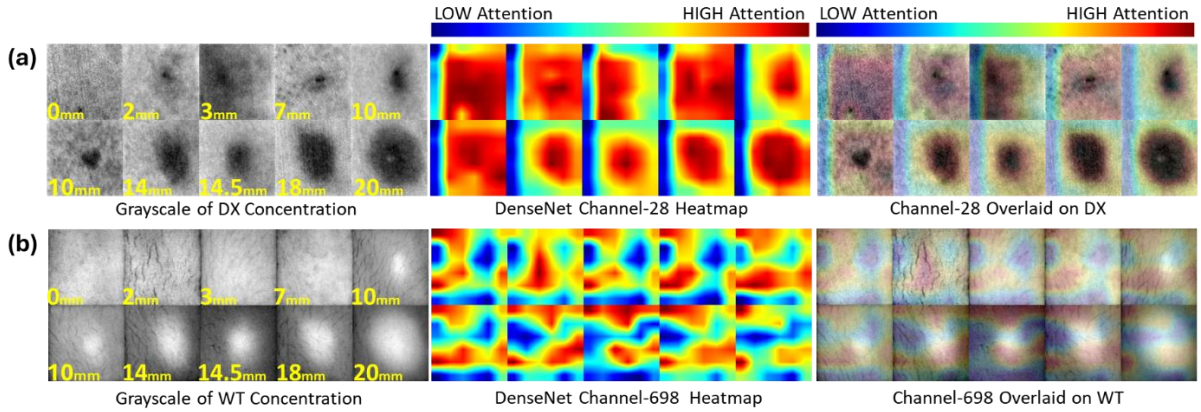


Figure 5.22: Visualisation of best DenseNet-121 deep feature pair for ten representative Vietnamese participants showing: (a) heatmap of feature channel-28 for DX, and (b) heatmap of feature channel-698 for WT

The results of the application of the PCA-enabled induration masks to each chromophore map before feeding them to the pretrained DenseNet-121, are given in Table 5.6. The application of the masks resulted in FR as the best single chromophore across all participants and for each cohort. This is identical to the identification of FR for each cohort and across cohorts by the deterministic model after the application of the masks. This validates the predictive accuracy of FR for TST. It also highlights the presence of cohort-specific characteristics in the normal skin regions which are excluded by the induration mask. As seen in the table, DenseNet features from induration-masked WT complement the FR features to produce the best accuracies for all participants and for the Vietnamese participants. A feature channel from induration-masked OX complemented an FR feature to produce the best accuracy for the South African participants. This validates the proposition that FR, WT, and OX are the most predictive chromophores for TST-based LTBI diagnosis.

Table 5.6: Accuracy Scores of the Best DenseNet-121 Deep Radiomics Features for Induration-masked Chromophore Maps

	South African and Vietnamese Participants			South African Participants Only			Vietnamese Participants Only		
	Feature	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**
BSF	FR (FMC-966)	85.7%	90.0%	FR (FMC-966)	90.0%	90.0%	FR (FMC-1020)	86.7%	93.3%
BFP	FR (FMC-966) + WT (FMC-981)	89.0%	90.0%	FR (FMC-966) + OX (FMC-862)	95.0%	95.0%	FR (FMC-1020) + WT (FMC-249)	95.0%	95.0%

Masking the chromophore maps focuses the attention of DenseNet-121 to the TST reaction site of all participants, as shown in Figure 5.23. For FR, the model tended to pay low attention to high concentrations of FR. Conversely, for WT, the model tended to pay positive attention to clusters of high WT concentration.

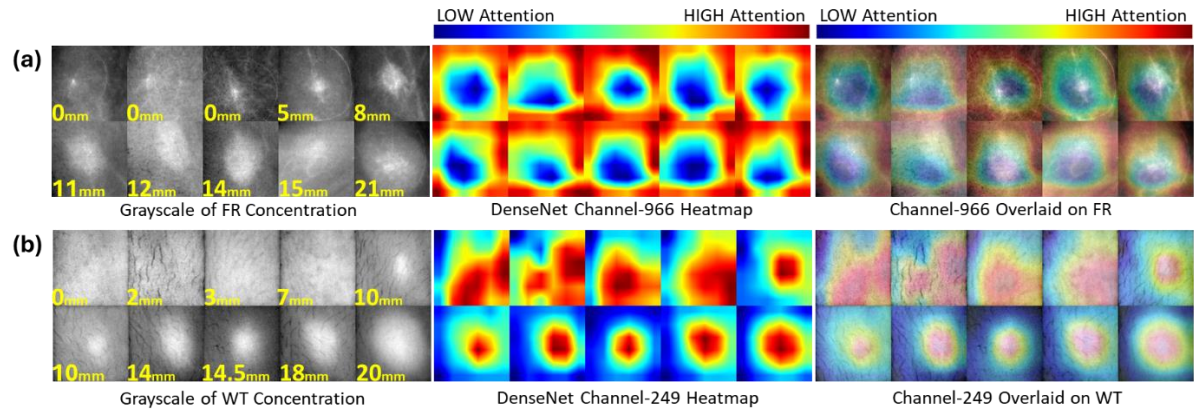


Figure 5.23: Visualisation of the heatmap of feature channels resulting from induration-masked chromophore inputs for (a) channel-966 of FR, and (b) channel-249 of WT

Applying deterministic radiomics to only the subset of 21 South African participants who have IGRA-based LTBI diagnosis results in Table 5.7. The utilization of IGRA labels instead of TST labels enables the investigation of the ability of the TST to predict IGRA via chromophore maps. As seen in the table, the GLDM dependence entropy feature of OX predicts the IGRA with 100% mean cross-validation and test accuracies. Though the predictive accuracy decreased to 90%, the position of this feature as the most predictive for IGRA did not change after applying induration masks to the chromophore maps.

Table 5.7: Best Deterministic Radiomics Feature for IGRA-based Classification

	Unmasked Chromophore Maps			Induration-Masked Chromophore Maps		
	Feature	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**
BSF	OX (GLDM dependence entropy)	100%	100%	OX (GLDM dependence entropy)	90.0%	90.0%

Applying deep radiomics, via DenseNet-121, to only the subset of participants with IGRA results in Table 5.8. Like the deterministic approach, the most predictive chromophore for IGRA-based LTBI diagnosis was OX, with cross-validation and test accuracy of 90%. The predictive performance of OX persists after the application of induration masks. The identification of OX as a most predictive chromophore for IGRA from a TST injection site by both deterministic and deep radiomics is a useful insight into the relationship between the TST and IGRA. However, there is a need to repeat the analysis on a larger diverse dataset of participants with IGRA readings to validate the predictive ability of OX.

Table 5.8: Best DenseNet-121 Deep Radiomics Feature for IGRA-based Classification

	Unmasked Chromophore Maps			Induration-Masked Chromophore Maps		
	Feature	**BMCA**	**MTA**	**Feature**	**BMCA**	**MTA**
BSF	OX (FMC- 272)	90.0%	90.0%	OX (FMC- 598)	90.0%	90.0%

The values per participant for the GLDM dependence entropy of OX, which was the most predictive deterministic feature for IGRA, is as shown by the number line in Figure 5.24. Participants with positive IGRA tended to have lower dependence entropy of OX than negative participants. This is also validated by the visualisation of the contribution maps in Figure 5.25 (a). The figure shows that the normal skin regions have higher entropy of OX compared to the TST reaction site. Furthermore, it shows that participants with positive IGRA diagnosis tended to have fewer pixels which contribute to entropy than participants with negative diagnosis. The heatmap of the most predictive DenseNet channel for IGRA is shown in Figure 5.25(b). The model tended to high attention to the normal skin region in contrast to the reaction site. This corroborates the high dependency entropy contribution of normal skin tissue identified using deterministic radiomics.

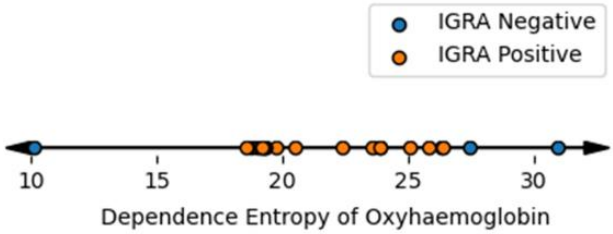


Figure 5.24: A number line for the best deterministic radiomics feature for IGRA prediction

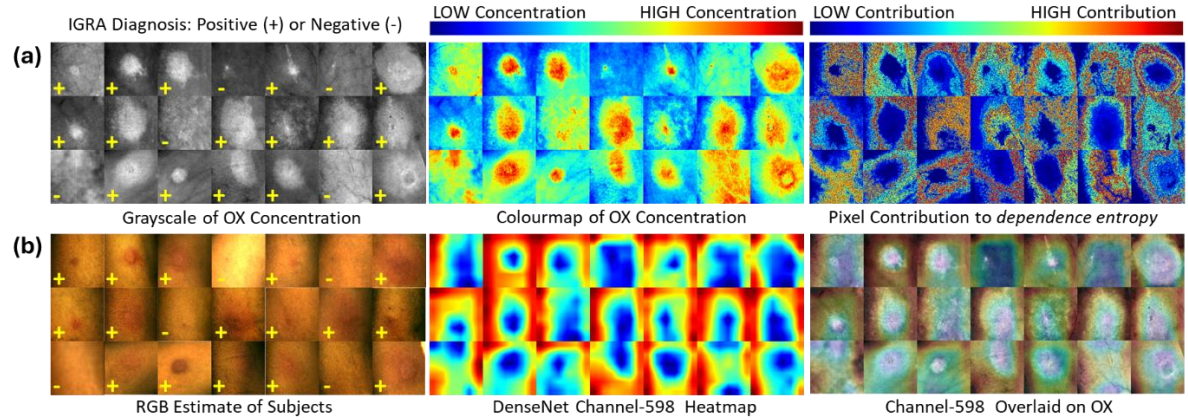


Figure 5.25: Visualisation of the best feature for IGRA prediction showing (a) the contribution maps for GLDM $dependency\ entropy$ of OX, (b) heatmap of feature channel-598 for OX

5.4. Discussion

This chapter set out to address the question on whether HSI can non-invasively capture features that are good representations of the underlying physiological responses to the tuberculin antigen on a TST injection site. Thus, an HSI framework was developed to identify biomarkers that accurately and consistently characterize TST indurations. The framework was validated using 70 participants (38 South African and 32 Vietnamese) to provide evidence of the accuracy and consistency of the identified biomarkers. The biomarkers were presented as digital representations of the pathophysiology of TST indurations with respect to normal skin physiology.

The framework was based on the assumption that each spectral signature from the participant hypercubes was composed of a linear combination of chromophores. It was also hypothesized that some of the constituent chromophores will be good representations of the pathophysiology of indurations. Based on the understanding of the physiology of the skin and of TST induration pathophysiology, a library of 12 chromophores was proposed as a superset of possible chromophores that may be present at the TST reaction site of the participants. With the aid of spectral unmixing and the induration masks generated in chapter 4, a subset of 6 candidate chromophores (oxyhaemoglobin, deoxyhaemoglobin, eumelanin, pheomelanin, water, and ferritin) were selected using sampled

spectral signatures from the reaction site of all 70 participants. The unmodelled residual chromophores for each participant (RES) was also estimated. The candidate chromophores presented evidence of their significant presence at the reaction site compared to other chromophores in the library. The evidence was in the form of a significant change in the median squared error with the inclusion of these chromophores in the unmixing model compared to other chromophores.

Applying spectral unmixing to the full hypercube of each participant, using only the candidate chromophores, revealed an array of chromophore maps per participant. The maps revealed clusters of relative high concentrations of chromophores as shown in Figure 15. There was a visually recognisable correlation between the appearance of these chromophore clusters and the perceived reaction site in the RGB estimates. Deterministic and deep radiomics features were generated to obtain quantitative estimates of the statistical and spatial relationships within each chromophore map. The ability of these features to predict the TST-based LTBI diagnosis of the participants was investigated using an SVM classifier.

The GLRM GLNU feature of FR and the total energy of WT were the best deterministic features that predict TST diagnosis. With a mean cross-validation accuracy and a mean test accuracy of 88.4% and 85.0%, as shown in Table 5.2. These features present evidence that the FR and WT chromophores are good representations of the pathophysiology of indurations across cohorts. The visualisation of the contribution maps in Figure 5.16 of the identified deterministic features of FR and WT, present a qualitative description of how the spatial distribution of the two chromophores can inform TST diagnosis. The figure suggests that the low level of nonuniformity of FR clusters is a good predictor of TST diagnosis. It also suggests that combining this FR feature with the energy of WT clusters significantly augments the predictive accuracy.

The higher cross-validation and test accuracies of 91.2% and 96.0% achieved for only the South African participants, using the MAD feature of OX and the variance of RES, is a sign of the influence of cohort-specific characteristics. The identification of deterministic features of WT as the best TST predictor for the Vietnamese cohort, which has a different skin tone compared to the South African cohort, suggests that skin tone might be the cause of the difference in the best predictive chromophore. The unmixing of eumelanin and pheomelanin from the other chromophores was expected to abstract the influence of skin tone from other chromophores. However, it appears that the higher levels of melanin in South African participants might have impacted the level of WT absorption captured by spectral signatures during imaging. This is highlighted by the low contrast in the WT maps for South African participants in Figure 5.15 compared to the Vietnamese participants. The orientation of induration diameter reading, which informed the assigned LTBI class labels, might have had a significant impact on the

differences between the results of the cohorts. The estimation of eumelanin and pheomelanin maps was meant to isolate the influence of skin tone. This leads to the possibility that the differences in reading orientation between the cohorts may be the biggest influence differences in the identified biomarkers. Further investigations with larger sample sizes of both cohorts will be needed in future research to validate this observation.

As shown in Table 5.3, the identification of FR and WT deterministic features as the best TST predictor for all participants and for each cohort, after the application of the induration mask to remove normal skin regions from the chromophore maps, suggests that FR and WT are good cohort-invariant predictors of the TST. However, the lower mean cross-validation accuracy (85.2% vs 88.4%) and mean test accuracy (83.3% vs 85.0%) of this approach compared to the unmasked chromophore map approach suggests that contrasting the chromophore characteristics of normal skin region against that of the reaction site is a better predictor of the TST than considering only the reaction site. This is further corroborated by the reduction in test accuracy for only the South African cohort (96.0% vs 81.3%), and the test accuracy for the Vietnamese cohort (96.7% vs 90.0%), after the application of the masks.

Deep radiomics was applied, in place of deterministic radiomics, to also generate features from the candidate chromophore maps that were validated against the LTBI diagnosis of participants. The higher predictive accuracies of the features generated by pretrained DenseNet-121, compared to ResNet-50 and Inception-V3, as shown in Table 5.4, showcases its superiority as a feature extractor for TST indurations. It also serves as evidence that the output channels of a deep learning model pretrained on a large radiological dataset, like RadImageNet, can be a good feature extractor for HSI-generated chromophore maps.

The identification of FR and WT as the chromophores whose feature maps generated the highest cross -validation and test accuracies (86.5% and 90.0%) for all participants, as shown in Table 5.5. corroborates the identification of these same features by the deterministic approach. The identification of DX as a good predictor at cohort level, with predictive accuracies of 95.0% and 96.7% for the South African and Vietnamese participants, is a new insight not provided by the deterministic approach. Visual interpretation of the dark pixel clusters in DX maps, compared to corresponding bright clusters in OX maps, suggests that the DX maps represent more of the relative absence of DX at the reaction site rather than the presence of DX. Comparing DX to the negative of OX revealed a significant, but not perfect, correlation between the two chromophore maps. Thus, the relationship between DX and OX, which is based on the oxygenation of haemoglobin molecules, suggests that DX maps may encode significant information on the negative form of OX maps.

The identification of OX and WT, for the South African and Vietnamese participants, respectively, as the best supporting chromophore maps to augment the predictive accuracy of DX, corroborates the identification of OX and WT as the best chromophores for each cohort. This suggests that the identification of DX by DenseNet is layered on the identification of OX and WT as the best cohort-specific predictors of the TST. The identification of FR and WT as the most TST-predicting chromophore across cohort by DenseNet suggests that these chromophores are the most generalisable and cohort-invariant representations of TST indurations. The OX chromophore, whose features are also significantly embedded in DX maps, provide cohort specific characterisation of the TST. Thus, the predictive accuracies of DenseNet output features for the TST, corroborates the identification of FR, WT, and OX by the deterministic approach as the best chromophore biomarkers for TST indurations.

The visualisation of the heatmaps of the most predictive feature channel of DenseNet for the TST of all participants, as shown in Figure 5.20, suggests that a contrast of the polarity of attention between the reaction site and normal skin region is a good LTBI predictor. This corroborates the identification of the contrast of the nonuniformity between the reaction site and normal skin by deterministic radiomics as a predictive spatial feature for the TST across all participants.

The identification of FR by DenseNet as the most predictive single chromophore for all participants and for each cohort, after the application of induration masks, as shown in Table 5.6, corroborates the performance of the deterministic approach after the application of the masks. The identification of WT as the best supporting chromophore for all participants and for the Vietnamese participants, and OX as the best supporting chromophore for the South African participants, is identical to the best supporting chromophores identified by the deterministic approach. It also supports the observation of the deterministic approach that the masking minimises cohort-specific characteristics and that FR and WT are the most cohort-agnostic chromophore predictors for the TST.

The consistent identification of OX as the most predictive chromophore for IGRA diagnosis, using TST indurations, by deterministic and deep radiomics, as shown in Tables 5.7 and 5.8, highlights its reliability as a good representation of IGRA characteristics within a TST induration. This presentation of a chromophore representation of TST indurations that can predict IGRA diagnosis may be a valuable contribution to knowledge on LTBI. The requirement for significant laboratory resources for the IGRA, which leads to its higher cost compared to the TST, has been a reason for the preference for TST in LMICs. Hyperspectral imaging's potential ability to predict IGRA diagnosis, which does not always agree with TST diagnosis Apriani et al. (2024), using the cheaper and easier TST could be of interest for LTBI screening. Based on these preliminary findings, it is conceivable that future research could yield low-cost HSI devices capable of generating chromophore maps to characterize TST indurations

and predict IGRA outcomes without the need for extensive laboratory infrastructure. Such an advancement would bridge the diagnostic gap between the TST and IGRA—a breakthrough not achieved in previous studies. Although the OX chromophore demonstrated high predictive accuracy for IGRA status in a cohort of 21 South African participants, validation in larger and more diverse cohorts is essential. Overall, these early results suggest that HSI holds promise as a cost-effective intermediary diagnostic tool between TST and IGRA.

The identification of the FR and WT as the best chromophore for predicting TST diagnosis for all participants, and OX as the best chromophore for predicting LTBI using both TST and IGRA methods in the South African participants, informs the conclusion that FR, WT, and OX are the best chromophore representations of the pathophysiology of indurations. The importance of white blood cells, like macrophages and CD4+, in the formation of indurations, as described in the induration biopsy study by Vukmanovic-Stejic et al. (2006), and the relationship between the cells and FR, as described in Recalcati and Cairo (2021) and Moreira et al. (2020), validates the identification of FR as a potential biomarker for the LTBI. The fact that macrophages, which process and show tuberculin antigen fragments to CD4+ cells, secrete FR, suggests that HSI can facilitate the non-invasive observation of the activity of the macrophages in an induration site via FR estimation. The fact that the activation and proliferation of CD4+ cells, which have been previously sensitized to *Mtb* in LTBI positive participants, is mediated by the iron content of FR, as described in Yarosz et al. (2020) and Kumar et al. (2024), suggests that HSI could facilitate the observation of the activity of CD4+ cells at an induration site. The spatial resolution of the Specim camera was not in the micrometre scale which would have been necessary for direct visualisation of macrophages and CD4+ cells. However, the ability to capture the relative concentration of FR – a molecule that is related to both macrophages and CD4+ cells, presents HSI as a modality that opens a novel window into the observation of TST indurations. The demonstration of the ability of FR, estimated via HSI, to predict TST diagnosis is a novel contribution of this research to LTBI literature. It offers a potential new way of observing TST indurations which could facilitate a more precise diagnosis that the current Mantoux method.

The presence of oedema at an induration site due to the vasodilation caused by tumour necrosis factor-α secreted by activated CD4+ cells, as described in Vukmanovic-Stejic et al. (2006), supports the identification of WT in this study as a chromophore biomarker for TST diagnosis. The ability of HSI to estimate the relative concentrations of WT on an induration site, presents HSI as a modality that can facilitate the quantification of the extent of oedema on the induration site. The TST study by Fiz et al. (2015) demonstrated the ability of thermography to estimate the temperature map of an induration site as a proxy for observing the inherent oedema. On the other hand, this study presents evidence

that HSI can enable similar insights into the oedema in an induration via the estimation of WT concentration.

The occurrence of vasodilation during the formation of induration suggests that red blood cells are present at an induration site. The presence of these cells on the site may not directly differentiate between positive and negative TST case as the subdermal injury caused by the injection of tuberculin can also lead to the presence of red blood cells at the site. This leads to the presence of erythema as a confounder at a TST site. However, it is expected that the cellular activities that occur over the 48-to-72-hour duration of the formation of an induration would lead to higher oxygenation levels. Thus, the identification of OX as a one of the biomarkers for LTBI testing using TST and IGRA diagnosis makes sense as a proxy for the activities of the immune cells at the induration site. The need for oxygen, delivered by OX content of red blood cells to tissue at the reaction site, for the metabolic activities of macrophages and other cells, conforms with the identification of OX as a biomarker.

Positive IGRA diagnoses of participants were obtained by exposing veinous blood to *Mtb* for 16 to 24 hours in a test tube to allow for activated CD4+ cells to secrete IFN-γ and identifying participants with IFN-γ concentration ≥0.35 IU/mL. On the other hand, the induration biopsy studies of Vukmanovic-Stejic et al. (2006) reveal that the CD4+ cells that are activated during the formation of an induration, secrete IFN-γ to activate more macrophages at the site. This means that a TST induration comprises of IFN-γ in addition to other immune cells and molecules. The identification of OX as the chromophore biomarker for the IGRA, via HSI, suggests that there could be a relationship between oxygenation levels at an induration site and IFN-γ concentration levels. The level of agreement between the TST and the IGRA diagnosis of the subset of 21 participants in this study was 62% which is lower than the 74.7% quoted in studies such as Apriani et al. (2024). The ability for OX to predict IGRA from the TST suggests that HSI could bridge the gap between the two tests. Future investigations employing larger cohorts are necessary to rigorously evaluate this proposition.

The HSI framework developed and validated in this chapter, presents FR, WT, and OX chromophore maps as good digital representations of the underlying pathophysiology of TST indurations. Radiomics features were generated from these chromophores to provide quantitative measures that were validated against results obtained by clinicians using the TST and IGRA tests. The features yielded predictive accuracies of 90.0% for all participants, 96.0% for South African participants, and 96.7% for Vietnamese participants in TST diagnosis. The features of OX perfectly predicted IGRA diagnosis, albeit with a smaller sample size. Contribution maps and heatmaps of the radiomics features were generated to visualise the regions in FR, WT, and OX that contribute to predictive ability of the chromophores.

5.5. Conclusion

The results demonstrate the ability of HSI, facilitated by the characterisation framework, to noninvasively offer good insights into the subdermal physiological processes of TST indurations. Ferritin, water, and oxyhaemoglobin were identified by the characterisation framework as potential biomarkers for LTBI. These chromophores potentially serve as digital proxies for observing cellular extravasation, vasodilation, immune cell metabolism and activation, and oedema at a TST induration site. The study by Fiz et al. (2015) measured temperature, using long-wave infrared, as a proxy for the physiological activities of a TST reaction. On the other hand, this chapter shows that HSI can facilitate a more direct *in vivo* observation of three molecules that are potentially constituents of a TST induration. This could potentially position HSI ahead of thermography as the leading modality for non-invasive characterisation of TST indurations. The results presented here is evidence of the viability of this proposition, a first of its kind in LTBI discourse, which can be further validated in future studies.

CHAPTER 6

Discussion

This thesis addresses a critical research gap in the screening of LTBI: the absence of an optical imaging solution that enables quantitative, non-geometric characterisation of TST indurations in a digital, standardised, automatable, precise, and reproducible manner. The conventional Mantoux method of induration assessment lacks these attributes. In response, this study proposes the use of HSI as a digital modality capable of capturing a comprehensive snapshot of TST indurations, thereby facilitating repeatable analysis and post-application validation.

The study demonstrates the utility of HSI through the generation of segmentation masks, induration diameter measurements, chromophore distribution maps, radiomics feature extraction, and machine learning-based classification metrics. These outputs collectively illustrate the enhanced reproducibility and analytical depth introduced by HSI in TST interpretation. Additionally, the spectral signatures of indurations, obtained from 38 South African and 32 Vietnamese participants, offer a valuable digital repository for future investigations aimed at validating or expanding upon the findings presented here.

The digital snapshot of indurations captured by HSI is more information-rich than the snapshot captured by RGB imaging and by thermography. The HSI hypercubes represented each pixel of induration tissue as 204 wavelength reflectance intensities in the VNIR range. Conversely, RGB imaging, in studies such as Zang et al. (2024) and Parihar et al. (2021), represented each induration pixel by three colour intensities in the visible light range. Similarly, thermography, as demonstrated in Fiz et al. (2015), represents each induration pixel as a temperature value. Thus, from the perspective of raw information representation, HSI captures more information from TST indurations than the current imaging approaches for TST reading. From a library of 12 chromophores via HSI linear spectral unmixing, this study identified ferritin, water, and oxyhaemoglobin as chromophore biomarkers for TST indurations with a mean test accuracy of 90.0%, 96.0%, and 96.7% for all, South African, and Vietnamese participants, respectively. The high information density of hyperspectral data cubes enables the potential application of alternative analytical methods in future research to extract additional chromophores, which may further enhance the characterisation of TST indurations. Conversely, thermography presents temperature as a single biomarker for characterising indurations. Furthermore, none of the RGB imaging studies for TST indurations demonstrated the ability for RGB imaging to generate any biomarker for TST indurations other than induration diameter. This positions

HSI as a viable candidate for image-based big data analytics for TST indurations – research that is currently absent from literature.

The use of transcriptomics, the study of all RNA transcripts, or the transcriptome, within a cell, tissue, or organism, providing insights into gene expression and cellular function, facilitated by skin biopsy, to profile the local and systemic immune response in an induration at RNA level was investigated in studies like Tomlinson et al. (2011) and Turner et al. (2024). The richness of TST hypercubes can potentially facilitate multimodal analysis to investigate the relationship between spectral signatures and RNA profiles of indurations. Such research can position HSI as an image-based alternative to transcriptomics. Skin biopsies are invasive which makes them impractical for largescale LTBI screening. The impracticality is also compounded by the requirement for substantial laboratory resources to preserve and process biopsy samples. The IGRA is presented in literature and clinical practice as a more precise but costlier alternative to the TST. Like biopsies, it also requires significant laboratory resources. It is not as invasive as skin biopsies and generates a single biomarker – the concentration of IFN-γ in processed peripheral blood. Despite the crudeness of the Mantoux method, a TST induration is characterised by a rich collection of immune cells and molecules (including the IFN-γ measured in IGRA) that are the product of 48-72 hours of localised physiological processes. This means that an induration is potentially a richer and more accurate representation of LTBI state than IGRA which is based on peripheral blood. The inherent crudeness of the Mantoux method undermines the complex physiological information embedded within a TST induration. This research therefore presents HSI as a non-invasive modality capable of accessing and capturing this richness with enhanced precision.

The significant human input required in biopsy-enabled transcriptomics and IGRA limits their ability as automatable screening tools. On the other hand, the non-invasiveness of HSI makes it a more practical modality for largescale LTBI screening. The ability of HSI to capture a TST hypercube in less than 5 minutes per participant, with minimal additional infrastructure and without the need for tissue samples, as demonstrated in chapter 3, can potentially facilitate the acquisition of rich induration data from hundreds of participants per day. The direct digital nature of TST hypercubes, unlike skin biopsies which require some human effort to digitize into RNA transcriptomic profiles, makes it a more viable modality for automating LTBI screening. The ability for mathematical models and machine learning models to make predictive diagnosis of LTBI from TST hypercubes, as demonstrated in chapters 4 and 5, means they can enable automated real-time diagnosis of LTBI. The predictive accuracies of 90.0%, 96.0%, and 96.7% reported in chapter 5, for three cohorts, using machine learning models, is evidence for this. Thus, this research presents HSI as a potential modality for largescale biopsies for TST indurations for big data analytics, and a modality for largescale automated LTBI screening.

The clinical adoption of HSI as a modality for largescale LTBI screening and biopsies required evidence of reproducibility. The possible clinical utility of HSI as a modality for large-scale LTBI screening and biopsy procedures necessitates demonstrable evidence of reproducibility. To address this requirement, Chapter 3 investigated the reproducibility of spectral signatures at TST injection sites under controlled illumination conditions, thereby providing the foundational validation for HSI's clinical utility. Also investigated were factors that negatively impact the reproducibility of the spectral signatures. Primary light source type and the spatial distribution of all secondary illumination were identified as factors that affect reproducibility. The investigation was done on a wide range of skin tones to ascertain that the illumination factors persist across skin tones. An imaging protocol derived from the protocol of the Specim IQ HSI camera, presented in Behmann et al. (2018), was developed to investigate the issue of spectra reproducibility as well as to mitigate the illumination factors that impact spectra reproducibility.

An experimental enclosure was developed to assess the impact of different illumination sources on the reproducibility of spectral signatures from TST injection sites. The findings revealed that the use of white LED lighting significantly compromised reproducibility, with a reduction in spectral correlation of 0.23 compared to the reference halogen-enabled signature. In contrast, broadband LED and xenon light sources produced minimal reductions in correlation—0.03 and 0.00, respectively—across a diverse range of skin tones. These results demonstrate that reproducible spectral signatures can be reliably acquired using halogen, xenon, or broadband LED as the primary illumination source.

The chapter also examined the influence of confounding secondary illumination, such as ambient light, on spectral reproducibility. The introduction of a single secondary illumination source led to a correlation drop of 0.10 relative to the reference spectra, highlighting the sensitivity of hyperspectral measurements to uncontrolled environmental lighting conditions.

