This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

KING'S College LONDON

Automated analysis in reflectance confocal microscopy images of skin anatomy and pathologies

Damato, Elaine

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 08. Oct. 2023

Division of Imaging Sciences and
Biomedical Engineering
School of Medicine

KING'S College LONDON

Automated analysis in reflectance confocal microscopy images of skin anatomy and pathologies

by Elaine DAMATO

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy of King's College London

May 2017

Supervisors:
Dr. Andrew COLEMAN
Dr. Graeme PENNEY

Examiners:
Dr. Catherine KENDALL
Dr. Robert SARKANY

Abstract

This thesis contributes to knowledge by developing algorithms that automatically detect and quantify structures of clinical interest in reflectance confocal microscopy (RCM) images, captured *in-vivo* and from excised skin tissue.

The first part of the thesis presents an algorithm that detects the dermal-epidermal junction (DEJ), characterised by papillae, in cubes of RCM images of *in-vivo* skin. A cube of images is a number of mosaic images captured at different depths parallel to the skin surface. A classifier, which makes use of texture and anatomical-based features was designed. The anatomical-based features are parameters that quantify the absence and presence of papillae across different images of the cube.

The second part of the thesis analyses RCM images of excised tissue collected during Mohs surgery. These tissue samples include basal cell carcinoma (BCC) and non-diseased samples. An algorithm was developed to differentiate between (i) cancerous regions, (ii) regions of inflammation, and (iii) non-diseased regions. A classifier based on texture and nuclei-concentration features was designed. The nuclei concentration in cancerous sites is different from that in non-diseased sites and thus can be used to distinguish the two.

The third part of the thesis analyses RCM video sequences of *in-vivo* skin imaged at the level of the DEJ. The boundaries of superficial skin capillaries can be delineated by visually observing the highly reflective red blood cells (RBCs) passing through the capillaries. An algorithm that automatically detects skin capillaries in RCM video sequences was developed. Additionally, an algorithm that quantifies the velocity of RBCs in cross-sectionally imaged capillaries is devised. The change in total capillary area (per unit frame area), individual capillary parameters and RBC velocity due to incremental ultra-violet radiation (UVR) doses are analysed in both fair and dark skinned volunteers.

The work presented in this thesis has the potential to increase the acceptance of RCM in the dermatology clinic, both for diagnosis and for assessing treatment response of skin conditions located at (or above) the DEJ. Additionally, the thesis enhances the potential of using RCM images of excised samples instead of preparing the tissue for histological examinations during surgery.

Acknowledgements

I would like to thank my supervisors, Andrew Coleman and Graeme Penney, for their support and guidance throughout my studies. Andrew was the mind behind this PhD and he has introduced the clinical VivaScope® RCM systems to St. Thomas' Hospital. Without his vision this PhD would not have been possible. In addition, he has provided me with valuable input and suggestions when I was at a cross-road in my projects. Graeme has provided me with helpful feedback with respect to designing and implementing my algorithms, and helped me shape the work I have done into a thesis.

I would like to thank Emma Craythorne who has given me the opportunity and training required to collect RCM images from St. John's Institute of Dermatology. Her warm nature made the transition to the clinical setting easier.

I am thankful of the in-depth discussions I had with Damilola Fajuyigbe and Anthony Young. Additionally, I have spent many hours collecting data with Damilola. Her presence and positive-outlook made the process more pleasant.

A special thanks goes to my close friends I have met at KCL: Yolanda, Sara, Kirsten, Alexis and Andreas. Their friendship made days at the office more enjoyable and lunch breaks full of discussions.

Throughout this journey I had the opportunity of understanding the true importance of family. My mother and late father have provided me with a beautiful childhood, where I could question and wonder. They taught me that the world is my oyster from my early years. Without this approach to life, this journey would not have been possible. I am thankful of my brother and Melanie, who provide continuous encouragement and kindness. I would like to thank my in-laws for believing in us and in each project we set our minds on. This thesis is dedicated to my husband who brightens my life. I'd like to thank him for his endless love and support.

Contents

1 Introduction **17**

1.1 RCM in the dermatology clinic 17

1.2 Summary of the Thesis 22

1.3 Original Contributions 24

2 Clinical Background and Skin Imaging **26**

2.1 Introduction 27

2.2 The skin 28

2.2.1 Skin layers 28

2.2.2 Micro-circulation and angiogenesis 30

2.2.3 UV therapy to induce erythema 32

2.2.4 Biological measures of RBCs and skin capillaries 33

2.3 Clinical diagnosis 35

2.3.1 Visual clinical examination 35

2.3.2 Dermoscopy 36

2.3.3 Histology 37

2.4 Non-invasive methods of analysing the skin 38

2.4.1 Optical imaging modalities for imaging the skin 38

2.4.1.1 Reflectance Confocal Microscopy 39

2.4.1.2 Optical Coherence Tomography 41

2.4.1.3 Multi-photon Imaging 42

2.4.1.4 Multi-spectral Imaging 42

2.4.1.5 Raman spectroscopy 43

2.4.1.6 Laser Speckle 44

2.4.1.7 Capillaroscopy 44

2.4.1.8 Laser Doppler Flowmetry and Imaging 45

2.4.2 Optical instrumentation for quantifying colour 45
2.4.3 Acoustic 47
2.4.4 Thermal 47
2.5 Conclusion 48

3 Literature Review **49**
3.1 Introduction 50
3.2 RCM for imaging *in-vivo* skin 50
3.2.1 Detection of skin sub-layers and DEJ 53
3.2.1.1 Kurugol et al. [1] [2] [3] 53
3.2.1.2 Somoza et al. [4] 56
3.2.1.3 Hames et al. [5] 58
3.2.2 Artefact removal 59
3.2.3 Detection of cancerous regions and segmentation of individual cells 61
3.2.3.1 Wiltgen et al. [6] [7] and Koller et al. [8] 61
3.2.3.2 Segmentation of cells 62
3.2.4 Discussion 62
3.3 RCM for imaging excised tissue 63
3.3.1 Introduction 63
3.3.2 Mohs surgery 64
3.3.3 Feasibility studies of using RCM in Mohs surgery 64
3.3.4 Tissue preparation for imaging in RCM fluorescence mode . 66
3.3.5 Imaging BCC in RCM fluorescence mode 66
3.3.6 Algorithms to analyse images of excised tissue 67
3.3.6.1 Dobbs et al. [9] 68
3.3.6.2 Legesse et al. [10] 69
3.3.7 Discussion 70
3.4 RCM video sequences for visualising skin capillaries 71
3.4.1 Introduction 71
3.4.2 Methods under which skin capillaries were analysed 71
3.4.3 Quantification of skin capillaries 72
3.4.4 Discussion 75

3.4.4.1 Protocol used to collect data 75
3.4.4.2 Techniques used to validate RCM video sequences 76
3.4.4.3 Motion in RCM video sequences 76
3.4.4.4 Repeatability of results and selecting ROI 77
3.4.4.5 VivaScope® 1500 frame rate and RBC velocity/flow 77
3.4.5 Discussion 78
3.5 Conclusion 79

4 Reflectance Confocal Microscopy 80
4.1 Introduction 80
4.2 Physics behind the RCM 81
4.3 Image acquisition from VivaScope® 82
4.4 VivaScope® Systems Safety Standards 86
4.5 VivaScope® 1500 88
4.5.1 RCM images of normal *in-vivo* skin 88
4.5.2 Artefacts in RCM images of *in-vivo* skin 91
4.5.2.1 Wrinkles and hair shaft artefacts 91
4.5.2.2 Artefacts due to unevenness in the skin surface . . 92
4.5.2.3 Loss of coupling artefacts 93
4.5.2.4 Poor technique artefacts 94
4.5.3 RCM videos of *in-vivo* skin 95
4.5.4 Artefacts of RCM video sequences 98
4.5.5 Power Alteration in RCM video sequences 98
4.6 VivaScope® 2500 99
4.6.1 Mohs surgery and RCM images of excised tissue 99
4.6.2 Artefacts in RCM images of excised tissue 101
4.6.2.1 Concentration of fluorescent dye 101
4.6.2.2 Epidermis visibility 102
4.6.2.3 Low signal regions 103
4.6.3 Comparing histology to RCM images 103
4.7 Conclusion 105

5 Automatic localisation of the dermal-epidermal junction in reflectance confocal microscopy image cubes 107

5.1 Introduction . . . 108
5.2 Algorithmic Method . . . 109
5.2.1 Introduction . . . 109
5.2.2 Layer classifier . . . 111
5.2.2.1 Texture based features . . . 112
5.2.2.2 Anatomical based features . . . 115
5.2.2.3 Feature Selection . . . 121
5.2.2.4 Classification . . . 122
5.2.3 *Find DEJ region* algorithm . . . 123
5.2.3.1 2D B-spline . . . 123
5.3 Experimental Method . . . 125
5.3.1 Data Acquisition . . . 127
5.3.2 Preparing the data for layer classification . . . 128
5.3.3 Selection of monogenic filter parameters . . . 131
5.3.4 Evaluating the layer classifier . . . 134
5.3.5 Evaluating the *find DEJ region* algorithm . . . 135
5.4 Results . . . 136
5.4.1 Intra-volunteer Layer Classifier . . . 137
5.4.2 Inter-volunteer Layer Classifier . . . 138
5.4.3 2D B-spline representation of results . . . 140
5.5 Discussion . . . 143
5.6 Future Work . . . 145
5.6.1 Quantification of papillae population and melanin . . . 146
5.6.2 Enhancing features . . . 147
5.6.3 Suppressing confounding structures . . . 148
5.6.4 Purer attributes and multi-labels . . . 148
5.7 Conclusion . . . 149

6 Identification of normal and pathological tissue in reflectance confocal microscopy images of excised tissue 151

6.1 Introduction . . . 152

6.2 Background . . . 154
6.3 Algorithmic Method . . . 158
6.3.1 Segmentation algorithm . . . 159
6.3.2 Selecting regions of interest (ROIs) . . . 161
6.3.3 Automatically segmenting a whole image into regions . . . 163
6.3.4 Feature Extraction: Nuclei concentration . . . 165
6.3.4.1 Detecting individual nuclei in RCM images . . . 165
6.3.4.2 Nuclei concentration features extracted from detected nuclei . . . 166
6.3.5 Feature Selection . . . 167
6.3.6 Classification . . . 167
6.4 Experimental Methods . . . 168
6.4.1 Data . . . 168
6.4.2 Labelling data . . . 169
6.4.3 Evaluating the classifier . . . 169
6.4.4 Experiment A: Classifier based on texture and nuclei concentration features . . . 170
6.4.5 Experiment B: Classifier based on texture features . . . 171
6.4.6 Experiment C: Classifier tested on whole images . . . 171
6.5 Results . . . 172
6.5.1 Results A: Classifier based on texture and nuclei concentration features . . . 172
6.5.1.1 Nuclei detected with the MSER algorithm . . . 172
6.5.1.2 Classification Results . . . 174
6.5.2 Results B: Classifier based on texture features . . . 175
6.5.3 Results C: Classifier tested on whole images . . . 177
6.6 Discussion . . . 181
6.6.1 Main observations from Experiment A . . . 181
6.6.2 Main observations from Experiment B . . . 184
6.6.3 Main observations from Experiment C . . . 184
6.6.4 Comparison with the literature . . . 185
6.6.5 Clinical context of results . . . 186
6.7 Future Work . . . 187

6.8 Conclusion 188

7 Automated detection of skin capillaries in reflectance confocal microscopy video sequences 190

7.1 Introduction 191
7.2 Background 194
7.2.1 Appearance of RBCs in RCM 194
7.2.2 Artifacts in RCM video sequences 196
7.2.2.1 Motion in RCM video sequences 196
7.2.2.2 Loss-of-coupling artefacts in RCM images 197
7.2.2.3 Laser power 197
7.3 Algorithmic Method 198
7.3.1 Introduction 198
7.3.2 Design of algorithm to detect skin capillary area 200
7.3.2.1 (a) Motion stabilisation and Denoising 200
7.3.2.2 (b) Windowed standard deviation volume 204
7.3.2.3 (c) Circle finder algorithm 206
7.3.2.4 (d) Connect areas 207
7.3.2.5 (e) Thresholding small regions 208
7.3.3 Calculation of individual capillary parameters 209
7.3.4 Design of algorithm to quantify RBC velocity across cross-sectionally imaged capillaries 210
7.4 Experimental Method 213
7.4.1 Data 215
7.4.1.1 UVR irradiation 215
7.4.1.2 Collecting confocal video sequences 217
7.4.2 Validation 220
7.4.3 Total capillary area vs UVR irradiated dose 220
7.4.4 Total capillary area vs a^* value 221
7.4.5 Calculating parameters from individual capillaries 222
7.4.6 RBCs velocity in cross-sectionally imaged capillaries 223
7.5 Results 224
7.5.1 Total capillary area vs UVR irradiated dose 224

7.5.2 Total capillary area vs a^* value . . . 230
7.5.3 Calculating parameters from individual capillaries . . . 232
7.5.4 RBCs velocity in cross-sectionally imaged capillaries . . . 237
7.6 Discussion . . . 240
7.6.1 Introduction . . . 240
7.6.2 Discussion of main findings of algorithm to detect skin capillary area . . . 241
7.6.3 Discussion of main calculations of individual capillary parameters . . . 242
7.6.4 Discussion of main findings of algorithm to quantify RBC velocity across cross-sectionally imaged capillaries . . . 244
7.6.5 Limitations in algorithm and methodology . . . 245
7.6.5.1 Intra- and inter-volunteer variability . . . 245
7.6.5.2 Confounding features in data . . . 245
7.6.5.3 Limitations in data collection . . . 249
7.6.5.4 Limitations in the development of the algorithm . 249
7.7 Future Work . . . 250
7.8 Conclusion . . . 251

8 Conclusion **253**
8.1 Summary and Context of the Contributions . . . 253
8.1.1 Detecting the dermal-epidermal junction in RCM cubes of images . . . 253
8.1.2 Detecting cancerous regions in RCM images of excised tissue 255
8.1.3 Detecting skin capillaries in RCM video sequences . . . 256
8.2 Future Work . . . 258
8.2.1 Enhancing the RBC signal . . . 258
8.2.2 Repeatability of RBC velocity measures . . . 259
8.2.3 Apply algorithms to diseased datasets . . . 260
8.3 Summary . . . 260

List of Figures

2.1 Skin schematic showing layers 30
2.2 Skin schematic showing layers, capillaries and vessels 31
2.3 H & E stained skin tissue 37

4.1 Reflectance confocal microscope technology 82
4.2 RCM for *in-vivo* and *ex-vivo* studies 83
4.3 Image acquisition with the VivaScope® RCM 83
4.4 Dermoscopic images of regions scanned by VivaScope® 1500 . . . 85
4.5 RCM images of the epidermal skin layers 89
4.6 RCM images of the DEJ and dermal layers 90
4.7 RCM images of wrinkle regions 92
4.8 RCM images of hair shafts 92
4.9 RCM imaging above the skin surface 93
4.10 RCM images of an uneven skin surface 94
4.11 RCM artefact images 94
4.12 Capillary levels that can be imaged with RCM 96
4.13 RCM image sequences of capillaries 97
4.14 A typical RCM scan of capillaries 97
4.15 Frames from RCM video sequence showing movement 98
4.16 Adjusting the RCM power for skin type VI 99
4.17 A frozen representation of excised tissue 101
4.18 Preparing excised tissue to be RCM imaged 101
4.19 Images of when the fluorescent dye is unevenly concentrated . . . 102
4.20 Flattening out the excised tissue for better quality images 103
4.21 Images with and without flattening out the excised tissue 103
4.22 RCM images showing missing tissue 104

4.23 Infiltrative BCC histology and RCM images . . . 104
4.24 Infiltrative BCC histology and RCM images . . . 105
4.25 Nodular BCC histology and RCM images . . . 105
4.26 Superficial BCC histology and RCM images . . . 106

5.1 Pipeline of steps to perform and evaluate layer classification . . . 110
5.2 Approximations and details of wavelet analysis . . . 116
5.3 Three level wavelet decomposition tree . . . 116
5.4 Hough circular detection pipeline . . . 117
5.5 Applying monogenic filter to DEJ image to enhance edges . . . 118
5.6 Applying the Hough circular transform to the edged image . . . 119
5.7 Detecting inner and outer radii of papillae . . . 120
5.8 Final detected circles using Hough circular transform . . . 120
5.9 2D B-spline grid . . . 125
5.10 X-Y representation of 2D B-spline output . . . 126
5.11 Y-Z representation of 2D B-spline output . . . 126
5.12 Intra-volunteer variability . . . 129
5.13 Inter-volunteer variability . . . 130
5.14 The effect of λ_{min} on the monogenic filter . . . 132
5.15 The effect of σ on the monogenic filter . . . 133
5.16 The monogenic filter applied to DEJ RCM images . . . 133
5.17 The monogenic filter applied to RCM images of all classes . . . 134
5.18 Layer classification height and thickness errors . . . 136
5.19 2D B-spline representation of a good case . . . 141
5.20 2D B-spline representation of an average case . . . 141
5.21 2D B-spline representation of a bad case . . . 142

6.1 An RCM image of BCC excised tissue . . . 155
6.2 RCM images of excised skin tissue showing glands . . . 156
6.3 RCM images of excised tissue of epidermis, muscle and fat . . . 157
6.4 RCM images of excised skin tissue showing tumour and inflammation . . . 157
6.5 Pipeline of steps to train the classifier on images of excised tissue . . . 158
6.6 Pipeline of steps to test the classifier on images of excised tissue . . . 159

6.7 Two half-discs used by the segmentation algorithm 160
6.8 Graphical user interface window 162
6.9 Segmentation threshold 163
6.10 Region growing algorithm 164
6.11 Generated mask 165
6.12 The output of the MSER algorithm 166
6.13 Nuclei concentration based features 168
6.14 Nuclei detected in RCM images of glands 173
6.15 Nuclei detected in RCM images of epidermis, muscle and fat 173
6.16 Nuclei detected in RCM images of tumour and inflammation 174
6.17 Good example of results from Experiment C 179
6.18 An average example of results from Experiment C 180
6.19 A poor example of results from Experiment C 182

7.1 RCM video sequence of capillary imaged cross-sectionally 195
7.2 RCM video sequence of capillary imaged along its length 196
7.3 Algorithmic Method Pipeline 201
7.4 Registration with a static reference image 202
7.5 Registration with a moving reference registration 202
7.6 Grids of control points used for the three-level resolution 203
7.7 Generation of a common registration mask for a video 205
7.8 Calculating the standard deviation volume 207
7.9 Intensity values inside and outside the capillary area 208
7.10 2D labelled map 210
7.11 Distance travelled between arterial and venous limb in capillaries . 211
7.12 Lag between arterial and venous limb signals 213
7.13 Cross-correlation lag calculation example 214
7.14 UVR irradiated inner arm of a fair skinned volunteer 216
7.15 UVR irradiated inner arm of a dark skinned volunteer 217
7.16 Minimal Erythema Dose for fair and dark skin 218
7.17 Dermoscopy images of sites exposed at 1 SED and 8 SED 219
7.18 Dermoscopy images of UVR exposed sites showing video locations 219
7.19 Fair skin visual results 226

7.20 Dark skin visual results 227
7.21 Fair skin total area detected vs dose (SED) 229
7.22 Fair skin total area detected vs dose (MED) 230
7.23 Dark skin total area detected vs dose (SED) 230
7.24 Dark skin total area detected vs dose (MED) 231
7.33 Arterial and venous intensity values for doses 0-11.3 SED 239
7.34 Arterial and venous intensity values 240
7.35 Lag for an RBC to travel from the arterial to venous limb 241
7.36 Reported increase in capillary diameter results vs literature values 243
7.37 Artefacts mistaken to be blood flow in a fair skinned volunteer . . . 247
7.38 Artefacts mistaken to blood flow in a dark skinned volunteer 248

List of Tables

2.1 Skin phototypes and description 29
2.2 Velocity of red blood cells 31
2.3 Non-invasive technologies to image skin 39
2.4 ITA°skin classification 46

3.1 Summary of RCM studies analysing *in-vivo* skin 52
3.2 Summary of RCM studies analysing skin capillaries 74

4.1 VivaScope® 1500 and 2500 specifications 85

5.1 Texture features used in this Chapter to classify RCM images based on Wiltgen et al. [6] 113
5.2 Collected datasets 127
5.3 Leave-one-out intra-volunteer layer classifier 131
5.4 Leave-one-out inter-volunteer layer classifier 131
5.5 Intra-volunteer layer classifier results 138
5.6 Intra-volunteer height and thickness errors 138
5.7 Inter-volunteer layer classifier 139
5.8 Inter-volunteer height and thickness errors 140

6.1 TP, PPV, NPV, sensitivity and specificity percentages of each class . 175
6.2 Percentage of correct and incorrect classified regions 175
6.3 Classifier's probability value 176
6.4 TP, PPV, NPV sensitivity and specificity percentages of each class . 176
6.5 Percentage of correct and incorrect classified regions 176
6.6 Classifier's probability values 177
6.7 TP, PPV, NPV sensitivity and specificity percentages of each class . 177

7.1 Volunteers irradiated with UVR dose. 217
7.2 Spearman's correlation between individual capillary area and SED 237
7.3 RBC velocity through cross-sectionally imaged capillaries 238

List of Abbreviations

2D	Two dimensional
3D	Three dimensional
BCC	Basal Cell Carcinoma
CARS	Coherent Anti-Strokes Raman Scattering
CFS	Correlation Feature Selection
CIE	Comission Internationale de l'Eclairage (International Commission on Illumination)
DEJ	Dermal-Epidermal Junction
DNA	Deoxyribonucleic Acid
DOCT	Doppler Optical Coherence Tomography
GUI	Graphical User Interface
HFUS	High Frequency Ultrasound
ITA	Individual Typology Angle
LDF	Laser Doppler Flowmetry
LDI	Laser Doppler Imaging
LM	Lentigo Maligna
MED	Minimal Erythema Dose
MSER	Maximally Stable Extremal Regions
NHS	National Health Service
NICE	National Institute for Clinical Excellence
OCT	Optical Coherence Tomography
PWS	Port Wine Stain
RBC	Red Blood Cell
RCM	Reflectance Confocal Microscopy
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SED	Standard Erythema Dose
SSM	Superficial Spreading Melanoma
TP	True Positive
UVR	Ultra-Violet Radiation
WEKA	Waikato Environment for Knowledge Analysis

Chapter 1

Introduction

Contents

1.1 RCM in the dermatology clinic 17

1.2 Summary of the Thesis 22

1.3 Original Contributions 24

The work presented in this thesis is focused on developing automated algorithms that detect structures and quantify parameters of clinical interest from skin images captured with reflectance confocal microscopy (RCM). This thesis is presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy of King's College London. This Chapter provides a brief summary of the motivations of this study, an outline of the thesis and the contributions towards science.

1.1 RCM in the dermatology clinic

Clinical dermatology has not traditionally relied upon imaging technologies. Expert visual examination of skin is capable of providing a high level of diagnostic accuracy and the skin is relatively easy to biopsy to gain histological confirmation of a diagnosis. Histology of biopsied tissue is considered the gold standard for diagnosis in dermatology [11]. The main limitations of histology are that it is invasive and that it destroys the site of interest. However, there is growing interest in the capability of new non-invasive imaging systems that can provide high

resolution images at depth in skin whilst keeping the site of interest intact [11]. These have the potential to reduce reliance on biopsy in some cases and to reduce the burden of biopsies that return negative, which are currently estimated to account for around 80% of examinations [12]. For a non-invasive modality to replace histology for a particular skin disease or condition, the non-invasive modality needs to achieve a sensitivity greater or equal than that achieved with histology and a similar level of specificity. For example, currently a sensitivity and specificity greater than 90% is achieved when analysing melanoma with histology [13]. Furthermore, a clinic needs to analyse whether the economic benefits associated with the non-invasive modality outweigh the costs before introducing the modality in the clinic.

One of the most promising of these technologies is reflectance confocal microscopy (RCM), which provides *optical sectioning* of tissue [14]. The contrast emerges from the refractive index differences between different skin structures. As a light beam propagates through the skin, the light is reflected at the boundaries of two structures with different refractive indices. The reflected light is then focused into a small pinhole and forms an image [15].

RCM is capable of visualising individual cells and organelles up to a depth of $200 - 250\mu m$, at a lateral resolution of $\sim 0.5 - 1\mu m$ and at an axial resolution of $\sim 3 - 5\mu m$ [16]. RCM has the potential of being a non-invasive diagnostic method as it has a lateral resolution comparable to that achieved by histology. A significant difference between RCM images and histology sections is the orientation; confocal images are captured in a horizontal (XY) plane parallel to the skin surface whereas histological sections are obtained from vertical (XZ) sectioning of biopsy samples [14]. This difference necessitates some retraining of staff to interpret the unfamiliar appearance of the anatomy of skin.

Unlike histology which destroys the site of interest, RCM can image the skin non-invasively and has the potential to allow the study of the skin in its native state [17]. This gives RCM an advantage that it can be used to monitor the progression of disease or treatment, where the imaging can be repeated as often as

required [11] [17]. RCM images also provide a useful addition to clinical screening and histology by helping a dermatologist make decisions such as whether, where, or when to biopsy a lesion [17]. Thus, RCM has the potential to reduce the need for invasive biopsies by facilitating non-invasive examinations of the skin [18]. In addition, RCM is also capable of non-invasively delineating tumour margins before surgery or non-surgical treatment [16].

Similarly, RCM can be used to analyse excised tissue and thus outline the tumour margin in Mohs surgery [16]. RCM images of excised tissue are produced in the same orientation as the histology slides; vertical slices perpendicular to the skin surface. As in histology, when imaging excised tissue with RCM in fluorescent mode, fluorescence dye can be absorbed to improve the visibility of particular structures. For example, acridine orange is used to colour cell nuclei and it results in a contrast between cell nuclei and cytoplasm. Imaging excised tissue with RCM, instead of analysing the tissue with histology, could eliminate the need of frozen sectioning [16]. This results in only needing ~ 10 minutes to stain and image an excised tissue, which is a fraction of the time needed ($\sim 30 - 45$ minutes) for a histologist to prepare a histology slide. RCM opens up new possibilities to reduce reliance on biopsies [16]. Thus, RCM has the potential of being used as a clinical technology that can provide pathological examinations of excised tissue [19].

RCM was developed by Marvin Minsky in 1955 [20]. The modality only gained popularity in the 1990s when technological advancements enabled the downsizing of the microscope from a bulky machine filling a room, to a mobile handheld device [21] [22]. The leading commercial RCM system, the VivaScope®, is produced by Lucid Inc. (USA) and marketed in Europe by Mavig GmbH (Germany). These commercial RCM systems have been available since 2004 when they were used in only a handful of key research centres in the USA, Germany and Italy. However, by 2015 the total number of VivaScope® RCM systems has increased to more than 600 and are located in dermatology clinics and hospital departments worldwide. Currently there are around 300 systems located in Europe, Middle East and Asia (EMEA), where there have been around 30 new

systems installed per annum since 2014 (Raphaela Kaestle, Mavig GmbH, personal communication, 2nd June 2016). Two VivaScope® systems are located at Guy's and St. Thomas' Hospital, London; the VivaScope® 1500 for *in-vivo* examinations and the VivaScope® 2500 used for analysing tissue samples.

The application of the RCM technology is to guide diagnosis and to aid in treatment planning. The main limitation is that currently the microscopes are only available in a minority of dermatology clinics and technology requires a learning curve to operate it and to interpret the images obtained [16]. Although there are strong similarities between RCM images and histology, inexperienced dermatologists still require a considerable amount of training to gain diagnostic confidence in RCM imaging. It is suggested that for a dermatologist to be become proficient in interpreting RCM images he/she needs to attend a one week training course organised by Mavig GmbH followed by online training which includes the reader to review 750 to 1000 clinical cases over a three month period (Andre Philipp, Mavig GmbH, personal communication, 5th April 2017). When readers undertake this training and have the same level of experience with RCM images, the inter-observer differences found across different readers is not significant. For example in a study undertaken by Rao et al. [23] two readers achieved a sensitivity and specificity above 90% and 60% respectively when compared against the histological representation of the specimen. However, a discrepancy of around 15% in specificity was present amongst these two readers and this was attributed to one reader having over than 9 years of experience in interpreting RCM images whereas the other reader had just over 1 year of experience. The most experienced reader achieved a sensitivity and specificity of 97.5% and 80.5% when comparing the results he achieved with RCM images against to those achieved with using the gold standard histological representation [23].

Various work has been carried out to make dermoscopic and histology skills more easily transferable to RCM, by emphasising the visual similarities between the two [24] [25] and even providing a digital staining tool for an RCM image to appear similar in colour to a histology slide [26]. However, a typical RCM examination generates 60-100 images [4] which require around 10 minutes to evaluate

in a clinical setting, and there is as yet no automated or semi-automated analysis that can be done routinely to help the clinician analyse this large dataset. As the user becomes more proficient in interpreting the images, the user can start analysing the images as they are being acquired, reducing the overall time needed to acquire and interpret the images.

The National Institute for Health and Care Excellence (NICE) suggest that RCM technologies are not yet recommended to decide whether to biopsy and excise skin lesions with suspected melanoma, basal cell carcinoma (BCC) or lentigo maligna (LM) [27]. NICE suggest that further research that displays sufficient evidence of the clinical effectiveness of defining tumour margins can boost the acceptance of RCM as a diagnostic tool and the technology being routinely adopted in the NHS [27].

This thesis focuses on developing algorithms that automatically analyse RCM images, with the aim of reducing the examination time and the level of training needed. For example, automated algorithms have the potential to highlight areas of particular concern, reduces the need for the dermatologist to view all images. Furthermore, automated algorithms give an objective analyses of RCM images and provide clinicians with quantitative measures that the clinician cannot extract when examining the images visually. The algorithms contribute towards a growing database of knowledge on RCM image interpretation, helping the clinician identify already recognized features of disease conditions using objective quantitative measures and of reducing the level of training needed.

Three algorithms have been designed and implemented in this thesis, these aim to provide:

- Detection of the dermal-epidermal junction based upon anatomical feature identification in RCM images of skin *in-vivo*;
- Detection of histological margins for surgery in RCM images of excised tissue;
- Detection and quantification of superficial skin capillaries in RCM video

sequences of skin *in-vivo*.

The methodologies adopted to develop such algorithms together with their contribution to knowledge is highlighted in subsequent Sections.

1.2 Summary of the Thesis

This thesis is structured into eight chapters, each of which is summarised below:

Chapter 2 gives detailed descriptions of the skin, including how its classified into layers, together with the function of skin capillaries, and how capillaries change in response to environmental insults. The importance of detecting and quantifying the dermal-epidermal junction, tumour regions and skin capillaries are identified. Additionally, an overview of the clinical practice of dermatology and of the non-invasive imaging modalities available to analyse the skin are given. RCM is compared to other non-invasive modalities and also to histology.

Chapter 3 provides a comprehensive literature review of the quantitative measurements which are currently being calculated from RCM images for dermatology and photobiology. Motivations and applications of devising semi- or fully-automated algorithms to extract quantitative measures from such images are identified. The literature review is subdivided into three parts: (i) RCM for imaging *in-vivo* skin (ii) RCM for imaging excised tissue and (iii) RCM video sequences for visualising skin capillaries. Thus, from this chapter the niche of where this project fits into prior literature will be discussed in depth.

Chapter 4 describes the RCM technology, together with the type of datasets that can be acquired from the commercial and clinical VivaScope® RCM systems used in this thesis. A distinction is made between the two RCM systems used in this thesis; VivaScope® 1500 and VivaScope® 2500 which are used to analyse *in-vivo* skin and excised tissue respectively. The protocol used to prepare the excised tissue for imaging in this study is given. Artefacts found in images captured using the two systems are identified and ways of mitigating such artefacts are discussed. Particular attention is given to artefacts as they can disrupt the

effectiveness of automated algorithms.

Chapter 5 devises a methodology for automatically detecting the dermal-epidermal junction (DEJ) in RCM images of *in-vivo* skin. The classification algorithmic method includes extracting texture and anatomical based features from RCM images. The anatomical based features characterise the presence and absence of papillae (which characterises the DEJ) across RCM images captured at different depths. The algorithm was tested on datasets captured from intra- and inter-volunteer sites. The Chapter concludes with a discussion of the results and possible future improvements.

Chapter 6 devises a methodology for differentiating between tumour, inflammatory, and non-diseased regions in RCM images of excised tissue acquired during Mohs surgery. The imaged samples contain basal cell carcinoma (BCC) and non-diseased tissue. The developed classification algorithm makes use of texture and nuclei concentration features to differentiate between the three mentioned classes. The results achieved with this algorithm are compared with the gold standard of frozen section histology as reported by the surgeon.

Chapter 7 describes an algorithm designed to automatically detect superficial skin capillaries located at the DEJ in RCM video sequences. The algorithm identifies the highly reflective red blood cells (RBCs) travelling though the capillary and uses these to delineate the capillary boundary. The velocity at which the RBCs travel inside the capillaries is calculated. The algorithm is tested on a dataset of video sequences captured at sites irradiated with incremental doses of ultra-violet radiation (UVR) in volunteers of both fair and dark skin types. Changes in vessel diameter and RBC velocity resulting from UVR induced erythema are reported and compared with photometric measurements of skin reddening.

Chapter 8 concludes the thesis with an overview, outlining the novel contributions towards knowledge achieved throughout the study. The clinical importance of this work is illustrated, and future research work, together with possible

improvements are identified.

1.3 Original Contributions

The original work described in this thesis can be summarised into three distinct contributions:

- The design and implementation of an automated algorithm to detect and quantify the dermal-epidermal junction (DEJ) thickness from *in-vivo* RCM cubes of images. The designed algorithm works automatically on standard clinical data; cubes of images with $8\mu m$ spacing. A total of 11 datasets, each with approximately 1400 tiles were used. The cube of images are representatives of clinical datasets, where each cube has a file size of around $3GB$ and can be stored either locally or on cloud storage. Despite clinics having sufficient storage facilities, clinicians' time is limited and an automated algorithm that speeds up the time needed for a clinician to analyse the images is required for the technology to gain wider clinical acceptance. No tiles or regions exhibiting artefacts were removed or masked out. To cater for the large number of tiles and clinically representative datasets, the designed algorithm extracts features from larger scale structures rather than individual cells. The DEJ thickness was located with an average error of $10.8\mu m$ and $20.1\mu m$ from the ground truth labels for intra- and inter-volunteer classifier models respectively. The ground truth labels were designated by myself after attending a one week in-person course and a three month online training course organised by Mavig GmbH. Each image was individually analysed and a label of epidermis, DEJ or dermis was assigned to it. DEJ images were characterised by edged or non-edge papillae, images above and below the DEJ were labelled as epidermis and dermis respectively.

  Impact of novelty: Such an algorithm could be used by dermatologists to reduce the number of images to be analysed before observing the DEJ, resulting in faster diagnoses.

- The design and implementation of an automated algorithm to differenti-

ate between cancerous, inflammatory and non-diseased regions in RCM images of excised tissue acquired during Mohs surgery. To improve algorithmic performance nuclei concentration features were automatically extracted. It was found that the introduction of nuclei concentration features increased the number of correctly classified regions classed as *tumour* and *inflammation* on average by 8.6% and 62.5% respectively.

Impact of novelty: This is the first automated algorithm designed to work specifically with RCM images of excised skin tissue. Although the results achieved with the designed algorithm are not sufficient to aid dermatologists in making a diagnosis, the work done in this thesis steps towards the acceptance of RCM being used to image excised tissue during surgery for diagnostic purposes.

- An automated algorithm has been designed to detect and quantify skin capillaries from RCM video sequences. The velocity at which red blood cells (RBCs) are travelling at inside cross-sectionally imaged capillaries is quantified. This is the first time an algorithm has been developed to automatically detect skin capillaries. Additionally, it is the first to quantify RBC velocity inside skin capillaries using RCM video sequences.

  Impact of novelty: This study points to a new use for RCM as a means of providing a more objective end-point than visually assessed erythema for photobiological experiments by measuring capillary area and RBC velocity in a area of tissue. A measure of capillary size and RBC velocity is beneficial to assess the response to phototherapy.

Chapter 2

Clinical Background and Skin Imaging

Contents

2.1 Introduction 27

2.2 The skin 28

2.2.1 Skin layers 28

2.2.2 Micro-circulation and angiogenesis 30

2.2.3 UV therapy to induce erythema 32

2.2.4 Biological measures of RBCs and skin capillaries 33

2.3 Clinical diagnosis 35

2.3.1 Visual clinical examination 35

2.3.2 Dermoscopy 36

2.3.3 Histology 37

2.4 Non-invasive methods of analysing the skin 38

2.4.1 Optical imaging modalities for imaging the skin 38

2.4.2 Optical instrumentation for quantifying colour 45

2.4.3 Acoustic 47

2.4.4 Thermal 47

2.5 Conclusion 48

2.1 Introduction

This chapter provides a brief overview of those aspects of skin anatomy that are used in the rest of the thesis. It also provides a review of the techniques that are available for analysing and imaging skin. The aim is to set in context the particular advantages and disadvantages of Reflectance Confocal Microscopy (RCM) as a tool for clinical imaging to aid the dermatologist in making a diagnosis and monitoring a treatment.

Currently when a dermatologist is faced with a suspicious mole they initially examine its physical shape and appearance with the naked eye, after which the clinician will look at it under dermoscopy. Where there is any doubt that the mole is malignant a biopsy is ordered to examine the histology. The histologist then provides a definitive diagnosis based on the appearance of cells and cell structures. The principal limitation of histology is that it is an invasive procedure.

There are various technologies that are used to image skin or attempt to quantify skin properties, such as blood flow. These modalities generate images or quantify parameters based on the optical, acoustic and thermal properties of the skin. Such technologies can only be used to examine a particular skin depth, and thus can only be used to analyse a subset of skin conditions and diseases. Furthermore, limitations in the resolution and image quality can limit the spectrum of conditions and diseases that can be analysed with the modality.

For a modality to be introduced in the clinical setting, the time needed to acquire and analyse the data need to be feasible to be completed during a patient's clinical visit. Additionally, the technology needs to have sufficient storage facilities that enable clinicians of viewing and comparing data acquired from past and current examinations to analyse the progression of the disease or condition. In cases where the modality examines ex-vivo skin, it is preferred if the sample preparations and dye staining are compatible to histology. Time needs to be invested by the clinic on deciding which technology works best for their spectrum

of patients. Moreover, the clinic needs to analyse whether the economic benefits outweigh the cost of the technology and of training clinicians in operating the technology and in interpreting the data obtained. Further training is needed to transfer knowledge across the different modalities and histology.

RCM can be used to diagnose benign and malignant skin lesions, such as melanoma, basal cell carcinoma and squamous cell carcinoma. Additionally, RCM can be used to monitor treatment such as assessing appropriate surgical margins ex-vivo in Mohs surgery and the effectiveness of imiquimod in actinic keratosis. Furthermore, RCM can be used to diagnose inflammatory conditions such as psoriasis and eczema [28]. In the mentioned skin conditions and diseases, the dermal-epidermal junction (DEJ) is an important feature to analyse in RCM when making an analysis. For example the DEJ flattens out and disarranged in early melanoma whereas malignant cells occupy the DEJ in more advanced melanoma. Additionally, the papillae (located at the DEJ) become non-rimmed in psoriasis [28].

2.2 The skin

2.2.1 Skin layers

The skin is typically classified into six phototypes defined in Table 2.1. The phototype number reflects the amount of melanin pigment present within the skin, and higher numbers reflect higher amounts of melanin. A lack of melanin, associated with phototypes I and II, results in the Reflectance Confocal Microscopy (RCM) images having less contrast.

Skin Phototype	Typical Features
I	Pale white skin, blue/hazel eyes, blonde/red hair
II	Fair skin, blue eyes
III	Darker white skin
IV	Light brown skin
V	Brown skin
VI	Dark brown or black skin

Table 2.1: Fitzpatrick Skin Phototype [29]

Skin tissue is typically described in terms of three main layers which are illustrated in Figure 2.1. The epidermis is shown by the purple coloured region; the dermis is shown by the pink coloured region and the DEJ is where the epidermis overlaps with the dermis. The main variation found between different anatomical locations is the epidermis thickness [30]. The epidermis thickness was estimated to be $56.6\mu m$, $70.3\mu m$ and $81.5\mu m$ at the dorsal forearm, shoulder and buttocks respectively [30]. When imaging the skin with RCM, sites with a thinner epidermis enable better visualisation of the different skin layers and thus better analysis of the dermis can be done, for example better visualisation of the different skin layers can be done at the forearm than at the buttocks.

The epidermis lies just below the skin surface, at a depth of approximately $0 - 100\mu m$. The epidermis can be sub-divided into four layers, where each layer occupies a distinct depth and has a distinct cell architecture. Keratinocytes are the dominant cell population found within the epidermis and are produced at the deepest epidermal layer (stratrum basalis) and they undergo maturation as they move upwards towards the skin surface [31]. The major function of the epidermis is to serve as a barrier to chemical penetration and microbiologic invasion from the environment [31]. The dermis is at a depth of roughly $100 - 150\mu m$ below the skin surface. The dermis provides structural and nutritional support [31].

The dermal-epidermal junction (DEJ) is the region between the epidermal and dermal layers, and this junction is categorised by upward finger-like projections, referred to as papillae, from the dermis into the epidermis layers. When visualis-

ing the skin with RCM, the DEJ shows both epidermal and dermal like structures. The DEJ is of clinical interest as a variety of diseases can originate from it [31]. These diseases typically then spread away from the DEJ into the epidermis and dermis [1]. A more detailed explanation of the skin layers and how each one is visualised in RCM is given in Chapter 4.

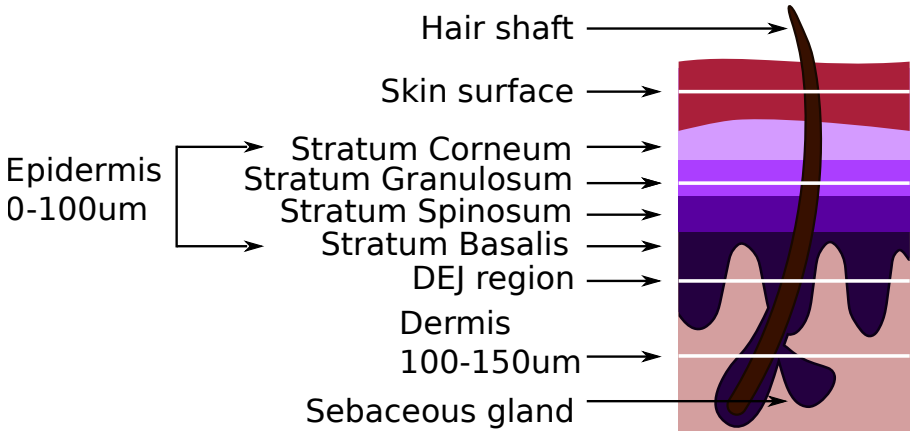


Figure 2.1: A skin schematic: the first group of layers of the skin below the skin surface is referred to as the epidermis which is located at a depth of approximately $0 - 100\mu m$. The epidermis layer is sub-divided into four: Stratum Corneum, Stratum Granulosum, Stratum Spinosum and Stratum Basalis. The deepest set of layers in the skin are referred to as the dermis and is located at a depth of roughly $100 - 150\mu m$. Intermediate between these two regions is the dermal-epidermal junction. Hair shafts emerge from the dermis layer to the skin surface whereas sebaceous glands are located at the dermis layer. The four white lines indicate RCM image acquisitions taken above the skin surface, within an epidermis layer, at the DEJ and at the dermis respectively.

2.2.2 Micro-circulation and angiogenesis

Microcirculation plays an important role in passing nutrients to the skin and regulating body temperature. Microcirculation takes place at different depths of the skin; arterial and venous capillaries at the DEJ are connected to arterioles and venous vessels in dermis [32]. A basic schematic of these structures is illustrated in Figure 2.2 and the variation in red blood cell velocity inside these locations are listed in Table 2.2 [33]. Each capillary sits within a particular dermal-epidermal papilla. From the muscles and fat at the deeper dermis, arteriole and venous vessels pass towards the mid dermis, and blood is supplied to hair bulbs and sweat glands [34].

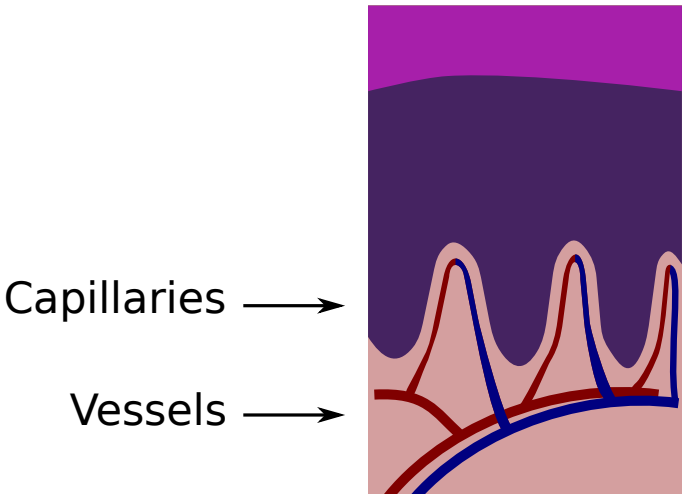


Figure 2.2: A depiction of the capillaries and vessels, which are located at the DEJ and dermal layers respectively. The arterial end of the capillary (red) is connected to the arteriole vessel while the venous end of the capillary (blue) is connected to the venous vessel.

The blood flow at superficial layers is regulated from the deeper dermis and is a mechanism of thermal regulation [34]. In response to environmental provocation, for example, cold weather, the blood capillaries within the papillary dermis and arrector pili muscles within the deep dermis contract (vasocontraction), which in turn causes the hair shaft to become erect. An erected hair shaft traps air between the hair and skin surface, creating a natural insulator to prevent heat loss. Conversely, in hot weather the blood capillaries dilate (vasodilation) and the arrector pili muscles relax, enabling blood to flow close to the skin surface, from which heat is lost. Changes in vessel diameter, corresponding to either vasodilation or vasoconstriction [35], is a valuable physiological end-point in many photobiological studies.

Location	**Velocity (m/s)**
Artery	0.45
Arteriole vessels	0.05
Capillary	0.001
Venous vessels	0.002
Vein	0.1

Table 2.2: Red blood cell velocity (m/s) inside arteries/veins, vessels and capillaries [33].

Imaging vessels can be clinically useful since changes in vasculature are often

a marker of disease. Angiogenesis is the process through which new blood vessels form from pre-existing vessels [36] and the remodelling of existing vessels through circumferential growth [31]. Angiogenesis is a characteristic feature of inflammatory skin disorders, such as blistering diseases, psoriasis and contact dermatitis [31]. Pronounced angiogenesis is associated with cancer growth including melanoma. Prominent blood vessels are also a characteristic of rosacea and basal cell carcinoma (BCC) [31]. Quantifying the number and size of capillaries and vessels may provide a means to identify the presence and/or extent of skin diseases and cancers. Angiogenesis is essential for tumour progression, in fact various anti-cancer treatments aim to inhibit it [31]. On the other hand, inflammatory skin diseases which are linked to a change in capillary population and size, such as psoriasis, are normally treated with UV therapy which will be discussed in more depth in 2.2.3.

2.2.3 UV therapy to induce erythema

In phototherapy the skin is exposed to ultra-violet radiation (UVR) to treat various skin conditions, including psoriasis, vitiligo, acne, cancer, and to aid wound healing. At present, individual sensitivity to UVR is determined by measuring the minimal erythema dose (MED); the lowest UVR dose necessary to provoke mild reddening. The MED is currently measured either visually in terms of skin redness or spectrophotometrically (Section 2.4.2) by quantitatively comparing the colour (redness) of skin at exposed skin sites. As outlined by Heckman et al. [37] MED testing is best performed $24 - 48$ hours after UVR exposure. Red or pink skin indicates erythema or burning, and erythema skin exposed to the shortest duration of UVR represents the MED site. When measuring skin redness using a spectrophotometer, place the aperture in the centre of the site being measured. The larger the a^* spectrophotometer value, the redder the site and the UVR dose it was exposed to [37]. In dark skinned subjects, colour estimates of erythema degree are notoriously unreliable [38]. In addition, skin redness is only an indirect measure of physiological changes related to changes in blood flow within the superficial layers of skin [34].

Various studies have been performed to define the changes experienced by the skin when exposed to UVR. An immediate effect of UVR exposure is an increase in blood flow, which is related to an increase in vessel area [39]. UVR reaching the skin can be divided into two categories: UVA (320-400 nm) and UVB (280-320 nm) [40]. UVB penetrates only the upper epidermal layers whilst UVA can penetrate to the deeper dermal layers.

As the skin is exposed to several UV exposures, the skin can experience delayed tanning and persistent pigment darkening [40]. In addition, there is a significant increase in melanin production by the dendritic melanocytes at the basal layer [41]. Melanosomes gradually migrate to the upper skin layers, closer to the skin surface, via these dendritic cells, resulting in an increase in epidermal thickness [42]; the thickening of the epidermal layer is a natural mechanism of sun protection as it is then harder for the UVA rays to penetrate to the dermal layers below and cause DNA damage [41].

This thesis will explore the capability of RCM to visualise skin capillaries and to measure some properties such as their diameter and the blood flow within them. Accurate identification of capillaries offers the potential to improve the understanding of skin physiology, skin disease and response to treatment [34]. By providing a direct measure of blood flow in the smallest and most superficial capillaries, RCM may be able to improve photobiological studies reliant on quantification of erythema-related end-points.

2.2.4 Biological measures of RBCs and skin capillaries

This section will be divided into a synopsis of relevant physiology knowledge on red blood cells (RBCs), skin capillaries and blood flow.

Blood can be divided into 55% plasma and 45% cells, of which around 40% are RBCs [43]. The fraction of RBC volume to the total blood volume is referred to as haematocrit. It is believed that the haematocrit can fall from $\sim$ 40% to $\sim 10-20\%$ in skin capillaries due to changes in capillary diameter [44]. As

red and white blood cells have a similar diameter and are both refractive, it is difficult to distinguish them in RCM [45]. However, as red blood cells are more abundant than white cells, at a ratio of 600-700 red to white blood cells, the blood flow visualised in RCM image sequences mainly corresponds to RBCs.

A red blood cell (RBC) has a disk diameter of approximately $6.2 - 8.2\mu m$ [46], a thickness at the thickest point of $2 - 2.5\mu m$ and a minimum thickness in the centre of $0.8 - 1\mu m$. These cells have an average volume of about $90\mu m^3$ [47] with a surface area of about $136\mu m^2$. The movement of RBCs is dependent on various parameters, such as: deformability, viscosity, hematocrit, blood pressure and vessel diameter [48].

Skin capillaries are located in the dermal-epidermal junction (DEJ) and each capillary consists of two parallel capillary limbs (arterial and venous limbs) with a spacing in between them. Each limb diameter can range between $10 - 35\mu m$ at the DEJ [49]. As the limbs are visualised deeper in the skin, they are referred to as vessels rather than capillaries, and their diameter increases to $40 - 50\mu m$ in mid to deep dermis [49] (Figure 2.2). Over the depth visualised by RCM the capillary is best defined up to a depth of $100\mu m$. One does not expect a large variation in the capillary diameter over this depth [50]; in fact a constant circumference has been recorded in different skin depths visible in RCM [51].

In healthy skin the capillary loops are arranged regularly and approximately vertically [52], with a density of between 9 and 14 (average 10) capillaries per mm^2 [53]. In addition to this, the apices (the points where the capillary limbs merge) are located at similar depths in healthy skin at approximately $40 - 60\mu m$ below the skin surface [52].

Monetal et al. [54] estimated that the cell's width occupies around 68% of the total capillary lumen diameter, and that the cell's width remains a similar proportion of the total capillary diameter for the RCM imaging depths [49]. Thus, the smallest superficial capillaries, located at the DEJ, are typically wide enough to allow only one RBC at a time. In the case of an increase in the haematocrit

(also known as packed cell volume), compact blood cell columns (chain of RBCs) may be present, making it difficult to distinguish individual cells with imaging systems of $\sim 1\mu m$ resolution [55].

2.3 Clinical diagnosis

This section sets out the current clinical practice used in making a diagnosis of a suspicious skin lesion which can be subdivided into three sections: (i) visual clinical examination, followed by a (ii) dermoscopic view of the lesion, and if the lesion looks suspicious (iii) a biopsy with histology. Histology is considered the gold standard as its capable of viewing the skin tissue with cellular resolution. However, at present around 80% of the histological examinations result negative [12]. Due to the large number of false positive excisions, there is clinical interest to examine the skin in its native state with other non-invasive modalities to examine the skin before deciding whether to excise. Such non-invasive examinations have the potential of reducing the number of false positive excisions. Furthermore, such technologies have the potential of replacing histology to examine the progression of a skin condition or disease. For the latter to be possible, clinicians need to achieve comparable levels of sensitivity and specificity with the modality as to those achieved with histology.

2.3.1 Visual clinical examination

Dermatologists are trained primarily to analyse skin conditions by visual examination, and use algorithms that have been developed to formalise the findings from clinical experience in identifying skin lesions, including in particular melanoma. The *ugly duckling* rule [56], for example, illustrates that a greater level of suspicion should be directed at a lesion that differs from those over most of the patient's body.

Another algorithm often employed is the ABCD-(E) rule developed by Friedman et al. [57] as an aid to visual examination for the identification of melanoma. This rule incorporates the key clinical findings associated with melanoma, in-

cluding asymmetry (A) of the lesion, the irregularity of the border (B), the variation in colour (C) across the lesion, the diameter (D) and the growth or evolution (E) of the lesion. The ABCD-(E) algorithm combines these observations such that the more findings that are present and the more marked these findings are, the greater the level of suspicion that the lesion is melanoma. The diagnostic accuracy achieved with clinical visual using the ABCD-(E) rule lies between 60% and 80% for malignant melanoma [58]. Visual examination still forms the largest basis for skin lesion examination, even in the secondary care setting.

2.3.2 Dermoscopy

Visual examination relies on photons reflected from the skin surface. The majority of such photons, however, carry no diagnostic information and simply represent glare direct from the skin surface. In the 1990's Pehamberger et al. [59] developed what is now known as dermoscopy, which is a non-invasive *in-vivo* examination of the skin. It uses incident light delivered from an acute angle of around 20°, which passes through a refractive index matching fluid (oil), enabling sub-surface skin structures (at a depth of about $20\mu m$) to be examined. The image from a dermatoscope, therefore, contains information from the underlying tissue or lesion not seen visually. The simplest implementation of the dermatoscope is a magnifying glass that is either coupled to the skin or alternatively is fitted with polarising filters that remove glare.

It is now considered standard practice amongst dermatologists that a dermatoscope should be used to examine the skin [60]. Dermoscopy technology is advancing and current systems provide digital images that can, if desired, be analysed by software, including software that implements algorithms such as the ABCD-(E) rule automatically [61].

This technology introduces some additional information to the basic information achieved using the ABCD-(E) rule, which can help improve the accuracy of diagnosing melanoma [62]. However, as reported by Argenziano et al. [63], the specificity of melanoma detection using dermoscopy is still low, resulting in a

large number of unnecessary excisions due to false-positive identifications. The number of moles needed to be excised in order to detect one melanoma ranged from 8.70 to 29.4 in specialised and non-specialised centres, respectively [63].

2.3.3 Histology

As discussed in Section 2.3.1, a clinician initially examines a suspicious mole with the naked eye, together with a dermatoscope. In the case that clinical examination confirms the clinician's suspicion then a biopsy is indicated in order to provide a definitive diagnosis. The biopsy may remove the entire lesion or indicate the need for more extensive surgery.

Histology involves the study of signs of disease using the microscopic examination of cut sections of a specimen placed onto glass slides. The different components of the specimen are typically analysed by dyeing the section with hematoxylin and eosin (H & E). The advantage of histology is that the preparation process preserves the underlying tissue architecture, while the histology slides provide a comprehensive view of the disease and its effect on tissue (Figure 2.3). Interpretation of histological slides is based largely on qualitative visual examination rather than computer aided interpretation. Indeed, histological images are normally not digitised but are viewed by eye using a microscope. This is changing as increasing numbers of digital systems are being used to capture such images.

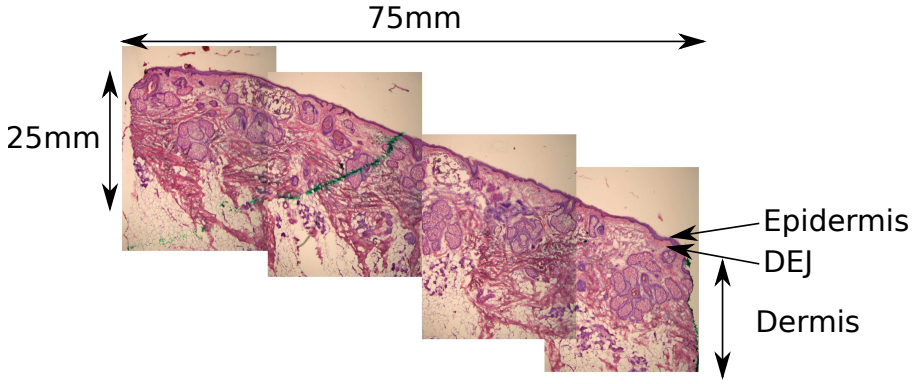


Figure 2.3: A composite of four digital images of H & E stained skin tissue.

2.4 Non-invasive methods of analysing the skin

This section describes non-invasive imaging modalities available to dermatologists. Any new imaging modality, if it is to be adopted, needs to add significantly to the accuracy of diagnosis and improve the sensitivity and specificity of an examination. In this section the non-invasive modalities listed in Table 2.3 are discussed in depth, covering optical, acoustic and thermal based technologies.

In Table 2.3 all the imaging technologies discussed in this Section are compared to histology, with respect to *orientation*, *depth* and *resolution*. *Orientation* describes the orientation of the captured image; $\perp$ represents images are obtained perpendicular to the skin surface (like histology) whereas $\parallel$ indicate that the technology captures images parallel to the skin surface. The *depth* indicates the skin depth used to generate the image; a + indicates that the depth extracts information from deep dermis; whilst a - indicates that the depth used is less than deep dermis; whereas surface indicates that only information at skin surface is used. Similarly, *resolution* compares the resolution of the image to that of histology; where a + indicates that the resolution is similar to histology and that individual cells can be distinguished, and - indicates that the resolution is poorer than histology and cells cannot be individually distinguished.

2.4.1 Optical imaging modalities for imaging the skin

In non-invasive optical imaging systems, the skin is exposed to near infra-red laser beams of wavelengths ranging between 700 and 1300nm, capable of penetrating up to the dermal region in the skin. This range of wavelengths results in a relatively low absorption and as the skin is an effective scatterer of photons; photons are scattered more than once [64]. Only a small fraction of the original beam will consist of reflected photons having not suffered some scattering. Various non-invasive optical modalities have been designed based on these optical properties of the skin to capture images. Such technologies can aid dermatologists to diagnose the skin and to map surgical margins.

In this section the different optical modalities that can be used in a clinical

Modality	Orientation	Depth	Resolution
Reflectance Confocal Microscopy	∥	-	+
Optical Coherence Tomography	⊥	+	-
Multi-Photon	∥	+	+
Multi-Spectral	∥	surface	-
Raman Spectroscopy	∥	-	+
Spectrophotometer	✗	surface	-
Laser Speckle	∥	surface	-
Capillaroscopy	∥	-	+
Laser Doppler Imaging	∥	+	-
Doppler Ultrasound	⊥	+	-
Thermal	∥	surface	-

Table 2.3: Overview of non-invasive technologies capable of providing images or quantitative measures of skin anatomy. Orientation refers to the orientation of the image, where ⊥ indicates that the image is captured perpendicular to the skin surface, as in histology, whereas ∥ represents that the image is captured parallel to the skin surface. The depth indicates whether the modality can extract information from the deep dermal layers, a + sign indicates yes whilst a - indicates no. Surface indicates that the modality can only obtain information from the surface of the skin. A + and - for resolution indicates that individual cells can or cannot be distinguished respectively.

environment to analyse skin structures and blood circulation are listed and their relevance to our study is identified.