To mitigate the effects of confounding illumination, the imaging protocol was modified through the incorporation of reflectors flanking the participant's arm within the enclosure and the strategic placement of a white reference panel between the reflectors. This enhancement significantly improved spectral consistency, reducing the difference in correlation from 0.10 under a single confounding illumination source to 0.00 even in the presence of three such sources.

The chapter thus presents evidence that HSI can produce reproducible digital snapshots of TST injection sites when supported by a carefully designed imaging protocol. These findings further imply that the reproducibility of downstream TST readings can be sustained through the application of mathematical correction techniques and machine learning algorithms.

Establishing HSI as a potential new approach for TST induration assessment necessitates evidence demonstrating that its outputs are comparable to those of clinician-derived readings. To address this, Chapter 4 investigates the capability of HSI to generate induration measurements consistent with clinical assessments, thereby providing critical validation of its suitability as a replacement for the traditional Mantoux method. The induration diameter measurement from the Mantoux method is a crude one-dimensional approximation of the underlying subdermal coagulated volume of a wide range of molecules and immune cells. The goal of the chapter was to demonstrate that HSI can also generate a similar but more precise one-dimensional approximation. The subjectivity of Mantoux readings is borne out of the difficulty faced by clinicians to precisely identify the boundaries of an induration. A PCA-based model was developed to cluster the spectral signatures of TST sites with the goal of identifying a consistent cluster across participants that correspond to the underlying volume of molecules and immune cells. A candidate cluster was identified and its viability as a representation of underlying induration was validated by the cross-validation and test accuracies of 82% and 93%, respectively - significantly higher than the accuracies produced by other clusters. The segmentation of the candidate cluster across participants resulted in the generation of masks that represent a reproducible visualisation of the perimeter of an induration. This, in itself, is a useful output of this research as it gives clinicians a precise visualisation of induration perimeter that they imprecisely access via palpations and tactile cues.

During Mantoux readings, there is typically variation in axis of measurement, even by the same clinician. This also contributes to this subjectivity – given that the perimeter of indurations is rarely perfectly circular. Induration segmentation masks enabled the use of a repeatable axis of diameter measurement. It also enabled the generation of a more robust diameter measurement in the form of the median diameter of a set of diameters 30^{o} about the measurement axis. The mean and median difference of 0.29 mm and 0.85 mm between the HSI-enabled diameter estimations and corresponding Mantoux readings, respectively, for 70 participants, was evidence of comparability. These estimations are within the inter-observer deviation of 3.7 mm reported in Pouchot et al. (1997), which validates the comparability. The ICC score of 0.80, which implies a good level of agreement, is additional evidence of comparability. Furthermore, a Pearson's correlation of 0.664 was obtained between the readings, which is similar to the correlation score of 0.65 obtained in the thermography approach in Fiz et al. (2015). This research presents HSI-based induration diameter estimations as a more precise and repeatable method for distance estimations of TST indurations, in place of the Mantoux method. This study explores the potential of HSI to produce precise measurements of TST induration diameters, offering a possible advancement in the scientific discourse surrounding LTBI screening and the limitations of the traditional Mantoux method.

The presentation of HSI as a potential modality to facilitate further investigations in TST skin biopsies and the IGRA required evidence to demonstrate its ability to generate biomarkers that provide consistent insights on the pathophysiology of TST indurations. Skin biopsies reveal that the activity of macrophages and CD4+ T cells leads to the production of cytokines like IFN-γ and TNF-α, which lead to the collection of water, red blood cells, and other immune cells at the injection site. In chapter 5, linear spectral unmixing was utilized to generate candidate chromophores from TST hypercubes of 70 participants. Radiomics and SVM-based feature selection was utilized to identify ferritin, water, and oxyhaemoglobin as the chromophores with the highest predictive accuracy for LTBI. They achieved accuracies of 90.0%, 96.0%, and 96.7% for all participants, 38 South African participants, and 32 Vietnamese participants, respectively. The relationship between these chromophores and the induration pathophysiology described by skin biopsy is evidence of the viability of HSI as a potential imaging modality to noninvasively facilitate largescale biopsies of indurations. The relationship between ferritin and macrophages—specifically, that ferritin is secreted by macrophages—has been reported in the literature (Recalcati & Cairo, 2021). The relationship between ferritin expression and $CD4^+$ T cell activity—namely, that the iron content of ferritin mediates the activation and proliferation of $CD4^+$ cells—has been highlighted in immunological studies (Yarosz et al. (2020); Kumar et al. (2024). The identification of water as a chromophore biomarker is supported by the skin biopsy description of the presence of oedema in indurations. The identification of oxyhaemoglobin as a chromophore biomarker is also supported by the presence of red blood cells in indurations due to the vascular permeability caused by TNF-α. The ability to noninvasively measure concentrations of these chromophores from TST indurations on a largescale is an edge that HSI potentially has over skin biopsies. Evidence for the viability of this proposition is the major contribution of this doctoral research to scientific discourse around LTBI management.

There is a need to directly validate the identified chromophore biomarkers against corresponding biopsy-generated molecular concentrations per participant in future studies. However, the relatability of the identified chromophores to induration pathophysiology encountered in literature presents compelling evidence of their viability for TST characterisation. With the application of the segmentation masks from chapter 4, ferritin was identified as the most predictive single chromophore for the South African and Vietnamese participants. Without the mask, there is a variation in the most predictive single chromophore for cohorts – oxyhaemoglobin for the South African and water for the Vietnamese participants. This variation opens a research question for future studies on the investigation of cohort-specific characteristics like baseline skin tone. However, the peak predictive accuracy of 90% achieved for both groups combined, using ferritin and water maps without the segmentation mask, suggests that the combination of these chromophores may account for cohort-

specific differences. The generation and visualisation of these features is a digital, standard, automatable, precise, and reproducible representation of indurations that potentially enables accurate LTBI screening that is independent of inter-observer variability.

The orientation of induration diameter reading by the South African clinician was along the longitudinal axis of the forearm and the orientation used by the Vietnamese clinicians was transverse. These differences in reading orientation, which informed the assigned LTBI class labels, might have had a significant impact on the differences between the results of the cohorts. The estimation of eumelanin and pheomelanin maps was meant to isolate the influence of skin tone. This suggests that there may be a significant influence of reading orientation on the target class of the feature selection algorithm and the downstream estimated biomarkers per cohort. Further investigation is required with larger cohort sizes to validate the effect of differences in measurement orientation.

The HSI approach proposed in this research can potentially be applied for the characterisation of indurations from the newer TBST whose development was motivated by the BCG-sensitivity of the TST. Like the TST, the TBST is currently read using the Mantoux reading method. Studies, such as To et al. (2024) and Lu et al. (2023), show that the same reading thresholds are used for both tests, which suggests that the HSI-based approach proposed in this research can potentially be applied to TBST indurations. However, future investigations are required to validate this proposition.

The novel demonstration of the ability of oxyhaemoglobin to predict IGRA diagnosis, on a subset of 21 South African participants, suggests that the gap in agreement between the TST and the IGRA, which is reported to be around 74.7% (Apriani et al., 2024), could be bridged using HSI. This research presents HSI as a modality that can potentially enable the estimation of IGRA diagnosis from TST indurations. Since IGRA involves the measurement of the concentration of IFN-γ in veinous blood exposed to *Mtb* antigens and IFN-γ is also present in a TST induration, the merits of both tests could potentially be combined using HSI. The chromophore-based evidence presented in this study opens the door to future research on the direct estimation of IFN-γ from the HSI images of indurations. The ability to directly estimate IFN-γ levels from a TST induration using low cost HSI devices, would eliminate the laboratory infrastructure required for the IGRA while preserving the simplicity and affordability of the tuberculin injection process of the TST. This opens the possibility of future research into the development of low-cost peripherals that can enable smartphones to capture usable hyperspectral images of indurations.To lay the groundwork for future research on the integration of transfer learning to merge HSI and RGB imaging for automated LTBI screening, an initial investigation was conducted on the utilization of telemedicine to transmit images to a remote server housing a diagnostic model. A telemedicine network architecture was proposed by the author in Oladokun et al.

(2022) for the secure and reliable transmission of hypercubes from the point of acquisition to a remotely located server. The study investigated the effect of batch transmission and band-by-band transmission of hypercubes on the performance of a set of mobile networks. The network architecture was modified by the author and reported in Oladokun et al. (2023) to include a hypercube compression model to enable the architecture to dynamically compress the spatial domain of hypercubes based on current network upload speed. This telemedicine system can include a branch in future research for the delivery of smartphone images of indurations to a remote server. The evidence of HSI-enabled precise diameter estimations and chromophore biomarker characterisation presented in this thesis suggests that future research on the merging of HSI with smartphone imaging of TST indurations using transfer learning could facilitate automated LTBI screening. This part of the work also addressed the need for improving accessibility to healthcare for rural, remote and underserved communities by leveraging the ubiquitous nature of cellular networks and the capabilities of telemedicine systems.

Evidence of the feasibility of transferability between HSI and RGB imaging in future studies can be gleaned from studies in other domains such as applied earth observations and remote sensing. Many studies on feature transfer between HSI and RGB, such as Yuan et al. (2017) and Wu et al. (2024), focus on hyperspectral super resolution. This involves transferring features from high resolution RGB images to improve the spatial resolution of hyperspectral images. The desired direction of feature transfer is the reverse for the TST application. Here it is desired that features be transferred from the spectral axis of hyperspectral images to improve the spectral resolution of 3-channel RGB images. The meta-transfer learning proposed by Cheng et al. (2024), which involves the development of a generalizable framework to transfer features between image domains of arbitrary spectral ranges, presents evidence of the feasibility of future applicability for the TST. This is further supported by the domain adaptation model by Chen et al. (2025) who propose a bilateral adaptation network that transfers knowledge in the forward and backward direction between a source and a target domain. There may exist a spectrally down-sampled configuration of the TST hypercubes with 204 bands that still enables the generation of the diameter estimation and induration characterisation results in chapters 4 and 5. Future investigations are required to identify this configuration.

A limitation of this research is the lack of a sample size that is large enough to enable a more rigorous clinical validation. Such a large sample size would also have enabled the utilization of more recent transformer-based deep learning models to extract features from hyperspectral images of indurations. A sample size in the thousands would be required to improve the clinical and methodological rigour of this research. An additional limitation of this research is the use of only two cohorts of participants. This is a noteworthy limitation due to the observed influence of a cohort-

specific characteristic, like skin tone, on the predictive accuracy of the identified HSI-enabled biomarkers. A larger number of cohorts would be required to obtain hyperspectral images of a wider range of baseline skin tones. This would facilitate a more rigorous search for HSI-enabled biomarkers for LTBI that are skin tone agnostic and generalise well for a larger population. Moreover, realizing the potential of HSI to support large-scale induration biopsies or to predict IGRA results from indurations would necessitate a larger sample size for rigorous clinical validation with high statistical power. Despite these limitations, this research provides preliminary but novel evidence of the potential of HSI for LTBI diagnosis which can be explored in future studies. Although constrained by these limitations, this study offers novel preliminary evidence of the potential of HSI in LTBI diagnosis, inviting further exploration in future research.

This chapter describes and amalgamates the contributions of chapters 3, 4, and 5 to the aim of the study. It outlines the significance of the results obtained in the chapters to the completion of the objectives of developing an HSI protocol for spectral reproducibility, developing an HSI model for induration diameter estimation, and identifying HSI-enabled biomarkers for LTBI. These findings present preliminary evidence to support the hypothesis H_1 which states that *Hyperspectral imaging, aided by machine learning, can capture clinically important subdermal features of TST indurations indicative of the underlying tuberculin-triggered immune response, which could address the inter-observer variability in the Mantoux TST method.*

CHAPTER 7
Conclusion

This thesis provides compelling evidence to support the hypothesis that HSI, aided by ML, can capture clinically important subdermal features of TST indurations indicative of underlying tuberculin-triggered immune response. The presented methodology for chromophore biomarker identification tested the hypothesis, and the results were supporting evidence of the ability of HSI to address the subjectivity of the TST. An outcome of this research is the precise HSI-enabled diameter estimation of TST indurations – a first of its kind in literature. A potentially more significant research outcome is the novel non-invasive chromophore-based characterisation of indurations that provides more insights than Mantoux readings and thermal imaging. This positions HSI as a potential modality for image-based LTBI diagnosis.

The identification of ferritin, water, and oxyhaemoglobin as chromophore biomarkers for LTBI diagnosis in this study presents evidence of the viability of HSI-generated chromophore maps as digital representations of induration pathophysiology. The high information density of hyperspectral data cubes enables the potential application of alternative analytical methods in future research to extract additional chromophores, which may further enhance the characterisation of TST indurations. The non-invasiveness of HSI contributes to its potential scalability and introduces the possibility of big data analytics of TST induration hypercubes. Furthermore, the relationship between the identified chromophore biomarkers and the induration pathophysiology described by skin biopsy is evidence of the viability of HSI as a potential imaging modality to noninvasively facilitate largescale biopsies of indurations. The potential of HSI to enable new investigations in LTBI is further strengthened by the preliminary evidence presented in this study of the ability of HSI to bridge the gap between the TST and the IGRA tests.

The generation and visualisation of HSI-enabled features is a digital, standard, automatable, precise, and reproducible representation of indurations that potentially enables accurate LTBI screening that is independent of inter-observer variability. The hyperspectral images of TST indurations of 70 participants, from two ethnicities, represents a digitized LTBI screening dataset for repeat analysis and future novel research approaches. The potential impact of this research is improved health outcome for at-risk TB groups via enhanced LTBI screening accuracy and precision. This study also includes preliminary work to address the need for improving accessibility to healthcare for rural, remote and

underserved communities by leveraging the ubiquitous nature of cellular networks and the capabilities of telemedicine systems.

A limitation in the approach used in current RGB imaging TST studies is the use of clinician annotations on RGB images of indurations to perform segmentation. The method presented in chapter 4 can potentially be used to generate HSI-enabled induration masks which can serve as ground truth for RGB imaging approaches for TST. Chapter 4 presents evidence that the HSI masks capture subdermal induration features which may be absent in clinician annotations on RGB images. Thus, RGB imaging of TST indurations can potentially be made more accurate using HSI masks as ground truth for model training. Furthermore, using the chromophore maps presented in chapter 5 as ground truth can facilitate the development of models that approximate chromophores from RGB images of indurations. The presented evidence for the viability of HSI for TST characterisation can inspire future research on the development of low-cost HSI devices. This may be realised by developing low-cost peripherals that can be attached to a smartphone to enable it to capture hyperspectral images.

With a patient dropout rate of around 17%, as reported in Faujdar et al. (2022), due to a lack of follow-up for Mantoux readings after initial tuberculin injection, there is a need for future research that addresses the secondary visit requirement for the TST. The development of low-cost multispectral peripherals for smartphone cameras is a potential research area to address this need. Another future perspective in TB discourse worth considering is the application of HSI in sputum smear analysis for active TB diagnosis. Microscopic image analysis is the current approach for extracting diagnostic information from the sputum of active TB patients, as described in Saffari et al. (2017) and Karlita et al. (2024). The application of the methods of this research on the hyperspectral images of sputum smears in future studies has the potential to result in new characterisations and insights.

Cohort sample size is a limitation of this research. It is recommended that future research generate HSI dataset with more samples from wide range of ethnicities. This would enable training with the use of large deep learning models. It will also facilitate more investigation into the possibility of cohort-specific reactions to the TST.

REFERENCES

Abbasi, N. R., Shaw, H. M., Rigel, D. S., Friedman, R. J., McCarthy, W. H., Osman, I., Kopf, A. W., & Polsky, D. (2004). Early diagnosis of cutaneous melanoma revisiting the ABCD criteria. *JAMA*, *292*(22), 2771-2776. https://doi.org/10.1001/jama.292.22.2771

Abbaspour, N., Hurrell, R., & Kelishadi, R. (2014). Review on iron and its importance for human health. *J Res Med Sci*, *19*(2), 164-174.

Abdlaty, R., Abbass, M. A., & Awadallah, A. M. (2021). High Precision Monitoring of Radiofrequency Ablation for Liver Using Hyperspectral Imaging. *Annals of Biomedical Engineering*, *49*(9), 2430-2440. https://doi.org/10.1007/s10439-021-02797-w

Abdlaty, R., Doerwald-Munoz, L., Farrell, T. J., Hayward, J. E., & Fang, Q. (2021). Hyperspectral imaging assessment for radiotherapy induced skin-erythema: Pilot study. *Photodiagnosis and Photodynamic Therapy*, *33*, 102195. https://doi.org/10.1016/j.pdpdt.2021.102195

Aboughaleb, I. H., Aref, M. H., & El-Sharkawy, Y. H. (2020). Hyperspectral imaging for diagnosis and detection of ex-vivo breast cancer. *Photodiagnosis and Photodynamic Therapy*, *31*, 101922. https://doi.org/10.1016/j.pdpdt.2020.101922

Agiomavriti, A.-A., Nikolopoulou, M. P., Bartzanas, T., Chorianopoulos, N., Demestichas, K., & Gelasakis, A. I. (2024). Spectroscopy-based methods and supervised machine learning applications for milk chemical analysis in dairy ruminants. *Chemosensors*, *12*(12), 263. https://doi.org/10.3390/chemosensors12120263

Akinola, O. A., Orimolade, J. F., Afolabi, A. S., Shopeju, H. K., Adetiba, E., & Adewale, A. A. (2024). Analysis and measurement of tuberculin skin test induration using deep neural network. *International Journal of Online & Biomedical Engineering*, *20*(12). https://doi.org/10.3991/ijoe.v20i12.47773

Al Jahdali, H. H., Baharoon, S., Abba, A. A., Memish, Z. A., Alrajhi, A. A., AlBarrak, A., Haddad, Q. A., Al Hajjaj, M., Pai, M., & Menzies, D. (2010). Saudi guidelines for testing and treatment of latent tuberculosis infection. *Annals of Saudi Medicine*, *30*(1), 38-49. https://doi.org/10.4103/0256-4947.59373

Alkawaz, M. H., Mohamad, D., Saba, T., Basori, A. H., & Rehman, A. (2015). The correlation between blood oxygenation effects and human emotion towards facial skin colour of virtual human. *3D Research*, *6*, 1-16. https://doi.org/10.1007/s13319-015-0044-9

Amadasun, M., & King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(5), 1264-1274. https://doi.org/10.1109/21.44046

Andersen, P., Munk, M. E., Pollock, J. M., & Doherty, T. M. (2000). Specific immune-based diagnosis of tuberculosis. *The Lancet*, *356*(9235), 1099-1104. https://doi.org/https://doi.org/10.1016/S0140-6736(00)02742-2

Apriani, L., McAllister, S., Sharples, K., Aini, I. N., Nurhasanah, H., Ratnaningsih, D. F., Indrati, A. R., Ruslami, R., Alisjahbana, B., van Crevel, R., & Hill, P. C. (2024). Tuberculin skin test and Interferon-gamma release assay agreement, and associated factors with latent tuberculosis infection, in medical and nursing students in Bandung, Indonesia. *PLOS ONE*, *19*(3), e0299874. https://doi.org/10.1371/journal.pone.0299874

Arango Duque, G., & Descoteaux, A. (2014). Macrophage cytokines: involvement in immunity and infectious diseases. *Front Immunol*, *5*, 491. https://doi.org/10.3389/fimmu.2014.00491

Balas, C. (2009). Review of biomedical optical imaging—a powerful, non-invasive, non-ionizing technology for improving in vivo diagnosis. *Measurement Science and Technology*, *20*(10), 104020. https://doi.org/10.1088/0957-0233/20/10/104020

Bannister, S., Sudbury, E., Villanueva, P., Perrett, K., & Curtis, N. (2021). The safety of BCG revaccination: A systematic review. *Vaccine*, *39*(20), 2736-2745. https://doi.org/10.1016/j.vaccine.2020.08.016

Barberio, M., Maktabi, M., Gockel, I., Rayes, N., Jansen-Winkeln, B., Köhler, H., Rabe, S. M., Seidemann, L., Takoh, J. P., & Diana, M. (2018). Hyperspectral based discrimination of thyroid and parathyroid during surgery. *Current Directions in Biomedical Engineering*, *4*(1), 399-402. https://doi.org/10.1515/cdbme-2018-0095

Baumann, L., Bernstein, E. F., Weiss, A. S., Bates, D., Humphrey, S., Silberberg, M., & Daniels, R. (2021). Clinical relevance of elastin in the structure and function of skin. In *Aesthetic Surgery Journal Open Forum* (Vol. 3, pp. ojab019). Oxford University Press US. https://doi.org/10.1093/asjof/ojab019

Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., Bohnenkamp, D., Kuska, M. T., Jussila, J., & Salo, H. (2018). Specim IQ: Evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. *Sensors*, *18*(2), 441. https://doi.org/10.3390/s18020441

Berg, J., Blumberg, E. J., Sipan, C. L., Friedman, L. S., Kelley, N. J., Vera, A. Y., Hofstetter, C. R., & Hovell, M. F. (2004). Somatic complaints and isoniazid (INH) side effects in Latino adolescents with latent tuberculosis infection (LTBI). *Patient Education and Counseling*, *52*(1), 31-39. https://doi.org/10.1016/S0738-3991(02)00268-9

Bonnans, M., Fouque, L., Pelletier, M., Chabert, R., Pinacolo, S., Restellini, L., & Cucumel, K. (2020). Blue light: Friend or foe? *Journal of Photochemistry and Photobiology B: Biology*, *212*, 112026. https://doi.org/10.1016/j.jphotobiol.2020.112026

Botina-Monsalve, D., Galeano, J., Torres-Madronero, M. C., Zarzycki, A., Garzón, J., Murillo, J., Robledo, S. M., & Marzani, F. (2020). Analysis of cutaneous leishmaniasis hyperspectral images by means of an inverse modeling procedure. In *15th International Symposium on Medical Information Processing and Analysis* (Vol. 11330, pp. 252-261). SPIE. https://doi.org/10.1117/12.2542135

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition* (pp. 3121-3124). IEEE. https://doi.org/10.1109/ICPR.2010.764

Calin, M. A., Manea, D., Savastru, R., & Parasca, S. V. (2023). Mapping the Distribution of Melanin Concentration in Different Fitzpatrick Skin Types Using Hyperspectral Imaging Technique. *Photochem Photobiol*, *99*(3), 1020-1027. https://doi.org/10.1111/php.13725

Cámara, M., Fernández-Ruiz, V., Redondo, D. F., Sánchez-Mata, M. C., & Torrecilla, J. S. (2012). Radial basis network analysis to estimate lycopene degradation kinetics in tomato-based products. *Food Research International*, *49*(1), 453-458. https://doi.org/10.1016/j.foodres.2012.07.030

Carranza, C., Pedraza-Sanchez, S., de Oyarzabal-Mendez, E., & Torres, M. (2020). Diagnosis for latent tuberculosis infection: New alternatives. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.02006

Chee, C. B. E., Reves, R., Zhang, Y., & Belknap, R. (2018). Latent tuberculosis infection: Opportunities and challenges. *Respirology*, *23*(10), 893-900. https://doi.org/10.1111/resp.13346

Chen, J., Wang, X., Wu, Q., & Mo, J. (2020). Skin melanoma detection based on hyperspectral imaging and deep-learning techniques. In *Optics in Health Care and Biomedical Optics X* (Vol. 11553, pp. 115532U). https://doi.org/10.1117/12.2575269

Chen, W., Wen, Y., Zheng, J., Huang, J., & Fu, H. (2025). BAN: A Universal Paradigm For Cross-Scene Classification Under Noisy Annotations From RGB And Hyperspectral Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2025.3539499

Chen, Y., Chen, H., Chao, W., Fu, Y., Wu, S., Wu, J., Chen, H., Jen, K., & Lui, P. (2018). Non-invasive assessment of cellulitis from snapshot hyperspectral imaging - A primary study. *Skin Res Technol*, *24*(2), 343-346. https://doi.org/10.1111/srt.12428

Cheng, W., Sun, D.-W., Pu, H., & Wei, Q. (2018). Characterization of myofibrils cold structural deformation degrees of frozen pork using hyperspectral imaging coupled with spectral angle mapping algorithm. *Food Chemistry*, *239*, 1001-1008. https://doi.org/10.1016/j.foodchem.2017.07.011

Cheng, Y., Wang, X., Ma, Y., Mei, X., Wu, M., & Ma, J. (2024). General hyperspectral image super-resolution via meta-transfer learning. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2024.3387970

Chiang, N., Jain, J. K., Sleigh, J., & Vasudevan, T. (2017). Evaluation of hyperspectral imaging technology in patients with peripheral vascular disease. *Journal of Vascular Surgery*, *66*(4), 1192-1201. https://doi.org/10.1016/j.jvs.2017.02.047

Cho, S., Shin, M. H., Kim, Y. K., Seo, J.-E., Lee, Y. M., Park, C.-H., & Chung, J. H. (2009). Effects of Infrared Radiation and Heat on Human Skin Aging in vivo. *Journal of Investigative Dermatology Symposium Proceedings*, *14*(1), 15-19. https://doi.org/https://doi.org/10.1038/jidsymp.2009.7

Chu, A., Sehgal, C. M., & Greenleaf, J. F. (1990). Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, *11*(6), 415-419. https://doi.org/10.1016/0167-8655(90)90112-F

Cui, R., Yu, H., Xu, T., Xing, X., Cao, X., Yan, K., & Chen, J. (2022). Deep learning in medical hyperspectral images: A Review. *Sensors*, *22*(24), 9790. https://doi.org/10.3390/s22249790

Daeschlein, G., Langner, I., Wild, T., von Podewils, S., Sicher, C., Kiefer, T., & Jünger, M. (2017). Hyperspectral imaging as a novel diagnostic tool in microcirculation of wounds. *Clinical Hemorheology and Microcirculation*, *67*(3-4), 467-474. https://doi.org/10.3233/CH-179228

Darvin, M. E., Sterry, W., Lademann, J., & Vergou, T. (2011). The role of carotenoids in human skin. *Molecules*, *16*(12), 10491-10506. https://doi.org/10.3390/molecules161210491

Deda, A., Lipka-Trawińska, A., Błońska-Fajfrowska, B., Odrzywołek, W., Lebiedowska, A., Hartman-Petrycka, M., Wcisło-Dziadecka, D., & Wilczyński, S. (2024). Methods of quantitative assessment of the response of dilated skin blood vessels to high-energy light treatments. *Journal of Clinical Medicine*, *13*(24), 7547. https://doi.org/10.3390/jcm13247547

Dendere, R., Mutsvangwa, T., Goliath, R., Rangaka, M. X., Abubakar, I., & Douglas, T. S. (2017). Measurement of Skin Induration Size Using Smartphone Images and Photogrammetric Reconstruction: Pilot Study. *JMIR Biomed Eng*, *2*(1), e3. https://doi.org/10.2196/biomedeng.8333

Denholm, J. T., McBryde, E. S., Eisen, D. P., Penington, J. S., Chen, C., & Street, A. C. (2014). Adverse effects of isoniazid preventative therapy for latent tuberculosis infection: a prospective cohort study. *Drug, Healthcare and Patient Safety*, *6*, 145-149. https://doi.org/10.2147/DHPS.S68837

Di Cecilia, L., Marazzi, F., & Rovati, L. (2018). Spectral repeatability of a hyperspectral system for human iris imaging. In *2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI)* (pp. 1-5). IEEE. https://doi.org/10.1109/RTSI.2018.8548513

Draper, B. A., Baek, K., Bartlett, M. S., & Beveridge, J. R. (2003). Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding*, *91*(1), 115-137. https://doi.org/10.1016/S1077-3142(03)00077-8

Dunn Iii, W., Scott, D., & Glen, W. (1989). Principal components analysis and partial least squares regression. *Tetrahedron Computer Methodology*, *2*(6), 349-376. https://doi.org/10.1016/0898-5529(89)90004-3

Fabelo, H., Melián, V., Martínez, B., Beltrán, P., Ortega, S., Marrero, M., Callicó, G. M., Sarmiento, R., Castaño, I., & Carretero, G. (2019). Dermatologic hyperspectral imaging system for skin cancer diagnosis assistance. In *2019 XXXIV Conference on Design of Circuits and Integrated Systems (DCIS)* (pp. 1-6). IEEE. https://doi.org/10.1109/DCIS201949030.2019.8959869

Farao, J., Malila, B., Conrad, N., Mutsvangwa, T., Rangaka, M. X., & Douglas, T. S. (2020). A user-centred design framework for mHealth. *PLOS ONE*, *15*(8), e0237910. https://doi.org/10.1371/journal.pone.0237910

Faujdar, S. S., Singh, U., Kumar, S., Mehrishi, P., Sharma, A., & Dutta, A. (2022). Mantoux test defaulters in rural population attending tertiary care hospital in a tuberculosis endemic area. *Journal of Family Medicine and Primary Care*, *11*(2), 677-679. https://doi.org/10.4103/jfmpc.jfmpc_1355_21

Fiz, J. A., Lozano, M., Monte-Moreno, E., Gonzalez-Martinez, A., Faundez-Zanuy, M., Becker, C., Pons-Rodriguez, L., & Ruiz Manzano, J. (2015). Tuberculine reaction measured by infrared thermography. *Computer Methods and Programs in Biomedicine*, *122*(2), 199-206. https://doi.org/10.1016/j.cmpb.2015.08.009

Flynn, J. L., Chan, J., & Lin, P. L. (2011). Macrophages and control of granulomatous inflammation in tuberculosis. *Mucosal Immunol*, *4*(3), 271-278. https://doi.org/10.1038/mi.2011.14

Giannoni, L., Lange, F., Sajic, M., Smith, K. J., & Tachtsidis, I. (2021). A hyperspectral imaging system for mapping haemoglobin and cytochrome-c-oxidase concentration changes in the exposed cerebral cortex. *IEEE Journal of Selected Topics in Quantum Electronics*, *27*(4), 1-11. https://doi.org/10.1109/JSTQE.2021.3053634

Goel, D., Mantan, M., & Sethi, G. R. (2018). Mantoux test revisited: Variability in reading tuberculin test in pediatric population. *The Journal of Infection in Developing Countries*, *12*(08), 625-630. https://doi.org/10.3855/jidc.9601

Gong, W., & Wu, X. (2021). Differential diagnosis of latent tuberculosis infection and active tuberculosis: a key to a successful tuberculosis control strategy. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.745592

Grambow, E., Dau, M., Sandkühler, N. A., Leuchter, M., Holmer, A., Klar, E., & Weinrich, M. (2019). Evaluation of peripheral artery disease with the TIVITA® Tissue hyperspectral imaging camera system. *Clinical Hemorheology and Microcirculation*, *73*(1), 3-17. https://doi.org/10.3233/CH-199215

Gualano, G., Mencarini, P., Lauria, F. N., Palmieri, F., Mfinanga, S., Mwaba, P., Chakaya, J., Zumla, A., & Ippolito, G. (2019). Tuberculin skin test – Outdated or still useful for Latent TB infection screening? *International Journal of Infectious Diseases*, *80*, S20-S22. https://doi.org/https://doi.org/10.1016/j.ijid.2019.01.048

Hald, M., Thyssen, J. P., Zachariae, C., Røpke, M. A., Carstensen, J. M., Schultz, N., & Johansen, J. D. (2019). Multispectral imaging of hand eczema. *Contact Dermatitis*, *81*(6), 438-445. https://doi.org/10.1111/cod.13377

Halicek, M., Fabelo, H., Ortega, S., Callico, G. M., & Fei, B. (2019). In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: revealing the invisible features of cancer. *Cancers*, *11*(6), 756. https://doi.org/10.3390/cancers11060756

Hamedani, G. G., & Tata, M. N. (1975). On the determination of the bivariate normal distribution from distributions of linear combinations of the variables. *The American Mathematical Monthly*, *82*(9), 913-915. https://doi.org/10.2307/2318494

Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

He, R., Xiao, K., Pointer, M., & Westland, S. (2020). Assessing skin tone heterogeneity under various light sources. In *Proceedings of IS&T London Imaging Meeting* (Vol. 1, pp. 5-9). Society for Imaging Science & Technology. https://doi.org/10.2352/issn.2694-118X.2020.LIM-15

Hetz, M. J., Garcia, C. N., Haggenmüller, S., & Brinker, T. J. (2024). Advancing dermatological diagnosis: Development of a hyperspectral dermatoscope for enhanced skin imaging. *arXiv preprint arXiv:2403.00612*.