2.4.1.1 Reflectance Confocal Microscopy

Reflectance confocal microscopy (RCM) is showing increasing potential as a non-invasive diagnostic tool in dermatology, not least because it is able to resolve individual cells and organelles to a depth of $200\mu m$, providing high resolution ($< 1\mu m$) images of near histological quality but without the need for biopsy. In RCM the skin is exposed to a near infra-red laser beam (of wavelength $830nm$), and an image is generated from the reflected light. As opposed to a conventional microscope, a pinhole is implemented inside the RCM, which rejects light from out-of-focus planes. Image contrast results from variations in the refractive index at different skin layers [65]. The physics behind RCM technology is discussed in more detail in Section 4.2.

The VivaScope® 1500 (Lucid Inc., Rochester, New York, USA), used in this study, can capture stacks of images covering the same $0.5 \times 0.5mm^2$ area at dif-

ferent depths up to a depth of $200\mu m$, which corresponds to the dermal region. In addition, it can form mosaics which cover a maximum area of $8 \times 8mm^2$ up to a depth of $200\mu m$. The different imaging options possible with the RCM in this study are discussed in more depth in Section 4.2. The VivaScope® 1500 also has the ability to capture videos of an area of $0.5 \times 0.5mm^2$ at a frame rate of 9fps. In these videos one can visualise blood flow in capillaries located at the DEJ. Individual RBCs can be seen in RCM video sequences as they move through capillaries.

The main difference between *in-vivo* RCM images of skin and histology is that the RCM images are obtained in a plane parallel to the skin surface, whereas histology traditionally is orientated in a plane normal to the skin surface. Due to this difference in orientation, initially dermatologists find it difficult to interpret *in-vivo* RCM images.

When imaging excised tissue, RCM has an advantage over histology in that tissue can be prepared faster and without the need of a wet lab. As depicted in Section 4.6.1, the tissue preparation for RCM imaging can be performed in less than 5 minutes, after which the confocalist can image the tissue. Thus, RCM has applications for both *in-vivo* examinations to act as a diagnostic tool or to monitor therapy response, and for *ex-vivo* examinations to aid dermatologists during surgery to identify tumour margins. There is growing literature on the use of RCM in dermatology as well as photobiology (refer to Chapter 3).

One limitation of RCM in its clinical use is that it generates large amounts of data per scan. For example, a single RCM examination with the VivaScope® 1500, covering an area of $0.5 \times 0.5mm^2$, generates about 60-100 high resolution images [4]. This thesis is concerned with the extraction of quantitative information from images obtained by RCM of normal skin and skin lesions. The aim is to provide the users with software tools which can minimise the effort of examining such large datasets and which can be used interactively to highlight suspicious areas and to measure fundamental skin parameters such as the thickness of the DEJ and blood flow in the superficial capillaries. Detection of sub-clinical his-

tological changes are likely to be important in early diagnosis and in aiding the making of pre-surgical margins. These automated algorithms may also provide researchers with more quantitative measures and end-points for photobiology and other research. Validation of these tools using clinical data will be demonstrated.

2.4.1.2 Optical Coherence Tomography

Optical Coherence Tomography (OCT) is another non-invasive modality used to visualise the skin *in-vivo*. The main advantage of OCT over RCM is that it offers deeper depth visualisation (1 to 2 mm). However, the OCT spatial resolution ($10 - 20\mu m$) is roughly 20 times larger than that of RCM [66]. The OCT resolution is sufficient to provide images of architectural features in skin. The principal limitation of OCT is that it is not capable of visualising typical and atypical cell structures, which are important for diagnosing cancer in its early stages.

OCT, in particular the VivoSight (Michelson Diagnostic Limited), works by initially directing an optical beam of near-infrared light with relatively long wavelengths, of $1300nm$, at the tissue under observation. The light is split into two paths, referred to as the reference and sample arms [64]. The length of the reference arm sets the depth of the imaging. The sample arm's light is focused onto the superficial skin layers [67], backscattered and recombined again with the reflected reference signal [64]. A cross-sectional image is then estimated from the interference signal between the reflected reference beams. An image is produced when a strong interference is present; this is when the reference and sample arms have travelled the same optical distance and thus both waves are coherent [67]. The ability to alter the axial distance of the reference arm enables optical sectioning of the sample (A-scan), resulting in an optical reflectivity value versus depth. A B-scan is produced by assembling a collection of neighbouring A-scans.

Using OCT one can identify changes from normal skin architecture, together with some pathologies [68]. Early pathological changes at the cellular level are not seen in OCT images. Therefore, OCT is more beneficial in delineating the

extent of disease that has already caused significant architectural change.

Port wine stains (PWSs), which are pathologically characterised by superficial dermal vessels [69], were examined using OCT which enables the visualisation of dermal vessels. A limitation of OCT is that it cannot visualise vessels which are not dilated [69], due to its limited spatial resolution of $10 - 20\mu m$. Further studies have been done to combine OCT with Doppler ultrasound (DOCT) to visualise blood flow in conjunction with skin architecture. DOCT provides quantitative information of blood flow together with the structural anatomy [70].

2.4.1.3 Multi-photon Imaging

Multi-photon imaging is a type of fluorescence imaging that has a penetration depth of about $1mm$ in tissue. The resolution obtained by multi-photon imaging is comparable to that obtained by RCM and the images are acquired using a microscope. This technology is based on a two-photon excitation processes which simultaneously absorbs two photons in the near infra-red wavelength, resulting in an electronic transition that normally requires an ultraviolet photon [71]. This technique can induce fluorescence emission at much shorter wavelengths than the original photon wavelength, therefore eliminating the need to use ultraviolet excitation [71]. The advantage of multi-photon imaging over RCM is that it can image additional structures, providing better cellular discrimination in the stratum corneum and visualising the deeper collagen and elastin fibres located in the dermis layer [72], which are not normally clear in RCM. However, these advantages do not overcome the fact that multi-photon imaging causes thermal damage (photodamage) in highly pigmented cells [72], which can lead to DNA damage [73]. Similarly to RCM, multi-photon imaging can be used to examine excised tissue.

2.4.1.4 Multi-spectral Imaging

Multi-spectral imaging makes use of light at visible and infra-red wavelength regions, ranging from 400 to $970nm$ [74]. The main goal of multi-spectral imaging is to obtain a set of images of the same object taken at different specific spec-

tral bands [75]. Image processing is used to identify the band or cumulation of bands that best detect disease. This imaging modality is useful in identifying melanoma [68] and has been shown to detect melanoma with a sensitivity and specificity of 100% and 85% respectively [13]. Similar to dermoscopy, multispectral imaging does not provide images of structures beneath the skin surface or of the micro-circulation in skin capillaries.

2.4.1.5 Raman spectroscopy

In contrast to the above mentioned techniques that analyse reflected light, Raman spectroscopy analyses inelastically scattered light. Raman light is scattered light which has a frequency different from that of the incident light (inelastic scattered light) [76]. Although the inelastic signal is very weak, Raman light gives information about the vibrations of molecular constitutes in the skin. The main disadvantage of this technology which limits its translation into the clinic, is the speed at which an image is captured [77].

Coherent anti-Stokes Raman scattering (CARS) refers to scattered photons having a higher energy state (higher frequencies) than the incident photons. This population of photons at a higher energy state will be smaller than those in a lower state [10]. As the signal will have a higher energy than the background fluorescent signal it enables enabling real-time visualisation at a high resolution [78]. The resolution obtained using CARS microscopy is $\sim 0.5\mu m$ and thus comparable to that achieved using RCM [10]. CARS microscopy enables visualising blood flow at 20fps [79], and it can be also be used to image excised tissue. In the literature, the Raman technology was used to produce stack images of the skin in-vivo at various depths [80].

In some cases Raman spectroscopy is combined with another imaging modality, such as the OCT, to give a more comprehensive view of the skin. In the case of Raman spectroscopy combined with OCT an image incorporating morphological and biochemical characterisation of the skin is achieved [81].

2.4.1.6 Laser Speckle

Laser speckle contrast imaging is used to visualise tissue blood perfusion. Motion of particles inside the illuminated tissue can be detected by changes in the amount of backscattered light accumulated over a certain time interval (the exposure time). Thus, the accumulated backscattered light will be averaged by the exposure time, resulting in movement having an effect of the speckle obtained. A relationship is obtained between the speckle and the speed of the moving particles [82]. If a stationary object is imaged, there will be no speckle effect and a sharp image is obtained. In contrast, when blood flow is imaged, the speckle image changes over time. The environment in which laser speckle recordings are made may effect the results obtained. The depth from which the image is generated is not clearly known and 95% of the backscatter originates from the top $700\mu m$ of the tissue [83]. Thus, it is difficult to differentiate the perfusion measure of capillaries from the underlying vessels and tissues. This technique only provides the user with arbitrary perfusion units that cannot be compared against other perfusion measures. In addition, this imaging technique cannot provide any information about the underlying morphology of the skin.

2.4.1.7 Capillaroscopy

Capillaroscopy makes use of a lens of magnification $\times 10$ to $\times 100$ to investigate capillaries in the nail-fold under a microscope [34]. This technique is limited to the nail-fold as it requires the capillaries to be located parallel to the skin surface, whereas the majority of capillaries in body sites run perpendicular to the skin surface [84]. It was reported by [34] that the capillary diameter recorded using this method were of $5 - 7\mu m$. Capillaroscopy enables the visualisation of the whole capillary bed which can be used to investigate the capillary density and the regularity of the capillaries shapes [34]. This technique is especially useful to analyse systemic sclerosis which is pathologically characterised by microvascular abnormalities in the fingers and toes (Raynaud's phenomenon) [52].

2.4.1.8 Laser Doppler Flowmetry and Imaging

Laser Doppler flowmetry (LDF) uses a low-power near infra-red beam. The scattered light from the skin surface is detected and the Doppler wavelength shift of this collected scattered light produces the LDF signal [34]. The scattered photons that previously collided with static structures do not experience a change in wavelength/frequency, whilst those photons colliding with moving red blood cells (RBCs) will experience a Doppler shift. LDF provides measurements related to changes in blood volume and velocity from a small sampling volume. RBCs moving at low speeds (e.g. $0.1mm/s$) typically found in the smallest capillaries, generate only a small Doppler shift and thus can be difficult to distinguish from noise [84]. LDF measures provide only relative changes of flow with time and thus different areas cannot be compared using this technique. However, these flow measures have been shown capable of discriminating between melanoma and benign pigmented skin lesions as vascularisation gradually increases during the transition of a lesion from benign to malignant melanoma [85]. A further limitation is that LDF measures do not provide depth information, therefore one cannot know the depth of measurement where perfusion is occurring [86].

Laser Doppler Imaging (LDI) has been introduced as an extension of LDF, it is a contactless technology and is capable of imaging a larger sampling volume. LDI enables the comparison of blood perfusion across different sites and across different scanning times [34].

2.4.2 Optical instrumentation for quantifying colour

In this Section the spectrophotometer is discussed in depth. Although this technology does not produce images, it is discussed in depth as its measurements of colour are used in Chapter 7.

In this thesis the CIE LAB colour space was used to measure the colour of a skin site using a Minolta Chroma Meter CM-700d spectrophotometer. As opposed to the human eye, which has three main cones (red, green and blue) to discriminate colour, the Minolta spectrophotometric instrument has multiple sen-

Skin Type	Skin Classification	ITA°
I	Very light	$ITA° > 55°$
II	Light	$41° < ITA° < 55°$
III	Intermediate	$28° < ITA° < 41°$
IV	Tan	$10° < ITA° < 28°$
V	Brown	$-30° < ITA° < 10°$
VI	Dark	$ITA° < -30°$

Table 2.4: ITA°skin classification as reported by Del Bino et al. [88]

sors which are used to measure light in a strictly defined region of the visible light wavelength range ($0 - 700nm$). This enables spectrophotometers to achieve a high accuracy and the ability to measure absolute colours, resulting in an ability to discriminate between more colours than the human eye.

The main outputs of the spectrophotometer when using the CIE LAB space are L^*, a^* and b^* values. The L^* value indicates the lightness of the colour, where it records the darkest black at $L^* = 0$ and the brightest white at $L^* = 100$. The a^* and b^* components give the chromaticity coordinates, where the red-green component is defined by the a^* axis. A positive a^* value indicates red while a negative a^* value represents green. Similarly, the b^* axis represent the yellow-blue component, with a positive b^* value indicating yellow and a negative b^* value representing blue. In the case that both a^* and b^* are equal to 0, this indicates achromaticity which can be characterised as a neutral grey colour. In literature related to skin colour, the a^* value is used to determine the reddening factor of the skin.

These values can also be used to calculate the individual typology angle ($ITA°$) which can be used to classify skin as defined by Del Bino et al. [87] [88]. As illustrated in Table 2.4, the $ITA°$ allows skin colour classification into six types, from very light to dark skin, corresponding to the skin classifications identified in Table 2.1. The $ITA°$ makes use of colorimeter parameters L* and b* and is calculated as follows:

$$ITA° = \tan^{-1} \frac{L* - 50}{b*} \times \frac{180}{\pi} \tag{2.1}$$

2.4.3 Acoustic

High frequency ultrasound (HFUS) scanners, with frequencies ranging from 20 to 50MHz and short wavelengths, are capable of visualising the skin [84] with high axial resolution [89]. A variety of tumours can be evaluated using HFUS and generally this technology is used to map the tumour thickness. The thickness of very thick and thin tumours is overestimated and underestimated respectively, as the differentiation between tumour and inflammation surrounding the tumour is limited in ultrasound [68]. Ultrasound imaging can be enhanced through techniques such as Doppler blood flow measurement [90]. Doppler imaging can differentiate between blood moving towards and away from the transducer. The blood flow measures are calculated from the Doppler shift, which combines information from both the deeper dermal and the capillaries located close to the DEJ [91]. The resolution achieved with Doppler ultrasound is not sufficient to image small vessels and the velocity component cannot originate from more than one moving structure [86].

2.4.4 Thermal

The temperature of the skin surface can be used as a regional indicator of blood flow of vessels located in the dermis. However, the principal disadvantage of this methodology is that it is difficult to differentiate heat generated from the environment or vessels from underlying tissues and organs [84]. Contact or non-contact devices can be used to measure the temperature distribution of the skin surface. Contact thermography measures the heat of the surface by heat conduction whereas non-contact thermography the thermal radiation emitted by the skin changes the state of the detector [34]. The temperature measured by contact thermography is lower as the contact of the detector reduces the temperature of the skin surface while non-contact thermography also measures surrounding radiation [34]. Thermal conductivity makes use of a probe that raises the skin surface temperature by several degrees and the temperature difference with and without the probe is recorded. The smaller the temperature difference the larger the blood flow [84].

2.5 Conclusion

In this Chapter, the imaging modality used in this thesis, RCM, is compared against other non-invasive methods of analysing the skin. The main benefits of RCM is that its provides a near histological resolution and can image skin up to a depth of around $200\mu m$. This penetration depth is sufficient for dermatologists to make a diagnosis. As opposed to histology slides, RCM images are obtained parallel to the skin surface, therefore dermatologists need to familiarise themselves with the new orientation and grey-scale images prior using RCM images to make a diagnosis. Additionally RCM can be used to analyse skin *in-vivo* and excised tissue samples as it can be operated in reflectance and fluorescent modes.

Chapter 3

Literature Review

Contents

3.1 Introduction 50

3.2 RCM for imaging *in-vivo* skin 50

3.2.1 Detection of skin sub-layers and DEJ 53

3.2.2 Artefact removal 59

3.2.3 Detection of cancerous regions and segmentation of individual cells 61

3.2.4 Discussion 62

3.3 RCM for imaging excised tissue 63

3.3.1 Introduction 63

3.3.2 Mohs surgery 64

3.3.3 Feasibility studies of using RCM in Mohs surgery 64

3.3.4 Tissue preparation for imaging in RCM fluorescence mode 66

3.3.5 Imaging BCC in RCM fluorescence mode 66

3.3.6 Algorithms to analyse images of excised tissue 67

3.3.7 Discussion 70

3.4 RCM video sequences for visualising skin capillaries 71

3.4.1 Introduction 71

3.4.2 Methods under which skin capillaries were analysed . . 71

3.4.3 Quantification of skin capillaries 72

3.4.4 Discussion 75
3.4.5 Discussion 78
3.5 Conclusion 79

3.1 Introduction

The primary motivation of this literature review is to set the work done in this thesis in context. In this Chapter published algorithms that have been developed to analyse Reflectance Confocal Microscopy (RCM) images automatically or semi-automatically are outlined. Similarly, procedures to extract quantitative values manually from RCM images are identified. The motives for developing such algorithms and the applications of the extracted quantitative values are discussed.

The review is divided into three sections: (1) algorithms developed for analysing RCM images captured from *in-vivo* skin, (2) algorithms developed for analysing cancerous regions in RCM images of excised tissue, and (3) algorithms developed and procedures for analysing skin capillaries in RCM videos. This Chapter is concluded with a discussion outlining the main achievements.

3.2 RCM for imaging *in-vivo* skin

Since the 1990s [15] RCM has become clinically available to dermatologists to analyse skin *in-vivo*. Dermatologists have devoted time to analyse skin conditions and diseases using the non-invasive histological resolution modality. Scoring methods [92], have been defined to quantify skin conditions and diseases, to aid the dermatologist community in interpreting and making a diagnosis. RCM has also been used for analysing skin ageing [93] by the cosmetics industry, in studies aimed at understanding how skin changes due to ageing can be delayed or reversed. Although knowledge has been acquired in interpreting RCM images, there is a steep learning curve and the clinician needs to devote a con-

siderable amount of time to achieving confidence. For example, MAVIG©, the supplier of the VivaScope® RCM technology in Europe, suggests a one week training course and 6 months online practice prior to dermatologists using the RCM for clinical purposes.

A typical RCM stack results in 60-100 images [4]; the process of analysing these images is time consuming as it requires a human expert to evaluate the images individually [5]. In order to reach the full potential of non-invasive RCM for detecting changes in skin conditions and diseases, together with acting as a diagnostic modality, additional tools are required to provide clinicians with a faster way of analysing the images [5].

Automated based algorithms have the potential to give a more quantitative and objective approach to analysing the skin and removes the inter-variability found across different clinicians [1]. This has created a niche for imaging scientists in developing algorithms and providing quantitative parameters to aid dermatologists in interpreting the images.

The papers discussed in this Section have developed semi or fully automated algorithms that analyse the morphological architectures and anatomical structures of the skin at different skin depths imaged using RCM. The discussed papers will cover work done in three main research areas:

- Automatic detection of skin layers and the dermal-epidermal junction (DEJ) from a vertical stack of images (Kurugol et al. [1] [2] [3], Somoza et al. [4], Hames et al. [5]);
- Artefact removal as a pre-processing step by detecting wrinkle areas (Sourati et al. [94] [12]);
- Detection of cancerous regions (Wiltgen et al. [6] [7]), segmentation of individual cells to analyse disease (Gareau et al. [95] [96]) and photoageing (Raphael et al. [97]) in individual image tiles.

An overview of the main achievements of these papers is given in Table 3.1.

Author	Year	Motive	Data	Skin type	Body site	No. of volunteers	Age range	Method	Pre-processing	sub-images area (pixels)	Accuracy	Error (μm)
Kurugol et al. [1]	2011	DEJ	4 stacks, $\times$60 images, $1\mu m$ spacing	I-II				texture features	masked out artefacts & registered tiles	50×50	90%	8.5
Kurugol et al. [2]	2012	DEJ	24 (15F/9D) stacks, $1\mu m$ spacing	I-II, III-VI				contrast and texture features	registering of tiles			~ 6.4, 5.3 ± 4.27
Kurugol et al. [3]	2014	DEJ	30 stacks, $1\mu m$ spacing	I-II, III-VI	forearm	30		contrast and texture features	registering of tiles	50×50, 32×32	58.6%, 88.0%	~ 7.95, 7.9 ± 6.4
Somoza et al. [4]	2014	5 skin layers	3 stacks, $\times$65 images, $1\mu m$ spacing		forearm	3		filters	masked out artefacts	25×25		
Hames et al. [5]	2015	4 skin layers	308 stacks, $\times$50 images, $2\mu m$ spacing	mainly I-III	dorsal & volar forearm	54	20-29, 50-59	features	normalising	7×7	85.6%	4.8 ± 4.8
											Sensitivity	**Specificity**
Sourati et al. [94]	2012	wrinkle-detection	40 images		arm, knee, abdomen, forehead, back			region growing			88%	97%
Sourati et al. [12]	2013	wrinkle-detection	3 stacks					probabilistic, texture		20×20		
Wiltgen et al. [6]	2008	melanoma detection	100 images					texture features		256×256	97.0%	96.0%
Wiltgen et al. [7]	2011	melanoma detection	857 images					texture features		512×512	84.1%	78.0%
Koller et al. [8]	2009	melanoma detection	16 269 images			178		texture features			58.5%	48.5%

Table 3.1: A summary of RCM studies analysing *in-vivo* skin and are discussed in this literature review. The true optical sectioning thickness of the VivaScope® 1500 *in-vivo* RCM, the minumum depth a tile can represent is $\sim 3\mu m$. Thus, for each acquired tile, the middle z-depth (the $1.5\mu m$ mark) is incremented by $1\mu m$; resulting in image overlapping by $0.5\mu m$ from the image above/below it when a spacing of $1\mu m$ is used. The sub-images used in Kurugol et al. were of 50×50 and 32×32 for fair and dark skin types, to incorporate information from neighbouring regions in fair skin and to represent the area of 1-2 basal cells in dark tones. A window size of 25×25 was chosen in [4] to replicate the size of keratinocytes.

3.2.1 Detection of skin sub-layers and DEJ

Clinical RCM images can be acquired as a stack of parallel image planes at increasing depth, which can vary in their spacing between $1 - 5\mu m$. A key step in any automation is to identify the image plane of interest. In this Section work done by a number of research groups aimed at generating semi or fully automated algorithms that sub-divide a stack of images into the various layers is presented. All studies used the VivaScope® 1500 RCM system.

3.2.1.1 Kurugol et al. [1] [2] [3]

Motive Kurugol et al. addresses the problem of localising the DEJ from an RCM image stack based on the different contrasts and texture features characteristic of skin layers. Two algorithms were developed to detect the DEJ in fair and dark skin tones. In this work, the DEJ in fair skin is characterised by a transition zone between the epidermal and dermal layers, whereas in dark skin the DEJ is characterised as the interface between the epidermal and dermal layers. Therefore, the skin is subdivided into three segments (epidermal, transition and dermal) for fair skin, and into two (epidermal and dermal) for dark skin.

Algorithmic Method The classification algorithm used in fair skin is defined in [1]. The main pre-processing steps in this algorithm include manually masking out of wrinkles and other artefacts, and registering images of the stack across the depth axis to compensate for any lateral movement. Each image was then sub-divided into 20×20 sub-regions of size 50×50 pixels, creating a stack of sub-regions referred to as sub-stack.

170 texture features were extracted from each sub-region; including co-occurrence matrix features, statistical metrics and features from a wavelet decomposition. A fast filter feature selection algorithm was used to select the most discriminative and least redundant subset of features from each class. Two subsets of features were chosen: one representing the epidermis class versus the rest, and another representing the dermis class versus the rest. The two selected feature subsets were the input to the classifiers, while the union of the two subsets were the in-

put to the sequence segmentation algorithm.

The sequence segmentation algorithm which is applied to each stack, incorporates the fact that if adjacent tiles have similar features, then they belong to the same class. Initially the stack of images were segmented into k homogeneous and sequential segments of images, where k ranges from 3 to 8 and each segment's features were approximated using an affine algorithm. The model parameters together with the segments' boundaries were optimised using a cost reduction function that minimises the difference between the segments' features and the modelled features.

The boundaries of the sequence were initially set as potential candidates of the layer boundaries. A support vector machine classifier was then used to give a probabilistic measure of belonging to a class. These classification results were used to determine which two end-points represent the boundary locations of the epidermal to transition and the transition to dermal interfaces, i.e. these boundaries enclose the DEJ region. Once the two boundary locations were selected for each sub-stack, the surfaces representing the epidermal to transition and transition to dermis boundaries were generated across the 20×20 sub-stacks. A Gaussian smoothing filter of size 5×5 tiles with a standard deviation of 0.75 was applied to the two surfaces.

Another algorithm detects the DEJ in darker skin types [2] [3]. The main difference between fair and dark skin types is the amount of skin pigmentation. In darker skin strong backscatter from the pigment melanin causes basal cells above the DEJ to appear with high contrast. When designing this algorithm, Kurugol et al. took advantage of this difference in skin composition. The appropriate peak of the smoothed average intensity depth profile of each sub-stack was detected; the peak corresponding to the basal layer was not always the brightest peak. Therefore texture features were introduced to the algorithm to ensure that the correct peak was selected. The DEJ was defined to be the surface layer below the bright basal cells.

Validation The results achieved were compared to ground truths defined by clinical experts. The transition region in fair skin is not clearly defined in [2]. In fact, it is reported that a correlation of 0.81 was achieved between the ground truths segmented by three experts [2]. This inter-expert variability might be attributed to the difficulty of clearly defining the transition region and for its definition to be variable across different experts.

Results As outlined in Table 3.1, the main results for the fair and dark skinned algorithms are an accuracy of 58.6% and 88.0% respectively [3]. The fair skin algorithm achieved an average error of $8.3 \pm 5.8\mu m$ and $7.6 \pm 5.6\mu m$ from the clinician's ground truth for the epidermis to transition zone and for the transition zone to dermis boundaries respectively. The dark skin algorithm achieved an average error of $7.9 \pm 6.4\mu m$ from the clinician's ground truth for the epidermis to dermis boundary.

Discussion To summarise, the approach adopted by Kurugol et al. has some clear advantages as it caters both for fair and dark skin tones, and has been tested on 24 datasets from both fair and dark skin tones. However, as the fair skin algorithm is subdividing the stack of images into three classes whilst the dark skin algorithm is subdivides the stack into two classes, these algorithms do not enable ease of comparison between fair and dark skin stacks. In addition, at present there is no algorithm that differentiates between fair (I-II) and dark (III-VI) skin tones. Thus, the user needs to input the fair/dark stack into its corresponding algorithm.

Another limitation of the algorithm is that the stacks contain images of the epidermis through the dermis; thus, they do not cater for stacks which start above the skin surface (located above the epidermal layers). Normal clinical data acquisition would start at the skin surface and terminate within the dermis. As outlined in Table 3.1 the stacks used for this experiment have a $1\mu m$ spacing, this aids registration as sequential images have overlapping information as the imaging slice thickness is approximately $5\mu m$. However, this spacing does not reflect the spacing used in the clinical setting, where a spacing of at least $2.8\mu m$

is generally used. These points show that the algorithm is not being tested over realistic clinical stacks.

As a pre-processing step the authors are manually masking out large areas which they describe as wrinkles. This reduces the complexity of the problem as the algorithm does not cater for wrinkle or other artefacts, which are present in confocal images. This requires user intervention which is undesirable in a clinical setting.

The algorithm that detects the DEJ in fair skin smooths across the (x, y) axis of the image and also along the depth direction (by means of using the sequential segmentation algorithm) of the stack to ensure consistency of the results achieved. Similarly, the algorithm that detects the DEJ in dark skin smooths across the (x, y) axis. In addition, a window size of 1-2 basal cells is being used in the latter algorithm, and thus the accuracy of the algorithm is highly dependent on whether the window is placed on a basal cell or inside the darker papillae.

3.2.1.2 Somoza et al. [4]

Motive Somoza et al. [4] aimed to differentiate three layers of the epidermis region, (i) stratum corneum, (ii) stratum granulosum and (iii) stratum spinosum, together with the DEJ and dermal layer from RCM stacks of images.

Algorithmic Method The images were convoluted with 10 filters of size 25×25 pixels, that represent the size of a keratinocyte, which is a predominant cell in the epidermal skin. An image from each of the 5 skin sub-layers from the stack available was used for training purposes. Ten filter responses were applied to each of the images and histogram parameters were calculated from these responses. For each image the 3 most discriminative responses were chosen from 10 responses using principal component analysis, resulting in a total of 15 distinctive responses. A classifier based on K means clustering was used to cluster these 15 responses into 5 classes (representing the 5 layer classes).

Validation The results achieved were compared against the ground truth defined by a clinical expert.

Results The algorithm was applied to the stack used for training and an additional of three image stacks. The results are represented in correlation values between the classification class and ground truth label, varying from 0.84 to 0.95. These results cannot be compared to the rest of the results achieved in other papers as they do not give measures of the number of correctly classified images, nor sensitivity or specificity values. However, from the illustrated plots it can be deduced that the amount of correctly classified images using this method were $\sim 50\%$.

Discussion The algorithm aims to differentiate between 5 sub-layers of the skin using filters that replicate keratinocyte cells for detection of sub-layers. Thus, the approach adopted in this paper includes morphological information.

By using filters based on keratinocytes that are more present in the epidermal layers, the algorithm is biased to be more capable of distinguishing epidermal layers than dermal layers. This is illustrated in the presented results, where a more dominant misalignment is present in the last two layers, which do not have a strong presence of keratinocytes.

The main limitation of this experiment is the limited data used for training purposes; only 5 images were used to train the classifier. Similarly, the algorithm was only tested on an additional 3 stacks. The stacks had a $1\mu m$ spacing, which results in sequential images appearing very similar.

This algorithm does not incorporate the fact that deeper depths should be represented by deeper sub-layers. Therefore with this approach, deeper images can be assigned to sub-layers closer to the surface than previous images (images at a lower depth).

No quantitative accuracy results were presented in this paper.

3.2.1.3 Hames et al. [5]

Motive Hames et al. [5] differentiated RCM images into 4 skin layers, by grouping the stratum granulosum and stratum spinosum into one layer and labelling it as the viable epidermis. Therefore, they aim to differentiate between stratum corneum, viable epidermis, DEJ and dermis. The algorithm aims to differentiate thickness across the different layers due to ageing, by making use of volunteers in the 20-29 and 50-59 age groups.

Algorithmic Method The authors excluded cases which had artefacts from their study. Regions of each image were chosen randomly and features were extracted. The logistic regression classifier algorithm was trained on stacks belonging to 36 of the volunteers and tested on the stacks belonging to the rest of the 18 volunteers.

Validation The results achieved were compared against the ground truth defined by a clinical expert. Each stack was labelled twice by the same dermatologist, achieving an intra-observer agreement of 97.4%.

Results As shown in Table 3.1 the overall accuracy (percentage of correctly classified images) achieved from this experiment is of 85.6%, the classes achieving the highest accuracy being the stratum corneum (93.5%) and DEJ (81.1%) respectively.

The algorithm achieved an average error of $4.8 \pm 4.8 \mu m$ from the generated interface based on the classification results to the interface selected by the clinician. The largest error of $6.0 \pm 5.3 \mu m$ being at the viable epidermis-DEJ interface, while the lowest error was reported at the stratum corneum-viable epidermis layer. This is due to the fact that the stratum corneum is the most homogeneous sub-layer across different skin types and across different age groups. Ageing affects deeper sub-layers whereas different skin types are mainly different at the DEJ where the majority of melanin is located.

Discussion This paper [5] included a large dataset covering 54 volunteers of different age groups.

The stacks used for this experiment have a spacing of $2\mu m$ and thus sequential images are quite similar to each other providing overlapping structures.

The algorithm focused on fair skin types (skin types I-III).

3.2.2 Artefact removal

Motive Artefacts due to poor contact, such as wrinkles, are present in *in-vivo* RCM images. Artefacts are a great hindrance to automated algorithms, Kurugol et al. [1] and Somoza et al. [4] removed artefact areas from part of their dataset, while Hames et al. [5] excluded datasets that were noisy from the study.

Sourati et al. [94] developed an algorithm that detects wrinkle artefacts in RCM images of human skin to facilitate subsequent automated algorithms. Wrinkle detection will lead to better DEJ and skin layer detection, as wrinkles influence the shape of nearby skin layers, especially the epidermal and DEJ regions. In addition, the wrinkle detection algorithm proposed will eliminate the manual pre-processing step of masking out wrinkles.

Wrinkles are described as 3D valleys with less defined boundaries than other tubular structures, such as airways. Generally, RCM images of wrinkles lack texture and are dark. However, some wrinkled regions are highly textured and thus they resemble non-wrinkled regions, making it harder to correctly classify them.

Algorithmic Method Sourati et al. [94] initialised the wrinkle position by selecting a number of low-intensity points. These points were fed into a watershed region growing algorithm, and regions smaller than 150 pixels were set to non-wrinkle regions. The selected threshold ensures that the resulting region has moderate shrinkage. Finally they expanded this region using a region growing

algorithm.

Sourati et al. [12] incorporated an extension of the wrinkle detection algorithm [94] to Kurugol's et al. [1] DEJ localisation algorithm for obliquely oriented RCM stacks. This extended algorithm makes use of a probabilistic model that takes into account skin depth, texture information and morphological features. This probabilistic model also takes into account the morphological characteristic of the skin structure. After the skin surface, one visualises the epidermal layers, after which the dermal layers are seen, resulting in the probability that a slice belonging to epidermis decreases as skin depth increases. In addition, the epidermis layer around a wrinkle is thinner than for non-wrinkled regions. Thus, the probabilistic model will take this into account, resulting in the probability of being an epidermal image to fall sharply around wrinkles. The algorithm also integrates classification results based on texture features from the non-wrinkled regions.

Validation The results achieved by the wrinkle detection algorithm [94] were compared to wrinkle labels selected by a clinical expert.

Results The basic wrinkle detection algorithm [94] achieved a sensitivity and specificity of 88% and 97% respectively. No quantitative results for the probabilistic model defined in [12] were presented.

Discussion The main contributions of the former paper [94] is that an automated wrinkle detection algorithm was developed, which was extended in [12] to integrate morphological information extracted from layer detection.

The wrinkle detection algorithm was only tested on individual images and not on stacks [94]. Thus, from the results presented one cannot analyse the progression of detected wrinkle area across images in a stack.

3.2.3 Detection of cancerous regions and segmentation of individual cells

3.2.3.1 Wiltgen et al. [6] [7] and Koller et al. [8]

Motive Wiltgen et al. [6] [7] and Koller et al. [8] focus their research on distinguishing diseased images from normal skin images, by making use of texture features and classification algorithms. These groups focus on melanocytic skin tumours, which can be categorised into benign or melanoma depending on their anatomical structure. Compared to melanoma cases, benign cases appear to have more symmetrically round cells and structured architecture. Due to difference in structure, benign cases exhibit more contrast than melanoma cases, which appear more homogeneous.

Algorithmic Method Wiltgen et al. [6] [7] analysed various texture features, including histogram, co-occurrence matrix, wavelet and Fourier transforms. In addition a number of classification algorithms, covering Bayes, trees, rule, function and lazy classifiers were compared. In addition, the influence of the square element size on classification results was analysed by ranging the size from 64×64 to 512×512 pixels.

Validation The results achieved by these algorithms were compared to the ground truths defined by a clinical expert.

Results The main results achieved by Wiltgen et al. [7] was an overall sensitivity and specificity of 84.1% and 78.0% on the test set using trees classification. Similarly Koller et al. [8] achieved 58.5% and 48.5% for sensitivity and specificity on the test set using tree classification.

Discussion The main contributions of this paper is that it provides a framework of different texture features and classification methods that can be used to analyse RCM images.

One limitation in the study presented by Wiltgen et al., is that the images used were extracted from the DEJ region, and thus do not cover epidermal nor dermal

images [6] [7]. Additionally, images only cover melanocytic skin tumours, and thus do not cover images of normal body sites. To summarise, effort was made to preselect which images were inputted in this study [6].

3.2.3.2 Segmentation of cells

Gareau et al. [95] also generated automated algorithms that detect pagetoid cells linked to superficial spreading melanoma (SSM). Characterisation of pagetoid cells together with the disarray of the DEJ was used to quantify early SSM. In addition Gareau [96] developed an algorithm that detects keratinocytes, which is a predominant cell found at the epidermal layer in RCM images. The motive in identifying the density of keratinocytes is that as the absence of healthy keratinocyte patterns can indicate disease or photoageing. The presence of pagetoid cells and the disorganisation of keratinocyte cells are normally combined with the disarray of the epidermal layer and DEJ. Similarly, Raphael et al. [97] developed a computationally based photo-ageing score based on RCM images that is significantly correlated with other grading approaches. The authors computed and combined several parameters based on intensity and Fourier based texture features to generate an RCM score. However, these parameters were extracted from a threshold representation of the RCM image, which in turn was used to segment out cells. The segmented cells are highly dependent upon the chosen threshold.

3.2.4 Discussion

As illustrated in this review, there is great interest in developing semi or fully automated algorithms that can provide clinicians with objective quantifications. Such algorithms have the potential of eliminating the inter-observer variability, together with speeding up the process of analysing the large amount of RCM images generated from a typical *in-vivo* skin scan.

The designed algorithms have incorporated morphological information and mimicked the methodology adopted by a clinician to assess RCM images. The large majority of automated algorithms aimed at differentiating the different lay-

ers of the skin, whereas others aimed in quantifying cells located at a particular layer of the skin. Individual cells are segmented from images captured from a particular skin layer (and skin depth). Therefore, these two set of algorithms can be combined, by first locating the different skin layers and then searching for particular cells or disarray at a particular layer.

Furthermore, artefacts are present across the majority of RCM scans, although artefacts due to poor imaging technique can be mitigated, wrinkle artefacts cannot be eliminated. Therefore, methods of detecting such artefacts can be usefully employed in automation based algorithms of RCM images.

3.3 RCM for imaging excised tissue

3.3.1 Introduction

In this Section the application of RCM to image excised tissue is discussed in depth. Currently in the clinical setting RCM is more popular as a non-invasive modality to image *in-vivo* skin, rather than to examine excised tissue. Therefore, there is less literature on RCM for *ex-vivo* examinations, and less work has been done to extract information from such images automatically. In this Section a brief overview of Mohs surgery, which at present is the surgery mostly studied to examine the feasibility of using RCM images of excised tissue for diagnostic purposes, is given. In addition, these feasibility studies are highlighted and their findings discussed. The different protocols used across clinics in imaging excised tissue and possible shortcomings of image quality are reviewed.

As RCM can be applicable to image pathologies from breast, colon, thyroid, ovarian, head and neck, esophageal, liver, bladder and other tissues [98] [99] [26], automatic algorithms developed to analyse images of excised tissue from different body sites are discussed. Similarly, in this review I will be reviewing algorithms using images of excised tissue not captured using an RCM technology.

3.3.2 Mohs surgery

RCM has great potential in the surgery setting, as it can be used instead of histology for detecting tumour margins in an excised tissue sample. Bennàssar et al. [100] suggests that RCM is an attractive alternative to frozen histology, especially for the case of surgery, as it does not require a wet lab and can be situated in the surgery room. It can give pathological information particularly when used in fluorescence mode.

Mohs surgery is a method that is used to treat skin cancers, notably basal cell carcinoma (BCC), in complex or recurrent tumours and in aesthetically sensitive body sites, such as the head and neck [98]. BCC is often categorised into sub-categories which include; micronodular, nodular, infiltrative and superficial BCCs. Mohs surgery involves multiple excisions of the tumour which enables correct tumour margin mapping and ensures that the smallest amount of normal skin is removed; maximising the likelihood of preserving function and minimising scarring [101]. The cure rate of Mohs surgery is up to 99% [102] and it is considered the highest cure rate for skin cancers.

3.3.3 Feasibility studies of using RCM in Mohs surgery

Currently Mohs surgery is highly dependent on real-time histology, as at each excision the pathological structure of the excised tissue is diagnosed from a frozen section the tissue [98] which can take up to 45 minutes [100]. This is a high cost form of surgery with specially trained histopathologists required. RCM has the potential to reduce the duration of surgery by two-thirds [100]. It has been estimated that using RCM instead of frozen sectioning has the potential to reduce the cost of Mohs surgery by $\sim$ 15% [99]. One possible advantage of RCM images is that it is easy to examine the complete surgical margins of the excised tissue being examined. Generally some of the tissue specimen is wasted in histology slide preparation, so one cannot guarantee that some tumour has not been missed [100]. Most literature related to Mohs surgery focuses on BCC as it accounts for 80% of skin cancer cases [101] [99].

There are two substantial studies that calculate the sensitivity and specificity achieved between RCM images and histology slides of the same excised tissue collected during Mohs surgery. For these studies, the diagnoses achieved from the RCM images are compared to those achieved from the histology slides, as the histology slides are considered the ground truth. The sensitivity gives a measure of the percentage of RCM images correctly diagnosed with tumour with respect to the total number of tumour cases. Similarly, the specificity gives a measure of the percentage of RCM images correctly diagnosed as tumour-free out of the total number of tumour-free cases. Bennàssar et al. [100] reported a sensitivity and specificity of 88% and 99% respectively, whereas Karen et al. [99] reported 96.6% 89.2%, 92.3% and 94.7% for sensitivity, specificity, positive and negative predictive values. Throughout these studies [100] [99], the observers made an effort to use a magnification similar to that used when viewing histology slides ($\times 2$ or $\times 4$). Gareau has argued that a sensitivity of 99% needs to be achieved for RCM to gain a widespread acceptance as a bedside diagnostic technique [26].

Chung et al. [103] reported that 25% of the samples imaged were of poor image quality and did not enable a good comparison with histology. Images can be classed as poor quality if they have poor staining, which can either be appear as saturated (bright regions in image) or having poor illumination (poor contrast in image). In addition, RCM images can sometimes not capture the whole tissue sample due to the tissue being too thick for the dye to penetrate evenly or because of poor preparation of the sample.

There are clearly training needs in both generating RCM images of excised tissue and in interpreting them. Karen et al. [99] noted that one observer, who had experience of RCM images of *in-vivo* skin, needed training to interpret skin structures in fluorescence mode. Another observer's sensitivity and specificity values increased as the study went along, reflecting increased confidence in interpreting the images. In addition, Bennàssar et al. [100] also reported that the missed cases corresponded to a lack of training of the observers in a specific type of tumours (infiltrative BCC).

3.3.4 Tissue preparation for imaging in RCM fluorescence mode

Some RCM studies examined fresh tissue [100] whereas others examined thawed frozen sections [99], or a combination of both fresh and thawed frozen tissue [98]. Bennàssar et al. [100] reported that if the tissue is freshly examined, it can be subsequently examined using frozen histology. Research indicates that it is easier to examine thawed frozen sections, as they are thinner than freshly excised tissue, resulting in better dye distribution and image quality [104]. However, this advantage cannot be met if RCM is used to image fresh excised tissue as a bedside pathological examination; i.e. to generate the RCM image inside the operating theatre during surgery.

Care needs to be taken to ensure that the excised tissue being prepared for imaging is less than 3 mm thick [104]. The tissue is stained with a dye, such as acridine orange. The dye gives rapid staining and the excised tissue can be imaged immediately after being stained [99]. Fluorescence RCM enables the visualisation of individual nuclei, and this can be considered as the major advantage over reflectance RCM. Visualisation of individual nuclei enables better comparison with histology. In general, tumour nests have both tumour and reactive cells and thus are characterised by a high concentration of fluorescence [100].

Harman et al. [104] notes that RCM system VivaScope® 2500 has the advantage of enabling users to flick between fluorescence and reflectance modes for additional help in assessing tissue. However, the reflectance mode is only capable of detecting larger tumours [98], such as nodular BCC, and is not suitable for micro-nodular and infiltrative BCCs which are characterised as small and sparsely nucleated tumour regions [99].

3.3.5 Imaging BCC in RCM fluorescence mode

Although great efforts have been made to define the different structures and patterns associated with BCC [105] [98], there is still ambiguity in distinguishing some structures from tumour regions in RCM images of excised tissue. For example, when sebaceous glands are imaged cross-sectionally they can easily be

mistaken to nodular BCC [98]. It is possible to change from fluorescence to reflectance mode to differentiate between sebaceous and nodular BCC. In addition, inflammation can easily be mistaken to infiltrative BCC and vice versa [98]. In general, infiltrative BCC is more densely populated with nuclei when compared to inflammation.

Poor quality images which miss the epidermis layer make superficial BCC, which originates at the epidermis layer [99] [98], notoriously difficult to detect. This limitation is more common in thicker fresh tissues than in thawed frozen sections, and is discussed in more depth in Chapter 4.

3.3.6 Algorithms to analyse images of excised tissue

The papers reviewed in this Section and the application of Mohs surgery is a great motivation in making the process of analysing images of excised tissue more automatic. Although the process of obtaining RCM images of excised tissue is much faster than obtaining histology slides, it does not necessary translate in the overall surgery being quicker. Surgeons need time to learn how to produce and interpret these images. Gareau [26] produced a mechanism of digitally staining fluorescence and reflectance RCM images of excised tissue to mimic histopathology appearance and facilitating diagnosis based on RCM images. This technique makes use of the fluorescence image to represent hematoxylin (purple) and to label nuclei and the reflectance image acts like eosin (pink) to label cytoplasm and dermal collagen. Algorithms like the ones developed by Gareau make RCM more clinically acceptable.

Gareau et al. [106] also developed a technique to stitch together the individual image tiles making up the mosaic image (imaged with the VivaScope® 2500 technology). Adjacent images in a mosaic of images normally have an offset between them, resulting in structures appearing as discontinued when zoomed with a sufficient magnification. The proposed algorithm in [106] enables better visualisation of individual cells at the borders of image tiles, which in turn enables clinicians visualising smoother mosaic images, and also provides a better

basis on which semi or fully automated algorithms can be based on. Refer to Chapter 4 for detailed description of tile and mosaic images.

A review of Dobbs et al. [9] and Legesse et al. [10] are given below. In their work they have developed algorithms that automatically analyse images of excised tissue.

3.3.6.1 Dobbs et al. [9]

Motive The aim of Dobbs et al. [9] is to distinguish between benign and malignant breast tissue in fluorescence RCM images of excised breast tissue.

Algorithmic Method Dobbs et al. [9] incorporated the knowledge that tumour regions are more densely nucleated than normal tissue and thus detected and segmented individual nuclei using a maximally stable extremal regions (MSER) algorithm. This algorithm is discussed in depth in Chapter 6. From these segmented nuclei, parameters indicating the pattern of nuclei clusters were extracted, together with a measure of how densely or sparsely they were packed.

In addition, the authors developed a gland segmentation algorithm that requires the user to choose an interactive threshold, to segment out individual glands from the nuclei patterns. Statistical feature parameters were extracted from the segmented glands.

The parameters extracted from both these algorithms were fed into a classifier.

Validation A breast pathologist compared the RCM fluorescence images to histology slides stained with H& E, and classified each $750 \times 750 \mu m^2$ image tile into benign or malignant breast features. A limitation of generating ground truths based on a single pathologist's opinion is that true positives cases can be missed while true negatives can be misclassified as positive. Repeatable ground truths are generated based upon more than one opinion.

Results A sensitivity and specificity of 75% and 93% were achieved respectively in differentiating benign and malignant tissue, with an overall accuracy of 90%. The invasive lobular carcinoma (ILC) feature was correctly classified 35% of the cases.

The classification of the gland sites achieved a sensitivity and specificity of 46% and 100% when using features extracted based on nuclear parameters. These values increased to 88% and 88% when using features solely extracted from the gland segmentation algorithm.

Discussion The algorithm illustrates that morphological based features have the potential of improving classification results. In addition, it highlighted the need of introducing features that target a particular class that is poorly performing (in this cases the gland class).

The main limitation of this algorithm is that there is a high reliance on the chosen threshold of the gland segmentation algorithm.

3.3.6.2 Legesse et al. [10]

Motive Legesse et al. [10] developed a classifier to distinguish between malignant and benign structures in images of BCC samples collected using Coherent anti-Strokes Raman scattering (CARS). The classifier is trained to classify 5 distinct classes: skin dermis, startum granulosum, image background, tumour region and adipose tissue. Thus, three of these classes correspond to healthy tissue.

Algorithmic Method Each class is represented by a total of 35 tile sub-images (of size 105×105 pixels), resulting in this study being based on a total of 175 sub-images extracted from 5 mosaic images.

Texture features were extracted from each sub-image and a classifier is trained. Unlike previous studies, the classifier is trained with a permutation set of features. The set of features producing the best results are chosen.

Validation The results achieved were compared to the ground truth labels selected by a clinical expert.

Results An overall sensitivity and specificity of 88% and 89% were achieved in differentiating malignant and benign structures.

Discussion This paper aims to classify images from excised tissue of the skin, covering regions of BCC. The main limitation of the imaging modality (CARS) used is that is only sufficient in visualising larger tumours, as its spatial information is not capable of visualising smaller BCCs [10].

One main limitation of this study is that the classes were trained and tested using only 175 cropped out sub-images from the larger image (35 sub-images per class). In addition, the features used in the algorithm were selected based upon the classifier performance they achieve.

3.3.7 Discussion

As can be noted from this literature review, clinicians are examining whether RCM can be used to image excised tissue and if it has the potential to be used as a bedside pathology examination unit in Mohs surgery. However, for it to be accepted by clinicians, it needs to achieve $\sim$ 99% sensitivity in the detection of cancer, which is currently being obtained by the traditional Mohs surgery using frozen histology slides [26]. This can only be achieved in RCM by incorporating additional information to the surgeon, to speed up their learning curve and their confidence in interpreting RCM images of excised tissue. It is suggested that this can be done by incorporating information from across different modalities, such as using both reflectance and fluorescence RCM modes [26]. In addition to this, further improvements need to be achieved in imaging quality, avoiding poor quality images due to dye penetration issues and due to missed epidermis region. Caliber, the developers of VivaScope® 2500, are already looking into ways of improving the RCM system to image the full view of the epidermis layer. Unless this is achieved, superficial BCCs will be difficult to detect and thus it will be difficult for the system to gain full acceptance as a diagnostic tool.

3.4 RCM video sequences for visualising skin capillaries

3.4.1 Introduction

The aim of this Section is to identify the key measures extracted from RCM video sequences of *in-vivo* skin and to highlight the motivations of extracting such parameters. Skin capillaries are delineated from RCM videos by visually identifying individual RBCs flowing through the capillaries. The studies covered in this Section analyse the use of RCM in monitoring vasodilation and vasocontraction due to environmental conditions or due to a provocation method.

The main quantification parameters extracted from skin capillary in RCM video sequences are reported and discussed. Variations in data collection methods and validation techniques are highlighted, together with shortcomings of approaches adopted.

3.4.2 Methods under which skin capillaries were analysed

In the literature RCM video sequences were used to examine the level of micro-circulation due to environmental conditions and provocation methods. Such provocation methods are used to cause vasodilation and in turn vasocontraction of the DEJ capillaries. Altintas et al. [107] analysed micro-circulation under temperature differences, they exposed volunteers to either local heating or local cold stress. Cinotti et al. [45] analyse micro-circulation after applying a vascular occlusion by placing a cuff around the arm. Căruntu et al. [39] exposed the skin to capsaicin, the hot ingredient of red pepper, to cause cutaneous inflammation, whereas Gambichler et al. [108] exposed the skin to incremental UVB doses.

Other experiments focus on analysing the difference in micro-circulation at different body sites, for example, Hegyi et al. [51] analyse the difference between the dorsal and ventral sides of the forearm, while Altintas et al. [109] focus on differentiating micro-circulation in the forearm and thigh. One motive of analysing the micro-circulation at different anatomical sites is to aid free flap reconstruc-

tion surgery, which involves the transplantation of tissues from one body site to another, in being more successful. It is reported that the higher the level of micro-circulation and the amount of oxygen in the tissue, the higher the success rates for transferred tissue [109]. These studies [45] [39] [107] [51] only analyse capillaries imaged cross-sectionally and thus represented by two separate limbs, and do not cater for capillaries imaged along their length.

Others focus on the changes that occur in micro-circulation due to diseases; Archid et al. [50] analyse the difference between psoriasis patients and healthy volunteers, whereas Venturini et al. [110] differentiate different degrees of systemic sclerosis (SSc). Both psoriasis and SSc are characterised by an increase in the capillary diameter. In addition, psoriasis is pathologically characterised by elongated and tortuous capillaries [50]. As RCM cannot image the changes occurring due to SSc in the deeper dermis, Venturi et al. [110] suggest that for now, until further research is done across different stages of SSc before and during treatment, RCM should be used as a research aid in understanding SSc cases rather than a diagnosis technique. These studies [50] [110] show that certain skin conditions and diseases have the potential of disrupting the distribution of capillaries and of changing their parameters.

This array of papers illustrate the potential of identifying and quantifying capillaries at the DEJ visualised using RCM video sequences. Apart from aiding dermatologists and photo-biologists in understanding the mechanisms of micro-circulation, it can be used to monitor progression of disease and response of therapy.

3.4.3 Quantification of skin capillaries

The mentioned papers aim to quantify parameters of capillaries in *in-vivo* skin imaged using RCM video sequences. For example, the number of RBCs detected in a particular capillary per unit time is a measure of blood flow, whilst measuring the cross-sectional area of a single capillary or capillary concentration (number of capillaries per unit area) are measures of vasodilation. In Table 3.2 a com-

parison of the main results achieved in the reviewed papers is given.

In conjunction with quantifying parameters related to blood flow and vasodilation, some studies [109] [108] [110] also measured changes in the epidermal layer thickness. The epidermal thickness can be an indication of disease progression [110], or the response to UVR [108]. After UVB exposure, the epidermal thickness increases with time due to the increase in melanin production which is provoked by the exposure. None of the mentioned papers have developed a fully automatic mechanism in extracting these measures.

In addition to the results presented in Table 3.2, Gambichler et al. [108] also provided comparisons between the colorimeter and *in-vivo* RCM results achieved at two UVB doses: $1\times$ and $3\times$ MED (Minimal Erythema Dose). Focusing on the results achieved for 3 MED, the vessel diameter increased by 80% after 24 hours, and reduced to an increase of 74% after 72 hours. The erythema colorimeter a^* results increased by 159% after 24 hours, and dipped to an increase of 139% after 72 hours. For the two exposed sites, both the vessel diameter and erythema a^* reached a peak at 24 hours, followed by a dip at 72 hours. On the other hand, Căruntu et al. [39] reported that the capillary area (diameter) increased by 16% (7%), 23% (11%) and 31% (14%) with respect to the baseline capillary area (diameter), at time intervals 10, 25 and 40 minutes after provoking the skin with capsaicin. Similarly, Altintas et al. [107] reported an increase of capillary diameter of 16% and 20% directly after and after 5 minutes of inducing heat to the skin surface.

Altintas et al. [107] noted that heat-induced vasodilation is characterised by an increase in blood cell flow of over 37% across 5 minutes. On the other hand, induced local cooling causes a decrease in blood cell flow of 31% immediately after cooling, followed by an overall increase of 37% in blood flow after 5 minutes from the initial flow, representing reactive hyperemia. Altintas et al. [109] reported a capillary density of 38 ± 3 and 22 ± 2.2 capillaries/mm^2 for the forearm and thigh sites respectively, whereas Venturini et al. [110] reported a ~ 42 capillaries/mm^2 in the forearm. In addition, [110] reported a decrease in blood

Author	Year	Provocation	Disease	Video/Stack	Duration/Depth	Volunteers	Skin Type	Site	Capillary Diameter (μm)	Flow (cells/min)	Other technique
Gambichler et al. [108]	2006	UVB		stack		10	II-III	ventral forearm	13.7 ± 3.8		colimeter
Hegyi et al. [51]	2009		psoriasis	stack		30		ventral & dorsal forearm	17.8 ± 5.3		
Altintas et al. [107]	2010	temperature		video	30s	22		ventral forearm	9.57 ± 0.25	60.9 ± 3.1	
Altintas et al. [109]	2011		transplantation	video		14	III	ventral forearm & thigh		62 ± 3.1 51 ± 2.4	
Căruntu et al. [39]	2012	capsaicin		video	10s	29		dorsal forearm	7.9 ± 1.5		
Archid et al. [50]	2012		psoriasis	stack	$5\mu m$	13	I-III	forearm & lower leg	9.53 ± 1.8		
Cinotti et al. [45]	2014			video	20s	11		interdigital spaces		57.7	
Venturini et al. [110]	2014		SSc	video		20		ventral & dorsal forearm	3.9 ± 0.8	48	video-capillaroscopy

Table 3.2: The values represented for papers [108] [51] [107] [45] have been averaged out from the set of results represented in the paper. In case that a provocation method was used or a disease is being examined, the baseline values or the healthy control results are being represented. [51] [39] capillary diameter represents the Feret's diameter. [109] state that the frame rate of VivaScope 1500 is of 20 fps. For the study designed by Căruntu et al., the video duration was deduced from the videos made available in [39].

flow by nearly half for SSc cases, a flow of ~ 24 cells/minute, whereas the capillary diameter increased by 71.8% to $6.7 \pm 0.9\mu m$. Archid et al. [50] also reported an increase in capillary diameter in diseased skin, this time focusing on psoriasis, the capillary increased by 159.9% to $24.39 \pm 2.34\mu m$.

3.4.4 Discussion

3.4.4.1 Protocol used to collect data

Hegyi et al. [51] analysed the diameter of the same capillary across incremental depths of $5 - 10\mu m$ imaged by RCM and found that capillary parameters remained quite constant across different skin depths. Irrespective of this, Căruntu et al. [39] report of using the same depth of scanning throughout the scans. Heygyi et al. [51] and Archid et al. [50] also report of manually altering the power to achieve the best contrast images, however the exact protocol on how the power was chosen was not illustrated in neither case. VivaScope® 1500 chooses an automatic varying power, which fluctuates depending upon the depth that is being imaged and the contrast achieved with the refractive indices of the imaged structures. Increasing the power results in more visible blood flow passing through the capillaries. However, it also results in melanin being represented with a higher intensity value. If the power is increased much further, the melanin intensity value over saturates.

One limitation when analysing capillaries using RCM video sequences is that it is difficult to scan the same capillaries over different time points. Cinotti et al. [45] overcame this by imaging the interdigital spaces of the fingers by using a handheld version of the *in-vivo* RCM. Căruntu et al. [39] indicates that the same area is scanned throughout the experiment, however, it is not clear whether the RCM was connected to the skin surface throughout the whole experiment or if it was removed and reconnected after each time point. The time points intervals for this study were $0, 10, 25, 40$ minutes.

Gambichler et al. [108] analysed time points further apart, at $0, 24, 72$ hours, thus it is impossible in this case to keep the RCM connected to the skin. In this

case [108], made use a permanent marker to delineate the area to be examined. Altintas et al. [107] marks the area to be examined at the various time points by using an adhesive waterproof sheet.

Although the latter studies [39] [108] [107] scan the same area throughout their study, it cannot be interpreted that the same capillaries were scanned across different scanning points. The VivaScope® 1500 gives a field of view $8 \times 8mm^2$, from which a $0.5 \times 0.5mm^2$ area is chosen to image a video. Thus it is difficult to image the same capillaries once the RCM window is removed and reconnected.

3.4.4.2 Techniques used to validate RCM video sequences

Gambichler et al. [108] made use of a colorimeter to validate the results achieved using RCM video sequences. The colorimeter provides parameters that give an indication of colour and it is used in erythema studies as the redness parameter (a^*) its value is related to the level of micro-circulation below the skin surface. The main parameters of the colorimeter that are used in this study [108] are the brightness (L^*) and erythema (a^*) values.

Venturini et al. [110] compared the results achieved with RCM video sequences with the gold standard of classifying SSc cases, video-capillaroscopy.

3.4.4.3 Motion in RCM video sequences

When capturing RCM video sequences, the volunteer's movement may affect the contact between the microscope and the skin surface. The longer the video, the more movement will be present between the first and last frame of the video. However, the longer the video, the more the video would give an accurate representation of the blood flow inside the capillaries across time. Analysing a longer video, however, allows any cyclical variations to have a higher chance of being detected.

Archid et al. [50] and Altintas et al. [107] both made use of a zooming function that enables capillaries being imaged at a magnified version, and using this mag-

nified capillary to calibrate them using common points. Archid et al. [50] make use of the outer border of the capillary, which is the outerline of the thin halo outside the RBCs, to calibrate whilst Altintas et al. [107] makes use of common points found across frames to calculate repeated readings of the same capillary. However, none of the studies mention registering the frames as a pre-processing step.

3.4.4.4 Repeatability of results and selecting ROI

Cinotti et al. [45] illustrates the difference in manual and semi-automated measures of blood flow. The authors aimed to illustrate the differences in measures between the manual count of the number of bright RBCs passing through the arterial limbs of the capillaries and the peaks detected inside the arterial limb using ImageJ. The region of interest, the arterial limb of the capillary, was manually selected by the authors. It is notable difficult to distinguish between the arterial and the venous limbs visually. The manual count method, which incorporated counts by four doctors, did not achieve intra-capillary and inter-observers reproducible, whereas the semi-automated peak detector achieved intra-capillary reproducibility. This finding is one of the main motivations in generating a fully automatic capillary finder algorithm.

Hegyi et al. [51] and Căruntu et al. [39] identify the region of interest from which capillary measures are calculated by visual inspection.

3.4.4.5 VivaScope® 1500 frame rate and RBC velocity/flow

Papers analysing RCM video sequences make use of the VivaScope® 1500 technology, which has a frame rate of 9fps. None of the papers note that this frame rate might be insufficient in visualising all the RBCs passing through the capillaries. Norman et al. [55] reported a velocity of $0.42mm/s$ and $0.32mm/s$ for the arterial and venous limbs in the nailfold capillaries of newborn infants measured using capillaroscopy. Stücker et al. [111] also reported that the resting blood velocity is $0.47 \pm 0.37mm/s$, ranging between 0.14 and $0.93mm/s$, in capillaries. These recordings were measured using a technology based on laser Doppler and

from these values it can be noted that a frame rate of 9fps is not sufficient to visualise all RBCs going through a capillary.

Quantitative blood flow in the literature is calculated in [45] [109] [107] by counting the number of differentiated RBCs. It is difficult to have a common understanding between different users of what can be classified as an RBC and what cannot. Cinotti et al. [45] denoted some rules used when counting the manual blood count; the first cell that was present inside the capillary at the start of the video was not counted, only cells that preceded in light (representing the plasma) were counted, and a chain of RBCs found across consecutive frames are counted as one cell.

3.4.5 Discussion

As can be noted from this literature overview, RCM video sequences have great potential in analysing blood flow in capillaries at the DEJ. As illustrated in this review, RCM can also be used to indicate the level of erythema of a volunteer. Parameters such as capillary diameter and blood flow (counts/min) extracted from different levels of erythema sites can aid better planning of UVR therapy. This is particularly useful in darker skin types (skin types IV-VI) where it is difficult to visually estimate the patient's response to the UVR, due to lack of visibility in changes of skin redness. In addition to this application, it has a great application in tracking changes in micro-vasculature due to a skin condition or disease, such as psoriasis or SSc.

RCM video sequences are valuable in tracking treatment of such conditions, as unlike histology they do not disrupt the healing process and provide the dermatologist with a mechanism of analysing the structure of capillaries, and that the structure has returned to its normal state before stopping treatment [50]. It is believed that if the capillaries return to their normal diameter, there is a less likelihood of the skin condition, such as psoriasis, returning [50]. In addition, RCM video sequences has the capability of detecting changes in micro-vasculature before other pathological features start emerging. As noted by [110], changes in

capillaries occur before any other pathological features in SSc patients.

This review motivates the need of an automatic identification of a capillary finder algorithm that is capable of extracting capillary parameters that can aid diagnosing skin conditions and diseases; identify the patient's response to therapy and aiding better planning of UVR therapy for all skin types.

3.5 Conclusion

This review has illustrated novel studies that detect and quantify skin structures in RCM images of *in-vivo* skin and excised tissue, together with RCM video sequences of *in-vivo* skin. Literature that is highly relevant to this thesis was discussed in-depth.

Classification based algorithms that differentiate between the various skin layers in stacks of RCM images of *in-vivo* images are discussed, together with algorithms that highlight regions having a high likelihood of being diseased in fluorescence images. Parameters extracted from RCM video sequences of *in-vivo* skin capillaries are discussed.

Attention is given to the developed algorithms, likewise to their applications. The benefits to the dermatology practice associated with the RCM technology are displayed throughout this thesis.

Chapter 4

Reflectance Confocal Microscopy

Contents

4.1 Introduction 80

4.2 Physics behind the RCM 81

4.3 Image acquisition from VivaScope® 82

4.4 VivaScope® Systems Safety Standards 86

4.5 VivaScope® 1500 88

4.5.1 RCM images of normal *in-vivo* skin 88

4.5.2 Artefacts in RCM images of *in-vivo* skin 91

4.5.3 RCM videos of *in-vivo* skin 95

4.5.4 Artefacts of RCM video sequences 98

4.5.5 Power Alteration in RCM video sequences 98

4.6 VivaScope® 2500 99

4.6.1 Mohs surgery and RCM images of excised tissue 99

4.6.2 Artefacts in RCM images of excised tissue 101

4.6.3 Comparing histology to RCM images 103

4.7 Conclusion 105

4.1 Introduction

In this Chapter an overview of the Reflectance Confocal Microscopy (RCM) technology and systems used in this thesis are covered. Initially the general physics

behind the principles of an RCM technology are discussed. Additionally, the RCM systems VivaScope® 1500 and VivaScope® 2500 which are used in this thesis to image *in-vivo* skin and excised tissue are reviewed in depth. An overview of the possible image acquisition techniques available with the VivaScope® system are given. The potentials and shortcomings of each system are illustrated.

4.2 Physics behind the RCM

The RCM technology is depicted in Figure 4.1. A low power ($\leq 35mW$) laser beam is directed in the near infra-red range ($830nm$) through a beam splitter and a $30\times$ interconnected lens system with a numerical aperture of 0.9, onto the area of the skin to be examined. The $\leq 35mW$ laser power does not damage the tissue and light of this wavelength penetrates only $200 - 300\mu m$ in tissue, which corresponds to the uppermost layer of the dermis (papillary dermis) in *in-vivo* scans. When imaging in fluorescence mode, an additional limitation to the imaging depth is the depth up to which the fluorescence dye is absorbed. As denoted in Figure 4.1, the reflected light goes through a small pinhole and forms an image in the detector, where the contrast in the images results from variations in the refractive indexes of microstructures within the skin. The pinhole admits light only from the depth under observation, rejecting light from out-of-focus planes [14]. This is the significant difference between confocal microscopy and conventional microscopy. In order to obtain k different depths, corresponding to parallel sections of the skin, the focus point is altered using a stepwise vertical movement. At each depth a grey-scale image is obtained with a field of view of at least $500 \times 500\mu m^2$ at a resolution of $\geq 0.5\mu m$. No filtering or post-processing is routinely applied to the generated images.

As illustrated in Figure 4.2, the RCM systems for in-vivo and ex-vivo tissue examination are fundamentally identical, differing only in the coupling into the tissue. To optimise imaging conditions in *in-vivo* imaging (Figure 4.2A), oil or an index matching fluid is used as an immersion media between the skin surface and an adhesive window. The oil's refractive index (1.50) is close to that of top most layer of the skin, stratum corneum (1.55). To minimise motion artefacts, a

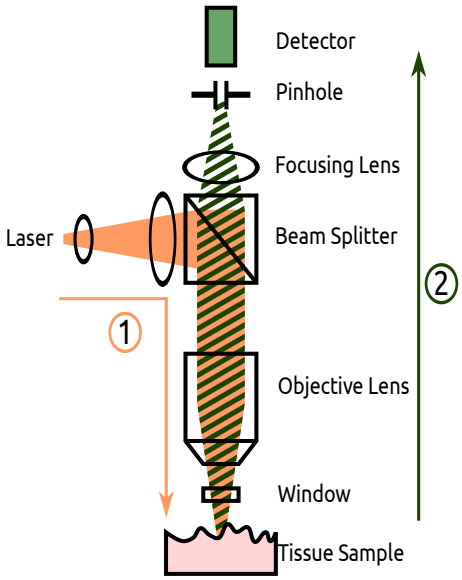


Figure 4.1: Reflectance confocal microscope technology: the orange beam represents the low power laser beam, and the green beam indicates the reflected light collected by the detector.

metal ring is attached to the adhesive window, which in turn is magnetically connected to the lens. In addition to this, an ultrasound gel is occasionally used to fill nodular skin lesions and it is applied at the window's and lens' interface [15].

Excised tissue is imaged using the technique identified in Figure 4.2B; the dyed tissue is placed between two microscopic slides (referred to as a double slide). The double slide is screwed to a board for stability. Ultrasound gel is placed above the double slide, at the side that is going to be imaged with the lens.

4.3 Image acquisition from VivaScope®

For the work described in this thesis two commercial RCM systems (VivaScope® 1500 and VivaScope® 2500, Lucid Inc., Rochester, New York, USA) were used. VivaScope® systems operate by sequentially capturing optical sections at increasing depths of the skin. As illustrated in Figure 4.3, the VivaScope systems enable different imaging acquisition methods: either (A) a vertical images (stack), or (B) a sequential horizontal (block/mosaic), or (C) a combination of the two (cube).

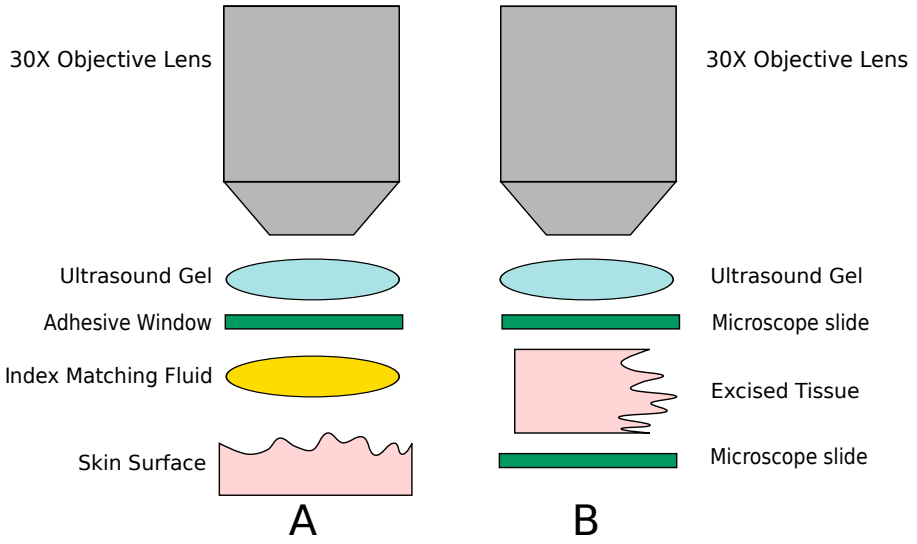
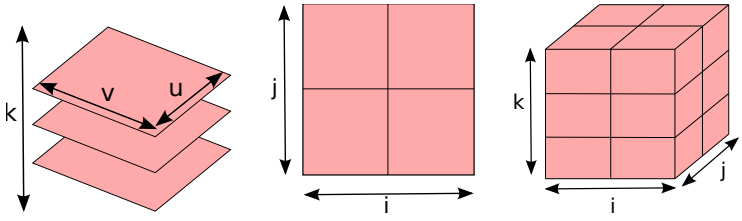


Figure 4.2: Sample preparation for *in-vivo* and *ex-vivo* imaging of skin tissue: (A) an RCM system imaging *in-vivo* skin and (B) an RCM system imaging excised tissue.



(A) Stack: 3 z-depths of basic images (tiles)

(B) Block/ mosaic: 2×2 basic images (tiles)

(C) Cube: 2×2 basic images in 3 layers

Figure 4.3: Different imaging options with VivaScope® 1500 and VivaScope® 2500, where a basic image is of size $0.5 \times 0.5mm^2$ and $0.63 \times 0.63mm^2$ for the respective technologies.

A VivaScope® system provides basic monochrome images, referred to as tiles (T_{ijk}), in the lateral (i, j) direction at a preselected imaging depth (k). Each tile contains 1000×1000 ($N_u \times N_v$) pixels with 8-bits per pixel intensity I. The notation (i, j, k) is used to define a tile's position within the cube of data, and (u, v) defies pixel position within a tile. $T_{ijk}(u, v)$ is used to define the image intensity of a pixel located at (u, v) in tile (i, j, k), i.e.:

$$T_{ijk}(u, v) = I; \Big\{ I = [0, 255], u \in \{1, .., N_u\}, v \in \{1, .., N_v\} \Big\} \tag{4.1}$$

In addition, a tile can be imaged over a duration of time, referred to a video sequences ($V_{ijk}(t)$). The frame rate of the VivaScope® system is of $9fps$ and is capable of imaging for a maximum duration of 1 minute, corresponding to a total of 540 frames. Thus, a video is a sequence of tiles captured at the same horizontal position (i, j) and at the same preselected depth k, but over a time sequence (frame number) t, i.e.:

$$V_{ijk}(t) = \left\{ T_{ijk}, t = \{1, .., N_t\} \right\} \tag{4.2}$$

A stack S_{ij} (Figure 4.3A), is a sequence of tiles captured at the same horizontal position (i, j) but at different imaging depths (k), i.e.:

$$S_{ij} = \left\{ T_{ijk}, k = \{1, .., N_k\} \right\} \tag{4.3}$$

On the other hand, a mosaic (or block) M_k (Figure 4.3B), is a high-resolution composite of tiles taken sequentially in the horizontal plane by an automated stepper and stitched together by dedicated software. Each individual tile is obtained using automatic image control, allowing the best possible contrast for each tile. This results in neighbouring tiles representing the same structure using different grey scale values, and thus stitched tiles sometimes exhibit different grey scale values at the border.

$$M_k = \left\{ T_{ijk}, i = \{1, .., N_i\}, j = \{1, .., N_j\} \right\} \tag{4.4}$$

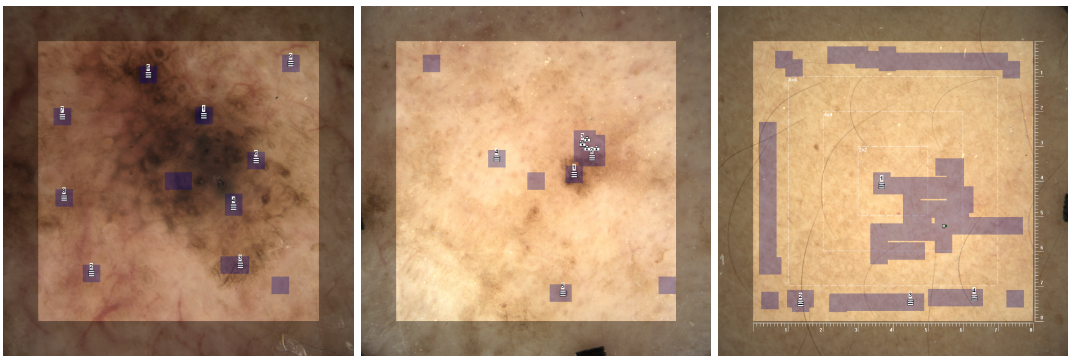
A set of mosaics at different depths make up a cube (Figure 4.3C), thus a cube consists of a series of mosaics covering the same (i, j) region, each having a different (k) depth.

The main specification differences between VivaScope® 1500 and VivaScope® 2500 are outlined in Table 4.1. In this thesis when acquiring data using the VivaScope® 2500 the $488nm$ laser beam was used together with acridine orange

Variable	VivaScope® 1500	VivaScope® 2500
Application	*in-vivo*	*ex-vivo*
Laser beam	$830nm$	$488nm, 658nm$ (fluorescence modes) $830nm$ (reflectance mode)
Depth	$150 - 200\mu m$	dependent upon fluorescence absorption
Detector	Avalanche photodiode	Photomultiplier tube
Pinhole ϕ	$125\mu m$	$75\mu m$ (with $488nm$ laser) $125\mu m$ (with $658nm, 830nm$ lasers)
Field of view (tile)	$500 \times 500\mu m^2$	$630 \times 630\mu m^2$
Maximum $N_i \times N_j$	16×16	32×32
Maximum area	$8 \times 8mm^2$	$20.16 \times 20.16mm^2$
Restrictions	$N_i = N_j$	$N_i = N_j$ or $N_i \neq N_j$
Lateral resolution	$0.50\mu m$	$0.63\mu m$
Axial resolution	$3 - 5\mu m$	$3 - 5\mu m$

Table 4.1: Specifications of VivaScope® 1500 and VivaScope® 2500.

fluorescence dye.



(A) Stacks inside and outside the mole (B) Stacks and image tiles inside the mole (C) Stacks, mosaic and video

Figure 4.4: This figure shows three examples of what data was acquired by the confocalist in the clinical setting. Dermoscopic images taken by the VivaCam covering the total available scanning area of $8 \times 8m^2$. The purple shaded area indicates the area that the confocalist searched in and the purple shaded squares having a corresponding label indicate that a stack, tile or video were captured inside this $0.5 \times 0.5mm^2$ square area. From these examples, one can note that the confocalist makes a quicker decision of which areas need to be scanned in pigmented lesions (Figure 4.4A) and pink lesions (Figure 4.4B), as opposed to non-pigmented skin (Figure 4.4C). In the latter example, the confocalist searched a larger percentage of the available area before finding regions of interest to scan. The dermascopic image in Figure 4.4C has a grid placed on top of the image, indicating the 4×4, 8×8 and 12×12 basic images borders, in which a mosaic of size $2 \times 2mm^2$, $4 \times 4mm^2$ and $6 \times 6mm^2$ can be scanned in. In addition to a grid, a ruler is placed at the bottom and at the right hand axes to give an indication of dimensions.

Three data acquisition examples using VivaScope®1500 are illustrated in Figure 4.4. The displayed dermoscopic images are acquired with the VivaCam, which is part of the VivaScope®1500 system. The dermoscopic image images are used by the user to track which $0.5 \times 0.5mm^2$ region of the $8 \times 8mm^2$ field of view is being imaged. In the example illustrated in Figure 4.4A, 9 stack of images were acquired from inside and outside the mole area of the skin. The purple shaded areas without any labels indicate areas the confocalist checked however decided not to scan in. In the case indicated in Figure 4.4B, the confocalist centred the stacks and image tiles inside the mole. In Figure 4.4C the confocalist searched the area outside the covered mosaic area to check for any missed interesting regions and decided to take three stacks at the bottom of the maps. An additional stack and video were taken inside the mosaic region. The functionality and purpose of taking RCM videos will be discussed later on. An RCM video captures basic images from the same location at a frame rate of 9 frames per second.

4.4 VivaScope® Systems Safety Standards

The VivaScope® 1500 System has been classified as a Class I laser product as defined by and in compliance with 21 CFR sub-chapter J of the Code of Federal Regulations, Center for Devices & Radiological Health, United States Food and Drug Administration (CDRH USFDA) [112].

The VivaScope® 1500 System has been classified as a Class 1M laser product as defined by and in compliance with the International Electrotechnical Commission (IEC) international standard 60825-1 [112].

The VivaScope® 1500 System is classified as Class I Type B equipment with respect to electrical safety [112].

The VivaScope® 2500 System is a Class II laser product and complies with 21 CFR sub-chapter J of the Code of Federal Regulations, Center for Devices & Radiological Health, Food and Drug Administration. For European sales, the

VivaScope® 2500 System is a Class 2 laser product and complies with IEC 60825-1 [113].

The VivaScope® 2500 System is classified as Class II Type B equipment with respect to electrical safety [113].

The guidelines for minimising laser exposure for FDA Class I and II (or IEC Class 1M and 2) laser products include protecting against unqualified use of the System and that users should be adequately trained in laser safety applicable for the System's laser product. User should not stare into the beam or at the beam reflections when the System is in use. Furthermore, reflective objects should be removed from around the System. This includes users and patients removing highly reflective jewellery. The objective lens should not be removed or replaced by the user; instead the user should contact the distributor, Caliber I.D., in case the lens needs changing. When operating the System the area needs to be designated as *Limited Access Area* with signs [112] [113].

Protective eye-wear is not required for FDA Class I and IE Class 1M but required for FDA Class II and IEC Class 2 laser products. Although not required for Class I and Class 1M, protective eye-wear that blocks laser radiation may be worn. The laser should not be viewed with binoculars, telescopes or microscopes [112] [113].

The electrical safety is achieved by use of a protective metal housing, which is grounded by a separate protective earth conductor. As designed, the System meets leakage current guidelines. Furthermore, the System is equipped with a hospital-grade attachment plug. To ensure grounding reliability, connect it to an equivalent hospital-grade socket. The System is installed in accordance with local and National Electrical Codes. Other devices should not be connected to the System because this may result in leakage current exceeding safe levels. Extension cords should not be used to provide power to the System as it may compromise the Systems grounding reliability. Before connecting and disconnecting any cables, shut down the System and the control computer. Failure to do so

may result in damage to the System and computer [112] [113].

4.5 VivaScope® 1500

4.5.1 RCM images of normal *in-vivo* skin

In this Section images captured using the VivaScope® 1500 from the epidermis, DEJ and dermis layers are illustrated.

Epidermal layers:

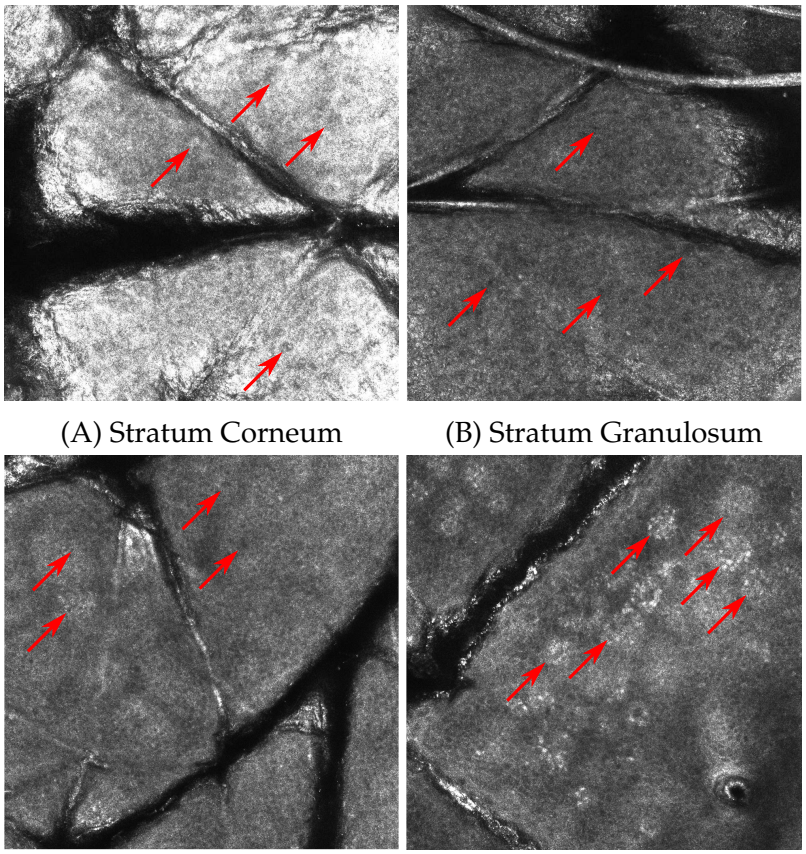
- Stratum Corneum (Figure 4.5A): approximately located $0 - 15\mu m$ below the skin surface and appears as a highly reflective surface surrounded by visible skin folds [114] . This layer is considerably brighter than the other epidermal layers because of back-scattered light caused by different refractive indexes at the interface between the microscope immersion medium (oil) and the skin [115]. The visible cells are referred to as corneocytes which are polygonal-shaped anucleated large cells, each with a dimension of $10 - 30\mu m$ [114]. When acquiring images using the VivaScope® 1500, the top most stratum corneum image is flagged set to the zero mark, i.e. the skin depth value is calibrated at this level.

- Stratum Granulosum (Figure 4.5B): approximately located $15 - 20\mu m$ below the skin surface and appears with large and sparsely located nuclei [65]. The granular cells are polygonal keratinocytes with a dimension of $25 - 35\mu m$ [116]. They have a well demarcated outline that form a grid that resembles a honeycombed pattern [115].

- Stratum Spinosum (Figure 4.5C): approximately located $20 - 100\mu m$ below the skin surface [65], having keratinocytes of size $15 - 25\mu m$ [116] arranged in a honeycombed pattern. As illustrated in Figure 4.5C it appears similar to the granulosum layer.

- Stratum Basalis (Figure 4.5D): located at an average depth of $50 - 100\mu m$ below the skin surface [65], and is characterised with basal cells. Basal cells are highly refractive due to the high refractive index of the melanin

present and are about $7 - 12\mu m$ [116]. In skin phototypes II-IV these cells are arranged in a cobblestone pattern - in skin phototype I the basal cells show low refractility and are difficult to see clearly [114] [65].



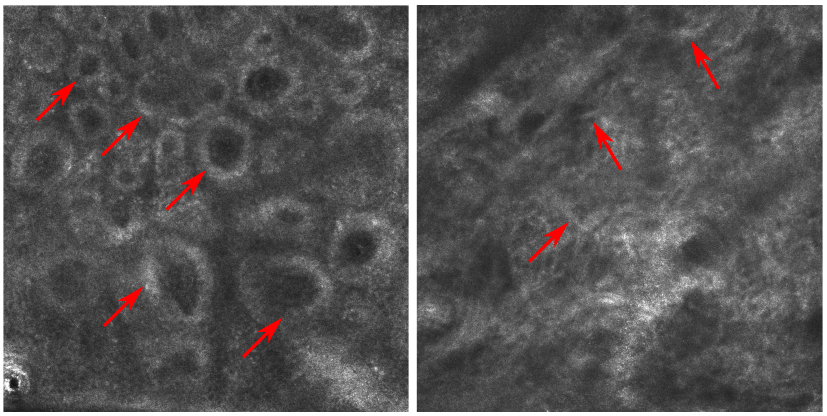
(A) Stratum Corneum (B) Stratum Granulosum

(C) Stratum Spinosum (D) Stratum Basalis

Figure 4.5: Confocal images of the epidermal layers of normal skin in a phototype III patient each covering an area of $500 \times 500\mu m^2$. The red arrows represent: (A) corneocytes which are anucleated large cells and polygonal-shaped, each with a dimension of $10 - 30\mu m$, (B) granular keratinocytes which are polygon shaped and each with a dimension of $25 - 30\mu m$, (C) kernatinocytes which are arranged in a honeycomb pattern, each with a dimension of $15 - 25\mu m$ and (D) basal cells which are highly refractive (due to melanin) and of dimension $7 - 12\mu m$.

Dermal-Epidermal Junction (DEJ) (Figure 4.6A): The dermis forms upward finger-like projections into the epidermis called papillae [114]. As the confocal sections scan deeper through the papillae, dark round oval structures or features corresponding to dermal papillae appear. These rings increase in size until neighbouring oval features touch each other tangentially, indicating that the base of the finger-like projections had been reached [114] [65].

Dermal Layers (Figure 4.6B): The dermal layers are located at an average



(A) Dermal-Epidermal Junction (B) Superficial Dermal layer

Figure 4.6: Confocal images of the DEJ and dermal layers of normal skin in a phototype III patient $[500 \times 500\mu m^2]$. The red arrows represent: (A) papillae which are circular in shape and (B) a network of fibres (diameter of $1-5\mu m$) and bundles (diameter of $5-25\mu m$).

depth of $100-150\mu m$ below the skin surface, visualising blood vessels which appear as dark tubular structures with differing refractive indexes [65]. Also, at average depths of $100-350\mu m$ below the skin surface, a network of bright fibres (diameter of $1-5\mu m$) and bundles (diameter of $5-25\mu m$) can be seen [114]. Figure 4.6B illustrates a bundle of fibres found in normal skin of a phototype III patient.

The above image examples were captured from non-diseased skin sites and thus do not display any atypical characteristics. Features associated with early melanoma include; disarray of the DEJ, papilla become less defined and atypical cells appear around papilla. As the melanoma advances additional features appear including; malignant cells occupy the entire DEJ and loose bright cells appear within the epidermis. Basal cell carcinoma characteristics include tightly packed cells with peripheral palisading combined with prominent blood vessels. On the other hand, features associated with psoriasis include non-rimmed papillae and an increase in capillary population at the DEJ, together with thickening of the epidermis [28]. Therefore, each skin condition or disease is characterised by a set of atypical features covering changes at the epidermis, DEJ and dermis layers.

4.5.2 Artefacts in RCM images of *in-vivo* skin

Optical contact between the RCM probe and skin can be removed by unavoidable features of skin anatomy, such as wrinkles, hair shafts, or unevenness in the skin surface. Additionally imaging artefacts can result from poor technique, such as the trapping of air bubbles in the coupling medium. Examples of these scenarios and methods that can be used to mitigate artefacts are highlighted in this Section.

To keep image artefacts to a minimum Mavig GmbH (the EMEA distributors of the VivaScope® System) do internal acceptance tests for laser power, field of view and flatness before installing a System. Additional tests are made to analyse the final image quality (Andre Philipp, Mavig GmbH, personal communication, 5th April 2017). Moreover, these tests are repeated on a yearly basis when a Mavig GmbH representative visits St. Thomas' Hospital to calibrate the System and test for image quality.

Furthermore, minor differences exist between inter-systems; however, efforts are made by Mavig GmbH to keep these differences to a minimum. These differences are minor and do not affect the clinical application of an automated algorithm. Nonetheless, differences in images can arise from the current state of the skin; for example, the same anatomical site can appear quite different across different time points depending on hydration and recent sun exposure (Andre Philipp, Mavig GmbH, personal communication, 5th April 2017). Thus the latter increases the variability of possible image representations and for any automated algorithm that will be applied in a clinical setting need to be able to cater for all possible image acquisitions.

4.5.2.1 Wrinkles and hair shaft artefacts

The main limitations with *in-vivo* RCM images are that when there is a wrinkle fold in the skin surface and when there are hair shafts, there is a lack of signal. Wrinkles are characterised by dark groves and reduce in darkness as one scans deeper in the skin, as seen in Figure 4.7. Wrinkles are normally present in

consecutive images in a stack, and in consecutive layers of a mosaic, and can remain present up to the dermal-epidermal junction. Hair shafts appear as bright threads and can be found in consecutive images in a stack and in consecutive layers in a mosaics, as seen in Figure 4.8. However, as can be noted, the visibility of the hair shafts drastically reduces across a depth of $15\mu m$. At a depth larger than $30\mu m$ below the skin surface, the hair shaft only appears as a shadow like structure. When comparing the wrinkles shown in Figure 4.7 to those in Figure 4.8, we can note that the wrinkles in the first example go deeper inside the skin, this is mainly due to the age difference of the volunteers.

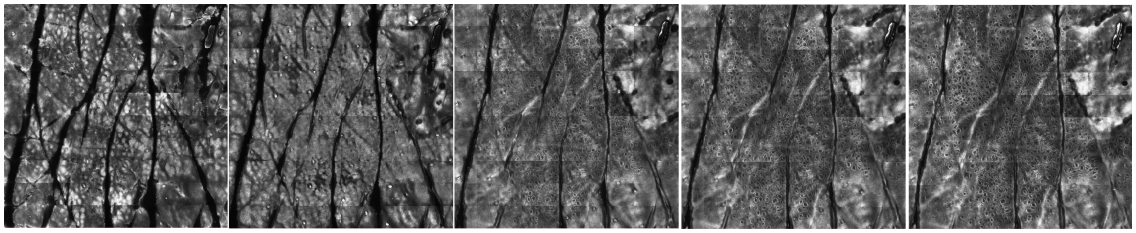


Figure 4.7: A mosaic of size 10×10 basic images, covering an area in the arm of $5 \times 5mm^2$, each consecutive mosaic is at a spacing of $15\mu m$ in depth. The volunteer is a type II skin type and is 50+ years old. As can be noted wrinkles appear as dark groves in the initial image, as the scan goes deeper inside the skin, the wrinkles are still present but as a less dark and smaller in width.

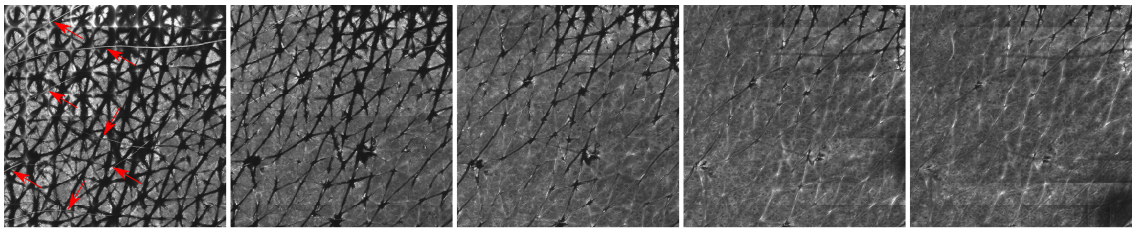


Figure 4.8: A mosaic of size 10×10 basic images, covering an area in the arm of $5 \times 5mm^2$, each consecutive mosaic is at a spacing of $15\mu m$ in depth. The volunteer is a type III skin type and is 25 years old. As can be noted hair shafts (pointed to with the red arrows) appear bright in the first two mosaics, after which they appear as a grey like shadow structure.

4.5.2.2 Artefacts due to unevenness in the skin surface

As the skin surface is naturally uneven it sometimes results in one end of a mosaic being inside the skin whilst the other is above the stratum corneum (as seen

in Figure 4.9). This is shown in the example illustrated in Figure 4.10, where the solid turquoise squares are imaging above the skin surface. This leads to consecutive mosaics imaging greater depths of the skin in these specific regions (dashed turquoise regions) than the general skin depth (dashed orange regions). This limitation can be mitigated by trying to align the adhesive window, metal ring and the VivaScope® 1500 microscope parallel to the skin surface as much as possible. However, there are cases that cannot be aligned due to anatomical restrictions (e.g: when imaging a protruding mole that is not flat) and when imaging areas which are at an angle (e.g: when imaging facial areas such as the nose or forehead). In the case that is illustrated in Figure 4.10, the region depicted by the solid green line in the dermoscopic image and in the confocal image, is slightly more elevated than the rest of the area, resulting in the area being imaged not being flat.

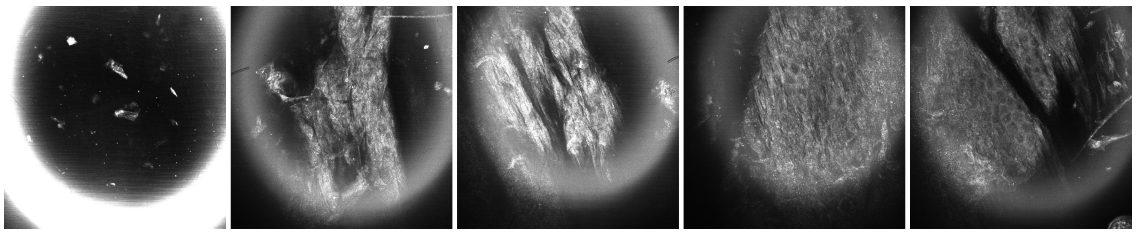


Figure 4.9: Five RCM basic images obtained above the skin surface. Each image covers an area of $0.5 \times 0.5mm^2$. When imaging above the skin surface there is a halo around the image which represents the metal ring that connects the VivaScope® 1500 to the skin surface. This halo effect is useful in helping to identify images that are not correctly positioned within skin.

4.5.2.3 Loss of coupling artefacts

When imaging for a duration of time on the same location, the adhesive window loses its connective properties, resulting in the images not being acquired from the correct (x, y) coordinates indicated on the dermascopic map. If one looks closely to the green area indicated in Figure 4.10, one can note that the region is shifting sideways along the horizontal x-axis across consecutive mosaics. A discrepancy of around one basic image ($\sim$0.5mm) can be noted between the first and last mosaic. When this happens it is either because of the adhesive window and metal ring sliding across the skin surface or due to different pressures exerted

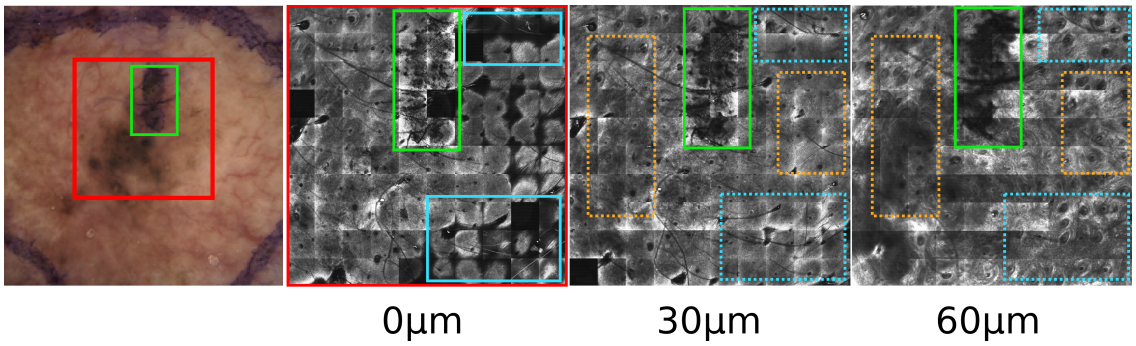


Figure 4.10: A mosaic of size 10×10 basic images, covering an area of $5 \times 5mm^2$, each consecutive mosaic is at a spacing of $30\mu m$ in depth. The dermoscopic image depicts the area that can be imaged by the confocalist, the chosen area that will be imaged with a mosaic is indicated by the solid red square. The green area depicts a part of the mole that is slightly elevated with respect to the rest of the area being imaged. The solid turquoise regions are imaging above the skin surface, resulting in the consecutive dashed turquoise regions imaging higher levels of the skin than the the regions indicated by the dashed orange regions.

by the volunteer across the scan. If correct correspondence between the dermoscopic and RCM images is critical, then the confocalist should halt the scanning once a misalignment is noted and restart scanning using a new adhesive window. Prior to resuming scanning the confocalist needs to ensure that the skin surface is cleaned.

4.5.2.4 Poor technique artefacts

Figure 4.11 show artefacts generated from air bubbles in the ultrasound gel or due to a poor imaging technique. These artefacts can be mitigated by putting the right amount of index matching fluid and applying a consistent amount of ultrasound gel across the adhesive window.

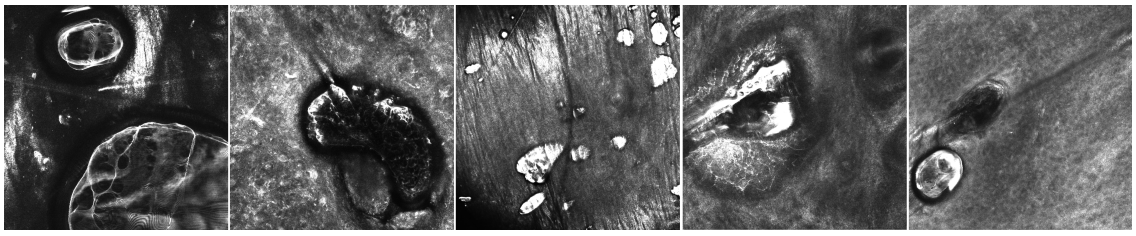


Figure 4.11: Five basic RCM images illustrating artefacts present due to air bubbles in the ultrasound gel and poor optical coupling in *in-vivo* RCM imaging, each covering an area of $0.5 \times 0.5mm^2$.

4.5.3 RCM videos of *in-vivo* skin

RCM has the capability of capturing video sequences of an area covering the basic image size, $0.5 \times 0.5mm^2$, at a frame rate of 9 frames per second (fps). Capturing time sequences of images from the same area gives the ability to image blood flow going through skin capillaries. As RBCs are highly reflective they appear as bright moving objects across RCM image sequences and will be referred to as flashes in this thesis. RBCs are best visualised at a depth not greater than $80\mu m$ below the skin surface. Although the depth that can be visualised using the *in-vivo* RCM is papillary dermis, corresponding to $150 - 200\mu m$ [15], beyond $80\mu m$ the vessels cannot be clearly distinguished and the blood flow going through is not visible.

The most superficial vessels in skin are the capillary loops located within each papillae. A protocol was established for this thesis to capture RCM sequences from these loops to ensure the sequences are all obtained in approximately the same plane. Initially the user needs to locate a $0.5 \times 0.5mm^2$ area which is characterised by capillaries. A video sequence is captured at this area. Once the tile location is decided upon, the user searches the k imaging depth to locate the depth at which the overall apexes of the capillaries are seen. The protocol requires the $30s$ image sequence to be captured in a plane located $20\mu m$ below the apexes of the visible capillaries.

As indicated in the RCM videos protocol, effort was made to capture the capillaries cross-sectionally in the middle, and thus have the capillary characterised by two separate circles, representing the arterial and venous loops of the capillary. This instance is indicated by line B in Figure 4.12, and the blood flow appears as flashes in two separate circles, as indicated by Figure 4.13 B. The arterial and venous loops can be differentiated as one having blood flowing towards and the other having blood moving away from the slice being scanned. The difference in direction can be distinguished in the intensity and frequency of the RBCs passing through them. In addition to the direction of movement, the loop displaying blood coming towards the slice represents oxygenated blood, whilst

the loop exhibiting blood moving away from the slice represents deoxygenated blood, thus both loops have a different haemoglobin distribution [34]. Also, an increase in haematocrit results in compact blood cell columns, characterised as a chain of RBCs without the presence of plasma in between the cells - resulting in the cells' boundaries not being individually distinguishable.

When capturing capillaries cross-sectionally at the apex of the capillary loop, where the arterial and venous limbs meet (line A in Figure 4.12), blood appears as flashes, as indicated in Figure 4.13 A. In other instances, capillaries are captured along their length (in-plane), as illustrated by the solid yellow line in Figure 4.12. Blood in in-plane capillaries appears as flashes similarly to those illustrated in Figure 4.13 C. As opposed to blood flow sequences in Figure 4.13 A and B, the flashes in C appear at different (x, y) coordinates of the image and make it harder to track. Clinicians find capillaries captured cross-sectionally easier to identify and track than capillaries captured in-plane. The two main reasons are: (1) in-plane capillaries have the flash located across different (x, y) coordinates at different frames, and (2) in-plane capillaries are surrounded by tissue and not by papillae, which is darker and gives a better contrast to the flash.

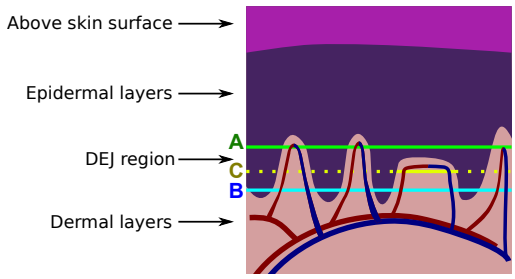


Figure 4.12: The different skin layers that can be visualised using RCM. The lines indicate at which levels capillaries can be captured at in RCM video sequences. Lines A and B illustrate two examples of capturing the capillary cross sectionally. The solid part of line C depicts an example of capturing the capillary in-plane.

The orientations of capillaries are highly variable, even in an area of a basic image of $0.5 \times 0.5mm^2$, one can have both capillaries oriented cross-sectionally and in-plane. A typical scenario is illustrated in Figure 4.14. When deciding upon the depth at which the video will be captured, one tries to maximise the amount of capillaries that are captured cross-sectionally in the middle, whilst

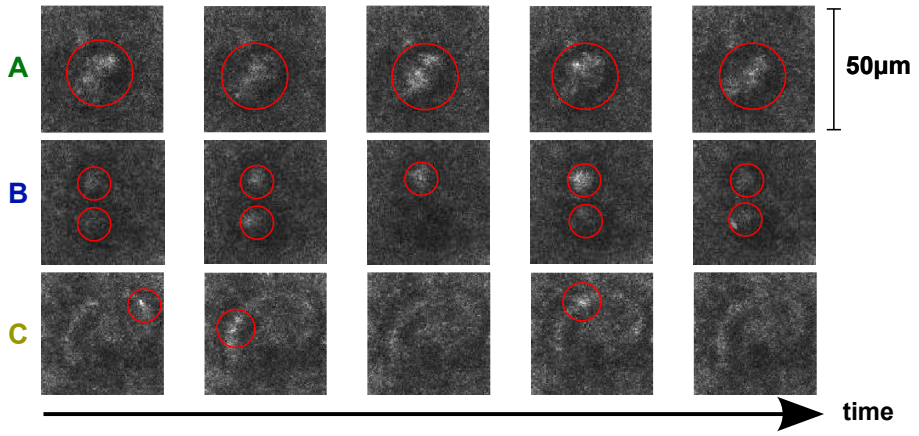


Figure 4.13: Skin capillaries can be captured at several orientations depending on whether the image captures the capillary cross-sectional or in-plane. (A) Capillary captured cross-sectional at its peak, (B) Capillary captured cross-sectional across the middle, (C) Capillary captured in-plane as illustrated by the continuous yellow line in Figure 4.12. Red circles indicate blood flow flashes.

minimising the number of capillaries captured at the peak or in-plane. In some cases, volunteers exhibit in-plane capillaries more frequently than cross-sectional capillaries, and thus have most videos characterised by capillaries imaged in-plane. In addition to this, sometimes the depth at which the capillaries are located is deeper than average, at a depth which cannot be clearly imaged by the RCM. In such cases, a preference to imaging the capillaries at their peak is given over imaging the capillaries cross-sectionally in the middle with reduced image quality.

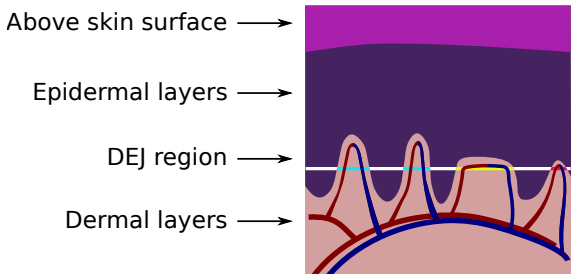


Figure 4.14: The white line indicates a typical scan, where capillaries are captured both cross-sectionally and in-plane. The depth of the white line was decided upon as it is located $\sim 20\mu m$ below the peak of the first two capillaries. However, at this depth another two capillaries emerge. The turquoise regions indicate cross-sectional capillaries that will be represented by two circular structures. The pink line indicate a capillary that was captured at its peak. The yellow line indicates a capillary captured in-plane.

RCM videos captured with the VivaScope® 1500 technology have a frame

rate of $9fps$. This technology does not provide a mechanism of improving the frame rate by capturing a smaller area than $0.5 \times 0.5mm^2$. The longest video that can be acquired is 1 minute, if a user would like to take a longer video, one can capture consecutive videos and concatenate them together.

4.5.4 Artefacts of RCM video sequences

The main artefact in RCM video sequences is the motion present across the duration of the video. In Figure 4.15 motion artefact is shown in four frames of a video. In this example a wrinkle was chosen to illustrate the movement, as it is the easiest skin structure to track visually across time. The location of the red lines, which correspond to the wrinkle location in the first frame, are kept constant across the four frames of Figure 4.15. When contrasting the first and fourth frame, it can be seen that the wrinkle moved towards the right and the wrinkle is not within the line boundaries in the fourth frame.

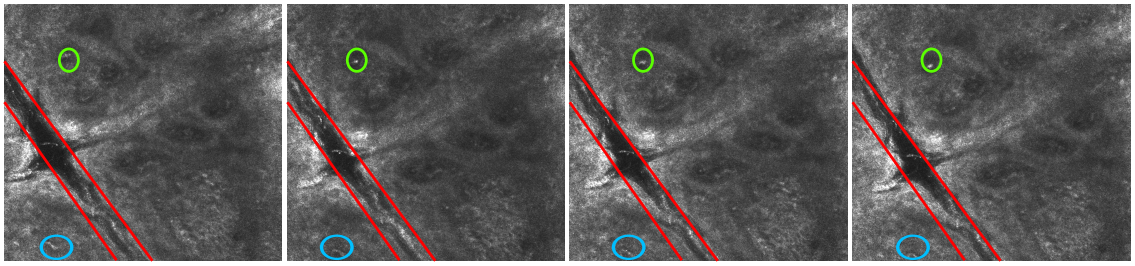


Figure 4.15: 4 frames from the same video covering the same spatial area at 10 second intervals. The red lines depict the wrinkle location in the first frame and the location of these lines are kept constant across all frames. If we analyse the wrinkle encompassed in the red region, we can note that it is moving towards the right across with time. The green and turquoise represent two capillaries captured at the peak and in-plane respectively.

4.5.5 Power Alteration in RCM video sequences

This thesis aims to detect and quantify capillaries across different skin types. The literature does not suggest that there is a difference in the amount of skin capillary per unit area with skin type, however, it has been noticed that fewer skin capillaries are visible in darker skin tones. One explanation for this is that

RBCs are less visible in darker skin tones due to the high reflectance of melanin at the periphery of papillae. To compensate for this, the power value was manually altered for volunteers with skin types IV-VI. As illustrated in Figure 4.16, as the power was increased from $0.6mW$ to $1.8mW$, the signal from RBCs was amplified together with the signal refracted from papillae.

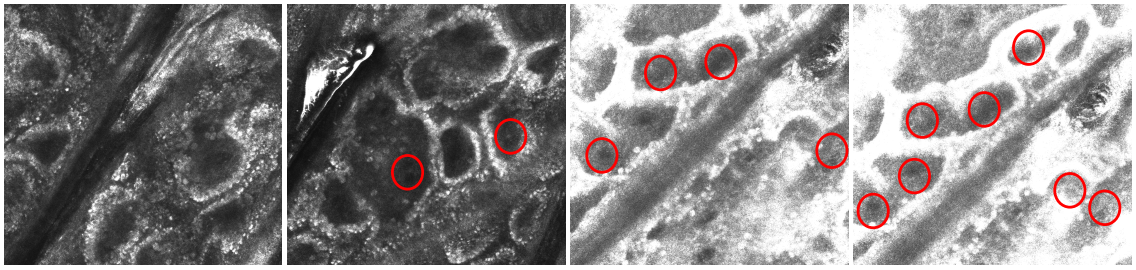


Figure 4.16: An unexposed forearm of a volunteer of skin type VI. As the RCM power was increased, more skin capillaries with RBCs going through them were noticed. The red circles indicate the regions which show RBCs inside them. The disadvantage of increasing the power is that the melanin over-saturates easily with an increase in power.

4.6 VivaScope® 2500

4.6.1 Mohs surgery and RCM images of excised tissue

In the clinical setting, RCM is not as widely used to examine excised tissue of the skin as it is used to analyse *in-vivo* skin. However, as illustrated in Chapter 3, there is an increasing interest in its role of imaging freshly excised tissue during an operation to examine whether the tissue is diseased. In this thesis tissue excised during Mohs surgery is examined.

There is not one standardised way of preparing excised tissue for RCM imaging in the literature. In this section the protocol used by St. John's Institute of Dermatology at St. Thomas' Hospital will be described. The images used in this thesis were provided by this clinic and collected during Mohs surgery. The images were only collected for research purposes and no diagnoses were made based on the RCM images. Similarly to the study identified in [117], RCM imaging was done on the discarded frozen tissue leftover from the Mohs surgery. The

RCM imaging protocol of excised tissue was designed around the Mohs surgery protocol and it ensured that no disruptions occur to the surgery procedure.

Mohs surgery is a widely used procedure to remove skin cancers in the facial and other aesthetically sensitive body sites. The main objectives of Mohs surgery is to achieve histological clear margins. When removing skin tumours, dermatologists also excise a small margin around the tumour to ensure the tumour is completely removed and it does not reappear. However, as opposed to the standard excision of skin cancers, which takes a 4 to 6 mm margin, Mohs surgery only uses a 1 to 1.5 mm margin. This margin is analysed and if cancerous cells are present a further small excision is performed near the diseased region. This procedure is repeated until no further cancerous cells are found.

The Mohs technician does not need all the excised tissue, and generally discards a section of the frozen tissue at each excision. A frozen section of tissue excised during Mohs surgery is shown in Figure 4.17. A confocal technician used the discarded part of this tissue for imaging using RCM. Before imaging, the tissue needs to be stained with a fluorescent dye using the staining technique defined below:

- Put the tissue in acetowhitening 10% citric acid for 1 minute
- Followed by buffered saline for 1 minute
- Followed by the fluorescent dye 10% acridine orange for 1 minute
- Finished by buffered saline for another 1 minute to remove the excess dye

After staining, the tissue is placed between two microscopic slides. This manoeuvre flattens out the tissue, resulting in better quality images as most or all the tissue can be imaged once the tissue is flattened out. A drop of ultrasound gel is placed above the tissue on the double-decker slide, on the side that is going to be imaged. Then the double slide is is placed on the microscope and screwed to it to make it steady, as illustrated in Figure 4.18.

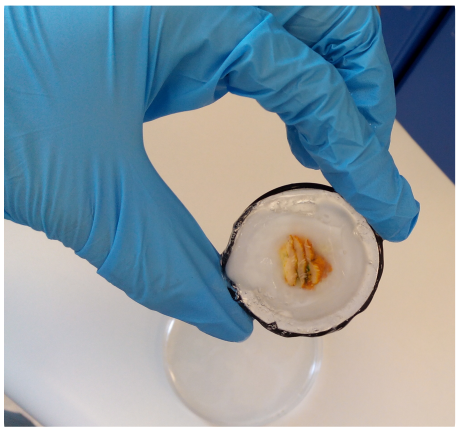


Figure 4.17: A frozen representation of tissue excised during Mohs surgery. Sections of this sample are stained and mounted on slides for histological and RCM examinations.

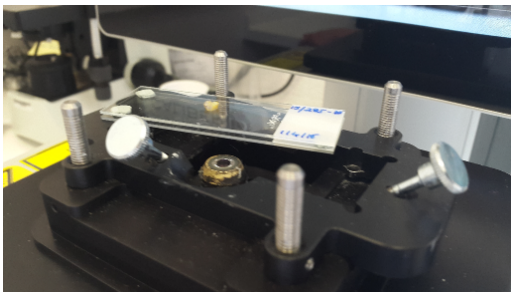


Figure 4.18: The double slide placed on the microscope which in turn is screwed to the microscope using the two screws. In cases of thicker tissues it is difficult to tighten the screws to the top slide, and thus when imaging thicker tissues the screws are fixed to the bottom slide leaving the top slightly unsupported. This arrangement can result in poor quality images, especially at the epidermis side.

4.6.2 Artefacts in RCM images of excised tissue

The main artefacts present in RCM images of excised tissue occur due to either poor distribution of fluorescent dye or parts of the tissue not being imaged.

4.6.2.1 Concentration of fluorescent dye

In some cases fluorescent dye is concentrated in one region due to poor staining. In such cases, and as illustrated in Figure 4.19A, it is difficult for the clinician to make a diagnosis from the bright structures in the image. Therefore the confocal technician retakes the image with a lower laser power (Figure 4.19B).

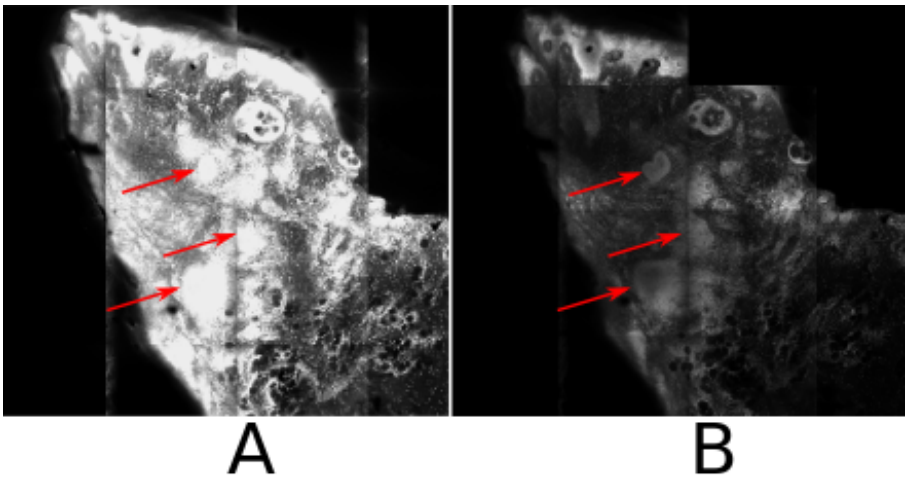


Figure 4.19: Fluorescent dye concentrated at the tip of the tissue and makes it difficult to distinguish whether the structures highlighted by the arrow are tumour or not. By using a lower laser power, the confocalist confirmed that these regions are tumour.

4.6.2.2 Epidermis visibility

The epidermis is not always visible when imaging excised tissue. This is of particular concern when making a diagnosis as the epidermis needs to be checked to verify that it is tumour free. Superficial BCC normally originates along the epidermis and thus if the epidermis is not captured in the image it cannot be checked for superficial BCC or other tumours.

The epidermis is not always visible as the dermis end of the tissue is thicker, which results in the epidermis not being flattened evenly between the slides. The conofocal technician collecting the images attempted to put a slight pressure on the epidermis side of the slide to enhance the visibility of the epidermis, as illustrated in Figure 4.20. In some cases this methodology results in better visibility of the epidermis, as shown in Figure 4.21. However, this does not always work and in some cases does not guarantee the full visibility of the epidermis around the tissue specimen. An advantage of using frozen tissue rather than fresh tissue is that the frozen sectioning flattens out the tissue. However, frozen sectioning also results in parts of the tissue being lost in the process.

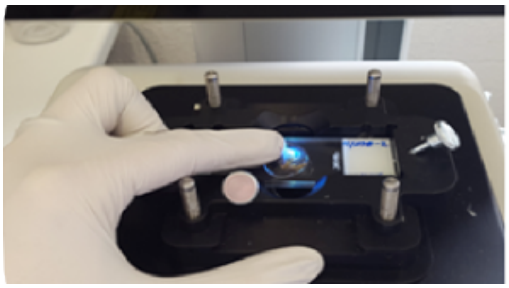


Figure 4.20: The confocal technician applying pressure to the epidermis end of the tissue. This methodology reduces the thickness between the microscopy slides at the epidermis end, resulting in a better image representation of the epidermis.

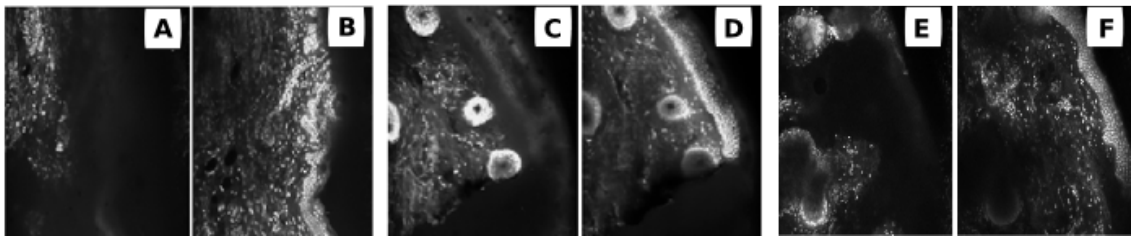


Figure 4.21: Three examples of imaging the epidermis before (A, C, E) and after (B, D, F) applying pressure at the epidermis end. A significant improvement in the visibility of the epidermis can be noticed after applying pressure.