Hirano, G., Nemoto, M., Kimura, Y., Kiyohara, Y., Koga, H., Yamazaki, N., Christensen, G., Ingvar, C., Nielsen, K., & Nakamura, A. (2020). Automatic diagnosis of melanoma using hyperspectral

data and GoogLeNet. *Skin Research and Technology*, *26*(6), 891-897. https://doi.org/10.1111/srt.12891

Holmer, A., Marotz, J., Wahl, P., Dau, M., & Kämmerer, P. W. (2018). Hyperspectral imaging in perfusion and wound diagnostics – methods and algorithms for the determination of tissue parameters. *Biomedical Engineering / Biomedizinische Technik*, *63*(5), 547-556. https://doi.org/doi:10.1515/bmt-2017-0155

Hook, L. A., Santhana Vannan, S., Beaty, T. W., Cook, R. B., & Wilson, B. E. (2010). Best practices for preparing environmental data sets to share and archive. https://daac.ornl.gov/datamanagement/BestPractices-2010.pdf

Horton, L., Brady, J., Kincaid, C. M., Torres, A. E., & Lim, H. W. (2023). The effects of infrared radiation on the human skin. *Photodermatology, Photoimmunology & Photomedicine*, *39*(6), 549-555. https://doi.org/10.1111/phpp.12899

Hosking, A. M., Coakley, B. J., Chang, D., Talebi-Liasi, F., Lish, S., Lee, S. W., Zong, A. M., Moore, I., Browning, J., & Jacques, S. L. (2019). Hyperspectral imaging in automated digital dermoscopy screening for melanoma. *Lasers in surgery and medicine*, *51*(3), 214-222. https://doi.org/10.1002/lsm.23055

Hossain, S., Satter, S., Kwon, T.-H., & Kim, K.-D. (2022). Optical measurement of molar absorption coefficient of HbA1c: comparison of theoretical and experimental results. *Sensors*, *22*(21), 8179. https://doi.org/10.3390/s22218179

Houben, R. M. G. J., & Dodd, P. J. (2016). The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLOS Medicine*, *13*(10), e1002152. https://doi.org/10.1371/journal.pmed.1002152

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708). https://doi.org/10.1109/CVPR.2017.243

Huang, H.-Y., Hsiao, Y.-P., Karmakar, R., Mukundan, A., Chaudhary, P., Hsieh, S.-C., & Wang, H.-C. (2023). A review of recent advances in computer-aided detection methods using hyperspectral imaging engineering to detect skin cancer. *Cancers*, *15*(23), 5634. https://doi.org/10.3390/cancers15235634

Huck, C. W. (2014). Advances of vibrational spectroscopic methods in phytomics and bioanalysis. *Journal of pharmaceutical and biomedical analysis*, *87*, 26-35. https://doi.org/10.1016/j.jpba.2013.05.010

Hwang, K., Kim, H., & Kim, D. J. (2016). Thickness of skin and subcutaneous tissue of the free flap donor sites: A histologic study. *Microsurgery*, *36*(1), 54-58. https://doi.org/10.1002/micr.30000

Izrah, M. (2024). A spotlight on the tuberculosis epidemic in South Africa. *National Communication*, *15*(1), 1290. https://doi.org/10.1038/s41467-024-45491-w

Jacobs, R. E. A., Gu, P., & Chachoua, A. (2015). Reactivation of pulmonary tuberculosis during cancer treatment. *International Journal of Mycobacteriology*, *4*(4), 337-340. https://doi.org/10.1016/j.ijmyco.2015.05.015

Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., & Yu, S. (2021). A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing*, *448*, 179-204. https://doi.org/10.1016/j.neucom.2021.03.035

Johansen, T., Møllersen, K., Ortega, S., Fabelo, H., Garcia, A., Marrero Callico, G., & Godtliebsen, F. (2019). Recent advances in hyperspectral imaging for melanoma detection. *Wiley Interdisciplinary Reviews: Computational Statistics*. https://doi.org/10.1002/wics.1465

Johansen, T. H., Møllersen, K., Ortega, S., Fabelo, H., Garcia, A., Callico, G. M., & Godtliebsen, F. (2020). Recent advances in hyperspectral imaging for melanoma detection. *WIREs Computational Statistics*, *12*(1), e1465. https://doi.org/10.1002/wics.1465

Jung, S., & Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.*, *37*(6B), 4104 - 4130. https://doi.org/10.1214/09-AOS709

Karlita, T., Pastika, F. A., Ramadijanti, N., Kusumawati, R., Yuniarti, H., & Sigit, R. (2024). Mycobacterium Tuberculosis Detection on Sputum Smear Microscopic Images Using Attention-Based Multi-Scale Convolutional Neural Network. In *2024 International Electronics Symposium (IES)* (pp. 371-378). IEEE. https://doi.org/10.1109/IES63037.2024.10665871

Karthikeyan, B., George, D. J., Manikandan, G., & Thomas, T. (2020). A comparative study on k-means clustering and agglomerative hierarchical clustering. *International Journal of Emerging Trends in Engineering Research*, *8*(5). https://doi.org/10.30534/ijeter/2020/20852020

Kato, K., Nemoto, M., Kimura, Y., Kiyohara, Y., Koga, H., Yamazaki, N., Christensen, G., Ingvar, C., Nielsen, K., Nakamura, A., Sota, T., & Nagaoka, T. (2020). Performance improvement of automated melanoma diagnosis system by data augmentation. *Advanced Biomedical Engineering*, *9*, 62-70. https://doi.org/10.14326/abe.9.62

Kazianka, H., Leitner, R., & Pilz, J. (2008). Segmentation and classification of hyper-spectral skin data. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications* (pp. 245-252). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-78246-9_29

Kiazyk, S., & Ball, T. B. (2017). Latent tuberculosis infection: An overview. *Can Commun Dis Rep*, *43*(3-4), 62-66. https://doi.org/10.14745/ccdr.v43i34a01

Kohler, L. H., Köhler, H., Kohler, S., Langer, S., Nuwayhid, R., Gockel, I., Spindler, N., & Osterhoff, G. (2021). Hyperspectral Imaging (HSI) as a new diagnostic tool in free flap monitoring for soft tissue reconstruction: A proof of concept study. *BMC Surgery*, *21*(1), 222. https://doi.org/10.1186/s12893-021-01232-0

Konugolu Venkata Sekar, S., Bargigia, I., Dalla Mora, A., Taroni, P., Ruggeri, A., Tosi, A., Pifferi, A., & Farina, A. (2017). Diffuse optical characterization of collagen absorption from 500 to 1700 nm. *Journal of Biomedical Optics*, *22*, 015006. https://doi.org/10.1117/1.JBO.22.1.015006

Konugolu Venkata Sekar, S., Beh, J. S., Farina, A., Dalla Mora, A., Pifferi, A., & Taroni, P. (2017). Broadband diffuse optical characterization of elastin for biomedical applications. *Biophysical Chemistry*, *229*, 130-134. https://doi.org/10.1016/j.bpc.2017.07.004

Koonce, B. (2021). ResNet 50. In *Convolutional Neural Networks with Swift for Tensorflow* (pp. 63-72). https://doi.org/10.1007/978-1-4842-6168-2_6

Kounas, K., Dinh, T., Riemer, K., Rosenblum, B. I., Veves, A., & Giurini, J. M. (2023). Use of hyperspectral imaging to predict healing of diabetic foot ulceration. *Wound Repair and Regeneration*, *31*(2), 199-204. https://doi.org/10.1111/wrr.13071

Kreyszig, E., Stroud, K., & Stephenson, G. (2008). Advanced engineering mathematics. *Integration*, *9*(4).

Krutikov, M., Faust, L., Nikolayevskyy, V., Hamada, Y., Gupta, R. K., Cirillo, D., Mateelli, A., Korobitsyn, A., Denkinger, C. M., & Rangaka, M. X. (2022). The diagnostic performance of novel skin-based in-vivo tests for tuberculosis infection compared with purified protein derivative tuberculin skin tests and blood-based in vitro interferon-γ release assays: a systematic review and meta-analysis. *Lancet Infect Dis*, *22*(2), 250-264. https://doi.org/10.1016/s1473-3099(21)00261-9

Kumar, A., Ye, C., Nkansah, A., Decoville, T., Fogo, G. M., Sajjakulnukit, P., Reynolds, M. B., Zhang, L., Quaye, O., Seo, Y.-A., Sanderson, T. H., Lyssiotis, C. A., & Chang, C.-H. (2024). Iron regulates the quiescence of naive CD4 T cells by controlling mitochondria and cellular metabolism. *Proceedings of the National Academy of Sciences*, *121*(17), e2318420121. https://doi.org/10.1073/pnas.2318420121

Lai, M. I., Feng, L. W., Yap, B. K., George, E., & Abdullah, M. (2015). Rapid ferritin iron release using FMN reductase. *International Journal of Bioscience, Biochemistry and Bioinformatics*, *5*(2), 111-119. https://doi.org/10.17706/ijbbb.2015.5.2.111-119

Li, L., & Wei, X.-W. (2020). 5 - Recent advances in multifunctional nanoengineered biomaterials. In M. Mozafari (Ed.), *Nanoengineered Biomaterials for Advanced Drug Delivery* (pp. 91-106). Elsevier. https://doi.org/10.1016/B978-0-08-102985-5.00005-X

Li, Y., Shen, F., Hu, L., Lang, Z., Liu, Q., Cai, F., & Fu, L. (2023). A stare-down video-rate high-throughput hyperspectral imaging system and its applications in biological sample sensing. *IEEE Sensors Journal*, *23*(19), 23629-23637. https://doi.org/10.1109/JSEN.2023.3308394

Lihacova, I., Bondarenko, A., Chizhov, Y., Uteshev, D., Bliznuks, D., Kiss, N., & Lihachev, A. (2022). Multi-class CNN for classification of multispectral and autofluorescence skin lesion clinical images. *Journal of Clinical Medicine*, *11*(10), 2833. https://doi.org/10.3390/jcm11102833

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., & Tu, X. M. (2016). Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Arch Psychiatry*, *28*(2), 115-120. https://doi.org/10.11919/j.issn.1002-0829.216045

Liu, Z., Wang, H., & Li, Q. (2011). Tongue tumor detection in medical hyperspectral images. *Sensors*, *12*(1), 162-174. https://doi.org/10.3390/s120100162

Lu, B., Dao, P. D., Liu, J., He, Y., & Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, *12*(16), 2659. https://doi.org/10.3390/rs12162659

Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, *19*(1), 010901-010901. https://doi.org/10.1117/1.JBO.19.1.010901

Lu, P., Wu, K., Zhou, H., Yu, H., Yuan, J., Dong, L., Liu, Q., Ding, X., Lu, W., & Yang, H. (2023). Evaluation of ESAT6-CFP10 skin test for Mycobacterium tuberculosis infection among persons living with HIV in China. *Journal of Clinical Microbiology*, *61*(4), e01816-01822. https://doi.org/10.1128/jcm.01816-22

Maclean, S. (2020). Image analysis for a mobile phone-based assessment of latent tuberculosis infection. *University of Cape Town*. http://hdl.handle.net/11427/32471

Martinez, L., Cords, O., Liu, Q., Acuna-Villaorduna, C., Bonnet, M., Fox, G. J., Carvalho, A. C. C., Chan, P.-C., Croda, J., & Hill, P. C. (2022). Infant BCG vaccination and risk of pulmonary and extrapulmonary tuberculosis throughout the life course: a systematic review and individual participant data meta-analysis. *The Lancet Global Health*, *10*(9), e1307-e1316. https://doi.org/10.1016/S2214-109X(22)00283-2

Mayerhoefer, M. E., Materka, A., Langs, G., Häggström, I., Szczypiński, P., Gibbs, P., & Cook, G. (2020). Introduction to radiomics. *J Nucl Med*, *61*(4), 488-495. https://doi.org/10.2967/jnumed.118.222893

Mehraj, H., Taufique, T., Ona, A., Nusrat, A., & Jamal Uddin, A. (2014). Performance of gerbera cultivars under different wavelengths of solar spectrum. *Journal of Bangladesh Academy of Sciences*, *38*(1), 27-37. https://doi.org/10.3329/jbas.v38i1.20202

Mei, X., Liu, Z., Robson, P. M., Marinelli, B., Huang, M., Doshi, A., Jacobi, A., Cao, C., Link, K. E., Yang, T., Wang, Y., Greenspan, H., Deyer, T., Fayad, Z. A., & Yang, Y. (2022). RadImageNet: an open radiologic deep learning research dataset for effective transfer learning. *Radiology: Artificial Intelligence*, *4*(5), e210315. https://doi.org/10.1148/ryai.210315

Micheel, C. M., Nass, S. J., Omenn, G. S., & CROBTPPOCT. (2012). Omics-based clinical discovery: Science, technology, and applications. In *Evolution of Translational Omics: Lessons Learned and the Path Forward*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK202165

Mir, M. A., Mir, B., Kumawat, M., Alkhanani, M., & Jan, U. (2022). Manipulation and exploitation of host immune system by pathogenic Mycobacterium tuberculosis for its advantage. *Future Microbiol*, *17*, 1171-1198. https://doi.org/10.2217/fmb-2022-0026

Moayedi-Nia, S., Barss, L., Oxlade, O., Valiquette, C., Ly, M.-X., Campbell, J. R., Lan, Z., Nsengiyumva, P., Fregonese, F., Lisboa Bastos, M., Sampath, D., Winters, N., & Menzies, D. (2019). The mTST – An mHealth approach for training and quality assurance of tuberculin skin test

administration and reading. *PLOS ONE*, *14*(4), e0215240. https://doi.org/10.1371/journal.pone.0215240

Moosaei, H., Mousavi, A., Hladik, M., & Gao, Z. (2023). Sparse L1-norm quadratic surface support vector machine with Universum data. *Soft Computing*, *27*, 1-20. https://doi.org/10.1007/s00500-023-07860-3

Moreira, A. C., Mesquita, G., & Gomes, M. S. (2020). Ferritin: an inflammatory player keeping iron at the core of pathogen-host interactions. *Microorganisms*, *8*(4), 589. https://doi.org/10.3390/microorganisms8040589

Nachbar, F., Stolz, W., Merkle, T., Cognetta, A. B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., & Plewig, G. (1994). The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, *30*(4), 551-559. https://doi.org/10.1016/S0190-9622(94)70061-3

Nanni, L., Brahnam, S., Ghidoni, S., Menegatti, E., & Barrier, T. (2013). Different approaches for extracting information from the co-occurrence matrix. *PLOS ONE*, *8*(12), e83554. https://doi.org/10.1371/journal.pone.0083554

Naraghi, S., Mutsvangwa, T., Goliath, R., Rangaka, M. X., & Douglas, T. S. (2018). Mobile phone-based evaluation of latent tuberculosis infection: Proof of concept for an integrated image capture and analysis system. *Computers in Biology and Medicine*, *98*, 76-84. https://doi.org/10.1016/j.compbiomed.2018.05.009

Nayak, S., & Acharjya, B. (2012). Mantoux test and its interpretation. *Indian Dermatol Online J*, *3*(1), 2-6. https://doi.org/10.4103/2229-5178.93479

Neittaanmäki, N., Salmivuori, M., Pölönen, I., Jeskanen, L., Ranki, A., Saksela, O., Snellman, E., & Grönroos, M. (2017). Hyperspectral imaging in detecting dermal invasion in lentigo maligna melanoma. *British Journal of Dermatology*, *177*(6), 1742-1744. https://doi.org/10.1111/bjd.15267

Nijhawan, A. E., Iroh, P. A., Brown, L. S., Winetsky, D., & Porsa, E. (2016). Cost analysis of tuberculin skin test and the QuantiFERON-TB Gold In-tube test for tuberculosis screening in a correctional setting in Dallas, Texas, USA. *BMC Infectious Diseases*, *16*(1). https://doi.org/10.1186/s12879-016-1901-8

Nozari, H., & Sadeghi, M. E. (2021). Artificial intelligence and Machine Learning for Real-world problems (A survey). *International Journal of Innovation in Engineering*, *1*(3), 38-47. https://doi.org/10.59615/ijie.1.3.38

Nunez, J. H., Park, C., Clark, A., Akarichi, C., Arnoldo, B., Mandell, S. P., Carlson, D. L., Huebinger, R. M., Chan, R. K., & Goverman, J. (2022). 533 Human Case Characterizations of Skin Burn Using Novel Multi-Spectral Short Wave Infrared Imaging. *Journal of Burn Care & Research*, *43*(Supplement_1), S101-S102. https://doi.org/10.1093/jbcr/irac012.162

Oladokun, A. S., Malila, B., Campello, V. M., Shey, M., & Mutsvangwa, T. E. (2024). SpeChrOmics: A biomarker characterization framework for medical hyperspectral imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 745-756). Springer. https://doi.org/10.1007/978-3-031-72384-1_70

Oladokun, A. S., Malila, B., & Mutsvangwa, T. (2022). A Secure and Reliable Hyperspectral Imaging-Based Telemedicine Network Architecture. In *Southern Africa Telecommunication Networks and Applications Conference (SATNAC)*. SATNAC. https://doi.org/10.25375/uct.26339719.v1

Oladokun, A. S., Malila, B., & Mutsvangwa, T. (2023). A Network Resource-Aware Hyperspectral Imaging-Based Telemedicine Architecture for Rural, Remote, and Underserved Communities. In *Southern Africa Telecommunication Networks and Applications Conference (SATNAC)* (pp. 131-136). SATNAC. https://doi.org/10.25375/uct.26339734.v1

Oliveira, R., Ferreira, J., Azevedo, L. F., & Almeida, I. F. (2023). An overview of methods to characterize skin type: focus on visual rating scales and self-report instruments. *Cosmetics*, *10*(1), 14. https://doi.org/10.3390/cosmetics10010014

Oltulu, P., Ince, B., Kokbudak, N., Findik, S., & Kilinc, F. (2018). Measurement of epidermis, dermis, and total skin thicknesses from six different body regions with a new ethical histometric technique. *Turkish Journal of Plastic Surgery*, *26*(2). https://doi.org/10.4103/tjps.TJPS_2_17

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, *11*(285-296), 23-27.

Pahal, P., & Sharma, S. (2020). PPD skin test. In *StatPearls*. StatPearls Publishing.

Pardo, A., Gutiérrez-Gutiérrez, J. A., Lihacova, I., López-Higuera, J. M., & Conde, O. M. (2018). On the spectral signature of melanoma: a non-parametric classification framework for cancer detection in hyperspectral imaging of melanocytic lesions. *Biomedical Optics Express*, *9*(12), 6283-6301. https://doi.org/10.1364/BOE.9.006283

Parihar, G., Christopher, J., Joyce, Y. S., Lazarus, Y. B., & Dayalan, J. (2021). Measurement of skin test wheals using image segmentation approaches. In *2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA)* (pp. 395-400). https://doi.org/10.1109/ICCCA52192.2021.9666435

Paulus, S., & Mahlein, A.-K. (2020). Technical workflows for hyperspectral plant image assessment and processing on the greenhouse and laboratory scale. *GigaScience*, *9*. https://doi.org/10.1093/gigascience/giaa090

Pecha, M., & Horák, D. (2020). Analyzing l1-loss and l2-loss support vector machines implemented in PERMON toolbox. In *AETA 2018-Recent Advances in Electrical Engineering and Related Sciences: Theory and Application* (pp. 13-23). Springer International Publishing. https://doi.org/10.1007/978-3-030-14907-9_2

Pooransingh, S., & Sakhamuri, S. (2020). Need for BCG vaccination to prevent tb in high-incidence countries and populations. *Emerg Infect Dis*, *26*(3), 624-625. https://doi.org/10.3201/eid2603.191232

Pouchot, J., Grasland, A., Collet, C., Coste, J., Esdaile, J. M., & Vinceneux, P. (1997). Reliability of tuberculin skin test measurement. *Annals of Internal Medicine*, *126*(3), 210-214. https://doi.org/10.7326/0003-4819-126-3-199702010-00005

Prahl, S. (2017a). *Beta-carotene*. Retrieved 15/01/2025 from https://omlc.org/spectra/PhotochemCAD/html/041.html

Prahl, S. (2017b). *Bilirubin*. Retrieved 2024/07/03 from https://omlc.org/spectra/PhotochemCAD/html/119.html

Prahl, S. (2018). *Assorted Spectra*. Retrieved 2024/07/03 from https://omlc.org/spectra/index.html

Ramirez-GarciaLuna, J. L., Martinez-Jimenez, M. A., Fraser, R. D., Bartlett, R., Lorincz, A., Liu, Z., Saiko, G., & Berry, G. K. (2023). Is my wound infected? A study on the use of hyperspectral imaging to assess wound infection. *Frontiers in Medicine*, *10*, 1165281. https://doi.org/10.3389/fmed.2023.1165281

Randeberg, L. L., & Hernandez-Palacios, J. (2012). Hyperspectral imaging of bruises in the SWIR spectral region. In *Photonic Therapeutics and Diagnostics VIII* (Vol. 8207, pp. 129-138). SPIE. https://doi.org/10.1117/12.909137

Recalcati, S., & Cairo, G. (2021). Macrophages and Iron: A Special Relationship. *Biomedicines*, *9*(11). https://doi.org/10.3390/biomedicines9111585

Rey-Barroso, L., Burgos-Fernández, F. J., Delpueyo, X., Ares, M., Royo, S., Malvehy, J., Puig, S., & Vilaseca, M. (2018). Visible and extended near-infrared multispectral imaging for skin cancer diagnosis. *Sensors (Basel)*, *18*(5). https://doi.org/10.3390/s18051441

Saffari, M., Jolandimi, H. A., Sehat, M., Nejad, N. V., Hedayati, M., Zamani, M., & Ghasemi, A. (2017). Smear grading and the Mantoux skin test can be used to predict sputum smear conversion in patients suffering from tuberculosis. *GMS hygiene and infection control*, *12*, Doc12. https://doi.org/10.3205/dgkh000297

Sandgren, A., Vonk Noordegraaf-Schouten, M., van Kessel, F., Stuurman, A., Oordt-Speets, A., & van der Werf, M. J. (2016). Initiation and completion rates for latent tuberculosis infection

treatment: a systematic review. *BMC Infectious Diseases*, *16*(1), 204. https://doi.org/10.1186/s12879-016-1550-y

Sârbu, C., & Pop, H. (2005). Principal component analysis versus fuzzy principal component analysis: a case study: the quality of Danube water (1985–1996). *Talanta*, *65*(5), 1215-1220. https://doi.org/10.1016/j.talanta.2004.08.047

Shah, M., & Dorman, S. E. (2021). Latent tuberculosis infection. *New England Journal of Medicine*, *385*(24), 2271-2280. https://doi.org/10.1056/NEJMcp2108501

Shaikh, S., Akhter, N., & Manza, R. (2019). Current trends in the application of thermal imaging in medical condition analysis. *Int. J. Innov. Technol. Explor. Eng*, *8*(8), 2708-2712.

Sharma, S. K., Sharma, A., Kadhiravan, T., & Tharyan, P. (2014). Rifamycins (rifampicin, rifabutin and rifapentine) compared to isoniazid for preventing tuberculosis in HIV-negative people at risk of active TB. *Evidence-Based Child Health: A Cochrane Review Journal*, *9*(1), 169-294. https://doi.org/10.1002/ebch.1962

Shmilovici, A. (2010). Support vector machines. In O. Maimon, Rokach, L. (Ed.), *Data Mining and Knowledge Discovery Handbook* (pp. 231-247). Springer. https://doi.org/10.1007/978-0-387-09823-4_12

Shukla, R., Dubey, A., Pandey, V., Golhani, D., & Jain, A. P. (2012). Chromophore-an utility in uv spectrophotometer. *Inventi Rapid: Pharm Ana & Qual Assur*, *2012*(3), 334.

Singla, K., MP, S. K., Bhattacharjya, A., Saxena, R., Choudhary, N., & Goyal, B. (2023). Bilirubin in wound healing: A double-edged sword. *Cell Biochemistry and Function*, *41*(8), 953-958. https://doi.org/10.1002/cbf.3849

Stamatas, G. N., & Kollias, N. (2006). Noninvasive quantitative documentation of cutaneous inflammation in vivo using spectral imaging. In *Photonic Therapeutics and Diagnostics II* (Vol. 6078, pp. 111-118). SPIE. https://doi.org/10.1117/12.647078

Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-77242-4

Sterling, T. R., Njie, G., Zenner, D., Cohn, D. L., Reves, R., Ahmed, A., Menzies, D., Horsburgh, C. R., Crane, C. M., Burgos, M., LoBue, P., Winston, C. A., & Belknap, R. (2020). Guidelines for the treatment of latent tuberculosis infection: Recommendations from the National Tuberculosis Controllers Association and CDC, 2020. *American Journal of Transplantation*, *20*(4), 1196-1206. https://doi.org/10.1111/ajt.15841

Sun, C., & Wee, W. G. (1983). Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, *23*(3), 341-352. https://doi.org/10.1016/0734-189X(83)90032-4

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826). https://doi.org/10.1109/CVPR.2016.308

Thibault, G., Angulo, J., & Meyer, F. (2013). Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Transactions on Biomedical Engineering*, *61*(3), 630-637. https://doi.org/10.1109/TBME.2013.2284600

To, K. W., Zhang, R., & Lee, S. S. (2024). Is the new tuberculous antigen-based skin test ready for use as an alternative to tuberculin skin test/interferon-gamma release assay for tuberculous diagnosis? A narrative review. *International Journal of Infectious Diseases*, 106992. https://doi.org/10.1016/j.ijid.2024.106992

Toivgoogiin, A., Toyota, M., Yasuda, N., & Ohara, H. (2005). Validity of using tuberculin skin test erythema measurement for contact investigation during a tuberculosis outbreak in schoolchildren previously vaccinated with BCG. *Journal of Epidemiology*, *15*(2), 56-64. https://doi.org/10.2188/jea.15.56

Tomlinson, G. S., Cashmore, T. J., Elkington, P. T., Yates, J., Lehloenya, R. J., Tsang, J., Brown, M., Miller, R. F., Dheda, K., Katz, D. R., Chain, B. M., & Noursadeghi, M. (2011). Transcriptional

profiling of innate and adaptive human immune responses to mycobacteria in the tuberculin skin test. *Eur J Immunol*, *41*(11), 3253-3260. https://doi.org/10.1002/eji.201141841

Torti, E., Leon, R., La Salvia, M., Florimbi, G., Martinez-Vega, B., Fabelo, H., Ortega, S., Callicó, G. M., & Leporati, F. (2020). Parallel classification pipelines for skin cancer detection exploiting hyperspectral imaging on hybrid systems. *Electronics*, *9*(9), 1503. https://www.mdpi.com/2079-9292/9/9/1503

Turner, C. T., Rosenheim, J., Thakker, C., Chandran, A., Wilson, H., Venturini, C., Pollara, G., Chain, B. M., Tomlinson, G. S., & Noursadeghi, M. (2024). Single-cell transcriptome and T cell receptor profiling of the tuberculin skin test. *bioRxiv*, 2024.2006.2025.600676. https://doi.org/10.1101/2024.06.25.600676

Twilt, M. (2023). mHealth applications are they the future or another burden for our patients? *Mhealth*, *9*, 9. https://doi.org/10.21037/mhealth-23-5

van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillion-Robin, J.-C., Pieper, S., & Aerts, H. J. W. L. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, *77*(21), e104-e107. https://doi.org/10.1158/0008-5472.Can-17-0339

van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H., & Baessler, B. (2020). Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights into Imaging*, *11*(1), 91. https://doi.org/10.1186/s13244-020-00887-2

Vasefi, F., MacKinnon, N., & Farkas, D. L. (2016). Chapter 16 - Hyperspectral and Multispectral Imaging in Dermatology. In M. R. Hamblin, P. Avci, & G. K. Gupta (Eds.), *Imaging in Dermatology* (pp. 187-201). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-802838-4.00016-9

Vibhute, A. D., Kale, K., Dhumal, R. K., & Mehrotra, S. (2015). Hyperspectral imaging data atmospheric correction challenges and solutions using QUAC and FLAASH algorithms. In *2015 International Conference on Man and Machine Interfacing (MAMI)* (pp. 1-6). IEEE. https://doi.org/10.1109/MAMI.2015.7456604

Villaret, A., Ipinazar, C., Satar, T., Gravier, E., Mias, C., Questel, E., Schmitt, A. M., Samouillan, V., Nadal, F., & Josse, G. (2019). Raman characterization of human skin aging. *Skin Research and Technology*, *25*(3), 270-276. https://doi.org/10.1111/srt.12643

Vollmer, M., Möllmann, K.-P., & Shaw, J. (2015). The optics and physics of near infrared imaging. In *Education and Training in Optics and Photonics* (pp. TPE09). Optica Publishing Group.