4.6.2.3 Low signal regions

Low signal regions, indicated by the arrows in Figure 4.22, can be due to missing tissue, which also correspond to a hole in the histology slide. However, they can also be due to the dye being taken up by a higher depth-value of the tissue. Thus, when faced with low signal regions the technician needed to confirm whether they correspond to missing tissue or to tissue that cannot be visualised at that depth with that particular laser. This was confirmed by going inside the region with different laser powers in a stack manner. In some cases these regions corresponded to tumour, due to the dye being absorbed by the top layer of the tumour due to the nucleated nature of the tumour.

4.6.3 Comparing histology to RCM images

Basal cell carcinoma (BCC) can be sub-divided in classes, including: infiltrative, micronodular, nodular and superficial BCC. In this section these different classes are illustrated and both histology and RCM image representations are given. An

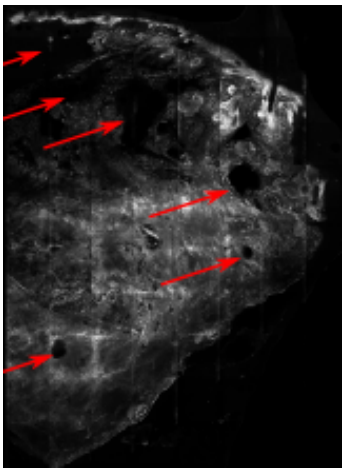


Figure 4.22: Low signal regions in RCM images can either correspond to actual missing tissue or due to the dye not penetrating parts of the tissue.

example of infiltrative BCC is shown in Figure 4.23 and as can be shown infiltrative BCC is characterised by BCC strands of different shapes and sizes. Examples of micronodular and nodular BCC are represented in Figures 4.24 and 4.25 respectively. These two types of BCC are similar in shape and appearance. However, micronodular BCC is made up of smaller components clustered together, whereas nodular BCC normally is exhibited as larger nodules, located sparsely across the tissue. Superficial BCC, depicted in Figure 4.26, is located closely to the epidermis and protrudes into the papillary dermis.

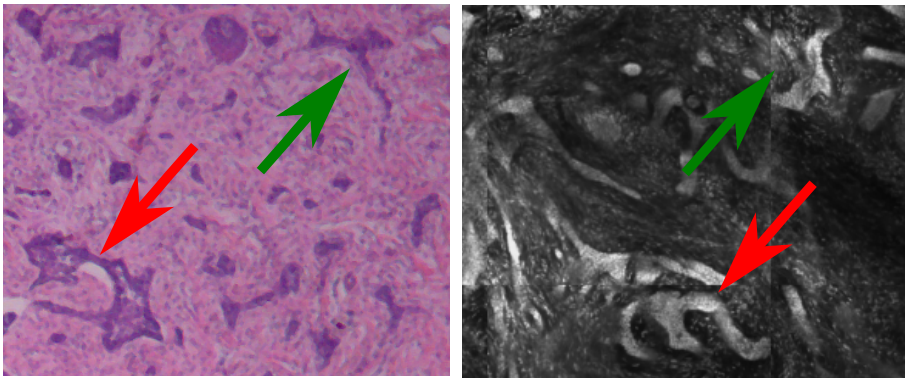


Figure 4.23: Histology and RCM image representations of an infiltrative BCC. These two images were collected from the same tissue, the images are extracted from tissue spacing difference of $\sim 5 - 10\mu m$. The shapes of infiltrative strands are not consistent across the different strands.

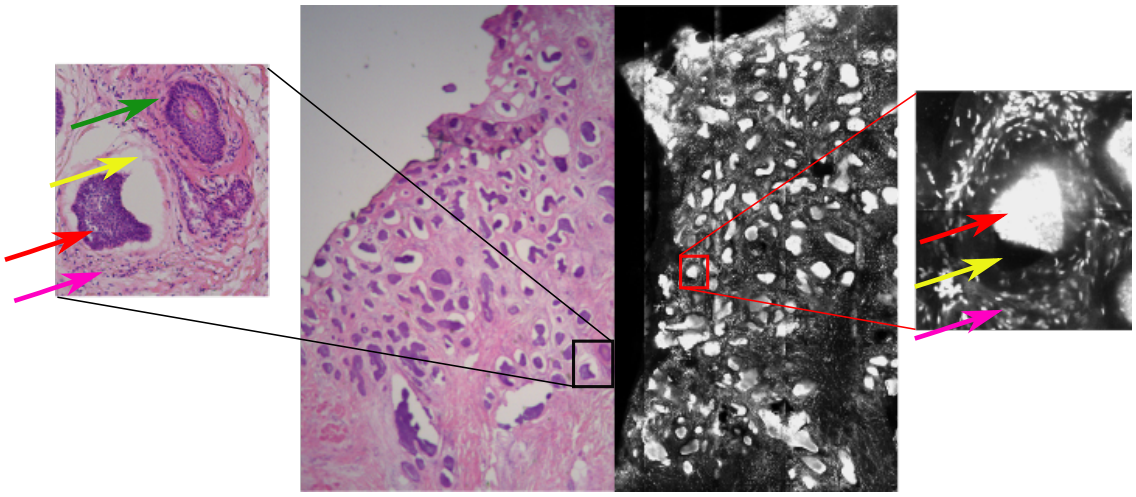


Figure 4.24: Histology and RCM image representations of a micronodular BCC. These two images were collected from the same tissue, the images are extracted from tissue spacing difference of $\sim 5-10\mu m$. The red arrow indicates tumour, surrounded by clefting (yellow arrow) which is represented by a white and black region in the histology and RCM images respectively. The tumour is normally surrounded by inflammation (pink arrow) which encapsulates the tumour. The structure with a green arrow is a hair follicle.

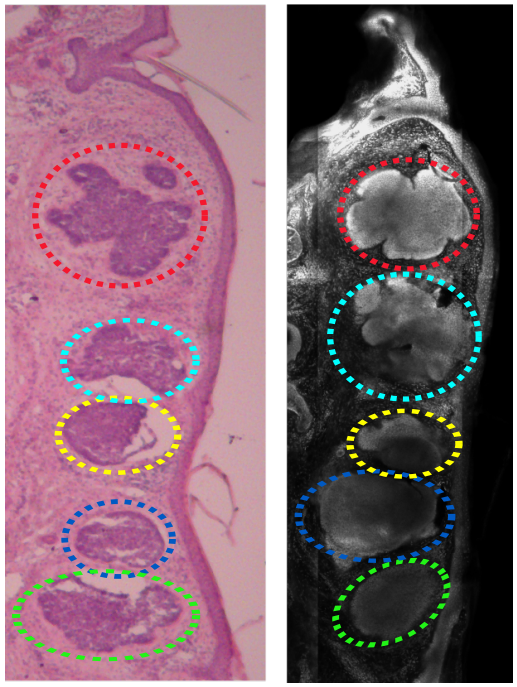


Figure 4.25: Histology and RCM image representations of a nodular BCC. These two images were collected from the same sample, with a tissue spacing difference of $\sim 5-10\mu m$. The circular colour matched circles represent the same tumour. Nodular BCC islands are larger than infiltrative BCC islands.

4.7 Conclusion

The main conclusion of this Chapter is that RCM has some remarkable capabilities to image skin. The potential of imaging *in-vivo* skin has been illustrated, in

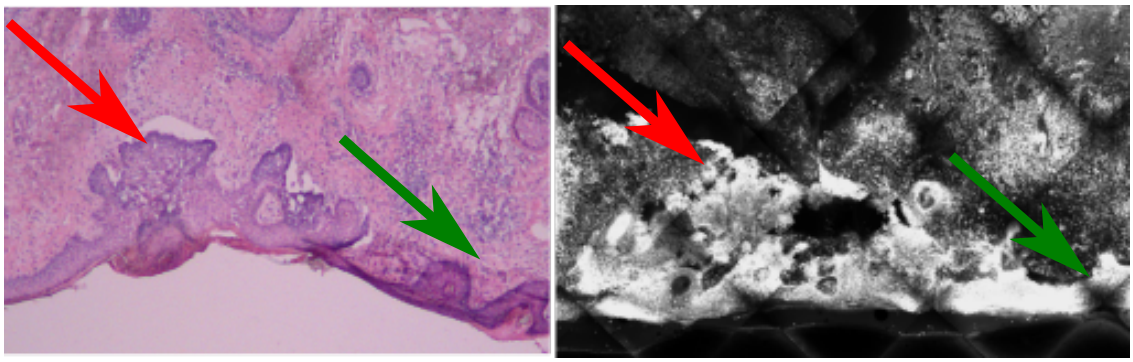


Figure 4.26: Histology and RCM image representations of a superficial BCC. These two images were collected from the same tissue, with a tissue spacing difference of $\sim 5 - 10\mu m$. The red arrows indicate tumour whereas the green arrows indicate healthy epidermis. As can be noted, superficial BCC originates from the epidermis and progresses into the papillary dermis.

particular the capability of imaging the flow of RBCs as they travel through a capillary loop has been highlighted. For this reason, automated algorithms that analyse RCM images and of detecting and quantifying skin capillaries are developed in this thesis. Also, the facility of imaging BCCs in excised tissue have been well documented. In this thesis an algorithm that analyses such images acquired during Mohs surgery is designed and developed. Both images acquired *in-vivo* and *ex-vivo* contain artefacts that can mislead automated algorithms. Therefore, careful attention to artefact removal during image acquisition is a critical component to this.

Chapter 5

Automatic localisation of the dermal-epidermal junction in reflectance confocal microscopy image cubes

Contents

5.1 Introduction 108

5.2 Algorithmic Method 109

5.2.1 Introduction 109

5.2.2 Layer classifier 111

5.2.3 *Find DEJ region* algorithm 123

5.3 Experimental Method 125

5.3.1 Data Acquisition 127

5.3.2 Preparing the data for layer classification 128

5.3.3 Selection of monogenic filter parameters 131

5.3.4 Evaluating the layer classifier 134

5.3.5 Evaluating the *find DEJ region* algorithm 135

5.4 Results 136

5.4.1 Intra-volunteer Layer Classifier 137

5.4.2 Inter-volunteer Layer Classifier 138
5.4.3 2D B-spline representation of results 140
5.5 Discussion 143
5.6 Future Work 145
5.6.1 Quantification of papillae population and melanin . . . 146
5.6.2 Enhancing features 147
5.6.3 Suppressing confounding structures 148
5.6.4 Purer attributes and multi-labels 148
5.7 Conclusion 149

5.1 Introduction

The aim of the work described in this Chapter is to describe the development and implementation of an algorithm which can both automatically locate the dermal-epidermal junction (DEJ) and quantify the DEJ thickness. The DEJ is of great interest to dermatologists, as skin cancers such as melanoma and basal cell carcinoma first appear here, and subsequently progress laterally or vertically as the disease becomes more aggressive [1].

The automatic location of the different skin layers has the potential to speed up the current procedure undertaken by clinicians to analyse RCM images. For example, if the clinician wants to focus their examination on changes occurring at the DEJ level, then the algorithm can supply the depth (μm) at which the DEJ is located and the clinician can immediately start analysing the subset of images within this range. Similarly, the developed algorithm could act as a platform to highlight the different layer depths to help train clinicians new to RCM.

Furthermore, an algorithm that automatically differentiates between different skin layers can be used as an initial step for subsequent automated algorithms. For example, an algorithm that aims to quantify a cell architecture located at a particular skin layer, can make use of the developed algorithm as a preprocess-

ing step to discard images not falling within the layer of interest.

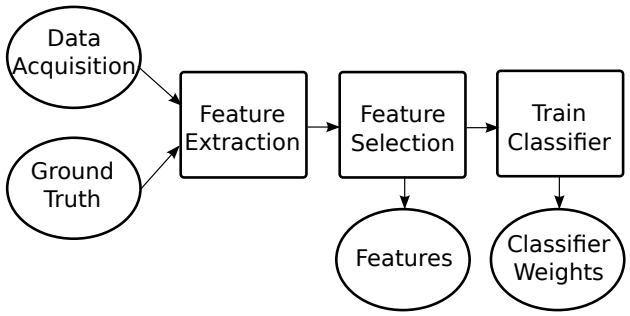
An algorithm was developed to classify RCM images captured using standard clinical protocols into the three distinct layers of the skin and is defined in Section 5.2. The key steps needed to design a classifier are (i) feature extraction, (ii) feature selection and (iii) classification. Both texture and morphologically based features are discussed and identified in the feature extraction section. The experimental method of this Chapter is presented in Section 5.3, which will cover the data collected, implementation of the algorithm, and validation techniques used. The results achieved are tabulated in Section 5.4 and discussed in Section 5.5. Possible future extensions to this experiment and conclusions are identified in Sections 5.6 and 5.7 respectively.

5.2 Algorithmic Method

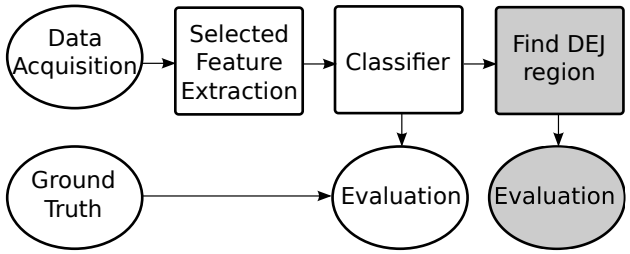
5.2.1 Introduction

In this section the algorithmic method developed for the purpose of locating and evaluating the DEJ is presented. This algorithm is segmented into two parts: (1) building a layer classifier algorithm to distinguish between skin layers using training data (Figure 5.1a) and (2) evaluating the classifier using testing data to extract quantitative information of the DEJ using the *find DEJ region* algorithm developed for this purpose (Figure 5.1b).

In Section 5.2.2 the layer classifier algorithmic method used to discriminate between the different skin layers is discussed. The motivation of the algorithm is to distinguish images classed as epidermal, DEJ and dermal. The three principle building blocks needed to perform classification will be thoroughly discussed: feature extraction covering both texture and anatomical parameters (Sections 5.2.2.1 and 5.2.2.2), feature selection (Section 5.2.2.3) and classifier algorithms (Section 5.2.2.4). In order to train the classifier a training set of images is acquired during the data acquisition stage together with the ground truth (labels set by myself) of each image. Thus, for each image a ground truth label, which



(A) Training the classifier



(B) Testing the classifier and applying the *find DEJ region* algorithm

Figure 5.1: A flowchart illustrating the image processing steps needed to perform layer classification (white) and those needed to find the dominant DEJ slice (grey). Oval shaped blocks illustrate input/output, whilst rectangular shaped objects indicate algorithmic processing. In stage (a) the classifier is trained on training datasets, these datasets are independent from the testing dataset used in stage (b) to test the classifier. The feature output in the training stage (a), indicates which features need to be extracted in the testing stage (b), whilst the classifier weights derived in the training stage are the weights used by the classifier in the testing stage.

indicates whether the image is of class epidermal, DEJ or dermal, is given. Before assigning ground truth labels, I have completed a one week in-person training course and online training exercises which certify that I am capable of interpreting RCM images. These courses were organised by Mavig GmbH, which is the company that distributes RCM VivaScope® systems in the EMEA region.

The classifier is evaluated on data which was not available to the classifier at the training phase. To summarise, the classifier is trained using a training data set, whilst it is evaluated on a different data set, referred to as testing data. The testing data is further evaluated using the algorithm *find DEJ region*, Section 5.2.3, which makes use of classification results to extract parameters of the DEJ (e.g. its

location and thickness).

5.2.2 Layer classifier

In this Section the steps involved in designing the classifier, identified in Figure 5.1(a), are described in detail. This Section will focus on the training side of the classifier, which will be eventually tested on testing data and discussed in Section 5.2.3.

As briefly outlined, the input needed to train a classifier is a dataset of images, together with the ground truth of each image. For the purpose of this Chapter, the images are classed by myself into either epidermal, DEJ or dermal, and the classifier will aim to differentiate between these three classes. Cubes of image data are used in this Chapter and are acquired using the RCM VivaScope®1500, refer to Section 4.3 (Chapter 4).

As illustrated in Section 4.5.1, different skin layers exhibit heterogeneous patterns and structures in RCM images. In this Chapter a wide range of features will be used to discriminate and characterise the different properties present in the range of patterns and structures. Texture and anatomical based features are extracted from each individual image (tile) with the aim of quantifying large scale structures, and can be sub-categorised into:

- Texture-based features (Section 5.2.2.1):
  - Spatial domain: histogram and co-occurrence based features;
  - Frequency domain: wavelet and Fourier transform based features.
- Anatomical-based features (Section 5.2.2.2): incorporating quantitative measures of the presence or absence of papillae that are located at the DEJ. These types of features have not been previously used to help automated DEJ location.

The feature extraction function, f, extracts n features from each image (tile) and these are represented by $(a_1, a_2, a_3, .., a_n)$. The features extracted from all tiles

were represented in a two dimensional matrix A, where each tile has a feature vector which characterises it. The vector indexed $[i + (j-1)N_i + (k-1)N_iN_j]$ represents the n extracted features of tile T located at (i, j, k) in the cube of images. $N_i \times N_j$ is the size of the mosaic and k represents the depth at which the mosaic is taken (refer to Section 4.3 to link notation to data acquisition methods):

$$A_{\{i+(j-1)N_i+(k-1)N_iN_j\}} = f(T_{ijk}) = (a_1, a_2, a_3, .., a_n)_{ijk} \tag{5.1}$$

Using features extracted from multiple tiles the most prominent features that are capable of discriminating between the three classes are selected using a feature selection algorithm (Section 5.2.2.3). The selected features, together with the ground truth associated with each set of features, are fed to train a classifier (Section 5.2.2.4). The output of the classifier are classifier weights that will be used to assign a class to unseen images (test data).

5.2.2.1 Texture based features

The texture features extracted for this Experiment are summarised in Table 5.1. The purpose of the texture features is to enable discrimination between different skin layers. As illustrated in Section 4.5.1, as images are acquired in the deeper dermal layers of the skin, the tissue structure is characterised by collagen fibres [15] (Figure 4.6B). This contrasts with the level of cell detail found in images acquired from epidermal layers (Figure 4.5). Quantification of this difference is used to classify the different layers.

Histogram based features Histogram features are based on the probability density function $h(g_i)$ of the grey levels g_i:

$$h(g_i) = \frac{N(g_i)}{K \times L} \tag{5.2}$$

where $N(g_i)$ is the number of pixels with value g_i (ranging from 0 to 255) and $K \times L$ is the size of the tile (total number of pixels).

Feature Domain	Based on	Details
spatial	histogram	mean value, mean square deviation, skewness, kurtosis and entropy of the grey level distribution
spatial	co-occurrence	angular second moment, difference moment, correlation, sum of squares, inverse difference moment, summed average, sum variance, entropy, sum entropy, difference entropy
frequency	Fourier transform	Fourier power spectrum at each frequency band
frequency	wavelet transform	standard deviation, energy, entropy at each frequency band

Table 5.1: Texture features used in this Chapter to classify RCM images based on Wiltgen et al. [6]

The mean value (m_B), mean square deviation (f_1), skewness (f_2), kurtosis (f_3) and entropy (f_4) of the grey level distribution are calculated using the below equations, which make use of values g_i and $h(g_i)$ calculated above. Variable i equates to the number of intensity values available [6]:

$$m_B = \sum_i g_i h(g_i) \tag{5.3}$$

$$f_1 = \sum_i (g_i - m_B)^2 h(g_i) \tag{5.4}$$

$$f_2 = \sum_i (g_i - m_B)^3 h(g_i) \tag{5.5}$$

$$f_3 = \sum_i (g_i - m_B)^4 h(g_i) \tag{5.6}$$

$$f_4 = (-1) \sum_i h(g_i) log_2(h(g_i)) \tag{5.7}$$

Co-occurrence based features Another interpretation to analyse intensity values is by using the co-occurrence matrix, which is based on the distribution of the grey levels of adjacent pixels as indicated in Equation 5.8.

$$w_{ij} = \frac{a_{ij}}{\sum_i \sum_j a_{ij}} \tag{5.8}$$

where the variable a_{ij} is the number of different grey level combinations $(g_i g_j)$ between two adjacent pixels, e.g. the actual pixel being considered and the pixel to its right, in the tile.

The angular second moment (f_5), difference moment (f_6), correlation (f_7), sum of squares (f_8), inverse difference moment (f_9), summed average (f_{10}), sum variance (f_{11}), entropy (f_{12}), sum entropy (f_{13}) and difference entropy (f_{14}) are calculated as indicated below using the following sum of row, columns and diagonals [6]:

$$P_x(i) = \sum_j w_{ij} \tag{5.9}$$

$$P_y(j) = \sum_i w_{ij} \tag{5.10}$$

$$P_{x+y}(k) = \sum_i \sum_{j_{i+j=k}} w_{ij} \tag{5.11}$$

$$P_{x-y}(k) = \sum_i \sum_{j_{i-j=k}} w_{ij} \tag{5.12}$$

$$u_i = \sum_i g_i P_x(i) \tag{5.13}$$

$$u_j = \sum_j g_j P_y(j) \tag{5.14}$$

$$\sigma_i = \sum_i (g_i - \mu_x) P_x(i) \tag{5.15}$$

$$\sigma_j = \sum_j (g_j - \mu_y) P_y(j) \tag{5.16}$$

$$f_5 = \sum_i \sum_j w_{ij}^2 \tag{5.17}$$

$$f_6 = \sum_k k^2 P_(x+y)(k) \tag{5.18}$$

$$f_7 = \frac{1}{\sigma_i \sigma_j} (\sum_i \sum_j (g_i - g_j) w_{ij} - \mu_x \mu_y) \tag{5.19}$$

$$f_8 = \sum_i \sum_j (g_i - m_B)^2 w_{ij}^2 \tag{5.20}$$

$$f_9 = \sum_k \frac{P_{x+y}(k)}{1+k^2} \tag{5.21}$$

$$f_{10} = \sum_k k P_{x+y}(k) \tag{5.22}$$

$$f_{11} = \sum_k (k - F_{10})^2 P_{x+y}(k) \tag{5.23}$$

$$f_{12} = (-1) \sum_i \sum_j w_{ij} log_2 w_{ij} \tag{5.24}$$

$$f_{13} = (-1) \sum_k P_{x+y}(k) log_2 P_{x+y}(k) \tag{5.25}$$

$$f_{14} = (-1) \sum_k P_{x-y}(k) log_2 P_{x-y}(k) \tag{5.26}$$

Fourier based features Fourier analysis breaks down a signal into constituent sine and cosine functions of different frequencies. When Fourier analysis is applied to an image, the image is decomposed into its frequency parts containing information about global properties. In this Chapter the Fourier power spectrum is used as a feature which characterises the global properties of texture pattern. The Fourier power spectrum is calculated from the Fourier decomposition of the image, by summing up the spectral values at circles with variable radii around the origin (of the decomposition). In general, an amplitude peak at a low frequency indicates a smooth texture, while an amplitude peak at higher frequencies indicates structured texture [6].

Wavelet based features In image analysis the discrete wavelet transform (DWT) is often used which transforms the image into approximations (A) and details (D). Approximations represent the high-scale, low-frequency components of the signal, while the details represent the low-scale, high-frequency components as illustrated in Figure 5.2. The decomposition process can have a number of levels (iterations), each iteration resulting in a lower resolution component. The example illustrated in Figure 5.3 exhibits a three level wavelet decomposition tree [118]. The advantage of decomposing the image into a number of levels is that different levels of detail can be examined.

For the purpose of this Chapter a wavelet decomposition of 4 levels was performed on each basic image using the bior3.7 wavelet in MATLAB. Quantitative values, including standard deviation, energy and entropy (Table 5.1) were calculated from each approximation and detail representation of each level.

5.2.2.2 Anatomical based features

In this Section the methodology used to incorporate morphological information to aid classification is described. These features give quantitative parameters of the presence (or absence) of papillae in an image by incorporating information across the imaging depth. As illustrated in Figure 4.6A, the DEJ is characterised

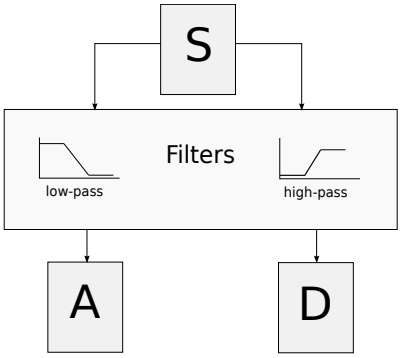


Figure 5.2: Approximations and Details in wavelet analysis

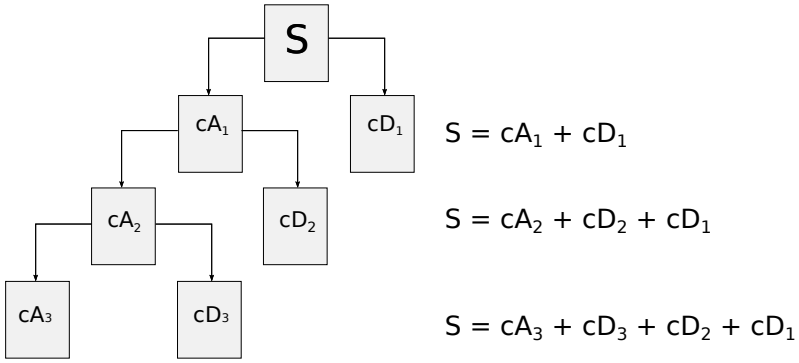


Figure 5.3: Three level wavelet decomposition tree

by circular bright halos which represent the melanin of the papillae imaged cross-sectionally. As the DEJ is sandwiched between the epidermal and dermal layers, an image having papillae has both epidermal and dermal like structures. The appearance of the DEJ depends upon the depth at which the DEJ is imaged, and the DEJ depth fluctuates greatly across different body sites (intra-patient variability) and across different volunteers (inter-patient variability).

Initially the edges of the basic images (tile) were enhanced using a monogenic filter, the output of which was fed into a circle finding algorithm to detect papillae like structures. These two steps are now described in more detail.

Monogenic filtering To enhance the visibility of papillae a monogenic filter [119] [120] was applied to each basic image. As opposed to classical edge detection approaches, which focus on high gradient (step) changes in intensity values, the monogenic filter encapsulates the fact that an edge can be represented somewhere between a step and a line [121]. The monogenic filter has been used in [122] to detect bright circular objects in ultrasound images. It was chosen for

use here as ultrasound images appear similar to RCM images in that they show a speckle pattern, have low contrast and low signal-to-noise ratio [123].

The two main parameters that define which edges are found by the monogenic filter are the wavelength which defines the width of the edge and σ which represents the width of the Gaussian kernel used as a bandpass filter. The output of the monogenic filter of the DEJ image in Figure 5.5A is illustrated in Figure 5.5B.

Hough circular detection Papillae are searched for in the enhanced edged representation using the Hough circular transform, the pipeline of which is shown in Figure 5.4. Firstly, the gradient of the edged image is calculated and pixels with a non-zero gradient are candidate pixels that are allowed to vote in the Hough circular transform accumulator array. The non-zero gradients of the edged papillae representation indicated by the red box in Figure 5.5B are shown in Figure 5.5C.

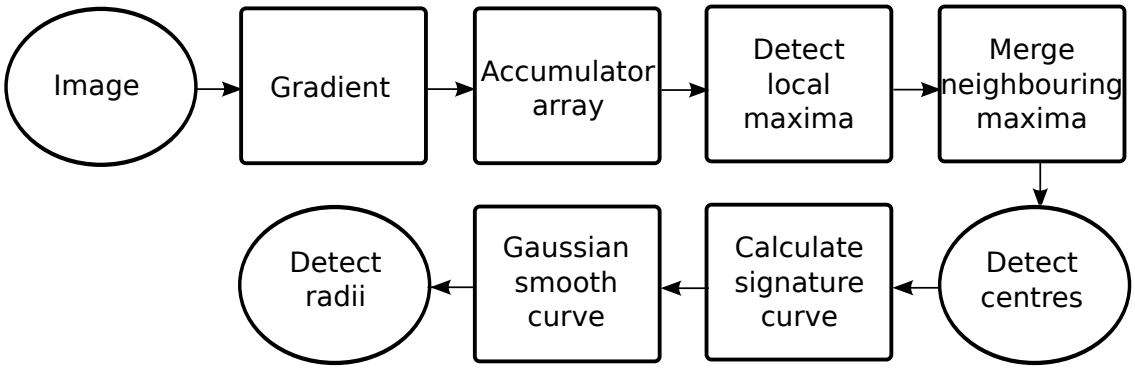


Figure 5.4: The gradient representation of the original image is calculated. Each non-zero gradient is a candidate to vote in the accumulation array, where the weight of each vote is determined by the pixel intensity. Local maxima regions within the accumulator array were searched for and neighbouring local maxima regions are merged together in one maximum, from which the circles' centres are calculated. For each detected centre, a signature curve is calculated which is initially Gaussian smoothed before being used to calculate the radius (or radii) associated with the centre. The output of the Hough circular detection algorithm is a list of the detected centre coordinates and the associated radii of each centre.

The gradient image is converted into an accumulator array through the mechanism of each non-zero gradient being a candidate to vote. The weight of each vote is determined by the pixel intensity, and thus high intensity locations are

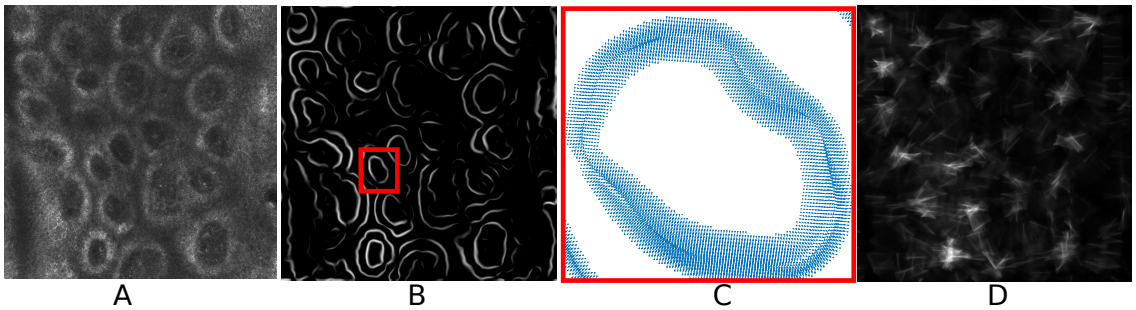


Figure 5.5: (A) An RCM image captured from the DEJ and characterised by papillae; (B) Papillae edges detected using the monogenic filter; (C) The gradient representation of the area indicated by the red box in (B); (D) The accumulator array of the RCM image in (A).

given a higher weight than lower intensity ones. Each pixel location will cast a vote into the accumulator array along a line equal to to the length of the maximum diameter being searched for, in the direction of the gradient orientation, starting at the pixel's location. Therefore one pixel will add values to the accumulator at several locations, equal to the length (in pixels) of the line. The accumulator illustrated in Figure 5.5D shows the superposition of several lines making up the accumulator. In case that a circular object lies in the edged representation, once all candidate pixels cast their vote, a peak will be located at the centre of the circular candidate.

Traditionally, an accumulator for each radius was used, resulting in a 3D accumulator representation, where the third dimension of the accumulator represents the various radii. In this case, the accumulator peak will be located at the correct centre and at the true radius. However, for speed reasons this concept has been adapted by Peng et al. [124] to represent the various votes for multiple radii in one 2D accumulator. The algorithm developed in [124], which was used in this Chapter, makes use of a 2D accumulator to detect a range of radii. Initially the local maxima regions were searched for on a thresholded version of the accumulator illustrated in Figure 5.6B. Areas above the threshold are illustrated in white and grey; where white represents the final local maxima regions and grey indicates the close neighbours (which are local maxima candidates). To avoid the same circle being detected as two (or more) separate regions, neighbouring maxima regions within a distance n are merged together into one maxima. Each detected final local maximum (white region) was grown using a region growing

algorithm, each corresponding to a detected circle in the image, where the centroid of each individual region is set to be the centre of a circle.

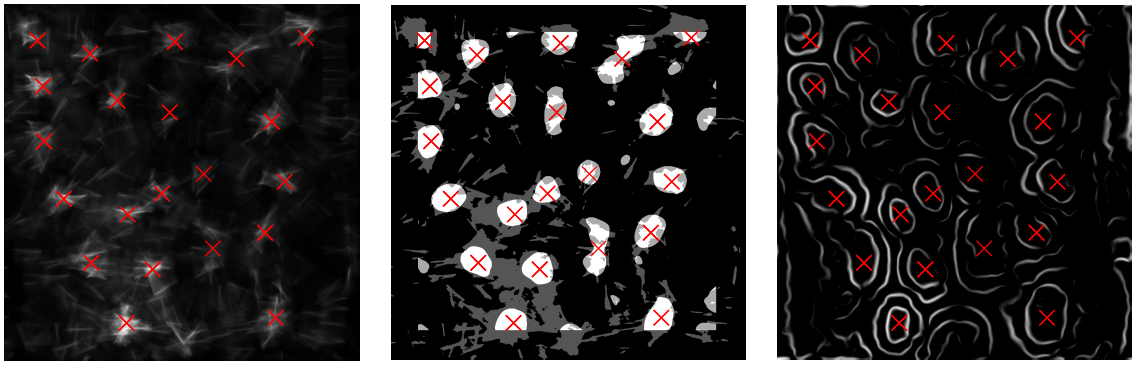


Figure 5.6: (A) Accumulator array (B) Thresholded accumulator indicated by grey and white shades. The white shades are a subset from the grey regions, which are selected to be the selected local maxima regions. The selected regions are circular in shape as a region growing algorithm was used to smooth them. (C) The selected centroids superimposed on the monogenic filter output image (the input to the circular detection algorithm).

Once the circles' centres are detected, the corresponding radius (or radii) of each centre is calculated using a signature curve [124]. The signature curve analyses potential radii and is calculated from the gradient representation of the image (Figure 5.5C). Local maxima will be located in the signature curve at true radii. A maximum is detected at radius r if pixels within the range of $r - \frac{\Delta r}{2}$ and $r + \frac{\Delta r}{2}$ from the circle's centre exhibit the largest gradient vectors from the circle's centre. In this Chapter Δr is equal to $0.5\mu m$ as r is increased by $0.5\mu m$ increments.

In Figure 5.7, the signature curves obtained from one papillae are analysed. Figure 5.7A shows the original calculated values at each potential radius [25, 80], whereas Figure 5.7B shows a Gaussian smoothed version of the one listed in Figure 5.7A. The three individual peaks, relate to three different radii found for the same centre, and are shown in Figure 5.7C. As the papillae is more elliptical in shape, rather than circular, more than two radii can be detected for the inner and outer border of the papillae. This is exemplified in Figure 5.8, where the detected circles were overlaid on the input edged image (Figure 5.8A) and on the original DEJ image (Figure 5.8B). As depicted, a papillae is detected by multiple consecutive radii due to the inner and outer edge borders. The experiment being covered

in this Chapter takes advantage of this, and will be discussed in more detail in the subsequent section.

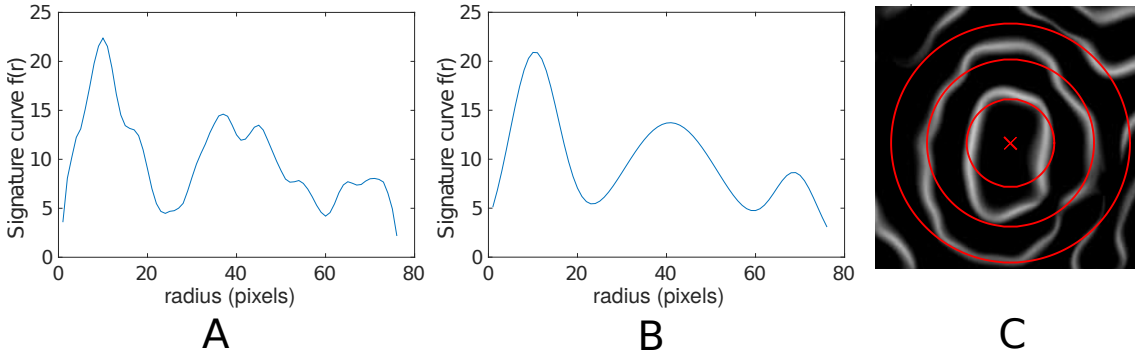
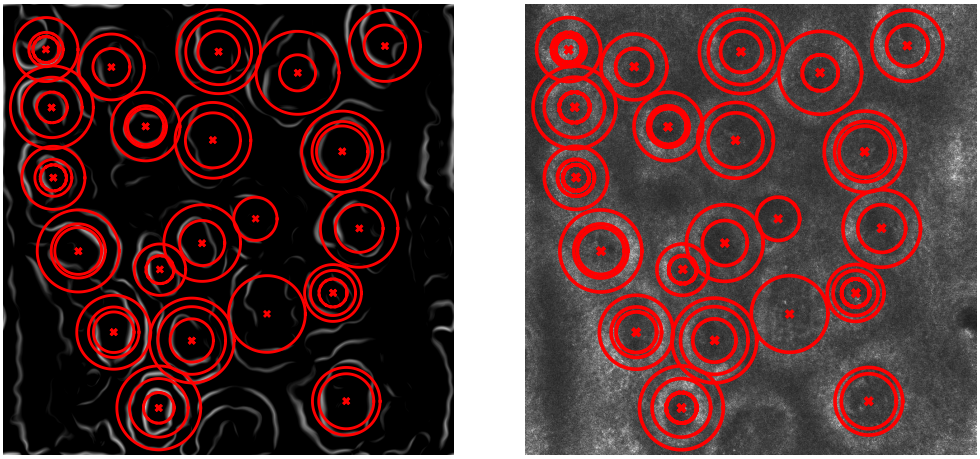


Figure 5.7: (A) The signature curve obtained for the papillae indicated by the red box in Figure 5.5A; (B) A Gaussian smoothed version of the signature curve depicted in (A); (C) The corresponding detected radii illustrated on the papillae image.



(A) The detected circles overlaid on the edged image

(B) The detected circles overlaid on the original DEJ image (from which the edged representation was generated from)

Figure 5.8: Final results of detected circles on both (A) the monogenic edged image and (B) original image

Sub-selecting detected circles The Hough circular transform parameters were relaxed as the papillae is not a perfect circle, but rather an ellipse. The relaxation of parameters ensures that the majority of circular arcs are detected, from which a subset of circles were eventually selected. The parameters were set to:

- Search for a large range of radii $25 - 100\mu m$;

- Apply a low threshold to the gradient image;
- Use a large distance ($40\mu m$) to merge neighbouring maxima;
- Use all peaks of the smoothed signature curve as possible radii.

The subset of circles need to fit the following criteria to be considered to be a papillae:

- At least two consecutive circles are located at the circle's centre;
- The consecutive circles had a range of radii of at least $20\mu m$ to represent the inner and outer ring of the papillae;
- For the circle to be considered to be a papillae, the detected papillae needs to be present across different layers in the stack;
- In addition, the circle cannot be located throughout the entire stack, nor at the first layer below the skin surface (epidermal layer), nor at the last layer below the skin surface (dermal layer).

5.2.2.3 Feature Selection

The inputs required by a supervised classifier are the extracted features and a ground truth classification for each instance. In general, each image is considered to be an individual instance having its own classification and extracted features. As high dimensionality data causes less accurate classification [125], the classifier is not trained on all of the features available, but on a subset of the extracted features. Extraction of this subset was achieved using the Correlation Feature Selection (CFS) algorithm [126]. In this algorithm a feature is considered to be relevant if it is correlated with the class, otherwise it is irrelevant. Features, which have one or more other features highly correlated with it, are considered to be redundant. Therefore, a good feature subset is one that contains features correlated with the class yet uncorrelated with each other. The n' selected discriminative features (A'), from a total of n features (A), that will aid the classifier to distinguish between the epidermal, DEJ and dermal layers, are defined as matrix B:

$$B(A) = A' = (a_1, a_2, a_3, .., a_{n'}) \quad \{A' \subset A, n' < n\} \tag{5.27}$$

5.2.2.4 Classification

The datasets are divided into two groups: training and testing datasets. The training set is used to train the classifier, while the testing set is used to evaluate the classifier on unseen data. The purpose of these machine learning algorithms is to classify an instance, the set of selected features (matrix B), into one of the predetermined classes depending upon the feature values inputted [127]. The Bayes' classifier was chosen to differentiate between epidermal, DEJ and dermal basic images.

The Bayes' classifier [128] is a probabilistic classifier, it calculates the likelihood that an element is assigned to a class, given a particular combination of features. The classifier assumes that each feature is equally important and that features are independent; that the presence (or absence) of a particular feature would not have an effect on any other features. Therefore even if the features depend on each other or upon the existence of other features, a Bayes' classifier considers all these features to independently contribute to the probability that a test image belongs to a particular class. This highlights the importance of performing feature selection before performing classification.

Once a Bayes' classifier (C) was trained to distinguish between epidermal layers, DEJ and dermal layers; it was tested on the testing data. The output of the classifier is a class (Epidermal/DEJ/Dermal) together with a probabilistic prediction value P ranging from zero to one, that reflects how strong the prediction is:

$$C(B) = \{class, P\} \quad \{class = Epidermal/DEJ/Dermal\}, \{P = [0, 1]\} \tag{5.28}$$

A P value of one (zero) indicates that the classifier is 100% (0%) confident in

its classification output, while any value in-between indicates an in-between certainty. The classifier's aim is to minimise the number of misclassified cases in the training data.

For both the feature selection and classification steps, the CFS and Bayes' classification algorithms of the Waikato Environment for Knowledge Analysis (WEKA) toolkit were used [125].

5.2.3 *Find DEJ region* algorithm

The *find DEJ region* algorithm aims to find a region exhibiting edged or non-edged papillae by establishing the epidermal-DEJ and DEJ-dermal boundaries for each stack in the cube. The input to this algorithm is the layer classifier output, where each image was given a *class*; i.e. whether the classifier assigns the image as epidermal, DEJ or dermal represented by 1, 2 and 3 respectively.

For this experiment, each tested cube was split into $N_i \times N_j$ stacks, and each stack was processed individually. As identified earlier, each stack will have an epidermal-DEJ and DEJ-dermal boundary associated to it. The epidermal-DEJ (DEJ-dermal) boundary tile is chosen to be the tile corresponding to the first (last) change in gradient in *class*. Therefore, the location of the two boundary tiles in the stack gives an estimate of the epidermal and dermal heights in the particular stack.

By feeding the epidermal and dermal heights at each $N_i \times N_j$ stack to a 2D B-spline, defined next in Section 5.2.3.1, the two epidermal-DEJ and DEJ-dermal boundaries were defined. Similarly, two boundaries were generated based on the ground truth data (instead of based on the classification prediction). The difference between the two set of boundaries are evaluated in Section 5.3.5.

5.2.3.1 2D B-spline

Based upon the classification results of the testing data, 2D B-spline surfaces representing the epidermal-DEJ and DEJ-boundaries are identified and the discrep-

ancy (μm) between the surfaces achieved using the ground truth and by using the classification results are given. The theory behind 2D B-spline surfaces is given in this Section.

2D B-splines [129] are used to represent the boundaries between the different layers of the skin; a boundary between the epidermal and DEJ and a second boundary between the DEJ and the dermal layer. Each boundary is represented using a uniform 2D B-spline grid. The weights of the B-spline basis functions are calculated by fitting the resulting surface to the height of the boundary from the skin surface at each stack of images, using a least squares approach. Therefore, if the mosaic being analysed is made up of $N_i \times N_j$ images at each layer, each boundary will be generated from $N_i \times N_j$ height values. The purpose of using B-splines instead of the raw height values, is to have a smooth approximation of the boundaries, which should more closely replicate the actual layer boundaries in the skin.

In more detail, a uniform B-spline grid consists of a regular grid of nodes, each node having a specific coordinate (k, l) and an associated weight w^{kl}. As 2D B-splines were used in this work, the grid needs to be larger than the area covered by $N_i \times N_j$ by at least 2 grid points at each border, as illustrated in Figure 5.9. This is because to generate the final surface, an individual 2D B-spline is generated at each grid node, each of which affects a neighbouring region of 2×2 grid nodes. The B-spline weights are calculated by minimising the error term E using the following equation:

$$E = min \sum_{x=1}^{N_i} \sum_{y=1}^{N_j} ||h_\beta(x, y) - h_{input}(x, y)||^2 \tag{5.29}$$

where h_β represents the 2D B-spline interpolation and h_{input} represents the heights established from the ground truth or the classification algorithm. In this experiment, the size of h_{input} is of size $N_i \times N_j$ which corresponds to the total number of tiles found in a mosaic layer.

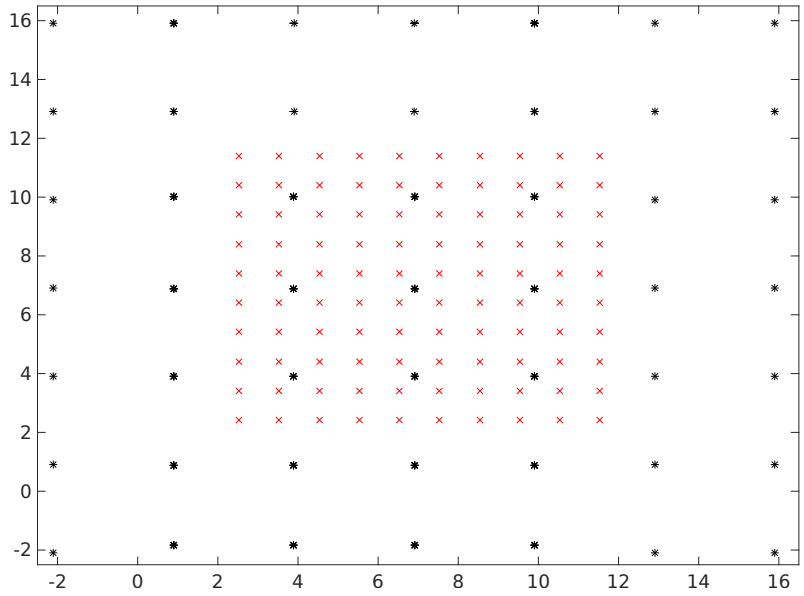


Figure 5.9: The red crosses indicate the $N_i \times N_j$ height inputs whereas the black dots represent the grid points. There needs to be at least 2 grid points outside each border of the input point, as a B-spline of degree 3 is used.

Visual representation of the implemented 2D B-spline surfaces are depicted in Figures 5.10 and 5.11. Figure 5.10 depicts the X-Y ($N_i \times N_j$) representation of the input ground truth and the estimated 2D B-spline. The B-spline provides an analytic expression which can be sampled as densely as required. For example, Figure 5.10 shows a three fold upsampling with respect to the input data. The plots in Figure 5.11 represent the Y-Z representation of the ground truth input data ($N_j \times h_{input}$) and of the estimated surfaces ($N_j \times h_\beta$). The red and blue dots in the latter plot indicate the ground truth input data of the dermal and epidermal layers respectively. It can be noted that the input points are not intersected, however, the 2D B-spline gives a close approximation of the boundary layers. In addition, from Figure 5.11, one can note that the skin boundaries are not always parallel to the imaging plane, mainly due to the posture of the patient during imaging.

5.3 Experimental Method

The algorithm developed in Section 5.2 is tested using the method defined in this Section. The algorithm has been trained and tested using clinical RCM image

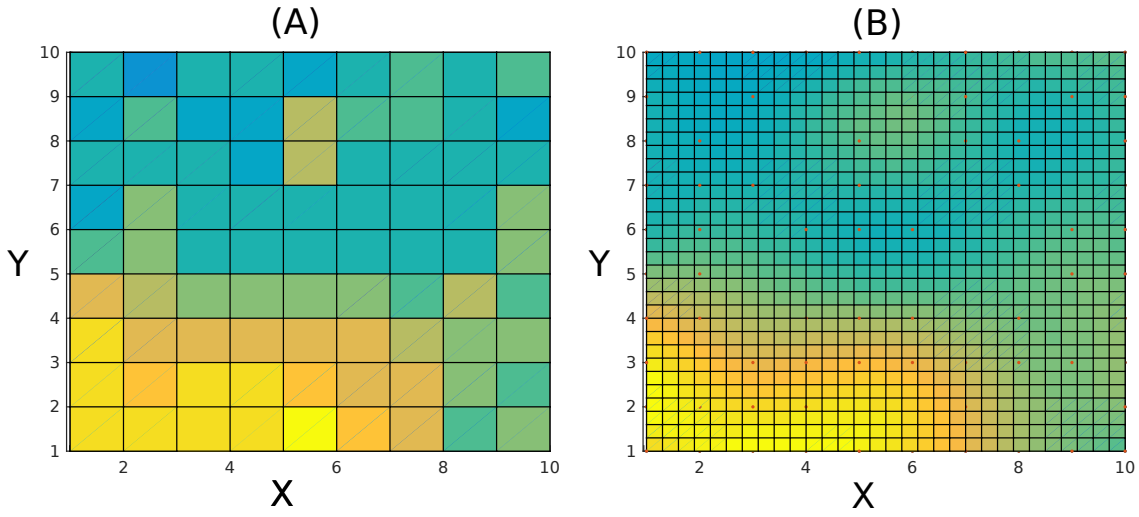


Figure 5.10: An X-Y ($N_i \times N_j$) representation of the dermal layer, where (A) The input data: depicting the ground truth heights of a dermal layer case, which were chosen by a clinician. (B) The 2D B-spline surface generated from the input defined in (A). A large set of interpolated values are available in the second plot. An area of 10×10 is being used as the $N_i \times N_j$ is of this size.

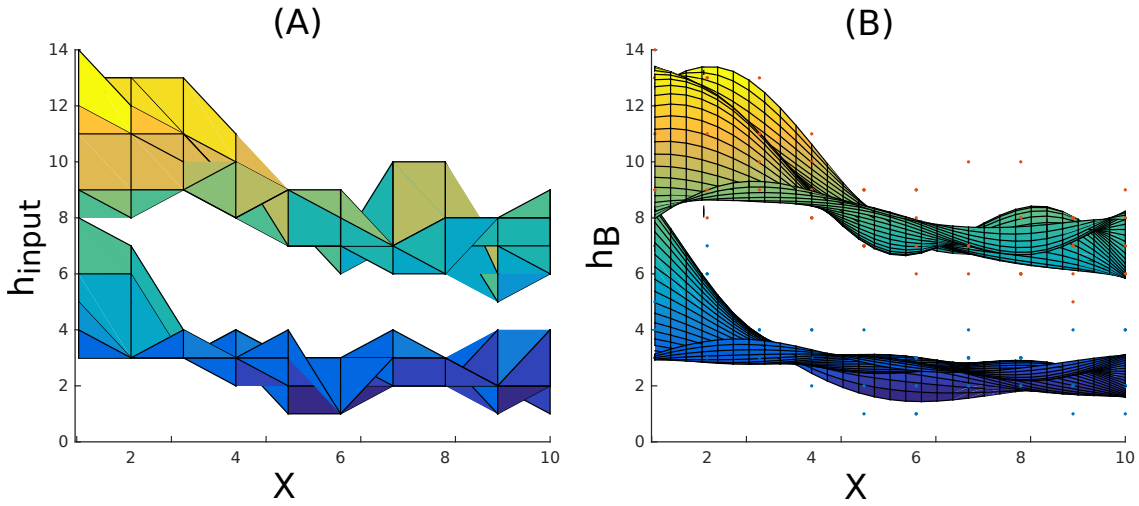


Figure 5.11: A Y-Z ($N_j \times h$) representation of the skin layer, where the surface shaded in blue and yellow represent the epidermal and dermal layers respectively. (A) The input data: depicting the ground truth heights (h_{input}) of an epidermal and dermal layer, which were selected by a clinician. (B) The two 2D B-spline surfaces generated from the estimated heights (h_β) of the input ground truth heights defined in (A). The red and blue dots depict the ground truth input data of the epidermal and dermal layers.

cubes acquired from different volunteers. Each individual image was examined by myself and a label of whether the image is epidermal, DEJ or dermal is given. These labels are taken as the gold standard and compared to the output of the classifier (at testing stage). A description of the data collected is given in Section 5.3.1. In this Chapter, a distinction is made between classifiers trained and tested

on data from the same volunteer (intra-volunteer classifiers) and those trained and tested on data taken from different volunteers (inter-volunteer classifiers). The details of how the data was prepared for these two distinct methodologies is given in Section 5.3.2. This is followed by an explanation on how the monogenic filter parameters were chosen (Section 5.3.3). The methodology used to validate the algorithm is described in Section 5.3.4.

5.3.1 Data Acquisition

Data was collected at St. John's Institute of Dermatology, Guy's and St. Thomas' Hospital, London. The acquisition process involved collecting cubes of images of healthy skin sites, covering different anatomical sites from different volunteers. A total of 11 cubes of images were collected; 7 of which belong to the same volunteer while the other 4 cubes belong to 4 different volunteers. The imaged anatomical sites included forehead, neck, arm, and wrist, as depicted in Table 5.2. The main motivation in acquiring datasets from different volunteers and from different volunteers was to understand the variations found across inter-volunteers and inter-sites and to design classifiers which are capable to handle this variation, which is present in a clinical setting.

Dataset number	Volunteer number	Phototype	Age	Anatomical location	Number of layers
1	1	III	20-35	forehead	13
2	1	III	20-35	neck	13
3	1	III	20-35	arm	10
4	1	III	20-35	arm	15
5	1	III	20-35	arm	15
6	1	III	20-35	arm	12
7	1	III	20-35	wrist	12
8	2	II	50-65	arm	18
9	3	III	20-35	forehead	17
10	4	IV	20-35	wrist	15
11	5	IV	20-35	arm	14

Table 5.2: Collected datasets

Following the widely accepted clinical RCM data acquisition protocol [15], each layer (mosaic) of the cube covered an area of $5 \times 5mm^2$, equivalent to 10×10

images. The depth (z-spacing) between each successive layer was set to $8\mu m$. The number of imaging depths (z) was roughly 14 layers ($\sim 110\mu m$), however this number varied across each volunteer, ensuring that each cube covered both epidermal and dermal layers. Hence, each cube was acquired by starting from an arbitrary location in the epidermis, capturing the first layer, and then successively moving the focal plane $8\mu m$ deeper, until the deep dermal layers were captured. The VivaScope® 1500 laser power was automatically adjusted during each image acquisition (refer to Section 4.5.5). Therefore, the data used for this Chapter resembles that acquired in a clinical setting.

Each tile (where each dataset has approximately 1400 tiles) was labelled by myself into one of three classes: epidermal layers, DEJ or dermal layers from a visual observation of the images on a workstation. The images are labelled as DEJ if they exhibit edged or non-edged papillae. These labels are used as our gold standard. Figures 5.12 and 5.13 illustrate the variability across intra-volunteer and across inter-volunteer sites. The main difference observed between cubes acquired from different anatomical locations is the epidermis thickness [30]. Therefore, differences in cubes of different anatomical sites are the number of layers represented by class *epidermis* rather than differences in individual tiles. Whereas the main differences found in cubes acquired from different volunteers is the level of pigmentation and most notably the variation in shapes of papillae. Although the intra-volunteer DEJ images were collected from different anatomical sites, the appearance of the papillae is more consistent in the intra-volunteer cases than in the inter-volunteer cases. This motivated two type of classifiers to be built: intra-volunteer and inter-volunteer layer classifiers.

5.3.2 Preparing the data for layer classification

A leave-one-out approach was used to generate the intra-volunteer and inter-volunteer classifiers. When using a leave-one-out approach the classifier is trained on all available datasets except one, and tested on the dataset which was not used in the training phase. For the purpose of this Chapter, the training datasets are altered until all datasets have served as testing datasets. A detailed explana-

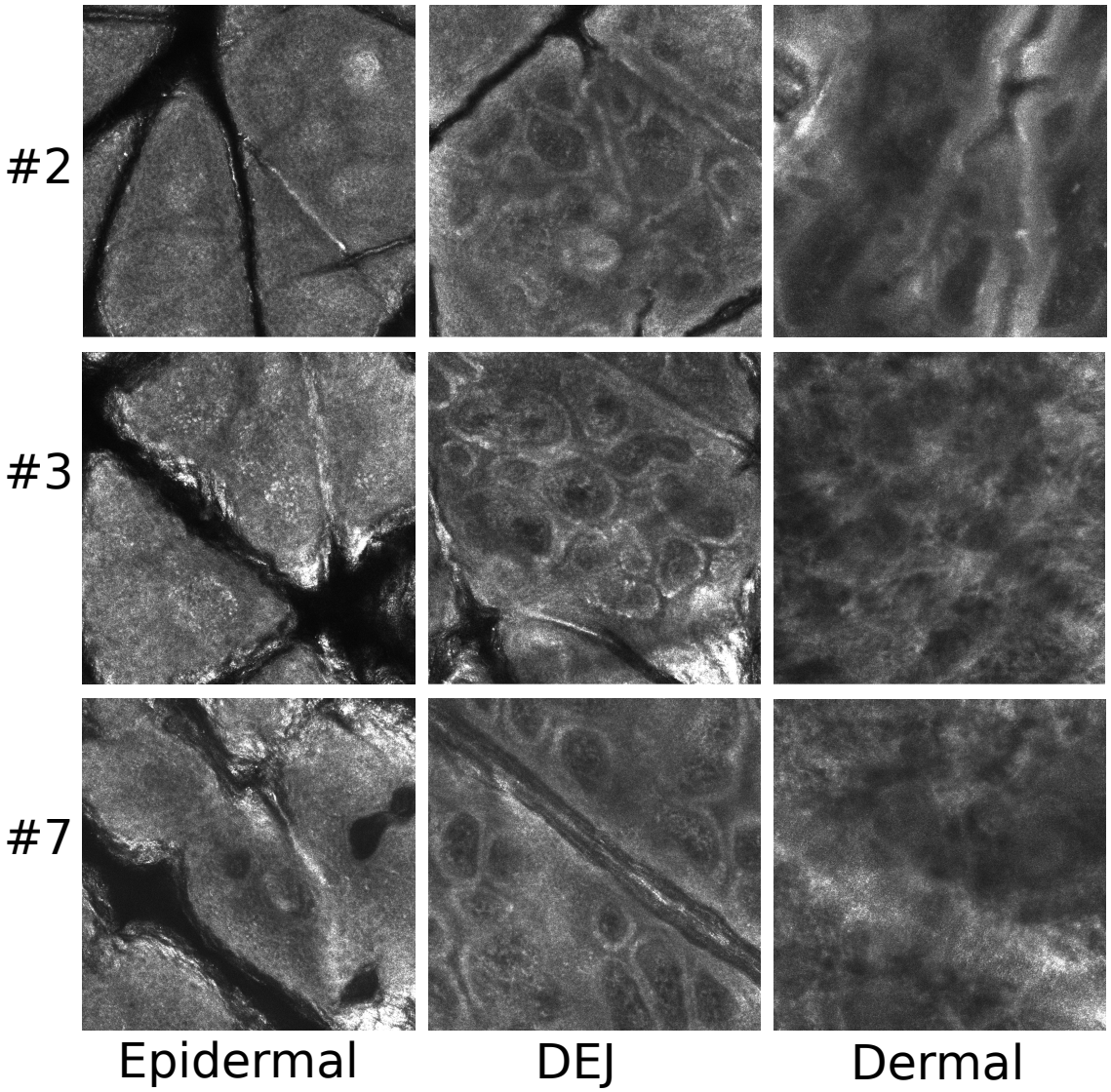


Figure 5.12: Intra-volunteer variability in the epidermal, DEJ and dermal classed images across different datasets. Each row represents the three classed images obtained from one dataset (datasets #2, #3, #7). Each dataset represents a particular anatomical location: neck, arm, wrist.

tion on how the data was divided across training and testing datasets, ensuring that each dataset serves as a testing dataset, is explained in more detailed in this Section. Unlike in the literature [1] [5], none of the acquired images which had artefacts were removed from the datasets, nor was manual or automatic masking out of regions was performed on the images.

The data was subdivided as shown in Tables 5.3 and 5.4 for the intra-volunteer and inter-volunteer classifiers respectively. The purpose of building these two set of classifiers is to evaluate whether classifiers trained and tested on datasets captured from the same volunteer are more robust than those trained and tested

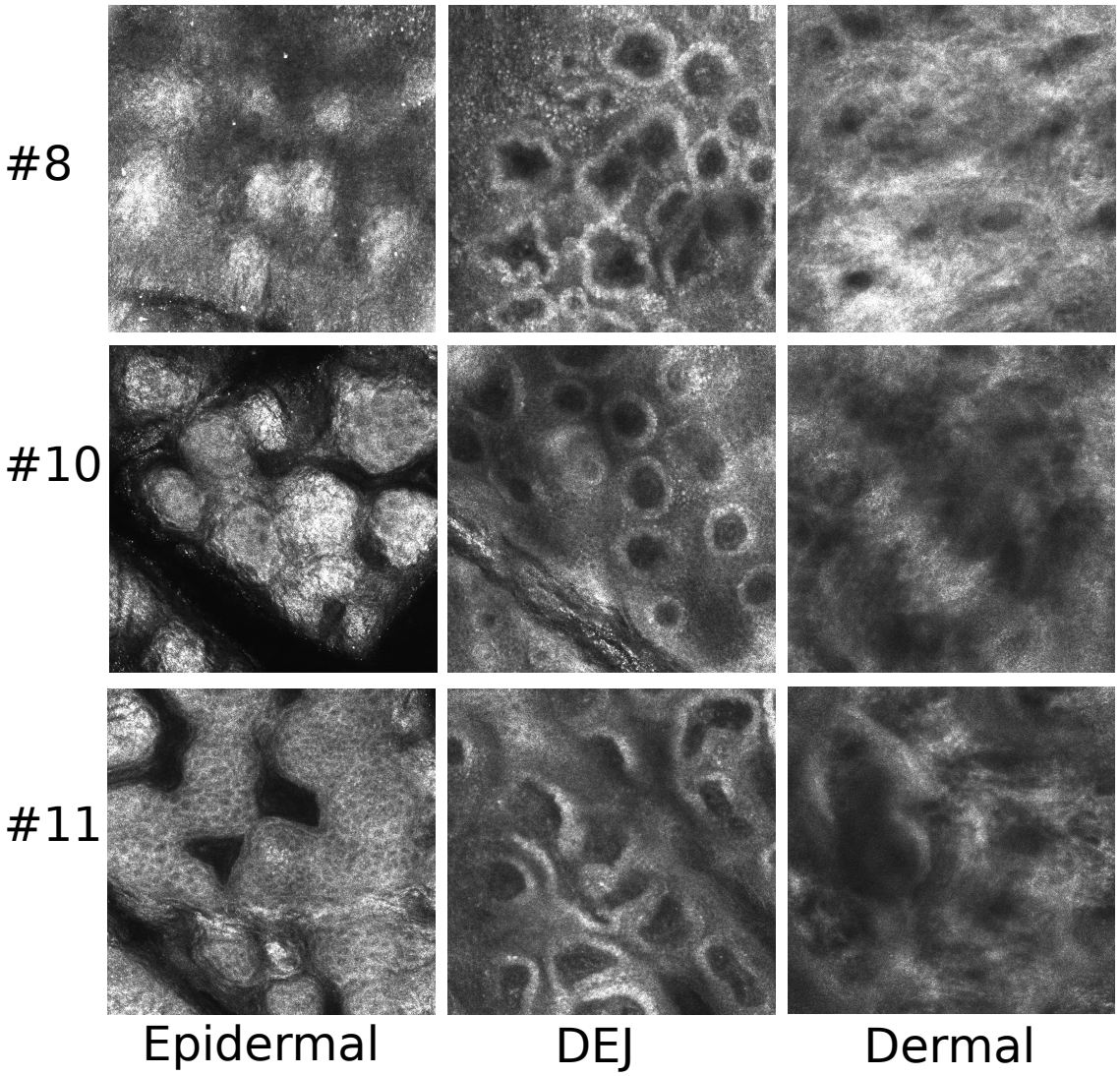


Figure 5.13: Inter-volunteer variability in the epidermal, DEJ and dermal classed images across different datasets. Each row represents the three classed images obtained from one dataset (datasets #8, #10, #11). Each class exhibits different characteristics across different volunteers, most notably are the papillae in the DEJ class, which vary in size and circularity across the three different volunteers.

on datasets captured from different volunteers. For the intra-volunteer classifier seven models (classifiers) were used, where each model was trained on six cubes, and tested on the remaining unseen cube.

As opposed to the above methodology, where the classifier is trained and tested on data belonging to the same volunteer, in the next part of the experiment the classifier was trained and tested on cubes belonging to different volunteers, whose phototypes ranged from II to IV. The inter-volunteer classifier was trained on 3 datasets belonging to 3 different volunteers and tested on a fourth cube (be-

Model no.	Train sets	Test set
i	2,3,4,5,6,7	1
ii	1,3,4,5,6,7	2
iii	1,2,4,5,6,7	3
iv	1,2,3,5,6,7	4
v	1,2,3,4,6,7	5
vi	1,2,3,4,5,7	6
vii	1,2,3,4,5,6	7

Table 5.3: Leave-one-out intra-volunteer layer classifier

longing to another volunteer). This methodology is presented in Table 5.4.

Model no.	Training sets	Testing set
viii	9,10,11	8
ix	8,10,11	9
x	8,9,11	10
iv	8,9,10	11

Table 5.4: Leave-one-out inter-volunteer layer classifier

Each classifier model defined in this Section is built twice; (i) using solely the texture based selected features (T_{CFS}) (ii) using both texture and anatomical based selected features (T_{CFS} and A_{CFS}). The CFS notation refers to the correlation based feature selection algorithm which selects the dominant features, refer to Section 5.2.2.3. The motive in building each classifier model twice is to be able to analyse the improvement in results achieved by using anatomical based features in conjunction with texture based features, as opposed to only using texture based features.

5.3.3 Selection of monogenic filter parameters

The output of the monogenic filter was highly dependent upon the input parameters fed into the filter. In this section an overview on how the parameters were selected to enhance the edges of the papillae, and to reduce the presence of confounding features is provided.

In the work defined in this Chapter a three-scale monogenic filter was used,

where each scale makes use of a distinct wavelength and the wavelength was increased by a scaling factor (λ_S) of 1.2 between consecutive scales. The minimum (maximum) wavelength was chosen to be 60 (86) pixels. By visually inspecting Figure 5.14, a minimum wavelength (λ_{min}) set to 60 segments the papillae edges. This corresponds to a wavelength range (λ_R) of $30 - 40\mu m$.

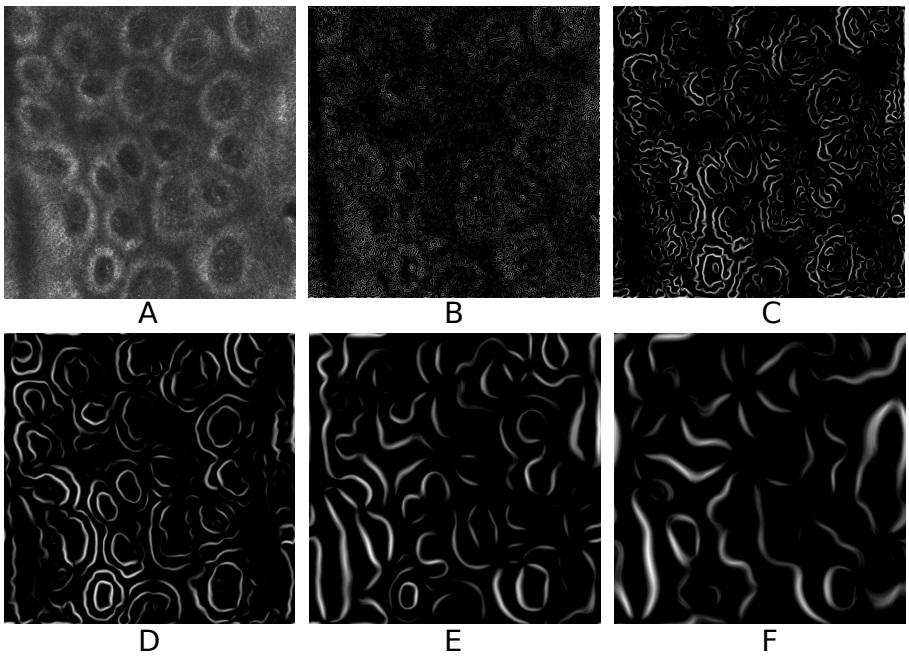


Figure 5.14: A 3-scale monogenic filter applied to the (A) original DEJ image with the three wavelengths being spaced by a factor of 1.2 and σ set to 0.6. λ_{min} was increased to (B) 10 pixels; (C) 30 pixels; (D) 60 pixels; (E) 120 pixels; (F) 200 pixels. As λ_{min} was increased from 10 to 60, the edges of the papillae became more prominent, after which the papillae edges merged together.

By inspecting Figure 5.15, one notes that as σ increases from 0.2 to 0.6, the papillae become sharper, as higher frequencies representing the papillae are incorporated. Whereas, as σ increases from 0.6 to 0.99 the frequency band increases significantly and the papillae are not well defined.

To summarise, for this Chapter a $\lambda_{min} = 60$, $\lambda_S = 1.2$ and $\sigma = 0.6$ were chosen. The choice of these parameters is further supported by Figure 5.16 where the same parameters were applied to other DEJ images from different volunteers. These parameters were additionally applied to the three different layers in a stack of the same volunteer in Figure 5.17. As can be noted, the detected edges in the epidermal and dermal images are less circular in shape than the edges

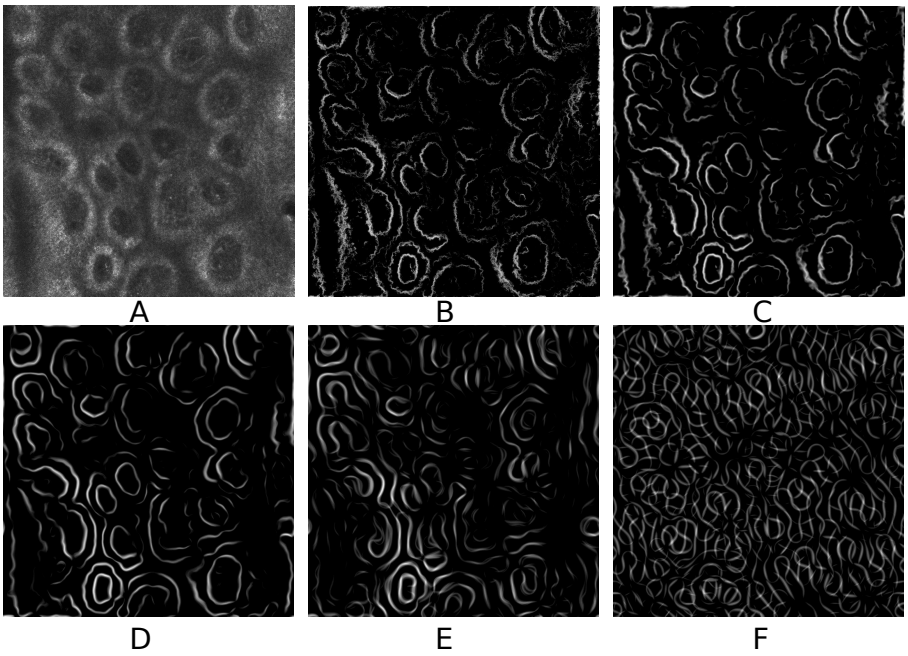


Figure 5.15: A 3-scale monogenic filter applied to the (A) original DEJ image with the minimum wavelength being 60 and a spacing factor of 1.2. The σ value was increased from (B) 0.2; (C) 0.4; (D) 0.6; (E) 0.8; (F) 0.99. As the σ value increased the image became sharper up to the point of $\sigma = 0.6$, beyond which the papillae were not properly defined.

found in the DEJ image. The detected edges in the DEJ image correspond to the papillae and are distinct in shape.

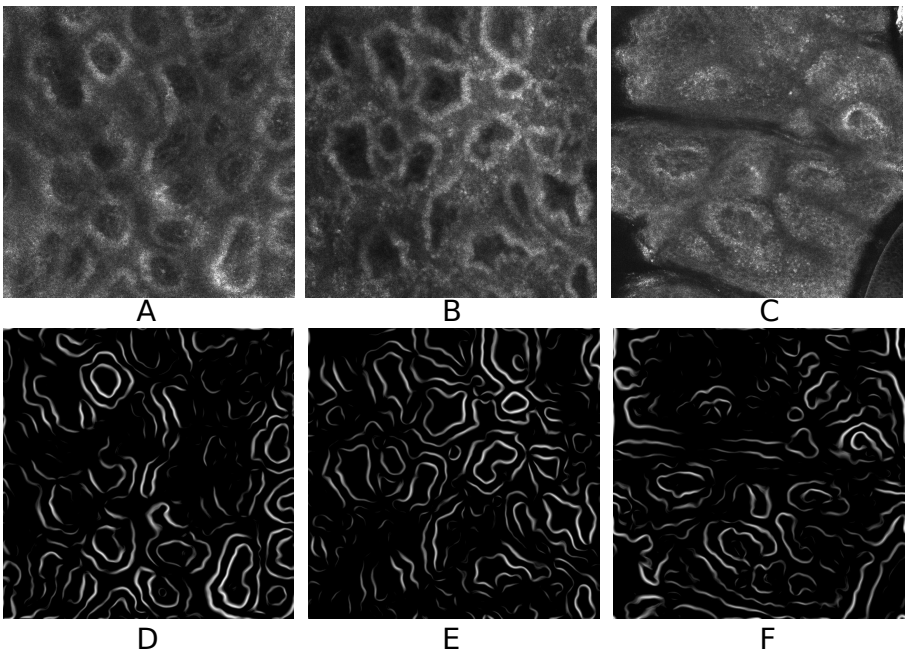


Figure 5.16: A 3-scale monogenic filter applied to three DEJ images, each of which was imaged from a different volunteer. λ_{min} was set to 60, and a wavelength spacing factor of 1.2 was used. σ was set to 0.6.

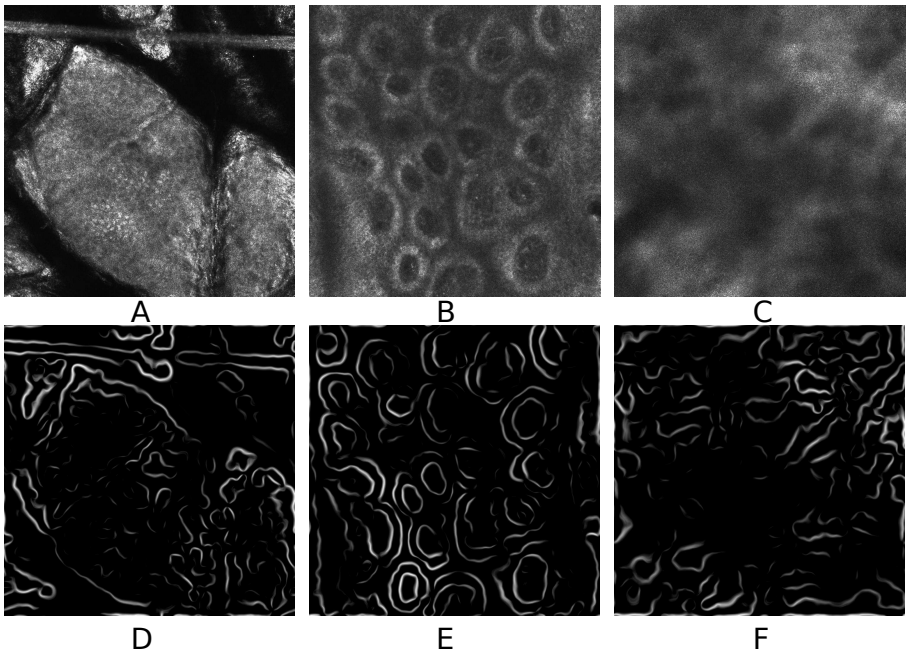


Figure 5.17: The feature asymmetry measure based on a 3-scale monogenic filter applied to three basic images,representing (A) the epidermal layer (B) the DEJ (C) the dermal layer, extracted from the same stack of images. λ_{min} was set to 60, and a wavelength spacing factor of 1.2 was used. σ was set to 0.6.

5.3.4 Evaluating the layer classifier

The classifiers were analysed using confusion matrices, which can be represented using the matrix in Equation (5.30). The *Actual* labels depict the gold standard, and can take values $\{0, 1\}$, whilst the *Class* labels represent the classification given by the layer classifier. The classifier assigns the class having the maximum classification probability (out of the three classes), therefore the output of the classifier can take values $\{0, 1\}$.

$$\begin{array}{c} \\ Actual_{Epi} \\ Actual_{DEJ} \\ Actual_{Der} \end{array} \begin{array}{c} \begin{array}{ccc} Class_{Epi} & Class_{DEJ} & Class_{Der} \end{array} \\ \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \end{array} \tag{5.30}$$

Values a, e and i are the number of correctly classified images of class epidermis, DEJ and dermis respectively. Values b and c are the number of epidermis images which were incorrectly classified as DEJ or dermis. Values d and f are the number of DEJ images which were incorrectly classified as epidermis or dermis.

In turn, values g and h are the number of dermis images which were incorrectly classified as epidermis or DEJ. The sensitivity and specificity percentages of each class were calculated from these values as defined in Equations 5.31. In addition the true positive (TP) was calculated to give a percentage of the total images correctly classified.

$$\begin{aligned} Sensitivity(Sn) &= \frac{TP}{TP+FN} \\ Specificity(Sp) &= \frac{TN}{TN+FP} \\ Sn_{Epi} = \frac{a}{a+b+c} \qquad & Sp_{Epi} = \frac{e+i}{e+i+d+g} \\ Sn_{DEJ} = \frac{e}{e+d+f} \qquad & Sp_{DEJ} = \frac{a+i}{a+i+b+h} \\ Sn_{Der} = \frac{i}{i+g+h} \qquad & Sp_{Der} = \frac{a+e}{a+e+c+f} \\ TP &= \frac{a+e+i}{a+b+c+d+e+f+g+h+i} \end{aligned} \tag{5.31}$$

5.3.5 Evaluating the *find DEJ region* algorithm

To evaluate how well the *find DEJ region* algorithm performs the parameters defined in Equations 5.32 - 5.36 are measured and evaluated. These measures are based upon the parameters epidermal height (h_{epi}), dermal height (h_{der}) and DEJ thickness (th) indicated in Figure 5.18. Equation 5.32 (5.33) measures Δh_{epi} (Δh_{der}), which is a measure of the mean difference in epidermal-DEJ (DEJ-dermal) boundary height location between the two 2D B-spline boundaries achieved using the ground truth $h_{epi_{GT}}$ ($h_{der_{GT}}$) and the classification results h_{epi_C} (h_{der_C}).

The DEJ thickness, which can be defined by the thickness between the epidermal-DEJ and DEJ-dermal boundaries, is an additional important measure for clinicians. The thickness error (Δth) measures the difference in the mean DEJ thickness when using the boundaries generated from the ground truth data (th_{GT}) and classification results (th_C).

$$\Delta h_{epi} = \frac{\sum_{j=1}^{N_j} \sum_{i=1}^{N_i} |h_{epi_{GT}}(i,j) - h_{epi_C}(i,j)|}{N_i \times N_j} \tag{5.32}$$

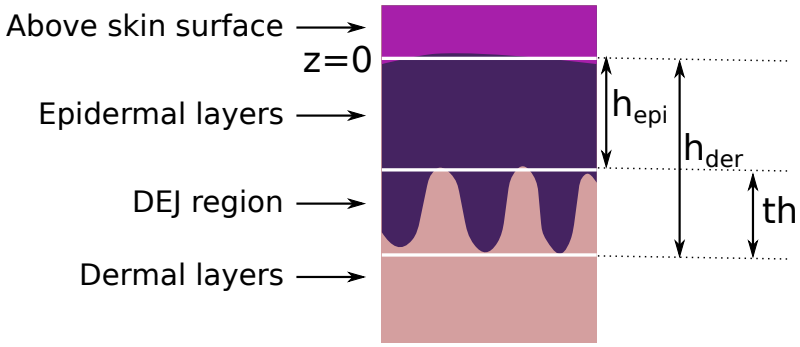


Figure 5.18: The datasets were collected from the epidermal layers (below the skin surface), from the point marked $z = 0$, up to a depth within the dermal layer. h_{epi} (h_{der}) mark the epidermal (dermal) height from the skin surface, whereas th represents the DEJ thickness, that can be calculated from the difference between h_{epi} and h_{der}.

$$\Delta h_{der} = \frac{\sum_{j=1}^{N_j} \sum_{i=1}^{N_i} |h_{der_{GT}}(i,j) - h_{der_C}(i,j)|}{N_i \times N_j} \tag{5.33}$$

$$th_{GT} = \frac{\sum_{j=1}^{N_j} \sum_{i=1}^{N_i} |h_{epi_{GT}}(i,j) - h_{der_{GT}}(i,j)|}{N_i \times N_j} \tag{5.34}$$

$$th_C = \frac{\sum_{j=1}^{N_j} \sum_{i=1}^{N_i} |h_{epi_C}(i,j) - h_{der_C}(i,j)|}{N_i \times N_j} \tag{5.35}$$

$$\Delta th = |th_{GT} - th_C| \tag{5.36}$$

where $N_i \times N_j$ are the number of stacks (and so the number of individual heights) the 2D B-spline boundaries were generated from.

5.4 Results

In this section the results achieved for the algorithm defined in Section 5.2 will be presented. The results are divided into two main parts: intra-volunteer (Section 5.4.1) and inter-volunteer (Section 5.4.2). Each section has two set of results; the layer classification results (calculated as defined in Section 5.3.4) and the height and thickness errors between the B-splines generated from the ground truth and the classification results (refer to Section 5.3.5). Tables 5.5 and 5.7 illustrate the change in results achieved by introducing the anatomical (A_{CFS}) features to the

selected texture features (T_{CFS}). Similarly, Tables 5.6 and 5.8 show the change in epidermal and dermal heights and in DEJ thickness error when introducing the anatomical (A_{CFS}) features to the selected texture features (T_{CFS}).

5.4.1 Intra-volunteer Layer Classifier

The results achieved using the intra-volunteer datasets (datasets #1-7) are shown in Tables 5.5 and 5.6. As discussed in Section 5.3.2, each Model was trained on 6 datasets and tested on the dataset not used for training purposes. As indicated in Table 5.3, Model (i) was tested on dataset #1, Model (ii) was tested on dataset #2, and so on.

As opposed to the other datasets, dataset #1 (on which Model (i) was tested) only has images acquired at the epidermal and DEJ, despite the cube of images covering a total depth of $\sim 100\mu m$. This is why no dermal results are presented for Model (i) in Tables 5.5 and 5.6. Anatomical differences are present across different body sites; dataset #1 is captured from the forehead, whereas the rest of the datasets were acquired from the neck or arm. Additionally, all datasets except dataset #1 have a strong wrinkle presence in the images classed as epidermal. This corresponds to the lowest epidermal sensitivity (Table 5.5) and highest error in epidermal height (Table 5.6) being achieved with Model (i).

From Table 5.5, it can be noted that all the mean results, except for Sp_{DEJ} and Sn_{Dermal}, increased by introducing the anatomical based features (A_{CFS}) in addition to the texture features (T_{CFS}). The introduction of A_{CFS} increased the mean TP rate from 58.0% to 63.2% of the images. The highest increase after introducing A_{CFS} was in the mean Sn_{DEJ}, where the results improved from 47.5% to 56.2%. When introducing A_{CFS} the largest reduction in sensitivity was in Sn_{Dermal}, where the mean value decreased from 61.9% to 54.6%.

From Table 5.6 one can note that when introducing the A_{CFS} features, the mean Δh_{epi}, Δh_{der} and Δth errors decreased by $1\mu m$, $3.8\mu m$ and $6.8\mu m$ respectively. The Wilcoxon signed rank test fails to reject the null hypothesis of zero

Model	Features	Epidermal		DEJ		Dermal		TP (%)
		Sn (%)	Sp (%)	Sn (%)	Sp (%)	Sn (%)	Sp (%)	
i	T_{CFS}	41.7	90.1	64.6	45.1	N/A	N/A	52.6
	$T_{CFS} + A_{CFS}$	43.0	92.6	72.7	45.6	N/A	N/A	57.2
ii	T_{CFS}	72.1	98.6	54.5	76.4	89.6	64.8	61.3
	$T_{CFS} + A_{CFS}$	67.5	99.4	93.9	61.2	32.5	95.7	83.2
iii	T_{CFS}	99.5	66.4	52.7	60.2	38.8	99.1	57.1
	$T_{CFS} + A_{CFS}$	98.6	75.1	63.3	63.1	43.8	98.7	63.0
iv	T_{CFS}	84.3	63.3	30.7	76.2	59.6	85.0	58.1
	$T_{CFS} + A_{CFS}$	87.9	72.2	46.1	74.6	59.8	90.3	63.3
v	T_{CFS}	65.6	99.4	41.5	86.2	95.1	56.7	69.3
	$T_{CFS} + A_{CFS}$	65.6	99.6	28.8	88.5	97.1	47.2	65.8
vi	T_{CFS}	98.4	38.4	30.2	54.9	20.5	99.5	42.6
	$T_{CFS} + A_{CFS}$	97.0	46.9	39.9	53.3	21.8	99.5	45.3
vii	T_{CFS}	77.3	90.2	58.1	71.6	68.0	72.7	65.0
	$T_{CFS} + A_{CFS}$	90.3	91.0	48.5	77.2	72.5	65.3	64.6
Mean	$\mathbf{T_{CFS}}$	**76.8**	**78.1**	**47.5**	**67.2**	**61.9**	**79.6**	**58.0**
	$\mathbf{T_{CFS} + A_{CFS}}$	**78.6**	**82.4**	**56.2**	**66.2**	**54.6**	**82.7**	**63.2**

Table 5.5: Intra-volunteer layer classifier results

median in the difference at a 5% significance level.

	T_{CFS}			$T_{CFS} + A_{CFS}$		
Model no.	$\Delta h_{epi}(\mu m)$	$\Delta h_{der}(\mu m)$	$\Delta th(\mu m)$	$\Delta h_{epi}(\mu m)$	$\Delta h_{der}(\mu m)$	$\Delta th(\mu m)$
i	37.4	N/A	N/A	36.0	N/A	N/A
ii	7.5	5.8	13.3	9.0	8.9	0.5
iii	14.0	28.5	42.5	10.1	0	10.1
iv	1.4	21.8	20.3	4.8	21.9	17.1
v	8.9	19.1	10.2	8.0	23.3	15.3
vi	26.0	37.7	11.7	20.2	37.1	17.0
vii	1.3	4.4	5.6	1.6	3.3	15.3
Mean	**13.8**	**19.6**	**17.3**	**12.8**	**15.8**	**10.8**
σ	**13.4**	**12.9**	**13.2**	**11.7**	**14.2**	**6.3**

Table 5.6: Intra-volunteer epidermal and dermal height errors and DEJ thickness errors

5.4.2 Inter-volunteer Layer Classifier

In this section the results achieved using inter-volunteer datasets (datasets #8-11) are shown in Tables 5.7 and 5.8. Similarly to the intra-volunteer models, each model was tested on 3 datasets and tested on the dataset not used for training the

classifier (e.g. model (viii) was tested on dataset #8). Despite the smaller number of datasets used in these classifiers, and the variability of the datasets (each dataset was captured from a different volunteer), the overall results achieved in Table 5.7 are similar to those achieved in Table 5.5.

Model (viii) was tested on a dataset captured on an older volunteer. All models except for model (ix), which was tested on a dataset captured from the forehead, were tested on datasets captured from the arm. Thus, datasets #8 and 9 were considerably different from datasets #10 and 11.

From Table 5.7, it can be noted that the mean TP rate increased from 59.4% to 62.1% when introducing the A_{CFS}. When comparing the mean results achieved using T_{CFS} to those achieved using $T_{CFS} + A_{CFS}$, most sensitivity and specificity values remained relatively constant after the introduction of A_{CFS}. However, similarly to the results achieved in the intra-volunteer classifiers, the mean Sn_{Dermal} decreased the most, from 64.1% to 57.1%, with the introduction of the anatomical based features. This illustrates that the anatomical based features are not aiding the classifier to correctly identify the images classed as dermal.

Model	Features	Epidermal		DEJ		Dermal		TP (%)
		Sn (%)	Sp (%)	Sn (%)	Sp (%)	Sn (%)	Sp (%)	
viii	T_{CFS}	23.5	99.1	67.2	47.1	66.0	73.1	50.5
	$T_{CFS} + A_{CFS}$	29.1	98.8	62.3	51.2	68.5	71.6	52.6
ix	T_{CFS}	69.6	85.0	36.8	74.4	70.8	66.2	56.1
	$T_{CFS} + A_{CFS}$	65.1	83.3	31.3	75.4	76.0	59.6	51.7
x	T_{CFS}	97.8	60.2	50.1	68.1	33.8	96.9	59.9
	$T_{CFS} + A_{CFS}$	97.4	63.3	48.2	75.2	48.3	95.1	63.4
xi	T_{CFS}	83.8	83.8	62.7	83.8	85.7	82.6	71.2
	$T_{CFS} + A_{CFS}$	80.8	88.6	81.3	79.6	35.7	94.3	80.6
Mean	$\mathbf{T_{CFS}}$	**68.7**	**82.0**	**54.2**	**68.4**	**64.1**	**79.7**	**59.4**
	$\mathbf{T_{CFS} + A_{CFS}}$	**68.1**	**83.5**	**55.8**	**70.4**	**57.1**	**80.1**	**62.1**

Table 5.7: Inter-volunteer layer classifier

As shown in Table 5.8, the mean errors with and without the anatomical based features are approximatively equal. The largest errors are found with Models (viii) and (ix), which are tested on a dataset captured from an older volunteer

and from the forehead respectively. The Wilcoxon signed rank test fails to reject the null hypothesis of zero median in the difference at a 5% significance level.

	T_{CFS}			$T_{CFS} + A_{CFS}$		
Model no.	$\Delta h_{epi}(\mu m)$	$\Delta h_{der}(\mu m)$	$\Delta th(\mu m)$	$\Delta h_{epi}(\mu m)$	$\Delta h_{der}(\mu m)$	$\Delta th(\mu m)$
viii	40.7	12.7	28.2	38.3	11.3	26.9
ix	44.9	12.4	31.3	46.3	14.9	30.2
x	20.2	0	20.2	18.6	0	18.6
xi	1.7	0	1.7	4.9	0	4.9
Mean	**26.9**	**6.3**	**20.3**	**27.0**	**6.6**	**20.1**
σ	**20.0**	**7.3**	**13.2**	**18.8**	**8.2**	**10.9**

Table 5.8: Inter-volunteer epidermal and dermal height errors and DEJ thickness errors

5.4.3 2D B-spline representation of results

In this Section some of the results tabulated in Sections 5.4.1 and 5.4.2 are illustrated with the 2D B-spline representations. In all these plots the h_β heights are shown in μm units and refer to the distance from the skin surface; the blue shaded 2D B-spline represents the epidermal-DEJ boundary whilst the yellow shaded 2D B-spline represents the DEJ-dermal boundary. In these plots, the blue (red) dots are the input data used to generate the epidermal-DEJ (DEJ-dermal) boundary. In these Figures the (A) plot represents the 2D B-splines generated using the ground truth, whereas the (B) plot represents the 2D B-splines generated using the classification results.

Figure 5.19 represents the 2D B-spline representation of the results achieved from Model #2 using both $T_{CFS} + A_{CFS}$. This example depicts the best overall model, where the mean Δth error is $0.5\mu m$ (refer to Table 5.6). When comparing the 2D B-splines achieved using the ground truth (Figure 5.19A) to those achieved using the classification results (Figure 5.19B), one can note that both the epidermal-DEJ and DEJ-dermal boundaries exhibit the same overall shape.

Figure 5.20 represents the 2D B-spline representations of the results achieved from Model #5 using both $T_{CFS} + A_{CFS}$. This example depicts an average repre-

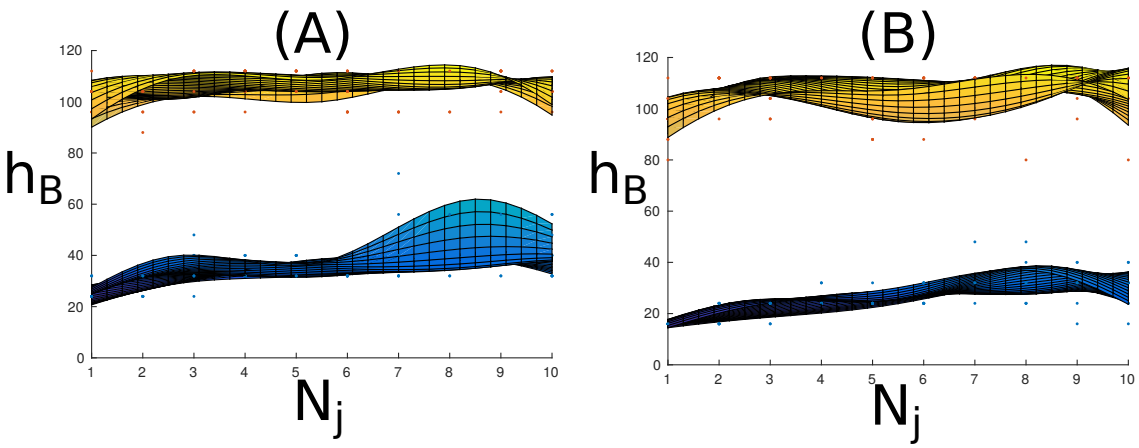


Figure 5.19: A Y-Z ($N_j \times h_\beta$) representation of the 2D B-splines achieved for Model # 2 using $T_{CFS} + A_{CFS}$. (A) The 2D B-splines generated using the ground truth labels (B) The 2D B-splines generated using the results achieved with the layer classifier. The errors achieved with this model are the lowest from all the models available, with a Δth of $0.5\mu m$.

sentation of the results achieved, where the mean Δth error is $15.3\mu m$ (refer to Table 5.6). When comparing the 2D B-splines achieved using the ground truth (Figure 5.20A) to those achieved using the classification results (Figure 5.20B), one can note that the epidermal-DEJ boundary exhibit the same overall shape. However, the DEJ-dermal boundary achieved using the classification results fluctuates more than that achieved using the ground truth labels.

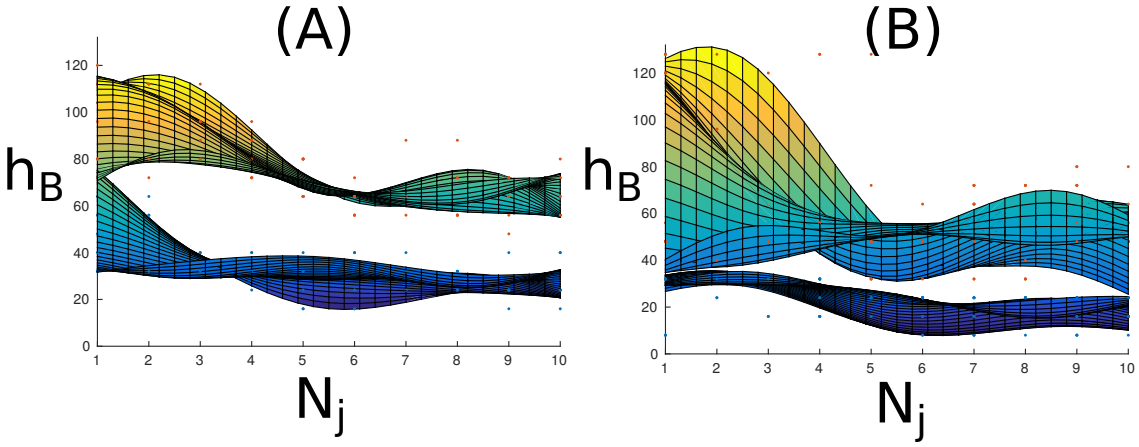


Figure 5.20: A Y-Z ($N_j \times h_\beta$) representation of the 2D B-splines achieved for Model # 5 using $T_{CFS} + A_{CFS}$. (A) The 2D B-splines generated using the ground truth labels (B) The 2D B-splines generated using the results achieved with the layer classifier. The errors achieved with this model are an average representation of the errors achieved from all the models available.

Figure 5.21 represents the 2D B-spline representations of the results achieved

from Model #8 using both $T_{CFS} + A_{CFS}$. This example depicts a model with one of the largest errors from the results achieved; the mean Δh_{epi} and Δth errors are 40.7μm and 26.9μm respectively (refer to Table 5.8). When comparing the 2D B-splines achieved using the ground truth (Figure 5.21A) to those achieved using the classification results (Figure 5.21B), one can note that the epidermal-DEJ and DEJ-dermal boundaries achieved using the classification results exhibit different shapes and mean values to those generated using the ground truth labels. The DEJ-dermal boundary achieved using the ground truth labels is relatively constant across the cube of images, whereas the same boundary achieved using the classification results fluctuates significantly across the cube of images. The shapes achieved for the epidermal-DEJ are more similar, however, the one achieved using the ground truth labels are deeper in the skin than that achieved using the classification results. Database #8, the database the Model was tested on, was captured from a volunteer in the age bracket 50-65, opposed to the rest of the volunteers whose age ranged from 20-35. As shown in Figure 5.13 the DEJ tile acquired from volunteer #8 depicts papillae which are less circular than those acquired from other volunteers shown in Figures 5.12 and 5.13.

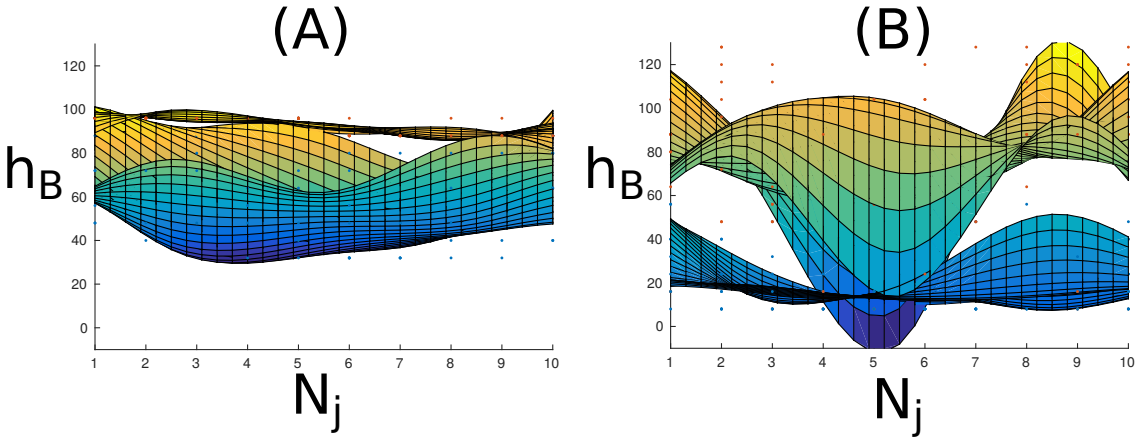


Figure 5.21: A Y-Z ($N_j \times h_\beta$) representation of the 2D B-splines achieved for Model # 8 using $T_{CFS} + A_{CFS}$. (A) The 2D B-splines generated using the ground truth labels (B) The 2D B-splines generated using the results achieved with the layer classifier. The errors achieved with this model are one the highest from all the models available, with a Δth of 26.9μm.

5.5 Discussion

The results achieved from the intra-volunteer datasets are comparable to those achieved using the inter-volunteer datasets. This illustrates that the variability of papillae size and shape together with differences in melanin content found between different volunteers does not hinder the classification. However, the variation found across different body sites had a large effect on performance.

When focusing on the inter-volunteer datasets, 3 out of 4 datasets (datasets #9-11) had prominent papillae which made the DEJ class more easily distinguishable from the rest. On the other hand, the datasets used for the intra-volunteer classifiers had less prominent papillae (referred to as non-edged papillae). Such papillae have less well defined borders that are more difficult for the Hough circular transform to detect.

When introducing the anatomical based features, the TP value (the percentage of correctly classified images) increased from an average of 58.0% (59.4%) to 63.2% (62.1%) for the intra-volunteer (inter-volunteer) cases. These values are comparable to 58.6% correctly classified images reported in [3] for datasets captured from fair skinned volunteers for the same application (that of differentiating between epidermal, DEJ and dermal images).

Based upon the results using texture and anatomical based features the mean DEJ thickness error is of $10.8\mu m$ and $20.1\mu m$ for the intra and inter-volunteer cases respectively. In the datasets used in this Chapter, each layer of the cube is displaced by $8\mu m$ from the previous layer, and therefore the mean error in terms of layers is of 1.4 and 2.5 layers respectively.

Kurugol et al. [1] achieved an average error of $8.5\mu m$ in locating the epidermal-DEJ and DEJ-dermal boundaries using stacks of $1\mu m$ spacing (error of 8.5 layers). Similarly, Hames et al. [5] reported average errors of $6.0\mu m$ and $5.5\mu m$ to locate the epidermal-DEJ and DEJ-dermal boundaries respectively using stacks of $2\mu m$ (errors of 3 and 2.75 layers respectively). The results achieved in this Chapter are

greater than the literature in absolute terms but smaller in terms of number of layers. Arguably using a smaller spacing results in smaller absolute errors due to the smaller changes across images captured sequentially. In addition, using stacks of data (instead of cubes) results in less variable data.

The results presented in Tables 5.6 and 5.8 are dependent on the performance of the layer classification algorithm. For each stack in the cube the epidermal-DEJ and DEJ-dermal boundaries are determined. The epidermal-DEJ (DEJ-dermal) is set to be the first (last) depth where the stack changes from an epidermal to a DEJ (from a DEJ to a dermal) image. This approach does not work well when there are images misclassified across the stack; as there will not be a clear classification gradient edge across the stack. However, the results indicate that it is easier to distinguish between the DEJ-dermal layer boundary than the epidermal-DEJ boundary; mainly because there is greater class overlap (i.e. an attribute occupying the same space/using similar values to represent different classes) between epidermal and DEJ classed images compared to DEJ and dermal classed images. Section 5.6.4 suggests possible future improvements to reduce class overlap and discusses the option of labelling an image which exhibits class overlap with multi-labels.

High intra and inter feature variability across different anatomical positions and across different volunteers, are a major factor of some misclassification instances. As skin fluctuations are present in small regions, neighbouring scanning regions from the same volunteer, i.e. neighbouring $5 \times 5mm^2$ regions, can produce very different pixel intensities from similar anatomical features. Furthermore, in all the available datasets, there are confounding features (e.g. wrinkles). Possible methods in removing confounding features are defined in Section 5.6.3. Moreover, additional anatomical based features could be added to improve classification (refer to Section 5.6.2).

In contrast to the experiments defined in [1] [4] [5] which made use of stacks with a $1 - 2\mu m$ spacing, the data used in this Chapter is closer to clinical data, as cubes of images having a $8\mu m$ spacing were used. In the work defined in [1] [4]

the authors manually masked out wrinkles whereas in the work defined in [5] noisy stacks were removed. In this Chapter no noisy tiles/datasets were removed and no regions were masked out.

The number of images used in this Chapter, where a total of 11 datasets were used each having $\sim$ 1400 images, is considerably larger than that used in Kurugol et al. [3] and Somoza et al. [4] where a total of 1800 and 195 images were used respectively. The number of images used in this Chapter is comparable to that used by Hames et al. [5] where a total of $\sim$ 15400 images were used. However, a limitation of the data collection methodology used in [5] is that the dataset includes two stacks from each body site; resulting in two stacks having very similar appearance. As the training and testing datasets were divided completely randomly [5], it might result in the training data including one of the stacks and the testing data including the other stack. In this case, the classifier is tested on data very similar to the data it was trained on. This contrasts with the methodology used in this Chapter, where the testing dataset was captured from a different site (or volunteer) from the rest of the data.

An additional way in which the work done in this Chapter distinguishes itself from prior work in the literature [1] [4] [5], is that large scale features are searched for. In the experiments in the literature [1] [4] [5] each tile was sub-divided into sub-tiles. For example, Kurugol et al. [1] sub-divided each tile into 20×20 sub-tiles. Additionally, [4] aimed to extract features characterising keratinocytes cells. In this Chapter texture and anatomical based features were calculated from tiles, rather than sub-tiles. The approach adopted in this Chapter requires less computational power and thus it is less computationally expensive for cubes of images to be examined using the method defined in this Chapter.

5.6 Future Work

In this Section possible improvements that can be made to the algorithm to improve the results achieved, together with an application of the algorithm are identified.

A possible application of the algorithm developed in this Chapter, is to design an extension to the algorithm that can detect papillae and quantify the melanin content inside the papillae (Section 5.6.1). However, for this extension to be possible the results achieved with the algorithm need to be improved using the methodologies identified below. The proposed extension would use the results achieved in the improved algorithm to firstly locate the DEJ region where papillae are located.

The results point to some key improvements that can be made to improve sensitivity and specificity in detecting the DEJ region in RCM images. In particular, the results point to the need for pre-processing methods to reduce variation in feature appearance and in image quality. A possible method to do this is to introduce additional clinical features that are known to be important (Section 5.6.2) whilst suppressing confounding features, which are currently hindering the classification performance (Section 5.6.3). These features can result from poor contact, wrinkles and hair. The proposed pre-processing steps will generate images whose attributes are purer, resulting in a smaller overlap across classes. In addition to this, variations to the classifier design can be made, which have the potential of improving classification results (these are discussed in Section 5.6.4).

5.6.1 Quantification of papillae population and melanin

On the basis that the results, more particularly the sensitivity of class DEJ, achieved from the algorithm developed in this Chapter can be improved. Then the algorithm could be used as a basis to extract parameters that quantify papillae population at the DEJ region and melanin content inside the papillae. The output of the improved developed algorithm can be used as input to the proposed algorithm extension. For example, images classed as DEJ can be fed to the algorithm that searches for a layer exhibiting edged or non-edged papillae. This can be achieved by using the information outputted by the monogenic filter and circle finder algorithm.

Once a layer exhibiting edged or non-edged papillae is identified, quantitative parameters characterising papillae can be extracted, such as number of papillae per unit area and circularity measure of papillae. By applying further image processing algorithms which are capable of identifying individual melanin inside the segmented papillae can provide clinicians with statistical quantifications of the melanin content. Such quantitative parameters have the potential of complementing clinicians' daily diagnostic decisions.

Possible methods which have the potential of improving the results achieved in this Chapter are identified in Sections 5.6.2 - 5.6.4.

5.6.2 Enhancing features

In this Chapter the monogenic filter was used as an edge detector; to enhance the visibility of edged and non-edged papillae. Features based upon the presence and absence of papillae were extracted from these edged representations (of the basic images), with the aim of differentiating the DEJ images from epidermal and dermal images. Papillae were searched for in the edged image using Hough circular detection. However, papillae are not always circular in shape and sometimes appear as ellipses. Possible improvements of papillae detection include adapting the Hough circular detection to cater for elliptical shapes too. A method to detect ellipses using the Hough circle transform was proposed by Cuingnet et al. [130], where an ellipse is detected as two proximal circles.

The edged image can be also used to estimate addition anatomical based features which will aid classification. For example: hair and wrinkles, which dominate the epidermal layers, can be searched for in the edged representation using Hough line detection. By applying this extension, epidermal layers can be characterised with linear structures; DEJ regions can be characterised by consecutive circles/ellipses; whereas dermal layers lack both linear and circular structures. However, as illustrated in Figure 5.17C, applying the edge detection on dermal images results in false edges being detected. This might result in false papillae being detected in dermal images. Additional features that aid in differentiating

between DEJ and dermal edges have the potential of aiding classification. Some literature strands suggest that an enhancing algorithm should only be applied to images belonging to the class characterised by the features it is enhancing, as the algorithm has the possibility of degrading structures characterising other classes [131].

5.6.3 Suppressing confounding structures

Calculating the texture features on images without any confounding structures results in purer attributes and the number of misclassification instances will reduce. The main confounding features found in RCM images originate from either poor contact or due to anatomical structures, such as wrinkles and hair (refer to Section 4.5.2). As discussed in Section 3.2.2, Sourati et al. [12] attempted to automatically locate wrinkles by using a probabilistic approach which utilises texture features and anatomical depth information.

The texture features used in this Chapter make use of all the pixels in the images and cannot be calculated on a random subset of pixels of an image. Whereas calculating texture features including masked out pixels results in less representative texture features. To cater for this, once the confounding feature region has been detected, an image completion method needs to be used to fill in the masked out region with data that matches the remainder of the image. Therefore, the use of a filling algorithm results in more distinctive texture features and better classification results. Komodakis and Tziritas [132] proposed an image completion method based on texture synthesis. The image completion algorithm requires a small input representing the feature of interest, and produces a larger output having the same visual characteristics as the smaller input.

5.6.4 Purer attributes and multi-labels

Another limitation of the current methodology is that some of the basic images used to train the classifiers can be characterised using two classes, however they are only assigned to the most dominant class. Xiong et al. [133] propose a number of methodologies on how to reduce the overlap across classes, one being

to discard data in overlapping regions, and train the classifier solely on non-overlapping data (images). Another suggested strategy is to generate a new class with overlapping regions, which will lead to the implementation of two classifiers. The first classifier indicated whether the image under consideration exhibits class overlap or not, whilst the second classifier determines which class(es) the image belongs to.

Boutell et al. [131] discuss the possibility of introducing multi-labels to instances that belong to more than one class. Knowledge about the class enables us to enhance certain qualities of this class. As opposed to Xiong et al., Boutell et al. believe that one should avoid introducing new classes, as it makes the problem more complex. Instead, when an image has characteristics of multiple classes, we should introduce it as a training instance in all the classes it characterises, resulting in more available training data. Using this method, one should train n-classifiers, where n is the number of base classes present. Each classifier discriminates between a class versus the rest, its output being a probability measure that calculates how likely an image belongs to the class it represents.

5.7 Conclusion

This Chapter describes an algorithm that automatically detects the DEJ region in a cube of RCM images using texture and anatomical based features. The main novelty of this Chapter that distinguishes this work from that discussed in Section 3.2.1, is that anatomical based features are incorporated. The method in which the anatomical features were extracted replicate the methodology used by clinicians when classifying an image into DEJ class; i.e. of locating individual papillae and tracking them across images captured at incremental imaging depths.

A potential application of this work could be to monitor changes across DEJ thickness across different body sites or across time (for the same body site). In addition, as discussed in Section 5.6.1, this algorithm could provide a framework for a papillae segmentation algorithm that provides a quantitative measure of

papillae population and melanin content.

Chapter 6

Identification of normal and pathological tissue in reflectance confocal microscopy images of excised tissue

Contents

6.1 Introduction 152

6.2 Background 154

6.3 Algorithmic Method 158

6.3.1 Segmentation algorithm 159

6.3.2 Selecting regions of interest (ROIs) 161

6.3.3 Automatically segmenting a whole image into regions . 163

6.3.4 Feature Extraction: Nuclei concentration 165

6.3.5 Feature Selection 167

6.3.6 Classification 167

6.4 Experimental Methods 168

6.4.1 Data 168

6.4.2 Labelling data 169

6.4.3 Evaluating the classifier 169

6.4.4 Experiment A: Classifier based on texture and nuclei concentration features 170
6.4.5 Experiment B: Classifier based on texture features 171
6.4.6 Experiment C: Classifier tested on whole images 171
6.5 Results 172
6.5.1 Results A: Classifier based on texture and nuclei concentration features 172
6.5.2 Results B: Classifier based on texture features 175
6.5.3 Results C: Classifier tested on whole images 177
6.6 Discussion 181
6.6.1 Main observations from Experiment A 181
6.6.2 Main observations from Experiment B 184
6.6.3 Main observations from Experiment C 184
6.6.4 Comparison with the literature 185
6.6.5 Clinical context of results 186
6.7 Future Work 187
6.8 Conclusion 188

6.1 Introduction

In this Chapter cancerous indicators and regions are searched for in reflectance confocal microscopy (RCM) images of excised tissue, collected during Mohs surgery. An algorithm was developed to classify regions of an image into three classes: (i) basal cell carcinoma (BCC) cancerous regions, (ii) regions of inflammation and (iii) non-diseased regions. The aim of the developed algorithm is to highlight areas in the sample which have a high likelihood of being cancerous, or of being indicators of disease. The classification technique makes use of texture and nuclei concentration features extracted from RCM image regions of tissue sample. Thus, as the algorithm has the potential of highlighting areas of concern to the clinician, it will aid the clinician in analysing the tissue sample faster.

The key aspects of Mohs surgery, which are discussed in depth in Section 3.3, are briefly outlined here. Mohs surgery is a surgical procedure that is used to excise skin cancers in aesthetically sensitive anatomical locations and to excise recurring tumours. In Mohs surgery once the tumour is removed, the borders of the excised tumour are analysed and if any part of the border is found positive of disease further excisions are done at the diseased boundary, resulting in $\sim 99\%$ of removed tumours not recurring [102]. Currently each excision is analysed with histology, which is time consuming and costly.

As outlined in Section 3.3, an alternative to histology slides is to image the excised tissue with RCM. The benefits of imaging the excised tissue instead of examining it with histology is that it eliminates the need of a wet lab and speeds up the process. By reducing the time required for Mohs surgery, this successful surgery can be extended to other tumours that are currently not being excised using this technique due to the costs involved.

Additionally, RCM permits the imaging of excised tissue using a fluorescent dye thus allowing the visualisation of individual nuclei, resulting in images comparable to histology. The advantage of imaging individual nuclei is that statistics on nuclei concentration can be extracted from the images and incorporated in the classification algorithm.

Two main methodologies which were designed around the clinical setting were implemented in this Chapter; (i) the user (clinician) can select a region of interest (ROI) which the classifier will classify, or (ii) the user (clinician) does not select a ROI and the classifier classifies the whole image.

A background and examples of RCM images of excised tissue is given in Section 6.2. The developed algorithmic method is discussed in Section 6.3. The experimental method is identified in Section 6.4, which covers the data collected, the developed algorithm used to label and train the data, and the experiments devised to test the classifier. The results achieved are tabulated in Section 6.5 and discussed in Section 6.6. Possible extensions to the proposed algorithm and

conclusions are identified in Sections 6.7 and 6.8 respectively.

6.2 Background

As opposed to the data used in Chapter 5 which made use of RCM images of *in-vivo* skin, here images of excised tissue collected during Mohs surgery are used. The excised tissue is generally less than $3mm$ thick and is normally extracted perpendicularly to the skin surface. Thus, the main difference between the two datasets is the orientation of the sample; excised tissue is imaged parallel to the sample in one 2D image, whereas *in-vivo* skin (as in Chapter 5) is imaged with multiple images parallel to the skin surface captured at different skin depths.

The dataset of images used in this Chapter are of BCC and non-diseased tissue samples collected during Mohs surgery as defined in Section 4.6.1. The non-diseased samples represent excisions which are found to be tumour-free. As outlined in Section 4.6.3 BCC classes include: infiltrative, micronodular, nodular and superficial BCC. Examples of RCM and histology images of these three BCC classes are illustrated in Figures 4.23, 4.24, 4.25 and 4.26 respectively. Due to the lack of visibility of epidermis in RCM images of excised tissue (Section 4.6.2.2), the number of RCM images depicting superficial BCC is limited and therefore this BCC class is excluded from this study.

In this Section the main skin structures found in RCM images of excised skin tissue are illustrated. These regions are classified as *other* to reflect non-diseased regions; *tumour* to reflect infiltrative, micro-nodular and nodular BCC; and *inflammation* to reflect the inflammatory cells found around a tumour.

A typical example of an RCM image of excised tissue is shown in Figure 6.1, where the regions enclosed by the red border represent tumour regions, covering a mixture of nodular and infiltrative BCC, whereas the regions enclosed by the turquoise border represent inflammatory response found around tumours. One key aspect to be noted from this image is that the user analysing such images can visually delineate the border of each structure.

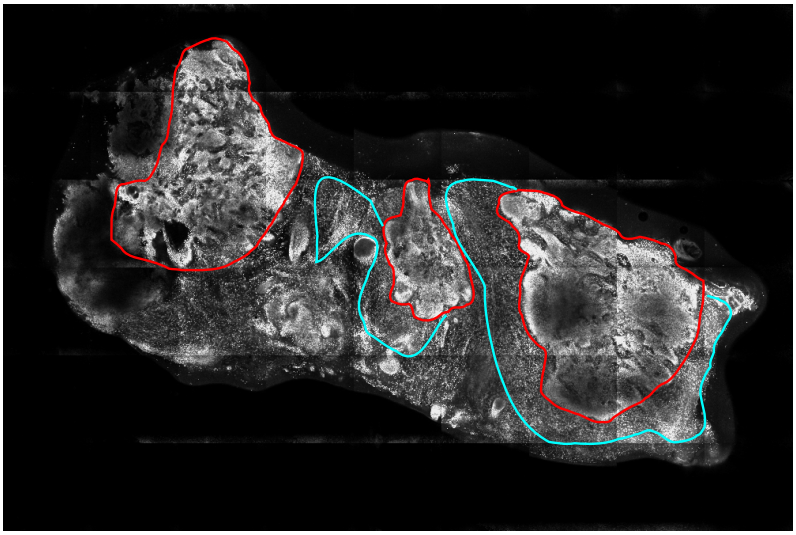


Figure 6.1: A typical RCM image of BCC excised tissue, displaying regions of class *tumour*, *inflammation* and *other*. The regions bounded by the red boundary are a mixture of infiltrative and nodular BCC, whereas the regions bounded by the turquoise boundary reflect inflammatory cells found around tumour regions. The rest of the tissue is non-diseased and can be classed as *other*.

The regions classed as *other* can be sub-divided into eccrine glands (Figure 6.2A and B); hair follicles (Figure 6.2C); sebaceous glands (Figure 6.2D); epidermis (Figure 6.3A); muscle (Figure 6.3B) and fat (Figure 6.3C). Eccrine glands are a coiled spiral duct; and represented by clusters of bright islands in RCM images. Hair follicles captured cross-sectionally are characterised by a dark inner circle (hair shaft) surrounded by a nuclear crowded bright region in RCM images [105]. Sebaceous glands can be distinguished by having an inner region with sparsely located and ordered nuclei and a brighter border in RCM images. The depicted sebaceous glands are clearly represented, except for the one indicated by the green border, which is captured cross-sectionally in the middle. In this case, the nuclei are not visible and difficult to distinguish if not located close to other sebaceous glands. A number of sebaceous glands are normally clustered together and are located close to hair follicles. The gland Figures in this Section are given in the best orientation that makes them easily distinguishable. For example, sebaceous glands imaged deeper cross-sectionally or hair follicles imaged in-plane are harder to discriminate.