Vukmanovic-Stejic, M., Reed, J. R., Lacy, K. E., Rustin, M. H. A., & Akbar, A. N. (2006). Mantoux Test as a model for a secondary immune response in humans. *Immunology Letters*, *107*(2), 93-101. https://doi.org/https://doi.org/10.1016/j.imlet.2006.08.002

Wasserman, L. (2004). *All of statistics: a concise course in statistical inference* (Vol. 26). Springer Science & Business Media.

Wawrzyk-Bochenek, I., Rahnama, M., Stachura, M., Wilczyński, S., & Wawrzyk, A. (2023). Evaluation of the reduction of skin hyperpigmentation changes under the influence of a preparation containing kojic acid using hyperspectral imaging—preliminary study. *Journal of Clinical Medicine*, *12*(7), 2710. https://doi.org/10.3390/jcm12072710

Weber, J. R., Cuccia, D. J., & Tromberg, B. J. (2006). Modulated imaging in layered media. *Conf Proc IEEE Eng Med Biol Soc*, *Suppl*, 6674-6676. https://doi.org/10.1109/iembs.2006.260918

WHO. (2022). *WHO consolidated guidelines on tuberculosis. Module 3: diagnosis. Tests for TB infection*. World Health Organization. https://www.who.int/publications/i/item/9789240056084

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*(1-3), 37-52.

Wu, J., Liu, A., Cui, J., Chen, A., Song, Q., & Xie, L. (2019). Radiomics-based classification of hepatocellular carcinoma and hepatic haemangioma on precontrast magnetic resonance images. *BMC Medical Imaging*, *19*(1), 23. https://doi.org/10.1186/s12880-019-0321-9

Wu, Y., Li, Z., Zhao, B., Song, Y., & Zhang, B. (2024). Transfer learning of spatial features from high-resolution RGB images for large-scale and robust hyperspectral remote sensing target detection. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1-32. https://doi.org/10.1109/TGRS.2024.3355184

Yang, Z., Albrow-Owen, T., Cai, W., & Hasan, T. (2021). Miniaturization of optical spectrometers. *Science*, *371*(6528), eabe0722. https://doi.org/doi:10.1126/science.abe0722

Yarosz, E. L., Ye, C., Kumar, A., Black, C., Choi, E. K., Seo, Y. A., & Chang, C. H. (2020). Cutting edge: Activation-induced iron flux controls CD4 T cell proliferation by promoting proper IL-2R signaling and mitochondrial function. *J Immunol*, *204*(7), 1708-1713. https://doi.org/10.4049/jimmunol.1901399

Yuan, Y., Zheng, X., & Lu, X. (2017). Hyperspectral image superresolution by transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), 1963-1974. https://doi.org/10.1109/JSTARS.2017.2655112

Zaidi, S. M., Coussens, A. K., Seddon, J. A., Kredo, T., Warner, D., Houben, R. M., & Esmail, H. (2023). Beyond latent and active tuberculosis: a scoping review of conceptual frameworks. *EClinicalMedicine*, *66*. https://doi.org/10.1016/j.eclinm.2023.102332

Zang, J., Cai, T., Zong, C., Yang, G., Zhang, Z., & Xue, C. (2024). Tuberculin skin test result detection method based on CSN-II and improved OTSU method. *Measurement*, *229*, 114409. https://doi.org/https://doi.org/10.1016/j.measurement.2024.114409

Zdrada-Nowak, J., Stolecka-Warzecha, A., Odrzywołek, W., Rusztowicz, M., Błońska-Fajfrowska, B., & Wilczyński, S. (2023). The assessment of moderate acne vulgaris face skin using blood perfusion and hyperspectral imaging–A pilot study. *Journal of Cosmetic Dermatology*, *22*(11), 3143-3151. https://doi.org/10.1111/jocd.15811

Zellweger, J. P., Sotgiu, G., Corradi, M., & Durando, P. (2020). The diagnosis of latent tuberculosis infection (LTBI): currently available tests, future developments, and perspectives to eliminate tuberculosis (TB). *Med Lav*, *111*(3), 170-183. https://doi.org/10.23749/mdl.v111i3.9983

Zheludev, V., Pölönen, I., Neittaanmäki-Perttu, N., Averbuch, A., Neittaanmäki, P., Grönroos, M., & Saari, H. (2015). Delineation of malignant skin tumors by hyperspectral imaging using diffusion maps dimensionality reduction. *Biomedical Signal Processing and Control*, *16*, 48-60. https://doi.org/https://doi.org/10.1016/j.bspc.2014.10.010

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells, W. M., 3rd, Jolesz, F. A., & Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*, *11*(2), 178-189. https://doi.org/10.1016/s1076-6332(03)00671-8

Zwanenburg, A., Leger, S., Vallières, M., & Löck, S. (2016). Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003*.

Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R. J., & Boellaard, R. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, *295*(2), 328-338. https://doi.org/10.1148/radiol.2020191145

APPENDIX A

Implementation of Principal Component Analysis (PCA) for Unsupervised Clustering of Spectral Signatures

Dataset Standardisation

The dataset of the TST hypercubes of participants, concatenated along the along the axis of wavelength bands, can be denoted as $\boldsymbol{D}$. Some wavelength bands in $\boldsymbol{D}$ may have relatively higher intensity range. This could lead to bias of the PCA model towards such wavelength bands. To avoid this bias, it is important to standardize the dataset $\boldsymbol{D}$ such that all wavelength bands across participants have the same intensity range while maintaining the intra-band relationships. The dataset $\boldsymbol{D}$ is standardised such that the skin spectral signature samples $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$ are mean-centred in the n-dimensional wavelength space around the origin (mean of 0 and standard deviation of 1), as shown in Figure A.1. The standardised wavelength band $\widetilde{\boldsymbol{B}}_{\boldsymbol{w}}$ for band w, across all training participants, is obtained from the measured wavelength band $\boldsymbol{B}_{\boldsymbol{w}}$, its mean μ_w and standard deviation σ_w, as shown in equation (A.1) (Kreyszig et al., 2008). The standardised dataset $\widetilde{\boldsymbol{D}}$ is as defined in equations (A.2) and (A.3).

$$\widetilde{\boldsymbol{B}}_{\boldsymbol{w}} = \frac{\boldsymbol{B}_{\boldsymbol{w}} - \mu_w}{\sigma_w} \tag{A.1}$$

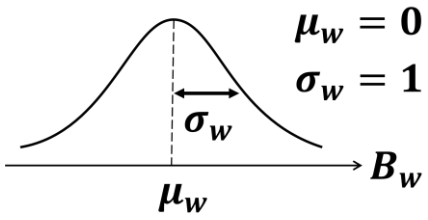


Figure A.1: Standardisation of wavelength band $\boldsymbol{B}_{\boldsymbol{w}}$

$$\widetilde{\boldsymbol{D}} = \begin{bmatrix} \widetilde{\boldsymbol{B}}_1 \\ \widetilde{\boldsymbol{B}}_2 \\ \vdots \\ \widetilde{\boldsymbol{B}}_{\boldsymbol{n}} \end{bmatrix} = \begin{bmatrix} \dfrac{\boldsymbol{B}_1 - \mu_1}{\sigma_1} \\ \dfrac{\boldsymbol{B}_2 - \mu_2}{\sigma_2} \\ \vdots \\ \dfrac{\boldsymbol{B}_{\boldsymbol{n}} - \mu_n}{\sigma_n} \end{bmatrix} = \begin{matrix} \textit{Observations (Samples)} \\ \begin{matrix} \vec{P}_1 & \vec{P}_2 & \cdots & \vec{P}_m \end{matrix} \\ \begin{bmatrix} P_{1B1} & P_{2B1} & \cdots & P_{mB1} \\ P_{1B2} & P_{2B2} & \cdots & P_{mB2} \\ \vdots & \vdots & \ddots & \vdots \\ P_{1Bn} & P_{2Bn} & \cdots & P_{mBn} \end{bmatrix} \end{matrix} \quad , \widetilde{\boldsymbol{D}} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{m}} \tag{A.2}$$

$$\widetilde{\boldsymbol{D}} = \left[\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}\right] \quad \text{(A.3)}$$

Projecting Spectral Signature Samples from Wavelength Band Space to the Principal Component Space

The spectral signature sample $\vec{\boldsymbol{P}}_{\boldsymbol{1}}$ can be described as a vector in the coordinate space of wavelength bands $\boldsymbol{B}_{\boldsymbol{v}}$ and $\boldsymbol{B}_{\boldsymbol{w}}$, such that the unit vectors of the coordinates are $\widehat{\boldsymbol{\beta}}_{\boldsymbol{v}}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}}$, and the scalar projections of the vector $\vec{\boldsymbol{P}}_{\boldsymbol{1}}$ to the axes are P_{1Bv} and P_{1Bw} respectively, as shown in Figure A.2. Similarly, all the spectral signature samples in the standardised training dataset $\widetilde{\boldsymbol{D}}$ can be represented as vectors $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \ldots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$ in the arbitrary ($\boldsymbol{B}_{\boldsymbol{v}}$, $\boldsymbol{B}_{\boldsymbol{w}}$) coordinate space, as shown in Figure A.3. It is assumed that there is a natural relationship between the samples such that the samples can be grouped into some groups of natural clusters. It is expected that the direction in the ($\boldsymbol{B}_{\boldsymbol{v}}$, $\boldsymbol{B}_{\boldsymbol{w}}$) coordinate space along which there is maximum variance between the samples will be the best direction along which to observe natural separations between the samples, as illustrated in a fictitious example in Figure A.2. This direction of maximum variance is defined here, and in (Sârbu & Pop, 2005), as the first principal component $\boldsymbol{P}_{\boldsymbol{c1}}$. The principal component $\boldsymbol{P}_{\boldsymbol{c1}}$ is described here as a new coordinate which has the unit vector $\widehat{\boldsymbol{p}}_{\boldsymbol{c1}}$. The unit vector $\widehat{\boldsymbol{p}}_{\boldsymbol{c1}}$ can be defined by the scalar projections L_{vBv} and L_{wBw} from the $\boldsymbol{B}_{\boldsymbol{v}}$ and $\boldsymbol{B}_{\boldsymbol{w}}$ coordinates respectively, as shown in Figure A.4. The scalar projections L_{vBv} and L_{wBw} are termed the "loadings" (or loading vector elements) of the principal component $\boldsymbol{P}_{\boldsymbol{c1}}$.

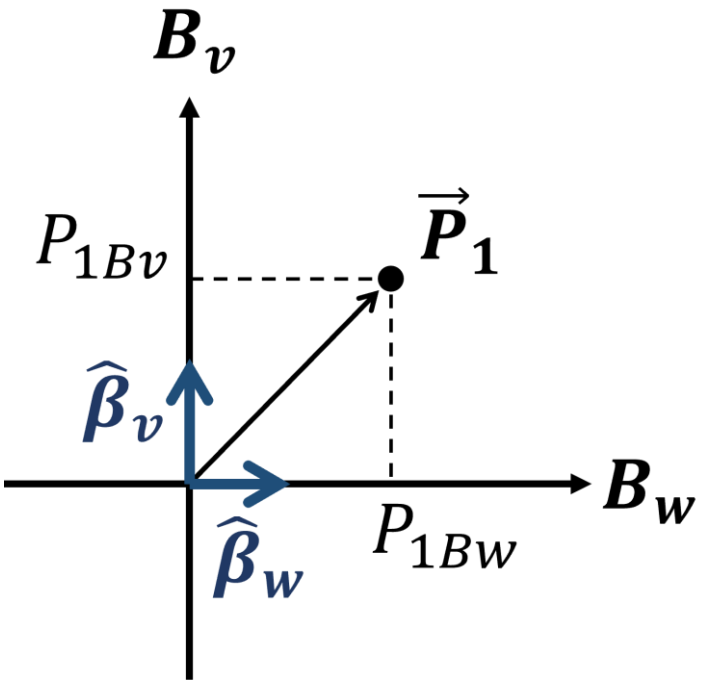


Figure A.2: Description of spectral signature $\vec{\boldsymbol{P}}_{\boldsymbol{1}}$ as a vector in the ($\boldsymbol{B}_{\boldsymbol{v}}$, $\boldsymbol{B}_{\boldsymbol{w}}$) coordinate space

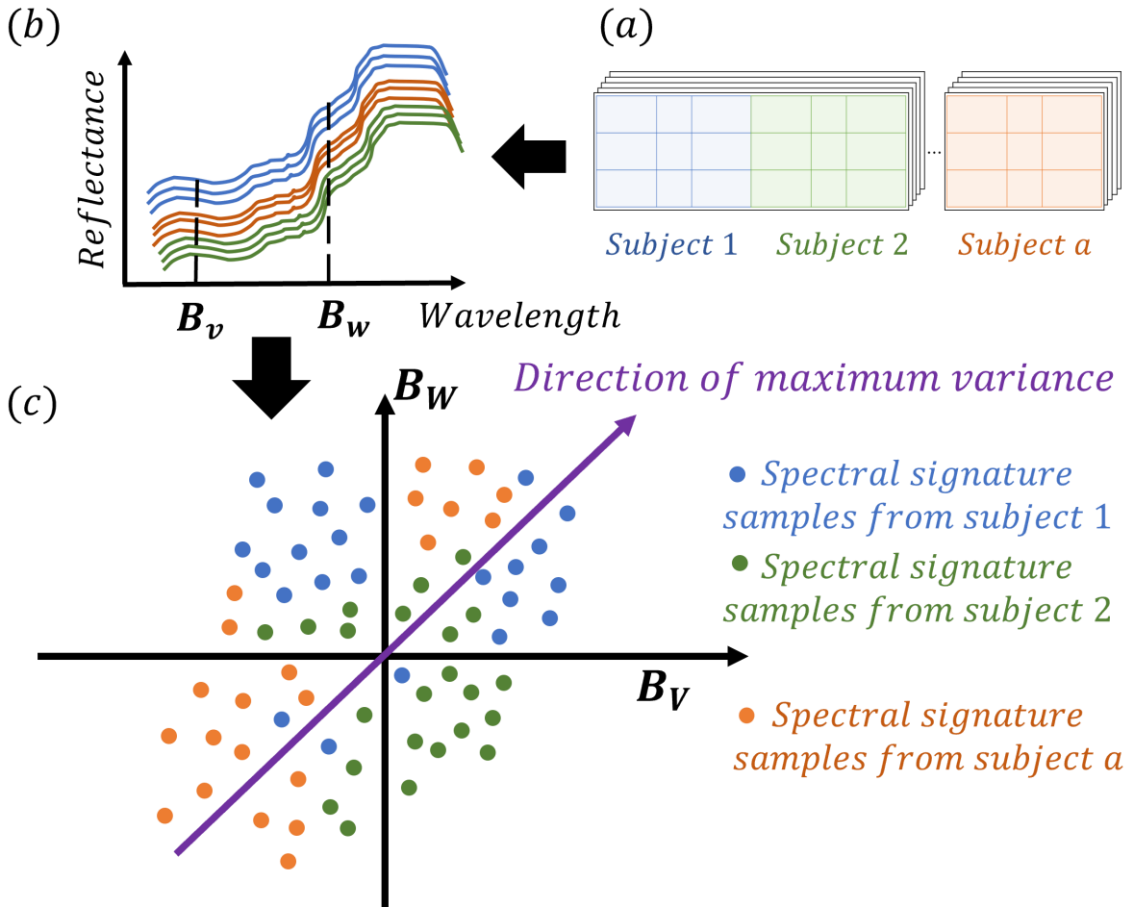


Figure A.3: Illustration showing the direction of maximum variance between samples in the ($\boldsymbol{B_v}$, $\boldsymbol{B_w}$) coordinate space showing representations of (a) hypercube (b) spectral signatures (c) wavelength band space

The perpendicular distances $d_1, d_2, \dots, d_m$ (also termed perpendicular residuals or offsets) between the samples $\vec{\boldsymbol{P}}_1, \vec{\boldsymbol{P}}_2, \dots, \vec{\boldsymbol{P}}_m$ and any line through the origin of the ($\boldsymbol{B_v}$, $\boldsymbol{B_w}$) coordinate space can be computed. The sum of squared perpendicular distances $SSD = d_1^2 + d_2^2 + \cdots + d_m^2$ are computed for each considered line as a measure of goodness-of-fit. The principal component $\boldsymbol{P_{c1}}$ is computed here as the best fitting line through the origin of the ($\boldsymbol{B_v}$, $\boldsymbol{B_w}$) coordinate space that minimizes the sum of squared perpendicular distances, as shown in equations (A.3) and (A.4). The scalar projections of the spectral signature samples $\vec{\boldsymbol{P}}_1, \vec{\boldsymbol{P}}_2, \dots, \vec{\boldsymbol{P}}_m$ to the principal component $\boldsymbol{P_{c1}}$ are denoted as $T_{1pc1}, T_{2pc1}, \dots, T_{mpc1}$. It is expected that these scalar projections will be clustered in a way that describes the most significant variation and relationship between the samples. The scalar projections $T_{1pc1}, T_{2pc1}, \dots, T_{mpc1}$ are reshaped from 1D to 2D such that the clustering of the samples can be more easily visualised, as shown in figure A.5. It is anticipated that these natural clusters will be useful in delineating spectral signature samples of induration from samples that are of only erythema and samples of normal skin tissue.

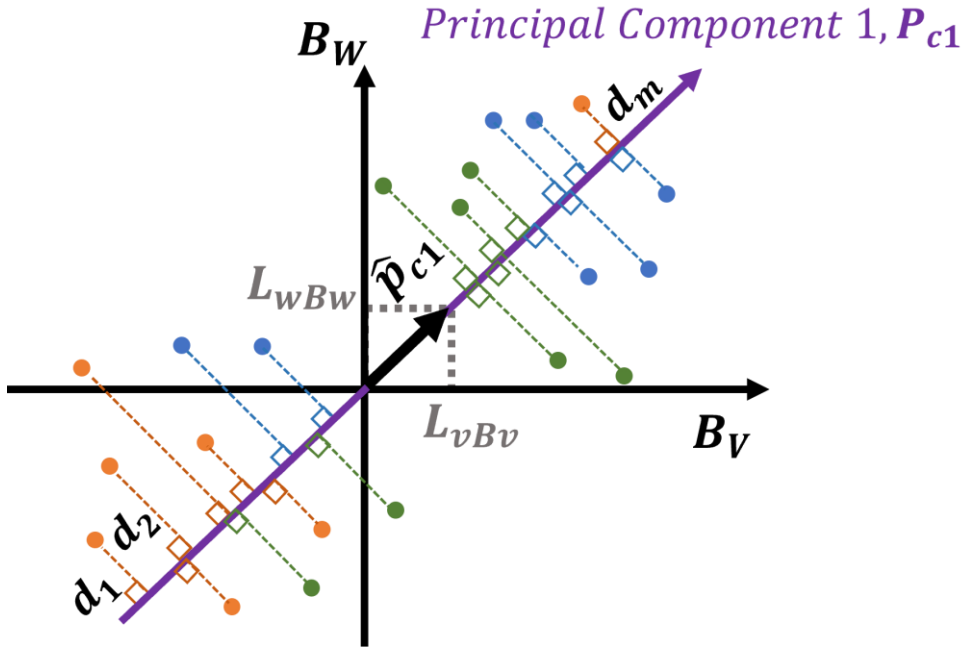


Figure A.4: Estimation of principal component 1 in the (B_v, B_w) coordinate space

$$\hat{p}_{c1} = [L_{vBv}, L_{wBw}] \tag{A.4}$$

$$\underset{\hat{p}_{c1}}{argmin}\, SSD = \underset{\hat{p}_{c1}}{argmin} \sum_{i=1}^{m} {d_i}^2 \tag{A.5}$$

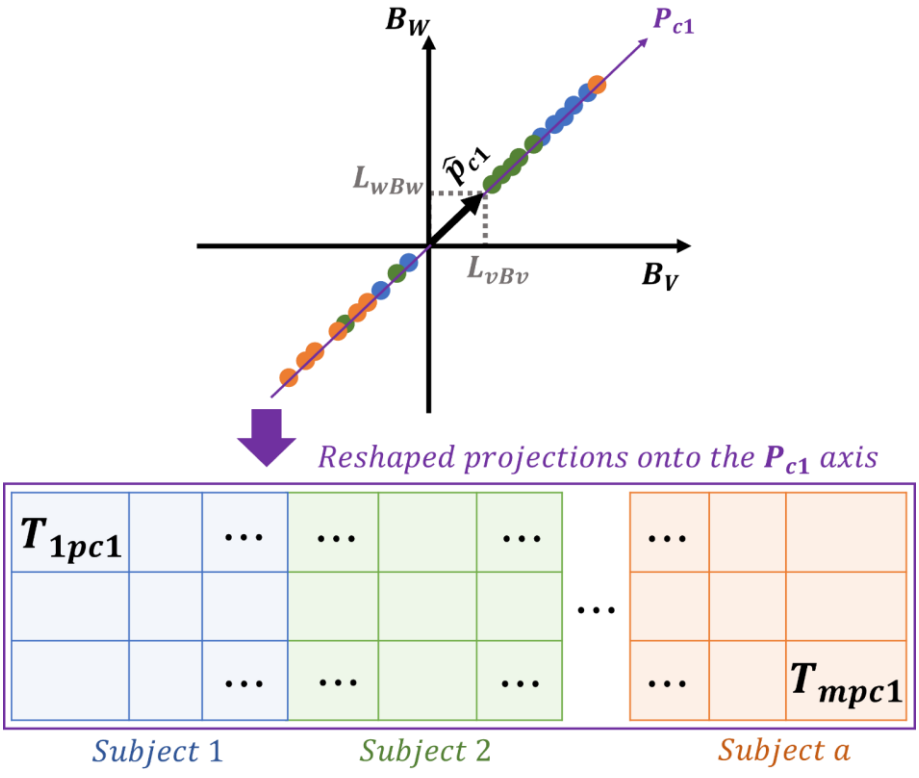


Figure A.5: Projection of samples to the principal component 1 in the (B_v, B_w) coordinate space

Though the projections $T_{1pc1}, T_{2pc1}, \dots, T_{mpc1}$ may explain the most significant variation and relationship between the samples of all the training participants, this variation may not precisely delineate induration from erythema and normal skin tissue. The underlying biological and physiological relationship between the skin tissues of the participants is not only an induration-to-erythema-to-normal skin relationship. The relationship also includes skin tone and skin texture variation. Thus, the most significant variation captured by $T_{1pc1}, T_{2pc1}, \dots, T_{mpc1}$ may be variations due to other skin properties. Computing a second principal component $\boldsymbol{P_{c2}}$ which is the second-best line of fit that minimizes the SSD and is orthogonal to $\boldsymbol{P_{c1}}$, would enable a projection onto a coordinate that clusters the samples in a way that explains the second most significant variation and relationship between the samples. It is possible that the scalar projections $T_{1pc2}, T_{2pc2}, \dots, T_{mpc2}$ on $\boldsymbol{P_{c2}}$ may better explain the induration-to-erythema-to-normal skin relationship than the projections on $\boldsymbol{P_{c1}}$. As shown in figures A.6 and A.7, the projections $T_{1pc2}, T_{2pc2}, \dots, T_{mpc2}$ can be reshaped for easier visualisation of the relationship between the samples that is explained by $\boldsymbol{P_{c2}}$. Thus, using PCA, the spectral signature samples can be projected from the ($\boldsymbol{B_v}$, $\boldsymbol{B_w}$) coordinate space to a 2D principal component space, as shown in Figure A.8, which clusters the samples in a way that the most significant and the second most significant variation between the samples can be more easily explained.

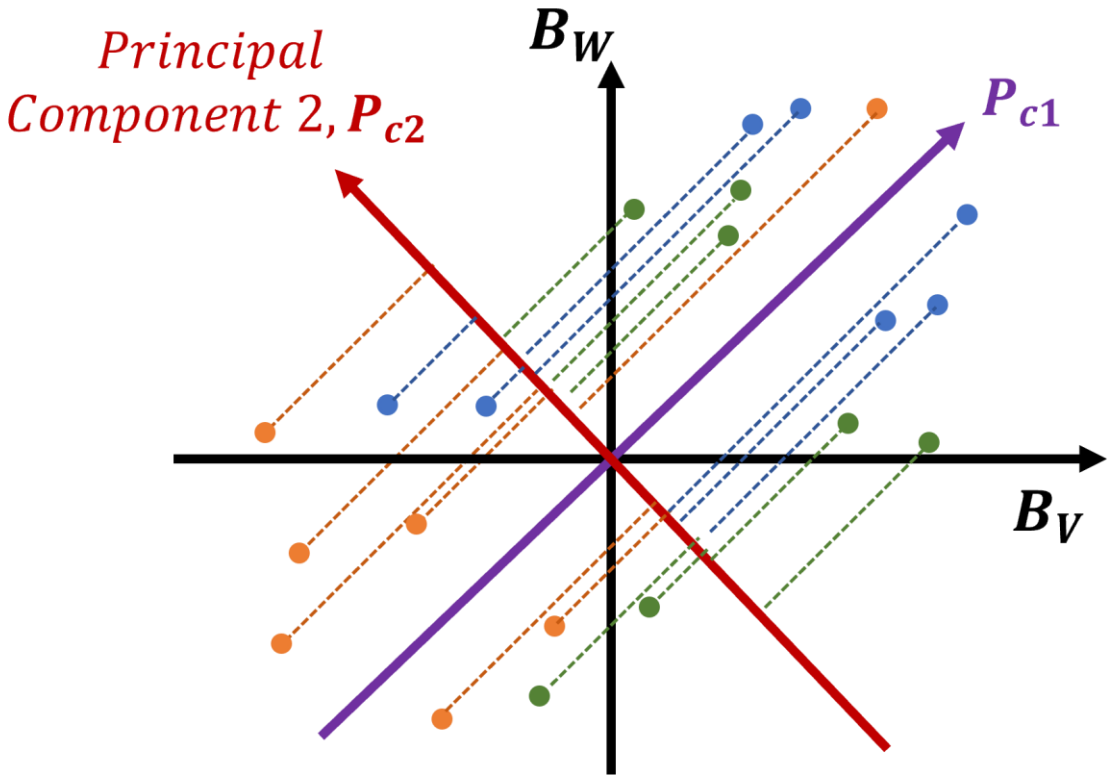


Figure A.6: Estimation of principal component 2 in the ($\boldsymbol{B_v}$, $\boldsymbol{B_w}$) coordinate space

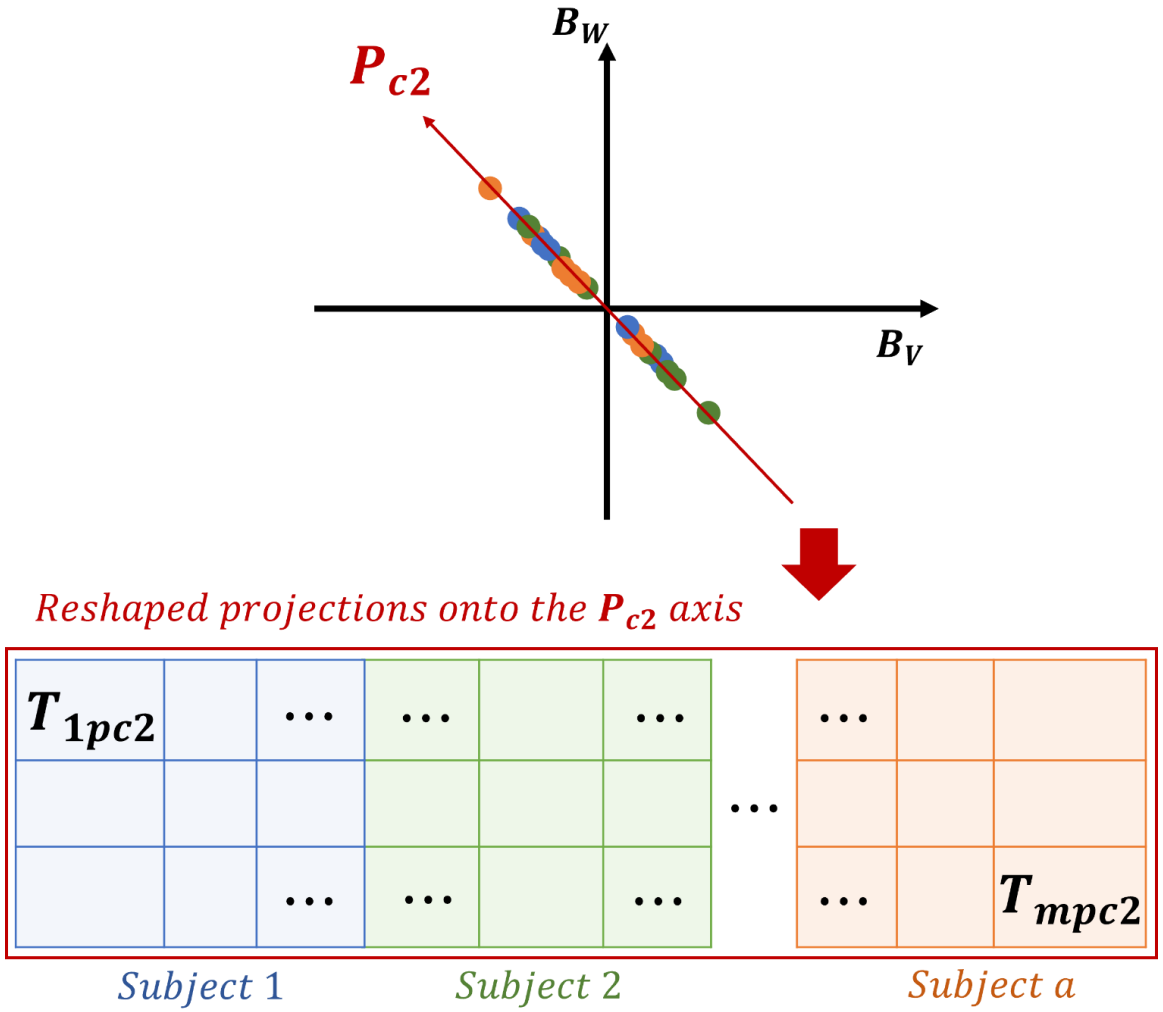


Figure A.7: Projection of samples to the principal component 2 in the (B_v, B_w) coordinate space

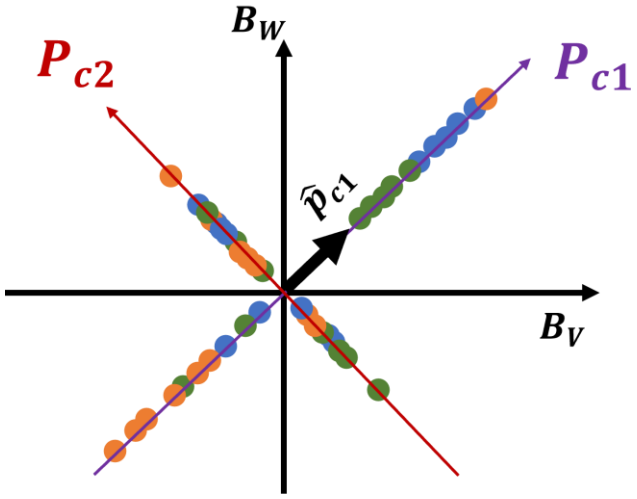


Figure A.8: Transformation of samples from the (B_v, B_w) coordinate space to the PCA space

n-Dimensional Computation of Principal Components

Using the full wavelength band range $B_1, B_2, \dots, B_n$ of the standardised training dataset $\widetilde{D}$, the spectral signature samples $P_1, P_2, \dots, P_m$ are represented as vectors $\vec{P}_1, \vec{P}_2, \dots, \vec{P}_m$ in an n-dimensional wavelength band coordinate space (B_1, $B_2, \dots, B_n$). In the n-dimensional wavelength

band space the vector $\vec{\boldsymbol{P}}_{\boldsymbol{1}}$ is defined in terms of its scalar projections P_{1B1}, $P_{1B2}, \dots, P_{1Bn}$ in the wavelength bands $\boldsymbol{B_1}$, $\boldsymbol{B_2}, \dots, \boldsymbol{B_n}$ and the unit vectors $\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}}$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}}, \dots, \widehat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ along the corresponding wavelength band axes. All the training spectral signature sample vectors $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$ can be defined similarly, as shown in equations (A.6) to (A.9).

$$\vec{\boldsymbol{P}}_{\boldsymbol{1}} = \vec{\boldsymbol{P}}_{\boldsymbol{1B1}} + \vec{\boldsymbol{P}}_{\boldsymbol{1B2}} + \cdots + \vec{\boldsymbol{P}}_{\boldsymbol{1Bn}} \tag{A.6}$$

$$\vec{\boldsymbol{P}}_{\boldsymbol{1}} = P_{1B1}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}} + P_{1B2}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}} + \cdots + P_{1Bn}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{n}} \tag{A.7}$$

$$\vec{\boldsymbol{P}}_{\boldsymbol{2}} = P_{2B1}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}} + P_{2B2}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}} + \cdots + P_{2Bn}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{n}} \tag{A.8}$$

$$\vdots$$

$$\vec{\boldsymbol{P}}_{\boldsymbol{m}} = \boldsymbol{P_{mB1}}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{1}} + \boldsymbol{P_{mB2}}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{2}} + \cdots + \boldsymbol{P_{mBn}}\,\widehat{\boldsymbol{\beta}}_{\boldsymbol{n}} \tag{A.9}$$

The computing hardware used in the implementation of all the models described in this chapter includes a 12th Gen Intel Core i9-12900 central processing unit (CPU) with 2.40GHz base speed, and a random-access memory (RAM) of 32 GB. The models were implemented using the python programming language and the Jupyter Notebook computing platform. Computing for n principal components by fitting n lines that minimise the SSD in n orthogonal directions in the $\boldsymbol{B_1}, \boldsymbol{B_2}, \dots, \boldsymbol{B_n}$ space is computationally intensive. A more computationally efficient way of generating n principal components in the $\boldsymbol{B_1}, \boldsymbol{B_2}, \dots, \boldsymbol{B_n}$ space is by computing the eigen decomposition of the covariance matrix of the training dataset $\widetilde{\boldsymbol{D}}$, based on the description in (Draper et al., 2003). A covariance matrix $\boldsymbol{C}$, which has the property of symmetry about its leading diagonal, is generated from the standardized training dataset $\widetilde{\boldsymbol{D}}$. It defines the covariance ${\sigma^2}_{BvBw}$ between every pair of wavelength band $\boldsymbol{B_v}$ and $\boldsymbol{B_w}$ for all samples, and the variance ${\sigma^2}_{BwBw}$ of all samples along the wavelength band axis $\boldsymbol{B_w}$, as shown in equations (A.20) to (A.12) and as described in (Wasserman, 2004). The covariances describe the level of correlation or statistical dependence that exists between every pair of wavelength band for all samples, and the variances describe the spread of the samples along each wavelength band axis. Wavelength bands with high variances skew the most significant principal components in their direction.

$$\boldsymbol{C} = \frac{1}{m-1}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{D}}^{T} = \frac{1}{m-1}\begin{bmatrix}\widetilde{\boldsymbol{B}}_{\boldsymbol{1}} \\ \widetilde{\boldsymbol{B}}_{\boldsymbol{2}} \\ \vdots \\ \widetilde{\boldsymbol{B}}_{\boldsymbol{n}}\end{bmatrix}\begin{bmatrix}\widetilde{\boldsymbol{B}}_{\boldsymbol{1}}{}^{T} & \widetilde{\boldsymbol{B}}_{\boldsymbol{2}}{}^{T} & \dots & \widetilde{\boldsymbol{B}}_{\boldsymbol{n}}{}^{T}\end{bmatrix} \tag{A.20}$$

$$\boldsymbol{C} = \begin{bmatrix} {\sigma^2}_{B1B1} & {\sigma^2}_{B1B2} & \cdots & {\sigma^2}_{B1Bn} \\ {\sigma^2}_{B2B1} & {\sigma^2}_{B2B2} & \cdots & {\sigma^2}_{B2Bn} \\ \vdots & \vdots & \ddots & \vdots \\ {\sigma^2}_{BnB1} & {\sigma^2}_{BnB2} & \cdots & {\sigma^2}_{BnBn} \end{bmatrix}, \boldsymbol{C} \in \mathbb{R}^{n \times n} \tag{A.11}$$

$$Where, {\sigma^2}_{B1B2} = \frac{1}{m-1} \widetilde{\boldsymbol{B}}_{\boldsymbol{1}} {\widetilde{\boldsymbol{B}}_{\boldsymbol{2}}}^T \; and \; {\sigma^2}_{B1B1} = \frac{1}{m-1} \widetilde{\boldsymbol{B}}_{\boldsymbol{1}} {\widetilde{\boldsymbol{B}}_{\boldsymbol{1}}}^T \tag{A.12}$$

If there exists an eigenvector $\vec{\boldsymbol{u}}_{\boldsymbol{w}}$ such that an operation on it using the covariance matrix $\boldsymbol{C}$ produces a multiple of itself scaled by τ_w, as shown in equation (A.13), then the matrix $\boldsymbol{C}$ can be decomposed into a set $\boldsymbol{U}$ of eigenvectors $\vec{\boldsymbol{u}}_{\boldsymbol{1}}, \vec{\boldsymbol{u}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{u}}_{\boldsymbol{n}}$ and a diagonal matrix $\boldsymbol{\Lambda}$ of corresponding eigenvalues $\tau_1, \tau_2, \dots, \tau_n$. The eigenvectors $\vec{\boldsymbol{u}}_{\boldsymbol{1}}, \vec{\boldsymbol{u}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{u}}_{\boldsymbol{n}}$ are equivalent to the unit vectors $\widehat{\boldsymbol{p}}_{\boldsymbol{c1}}, \widehat{\boldsymbol{p}}_{\boldsymbol{c2}}, \dots, \widehat{\boldsymbol{p}}_{\boldsymbol{cn}}$ of the principal components axes $\boldsymbol{P}_{\boldsymbol{c1}}, \boldsymbol{P}_{\boldsymbol{c2}}, \dots, \boldsymbol{P}_{\boldsymbol{cn}}$, as shown in equations (A.14) to (A.17) and described in (Dunn Iii et al., 1989). The eigenvalues $\tau_1, \tau_2, \dots, \tau_n$ are scalars that quantify the amount of variation in decreasing order in $\widetilde{\boldsymbol{D}}$ that can be explained by the principal component axes $\boldsymbol{P}_{\boldsymbol{c1}}, \boldsymbol{P}_{\boldsymbol{c2}}, \dots, \boldsymbol{P}_{\boldsymbol{cn}}$. It is expected that $\boldsymbol{P}_{\boldsymbol{c1}}$ captures the most significant and unique variation in the training dataset, while $\boldsymbol{P}_{\boldsymbol{cn}}$ captures the least significant variation in the dataset. It can be assumed that all the most significant and unique variations in the dataset are captured in the first r principal components such that $r < n$. This facilitates dimensionality reduction such that the n-dimensional dataset in the wavelength band space can be transformed to a smaller r-dimensional form in the PCA space where the most significant variations and relationships between the spectral signature samples are preserved, and less significant variations are discarded as redundancies. The array of the most significant principal component unit vectors in decreasing order of significance is defined as $\boldsymbol{U}_{\boldsymbol{R}} = [\widehat{\boldsymbol{p}}_{\boldsymbol{c1}}, \widehat{\boldsymbol{p}}_{\boldsymbol{c2}}, \dots, \widehat{\boldsymbol{p}}_{\boldsymbol{cr}}]$, as shown in equation (A.19). The principal component unit vector $\widehat{\boldsymbol{p}}_{\boldsymbol{c1}}$ can be defined in terms of loadings $L_{1B1}, L_{1B2}, \dots, L_{1Bn}$ which are scalar projections from the wavelength bands $\boldsymbol{B}_{\boldsymbol{1}}, \boldsymbol{B}_{\boldsymbol{2}}, \dots, \boldsymbol{B}_{\boldsymbol{n}}$. All the principal component unit vectors are then defined in this form, as shown in equations (A.20) to (A.22).

$$\boldsymbol{C}\vec{\boldsymbol{u}}_{\boldsymbol{w}} = \tau_w \vec{\boldsymbol{u}}_{\boldsymbol{w}} \tag{A.13}$$

$$\boldsymbol{C} \cdot [\boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \dots \quad \boldsymbol{u_n}] = [\boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \dots \quad \boldsymbol{u_n}] \cdot \begin{bmatrix} \tau_1 & 0 & \cdots & 0 \\ 0 & \tau_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tau_n \end{bmatrix} \tag{A.14}$$

$$\boldsymbol{CU} = \boldsymbol{U\Lambda} \tag{A.15}$$

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{T}, where\ \boldsymbol{C}, \boldsymbol{U} \epsilon \mathbb{R}^{n \times n} \tag{A.16}$$

$$\boldsymbol{U} = [\boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \cdots \quad \boldsymbol{u_n}] = [\boldsymbol{\hat{p}_{c1}} \quad \boldsymbol{\hat{p}_{c2}} \quad \cdots \quad \boldsymbol{\hat{p}_{cn}}] \tag{A.17}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \tau_1 & 0 & \cdots & 0 \\ 0 & \tau_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tau_n \end{bmatrix} \tag{A.18}$$

$$\boldsymbol{U_R} = [\boldsymbol{u_1} \quad \boldsymbol{u_2} \quad \cdots \quad \boldsymbol{u_r}] = [\boldsymbol{\hat{p}_{c1}} \quad \boldsymbol{\hat{p}_{c2}} \quad \cdots \quad \boldsymbol{\hat{p}_{cr}}], where\ r < n \tag{A.19}$$

$$\boldsymbol{\hat{p}_{c1}} = [L_{1B1} \quad L_{1B2} \quad \cdots \quad L_{1Bn}] \tag{A.20}$$

$$\boldsymbol{\hat{p}_{c2}} = [L_{2B1} \quad L_{2B2} \quad \cdots \quad L_{2Bn}] \tag{A.21}$$

$$\vdots$$

$$\boldsymbol{\hat{p}_{cr}} = [L_{rB1} \quad L_{rB2} \quad \cdots \quad L_{rBn}] \tag{A.22}$$

Projecting Spectral Signatures Samples from an n −Dimensional Wavelength Band Space to an r −Dimensional Principal Component Space

The spectral signature samples $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$ for all the training participants in the dataset $\widetilde{\boldsymbol{D}}$ are projected to the principal component axes $\boldsymbol{P_{c1}}, \boldsymbol{P_{c2}}, \dots, \boldsymbol{P_{cn}}$ by taking the dot product of the transpose of the dimension-reduced array of principal component unit vectors $\boldsymbol{U_R}$ and $\widetilde{\boldsymbol{D}}$, as shown in equations (A.23) to (A.25) and described in (Wold et al., 1987). This projection produces $\vec{\boldsymbol{T}}_{\boldsymbol{1}}, \vec{\boldsymbol{T}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{T}}_{\boldsymbol{m}}$ which is the transformed versions of spectral signature samples $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$ in the principal component space, as shown in equation (A.26). The transformed sample $\vec{\boldsymbol{T}}_{\boldsymbol{1}}$ is composed of scalar intensities $T_{1pc1}, T_{1pc2}, \dots, T_{1pcr}$ in the principal component space which is analogous to the scalar intensities $P_{1B1}, P_{1B2}, \dots, P_{1Bn}$ in the wavelength band space. The scalar intensities in the principal component space can be expressed as algebraic sums of the scalar intensities in the wavelength band space weighted by the loadings of the principal component axes, as shown in equations (A.27) to (A.28). The size of $\vec{\boldsymbol{T}}_{\boldsymbol{1}}$ is r while the size of the analogous $\vec{\boldsymbol{P}}_{\boldsymbol{1}}$ is n, where $r < n$. Thus, $\vec{\boldsymbol{T}}_{\boldsymbol{1}}, \vec{\boldsymbol{T}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{T}}_{\boldsymbol{m}}$ are dimension-reduced versions of the spectral signature samples $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$ that capture the most significant variations and relationships between $\vec{\boldsymbol{P}}_{\boldsymbol{1}}, \vec{\boldsymbol{P}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{P}}_{\boldsymbol{m}}$. The transformed dataset $\boldsymbol{T}$ can be reshaped from a $(r \times m)$ array to a $(k \times ak \times r)$, where a is the number of training participants ($a = 56$ in this

chapter) and $k = 200$ pixels for the hypercube spatial axes in this chapter. This reshaping enables the 2D visualisations of the natural clusters for each participant along the principal component axes $\boldsymbol{P_{c1}, P_{c2}, \dots, P_{cn}}$, where each adjacent $(k \times k)$ block is the projection for each participant, as shown in Figure A.9.

$$\boldsymbol{T} = \boldsymbol{U_R}^T \tilde{\boldsymbol{D}}, where\ \boldsymbol{U_R} \epsilon \mathbb{R}^{r \times n}, \tilde{\boldsymbol{D}} \epsilon \mathbb{R}^{n \times m}, \boldsymbol{T} \epsilon \mathbb{R}^{r \times m} \tag{A.23}$$

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{P_{c1}} \\ \boldsymbol{P_{c2}} \\ \vdots \\ \boldsymbol{P_{cr}} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{p}}_{\boldsymbol{c1}} \\ \hat{\boldsymbol{p}}_{\boldsymbol{c2}} \\ \vdots \\ \hat{\boldsymbol{p}}_{\boldsymbol{cr}} \end{bmatrix} \begin{bmatrix} P_{1B1} & P_{2B1} & \cdots & P_{mB1} \\ P_{1B2} & P_{2B2} & \cdots & P_{mB2} \\ \vdots & \vdots & \ddots & \vdots \\ P_{1Bn} & P_{2Bn} & \cdots & P_{mBn} \end{bmatrix} \tag{A.24}$$

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{P_{c1}} \\ \boldsymbol{P_{c2}} \\ \vdots \\ \boldsymbol{P_{cr}} \end{bmatrix} = \begin{bmatrix} L_1{}^T & L_2{}^T & \cdots & L_r{}^T \\ L_{1B1} & L_{1B2} & \cdots & L_{1Bn} \\ L_{2B1} & L_{2B2} & \cdots & L_{2Bn} \\ \vdots & \vdots & \ddots & \vdots \\ L_{rB1} & L_{rB2} & \cdots & L_{rBn} \end{bmatrix} \begin{bmatrix} \vec{P}_1 & \vec{P}_2 & \cdots & \vec{P}_m \\ P_{1B1} & P_{2B1} & \cdots & P_{mB1} \\ P_{1B2} & P_{2B2} & \cdots & P_{mB2} \\ \vdots & \vdots & \ddots & \vdots \\ P_{1Bn} & P_{2Bn} & \cdots & P_{mBn} \end{bmatrix} \tag{A.25}$$

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{P_{c1}} \\ \boldsymbol{P_{c2}} \\ \vdots \\ \boldsymbol{P_{cr}} \end{bmatrix} = \begin{bmatrix} \vec{T}_1 & \vec{T}_2 & \cdots & \vec{T}_m \\ T_{1Pc1} & T_{2Pc1} & \cdots & T_{mPc1} \\ T_{1Pc2} & T_{2Pc2} & \cdots & T_{mPc2} \\ \vdots & \vdots & \ddots & \vdots \\ T_{1Pcr} & T_{2Pcr} & \cdots & T_{mPcr} \end{bmatrix} \tag{A.26}$$

$$T_{1Pc1} = L_{1B1}P_{1B1} + L_{1B2}P_{1B2} + \cdots + L_{1Bn}P_{1Bn} \tag{A.27}$$

$$\vdots$$

$$T_{mPcr} = L_{rB1}P_{mB1} + L_{rB2}P_{mB2} + \cdots + L_{rBn}P_{mBn} \tag{A.28}$$

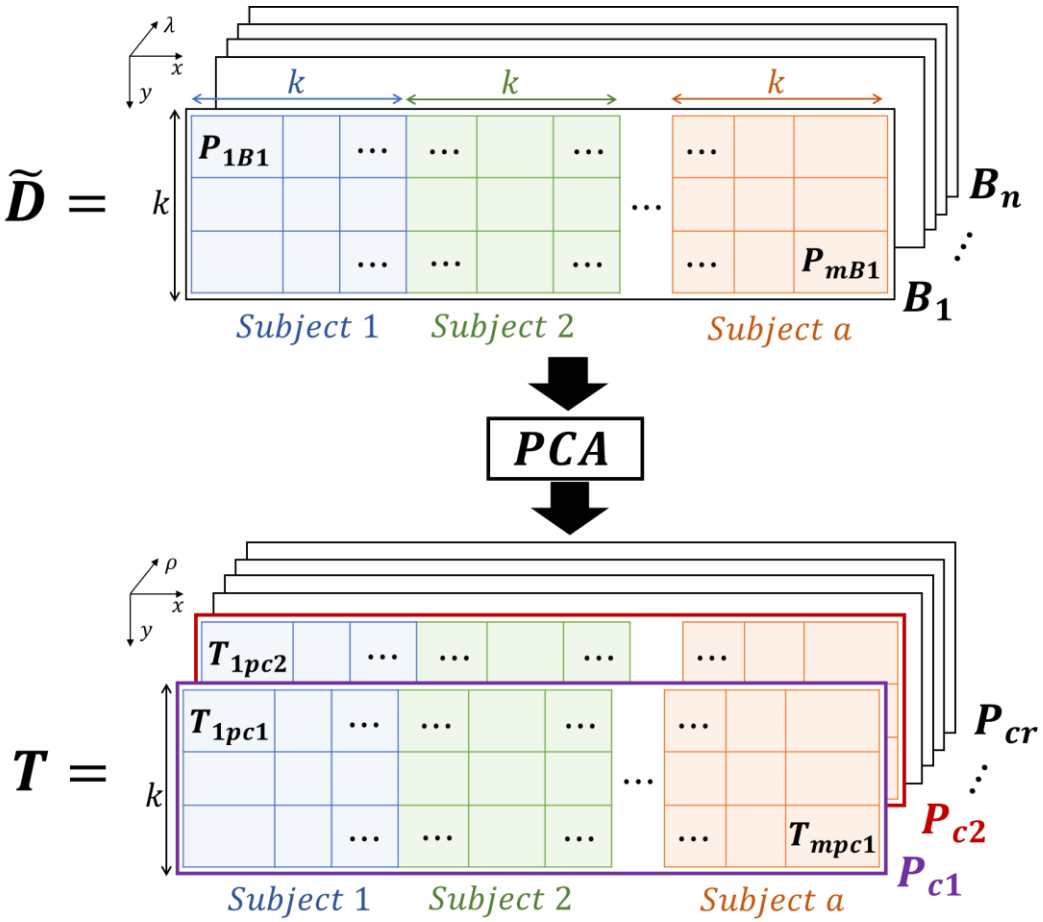


Figure A.9: Transformation of training dataset from wavelength-based hypercube to an array of PCA-based projections

PCA-Based Erythema Segmentation

Based on the hypothesis that indurations are always subsumed within erythema, it is desired to segment erythema samples from normal skin samples. As seen in (Hossain et al., 2022), erythema, which is characterized by oxyhaemoglobin, has two absorption peaks within the 500nm and 600nm wavelength window of visible light. This coincides with the green band of visible light. It can be expected that some of the natural clusters produced in $\boldsymbol{T}$ will be erythema related. Since the loadings $\boldsymbol{L} = \boldsymbol{L_1}, \boldsymbol{L_2}, \dots, \boldsymbol{L_n}$ of the principal component axes $\boldsymbol{P_{c1}}, \boldsymbol{P_{c2}}, \dots, \boldsymbol{P_{cn}}$ respectively, are projections from the wavelength band space, it is possible to validate the presence of erythema clusters in $\boldsymbol{T}$ by correlating $\boldsymbol{L}$ with the oxyhaemoglobin wavelength characteristic defined in (Hossain et al., 2022). To increase the likelihood of accurate segmentation of erythema from normal skin samples, the training dataset $\widetilde{\boldsymbol{D}}$ is cropped along the wavelength band axis from $\boldsymbol{\lambda} = 500nm$ to $\boldsymbol{\lambda} = 600nm$ to form the cropped hypercube $\widetilde{\boldsymbol{D}}_{\boldsymbol{500nm,600nm}}$, as shown in figure A.10. It is expected that applying PCA to $\widetilde{\boldsymbol{D}}_{\boldsymbol{500nm,600nm}}$ to generate $\boldsymbol{T}_{\boldsymbol{500nm,600nm}}$ will facilitate an accurate clustering and segmentation of erythema samples from normal skin samples.

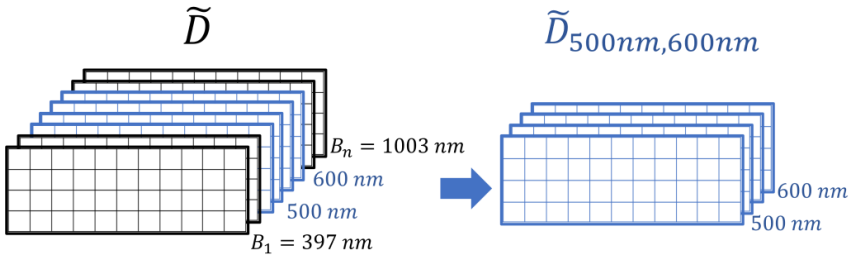


Figure A.10: Wavelength cropping of HSI dataset to erythema wavelength window

Based on the assumption that the erythema region in the TST skin site captured for each participant is surrounded by normal skin tissue, a bivariate Gaussian kernel-based vignette $\boldsymbol{f_v}$, defined in equation (4.1), and also described in (Hamedani & Tata, 1975), was applied to adjacent $(k \times k)$ blocks in each principal component axes of $\boldsymbol{T_{500nm,600nm}}$. Each adjacent $(k \times k)$ block in $\boldsymbol{T_{500nm,600nm}}$ represented a 2D shaped projection in the PCA space for a participant. Passing the output of the Gaussian function through a $(\gamma \times \gamma)$ median filter $\boldsymbol{f_{med}}$, to reduce noise, and an intensity thresholding function $\boldsymbol{f_{th}}$ enabled the generation of an array of erythema masks $\boldsymbol{M}$, as shown in figure A.11. The mask facilitated the segmentation of erythema samples from normal skin samples.

$$f_v(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu_x)^2+(y-\mu_y)^2}{2\sigma^2}} \tag{A.29}$$

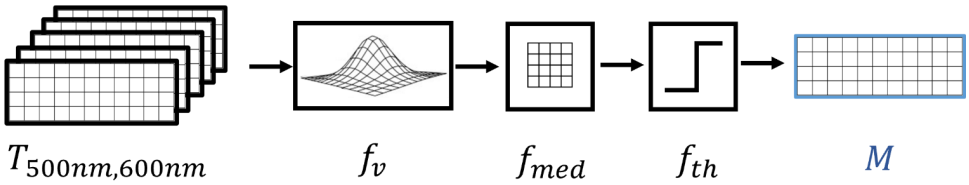


Figure A.11: Pathway to the generation of erythema mask array

PCA-Based Induration Segmentation from Erythema Region

With the segmentation of the erythema samples from the normal skin samples, the next goal is to segment the induration samples presumably subsumed amongst the erythema samples. Though wavelength cropping of the training dataset $\widetilde{\boldsymbol{D}}$ in the $500nm - 600nm$ wavelength window to generate $\widetilde{\boldsymbol{D}}_{\boldsymbol{500nm,600nm}}$ facilitates accurate clustering of erythema samples using PCA, a similar window with significant wavelength characteristics unique to induration samples is not currently known in literature. Thus, full wavelength range of the training dataset $\widetilde{\boldsymbol{D}}$ will be required for induration segmentation. Due to the perceived subtility of induration samples, it is expected that the induration samples will be clustered along with erythema samples in $\boldsymbol{T}$, the PCA projection of $\widetilde{\boldsymbol{D}}$. The presence of the normal skin samples in $\boldsymbol{T}$ will also reduce the likelihood of induration samples having a unique cluster in the principal component space. The normal skin samples can be removed from the training dataset $\widetilde{\boldsymbol{D}}$ by applying the erythema mask $\boldsymbol{M}$ to $\widetilde{\boldsymbol{D}}$ and performing Hadamard (elementwise)

product to generate $\breve{\boldsymbol{D}}_{\boldsymbol{e}}$ – the erythema-masked training dataset. It is expected that one of the most significant variations between the samples in $\breve{\boldsymbol{D}}_{\boldsymbol{e}}$ will be due to the presence of two unique clusters – the "erythema only" and "induration" clusters. Thus, projecting $\breve{\boldsymbol{D}}_{\boldsymbol{e}}$ to the principal component space should transform the samples to generate the PCA-transformed dataset $\breve{\boldsymbol{T}}$ such that the samples form the "erythema only" and "induration" clusters, and there is maximal delineation of the subsumed induration from surrounding erythema. The complete architecture of the described induration model is as shown in figure A.12.

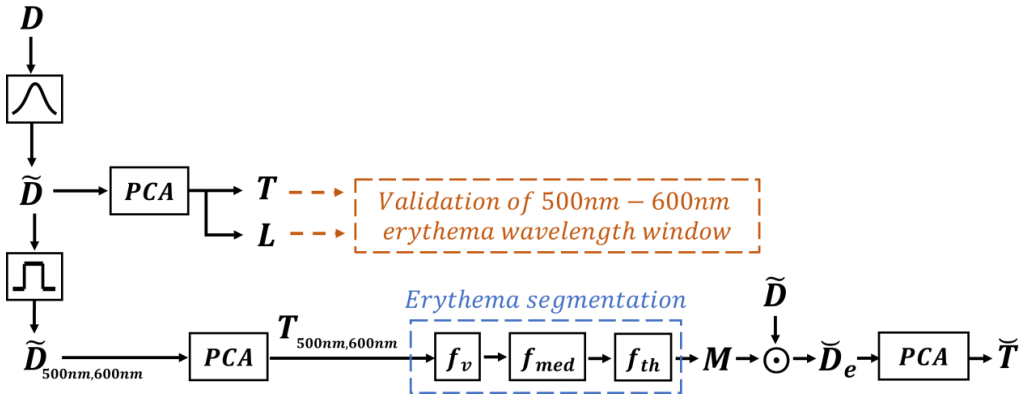


Figure A.12: Overview of Induration Segmentation Model

Pixelwise Labelling of TST Spectral Signature Samples

The spectral signature samples defined by the set $\widetilde{\boldsymbol{D}} \backslash \breve{\boldsymbol{D}}_{\boldsymbol{e}} = \{\boldsymbol{P} \in \widetilde{\boldsymbol{D}} | \boldsymbol{P} \notin \breve{\boldsymbol{D}}_{\boldsymbol{e}}\}$ are assigned the pixelwise label "normal skin". In other words, the "normal skin" label is assigned to the pixels of the spectral signature samples that are present in $\widetilde{\boldsymbol{D}}$ but not in $\breve{\boldsymbol{D}}_{\boldsymbol{e}}$. The spectral signature samples in $\breve{\boldsymbol{T}}$ are assigned the pixelwise labels of "induration" or "erythema only". Based on the expectation that the samples in $\breve{\boldsymbol{T}}$ are grouped into two unique clusters, applying an intensity threshold value of ψ should delineate the cluster of induration samples from the samples that can be labelled "erythema only". Thus, the spectral signature samples defined by the set $\{\boldsymbol{P} | \boldsymbol{P} \in (\breve{\boldsymbol{T}} > \psi)\}$ are assigned the "induration" pixelwise label. Similarly, the spectral signature samples defined by the set $\{\boldsymbol{P} | \boldsymbol{P} \in (\breve{\boldsymbol{T}} < \psi)\}$ are assigned the "erythema only" pixelwise label.