The epidermis is located at the boundary of the tissue and appears nucleated

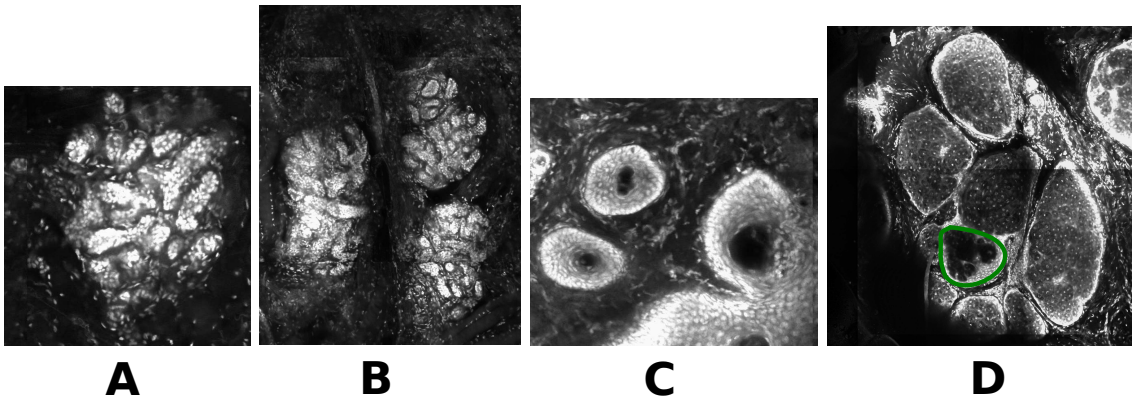


Figure 6.2: RCM images of excised skin tissue showing glands. (A) An individual eccrine sweat gland characterised by a coiled spiral duct and (B) a cluster of three eccrine glands. (C) Three cross-sectionally imaged hair follicles; the inner dark region represents the area from which the hair emerges, surrounded by a brighter region. (D) A group of sebaceous glands; sebaceous glands are characterised by ordered nuclei in the middle of the structure and with a bright border outside in RCM images. However, when captured cross-sectionally (e.g. the bottom left one indicated by the green border) the individual nuclei are not visible. Sebaceous glands can be connected to hair follicles.

in RCM images. A limitation when imaging excised tissue with the VivaScope® 2500 is that the epidermis layer is not always visible. This is mainly due to the fact that the epidermis-end of the tissue is less thick than the rest of the tissue. Muscle and fat occupy a large portion of the tissue sample. Muscle is characterised as elongated strands (Figure 6.3B) whereas fat cells are depicted as blobs gathered together (Figure 6.3C). Fat cells are easily segregated as they are not refractive except for their nuclei. To summarise, class *other* represents a range of structures with different nuclei concentration. The structure with the most concentrated nuclei in this class is the epidermis.

Regions classed as *tumour* cover micronodular, nodular and infiltrative BCCs and examples of class *tumour* are depicted in Figures 6.4 A and B. BCCs can at times be characterised as having nuclear crowding or increased nuclear density and these features can be used to distinguish BCC regions [134] [105]. In general, despite tumour regions being hyper fluorescent, individual nuclei cannot be individually differentiated in RCM images. At times, zooming in the image at a high magnification can aid in differentiating individual nuclei. Another characteristic of BCC is peripheral palisading which takes place at the the boundary of the tumour island [134]. Tumour regions can be surrounded by inflamma-

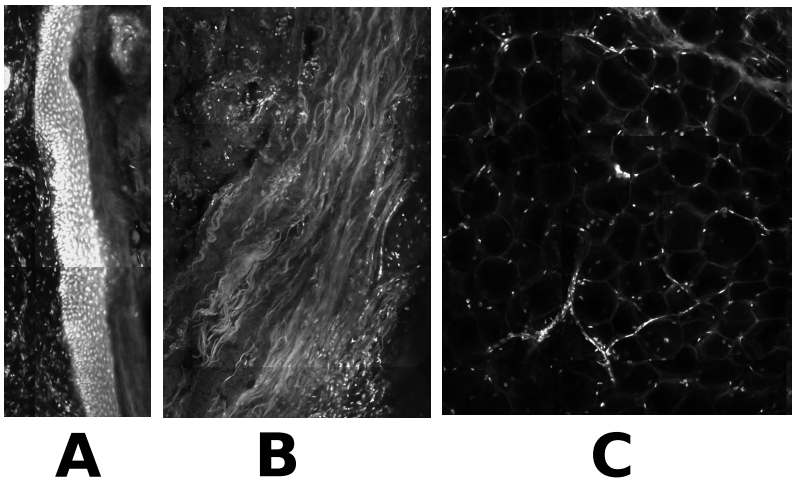


Figure 6.3: RCM images of excised skin tissue showing epidermis, muscle and fat. (A) Epidermis which is depicted as an elongated nucleated border. (B) Muscle is characterised by elongated strands and appear as different shades of grey in RCM images. (C) Fat is characterised by black blobs in RCM images. A number of individual nuclei are visible at the place where blobs connect together.

tion, which can aid in the identification of tumour. As indicated by Figure 6.4C, class *inflammation* is characterised by a concentration of nuclei as shown in Figure 6.4D.

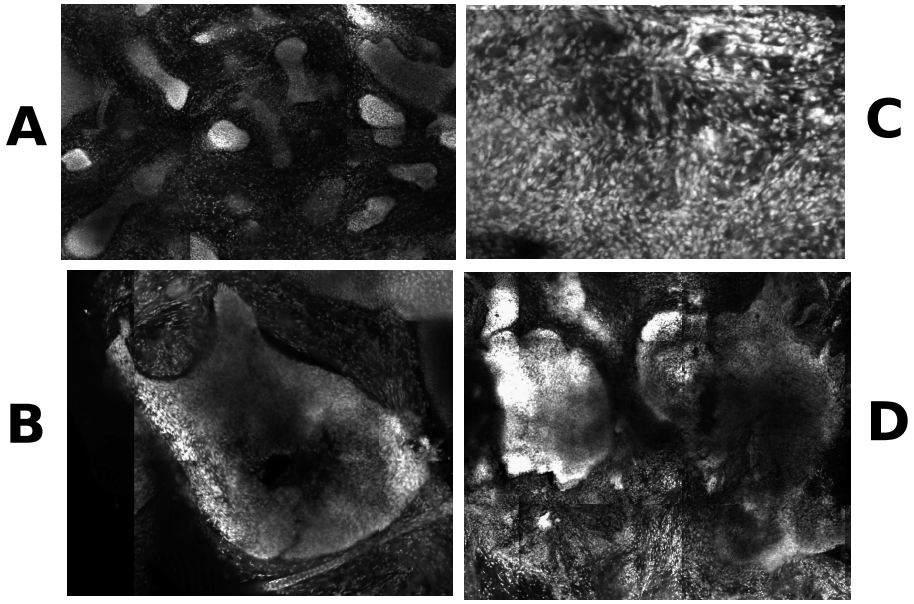


Figure 6.4: RCM images of excised skin tissue showing tumour and inflammation regions. (A) Micronodular BCC and (B) nodular BCC. Micronodular and nodular BCC islands are similar, however, the latter are larger in size. (C) Inflammation characterised by a concentration of nuclei. (D) Image illustrating how inflammation is interlinked to tumour.

6.3 Algorithmic Method

In this Section the algorithmic method developed to classify RCM images of excised tissue is explained. A block diagram of the steps involved in training and testing the classifier are depicted in Figures 6.5 and 6.6 respectively. The main differences between these pipelines to those used in Chapter 5 (Figures 5.1a and 5.1b), is that each image is segmented (Section 6.3.1) and regions of interest (ROIs) are selected from the segmented image (Section 6.3.2) prior to feature extraction. In addition, different anatomically based features, which in this case are based on nuclei concentration (Section 6.3.4), are extracted in this Chapter. The most discriminative features are selected by a feature selection algorithm (Section 6.3.5) and fed into a classifier (Section 6.3.6).

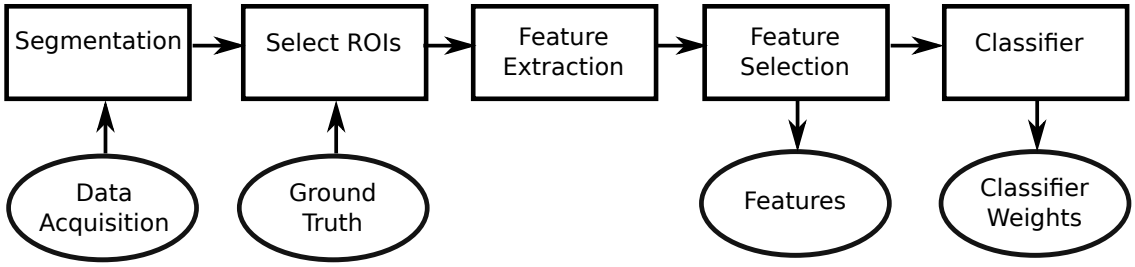


Figure 6.5: The pipeline of the algorithmic method to train the classifier which classifies RCM images of excised tissue. Oval shaped blocks illustrate input/output whilst rectangular shaped blocks indicate algorithmic processing. The additional steps in this pipeline to that used in Chapter 5 is that each acquired image is segmented and regions of interest (ROIs) are selected by the user using a Graphical User Interface (GUI) which was developed for this purpose. In addition, the GUI enables the user to assign a class (ground truth) to each selected ROI. These two additional steps (segmentation and select ROIs) cater for the fact that each image consists of regions belonging to different classes. Features are extracted from the selected ROIs, covering both texture and nuclei concentration based features. The most dominant features are selected and a classifier is built based on these selected features. The output of the trained classifier is a set of classifier weights which are then used to classify the testing datasets.

The additional steps (segmentation and select ROIs) are needed as the images used in this Experiment have regions belonging to more than one class. As outlined in Section 6.2, the images used in this Chapter are 2D images of the excised tissue. An image can be characterised by regions of tumour, inflammatory regions and non-diseased structures. In order to separate different parts (of an image) that belong to different classes a segmentation algorithm is used (Sec-

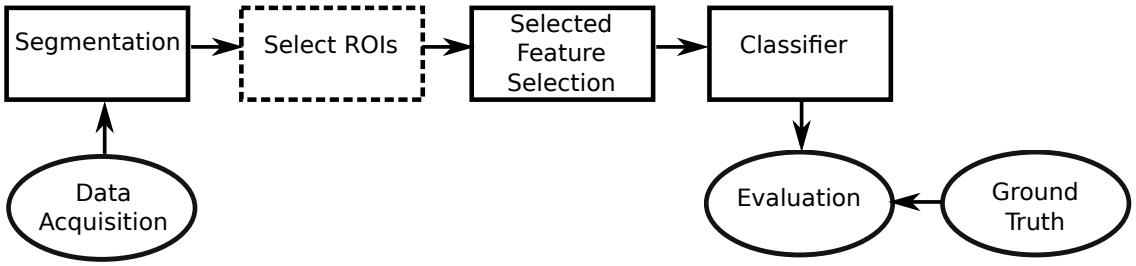


Figure 6.6: The pipeline of the algorithmic method to test the trained classifier illustrated in Figure 6.5. Each RCM image in the testing dataset is segmented. As opposed to the training step, the user can either feed a selected ROI or the whole image to the classifier. Thus the *select ROIs* step is optional and this is indicated by the dotted border. If the user wants the whole image to be classified, the image is automatically segmented in a number of regions. Features which were selected to be dominant features are extracted from each region and fed to the classifier. The classifier's output is a probability measure; for each region a probability measure of it belonging to each class is given and the region is assigned to the class having the largest probability. To evaluate the performance of the classifier on the testing dataset, the classifier's output is compared against the ground truth of each region.

tion 6.3.1). In the *select ROIs* step the user selects regions of interest from the segmented representation of the image, either for training or testing purposes. In addition, the user also has the option to test the classifier on a whole image, which is automatically segmented into regions.

6.3.1 Segmentation algorithm

As briefly outlined in Section 6.3 images of excised tissue are segmented in a number of regions. The purpose of the segmentation algorithm is to divide an image into regions where each region belongs to one class. The input to the segmentation algorithm is a 2D RCM image of excised tissue, and the output is a segmented representation of the image. This output will be the input to the subsequent algorithmic step defined in Section 6.3.2.

The segmentation algorithm used in this Chapter is based on contour detection developed by Arbelaez et al. and defined in [135]. The software written by the author is available online and implemented in MATLAB. For completeness I will give a summary of the algorithm in this Section. The algorithm makes use of brightness (grey scale intensity values) to define a gradient that defines the boundaries.

The boundaries of the image are detected using a circular disc split into two half-discs which was placed at each (x, y) location at eight equally spaced orientations (θ) in the interval of $[0, \pi)$. Four of these orientations are depicted in Figure 6.7. The gradient $G(x, y, \theta)$ is defined by the χ^2 distance between the two half-disc histograms g and h at orientation θ, as denoted in Equation 6.1. Value i is the intensity value and ranges from $[0, 255]$. $g(i)$ and $h(i)$ represent the number of pixels with intensity value i in histograms g and h respectively.

$$\chi^2(g, h) = \frac{1}{2} \sum_{i=0}^{255} \frac{(g(i) - h(i))^2}{g(i) + h(i)} \tag{6.1}$$

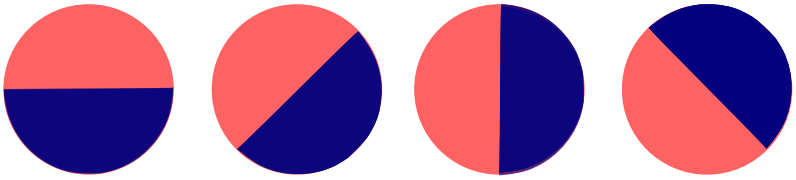


Figure 6.7: Orientated representations of the two half-discs used to calculate χ^2 at orientation $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$. The pixel values used to calculate histogram g (h) are indicated by the red (blue) half-circle.

For each (x, y) pixel location, the orientation (θ) that produces the strongest gradient (largest χ^2 distance) is considered to be the corresponding boundary strength (BS). The boundary strength BS normally does not result in closed sections. It is desirable for this Chapter that the segmented regions are closed. Edges which are close to each other, within a distance of 5 pixels, are connected together using a line. A segmented representation of Figure 6.8A is shown in Figure 6.8B. The effect of the BS value on the segmentation representation for the same RCM image is shown in Figure 6.9.

The output of the segmentation algorithm will be used as input to the *Select ROIs* algorithm and to a region growing algorithm (Section 6.3.2). In the subsequent algorithm (Section 6.3.2) a user generates a mask of a class by indicating (with a pointer) which bounded regions belong to the particular class. A region growing algorithm then makes use of the selected seeds to grow the region up to the closed boundary defined by this segmentation algorithm. Alternatively, the user can decide to classify a whole image which is automatically segmented into

regions as defined in Section 6.3.3.

6.3.2 Selecting regions of interest (ROIs)

A program together with a graphical user interface (GUI) has been developed for this thesis for clinicians to select ROIs from a segmented representation of the excised image. The GUI is used to select regions from an image to help produce the ground truths for training the classifier, i.e. this program will enable clinicians to assign regions to a label which will then be fed to training the classifier. Additionally, a user can select ROIs and assign them to an unknown label, after which the trained classifier can assign a label to them. The latter setting aims to simulate a potential clinical application for the work described in this Chapter where a user selects a ROI which he/she would like the algorithm to classify.

The purpose of the GUI, shown in Figure 6.8, is so that the the user can interact easily with the developed program. It gives a depiction of the RCM image being considered (A) and the segmented representation of the image (B). The boundary strength BS of the segmented representation can be varied by using the scroll bar labelled with (D). As the scroll bar is moved to the right more boundaries are visible, i.e. the threshold value which is used to threshold the segmentation image is reduced. Figure 6.9 illustrates the increased number of visible edges as the BS value is decreased from 0.25 to 0.15 to 0.05. The value is decreased until the specific value delineates the border of the ROIs. The scroll bar enables the BS value to be altered for each class. Once the threshold is set from the GUI shown in Figure 6.8, image (B) is converted into a binary image.

Once the user is satisfied by the strength of the boundaries, he/she selects the label of the ROIs using the side buttons. This will generate a pointer which the user can use to select the ROIs from the segmented image (B). Thus, the user will select n regions from image (B) that correspond to the selected label. The selected points (corresponding to seed points) are fed to a region growing algorithm which will grow the seeds to the selected boundary regions. Let the n seeds be represented using $S_1, S_2, ...S_n$.

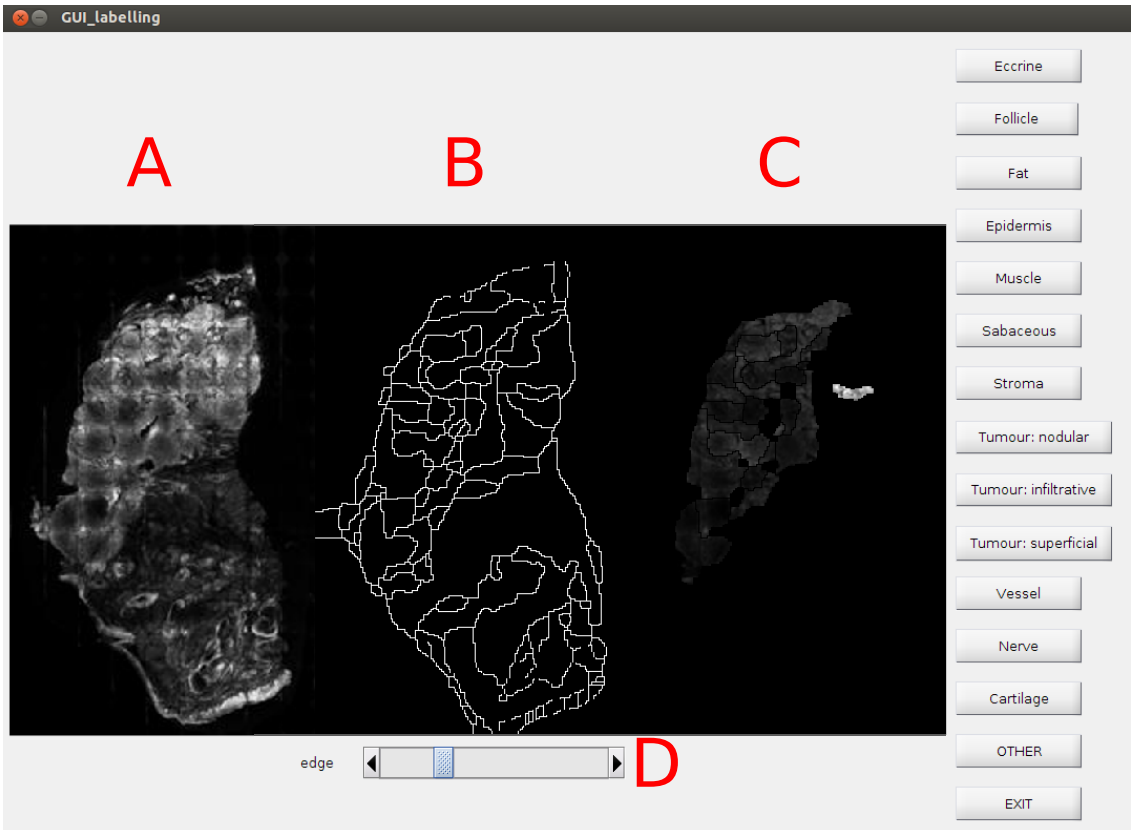


Figure 6.8: GUI window: (A) original RCM image of excised tissue (B) segmented image, the scroll bar indicated by (D) alters the boundary strength (BS) value. The user can scroll across the bar to choose a BS value that segments the structures in the best possible way, i.e. no structure that will be selected will belong to two classes. The effect of altering BS is depicted in Figure 6.9. The buttons on the right hand side depict the classes which the user can choose from. Once the user presses a button a pointing device appears and the user can select a number of regions in image (B). (C) represents the selected regions which will be saved as a binary mask.

Figure 6.10 depicts how three seeds S_1, S_2, S_3 are increased from three pixels to a collection of pixels over m iterations [136]. Initially, the seeds which are represented as black pixels in black in Figure 6.10A are assigned to mask M. The grey shaded regions around the seeds indicate the 8-connected neighbouring pixels. At each iteration, pixels which are a 8-connected neighbour to a pixel in M and which satisfy the homogeneous criteria are added to mask M. The homogeneous criteria ensures that the pixel to be added to mask M has a similar intensity value to the pixel to which it is connected to and which is already in mask M. The region stops growing at the boundary edge where the homogeneous criteria is not satisfied. Thus this algorithm ensures that the generated

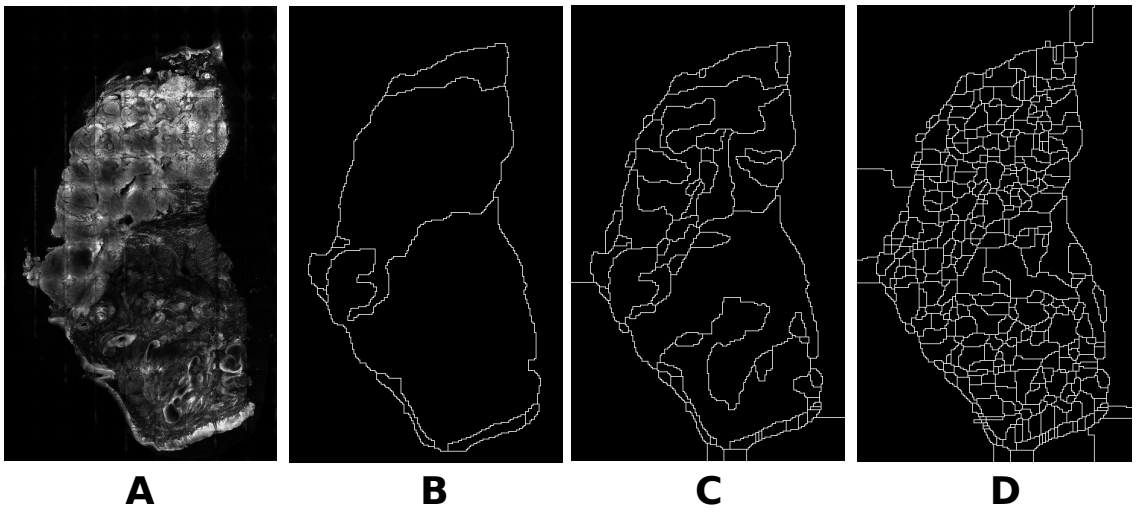


Figure 6.9: (A) original RCM image of excised tissue from which the segmentation is calculated. Thresholds 0.25, 0.15 and 0.05 are used for segmentations (B), (C) and (D).

mask initiated by seeds S_1, S_2, S_3 covers regions enclosed by the segmentation boundaries. To summarise, after m iterations the seeds in Figure 6.10A grow into the pixels in Figure 6.10B.

The area covered by the region growing algorithm is shown in image (C) in Figure 6.8. The main output of the GUI is a binary representation of image (C) with the associated class label. Figure 6.11 gives an example of the type of masks generated from the GUI illustrated in Figure 6.8. Figure 6.11 illustrates the mask selected in the example depicted in Figure 6.8; the mask indicates regions characterised by nodular BCC. This process is repeated for each class present in the image. From each region (of each mask and associated class) features are extracted, covering both texture features and nuclei population features.

6.3.3 Automatically segmenting a whole image into regions

In the cases where the user would like to test the classifier on a whole image, the test image is automatically segmented into regions. To generate the automatic testing regions, the testing image is segmented using a boundary strength BS value of 0.08. For computational processing, the generated regions are further sub-divided into square regions, this is illustrated in Section 6.5.3. The centre point of each segmented region was fed into a region growing algorithm, result-

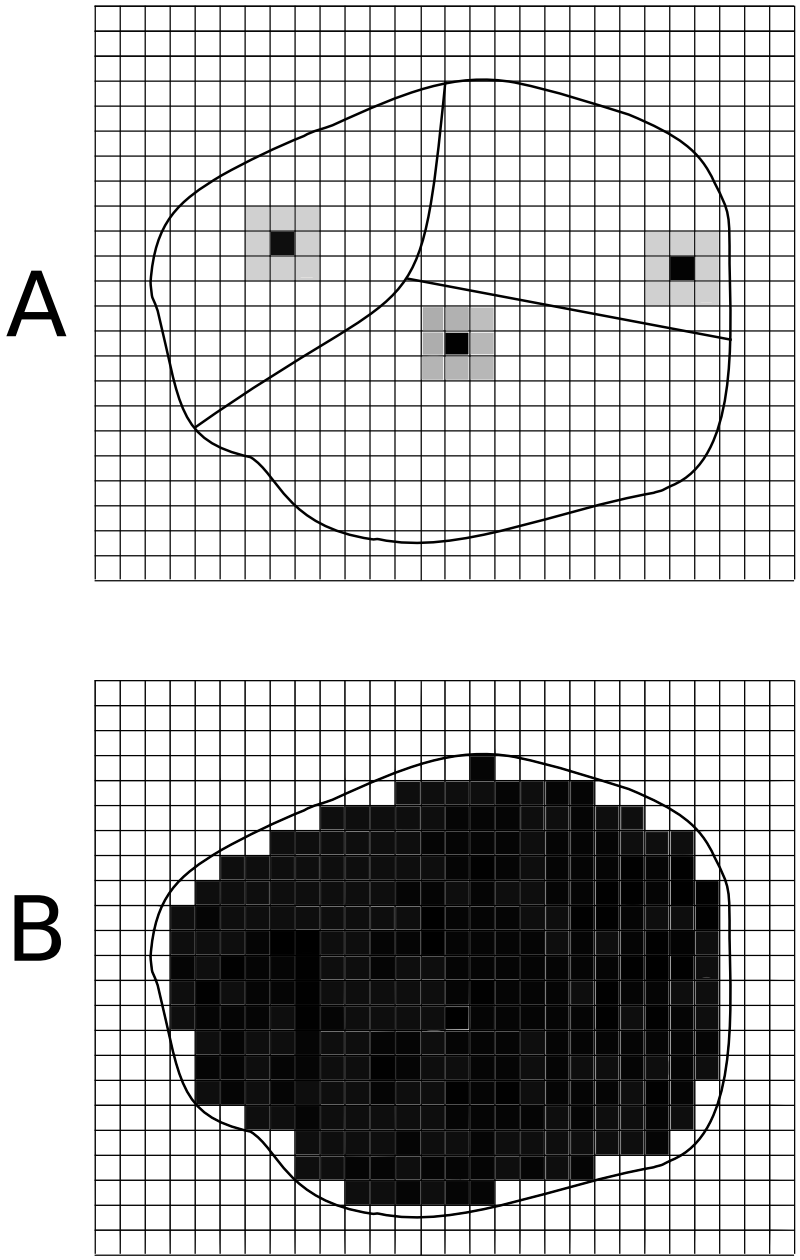


Figure 6.10: (A) Three seeds ($S_1 - S_3$) are selected within three regions (enclosed in a larger region), where S_1, S_2 and S_3 pixel locations are indicated by the black pixels. The 8-connected neighbouring pixels to the seeds are indicated by grey. (B) After m iterations, the pixels within the enclosed area are assigned to mask M. Pixels which are at the boundary and have a part of their area outside the boundary are not assigned to mask M; the homogeneous criteria is not satisfied for these pixels.

ing in each image being represented by a mask having a number of regions. The region centre points acted as seeds to the region growing algorithm (instead of the points selected by the user in Section 6.3.2).

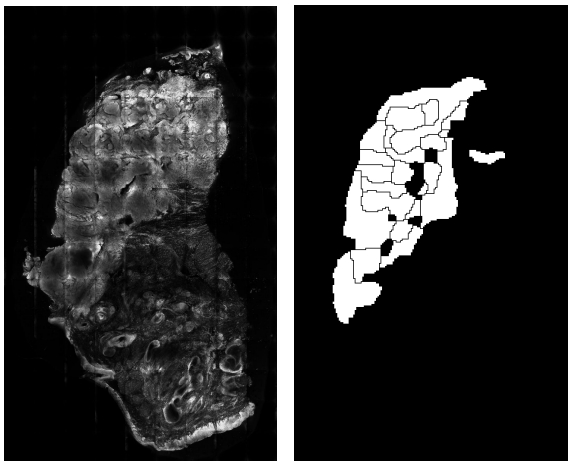


Figure 6.11: (A) original RCM image of excised tissue that is characterised by nodular BCC. This image is the same image illustrated in Figure 6.8A. (B) The mask generated in Figure 6.8C, which represents the areas having tumour of type nodular BCC. Features are extracted from each segmented region. For example, in this case the tumour is made up of 25 regions and features are extracted from each separate region.

6.3.4 Feature Extraction: Nuclei concentration

Features are extracted from each bounded region of each mask, covering both texture and nuclei concentration based features. The texture features extracted are the same as to those used for Chapter 5. However, the features quantifying the nuclei concentration are new to this Chapter and discussed in depth here. The purpose of detecting individual nuclei is to incorporate information about the different nuclei concentrations found across the three different classes. This Section is subdivided into two main parts: the first part describes the algorithm used to detect individual nuclei in RCM images and the second part identifies which features are extracted from the nuclei population detected.

6.3.4.1 Detecting individual nuclei in RCM images

The nuclei are individually detected using a maximally stable external regions (MSER) [137] algorithm which is based on the watershed segmentation algorithm and implemented in MATLAB. The MSER algorithm is widely used to detect blob-like structures. Although it makes use of intensity thresholds, isolated connected regions are not expected to be stable using the same threshold

value. Instead all thresholds are tested and the stability of the isolated region is evaluated [9]. A region is considered stable if its area changes only by a defined factor with the change of the threshold [138]. This is what distinguishes the MSER algorithm from the watershed algorithm [137]. Figure 6.12A shows a simplified image of a number of nuclei, and Figure 6.12B shows the output of the MSER algorithm to this image.

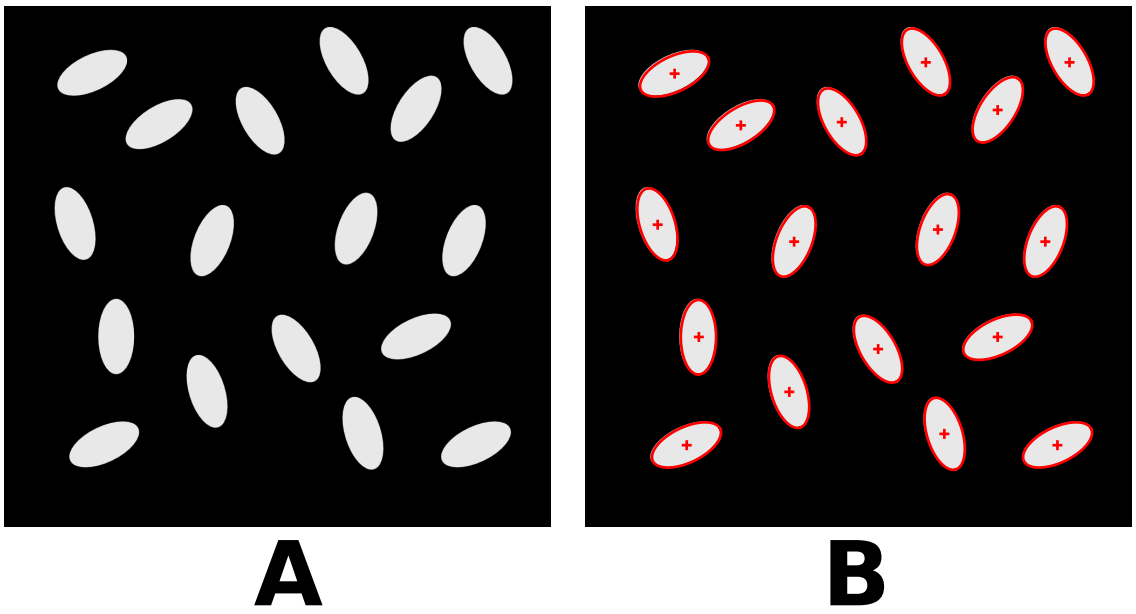


Figure 6.12: (A) A sample image of a number of nuclei (B) The MSER output of image (A). The red cross indicates the centre of each blob (nuclei) detected and the red border indicate the detected area.

6.3.4.2 Nuclei concentration features extracted from detected nuclei

Using the nuclei population detected by the MSER algorithm a number of parameters quantifying their concentration are extracted. Each region is taken in turn and a number of statistical parameters are calculated from the nuclei detected in the region. Focusing on the nuclei detected in a region the distance to the nearest neighbour (nucleus) is calculated for each nucleus (Figure 6.13A). Additionally, the mean point of the nuclei detected in the region is calculated, and the distances from each nucleus centre to the mean point are calculated (Figure 6.13B). Here the maximum distance d_{max} represents the distance between the furthest nucleus in the region to the mean point of the region. The minor and major axes are calculated from each detected nucleus as shown in Figure 6.13C. The values extracted from these calculations are:

- Number of nuclei detected per unit area (density);
- The mean, median, mode, interquartile range, range, standard deviation of the distance to the nearest nucleus. The red lines in Figure 6.13A represent the distance to the nearest nucleus.
- The percentage of nuclei located at a distance of $0- < 20\%$; $\geq 20- < 40\%$; $\geq 40- < 60\%$; $\geq 60- < 80\%$ and $\geq 80 - 100\%$ of the maximum distance to the mean point of the nuclei population. In Figure 6.13B, the green cross is the mean point and each line represents the distance from each nucleus centre to the mean point. The maximum distance is denoted by d_{max} in the figure.
- A set of nuclei having a neighbouring nucleus at a distance greater than 1 pixel and at a distance smaller than 20% of the maximum distance to the mean point are labelled as *inner* distances; the rest of the points are labelled as *outer* distances. The differences between the mean, median and mode of the inner and outer distances are used as parameters.
- The mean shape of each detected nucleus is quantified by dividing the length of the minor axis length by the major axis length (eccentricity). In Figure 6.13C the major and minor axis lengths are shown for each detected nucleus.
- The mean minor axis length and the mean major axis length of the detected nuclei are used as parameters.

6.3.5 Feature Selection

The most dominant features were selected using the correlated based feature selection [126] defined in Section 5.2.2.3.

6.3.6 Classification

Using a Bayes' classifier (defined in Section 5.2.2.4), the distribution of probabilities of the region belonging to class *tumour*, *inflammation* or *other* are given. Each

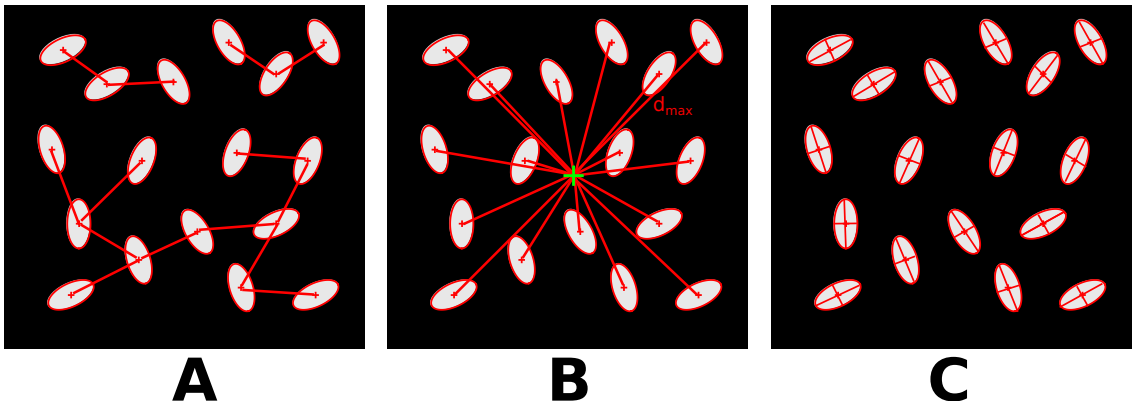


Figure 6.13: Representations of the distances and lengths from which the nuclei concentration based features are extracted: (A) For each nucleus in the region the distance to the nearest nucleus is calculated; (B) The distances from each nucleus in the region to the mean point of the population of nuclei (in the region) are calculated. Distance d_{max} represents the distance between the furthest nucleus in the region to the mean point of the region. (C) The major and minor axis are calculated from each nucleus in the region.

region is assigned to the class having the largest probability (out of the three classes).

6.4 Experimental Methods

6.4.1 Data

A confocalist technician at St. John's Institute of Dermatology prepared the tissues and *ex-vivo* images used for this study. As discussed in Section 4.6.1, the procedure used for the standard Mohs surgery was not disrupted. The confocalist imaged tissue samples characterised by tumour and samples without any disease. In this experiment 50 of these images were used:

- 30 images are tumour free
- 10 images are predominantly characterised by infiltrative BCC
- 10 images are predominantly characterised by micronodular or nodular BCC

Figures 4.23, 4.24 and 4.25 show examples of the infiltrative, micronodular and nodular BCC classes respectively. Some of the diseased cases have regions characterised by a mixture of BCCs, for example the tissue sample shown in Fig-

ure 6.1 is characterised by both infiltrative and nodular BCC. Due to the limited visibility of the epidermal layer outlined in Section 4.6.2, superficial BCC is not included in this study.

6.4.2 Labelling data

Each image was labelled using the algorithm and GUI identified in Section 6.3.2. When labelling an image reference was made to the diagnosis made by the surgeon and the corresponding histology slide is used as reference.

N.B. the user made use of the 14 available labels of the GUI when labelling the data. Therefore, the output of one image can be a maximum of 14 masked labels. However, a typical image only consisted of structures corresponding to 5-10 labels. The GUI was designed to cater for all 14 labels so that further Experiments could be performed.

For the purpose of subsequent steps, the masked regions are then grouped into three classes:

- *Tumour*: infiltrative, micronodular and nodular islands of BCC
- *Inflammation*: represents inflammatory response around tumour region [139]
- *Other*: representing non-diseased regions (any region not classed as disease or as inflammation)

In order to prevent obtaining a skewed classification, efforts were made so that each of the three classes is represented by a similar amount of regions.

6.4.3 Evaluating the classifier

In addition of calculating true positives, sensitivity and specificity values defined in Section 5.3.4, in this Chapter positive predictive value (PPV) and negative predictive value (NPV) are calculated as defined here. These values were calculating using confusion matrices, which can be represented using the matrix in Equation (6.2). The labels *Actual* depict the gold standard whilst the *Class* labels represent the classification given by the classifier.

$$\begin{array}{c} \\ Actual_{tumour} \\ Actual_{inflammation} \\ Actual_{other} \end{array} \begin{array}{c} \begin{array}{ccc} Class_{tumour} & Class_{inflammation} & Class_{other} \end{array} \\ \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \end{array} \tag{6.2}$$

Values a, e and i are the number of correctly classified images of class *tumour*, *inflammation* and *other* respectively. Values b and c are the number of *tumour* images which were incorrectly classified as *inflammation* and *other*. Values d and f are the number of *inflammation* images which were incorrectly classified as *tumour* or *other*. In turn, values g and h are the number of *other* images which were incorrectly classified as *tumour* or *inflammation*. The PPV and NPV of each class were calculated from these values as defined in Equations 6.3.

$$\begin{aligned} PPV &= \frac{TP}{TP+FP} \\ NPV &= \frac{TN}{TN+FN} \\ PPV_{tumour} = \frac{a}{a+d+g} \qquad & NPV_{tumour} = \frac{e+i}{e+i+b+c} \\ PPV_{inflammation} = \frac{e}{e+b+h} \qquad & NPV_{inflammation} = \frac{a+i}{a+i+d+f} \\ PPV_{other} = \frac{i}{i+c+f} \qquad & NPV_{other} = \frac{a+e}{a+e+g+h} \end{aligned} \tag{6.3}$$

6.4.4 Experiment A: Classifier based on texture and nuclei concentration features

In this Experiment a classifier is trained and tested on regions selected by the user. Thus the purpose of this experiment is to illustrate the potential of the algorithm for it to classify ROIs selected by the clinician using the developed GUI (defined in Section 6.3.2). As the user has selected the test regions, the user believes it is a good representation of a structure which should be classified into class *tumour*, *inflammation* or *other*.

From the 50 image cases labelled as described in Section 6.4.2, features cov-

ering both texture and nuclei concentration features were extracted from each bounded region. The classifier in this Experiment is tested using a leave-one-out approach, i.e. the classifier is trained on the most discriminative features extracted from 49 images and tested on the selected features extracted from the remaining test image. The test image is not used to train the classifier.

The results achieved from this Experiment are presented in Section 6.5.1. The results are represented in terms of True Positive, Positive Predictive Value, Negative Predictive Value, Sensitivity and Specificity.

6.4.5 Experiment B: Classifier based on texture features

The Experiment described in Section 6.4.4 is repeated but this time using only texture features (thus excluding nuclei concentration features). The motive of this Experiment is to investigate the effect on results of including nuclei concentration based features.

The results achieved from this Experiment are presented in Section 6.5.2. The results are represented in terms of True Positive, Positive Predictive Value, Negative Predictive Value, Sensitivity and Specificity.

6.4.6 Experiment C: Classifier tested on whole images

The purpose of this Experiment is to classify the whole image into *tumour*, *inflammation* and *other* regions. The application of such an Experiment is for a clinician to check whether the classifier highlights any regions of an image as *tumour* or *inflammation*. In this Experiment the performance of the classifiers on whole images (which are segmented automatically) is analysed. This Experiment has great potential in the clinical setting where time is limited; the clinician might not have sufficient time to select ROIs and would benefit from a map which highlights regions of concern.

For this Experiment all the 50 images are used as testing images on the classifiers trained in Section 6.4.4, which included features based on texture and nuclei

concentration. Therefore, each classifier that was built in Section 6.4.4 was tested on the whole image which was not used for training the classifier. The whole test image is automatically segmented into regions as defined in Section 6.3.3.

The output of this Experiment are visual representations (maps) highlighting regions in the image classed as *tumour*, *inflammation* and *other*. These maps are compared to the diagnosis made by the dermatologist of the tissue using both the RCM images and histology slides. However, there are no ground truths of the individual segmented regions and thus individual regions cannot be compared against a ground truth.

6.5 Results

6.5.1 Results A: Classifier based on texture and nuclei concentration features

The Results achieved from the Experiment defined in Section 6.4.4 are subdivided into two parts; (i) visual examples of individual nuclei detected with the MSER algorithm for each skin structure, and (ii) classification results achieved from the classifier identified in Section 6.4.4.

6.5.1.1 Nuclei detected with the MSER algorithm

In this Section examples of the nuclei detected with the MSER algorithm defined in Section 6.3.4 are shown. The images used in this Section are those used in Section 6.2 to define the main skin structures.

The nuclei detected in the gland examples are illustrated in Figure 6.14. The nuclei detected in eccrine, follicle and sebaceous glands are sparsely located. From Figure 6.14D, one can note that as opposed to the other sebaceous glands, the gland imaged cross-sectionally (indicated by the green border) has only a few nuclei detected inside it. Figure 6.15 gives an example of the nuclei detected in RCM images of the epidermis, muscle and fat. The concentration of nuclei in epidermis (Figure 6.15A) is larger than that detected in muscle and fat (Figures

6.15B and 6.15BC).

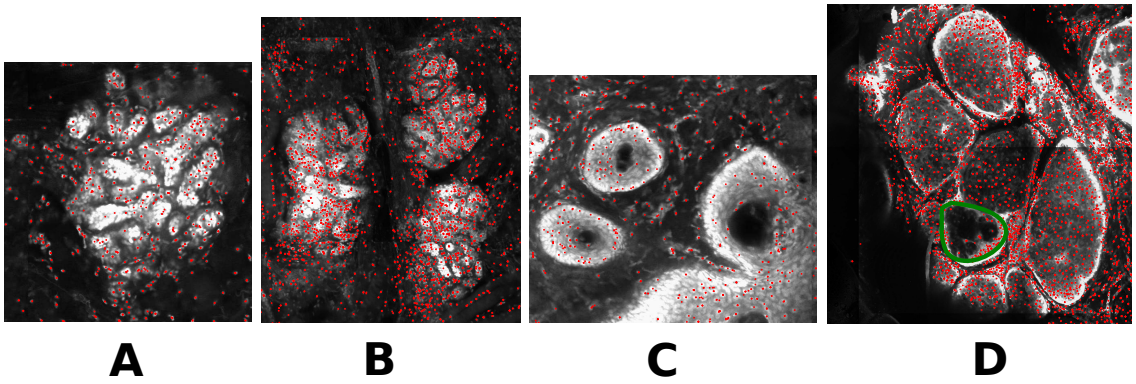


Figure 6.14: Nuclei detected in RCM images of excised skin tissue showing glands. (A) An indivdual eccrine gland and (B) a cluster of three eccrine glands. (C) Three cross-sectionally imaged hair follicles. (D) A group of sebaceous glands, one of which is captured cross-sectionally (indicated by the green border).

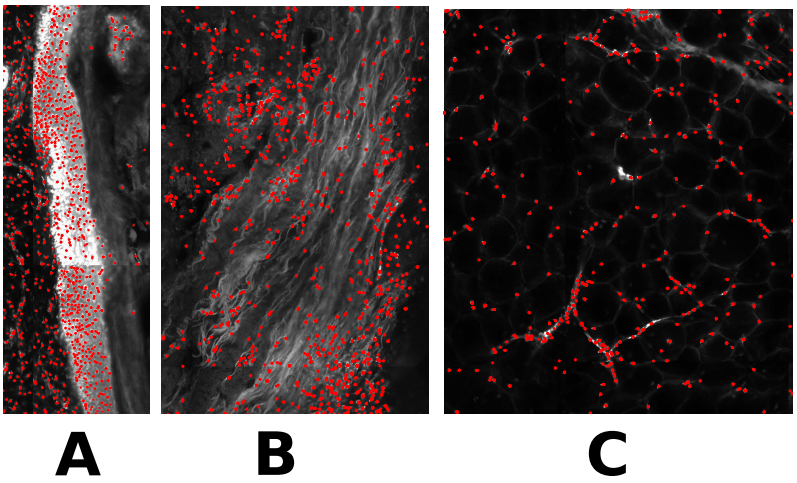


Figure 6.15: Nuclei detected in RCM images of excised skin tissue showing (A) epidermis, (B) muscle and (C) fat.

Figure 6.16 illustrate the nuclei detected in BCC (Figures 6.16A and B), together with the nuclei detected in inflammation (Figure 6.16C). As there is nuclei crowding in tumours, nuclei cannot be individually discriminated neither visually nor by the MSER algorithm. This observation is strongly present at the centre of the tumour island. Additionally, it can be noted that tumours (Figures 6.16A and 6.16B) are very similar visually and in terms of nuclei detected to sebaceous glands imaged cross-sectionally (indicated by the green region in Figure 6.14D).

Inflammation (Figure 6.16C) has a high concentration of detected nuclei, this concentration is generally higher than the concentration found in glands (Figure 6.14), muscle (Figure 6.15B) and fat (Figure 6.15C), however, it is similar to that detected in epidermis (Figure 6.15A). Figure 6.16D shows how the nuclei concentration around tumour regions is high due to inflammatory response. Therefore, the concentration of nuclei found in inflammation can be used as an indicator of disease.

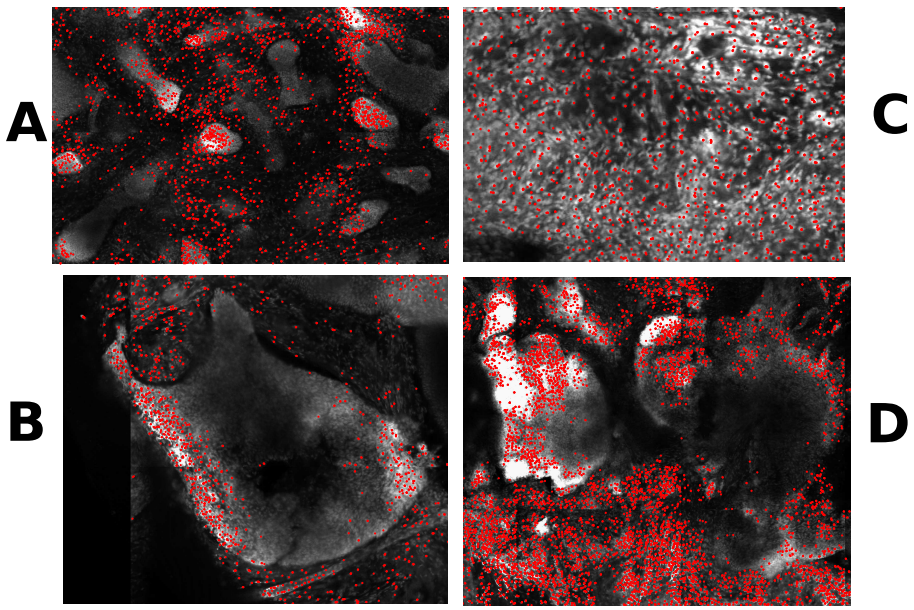


Figure 6.16: Nuclei detected in RCM images of excised skin tissue showing tumour and inflammation regions. (A) Micronodular BCC and (B) nodular BCC. (C) Inflammation characterised by a high concentration of nuclei. (D) Image illustrating how inflammation surrounds tumour.

6.5.1.2 Classification Results

In this Section the results achieved by the classifier described in Section 6.4.4 are tabulated. In Table 6.1 the True Positives, Positive and Negative Predictive Values, Sensitivity and Specificity of each class are tabulated. As the number of regions representing each class varies across the 50 images, a weighted average of these values is given.

In Table 6.2 the misclassified cases are looked at in more depth. The misclassification cases were grouped together based on the class they were misclassified as. The rows represents the true (actual) region class and the columns represents

class	TP	PPV	NPV	Sensitivity	Specificity
Tumour	58.3%	53.7%	81.8%	58.4%	92.8%
Inflammation	62.5%	35.9%	92.0%	62.5%	88.7%
Other	65.9%	81.6%	60.8%	65.9%	88.9%

Table 6.1: The true positive, positive predictive value, negative predictive value, sensitivity and specificity weighted percentages for each class across all 50 test images.

the classification result from the algorithm. For example, 58.3% of the *tumour* cases were correctly classified, 14.8% of the *tumour* cases were misclassified to *inflammation* and 26.8% were misclassified to *other*. The largest misclassified group was of *tumour* regions classified as *other*.

		Classification		
	class	*Tumour*	*Inflammation*	*Other*
Actual	*Tumour*	**58.3%**	14.8%	26.8%
	Inflammation	20.2%	**62.5%**	17.3%
	Other	16.9%	17.1%	**65.9%**

Table 6.2: The percentage of correct (bold) and incorrect classified regions across each class. The rows indicate the true class of the region, whereas the columns indicate the class that the classifier classed the region as.

The classifier assigns a probability of each test region being either of class *tumour*, *inflammation* or *other*. Currently, a region is assigned to the class having the largest probability. As depicted in Table 6.3, correctly classified regions are predicted with the highest probabilities. The average correctly classified *tumour*, *inflammation* and *other* regions have a mean probability of 0.87, 0.78 and 0.88 respectively.

6.5.2 Results B: Classifier based on texture features

In this Section the results achieved by the classifier described in Section 6.4.5 are tabulated. In Table 6.4 the True Positive, Positive and Negative Predictive Values, Sensitivity and Specificity of each class are tabulated. As the number of regions representing each class varies across the 50 images, a weighted average is given

		Classification		
	class	*Tumour*	*Inflammation*	*Other*
Actual	*Tumour*	**0.87**	0.73	0.83
	Inflammation	0.80	**0.78**	0.78
	Other	0.83	0.75	**0.88**

Table 6.3: The mean value of the classifier's probability across each class. The bold values indicate the probability value of correctly classified *tumour*, *inflammation* and *other*. The class column indicates the true class of the region, whereas the class row indicates the class that the classifier classed the region as.

of these values.

class	**TP**	**PPV**	**NPV**	**Sensitivity**	**Specificity**
Tumour	49.8%	55.0%	80.3%	49.7%	75.0%
Inflammation	0%	N/A	83.7%	0%	100%
Other	86.4%	69.2%	60.8%	86.5%	11.6%

Table 6.4: The true positive, positive predictive value, negative predictive value, sensitivity and specificity weighted percentages for each class across all 50 test images.

In Table 6.5 the misclassified cases are looked into more depth. The misclassification cases were grouped together based on the class they were misclassified as. The rows represent the true region class and the columns represent the classification result. For example, 49.7% of the *tumour* cases were correctly classified, 0% of the *tumour* cases were misclassified to *inflammation* and 50.2% were misclassified to *other*. The largest misclassified group was of *inflammation* regions classified as *other*, at a misclassification rate of 82.7% of the *inflammation* regions.

		Classification		
	class	*Tumour*	*Inflammation*	*Other*
Actual	*Tumour*	**49.7%**	0%	50.2%
	Inflammation	17.3%	**0%**	82.7%
	Other	13.5%	0%	**86.5%**

Table 6.5: The percentage of correct (bold) and incorrect classified regions across each class. The rows indicate the true class of the region, whereas the columns indicate the class that the classifier classed the region as.

The classifier assigns a probability of each test region being either of class

tumour, *inflammation* or *other* and the region is assigned to the class having the largest probability. As depicted in Table 6.6, correctly classified regions are predicted with the highest probabilities. The average correctly classified *tumour* and *other* regions have a mean probability value of 0.96 and 0.76 respectively.

		Classification		
	class	*Tumour*	*Inflammation*	*Other*
Actual	*Tumour*	**0.96**	-	0.75
	Inflammation	0.91	**-**	0.76
	Other	0.90	-	**0.76**

Table 6.6: The mean value of the classifier's probability across each class. The bold values indicate the confidence interval of correctly classified *tumour*, *inflammation* and *other*. The class column indicates the true class of the region, whereas the class row indicates the class that the classifier classed the region as.

The classifier defined in Section 6.4.5 was computed again using only a subset of the train and test datasets; only regions of class *tumour* and *other* were used to train and test the classifiers. The motive of doing this is to check whether the classification results improve when using a two-class classification model as opposed to a three-class one. In Table 6.7 the True Positive, Positive and Negative Predictive Values, Sensitivity and Specificity of the two classes are tabulated. As the number of regions representing the two classes varies across the 50 images, a weighted average is given of these values.

class	**TP**	**PPV**	**NPV**	**Sensitivity**	**Specificity**
Tumour	57.7%	77.1%	49.1%	57.7%	11.4%
Other	82.0%	85.8%	14.3%	82.0%	18.3%

Table 6.7: The true positive, positive predictive value, negative predictive value, sensitivity and specificity weighted percentages for each class across all 50 test images.

6.5.3 Results C: Classifier tested on whole images

In this Section visual representations of the results achieved in Experiment 6.4.6 are presented. In this Experiment the classifiers generated in Experiment 6.4.4 are

tested on regions generated automatically. The regions generated automatically cover the majority of the imaged tissue, and each region does not necessarily belong to only one class.

In Figures 6.17-6.19 the probability maps of three tissue samples are illustrated. In all Figures the (A) image represents the RCM image of the excised tissue, (B) represents the regions classed as *tumour* (red), (C) represents the regions classed as *inflammation* (blue), (D) represents the regions classed as *other* (green) and (E) the composite representation of the results. The brighter the colour the more certain the classifier is in its prediction. Note, although the classifier has a probability measure of each region belonging to each class (i.e. the output of the classifier for one region is three probabilities, corresponding to the three classes, which amount to a total of 1), it was decided to only plot the largest probability of each region as it is easier to interpret visually.

Figure 6.17 shows an image of excised tissue which has the majority of the area covered in infiltrative and nodular tumour islands. Areas of infiltrative and nodular tumour islands are enclosed by the red boundary. In examples such as this, the clinician can make a quick diagnosis as the *tumour* islands are visible throughout the specimen. This example presents a good classification result; the classifier classed the majority of the regions as *tumour*. In addition, the majority of the regions classed as *other* are tumour-free.

Figure 6.18 is an example of tissue exhibiting all three classes. The region enclosed by the red boundary in Figure 6.18A has a mixture of tumour and inflammation. It is difficult for the user to delineate the *tumour* and *inflammation* borders as both classes are intertwined. A large percentage of the area indicated by the red boundary is classed as *tumour* and *inflammation*, although some of this region is misclassified as *other*. Similarly, some regions outside the region delineated by the red boundary were misclassified to class *tumour* and *inflammation*. Therefore this classification depicts an example where the classifier did not achieve the desired classification.

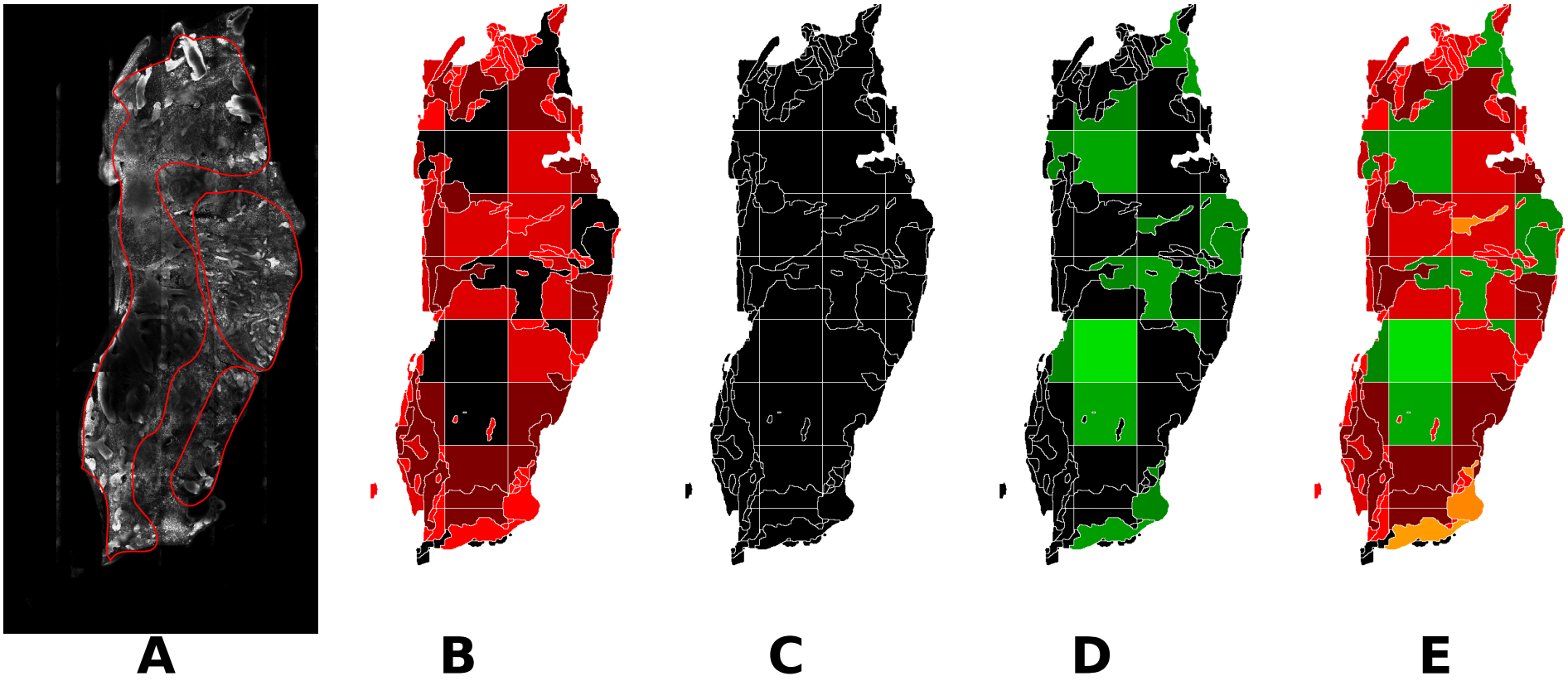


Figure 6.17: (A) An RCM image of an excised tissue covering regions of class *tumour* and *other*. The majority of the tissue is of class *tumour* (infiltrative BCC islands). The areas enclosed by the red boundaries cover regions of class *tumour*. Representations (B)-(D) represent the output of the classifier, black regions enclosed by a white boundary are the regions classified by the algorithm. (B) Regions classed as *tumour* by the classifier. (C) Regions classed as *inflammation* by the classifier. (D) Regions classed as *other* by the classifier. (E) The composite representation of plots (B), (C) and (D).

Figure 6.18: (A) An RCM image of an excised tissue covering regions of class *tumour*,*inflammation* and *other*. The area enclosed by the red boundary covers regions of class *tumour* and *inflammation*. Representations (B)-(D) represent the output of the classifier, black regions enclosed by a white boundary are the regions classified by the algorithm. (B) Regions classed as *tumour* by the classifier. (C) Regions classed as *inflammation* by the classifier. (D) Regions classed as *other* by the classifier. (E) The composite representation of plots (B), (C) and (D).

Figure 6.19 depicts a tissue sample which has a large number of glands and structures, the clinician needs to analyse each structure individually in order to examine whether the sample is diseased or not. This type of tissue sample is the most time consuming for the clinician to examine. Therefore, the clinician would benefit most if the generated masks for such samples are accurate and highlight true areas of concern. However, this sample was poorly classified; the only region that should have been classed as *tumour* is identified by the red circle in Figure 6.19A. However, one can note that the tumour islands generally had a large probability. One should note that in the examples depicted in Figures 6.17 and 6.19 no region was classed as *inflammation* (which matches the visual examination).

6.6 Discussion

In this Section the main conclusions from the results achieved in Section 6.5 are discussed. Overall, the classifier identified in Section 6.4.4 achieved better results than those achieved with the method identified in Section 6.4.5, this can be demonstrated by comparing Tables 6.1 and 6.4. For example, the TP rate of class *inflammation* increased from 0% to 62.5% by the introduction of nuclei concentration features.

6.6.1 Main observations from Experiment A

Focusing on the results achieved from the Experiment identified in Section 6.4.4, class *tumour* was misclassified as *other* for 26.8% of the cases; this is due to tumours having a similar visual representation as sebaceous glands. In addition, nuclei concentration across different tumours varies: in some tumour islands no nuclei can be individually segmented, whereas in other islands nuclei can be individually segmented out in the periphery of the island.

The results show that regions with small islands of infiltrative BCC are not segmented individually and have some background *inflammation*. Therefore, an infiltrative BCC region fed into the classifier consists of a group of infiltrative

Figure 6.19: (A) An RCM image of an excised tissue covering regions of class *tumour* and *other*. The area enclosed by the red boundary covers regions of *tumour* (infiltrative BCC islands). Representations (B)-(D) represent the output of the classifier, black regions enclosed by a white boundary are the regions classified by the algorithm. (B) Regions classed as *tumour* by the classifier. (C) Regions classed as *inflammation* by the classifier. (D) Regions classed as *other* by the classifier. (E) The composite representation of plots (B), (C) and (D).

islands with some background *inflammation*. This results in 14.8% of the *tumour* regions being misclassified as *inflammation* by the classifier, and 20.2% of *inflammation* regions being misclassified as *tumour* by the classifier.

Table 6.3 shows that misclassified probabilities are not different in their values than correctly classified ones. The two strongest mean probabilities correspond to correctly classified *tumour* and *other* cases, with values 0.87 and 0.88. The next two strongest probabilities, at a mean value of 0.83, are *tumour* regions misclassified as *other* and *other* regions misclassified as *tumour*. This illustrates that regions of class *tumour* and *other* are considered to be the most similar by the classifier.

The sensitivity and specificity values of 58.4% and 92.8% achieved by the algorithm (class *tumour*) can be compared against 96.6% and 89.2% achieved by Karen et al. [99]. Karen et al. [99] do not make use of an automated algorithm, but compare the clinicians' diagnoses based on the RCM image to the diagnoses made by histology slides. The comparison in results achieved highlights that the developed algorithm performs better than the clinician in detecting non-diseased regions correctly, whereas its performance is poorer than the clinician in detecting *tumour* regions in RCM images. Future possible extensions of this Experiment include putting more weight on the sensitivity rather than the specificity; i.e. the classifier puts more weight on correctly classifying a *tumour* region than correctly classifying non-diseased regions as the consequences of missing a *tumour* region are more severe than of misclassifying a non-diseased region.

The trade off between sensitivity and specificity can be analysed using a receiver operating characteristic (ROC) curve, where Sn is plotted against $1 - Sp$. Using ROC curves the user can choose the optimal sensitivity and specificity values of the classification algorithm for his/her application, which ideally corresponds to the highest sensitivity and specificity values available and which have the greatest clinical value in making a diagnosis [140]. A high sensitivity is value to detect true positive and not to miss true *tumour* regions, whereas a high specificity is also beneficial to avoid false positives; non-diseased regions being mistaken to *tumour*. Treating non-diseased regions as *tumour* regions is costly

and inconvenient to the patient.

6.6.2 Main observations from Experiment B

The results achieved from the Experiment identified in Section 6.4.5 illustrate the importance of including nuclei concentration based features to detect regions of class *inflammation*. Without nuclei concentration based features, no region was classed as *inflammation*, mainly because this class is strongly characterised by its nuclei concentration features, which were excluded in this study. Additionally the TP value of regions of class *tumour* decreased from 58.3% to 49.7% by the removal of nuclei concentration based features. Thus, one can conclude that features based on nuclei concentration aid the classifier, especially for class *inflammation*.

However, the TP of class *other* increased from 65.9% to 86.4% by the removal of nuclei concentration based features. Additionally, in this Experiment (Section 6.4.5) 50.2% and 82.7% of *tumour* and *inflammation* regions were misclassified as *other*. This increase is contributed to the classifier defined in Section 6.4.5 classifying most regions as *other* which contributes to the largest population of regions (out of the three classes).

6.6.3 Main observations from Experiment C

The probability maps generated in Experiment 6.4.6 show the potential of using the developed classifier algorithm in clinical practice. Such visual representations aid the user in interpreting the results achieved.

As discussed in Sections 6.4.6 the regions were generated automatically using a segmentation strength of BS equal to 0.08. These regions were further sub-divided in blocks for computational purposes. However, these regions are still considerably large. Therefore, future improvements include using a smaller threshold so that the image is subdivided into smaller regions which are more representative of a single class.

Another limitation of the algorithm generated to automatically segment the image is that it results in some regions not being included in the mask (depicted by white regions). Further improvements can ensure that the tissue is fully represented by the generated mask.

One main observation, is that using these large regions a relatively small number of images (7 out of 50 images) had regions classed as *inflammation*. This is because although a partition of the region exhibits areas of class *inflammation*, *inflammation* does not generally cover the whole area, and thus the calculated nuclei concentration measures calculated on the whole region reduce the probability of *inflammation* being detected.

6.6.4 Comparison with the literature

The most relevant paper aimed at detecting BCC regions was of Legesse et al. [10] with images captured using Coherent anti-Stokes Raman scattering (CARS) microscopy. Using texture analysis a sensitivity of 88% and specificity of 91% were achieved [10]. The main limitation of this study is that only five images (of five tissue samples) were used. From these images 35 sub-images of each class each of size 105×105 pixels were cropped, resulting in a total of 175 sub-images. In the Experiment identified in this Chapter the testing set of regions were extracted from a different image (of a tissue sample belonging to a different patient) than the images the training set of regions were extracted from. This contrasts the methodology used in Legesse et al. [10]. In [10] checks were not made to ensure that the testing dataset was extracted from a different tissue sample than those in the training dataset. Therefore, the results presented in [10] were based on testing sub-images extracted from the same images used to train the classifier. Furthermore, there is the possibility that all the 35 sub-images of class *tumour* could have been extracted from one tumour island (in one image). In this case, the classifier would have been trained and tested on sub-images extracted from the same tumour island for class *tumour*.

Dobbs et al. [9] made use of RCM images of excised breast tissue to segment

out parameters indicative of abnormalities. This study made use of nuclei concentration to differentiate between benign and malignant tissue. A sensitivity and specificity of 75% and 93% were achieved. As opposed to the Experiment in this Chapter, Dobbs et al. made use of the MSER algorithm to detect nuclei to identify which nuclei are connected together, i.e. which nuclei belong to the same structure. From the resulting connected components, a number of statistically significant parameters were extracted based on nuclei concentration and structure in order to differentiate between abnormal and benign structures. This approach could be implemented as individual nuclei can be detected in the breast abnormal structures. As BCCs typically have nuclear crowding [134], it is difficult to segment the individual nuclei in BCC RCM images and thus the approach defined in Dobbs et al. cannot be easily implemented for regions of class *tumour*. However, it is relevant to the process of identifying regions of classes *inflammation* and *other*, where individual nuclei can easily be segmented.

6.6.5 Clinical context of results

The results achieved in prior literature [99] [100] by dermatologists analysing RCM images of excised tissue as opposed to histology indicate that the information present in RCM images can be used to aid in making a diagnosis. For example, Karen et al. [99] achieved a sensitivity, specificity, positive and negative predictive values of 96.6%, 89.2%, 92.3% and 94.7% respectively. Similarly, Bennàssar et al. [100] reported a sensitivity and specificity of 88% and 99%. However, for RCM images of excised tissue to be used instead of histology to direct Mohs surgery, a 99% sensitivity value needs to be achieved [26].

At present the desired sensitivity 99% [26] is not being achieved by clinicians analysing RCM images of excised tissue. As discussed in Section 3.3.7, the desired sensitivity cannot be achieved unless information from a number of modalities (such as reflectance and fluorescence RCM modes) is combined together. Additionally, further improvements need to be made to the image quality achieved with the current RCM technology, including advancements to ensure the visibility of the epidermis layer. At present, RCM images cannot replace histology

examinations for all stages of Mohs surgery.

The RCM technology can be used to examine the skin *in-vivo* prior to the surgery to examine the tumour margin and also to estimate the depth of the tumour. In addition, a possible application of RCM images of excised tissue in the surgical setting is to analyse intermediate excisions of Mohs surgery using RCM. While histology would be used to analyse the final excisions and therefore to verify that the final margins are tumour-free. In the suggested application, RCM would be used to check whether the sample has tumour, and if yes a further excision is done, whereas if no tumour is found the sample would be prepared histologically to verify that the sample is tumour free. This set-up is possible as the sample can still be prepared for histological examination after being imaged with the RCM technology.

Using the application defined above, the clinician could examine the RCM image, after which the image is fed to the classification algorithm defined in this Chapter. Computer-aided diagnosis, where a computer algorithm serves as a second reader, are gaining acceptance in the medical field especially in screening mammography [141].

6.7 Future Work

When classifying a structure in an RCM image, clinicians sometimes make use of neighbouring information to make the classification. For example, if the clinician is not sure if a structure is a sebaceous gland or a nodular BCC, the clinician looks at neighbouring sebaceous glands to compare their appearance. The Experiment identified in this Chapter does not make use of neighbouring knowledge when classifying each region, instead it treats each region independent from the rest. Future work can include iterating classification results by running the classifier a second time, where classification results of neighbouring regions are fed into the classifier.

A possible extension of this algorithm is to sub-divide class *other* into a num-

ber of sub-classes. At present, class *other* might be hindering the classification results as the class consists of various skin structures which appear different. A possible way to divide class other is into three sub-classes covering glands, muscle and fat. These three structures are visually different and are located at different parts of the tissue. In general glands are located closer to the epidermis, whilst muscle and fat are located in deeper dermis. Thus, if an algorithm that orients the image can be developed, i.e. locating the epidermis and dermis sides of the image, the classifier can incorporate this spatial knowledge when making a classification decision on a ROI. The classifier can search for glands closer to the epidermis, whereas it can search for muscle and fat further away from the epidermis. Thus this extension has the potential of improving classification results of regions of class label *other*.

Mohs surgery involves multiple excisions of tissue layers from the same site. The surgeon diagnoses each excision; by checking whether it has regions characterised by tumour or not. In case tumour regions are present a further excision is taken. This process is repeated until it is verified that the site is tumour free at all margins. The tumour structures found across consecutive excisions are similar in appearance and in nature. A possible tool that has the potential of aiding clinicians in analysing subsequent images quicker is by developing a pattern recognition algorithm. The clinician selects tumour regions (ROIs) in the RCM image of the first excised tissue, and the algorithm finds regions similar to the ones selected by the clinician in subsequent layers. For this proposed algorithm, the the clinician will select ROIs from the first excision of each surgery. Thus, the proposed algorithm is patient specific, and removes the high inter-variability found across different patients. Such an algorithm has the potential of reducing the time clinicians take to make a diagnostic decision.

6.8 Conclusion

In this Chapter I developed an automated algorithm to highlight tumour regions and inflammatory regions in RCM images of excised skin tissue. According to my knowledge, this is the first attempt in the literature to apply automated analy-

sis to RCM images of excised tissue. This Chapter demonstrates the advantages of using nuclei concentration features in conjunction to texture-based features, as opposed to using texture-based features alone. With the introduction of nuclei based features, the number of correctly detected *tumour* and *inflammation* regions increased by 8.5% and 62.5% respectively. The potential of using RCM images of excised tissue as an alternative to histology for intermediate stages of Mohs surgery is highlighted. The benefits of using the developed algorithm in conjunction to the diagnosis made by the surgeon (on RCM images of excised tissue) has the potential of reducing the number of missed cases, which is estimated at 5.3% [99]. Eliminating the need for taking histology slides at each stage of Mohs surgery helps to reduce the duration and cost of the surgical procedure.

Chapter 7

Automated detection of skin capillaries in reflectance confocal microscopy video sequences

Contents

7.1 Introduction 191

7.2 Background 194

7.2.1 Appearance of RBCs in RCM 194

7.2.2 Artifacts in RCM video sequences 196

7.3 Algorithmic Method 198

7.3.1 Introduction 198

7.3.2 Design of algorithm to detect skin capillary area 200

7.3.3 Calculation of individual capillary parameters 209

7.3.4 Design of algorithm to quantify RBC velocity across cross-sectionally imaged capillaries 210

7.4 Experimental Method 213

7.4.1 Data 215

7.4.2 Validation 220

7.4.3 Total capillary area vs UVR irradiated dose 220

7.4.4 Total capillary area vs a^* value 221

7.4.5 Calculating parameters from individual capillaries . . . 222
7.4.6 RBCs velocity in cross-sectionally imaged capillaries . . 223
7.5 Results . . . 224
7.5.1 Total capillary area vs UVR irradiated dose . . . 224
7.5.2 Total capillary area vs a^* value . . . 230
7.5.3 Calculating parameters from individual capillaries . . . 232
7.5.4 RBCs velocity in cross-sectionally imaged capillaries . . 237
7.6 Discussion . . . 240
7.6.1 Introduction . . . 240
7.6.2 Discussion of main findings of algorithm to detect skin capillary area . . . 241
7.6.3 Discussion of main calculations of individual capillary parameters . . . 242
7.6.4 Discussion of main findings of algorithm to quantify RBC velocity across cross-sectionally imaged capillaries . . . 244
7.6.5 Limitations in algorithm and methodology . . . 245
7.7 Future Work . . . 250
7.8 Conclusion . . . 251

7.1 Introduction

Skin capillaries located at the dermal-epidermal junction (DEJ) are connected to vessels at the dermal layers, as illustrated in Figure 2.2. As denoted in Section 2.2.3, the flow of red blood cells (RBCs) passing through the superficial capillaries is dependent upon skin conditions and inflammatory response, whereas the flow in the deeper dermis is more dependent on the cardiac blood flow cycles. Clinically there is great interest in calculating quantitative values of the blood flow inside the capillaries located at the DEJ, to assess changes in micro-circulation due to environmental insults. For example, the micro-circulation at the DEJ depth will influence the level of erythema observed on the skin due to ultra-violet radiation (UVR) irradiation. Currently the minimal erythema dose (MED), which

represents the UVR dose needed to cause minimal changes of redness of the skin, is assessed visually or measured using a colour-meter. Clinicians and researcher would benefit from a measure of a more fundamental physiological change in response to UVR than a skin colour change. A measure of increased flow in the capillaries is the direct physiological change that causes erythema and, as such, may be a more valuable end-point than erythema. Currently there are various technologies that were discussed in depth in Chapter 2 that provide a quantitative measure of blood flow and levels of erythema. However, unlike RCM, they are not capable of distinguishing the blood flow between the different skin depths, nor do they provide any morphological information of nearby structures. For example, Laser Doppler Flowmetry (LDF) cannot discriminate between the flow going through the DEJ and deeper dermis, while it provides the user with a quantitative value it does not provide the user with any anatomical information. Capillaroscopy, on the other hand, enables visualisation and quantification of the skin capillary flow, however this method is limited to the nailfold site. RCM has the potential of imaging the micro-circulation of the most superficial skin capillaries located at the DEJ, together with providing anatomical information of the layer at which the capillaries are imaged.

Each capillary loop supplies a particular dermal-epidermal papillae and the arterial and venous capillary limbs are parallel to the papillae wall (Figure 2.2). In RCM video sequences capillaries are located at a number of orientations and different depths. As illustrated in Figure 4.14 a typical RCM scan can capture capillaries both cross-sectionally and in-plane. In turn, when capturing capillaries cross-sectionally it can either be at the peak, where the arterial and venous limbs merge, or across the middle of the capillary where the limbs are separated at a distance. These various representations, and how they are imaged in RCM video sequences, are illustrated in Figure 4.13. It is easier to track the flashes across a capillary imaged cross-sectionally than a capillary in-plane. In general, capillaries captured cross-sectionally show the flash of RBCs in approximately the same area of the image, while those imaged in-plane show the flash across different points of the larger in-plane capillary area seen. In addition to this, in-plane capillaries are encapsulated in tissue, whereas those captured cross-sectionally are

surrounded by papillae imaged cross-sectionally (darker regions), refer to Figure 4.13. The papers that were described in Section 3.4 focus on capillaries captured cross-sectionally (turquoise B line in Figure 4.12) and do not acknowledge the presence of in-plane capillaries.

In this Chapter an algorithm that detects skin capillaries, orientated either cross-sectionally or in-plane, located at the DEJ depth in RCM video sequences was developed. Capillary parameters extracted from the RCM video sequences include the percentage of frame area covered in skin capillaries and the individual skin capillary area. As opposed to the prior work described in the literature review [107] [39] [45] which quantify capillary size, the developed algorithm is fully automated, and is capable of locating both capillaries captured cross-sectionally and in-plane. A second algorithm has been developed which automatically calculates the velocity at which the RBCs travel at inside capillaries imaged cross-sectionally. This is the first study to quantify the RBC velocity using the RCM technology.

Furthermore, as opposed to the prior work described in the literature [108] [107] [50], the developed skin capillary detection algorithm were tested on video sequences collected from both fair and dark skinned volunteers, where each volunteer was irradiated by incremental UVR doses. By using the collected datasets the algorithm was tested on whether it detects capillary changes corresponding to vasodilation and whether it can detect capillaries in videos collected from both fair and dark skinned volunteers. Similarly, RBC velocity was quantified in capillaries captured from sites irradiated with incremental UVR doses. Thus the algorithm was tested on whether the algorithm detects RBC velocity changes corresponding to vasodilation.

Section 7.2 will provide additional RCM and clinical background information that is directly relevant to this Chapter. An explanation of how the detection algorithm was implemented is given in the experimental background in Section 7.3. A detailed overview of the data collected for this work and the validation techniques used to validate the algorithm is given Section 7.4. The results

achieved are tabulated and discussed in Sections 7.5 and 7.6.

7.2 Background

This Section expands upon the information provided in Chapters 2 and 4 to highlight aspects particularly relevant to the developed automated skin capillary detection algorithm. In RCM video sequences capillaries are characterised by abrupt fluctuations in pixel intensity over time (flashes); this variation corresponds to the presence and absence of RBCs travelling through the capillary. The appearance of RBCs in RCM video sequences will be incorporated in the developed algorithm; as circular bright objects will be searched for. However, variation in capillary spatial location across time due motion in the video (Section 7.2.2.1), and confounding features (Section 7.2.2.2) make it harder to locate capillaries automatically. Additionally, melanin, which has a higher concentration in volunteers with skin types IV-VI makes it harder to locate capillaries due to the melanin's high reflectance properties (Section 7.2.2.3).

7.2.1 Appearance of RBCs in RCM

RBCs appear as circular in shape in 2D RCM video sequences (refer to Section 2.2.4). Capillary lumen are not visible in RCM video sequences, however, clinicians can visually estimate the location, area and width of the capillary lumen by noting the highly reflective RBCs passing through the capillary. The passing RBCs appear as flashes of high pixel intensity in RCM video sequences as they flow in and out of the imaging field.

Capillaries can be imaged in various orientations in RCM video sequences, including capillaries imaged cross-sectionally and along their length (also referred to as in-plane in this thesis). As illustrated in Figures 7.1 and 7.2, cross-sectionally imaged capillaries appear more circular than capillaries imaged along their length, which are more elongated and oval in shape.

Figure 7.1 shows five frames of a capillary imaged cross-sectionally, each frame at an intervals of approximately 5s. The frames were chosen to illustrate the presence and absence of the flashing RBCs. The two consecutive green boundaries represents the edged papillae, whereas the areas represented by the red circled regions represent the arterial and venous limbs of the capillary. The top limb is the arterial limb and the bottom limb is the venous limb. In this example, the flashes in the arterial limb have a larger intensity value and are present across the five consecutive frames; the flashes in the venous limb are dimmer and only present in two out of the five frames (2nd and 4th frames). When RBCs are not present in the area enclosed by the red circles, the capillary limbs are darker than their surroundings.

Figure 7.2 represents five frames of a capillary along its length, each frame at an interval of approximately 5s. The frames were chosen to illustrate the presence and absence of the flashing RBCs across the capillary length. As opposed to capillaries imaged cross-sectionally, the capillary shown in 7.2 is encapsulated in epidermal like structures, making the latter more difficult to distinguish and delineate.

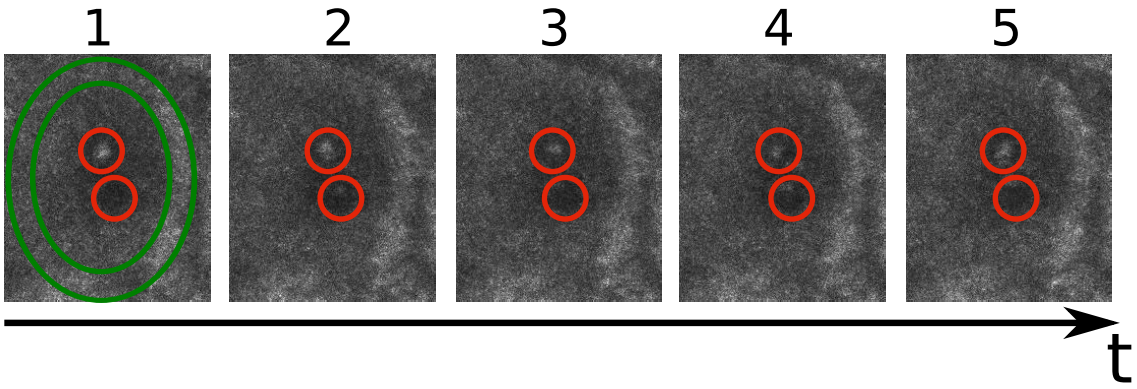


Figure 7.1: An example of an RCM video sequence of a cross-sectionally imaged capillary. Each frame is at an interval of approximately 5s and covers an area of $\sim 100 \times 150 \mu m^2$. The red circles represent the arterial and venous capillary limbs which are within the edged papillae (highlighted in green). The limbs can be distinguished from their surrounding as they are darker. The RBCs flow within the red area.

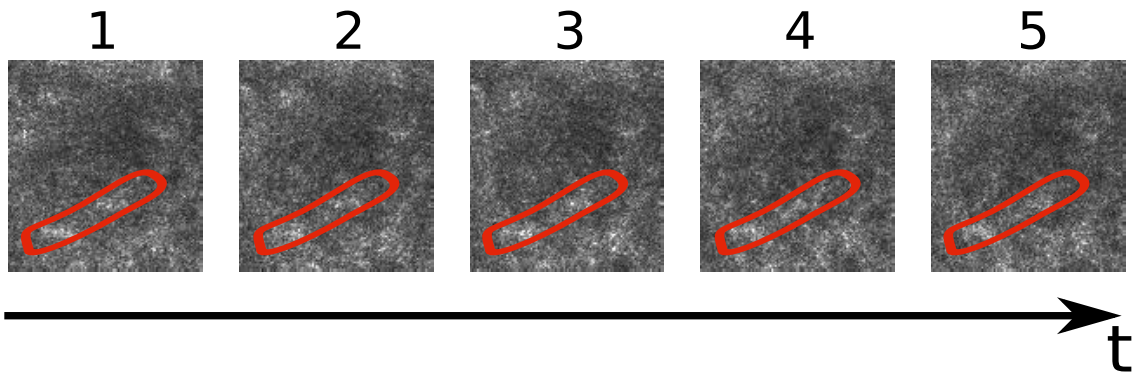


Figure 7.2: An example of an RCM video sequence of a capillary imaged along its length. Each frame is at an interval of approximately 5s and covers an area of $\sim 100 \times 100 \mu m^2$. The red shape represents the boundary of the imaged capillary surrounded by epidermal like structures. As opposed to the capillary imaged cross-sectionally, the boundary of the capillary imaged along its length cannot be clearly defined. As a chain of RBCs can be imaged in this example, a large part of the capillary area can be bright (representing the presence of RBCs).

7.2.2 Artifacts in RCM video sequences

In this Section the main artifacts present in RCM video sequences that confound the visualisation of RBCs are discussed.

7.2.2.1 Motion in RCM video sequences

In this Chapter capillaries are searched for in RCM video sequences. A video sequence is a representation of a single plane, covering the same $0.5 \times 0.5 mm^2$ area, over time. The designed and implemented algorithm analyses intensity temporal sequences, which highlight the presence and absence of RBCs, of the same pixel location. Thus, it is important that the sequences are extracted from the same spatial location. However, motion is present in RCM video sequences, resulting in the same pixel location representing neighbouring skin locations across time.

The motion is due to the volunteer's movement and due the metal ring and adhesive window losing its connectivity with time, resulting in video motion artifacts. The metal ring (and adhesive window) is placed on the skin surface and in turn connected to the microscope head to keep motion artifacts to an acceptable level (Section 4.2). An example of motion artefact across time is shown in Figure 4.15 (Section 4.5.4). Therefore, for the purpose of the study, it is important to correct for motion found in RCM video sequences.

7.2.2.2 Loss-of-coupling artefacts in RCM images

As discussed in Section 4.5.2, RCM imaging relies on optical coupling between the microscope and the skin surface. Poor coupling can result from wrinkles, or due to insufficient coupling oil, or due to a non-uniform distribution of ultrasound gel. Wrinkles appear in most RCM image datasets, and can be described as a crease in the skin, characterised by darker linear structures (Figure 4.7). Wrinkles are present across the epidermal, DEJ and dermal layers, and thus can be visualised across different scanning depths. As discussed in Section 3.2.2, Sourati et al. [94] generated an unsupervised wrinkle detection algorithm based on the Watershed algorithm. On the other hand, inadequate coupling fluid appears as dark regions whereas non-uniform distribution of ultrasound gel appears as large bright circular structures (Figure 4.11).

In areas occupied with loss of coupling artifacts one cannot visualise blood flow nor any other skin structure, thus limiting the area with possible capillaries. In addition to this, these areas have a large time varying intensity. This is expected to effect the algorithm's performance and ways to reduce these effects will be discussed in Section 7.3.2.

7.2.2.3 Laser power

The *in-vivo* RCM technology (VivaScope® 1500) enables videos to be captured with an automated laser power setting, or the power can be manually adjusted by the user. In the automated setting, the power is altered automatically so that the skin structures being imaged are captured with an intensity range of $0 - 255$. Therefore, an image captured using the automatic setting has an intensity range of $0 - 255$, whereas those captured by setting the power manually do not necessary cover the full range of possible intensities. Imaging is normally performed with the power set automatically since structures are well defined and the melanin content can be individually discretised. As illustrated in Section 4.5.5, the higher the power used to capture the image, the more blood flow can be visualised in the DEJ. However, as the power used to capture the sequence of images increases, the signal-to-noise ratio reduces together with some parts of

the image being saturated.

7.3 Algorithmic Method

7.3.1 Introduction

In this Section the algorithmic method developed to automatically detect skin capillaries by detecting the RBCs detected within them over a time sequence of RCM images is presented. The purpose of the developed algorithm is to generate a 2D map that highlights spatial locations (x, y) that are characterised by presence of RBCs in the time direction (t). The spatial locations of interest correspond to capillaries. In addition to detecting, the proposed algorithm also calculates the size of the capillaries, together with the velocity at which the RBCs are travelling through the capillary.

As previously discussed in Section 7.2.1, the capillaries are not visible in RCM video sequences, however, clinicians can visually outline their location and size by observing RBCs travelling through the capillaries. The proposed algorithm aims to emulate this, by searching for individual RBCs across a video sequence. As an introduction to the developed algorithm the steps involved are illustrated in Figure 7.3. Each step of the pipeline is explained in detail in following sections.

The input to the pipeline is a $\sim 30s$ RCM video sequence, which is captured from an a $0.5 \times 0.5mm^2$ *in-vivo* skin site using the VivaScope® 1500 RCM system. Initially, the video sequence is motion corrected using a motion stabilisation algorithm, the output of which is a video sequence of the same duration. Spatial areas not present throughout the duration of video sequence are masked out, resulting in the output video representing a spatial area smaller than $0.5 \times 0.5mm^2$. Motion correction is further described in Section 7.3.2.1. For the purpose of this explanation, in this Section the video sequence is referred to as a 3D volume where the depth axis represents the time domain.

Each individual frame of the motion corrected 3D volume is denoised using an anisotropic diffusion algorithm, the purpose of this step is to produce smoother images which will aid subsequent processing steps. The input to this is the motion corrected 3D volume, and the output is a 3D volume of the same size.

A windowing technique is used to emphasise the presence of RBCs and to lessen the visibility of skin structures (which do not vary in intensity with time). Using the windowing technique, which is discussed in more depth in Section 7.3.2.2, the standard deviation of the change in pixel intensity with time is calculated on the 3D volume generated by the denoising algorithm. Thus, the standard deviation volume at each time point incorporates knowledge from preceding and subsequent frames. The output of this step is a 3D volume, which will be referred to as the windowed standard deviation volume (L_{STD}).

Circles corresponding to the size of an individual RBC are searched for at each depth of the L_{STD} volume using the Hough circular transform (Section 7.3.2.3). The radii, centres and depth (time) of detected circles are recorded and the detected circles are mapped to a 2D map. The 2D map highlights potential capillary regions. This 2D map is inputted to a connect areas algorithm (Section 7.3.2.4) which groups neighbouring detected circles into one region, where each individual region is a potential capillary. The connect areas algorithm outputs a binary 2D map of potential capillary regions.

In order to discriminate between capillary and noise regions, a threshold is applied to the binary 2D map (Section 7.3.2.5). The threshold incorporates a minimum number of circles that need to be detected inside a region for it to be considered as a capillary region. The output of this step is a 2D map which identifies regions detected as capillaries by the proposed algorithm.

From the output map individual parameters of the detected capillaries are extracted, including capillary area and minor axis length. This is further discussed in Section 7.3.3.

An extension to this algorithm was produced to calculate the velocity of RBCs from capillaries imaged cross-sectionally using cross-correlation as the basis of the technique. Cross-sectionally imaged capillaries are defined as two circles of approximately equal size located at a distance of $15 - 25\mu m$. Such capillaries give a window into how long (in seconds) it takes for an RBC to travel from the arterial to the venous capillary limb. The distance travelled by the RBC is estimated based upon the protocol used to capture RCM videos. The technique is discussed in more detail in Section 7.3.4.

To summarise, a novel algorithm has been designed and implemented in this Chapter, which is capable of detecting and quantifying skin capillaries in RCM video sequences, together with calculating the velocity at which RBCs travel at inside the detected capillaries. The components of this algorithm are now presented in detail in the following sections.

7.3.2 Design of algorithm to detect skin capillary area

7.3.2.1 (a) Motion stabilisation and Denoising

A non-rigid registration [142] algorithm was used to calculate the amount of motion in the images and then the images were transformed to remove the detected motion.

For each individual video, all the frames were registered to a common reference which was selected to be the first frame of the video, as shown in Figure 7.4. This method was preferred to the method depicted in Figure 7.5, which has a moving reference that moves together with the source. The latter approach fails in cases where a poor registration takes place in one of the frames as the error propagates in subsequent frames.

The volunteer's movement causes different levels of pressure between skin surface and metal ring, resulting in non-rigid movement of the skin and artefact

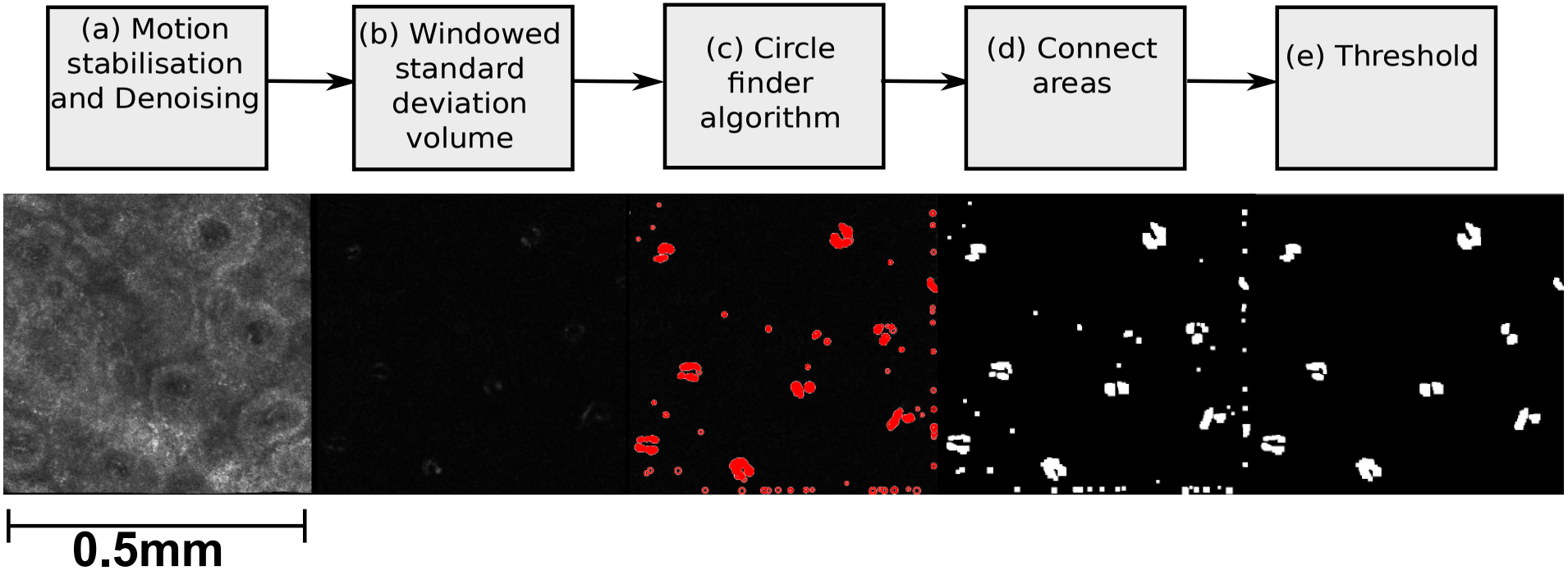


Figure 7.3: The algorithmic method pipeline designed and developed in this Chapter. (a) Initially the RCM video sequences were motion corrected using a non-rigid registration and each image was denoised using an anisotropic diffusion filtering algorithm. (b) A windowed standard deviation volume across the time domain was calculated to enhance the visibility of RBCs. (c) RBCs were detected using Hough circle detection, where the red individual dots represent the individual circles detected across all the video. (d) Neighbouring regions are connected together and (e) regions with less than 0.5 circles/s detected were removed.

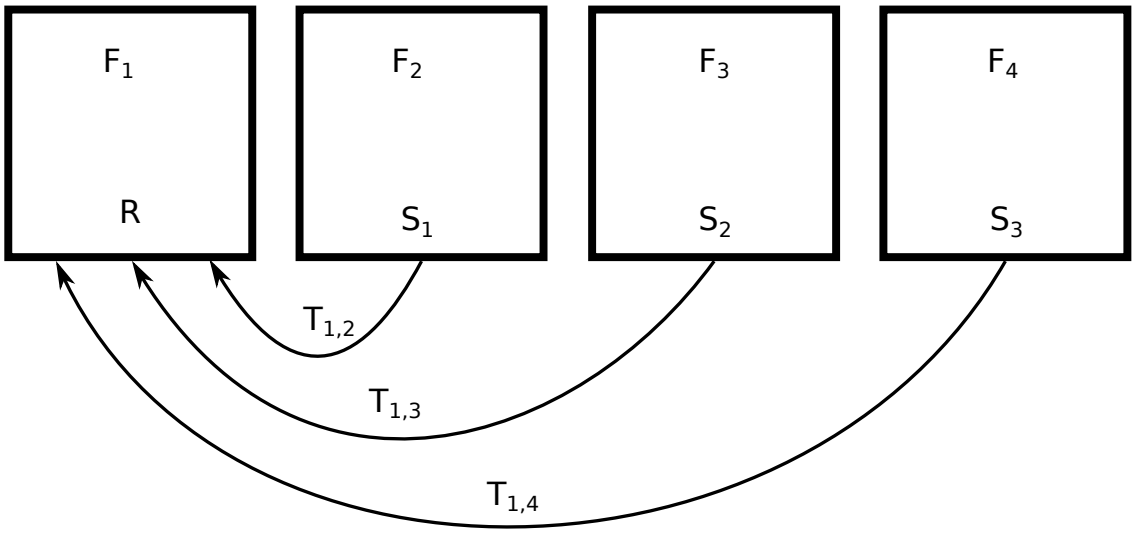


Figure 7.4: Registration with a static reference image: frames 1-4 ($F_1 - F_4$), where F_1 was set to the reference (R) image, and frames $F_2 - F_4$ were sequentially registered to frame F_1. Transforms $T_{1,2}$, $T_{1,3}$ and $T_{1,4}$ represent the transformation that needs to be applied to the source images S_1, S_2 and S_3 to obtain the motion corrected frames $F_2' - F_4'$.

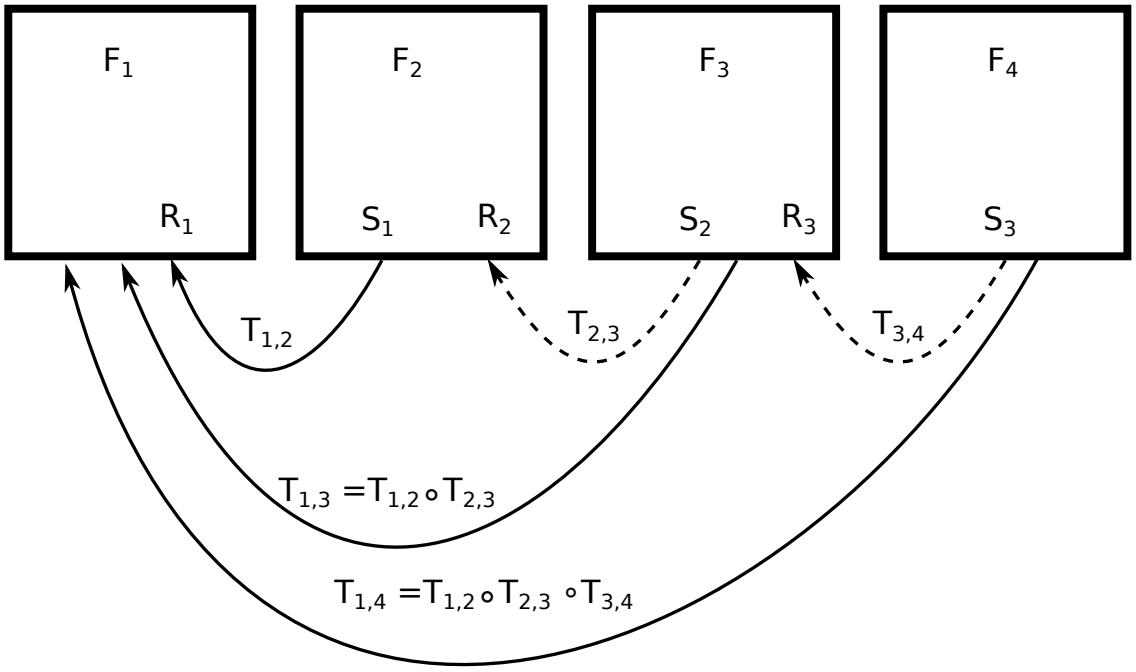


Figure 7.5: Registration with a moving reference: frames 1-4 ($F_1 - F_4$), where the reference (source) image was sequentially incremented to be $F_1 - F_3$ ($F_2 - F_4$). The dotted lines represent intermediate transforms, whereas solid lines represent the transformation between frames $F_2 - F_4$ and frame F_1. The transforms represented with the solid lines are equivalent to the transformations depicted with the solid lines in Figure 7.4.

structures. A non-rigid registration [142] algorithm was applied to all videos. The defined non-rigid registration algorithm makes use of free-form deformations that warp an image both globally and locally, to cater for smooth and less

smooth (localised) deformations [142]. Large-scale deformations require a mesh with coarse control points, whereas small-scale deformations need more refined spacing [142]. As a range of deformations are used, a multi resolution mesh representation is needed, where the mesh is progressively refined to capture a more localised deformation [142].

A three-level resolution non-rigid registration was applied: initially a coarse non-rigid registration, with control points having $200\mu m$ spacing, was applied; then $100\mu m$ pixels spacing; and the final level uses a $50\mu m$ spacing, shown in Figures 7.6A, B and C respectively. The initial spacing of $200\mu m$ captures coarse structural movement, whilst the final spacing of $50\mu m$ retrieves finer movement. The initial levels have a larger amount of blurring $5\mu m$ (as opposed to $0.5\mu m$ for the finest level) and use an image downsampled by a factor of 10, which aids in registering large-scale movement. As depicted in Figure 4.15, large-scale movement is normally associated with wrinkles as their movement is more pronounced than that of other structures. Normalised mutual information, which gives a measure of image alignment [143], was used as a similarity measure.

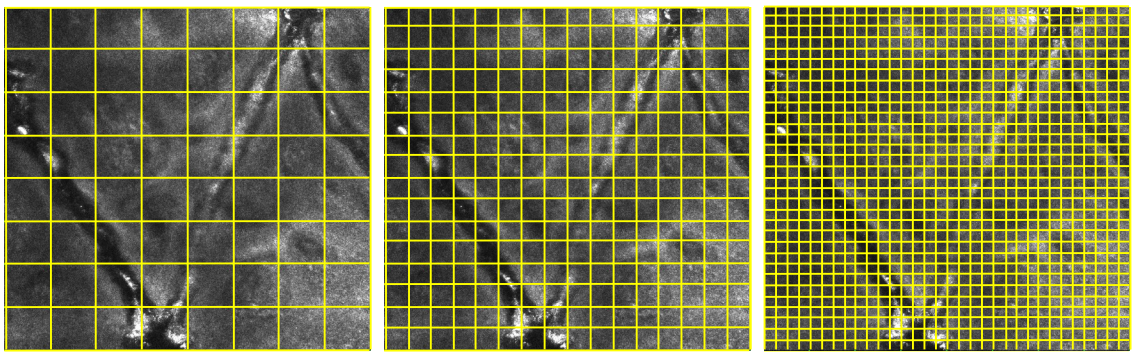


Figure 7.6: A multi-level resolution non-rigid registration makes use of control points having different spacings in order to capture coarse and fine movement. In this Chapter three spacings were used: (A) $200\mu m$, (B) $100\mu m$ and (C) $50\mu m$.

Due to motion some anatomical areas move out of the imaged field of view during the video. Using the non-rigid registration stabilisation algorithm described above, a mask M was generated for the common area visible throughout the duration of the video. As depicted in Figure 7.7, only the area visualised for the duration of the video will be considered as relevant to this study. The black

borders of images $F_1'' - F_4''$ show the area that was present in F_1 but has moved outside the field of view of the source image (i.e. $F_2..F_4$). The masked area defined by mask M, is calculated from the non-rigid registrations of all source frames to the reference frame. In addition, videos having a large motion which could not be corrected for using the non-rigid registration were discarded, whereas videos having a poor motion correction beyond a certain time point were truncated (in time). These steps ensured that the capillaries being extracted are present throughout the stabilised video.

Each frame of the motion stabilised video is denoised using an anisotropic diffusion algorithm defined in [144] [145] and implemented by Daniel Lopes in MATLAB. This algorithm aims to denoise an image without eliminating edges, and it gives a higher preference to high-contrast edges as opposed to low-contrast ones. This algorithm was chosen because RBCs appear as high-contrast edges with respect to the neighbouring skin structures, especially for the case of RBCs in capillaries imaged cross-sectionally (refer to Figures 7.1 and 7.2).

7.3.2.2 (b) Windowed standard deviation volume

As depicted in the cross-sectionally imaged capillary shown in frames $F_1 - F_{10}$ of Figure 7.8, as RBCs travel through a capillary they group together, resulting in consecutive frames containing RBCs within the capillary. Larger intensity variations are present within the capillary area (e.g. point A in Figure 7.8) than outside the capillary area (e.g. point B in Figure 7.8). The intensity variations across the duration of the video sequence ($30s$) of points A and B are shown in Figure 7.9. The larger intensity variations within the capillary are present as points within capillary area are imaging the presence and absence of RBCs across the video duration whereas points outside capillary area are imaging the same skin structure for the duration of the video sequence. To emphasize this distinctive characteristic and to highlight the presence of RBCs, the standard deviation of the time-varying pixel value was calculated.

A sliding window of width K_W was used to calculate L_{STD}, which will be referred to as the windowed standard deviation 3D volume and is calculated as

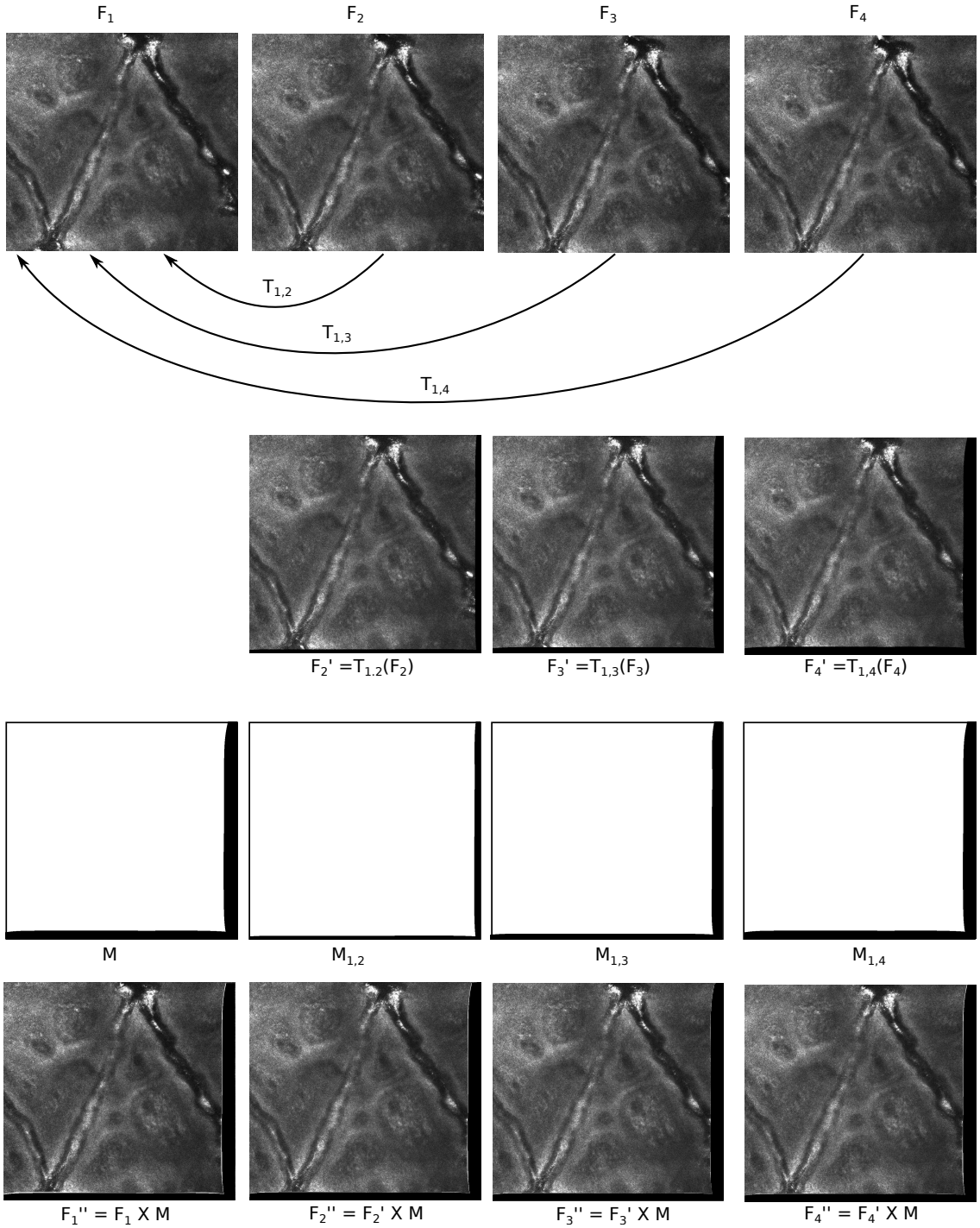


Figure 7.7: Methodology used in generating a common registration based mask for all frames in the video. Frames $F_1 - F_4$ represent four frames in a video, which are $\sim$ 10s apart. F_1 is the first frame of the video and is set to the reference R frame, to which all consecutive frames were registered to. Transforms $T_{1,2}$, $T_{1,3}$ and $T_{1,4}$ represent the non-rigid transforms between frames F_2, F_3, F_4 to the reference frame F_1. The black regions in masks $M_{1,2}$, $M_{1,3}$ and $M_{1,4}$ represent the areas in frames $F_2 - F_4$ that were not matched to areas in R frame. The common mask M is calculated by multiplying consecutive masks $M = M_{1,2} \times M_{1,3} \times M_{1,4}$. Mask M is then multiplied to all motion corrected frames ($F_2' - F_4'$) and to the reference frame (F_1), resulting in the available frame area which is common across all frames ($F_1'' - F_4''$) of the video. The capillary detection algorithm searches for capillaries in frames $F_1'' - F_4''$.

shown in Equation 7.1. In Equation 7.1, V represents the motion stabilised and denoised video, where (x, y) is the frame coordinate system, and t is the frame number. σ is the standard deviation of the set of values. The size of L_{STD} is the same size of the video and its aim is to highlight blood regions. It was noted that in general RBCs are typically present in less than 5 successive frames, and thus a width $K_W = 10$ frames was chosen. In the presence of RBCs a value of 10 frames is a sufficient width to cover both frames with and without RBCs, resulting in a high L_{STD} value that can be used to characterise the presence of RBCs.

$$L_{STD}(x, y, t) = \sigma \left\{ V(x, y, z + i) \right\}, i = -\frac{K_W}{2} + 1, ..., \frac{K_W}{2} \tag{7.1}$$

The example shown in Figure 7.8 shows one 2D slice of the L_{STD} volume calculated from the standard deviation of frames $F_1 - F_{10}$. As shown in the L_{STD} representation of Figure 7.8, the presence of blood in the capillary area (e.g. point A) is enhanced, whilst points outside the capillary area are diminished (e.g. point B) as their intensity does not vary significantly. Subsequent slices of the 3D L_{STD} volume are calculated from different sets of 10 frames (refer to Equation 7.1).

7.3.2.3 (c) Circle finder algorithm

Since RBCs are approximately spherical in shape, a Hough circular transform was used to detect circular RBCs moving through skin capillaries. Circles of radius 4 - 12 pixels, which correspond to a radius of 2 - 6μm were searched for. A minimum radius smaller than a RBC radius of 3μm was chosen as the RBC is not always captured cross-sectionally from the middle, and thus its apparent radius can appear smaller. A maximum radius corresponding to the average radius of a capillary limb is searched for. This value is used instead of the maximum radius of an RBC as the cross sectional area of the capillary can be covered with a number of RBCs travelling through the capillary simultaneously but which are not resolved as separate RBCs. The Hough circular transform implemented in MATLAB was used in this study. It must be noted that detecting RBCs as circles favours detection of blood flow in cross-sectionally imaged capillaries rather than capillaries imaged across their length.

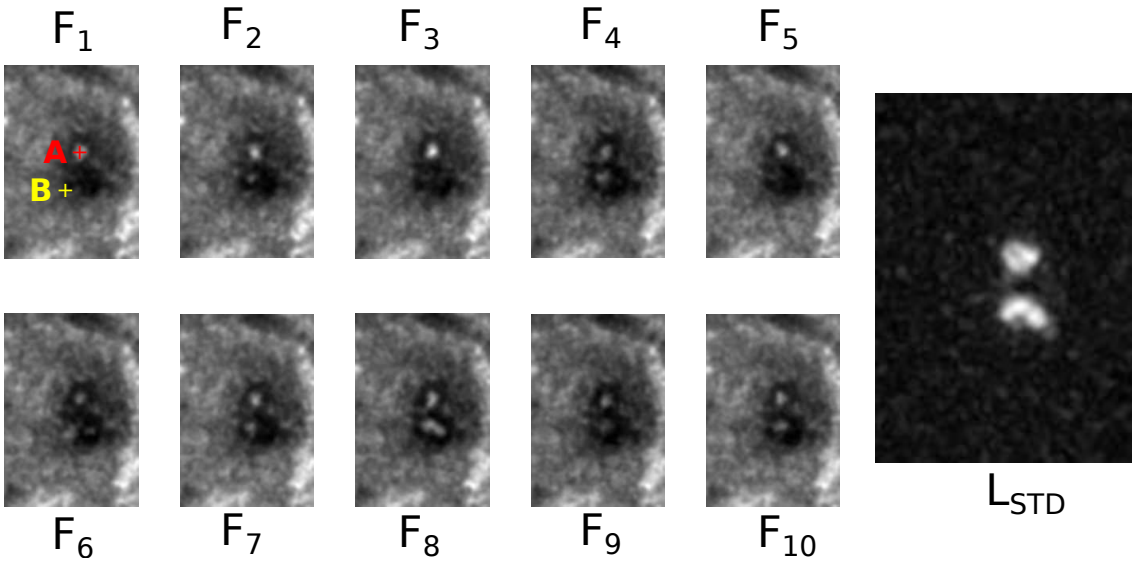


Figure 7.8: Calculating the standard deviation representation of frames $F_1 - F_{10}$ using a window width $K_W = 10$ frames. The frames are 10 consecutive frames of a cross-sectional capillary (within a papillae) captured at a site exposed with 11.3 SED. The presence of RBCs is highlighted in the L_{STD} image representation, whereas the bright papillae and tissue surrounding the capillary limbs are not visible in the L_{STD} as their intensity values do not fluctuate across time. Thus, the purpose of calculating L_{STD} is to enhance the identification of RBCs. L_{STD} is calculated from $F_1 - F_{10}$ and the standard deviation 3D volume is generated by iterating the frames as shown in Equation 7.1. Points A and B represent pixel coordinates inside and outside the capillary area respectively.

7.3.2.4 (d) Connect areas

In this section all the detected RBCs over the RCM video sequence are used to generate a 2D map highlighting which image regions are capillaries. The map's pixel values at each detected circle (RBC) location were selected to follow a Gaussian distribution centred at the circle's centre [146]. The input to this algorithm was the detected circles centres identified above, together with the individual radius of each detected circle. A 2-D Gaussian profile centred at C_x, C_y with standard deviation $\sigma = \frac{r}{3}$, where r represents the radius of the detected circle centred at C_x, C_y. The Gaussian profile is truncated at 3σ, which is equivalent to r. A map was generated from the sum of the Gaussian plots, where the intensity values of the map are proportional to the chance of probability a capillary positioned at that (x, y) location [146]. The purpose of generating Gaussian plots around each centre is to connect individual capillaries detected as several circles; circles within close proximity (equivalent the circles radii) were grouped together into one region. Individual regions which consist of pixels which have

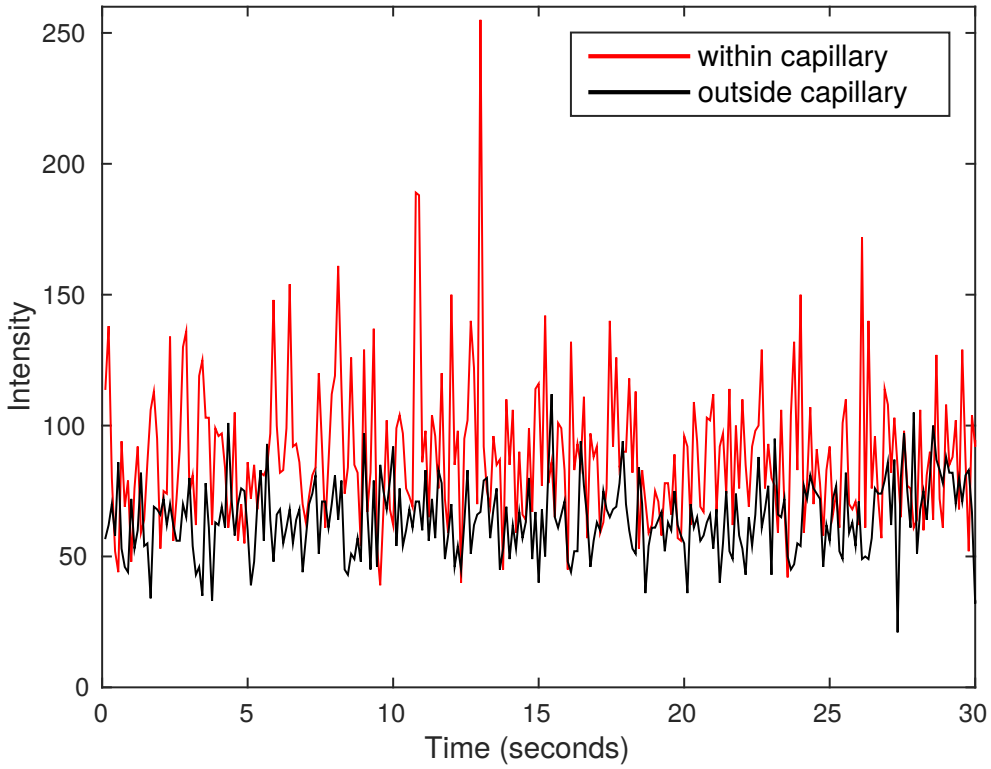


Figure 7.9: The intensity variations of points A and B (indicated in Figure 7.8) across the 30*s* video sequence. Points A and B represent pixel coordinates within and outside the capillary area respectively. Larger intensity variations are present inside the capillary area compared to areas outside of the capillary. This is due to points within the capillary are imaging the presence and absence of RBCs across the duration of the video sequence and some frames contain no RBCs, whereas points outside the capillary are imaging the same skin structure for the duration of the video sequence.

a positive value in the 2D map were distinguished and given an individual label. Each pixel in a region is an 8-pixel neighbour to another pixel in the region. Each connected region is a potential capillary region; it can either be an in-plane capillary, or a limb of a cross-sectional capillary, or a false positive capillary.

7.3.2.5 (e) Thresholding small regions

These connected regions were fed into the next and final step to generate a thresholded image. A threshold, defined as *region circle count*, corresponding to the number of circles that need to be detected within a region for it to be considered as a capillary was chosen from a visual inspection. The visibility of blood flow varies across different volunteers and across sites of the same volunteer exposed

to different UVR exposure doses. A possible explanation of this variability is that a range of capillary diameters, blood velocities and melanin amounts are found across different volunteers. The larger the capillary diameter and the blood velocity, the more visible the flow is, whereas large amounts of melanin (found in volunteers with skin types IV-VI) reduce the detectability of flow. Across different exposure sites of one volunteer, one can see more blood flow in high UVR exposed sites than in sites characterised by low UVR dose; due to larger capillary diameters and blood velocities associated with higher UVR doses. Due to these variations, the practical value of *region circle count* differed across volunteers and across different sites. The more flow that can be visualised and picked up by the algorithm, the larger *region circle count* can be set. For noisy data, a lower value of *region circle count* should be set as there is a poor differentiation between the blood signal and noise (false positive capillary regions). Although the threshold can be varied across different volunteers and sites, one threshold was chosen for all cases. A *region circle count* of 0.5 circles/s was used for all the experiment in this Chapter.