Validation of the PCA Features against Clinician Mantoux Readings to inform Downstream HSI-based Induration Segmentation

In the absence of ground truth pixelwise labels, it is important to correlate the pixelwise labelling of the perceived induration samples in $\breve{\boldsymbol{T}}$ with the hypercube-level labels $\boldsymbol{G} = \{g_1, g_2, \dots, g_a\}$, which are the clinician readings of the induration for each participant. If there are indeed two unique cluster of samples in $\breve{\boldsymbol{T}}$ per participant – the "induration" and "erythema only" clusters, then the means and standard deviations for each per-participant samples in $\breve{\boldsymbol{T}}$ can be useful features to validate the accuracy of the clusters against the clinician readings in $\boldsymbol{G}$. It can be expected that participant samples

with little or no measured clinically significant indurations ($< 10mm$) will have significantly different means and standard deviations compared to participants with measured clinically significant indurations ($\geq 10mm$). These features offer a per- participant description of the transformed data in $\tilde{\boldsymbol{T}}$ which can be compared with the per- participant clinician readings in $\boldsymbol{G}$.

These features are then fed to a classification model to classify the representation of the participants in $\tilde{\boldsymbol{T}}$ into "induration" and "erythema only" classes using the per- participant labels in $\boldsymbol{G}$ to validate the accuracy of the features and the underlying clustering. It is desired that a simple classification decision boundary be created between the two classes. The more complex the decision boundary the greater the likelihood of overfitting. Support vector machines is the classification model utilized in this study, and it is used to create simple and easily explainable decision boundaries between the classes.

APPENDIX B

Support Vector Machines (SVM) Implementation

The SVM classification can be implemented in a 2D feature space where the features are the mean of the samples in $\bar{\boldsymbol{T}}$ and the standard deviation of the samples in $\bar{\boldsymbol{T}}$. The vector representation of the training participants $i = 1,2,\dots,a$ in this feature space is denoted as $\vec{\boldsymbol{X}}_{\boldsymbol{1}}, \vec{\boldsymbol{X}}_{\boldsymbol{2}}, \dots, \vec{\boldsymbol{X}}_{\boldsymbol{a}}$, as shown in equation (B.1). The goal of SVM is to create a decision boundary defined as the optimal hyperplane, given by equation (B.3) derived in (Shmilovici, 2010), that maximizes the margin $2\varphi_h$ between the positive and negative classes. Where $\vec{\boldsymbol{z}}$ is the vector normal to the decision boundary, as shown in Figure B.1, and δ is the intercept. The positive class ($g_i = +1$) is the class of participants who have measured clinically significant TST induration $\geq 10\ mm$. Similarly, the negative class $g_i = -1$) is the class of participants who have clinically measured TST induration $< 10\ mm$, as shown in equation (B.2). The positive class hyperplane which defines the boundary of the positive class is defined by equation (B.4) derived from (Steinwart & Christmann, 2008). The negative class hyperplane which defines the boundary of the negative class is defined by equation (B.5) (Steinwart & Christmann, 2008). The vectors that lie on the positive and negative class hyperplanes are the support vectors, as shown in Figure B.1. The vector $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ for participant i is a support vector if equation (B.6) holds true and where g_i is the class label for the participant. The vector $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ for participant i is a non-support vector in the positive class (clinically measured induration $\geq 10\ mm$) if the inequality (B.7) holds true, and a non-support vector in the negative class if the inequality (B.8) holds true. The goal in SVM classification is to find the support vectors that facilitate the creation of a decision boundary that maximizes the margin between the classes.

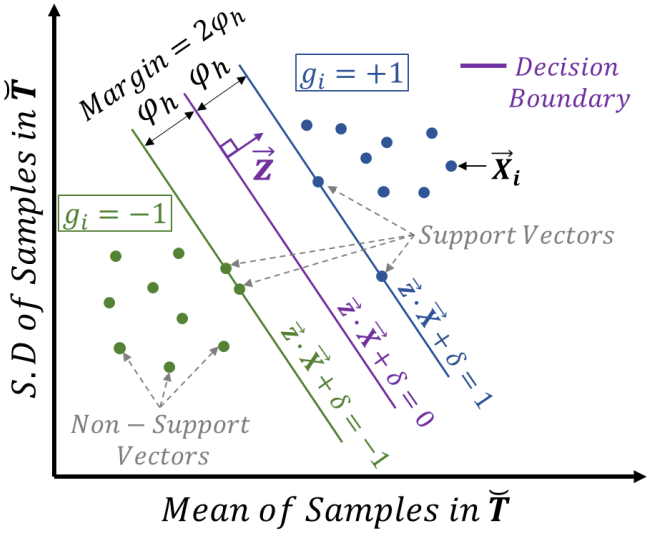


Figure B.1: Description of SVM classification in the PCA-based feature space

$$Training\ subjects = \begin{bmatrix} \vec{X}_1 \\ \vec{X}_2 \\ \vdots \\ \vec{X}_a \end{bmatrix} \tag{B.1}$$

$$, where\ \vec{X}_i = \begin{bmatrix} Mean\ of samples\ in\ \breve{T}\ for\ subject\ i \\ Standard\ deviation\ of samples\ in\ \breve{T}\ for\ subject\ i \end{bmatrix}$$

$$Class\ labels = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_a \end{bmatrix} \tag{B.2}$$

$$, where\ g_i = \begin{cases} +1, & If\ subject\ i\ has > 0mm\ clinically\ measured\ induration \\ -1, & If\ subject\ i\ has\ 0mm\ clinically\ measured\ induration \end{cases}$$

$$\vec{z} \cdot \vec{X} + \delta = 0 \tag{B.3}$$

$$\vec{z} \cdot \vec{X} + \delta = 1 \tag{B.4}$$

$$\vec{z} \cdot \vec{X} + \delta = -1 \tag{B.5}$$

For support vectors,

$$g_i(\vec{z} \cdot \vec{X}_i + \delta) = 0 \tag{B.6}$$

For non-support vectors in the positive class,

$$g_i(\vec{z} \cdot \vec{X}_i + \delta) > 1 \tag{B.7}$$

For non-support vectors in the negative class,

$$g_i(\vec{z} \cdot \vec{X}_i + \delta) < 1 \tag{B.8}$$

The distance between the decision boundary and the class hyperplanes in the direction of $\vec{z}$ is φ_h. φ_h is also described as the factor by which to scale the unit vector $\hat{z}$ in the direction of $\vec{z}$ such that the endpoint of this scaling lies on the positive class hyperplane $\vec{z} \cdot \vec{X} + \delta = 1$. This is as described in equations (B.9) to (B.16), as described in (Steinwart & Christmann, 2008). As shown in equation (B.17), the margin is inversely proportional to the magnitude of the vector $\vec{z}$. Thus, the margin is maximized by minimizing the magnitude of $\vec{z}$.

$$\hat{z} = \frac{\vec{z}}{\|\vec{z}\|} \tag{B.9}$$

$$\vec{z} \cdot \vec{X} + \delta = 1 \tag{B.10}$$

$$\vec{\boldsymbol{z}} \cdot \left(\vec{\boldsymbol{X}} + \varphi_h \hat{\boldsymbol{z}}\right) + \delta = 1 \tag{B.11}$$

$$\vec{\boldsymbol{z}} \cdot \left(\vec{\boldsymbol{X}} + \varphi_h \frac{\vec{\boldsymbol{z}}}{\|\vec{\boldsymbol{z}}\|}\right) + \delta = 1 \tag{B.12}$$

$$\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{X}} + \delta + \varphi_h \frac{\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{z}}}{\|\vec{\boldsymbol{z}}\|} = 1 \tag{B.13}$$

$$From\ equation\ (\mathrm{B.3}), \vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{X}} + \delta = 0 \tag{B.14}$$

$$0 + \varphi_h \frac{\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{z}}}{\|\vec{\boldsymbol{z}}\|} = 1 \tag{B.15}$$

$$\varphi_h = \frac{1}{\|\vec{\boldsymbol{z}}\|} \tag{B.16}$$

$$Margin = 2\varphi_h = \frac{2}{\|\vec{\boldsymbol{z}}\|} \tag{B.17}$$

Soft margin SVM is a variant of SVM classification that generates the optimal decision boundary hyperplane while allowing for misclassification of outliers in the feature space. The hinge loss function, defined in equation (B.18) and described in (Moosaei et al., 2023), is the cost function used in SVM to compute the error between the predicted class label of vector $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ and its true class label g_i. The hinge loss is defined such that if the predicted score $(\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{X}} + \delta)$ for participant datapoint vector $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ and its true class label g_i have the same polarity and the predicted score > 1, then the hinge loss is 0. This connotes correct class prediction for participant datapoint vector $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$. If the predicted score for $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ and its true class label g_i have different polarities, connoting a misclassification of $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$, the hinge loss will be inversely proportional to the predicted score, as shown in Figure B.2. If $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ is correctly classified but the predicted score $(\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{X}} + \delta)$ is between 0 and 1, then $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ lies within the margin $2\varphi_h$ and beyond its class hyperplane.

$$Hinge\ loss = max\left(0, 1 - g_i(\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{X}}_{\boldsymbol{i}} + \delta)\right) \tag{B.18}$$

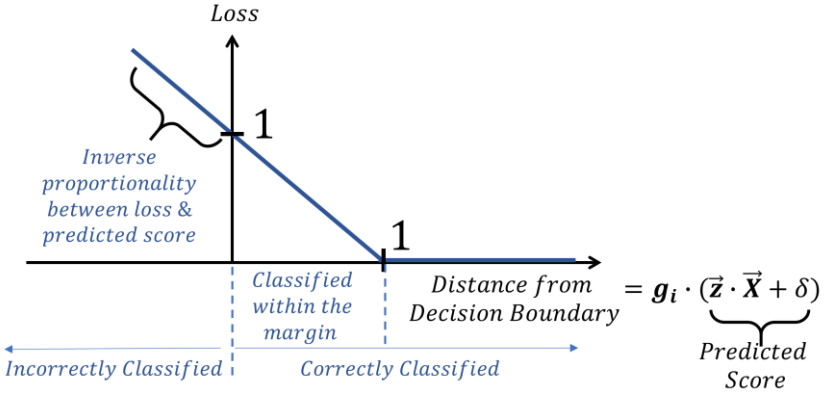


Figure B.2: Description of SVM hinge loss

The soft margin SVM can be defined as a search for the optimal vector $\vec{\boldsymbol{z}}$ and intercept δ in the feature space that minimizes the average hinge loss for all participants $i = 1,2, \dots, a$ and the magnitude of $\vec{\boldsymbol{z}}$, participant to the constraint that $\vec{\boldsymbol{X}}_{\boldsymbol{i}}$ for all $i = 1,2, \dots, a$ is correctly classified, as described in statements (B.19) and (B.20) obtained from (Pecha & Horák, 2020). Minimizing the magnitude of $\vec{\boldsymbol{z}}$ maximizes the width of the margin, as described in equation (B.17). The regularisation parameter ξ is a scalar that controls the balance between the objective of maximizing the margin width and the objective of minimizing the hinge loss.

$$\underbrace{\underset{z,\delta}{\operatorname{argmin}} \left[\left(\frac{1}{a} \sum_{i=1}^{a} max\left(0, 1 - g_i\left(\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{X}}_{\boldsymbol{i}} + \delta\right)\right)\right) + \xi \|\vec{\boldsymbol{z}}\|^2 \right]}_{Objective\ Function} \tag{B.19}$$

$$s.t. \quad \underbrace{g_i\left(\vec{\boldsymbol{z}} \cdot \vec{\boldsymbol{X}}_{\boldsymbol{i}} + \delta\right) \geq 1,}_{Constraint} \qquad \forall\, i = 1,2, \dots, a \tag{B.20}$$

APPENDIX C

Description of Deterministic Radiomics Features

GLSZM-based Features

The GLSZM-based features are generated from the GLSZM matrix $\boldsymbol{P^{szm}}$ of the image $\boldsymbol{C}$. The matrix quantifies the gray level zones in an image. A gray level zone can be defined as the number of connected pixels with the same gray level intensity. The $(j,k)^{\text{th}}$ element in $\boldsymbol{P^{szm}}$ is the number of zones with gray level j and zone size k that appear in image $\boldsymbol{C}$ (Thibault et al., 2013). The generation of the matrix $\boldsymbol{P^{szm}}$ from the image $\boldsymbol{C}$ is as illustrated in Figure C.1. If the number of discrete zone sizes in the image is denoted by N_s and N_z denotes the number of zones in the image, then N_z can be written as $\sum_{j=1}^{N_g}\sum_{k=1}^{N_s} P^{szm}(j,k)$, and the normalised GLSZM defined as $p^{szm}(j,k) = \frac{P^{szm}(j,k)}{N_z}$.

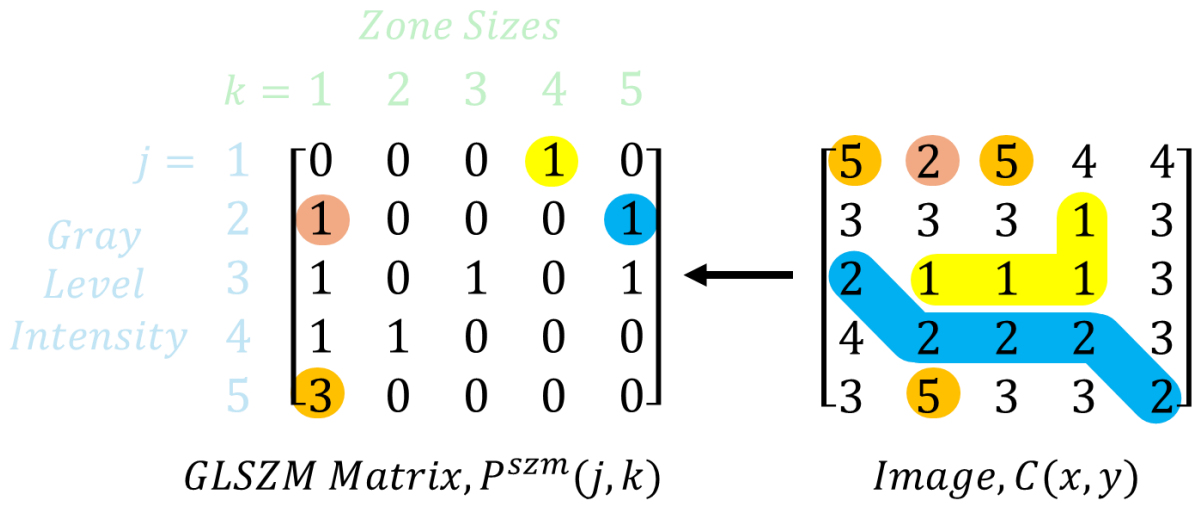


Figure C.1: Illustration of the generation of the GLSZM matrix from a given image.

The radiomics features generated from the GLSZM matrix $\boldsymbol{P^{szm}}$ are as outlined in Table C.1. The measure of the distribution of small size zones in $\boldsymbol{C}$ is captured by the small area emphasis feature. Conversely, the large area emphasis feature quantifies the distribution of large area zones in the image. The gray level non-uniformity feature, and its normalised form, measures the variability of intensity values in the image. The size-zone non-uniformity feature, and its normalised form, quantifies the variability of size zone areas in the image. A measure of the coarseness of the image texture is captured by the zone percentage (Zwanenburg et al., 2016). The gray level variance feature measures the variance in intensities per zone in an image. Conversely, the zone variance feature measures the variance in the zone sizes in the image. A measure of the randomness in the distribution of zone sizes and gray levels is captured by the zone entropy. The low gray level zone emphasis feature quantifies the distribution of lower gray-level size zones in the image. Conversely, the high gray level

zone emphasis feature quantifies the distribution of higher gray-level values in the image (van Timmeren et al., 2020). The proportion of the combined distribution of smaller zones with lower gray-levels in the image is measured by the small area low gray level emphasis feature. In contrast, the small area high gray level emphasis feature measures the proportion of the combined distribution of smaller zones with higher gray-levels. The large area low gray level emphasis feature measures the proportion of the combined distribution of larger size zones with lower gray levels. Conversely, the large area high gray level emphasis feature quantifies the proportion of the combined distribution of larger size zones with higher gray levels (Zwanenburg et al., 2020).

Table C.1: GLSZM-based Deterministic Radiomics Features (Zwanenburg et al., 2016)

$small\ area\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}\frac{P^{szm}(j,k)}{k^2}}{N_z}$	$large\ area\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}P^{szm}(j,k)\cdot k^2}{N_z}$
$gray\ level\ non-uniformity = \frac{\sum_{j=1}^{N_g}\left(\sum_{k=1}^{N_s}P^{szm}(j,k)\right)^2}{N_z}$	$gray\ level\ non-uniformity\ normalised = \frac{\sum_{j=1}^{N_g}\left(\sum_{k=1}^{N_s}P^{szm}(j,k)\right)^2}{{N_z}^2}$
$size-zone\ non-uniformity = \frac{\sum_{k=1}^{N_s}\left(\sum_{j=1}^{N_g}P^{szm}(j,k)\right)^2}{N_z}$	$zone\ percentage = \frac{N_z}{N_p}$
$size-zone\ non-uniformity\ normalised = \frac{\sum_{k=1}^{N_s}\left(\sum_{j=1}^{N_g}P^{szm}(j,k)\right)^2}{{N_z}^2}$	
$gray\ level\ variance = \sum_{j=1}^{N_g}\sum_{k=1}^{N_s}p^{szm}(j,k)\cdot(j-\mu_{szm})^2\ ;\ where\ \mu_{szm} = \sum_{j=1}^{N_g}\sum_{k=1}^{N_s}p^{szm}(j,k)\cdot j$	
$zone\ variance = \sum_{j=1}^{N_g}\sum_{k=1}^{N_s}p^{szm}(j,k)\cdot(k-\mu_{szm})^2\ ;\ where\ \mu_{szm} = \sum_{j=1}^{N_g}\sum_{k=1}^{N_s}p^{szm}(j,k)\cdot k$	
$zone\ entropy = -\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}p^{szm}(j,k)\cdot\log_2(p^{szm}(j,k)+\epsilon)$	$low\ gray\ level\ zone\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}\frac{P^{szm}(j,k)}{j^2}}{N_z}$
$high\ gray\ level\ zone\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}P^{szm}(j,k)\cdot j^2}{N_z}$	$small\ area\ low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}\frac{P^{szm}(j,k)}{j^2\cdot k^2}}{N_z}$
$small\ area\ high\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}\frac{P^{szm}(j,k)\cdot j^2}{k^2}}{N_z}$	$large\ area\ low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}\frac{P^{szm}(j,k)\cdot k^2}{j^2}}{N_z}$
$large\ area\ low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_s}P^{szm}(j,k)\cdot j^2\cdot k^2}{N_z}$	

GLRLM-based Features

The GLRLM-based features are generated from GLRLM matrix $\boldsymbol{P^{rlm}}$ of the image $\boldsymbol{C}$. The matrix quantifies the gray level runs in the image. A gray level run is the length, in number of pixels, of consecutive pixels along a specified direction that have the same intensity in $\boldsymbol{C}$. The $(j,k)^{\text{th}}$ element in $\boldsymbol{P^{rlm}}$ is the number of runs with gray level j and run length k that appear in image $\boldsymbol{C}$ (Zwanenburg et al., 2016). The generation of the matrix $\boldsymbol{P^{rlm}}$ from the image $\boldsymbol{C}$ for gray level runs in the horizontal direction is as illustrated in Figure C.2.

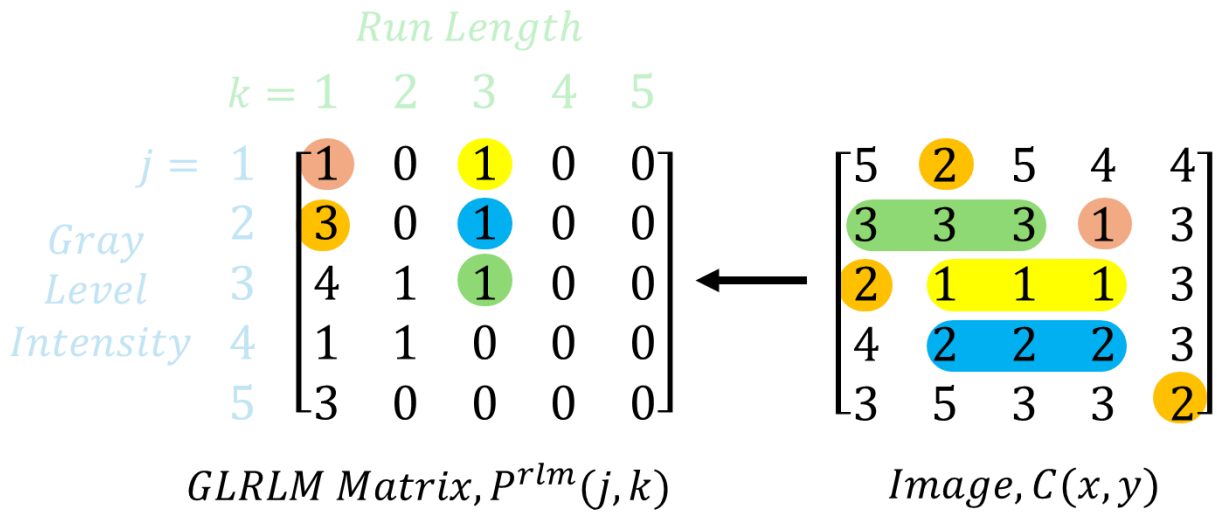


Figure C.2: Illustration of the generation of the GLRLM matrix from a given image.

The radiomics features generated from the GLRLM matrix $\boldsymbol{P^{rlm}}$ are as outlined in Table C.2. If the number of discrete run lengths in $\boldsymbol{C}$ is denoted as N_r, then the normalized run length matrix can be defined as $p^{rlm}(j,k) = \frac{P^{rlm}(j,k)}{N_r}$. The distribution of short run lengths in the image is captured by the short run emphasis feature. Conversely, the long run emphasis feature quantifies the distribution of long run lengths. The GLRLM gray level non-uniformity feature, and its normalised form, measures the similarity of the gray level intensities in the image. The similarity of run lengths in the image is quantified by the run length non-uniformity feature, and its normalised form. The run percentage feature measures the coarseness of an image. The variance in the gray levels per run length is captured by the gray level variance feature. Conversely, the variance in number of runs per run lengths in an image is captured by the run variance feature (Zwanenburg et al., 2016). Run entropy quantifies the randomness in the distribution of run lengths and gray levels. The low gray level run emphasis measures the distribution of the low gray level intensities. In contrast, the high gray level run emphasis feature quantifies the distribution of the higher gray level intensities (Chu et al., 1990). The combined distribution of shorter run lengths with lower gray levels are quantified by the short run low gray level emphasis feature. Conversely, the combined distribution of shorter run lengths with higher gray levels

are quantified by the short run high gray level emphasis feature. The long run low gray level emphasis feature measures the combined distribution of long run lengths with lower gray levels. In contrast, the long run high gray level emphasis feature measures the combined distribution of long run lengths with higher gray levels (Zwanenburg et al., 2016; Zwanenburg et al., 2020).

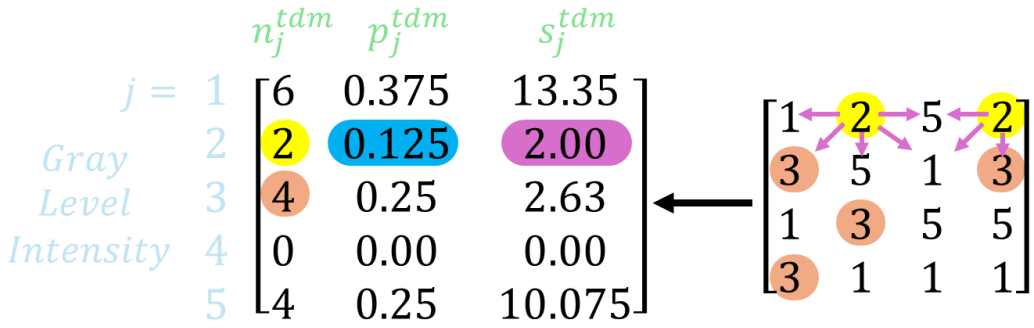
Table C.2: GLRLM-based Deterministic Radiomics Features (Zwanenburg et al., 2016)

$$short\ run\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}\frac{P^{rlm}(j,k)}{k^2}}{N_r}$$	$$long\ run\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\cdot k^2}{N_r}$$
$$gray\ level\ non-uniformity = \frac{\sum_{j=1}^{N_g}\left(\sum_{k=1}^{N_r}P^{rlm}(j,k)\right)^2}{N_r}$$	$$gray\ level\ non-uniformity\ normalised = \frac{\sum_{j=1}^{N_g}\left(\sum_{k=1}^{N_r}P^{rlm}(j,k)\right)^2}{{N_r}^2}$$
$$run\ length\ non-uniformity = \frac{\sum_{k=1}^{N_r}\left(\sum_{j=1}^{N_g}P^{rlm}(j,k)\right)^2}{N_r}$$	$$run\ length\ non-uniformity\ normalised = \frac{\sum_{k=1}^{N_r}\left(\sum_{j=1}^{N_g}P^{rlm}(j,k)\right)^2}{{N_r}^2}$$
$$run\ percentage = \frac{N_r}{N_p}$$	$$gray\ level\ variance = \sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\cdot(j-\mu_{rlm})^2$$ $$,where\ \mu_{rlm} = \sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\cdot j$$
$$run\ variance = \sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\cdot(k-\mu_{rlm})^2$$ $$,where\ \mu_{rlm} = \sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\cdot k$$	$$run\ entropy = -\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\log_2(P^{rlm}(j,k)+\epsilon)$$ $$,where\ \epsilon = 2.2\times10^{-16}$$
$$low\ gray\ level\ run\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}\frac{P^{rlm}(j,k)}{j^2}}{N_r}$$	$$high\ gray\ level\ run\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\cdot j^2}{N_r}$$
$$short\ run\ low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}\frac{P^{rlm}(j,k)}{j^2\cdot k^2}}{N_r}$$	$$short\ run\ high\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}\frac{P^{rlm}(j,k)}{k^2}}{N_r}$$
$$long\ run\ low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}\frac{P^{rlm}(j,k)\cdot k^2}{j^2}}{N_r}$$	$$long\ run\ high\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_r}P^{rlm}(j,k)\cdot j^2\cdot k^2}{N_r}$$

NGTDM-based Features

The NGTDM-based features are generated from the NGTDM matrix $\boldsymbol{P^{tdm}}$ of the image $\boldsymbol{C}$. The matrix quantifies the difference between the gray level intensity of a pixel and the average intensity of its neighbours (Amadasun & King, 1989). The generation of the matrix $\boldsymbol{P^{tdm}}$ from the image $\boldsymbol{C}$ for a

neighbourhood distance of one pixel is as illustrated in Figure C.3. The variables n_j^{tdm}, p_j^{tdm}, and s_j^{tdm} are also as defined in the Figure C.3.



$NGTDM\ Matrix, P^{tdm}(j)$ $Image, C(x,y)$

$$P^{tdm}(j) = \left[n_j^{tdm}, p_j^{tdm}, s_j^{tdm}\right]$$

$$p^{tdm}(2) = \frac{n_2^{tdm}}{No.\,of\ pixels} = \frac{2}{16} = 0.125$$

$$s^{tdm}(2) = \left|2 - \frac{(1+3+5+1+5)}{5}\right| + \left|2 - \frac{(5+1+3)}{3}\right| = 2.00$$

Figure C.3: Illustration of the generation of the NGTDM matrix from a given image.

The radiomics features generated from the NGTDM matrix $\boldsymbol{P^{tdm}}$ are as outlined in Table C.3. A measure of the average difference between the centre pixel and its neighbourhood is quantified by the coarseness feature. The contrast feature quantifies the spatial intensity change and the dynamic range of an image. A measure of the change in intensity from a pixel to its neighbour is captured by the busyness feature. The complexity variable measures the frequency of the occurrence neighbourhoods with varying intensities. The strength feature quantifies the rate of change in intensity in neighbourhoods of pixels (Amadasun & King, 1989).

Table C.3: NGTDM-based Deterministic Radiomics Features

$coarseness = \frac{1}{\sum_{j=1}^{N_g} p_j^{tdm} s_j^{tdm}}$	$busyness = \frac{\sum_{j=1}^{N_g} p_j^{tdm} s_j^{tdm}}{\sum_{j=1}^{N_g}\sum_{k=1}^{N_g} \lvert j \cdot p_j^{tdm} - k \cdot p_k^{tdm}\rvert}$
$contrast = \left(\frac{1}{N_{g,p}(N_{g,p}-1)} \sum_{j=1}^{N_g}\sum_{k=1}^{N_g} p_j^{tdm} p_k^{tdm} (j-k)^2\right)\left(\frac{1}{N_p}\sum_{j=1}^{N_g} s_j^{tdm}\right)$, $where\ N_{g,p} = number\ of\ gray\ levels\ where\ p_j^{tdm} \neq 0$	
$complexity = \frac{1}{N_p}\sum_{j=1}^{N_g}\sum_{k=1}^{N_g} \lvert j-k\rvert \frac{p_j^{tdm} s_j^{tdm} + p_k^{tdm} s_k^{tdm}}{p_j^{tdm} + p_k^{tdm}}, where\ p_j^{tdm}, p_k^{tdm} = 0$	

$$strength = \frac{\sum_{j=1}^{N_g}\sum_{k=1}^{N_g}(p_j^{tdm} + p_k^{tdm})(j-k)^2}{\sum_{j=1}^{N_g} s_j^{tdm}}$$

GLDM-based Features

The GLDM-based features are generated from the GLDM matrix $\boldsymbol{P^{dm}}$ of the image $\boldsymbol{C}$. The matrix quantifies the gray level dependencies that exist in the image. A gray level dependency is the number of connected pixels within a neighbourhood that are dependent on a centre pixel. The dependency is determined by a specified condition of equivalence between neighbouring pixels and a centre pixel (Sun & Wee, 1983). The $(j,k)^{\text{th}}$ element in $\boldsymbol{P^{dm}}$ is the number of times a pixel with gray level intensity j with k dependent pixels appear in image $\boldsymbol{C}$. The generation of the matrix $\boldsymbol{P^{dm}}$ from the image $\boldsymbol{C}$ for the dependency condition that a neighbouring pixel is equal in intensity to a centre pixel, is as illustrated in Figure C.4.

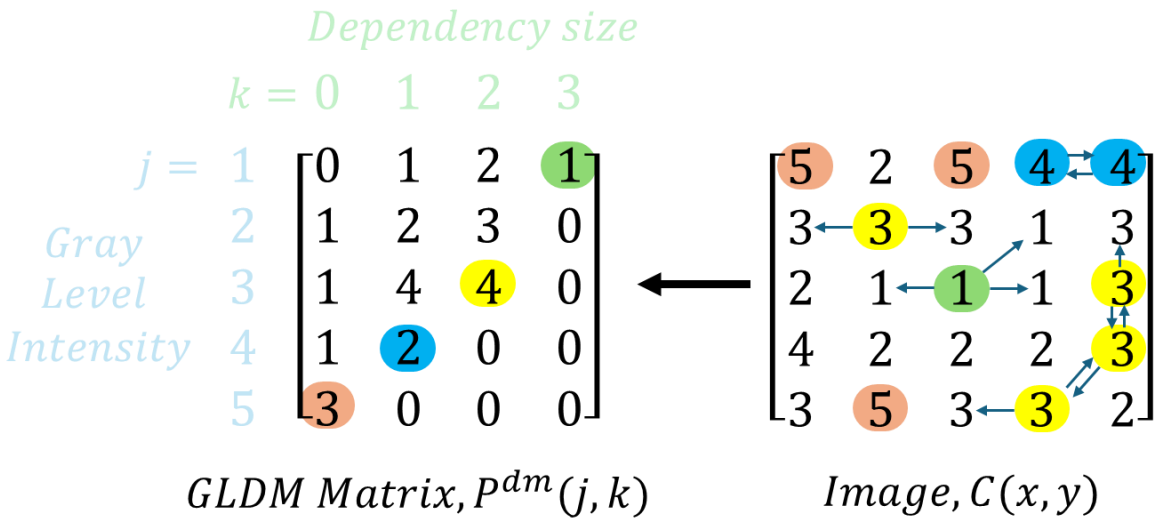


Figure C.4: Illustration of the generation of the GLDM matrix from a given image.

The radiomics features generated from the GLDM matrix $\boldsymbol{P^{dm}}$ are as outlined in Table C.4. If the number of discrete dependency sizes in $\boldsymbol{C}$ is denoted as N_d, then the number of dependency zones in the image can be defined as $N_z = \sum_{j=1}^{N_g}\sum_{k=0}^{N_d} P^{dm}(j,k)$ and the normalised GLDM matrix can be defined as $p^{dm}(j,k) = \frac{P^{dm}(j,k)}{N_z}$. The distribution of small dependencies in the image is captured by the small dependence emphasis feature. Conversely, the large dependence emphasis feature quantifies the distribution of large dependencies. The similarity of gray level intensities in the image is quantified by the gray level non-uniformity feature (Zwanenburg et al., 2016). The similarity of dependence zones through the image is measured by the dependence non-uniformity feature and its normalised form. The variance in the gray level intensities in the image is captured by the gray level

variance feature. In contrast, the variance in the dependency sizes in the image is quantified by the dependence variance feature. The dependence entropy feature quantifies the randomness in the distribution of dependence zones and gray levels. The distribution of low gray level intensities is measured by the low gray level emphasis feature. Conversely, the distribution of higher gray level intensities is measured by the high gray level emphasis feature (Zwanenburg et al., 2016). The small dependence low gray level emphasis feature quantifies the combined distribution of small dependence zones with lower gray level intensities. Conversely, the small dependence high gray level emphasis feature quantifies the combined distribution of small dependence zones with higher gray level intensities. The combined distribution of large dependence zones with lower gray level intensities is measured by the large dependence low gray level emphasise. Conversely, the combined distribution of large dependence zones with higher gray level intensities is quantified by the large dependence high gray level emphasis feature (Zwanenburg et al., 2016; Zwanenburg et al., 2020).

Table C.4: GLDM-based Deterministic Radiomics Features

$$small\ dependence\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}\frac{p^{dm}(j,k)}{j^2}}{N_z}$$	$$large\ dependence\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}p^{dm}(j,k)\cdot k^2}{N_z}$$
$$gray\ level\ non-uniformity = \frac{\sum_{j=1}^{N_g}\left(\sum_{k=0}^{N_d}p^{dm}(j,k)\right)^2}{N_z}$$	$$dependence\ non-uniformity = \frac{\sum_{k=0}^{N_d}\left(\sum_{j=1}^{N_g}p^{dm}(j,k)\right)^2}{N_z}$$
$$dependence\ non-uniformity\ normalised = \frac{\sum_{k=0}^{N_d}\left(\sum_{j=1}^{N_g}p^{dm}(j,k)\right)^2}{{N_z}^2}$$	
$$gray\ level\ variance = \sum_{j=1}^{N_g}\sum_{k=0}^{N_d}p^{dm}(j,k)(j-\mu_{dm})^2, where\ \mu_{dm} = \sum_{j=1}^{N_g}\sum_{k=0}^{N_d}j\cdot p^{dm}(j,k)$$	
$$dependence\ variance = \sum_{j=1}^{N_g}\sum_{k=0}^{N_d}p^{dm}(j,k)(k-\mu_{dm})^2, where\ \mu_{dm} = \sum_{j=1}^{N_g}\sum_{k=0}^{N_d}k\cdot p^{dm}(j,k)$$	
$$dependence\ entropy = -\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}p^{dm}(j,k)\log_2(p^{dm}(j,k)+\epsilon)$$	$$low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}\frac{p^{dm}(j,k)}{j^2}}{N_z}$$
$$high\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}p^{dm}(j,k)\cdot j^2}{N_z}$$	$$small\ dependence\ low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}\frac{p^{dm}(j,k)}{j^2\cdot k^2}}{N_z}$$
$$small\ dependence\ high\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}\frac{p^{dm}(j,k)\cdot j^2}{k^2}}{N_z}$$	
$$large\ dependence\ low\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}\frac{p^{dm}(j,k)\cdot k^2}{j^2}}{N_z}$$	
$$large\ dependence\ high\ gray\ level\ emphasis = \frac{\sum_{j=1}^{N_g}\sum_{k=0}^{N_d}p^{dm}(j,k)\cdot j^2\cdot k^2}{N_z}$$	

APPENDIX D

Principal Component Analysis Full Dataset Results

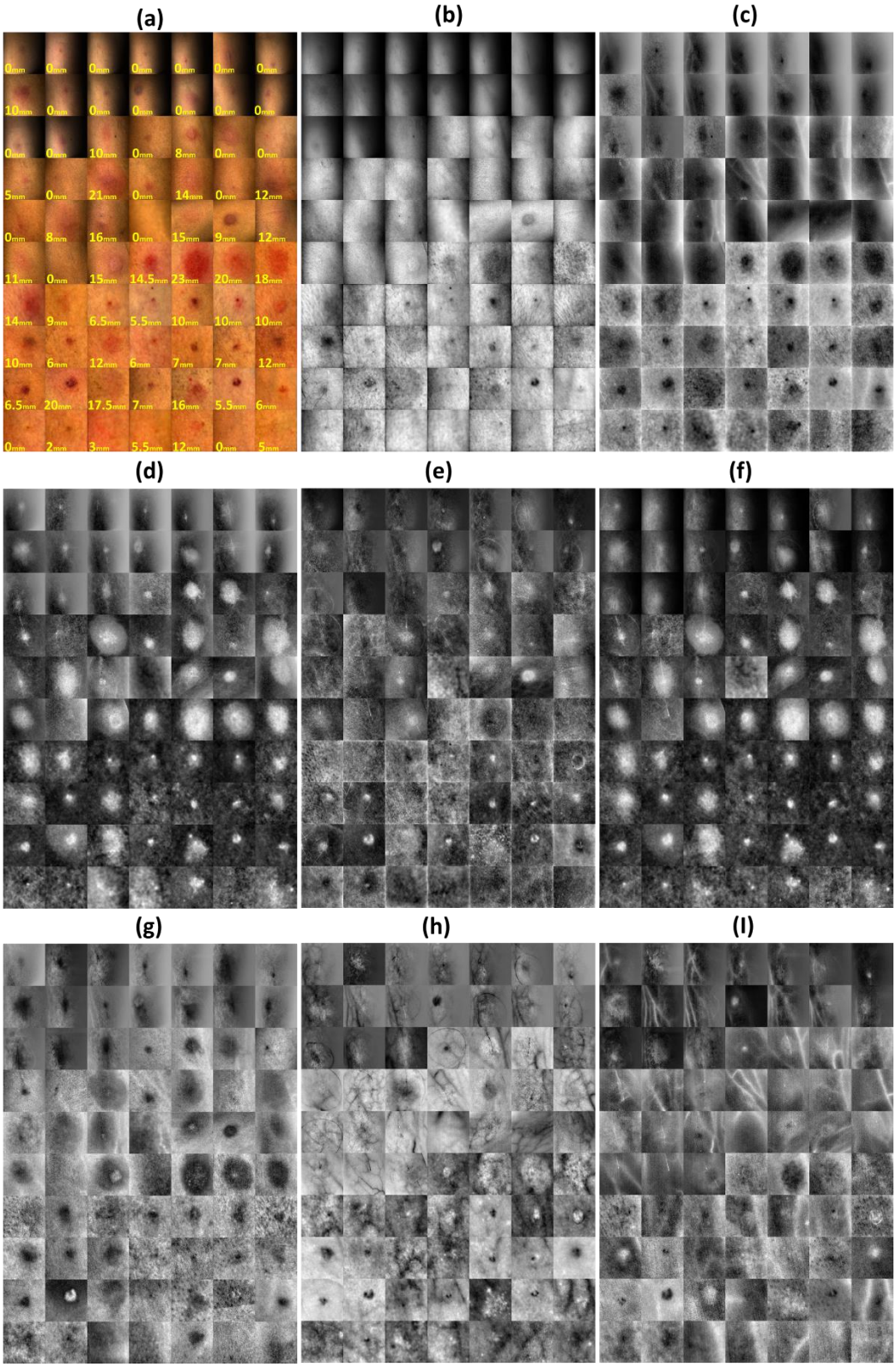


Figure D.1: Full dataset PCA outputs showing: (a) RGB images estimated from the hypercubes, and the projection of the corresponding hypercubes to the (b) P_{c1}, (c) P_{c2}, (d) P_{c3}, (e) P_{c4}, (f) P_{c5}, (g) P_{c6}, (h) P_{c7}, and (i) P_{c8} coordinates

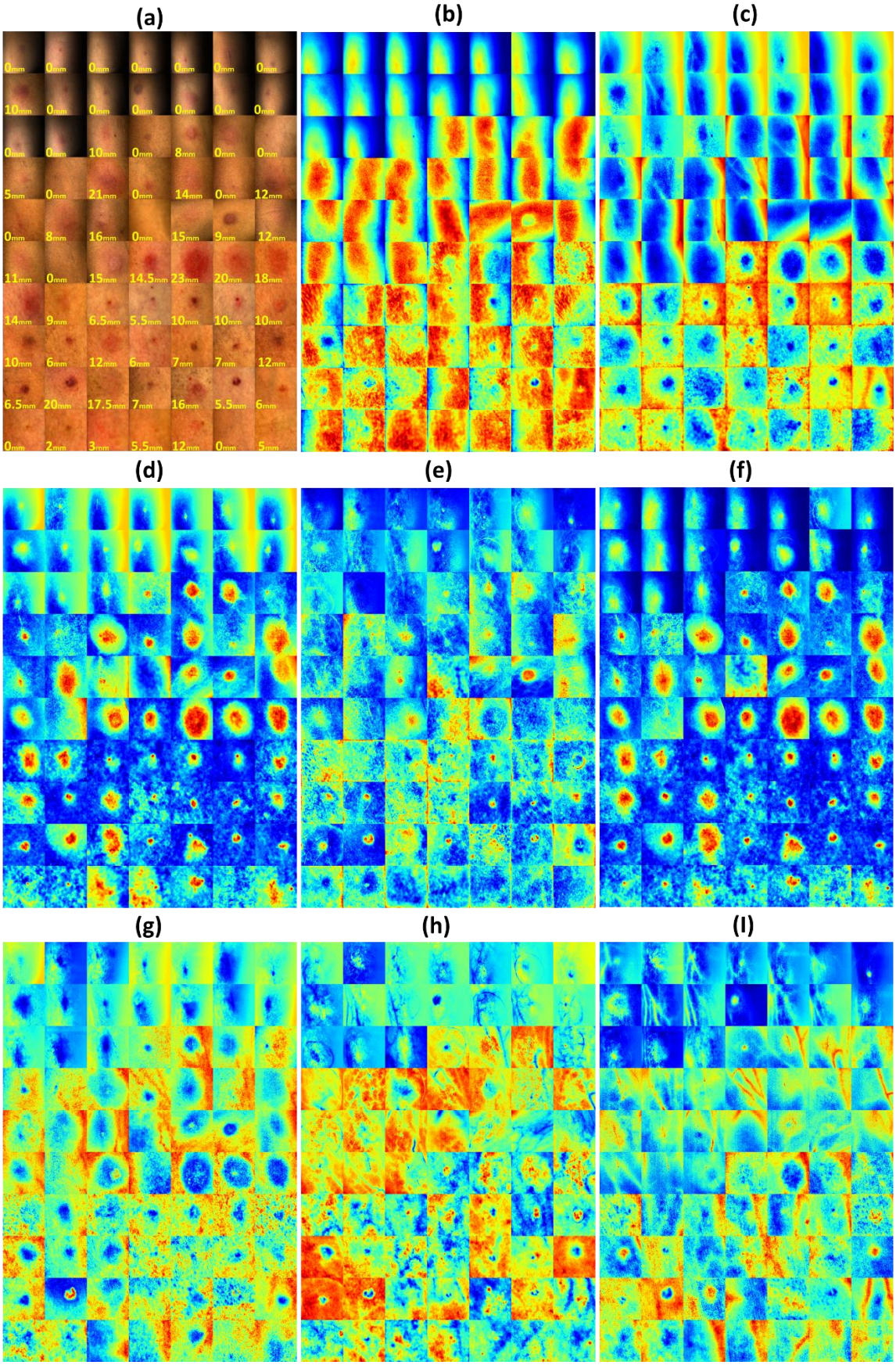


Figure D.2: Full dataset PCA colourmap outputs showing: (a) RGB images estimated from the hypercubes, and the colormaps of projection of the corresponding hypercubes to the (b) $\boldsymbol{P_{c1}}$, (c) $\boldsymbol{P_{c2}}$, (d) $\boldsymbol{P_{c3}}$, (e) $\boldsymbol{P_{c4}}$, (f) $\boldsymbol{P_{c5}}$, (g) $\boldsymbol{P_{c6}}$, (h) $\boldsymbol{P_{c7}}$, and (i) $\boldsymbol{P_{c8}}$ coordinates

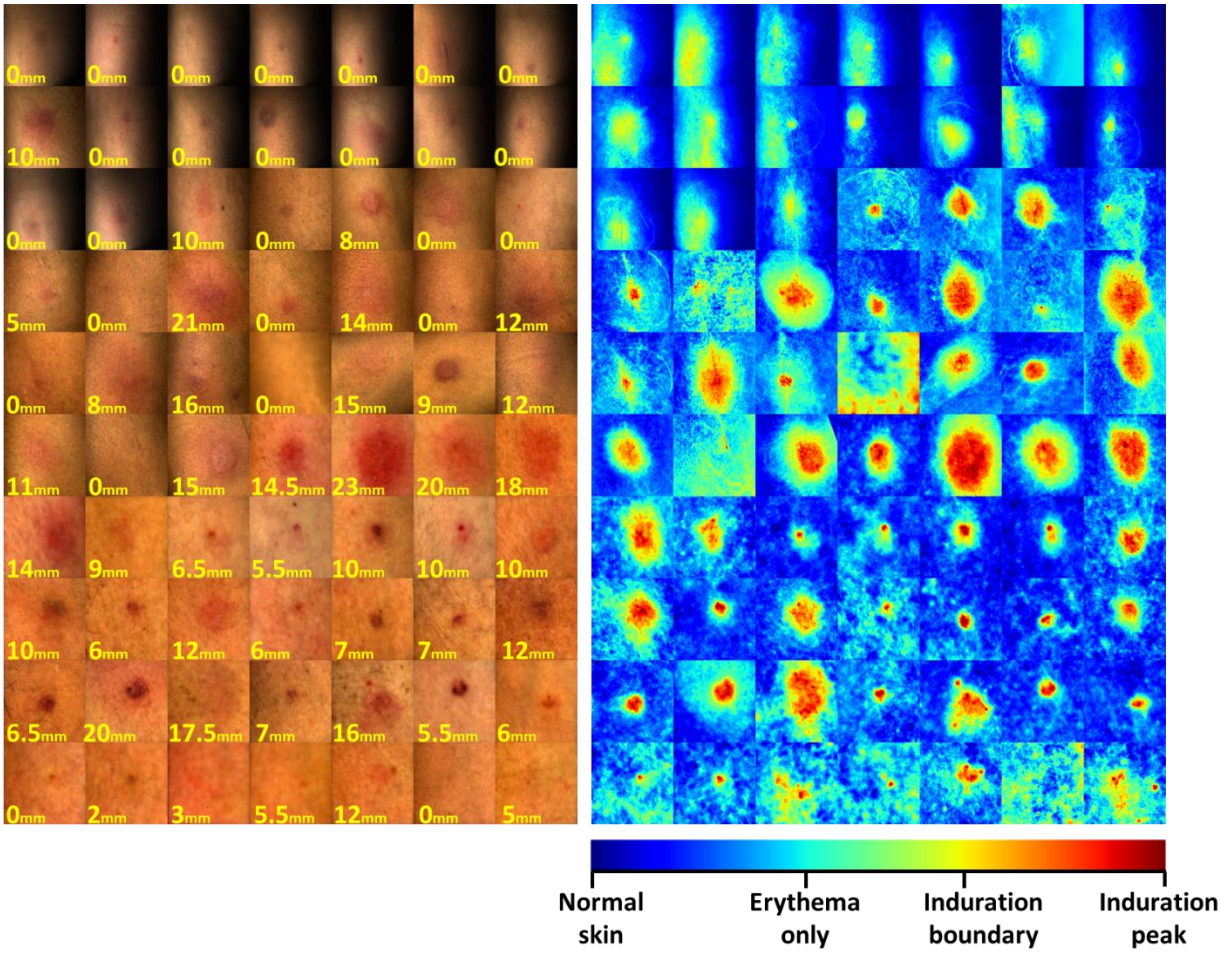


Figure D.3: Colourmap of P_{c5} for the full dataset which offers natural clusters of samples that can facilitate erythema-induration separation

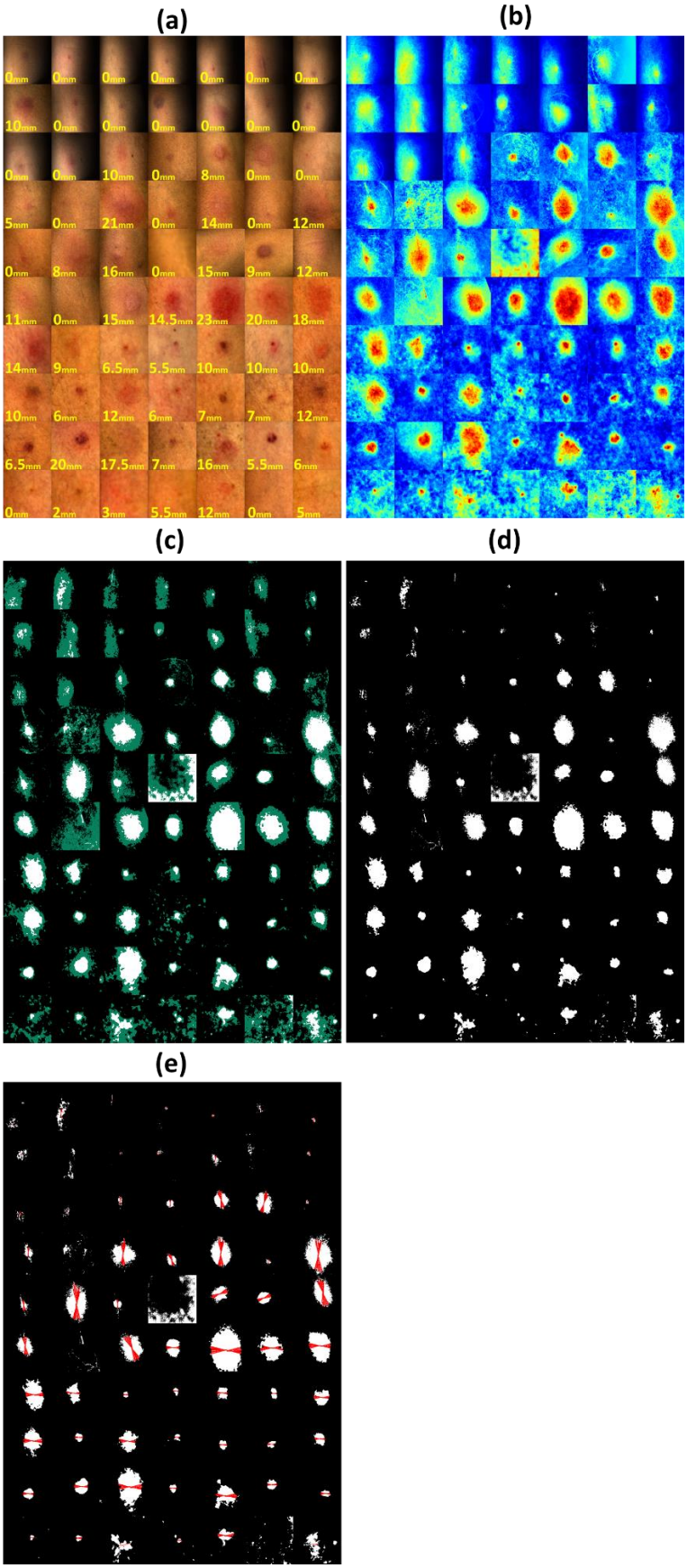


Figure D.4: Visualisation of the segmentation of induration from erythema for the full dataset showing (a) the RBG estimates, (b) transformation of samples in $\boldsymbol{P}_{c5}$, (c) thresholding of induration samples from erythema, (d) resulting induration masks, and (e) estimation of induration diameters

APPENDIX E

Full Dataset Results from Characterisation Framework

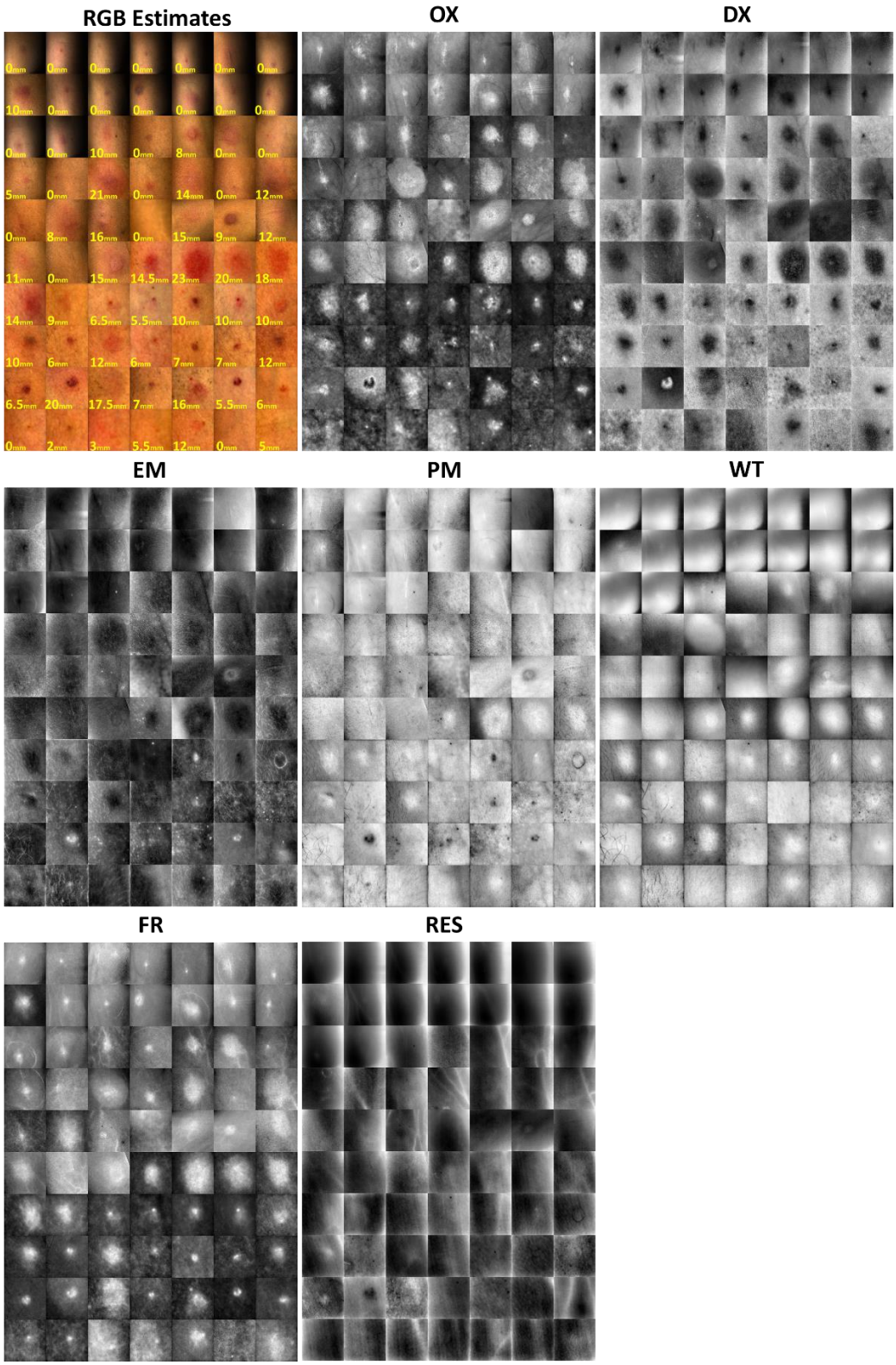


Figure E.1: RGB estimates, and grayscale chromophore maps for all participants showing

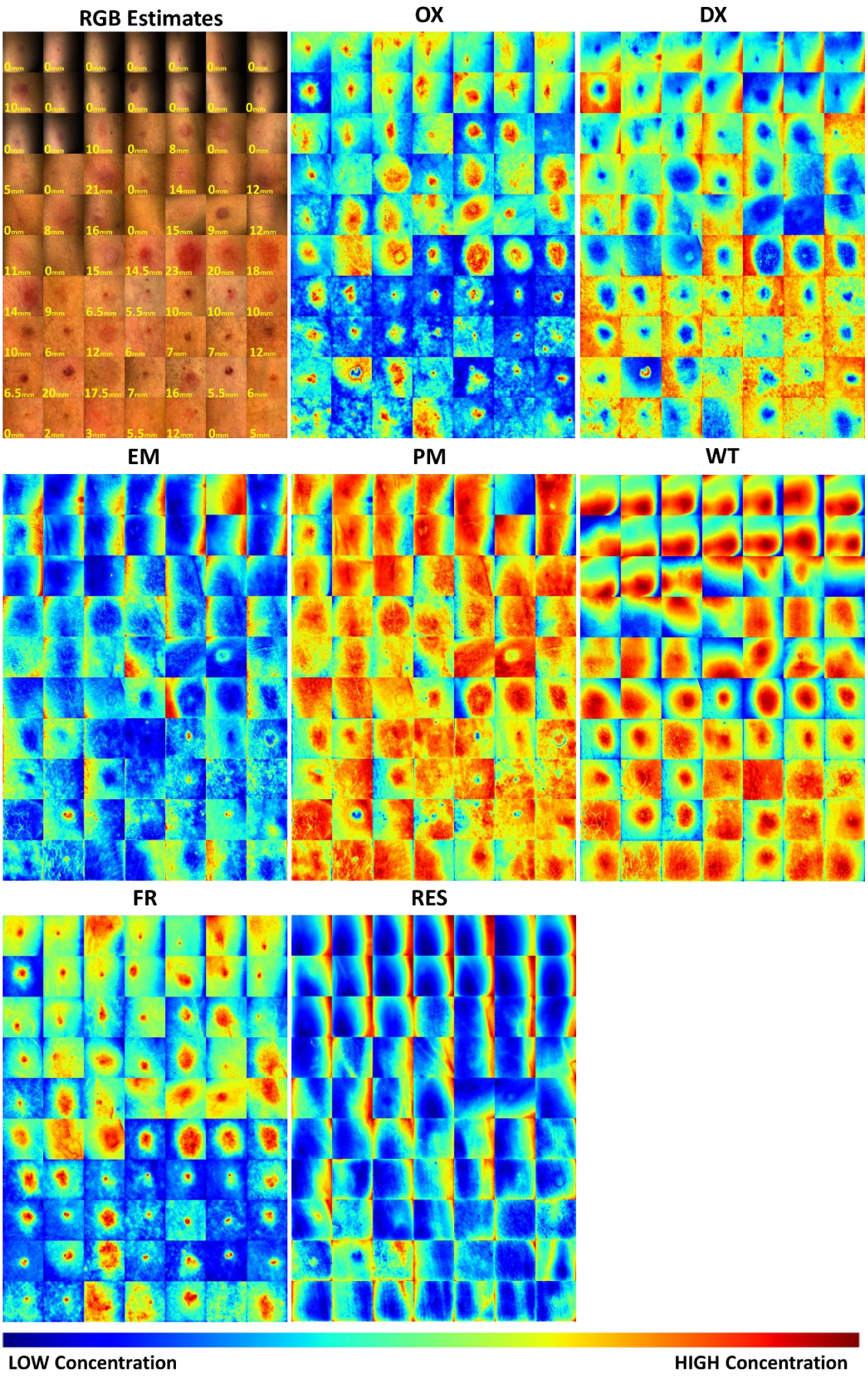


Figure E.2: RGB estimates, and colourmap of the chromophore maps for all participants showing

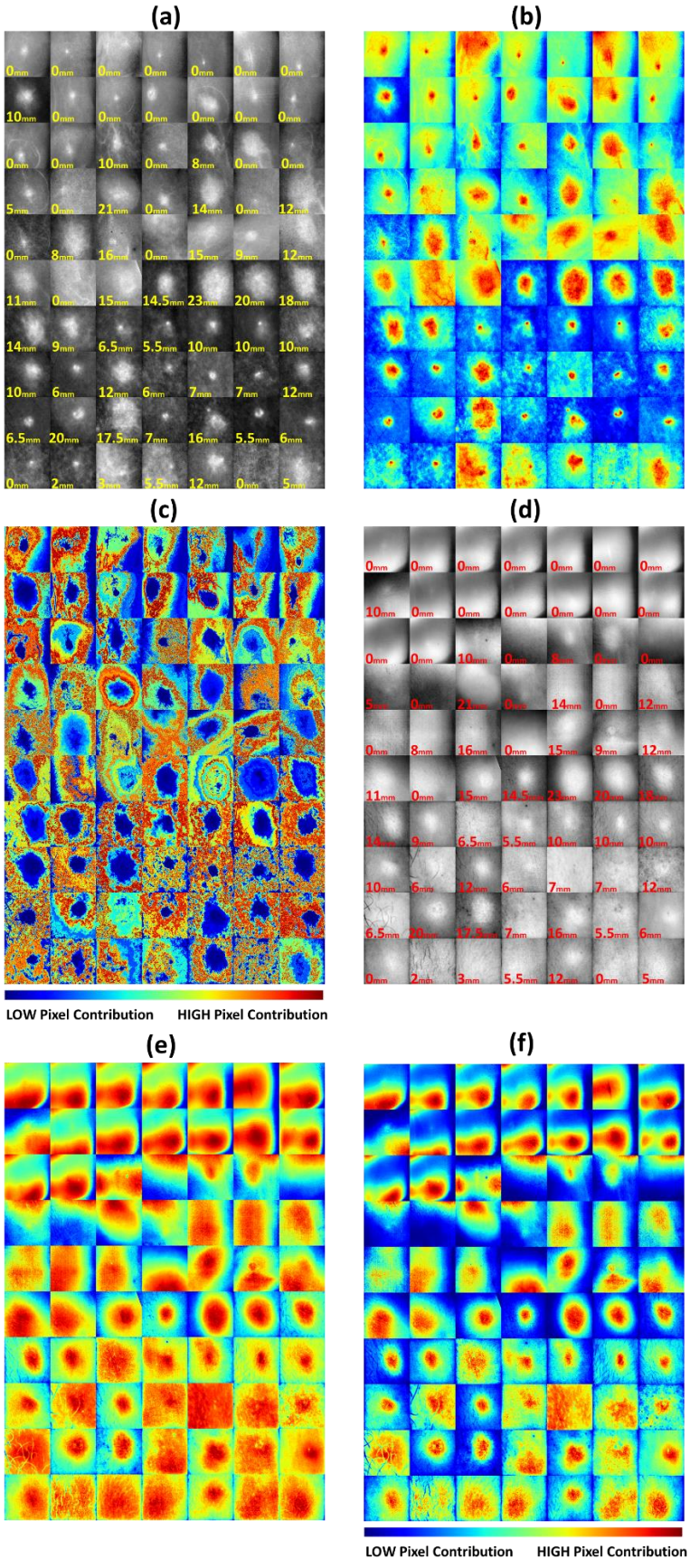


Figure E.3: Visualisation of best deterministic feature pair for all participants showing: (a) grayscale images of FR, (b) colourmap images of FR, (c) feature contribution map for the GLRLM $gray\ level\ nonuniformity$ of FR, (d) grayscale images of WT, (e) colourmap images of WT, (f) feature contribution map for the $total\ energy$ of WT

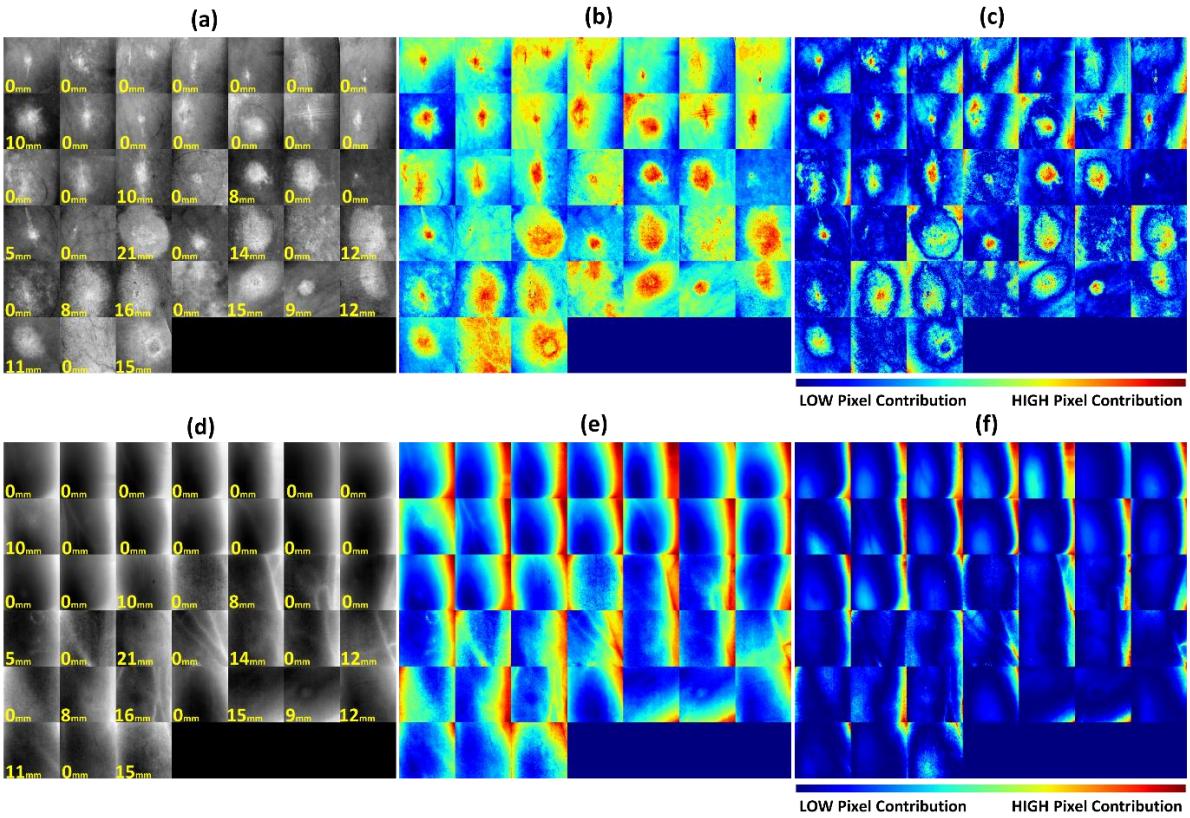


Figure E.4: Visualisation of best deterministic feature pair for the South African participants showing: (a) grayscale images of OX, (b) colourmap images of OX, (c) feature contribution map for the $mean\ absolute\ deviation$ of OX, (d) grayscale images of RES, (e) colourmap images of RES, (f) feature contribution map for the $variance$ of RES.

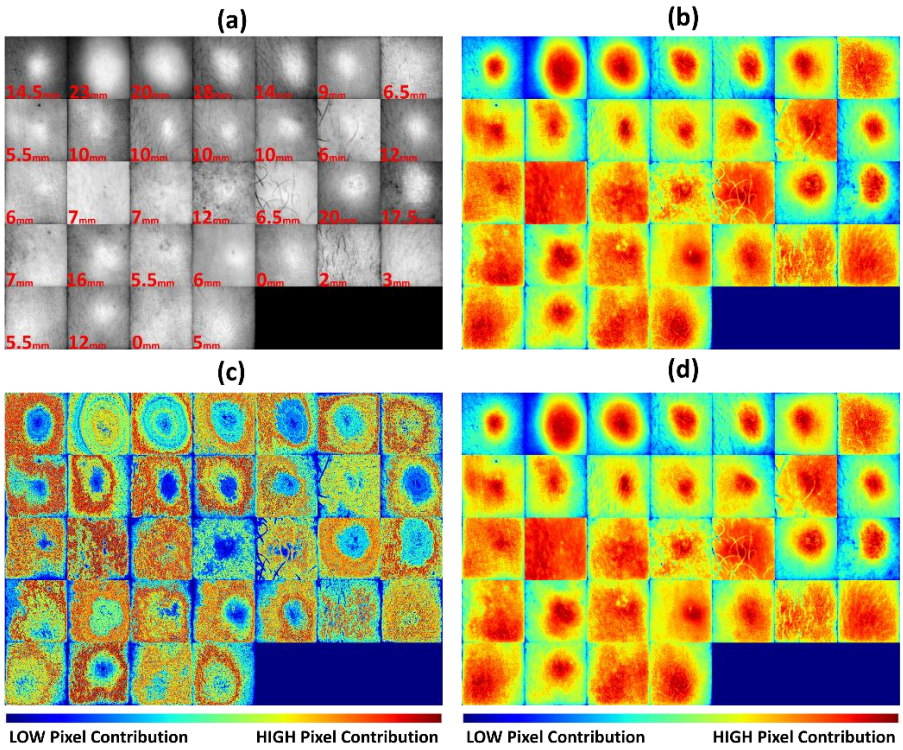


Figure E.5: Visualisation of best deterministic feature pair for the Vietnamese participants showing: (a) grayscale images of WT, (b) colourmap images of WT, (c) feature contribution map for the $high\ gray\ level\ run\ emphasis$ of WT, (d) feature contribution map for the $mean$ of WT.

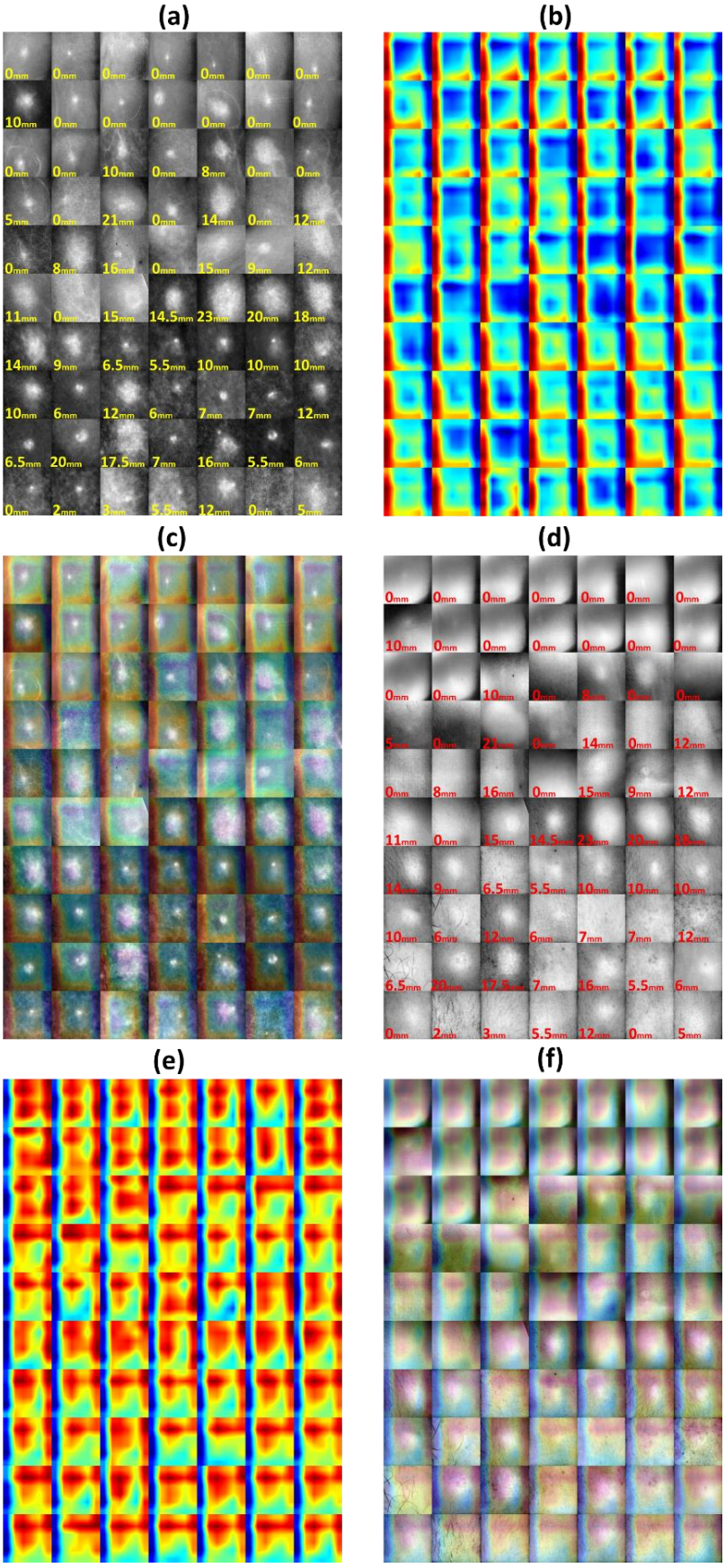


Figure E.6: Visualisation of best DenseNet-121 deep feature pair for the all participants showing: (a) grayscale images of FR, (b) heatmap of feature channel-523 for FR, (c) feature channel-523 overlaid on FR, (d) grayscale images of WT, (e) heatmap of feature channel-566 for WT, (f) feature channel-566 overlaid on WT.

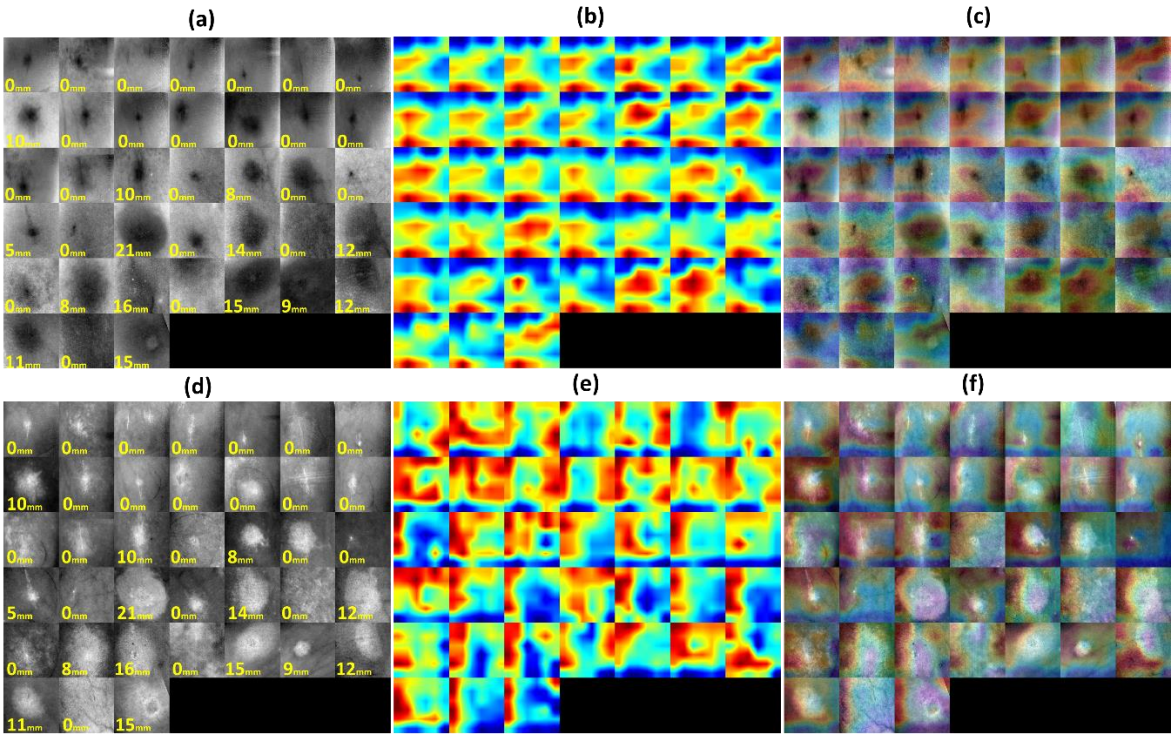


Figure E.7: Visualisation of best DenseNet-121 deep feature pair for only South African participants showing: (a) grayscale images of DX, (b) heatmap of feature channel-692 for DX, (c) feature channel-692 overlaid on DX, (d) grayscale images of OX, (e) heatmap of feature channel-188 for OX, (f) feature channel-188 overlaid on OX.

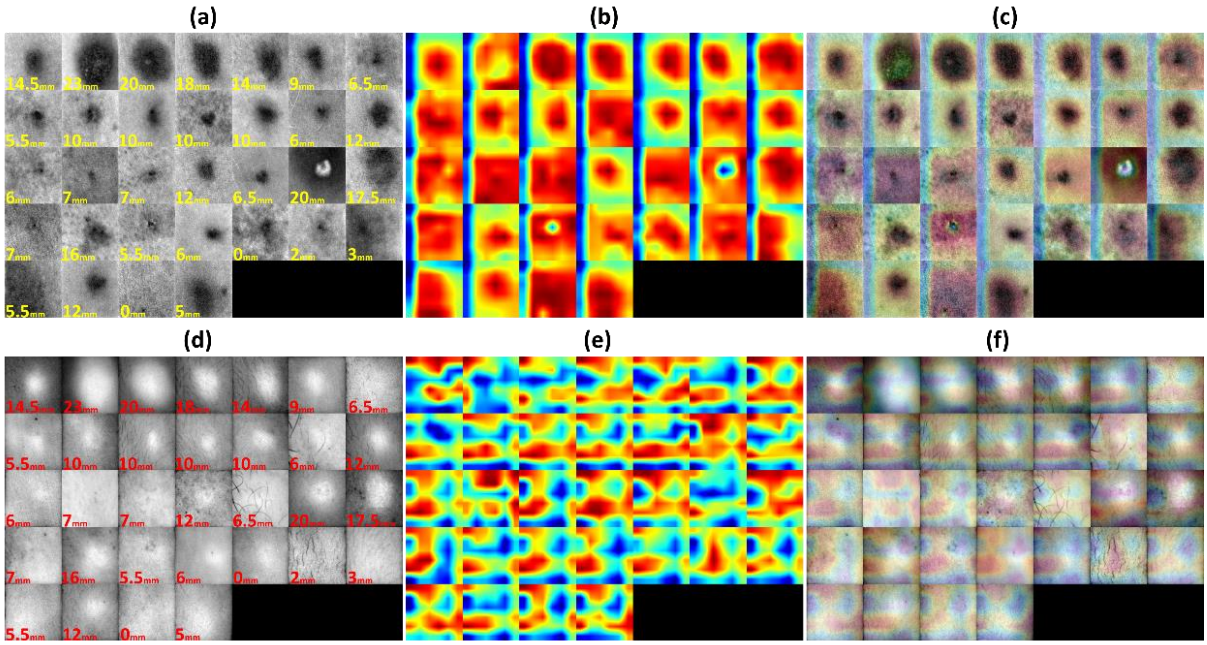


Figure E.8: Visualisation of best DenseNet-121 deep feature pair for only Vietnamese participants showing: (a) grayscale images of DX, (b) heatmap of feature channel-28 for DX, (c) feature channel-28 overlaid on DX, (d) grayscale images of WT, (e) heatmap of feature channel-698 for WT, (f) feature channel-698 overlaid on WT.

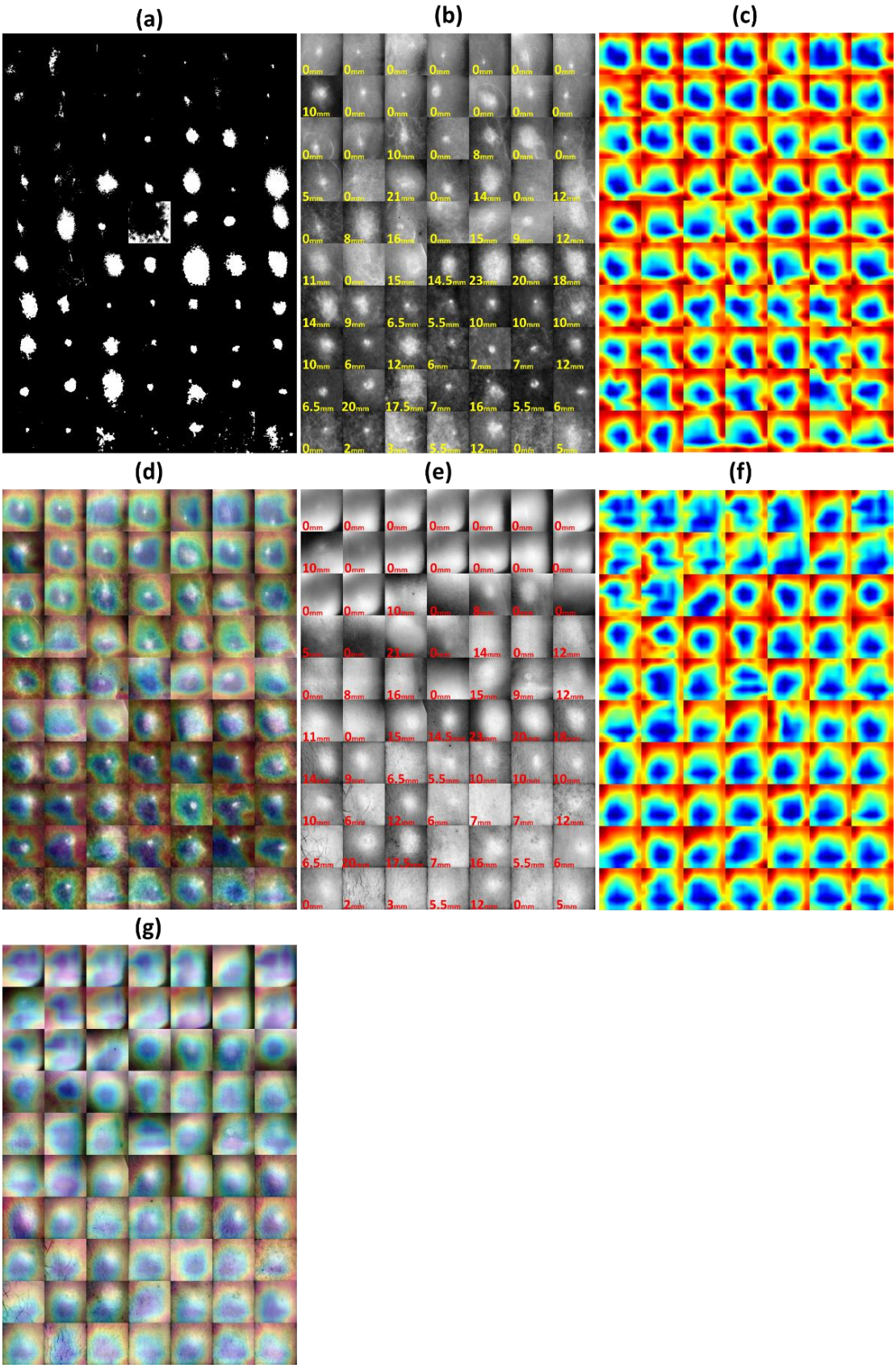


Figure E.9: Visualisation of best DenseNet-121 deep feature pair for the induration-masked chromophore maps of all participants showing: (a) induration masks, (b) grayscale images of FR, (c) heatmap of feature channel-966 for FR, (d) feature channel-966 overlaid on FR, (e) grayscale images of WT, (f) heatmap of feature channel-981 for WT, (g) feature channel-981 overlaid on WT.

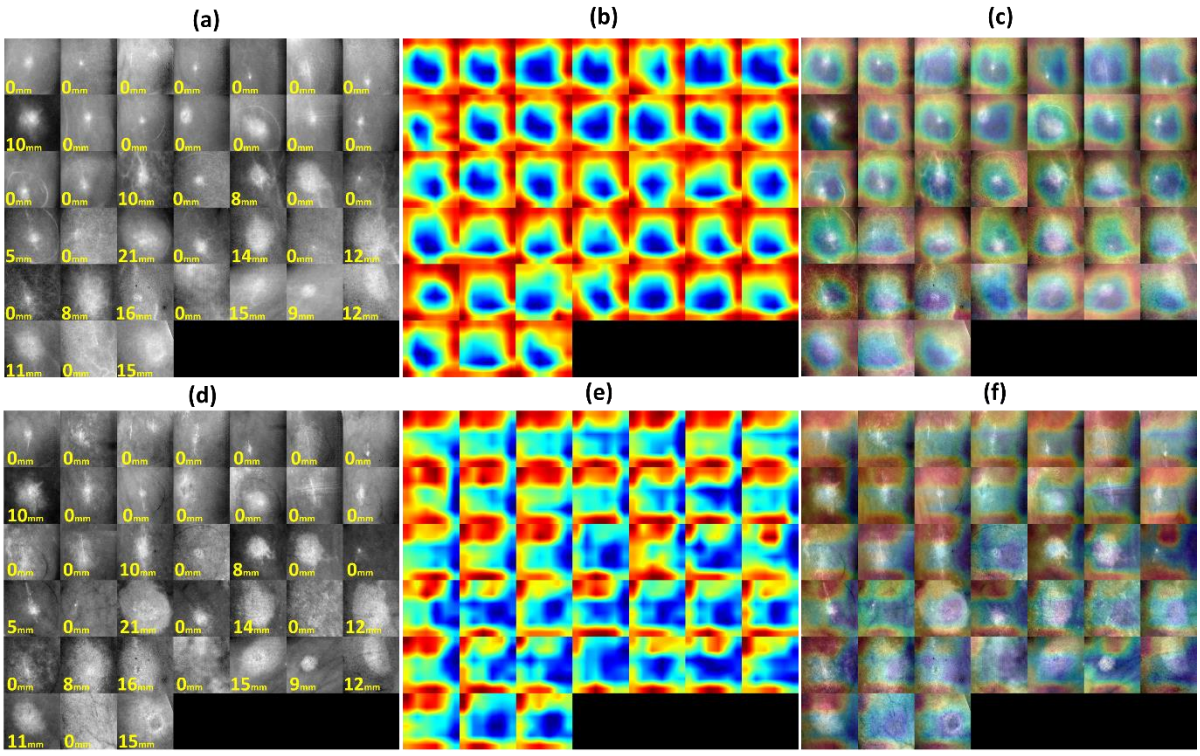


Figure E.10: Visualisation of the best DenseNet-121 deep feature pair for the induration-masked chromophore maps of only South African participants showing: (a) grayscale images of FR, (b) heatmap of feature channel-966 for FR, (c) feature channel-966 overlaid on FR, (d) grayscale images of OX, (e) heatmap of feature channel-862 for OX, (f) feature channel-862 overlaid on OX.

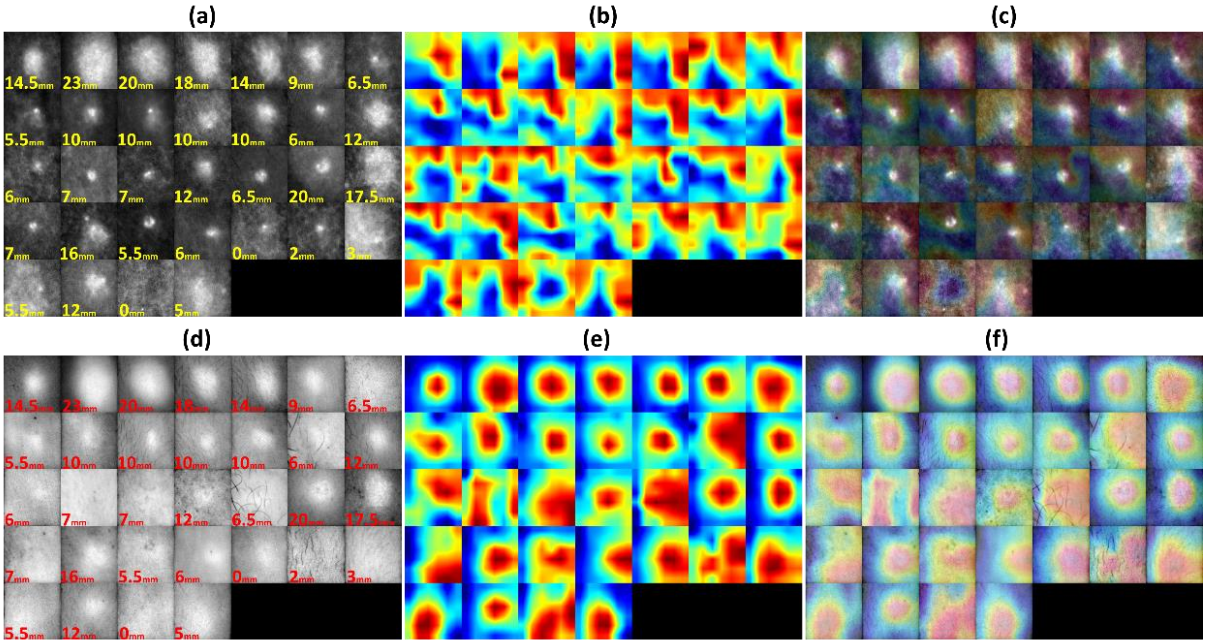


Figure E.11: Visualisation of the best DenseNet-121 deep feature pair for the induration-masked chromophore maps of only Vietnamese participants showing: (a) grayscale images of FR, (b) heatmap of feature channel-1020 for FR, (c) feature channel-1020 overlaid on FR, (d) grayscale images of WT, (e) heatmap of feature channel-249 for WT, (f) feature channel-249 overlaid on WT.