7.3.3 Calculation of individual capillary parameters

The output of the algorithm defined in Section 7.3.2 is a 2D map depicting regions detected as capillary areas. In the 2D map a number of capillaries are detected, each of which has its own size and shape. In this Section the methodology used to extract parameters from each detected capillary is outlined. Each individual capillary region in the 2D map is represented by a connected region of pixels. Initially each individual region is searched for and labelled using an 8-connected component algorithm [147]. The output of the connected component algorithm is another 2D map, however, here each capillary region is represented with an individual label as shown in Figure 7.10B. From each labelled region the capillary parameters, such as area and minor axis length, are extracted. The minor axis length is assumed to represent the capillary diameter.

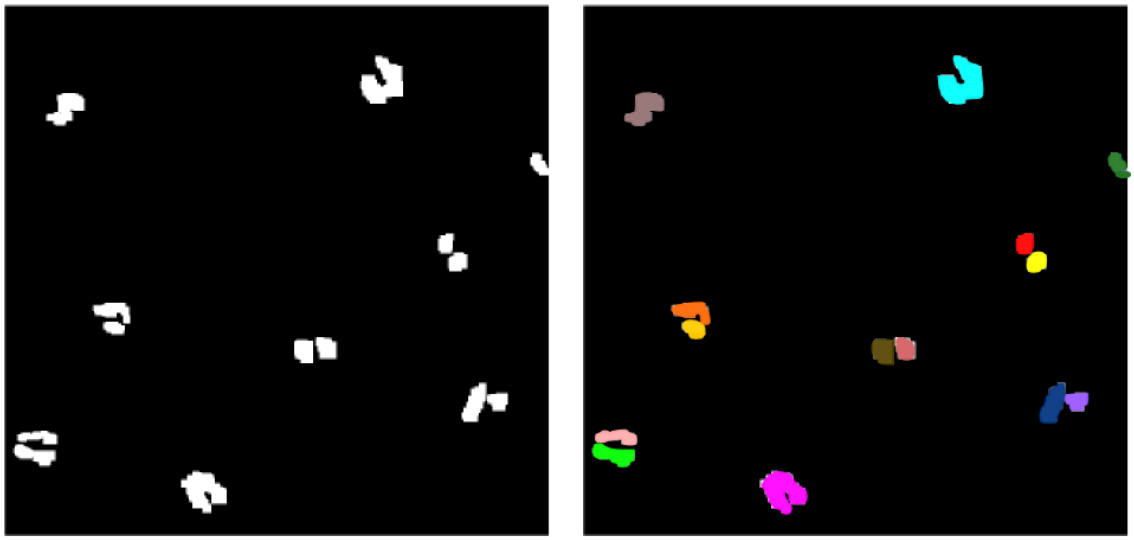


Figure 7.10: (A) The 2D map of regions detected as capillaries by the developed algorithm. (B) Each individual capillary region detected by an 8-connected algorithm and labelled. Each unique label is represented with a distinct colour in the output 2D map.

7.3.4 Design of algorithm to quantify RBC velocity across cross-sectionally imaged capillaries

In this section a methodology is defined to estimate the velocity of RBCs passing through capillaries imaged in cross-section. Here a single cross-sectional capillary is characterised as two separate adjacent circles, the centres of which are at a distance of $15 - 25\mu m$ apart, representing the separation of the arterial and venous limbs (refer to Section 2.2.4).

This study concentrates on measuring RBC velocity in capillary loops (capillaries imaged cross-sectionally). This is because these are the smallest capillaries in the human body and therefore demonstrate the unique capability of RCM of imaging single cells as they move through the loops. Additionally, the capillary loops in skin are the closest blood vessels to the skin surface and thus are likely to be the most sensitive to environmental insults such as UVR exposure to the skin and therefore the most physiologically interesting.

Figure 7.11 illustrates the geometry of a capillary with respect to the RCM image plane. This imaging plane is selected by first identifying the apex of the

capillary loop and then positioning the plan to lie at a distance D of $20\mu m$ below the apex, so that the plane intersects the capillary at two points, points A and B (shown in Figure 7.11). The nature of these capillary loops ensures that the distance between the arterial and venous limbs are between $15\mu m$ and $25\mu m$. Refer to Section 4.5.3 for protocol used to capture RCM video sequences of skin capillaries.

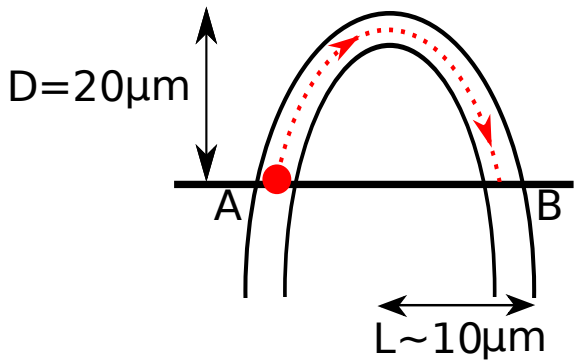


Figure 7.11: A diagram of a capillary loop where the dark line represents the imaging plane from which the video sequence is captured. As defined by the RCM protocol discussed in Section 4.5.3, cross-sectional capillaries were captured at distance $D = 20\mu m$ below the peak of the capillary (where the arterial and venous limbs merge). The red circle represents an RBC, which is initially visualised at the arterial limb (A), after which it will travel across the capillary to the venous limb (B). Distance $2L$ represents the separation of the arterial and venous limbs, where $L \simeq 10\mu m$.

The capillary loop is approximated as an ellipse with semi-major axis $20\mu m$ and the minor axis given by the separation of the arterial and venous limb, L. The distance travelled along the capillary loop by an RBC from A to B is then the ellipsoidal perimeter $\frac{P}{2}$ where;

$$P \approx 2\pi\sqrt{\frac{D^2 + L^2}{2}} \tag{7.2}$$

So for $L \simeq 10\mu m$, the distance travelled by an RBC between the arterial to the venous limb ($\frac{P}{2}$) is approximately $50\mu m$.

The devised algorithm calculates the time taken for a single RBC to progress from A to B as follows: the spatially average pixel intensity in the two regions of interest encompassing the capillary limbs in the image are plotted with respect

to time, as shown in Figure 7.12. The peak in intensity first seen in the arterial limb determines a time t_1, there is then a lag of Δt during transit; and a second peak is seen in the venous limb at t_2, where $\Delta t = t_2 - t_1$. The velocity of the RBC, v, is then simply:

$$v = \frac{P}{2\Delta t} \tag{7.3}$$

In practice Δt is better determined from a cross correlation of the time variation of mean pixel intensity which will be described here. To achieve this, the mean intensity values across limbs A and B are extracted, and indicated by signals $S_A(t)$ and $S_B(t)$ respectively. A typical example of signals $S_A(t)$ and $S_B(t)$ is illustrated in Figure 7.12. Whilst a single time sequence, $S_A(t)$ or $S_B(t)$, cannot be used to infer RBC velocity, the cross correlation of these signals provides velocity information by calculating the lag (Δt) for an RBC to travel from the arterial to venous limb. Depending on whether the cross correlation lag between $S_A(t)$ and $S_B(t)$ is negative or positive one can differentiate the arterial limb from the venous limb. If the lag is negative (positive) then $S_A(t)$ ($S_B(t)$) is representing the arterial limb, i.e. the limb where the RBC first appears. The following steps are used to calculate the cross correlation value and corresponding blood velocity:

- Subtract the mean and normalise by the standard deviation for each mean intensity signal:
  $S_A(t) = \frac{S_A(t) - \bar{S_A}}{\sigma_{S_A}}$
  $S_B(t) = \frac{S_B(t) - \bar{S_B}}{\sigma_{S_B}}$

- The signals $S_A(t)$ and $S_B(t)$ are effectively digitialised at 100ms due the RCM frame rate of 9fps. The signals were interpolated at 1ms intervals.

- Calculate the cross correlation between the arterial and venous signals of each capillary. The maximum cross correlation within the 0-0.4 s interval was considered to be the calculated lag (Δt); Δt represents the time lag for an RBC to progress around the capillary loop. The possible interval was truncated to 0.4 s to correspond to a minimum velocity of 0.12 mm/s [111].

- Calculate the velocity (v) by dividing the estimated distance by the Δt:
  $v = \frac{P}{2\Delta t}$

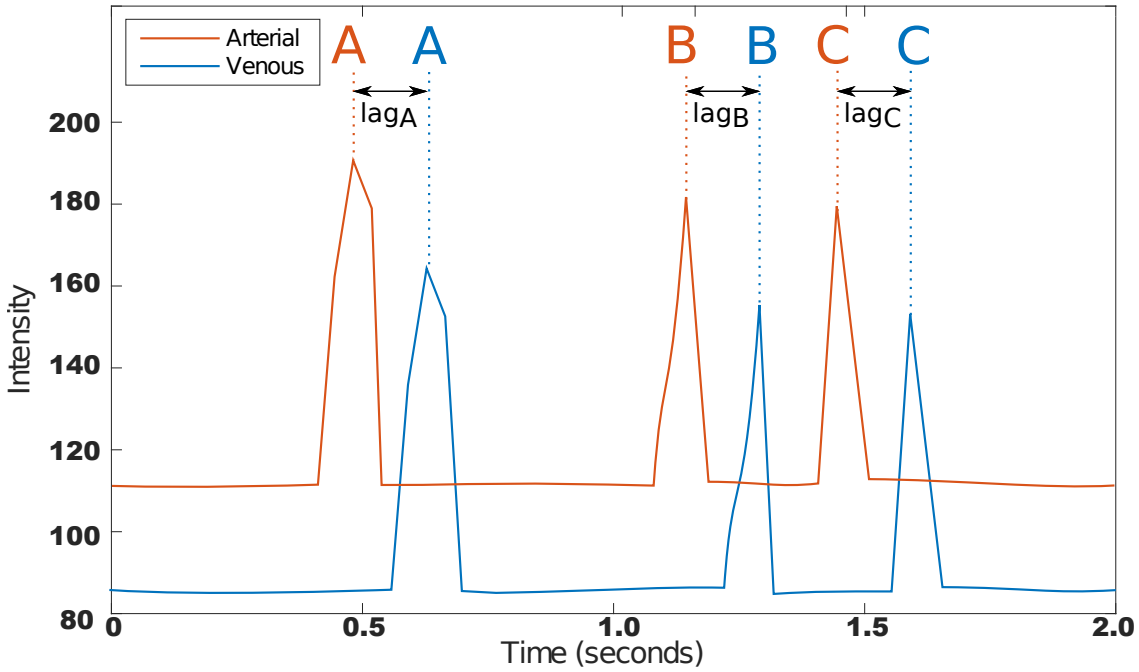


Figure 7.12: An example of the mean intensity signals of the arterial and venous limbs of a capillary imaged cross-sectionally. The arterial signal is leading; the venous signal is behind the arterial signal. Three corresponding peaks labelled A, B and C are shown, their corresponding lags are labelled as lag_A, lag_B and lag_C. The approximate mean lag value is $0.2s$. Each peak represents an RBC, for example, peak A in the arterial signal is an RBC which later appears as peak A in the venous signal.

Figure 7.13 illustrates each step of the above method. The signals shown in Figure 7.13A are the mean intensity signals extracted from the arterial and venous limbs of a cross-sectionally imaged capillary. Figure 7.13D represents the signals depicted in 7.13A after introducing the lag calculated using the steps identified above. When comparing Figures 7.13A to 7.13D one can note that the peaks of both signals are coincident.

7.4 Experimental Method

In this Section the experimental method, which covers the data collected, experiments devised and the validation technique used are described.

The algorithm defined in Section 7.3.2 is tested on RCM video sequences captured from anatomical sites exposed with incremental UVR doses, of both fair

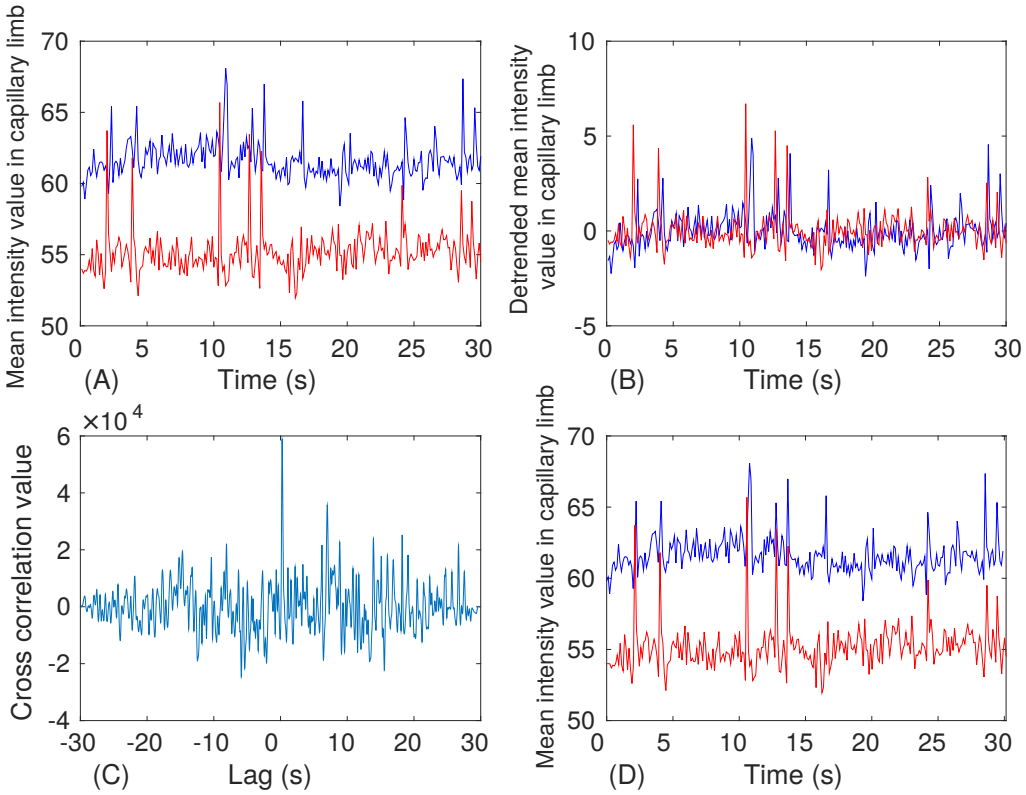


Figure 7.13: An example depicting the steps involved in calculating the cross-correlation lag between two signals extracted from the arterial and venous limb of a cross-sectionally imaged capillary. (A) the arterial (red) and venous (blue) mean intensity values across the 30s video duration. As illustrated the arterial peaks appear in time before the venous peaks. (B) The arterial and venous signals were detrended (the mean value was removed) and divided by the signal's standard deviation. (C) The cross correlation value response between both signals, a peak of value occurs at a lag of 0.242 s. (D) The red arterial signal was lag adjusted by 0.242 s and the corresponding peaks of the arterial and venous limbs occurs simultaneously in time.

and dark skinned volunteers. This dataset is chosen as an increase in capillary area, characterised as vasodilation, is expected with incremental UVR doses. Therefore, this data can be used to test whether the developed algorithm is capable of detecting increases in capillary area associated with incremental UVR dose. The data collected for this Chapter is described in Section 7.4.1.

The Experiments devised to test the algorithm are described in Sections 7.4.3-7.4.6. The total capillary area detected is compared to the UVR dose (Section 7.4.3) and a measure of the redness of the skin (Section 7.4.4). In addition, parameters that quantify the size of the capillary are extracted (Section 7.4.5). The Experiment to calculate RBC velocity within capillaries is identified in Section

7.4.6.

The total capillary area is validated against colour readings captured using a spectrophotometer, which give a measure of the level of reddening of the skin.

7.4.1 Data

7.4.1.1 UVR irradiation

As illustrated in Section 3.4 in Table 3.2, there are various ways of inducing skin erythema, including temperature changes, application of capsaicin based creams and ultra-violet radiation (UVR) exposure. UVR exposure is of particular clinical interest in photo-dermatology and the method used for the experiment in this Chapter.

Incremental doses of artificial UVR generated in the laboratory were applied to small areas of the ventral forearm. A monochromator, 300W-16S Xenon Arc Solar Simulator (Solar Light, Glenside, USA), was used to generate artificial light of wavelength 295-400nm, includes ultraviolet and visible light radiation. The doses used are in the range that can be obtained during a day when exposed to solar radiation. Volunteers were exposed using the guidelines indicated in London ethics reference number: 15/LO/0380 under the study: Molecular photoprotection of the skin by natural and synthetic sunscreens, enzymes and melanin and sub-study: Skin imaging and pain perception (MEL-1), Prof. Antony Young being the chief investigator. The UVR exposures are measured in standard erythema dose (SED) units; 1 SED is equivalent to $100J/m^2$. The individual sensitivity to the sunlight sources was assessed using a standard protocol in which the lowest dose that causes a just perceptible sunburn after UVR exposure. This dose is referred to as the minimal erythema dose (MED). Figures 7.14 and 7.15 illustrate the different UVR doses exposed to fair and dark volunteers. For the examples illustrated in these Figures, the volunteers' MED are 4SED and 30SED respectively. It is easier to detect the change in redness, and thus the MED, in fair skinned than in volunteers with dark skin.

In this experiment incremental UVR doses have been applied to the inner arm of 11 volunteers ranging from fair skin (skin types I-III) to darker skin (skin type VI), consecutive doses are incremented by a factor of $\sqrt{2}$, as shown in Table 7.1. The inner arm was selected as the best anatomical location to expose UVR irradiation as it is easily accessible and one of the areas that is least exposed to sunlight, and thus reddening occurs with lower UVR doses than other anatomical sites. The collected datasets are listed in Table 7.1 in the order they were captured. The tabulated MED refers to the lowest dose at which a change in reddening occurs visually. In the case of fair skin volunteers (volunteers # 1-7) a dose range of between 2 and 11.3 SED was used. For darker skin type volunteers (volunteers # 8-11), a larger range of UVR dose (15 to 120 SED) was used to cause comparable levels of erythema as the MED for darker skin tones are considerably larger than in fair skin volunteers.

A comparison between the MED values in fair and dark skinned volunteers is shown in Figure 7.16. Deciding upon which dose represents the volunteer's MED is not a straight forward task, it is difficult to detect changes in redness by eye, especially in volunteers with darker skin tones.

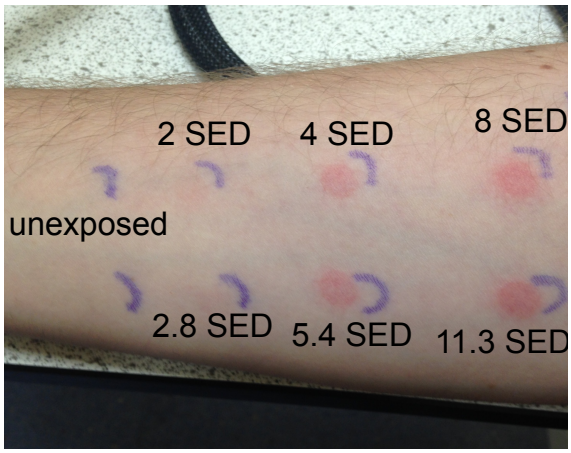


Figure 7.14: Fair skinned volunteer #5 16-24 hours after UVR exposure with a MED of 4 SED.

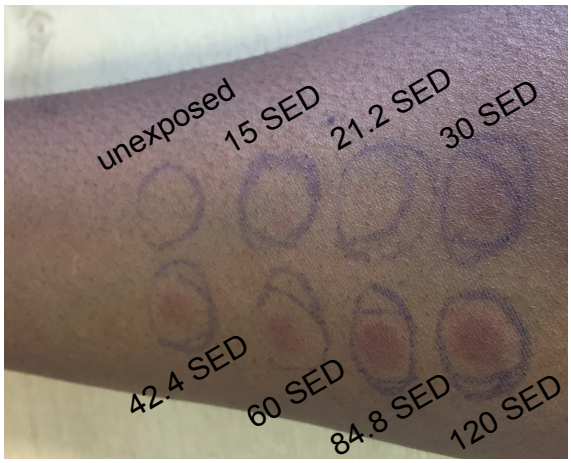


Figure 7.15: Dark skinned volunteer #9 16-24 hours after UVR exposure with a MED of 30 SED.

Volunteer	Skin Type	Sex	Doses (SED)	MED
# 1	III	Male	[0, 2, 2.8, 4, 5.6, 8, 11.3]	4
# 2	II	Female	[0, 2, 2.8, 4, 5.6, 8, 11.3]	4
# 3	III	Female	[0, 2, 2.8, 4, 5.6, 8, 11.3]	2.8
# 4	II	Male	[0, 2, 2.8, 4, 5.6, 8, 11.3]	4
# 5	II	Male	[0, 2, 2.8, 4, 5.6, 8, 11.3]	4
# 6	II	Male	[0, 2, 2.8, 4, 5.6, 8, 11.3]	2.8
# 7	II	Female	[0, 2, 2.8, 4, 5.6, 8, 11.3]	4
# 8	VI	Male	[0, 15, 21.2, 30, 42.4, 60, 84.8, 120]	42.4
# 9	VI	Male	[0, 15, 21.2, 30, 42.4, 60, 84.8, 120]	30
# 10	VI	Female	[0, 15, 21.2, 30, 42.4, 60, 84.8, 120]	30
# 11	VI	Female	[0, 15, 21.2, 30, 42.4, 60, 84.8, 120]	30

Table 7.1: Volunteers irradiated with UVR dose.

7.4.1.2 Collecting confocal video sequences

16-24 hours after UVR irradiation video sequences are captured with an RCM (VivaScope 1500®, Lucid Inc. Henrietta, New York) using the protocol identified in Section 4.5.3. Due to the high variability of capillaries present in the available RCM window area (refer to Section 4.5.3), around 8 video sequences were collected from each exposed and unexposed site. The captured videos were kept to a maximum duration of 30 s to keep motion to an acceptable level. Each video covers an area of $0.5 \times 0.5mm^2$ and it is selected from a maximum area of $8 \times 8mm^2$. Although the area that can be scanned from is of $8 \times 8mm^2$, videos are normally captured from the middle $4 \times 4mm^2$ as the middle area is has a larger response to the exposed dose, as illustrated in Figure 7.17. An effort was made to ensure that the selected $0.5 \times 0.5mm^2$ locations have capillaries and are located

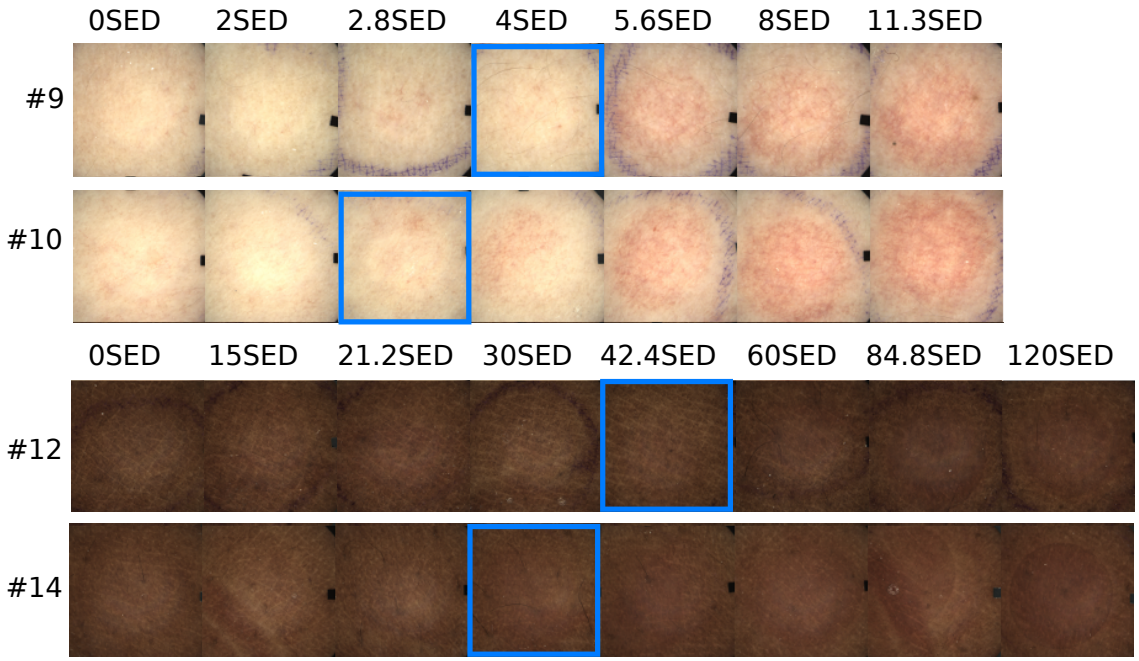


Figure 7.16: Dermoscopic images of UVR irradiated sites on 4 volunteers captured 16-24 hours after irradiation. Rows 1-2 are of a skin type II (volunteers # 9 and # 10) and rows 3-4 represent a skin type VI (volunteers # 12 and # 14). Each image in a row represents an incremental irradiation dose. Volunteers of skin type II were irradiated with doses [2, 2.8, 4, 5.6, 8, 11.3] SED and type VI were irradiated with [15, 21.2, 30, 42.4, 60, 84.8, 120] SED. The first column represents non-exposed control sites. The blue boxes indicate the visually assessed MEDs; 4, 2.8, 42.4, 30 SED respectively.

within the available exposed area (i.e. are not situated in the bordering regions that do not represent the exposed dose). In addition, efforts were made to keep artefacts to a minimum and attempts were made to capture videos in regions without wrinkles.

When collecting video sequences from the different exposed and control sites, it was noted that for sites with no or low UVR exposure, a larger percentage of the $8 \times 8mm^2$ area needed to be examined in order to find eight $0.5 \times 0.5mm^2$ areas exhibiting skin capillaries, than in sites exposed with higher UVR doses. This is depicted in Figures 7.17 and 7.18. In these plots, the purple shaded area illustrates the area which was searched to locate an RCM image showing capillaries. In general a larger area needed to be searched at lower doses.

Taking the considerations noted in Section 7.2.2.3 into account, it was decided

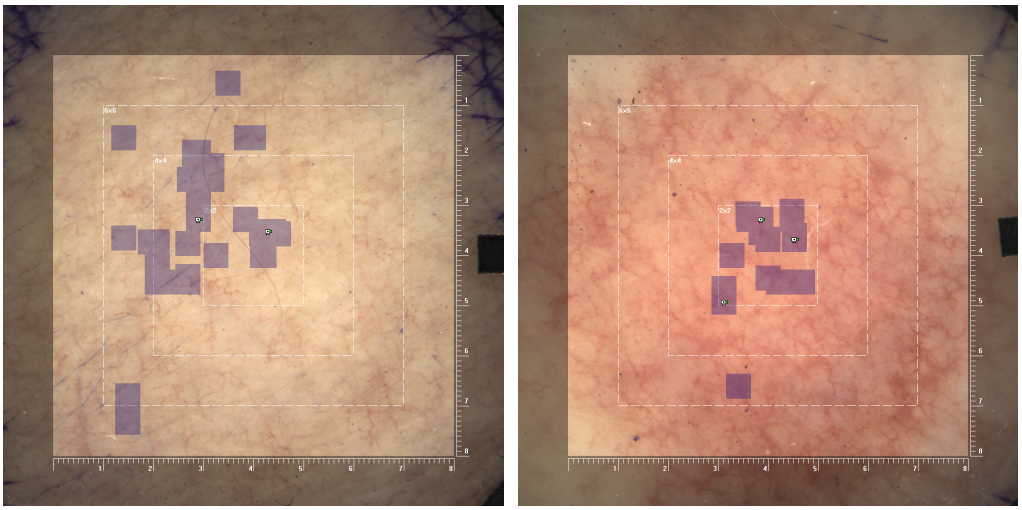


Figure 7.17: Dermoscopy images of UVR exposed sites showing regions at which RCM video sequences were obtained. The video maps illustrate 2 and 3 sites where videos were captured for the 2 SED and 8 SED sites for volunteer # 4. These maps were captured in the process of data acquisition, at the point of having 2 and 3 (out of the total 8) captured videos. The scale at the left hand side and at the bottom is a ruler in millimetres. Square regions indicate the boundaries of $2 \times 2mm^2$, $4 \times 4mm^2$ and $6 \times 6mm^2$ regions centred at the centre of the image. The purple shaded regions indicate where the user has searched for skin capillaries, whilst the video icon indicates that a video covering an area of $0.5 \times 0.5mm^2$ was captured around that centre. The main distinction between the two maps is that a larger area needs to be searched for in lower doses to find regions exhibiting skin capillaries. As illustrated in the 8 SED map, a higher level of reddening is present inside the $6 \times 6mm^2$ boundary, and thus, videos are captured inside the region exhibiting the highest response.

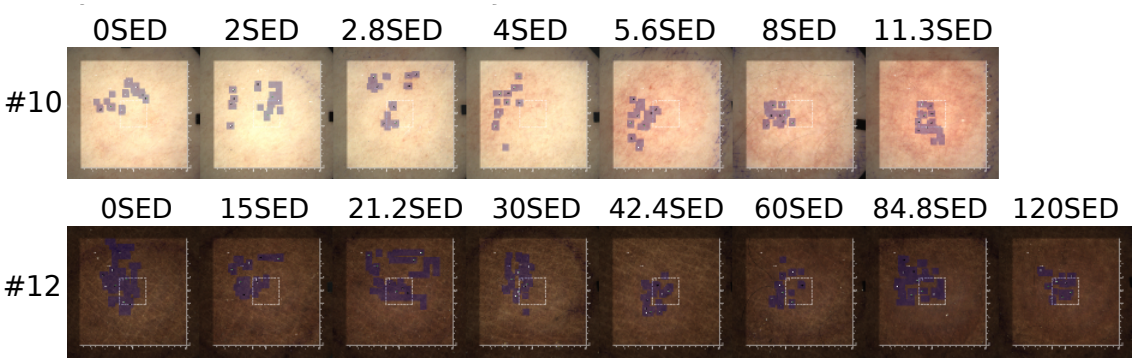


Figure 7.18: Dermoscopy images of UVR exposed sites showing regions at which RCM video sequences were obtained for a fair and dark skinned volunteer. Video maps of volunteers # 10 and # 12 at each exposed site and at the unexposed control site. The volunteers have an MED of 2.8 SED and 42.4 SED respectively. The purple shaded regions indicate where the user has searched for skin capillaries. The shaded area is considerably larger for doses below the MED, especially for volunteer # 12 (skin type VI).

that for fair skin types (types I-III) the automated power setting was used, as this replicates the standard clinical procedure, and also because for these skin tones

the automated power is sufficient to capture the flow of RBCs. However, for darker skin tones (skin types IV-VI) it was decided to manually alter the power value in the range of $2 - 2.5mW$ as here the chosen automated power is not sufficient to capture blood movement. The reason being that the RCM chooses a relatively lower level of power to image the highly refractive melanin content in darker skin tones. By manipulating the chosen power to a higher value, the RBCs in capillaries can be more clearly observed, at the expense of reduced melanin visualisation. However, a power value greater than $2.5mW$ generally results in poor quality (saturated) images.

7.4.2 Validation

Colour readings were calculated using the Minolta Chroma Meter CM-700d spectrophotometer at each site (Section 2.4.2). As discussed in Section 2.4.2, the spectrophotometer gives an a^* measure that represents the redness measure. Spectrophotometry is commonly used to assess skin reddening in photobiology [108]. The aim of using it in this study is to examine the relationship between the a^* value and the RCM detected capillary area. The relationship between a^* and total capillary area C_{Area}, which is calculated by summing the individual capillary areas of the N capillaries present in the video, is shown in Equation 7.4.

$$a^* \propto \sum_{i=1}^{N} C_{Area_i} \tag{7.4}$$

7.4.3 Total capillary area vs UVR irradiated dose

The aim of the Experiment detailed in this Section is to check whether the capillary area is correlated to the UVR exposure, i.e. whether more capillaries are detected by the algorithm at higher UVR doses. For this Experiment, data was acquired from 11 volunteers (volunteers #1 - # 11 in Table 7.1). Seven volunteers had a skin types II-III (referred to fair skinned volunteers) and 4 volunteers had a skin type VI (referred to dark skinned volunteers). The forearm of each individual was irradiated with incremental doses, in the range of 0-11.3 SED and 0-120SED for fair and dark skinned volunteers respectively. Eight RCM video sequences were captured from each exposed site and from an unexposed site with

the VivaScope®1500 system after 16-24 hours of exposure as outlined in Section 7.4.1.

Each RCM video sequence was used as input into the algorithm described in Section 7.3.2. The output from this algorithm for each video sequence is a 2D map identifying regions detected as capillaries. The total area of regions detected as capillaries per video will be referred to as total capillary area from now on. The total capillary area is compared to the UVR exposure, where the exposure dose is analysed in terms of SED units and in factors of the volunteer's personal MED.

When analysing the total capillary area with respect to the SED exposure dose, the results obtained are divided into fair and dark skinned volunteers. The reason for grouping the results into these two subgroups is because of the different UVR exposure doses used for fair and dark skinned volunteers. During data acquisition it was noted that it was difficult to find 8 sites which clearly showed capillaries for doses less than 30 SED in volunteers of skin type IV-VI. Therefore, some of the captured videos contain no visible capillaries. Due to this, the total capillary area detected for these doses is not representative of capillary area and so these datasets were removed and were not analysed any further.

The next step is to analyse the total capillary area with respect to MED. As opposed to SED, the MED enables all skin types to be analysed together. For each volunteer, the set of SED values are divided by the volunteer's personal MED. These values are then grouped into three groups: below MED, $1-2\times$ MED, and above $2\times$ MED. The total capillary area is compared against these groups. This extension was performed as the MED enables comparison between volunteers having the same skin type and across volunteers of different skin types.

7.4.4 Total capillary area vs a^* value

The aim of the Experiment detailed in this Section is to check whether the total capillary area is correlated to a^*, which quantifies the redness of the skin as

measured using a spectrophotometer. Thus, the purpose of this Experiment is to check whether the relationship identified in Equation 7.4 holds. The data used for this Experiment is the same as that used for the Experiment outlined in Section 7.4.3. However, in addition to the RCM video sequences, spectrophotometer readings were acquired from each exposed and unexposed site. Three readings from each site were acquired using the Minolta Chroma Meter CM-700d.

The total capillary area output is compared to the spectrophotometer validation data. Due to the variability across different volunteers the total capillary area was compared to $\bar{a}^*$. The $\bar{a}^*$ is calculated by normalising each a^* value by the a^*_{MED}, which corresponds to the a^* at the volunteer's MED site. As the range of a^* is greater for fair than dark skinned volunteers, the Results for this Chapter are subdivided into fair and dark skinned volunteers. For the case of fair skinned volunteers $\bar{a}^*$ was analysed in 0.5 increments, whereas for dark skinned volunteers $\bar{a}^*$ was analysed in 0.1 increments.

7.4.5 Calculating parameters from individual capillaries

The aim of the Experiment detailed in this Section is to check whether the individual capillary size increases as UVR irradiation dose increases, i.e. are individual capillaries dilating (increasing in size) as the irradiation dose increases. The data used for this Experiment is the same as that used for Experiment 7.4.3, and similarly individual capillary parameters are compared to SED and MED values. Note: each RCM video sequence imaged different capillaries, therefore this Experiment does not aim to track an individual capillary across different exposure doses but to quantify the median capillary size at each dose for fair and dark skinned volunteers.

Each video sequence was used as input to the algorithm identified in Section 7.3.2. The output of each video sequence is a 2D map which highlights different regions which were detected as capillary area. Using the algorithm identified in Section 7.3.3 the area and minor axis length of each capillary region is measured.

When analysing these individual parameters with respect to the SED exposure dose, the results obtained are divided into fair and dark skinned volunteers. Similarly to the Experiment identified in Section 7.4.3, doses less than 30 SED are not plotted for dark skinned volunteers.

The next step is to analyse the measured individual parameters with respect to MED. For each volunteer, the set of SED values are divided by the volunteer's personal MED. By visually inspecting the RCM videos, it was noted that in general the capillaries observed for fair skinned volunteers at high irradiation doses are generally larger than those observed for dark skinned volunteers; i.e. the capillaries present in RCM video sequences of fair skinned volunteers at 11.3 SED are larger than the capillaries present in RCM video sequences of dark skinned volunteers at 120 SED. Due to this, it was decided to split these results into fair and dark skinned subgroups.

7.4.6 RBCs velocity in cross-sectionally imaged capillaries

The aim of the Experiment detailed in this Section is to measure the velocity at which the RBCs are travelling at inside the capillaries and to quantify the change in RBC velocity attributed to an increase in UVR irradiation dose. As a much larger number of cross-sectional capillaries were detected than in videos captured from dark skinned volunteers, this study has just used data from fair skinned volunteers.

The data used for this Experiment is the data of the fair skinned volunteers (volunteers #$1-7$) outlined in the Experiment identified in Section 7.4.3. As in previous Experiments, each RCM video sequence is used as an input for the algorithm identified in Section 7.3.2. The output of this algorithm is a 2D map, in which imaged-capillaries are searched for. As outlined in Section 7.3.4, cross-sectionally imaged capillaries are characterised as two adjacent circles of similar size in the 2D map. An algorithm was developed to search for two circles, one of the circles needs to be at least 80% the size of the other one, and located within

$15 - 25\mu m$ apart. The output of this algorithm is another 2D map but this time only containing cross-sectionally imaged capillaries.

From this subset of detected capillaries, the arterial and venous signals (the mean intensity of each circle across time) is extracted. The velocity is calculated from these two signals using the cross-correlation methodology outlined in Section 7.3.4.

7.5 Results

This section presents the results achieved with the developed algorithm that automatically detects skin capillaries in RCM video sequences. The results are split into four Sections. The below list summarises specific questions which are addressed by each set of results. The following aspects will be covered:

1. Whether the algorithm is capable of automatically detecting skin capillaries in RCM videos (Section 7.5.1)
2. Whether the algorithm can be used to measure different levels of erythema based on the total detected capillary area (Section 7.5.1)
3. Whether the total detected capillary area is correlated with the a^* value captured by the spectrophotometer (Section 7.5.2)
4. Whether the algorithm can quantify individual skin capillaries, by extracting parameters such as area and minor axis length (Section 7.5.3)
5. Whether the algorithm can differentiate between the arterial and venous loops of the capillary (Section 7.5.4)
6. Whether the algorithm is capable of estimating the RBCs velocity in cross-sectionally imaged capillaries (Section 7.5.4)

7.5.1 Total capillary area vs UVR irradiated dose

This Section presents results related to calculation of the total area detected for each exposed dose in fair (skin types I-III) and darker skin (skin type VI). Figures

7.19 and 7.20 provide a visual example of the results achieved by the developed automated algorithm for fair and dark skin tones respectively. The highlighted regions are the regions detected as capillaries by the developed algorithm. The main difference between the results achieved for fair and dark skin are the number and size of capillaries detected. The overall capillary area detected increased from 0.25% to 1.20% in fair skinned volunteers as the exposure increased from 0 to 11.3 SED (Figure 7.21), whereas for dark skinned volunteers the area increased from 0.12% to 0.50% as the exposure increased from 0 to 120 SED (Figure 7.23). An example of this instance is depicted in Figures 7.19 and 7.20; the number of detected capillaries in Figure 7.19G (11.3 SED in fair skin) are nearly double to those detected in Figure 7.20H (120 SED in dark skin). In addition, the size of the detected capillaries are larger in Figure 7.19G.

As discussed in Section 7.4.1, different irradiation dose ranges were used to irradiate fair and dark skinned volunteers. The motive in doing so is to achieve comparable levels of vasodilation. However, as illustrated in Figures 7.19 and 7.20 the number of visible capillaries and size of capillaries is much smaller in the latter case. This supports the argument depicted in Section 7.4.3 of dividing the results into these two subgroups (fair and dark).

In the second part of this section, the total area detected as skin capillaries using the proposed algorithm is given as a percentage of the field of view (of $0.5 \times 0.5mm^2$). The area is given for each video sequence extracted at each site for each volunteer. The percentage of detected skin capillary area is analysed versus the dose in terms of SED and in factors of the personal volunteer's MED. As different doses were used for fair and dark skin, the MED provides a method to compare the results achieved for fair and dark skin.

Figure 7.21 shows the total area detected as capillary area from the fair skinned volunteers (volunteers # 1-7) against UVR exposure. The multiple values (blue dots) plotted at each dose show the area calculated from each captured RCM video at each exposure site. For each volunteer a positive correlation between detected area and dose is observed. Spearman's rank correlation was calculated

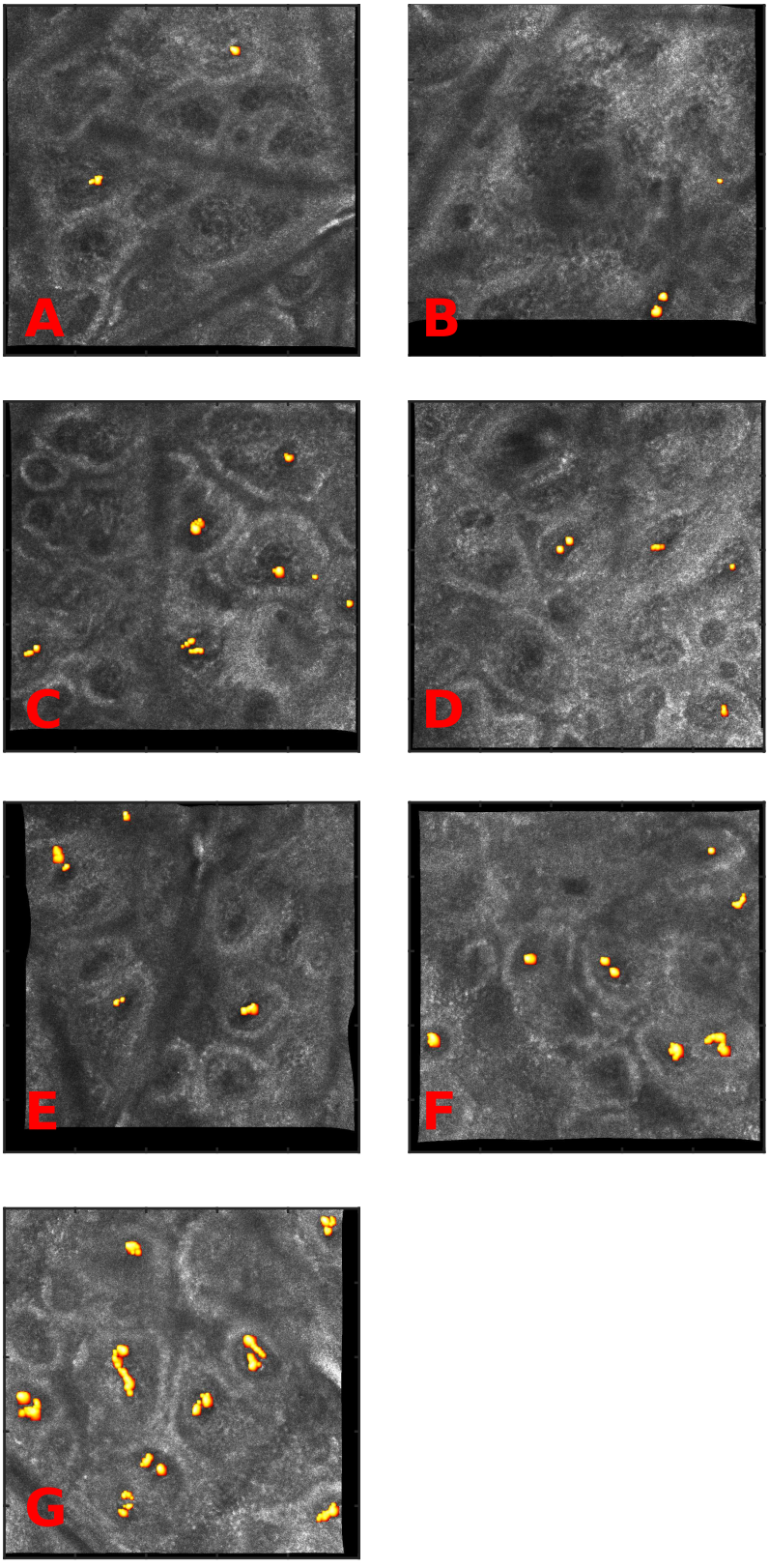


Figure 7.19: A representation of the detected capillary areas overlaid on RCM images for a fair skinned volunteer (volunteer # 5) at a range of UVR exposures: (A) 0 SED (B) 2 SED (C) 2.8 SED (D) 4 SED (E) 5.4 SED (F) 8 SED (G) 11.3 SED. Each image represents the first frame of a video and the corresponding detected capillary areas for the particular video. The dark regions are the areas masked out by the non-rigid registration. The MED for volunteer # 5 is 4 SED.

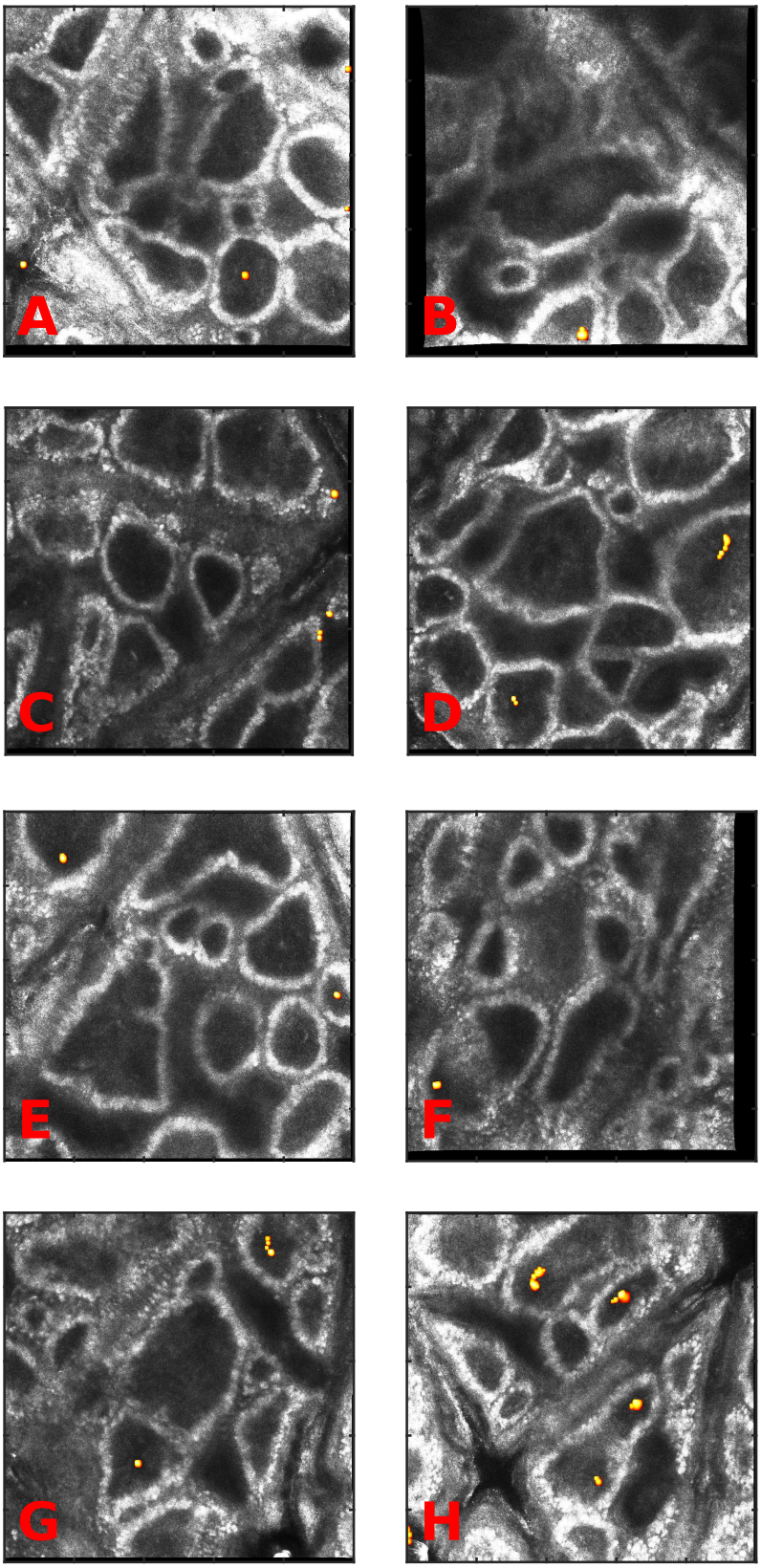


Figure 7.20: A representation of the detected capillary areas overlaid on RCM images for a dark skinned volunteer (volunteer # 8) at a range of UVR exposures: (A) 0 SED (B) 15 SED (C) 21.2 SED (D) 30 SED (E) 42.4 SED (F) 60 SED (G) 84.8 SED (H) 120 SED. Each image represents the first frame of a video and the corresponding detected capillary areas for the particular video. The dark regions are the areas masked out by the non-rigid registration. The MED for volunteer # 8 is 42.4 SED.

for each volunteer resulting in values between 0.48 and 0.74.

The box plot in Figure 7.21 shows the total capillary area results for all the fair skinned volunteers combined. A box plot is shown for each exposed dose level. As the exposure is increased from 0 to 11.3 SED the median percentage capillary area detected increases from 0.25% to 1.20%. This increase is statistically significant; the Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$). These differences can be seen in Figure 7.19; as the exposure dose increases there is an increase in number and size of capillaries.

Figure 7.22 shows these results after grouping based on the volunteer's personal MED. Results are collated into three groups: (i) below a volunteer's MED; (ii) above MED, but below 2× MED; (iii) and greater than 2× MED. The median percentage of total capillary area is increased from 0.25% to 0.48% to 1.15% for these respective groups. These increments are statistically significant; the Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$) for both increments.

The above calculations were repeated for volunteers having a darker skin tone (volunteers # 8 - 11). Figure 7.23 shows the total area detected as capillary from the dark skinned volunteers (volunteers #8-11) plotted against UVR exposure. The Spearman's rank correlation was calculated for each volunteer and the values achieved ranged from 0.37 to 0.64.

The box plot in Figure 7.23 show the capillary area results for all the dark skinned volunteers combined. As the exposure is increased from 30 to 120 SED, the median percentage capillary area detected increased from 0.12% to 0.50%. This increase is statistically significant; the Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$). These differences can be visually seen in Figure 7.20, when comparing sub-figures D and H, one can visually notice a larger number of capillaries detected and the capillaries have a larger diameter in H than in D. However, in dark skinned volunteers the change in detected area with incremental exposure is not as pronounced as in fair skinned volunteers.

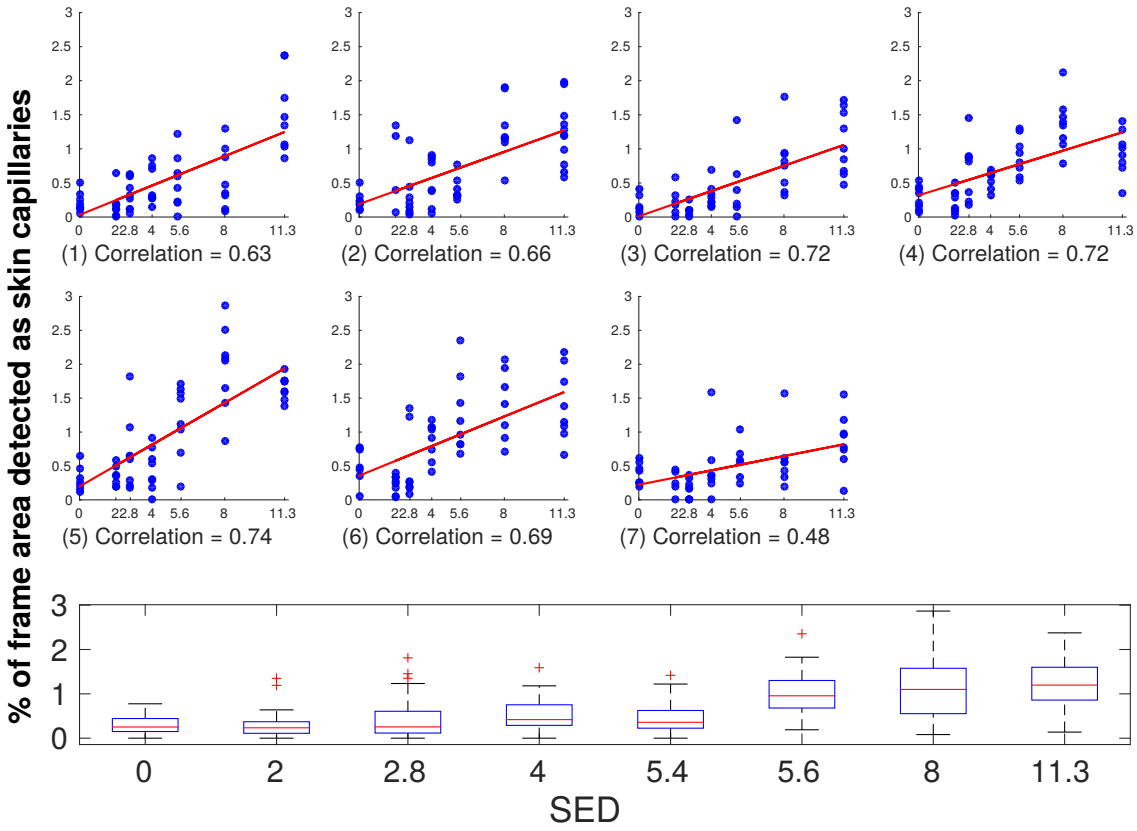


Figure 7.21: Fair skin: Total capillary area as a percentage (%) of the video's $0.5 \times 0.5mm^2$ field of view (FoV). Plots (1) to (7) show the results achieved for volunteer # 1 to # 7 respectively. The lower box plot combines all the volunteer results. In plots (1)-(7) each individual point represents the total capillary area obtained from a 30 s video sequence captured at a frame rate of 9 fps.
There is a large presence of wrinkles in the video sequences captured from volunteer # 7, which corresponds to the volunteer having the lowest correlation value (0.48) from fair skinned volunteers.
Lower box plot: The red line represents the median, and the blue lines represent the 25th and 75th percentiles of total capillary area (as a percentage of the video's FoV). The plus signs indicate outliers and are values larger or smaller than $\pm 2.7\sigma$. These boundaries are indicated by the upper and lower whiskers; beyond which values are considered to be outliers.

Figure 7.24 shows these results based on the volunteer's personal MED. Results were collated into three groups: below a volunteer's MED; above MED, but below $2\times$ MED; and greater than $2\times$ MED. The percentage of detected total capillary area decreased from 0.13% to 0.11% followed by an increase to 0.37% for the three respective groups. The change in detected area between the $1-2\times$ MED and above $2\times$ MED groups was found to be statistically significant; the Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$).

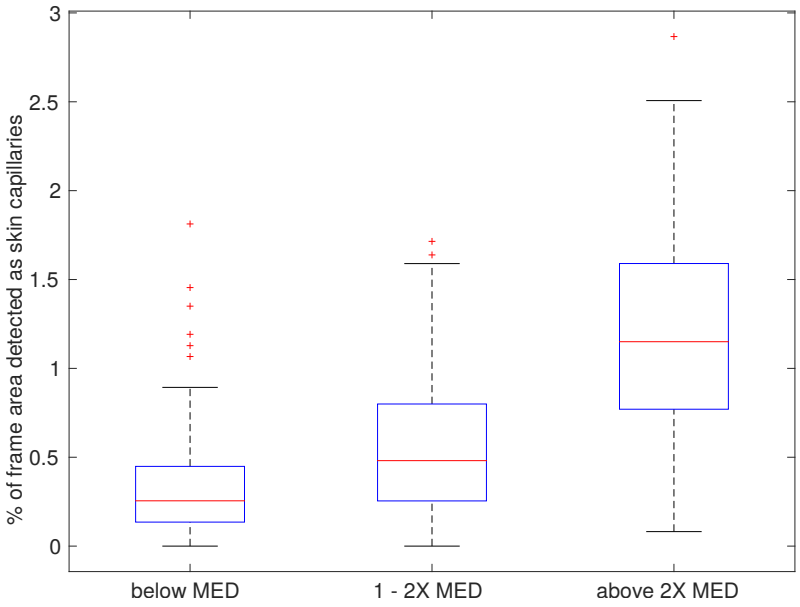


Figure 7.22: The variation in the percentage of total capillary area per field of view of the video ($0.5 \times 0.5mm^2$) with MED is shown. A differentiation is done between doses below MED; doses representing 1 to 2 $\times$ MED and doses greater than 2$\times$ MED.

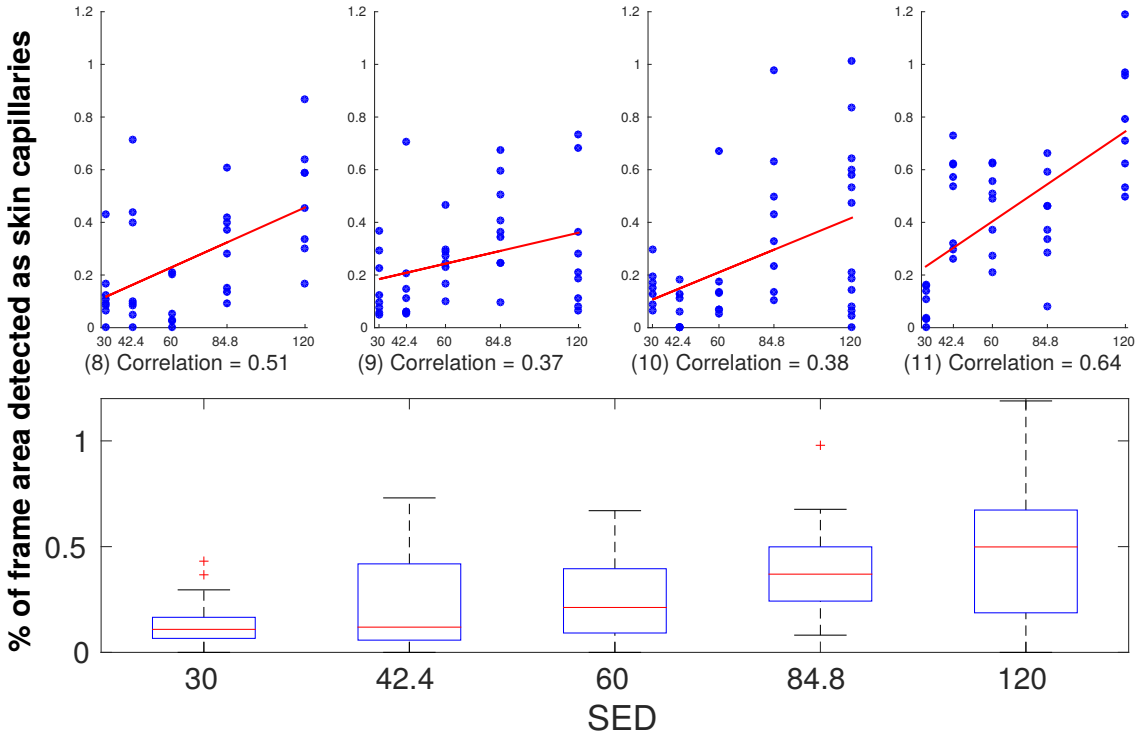


Figure 7.23: Dark skin: The percentage of total capillary area relative to the video's $0.5 \times 0.5mm^2$ field of view for the fair skin volunteers (skin types VI) vs the exposed dose in SED. Plots (8) to (11) represent the results achieved for volunteer # 8 to # 11 respectively, whereas the box plot combines all the volunteers' results together.

7.5.2 Total capillary area vs a^* value

In this section, the total skin capillary area is plotted against a^* value acquired from the Minolta Chroma Meter CM-700d spectrophotometer discussed in Sec-

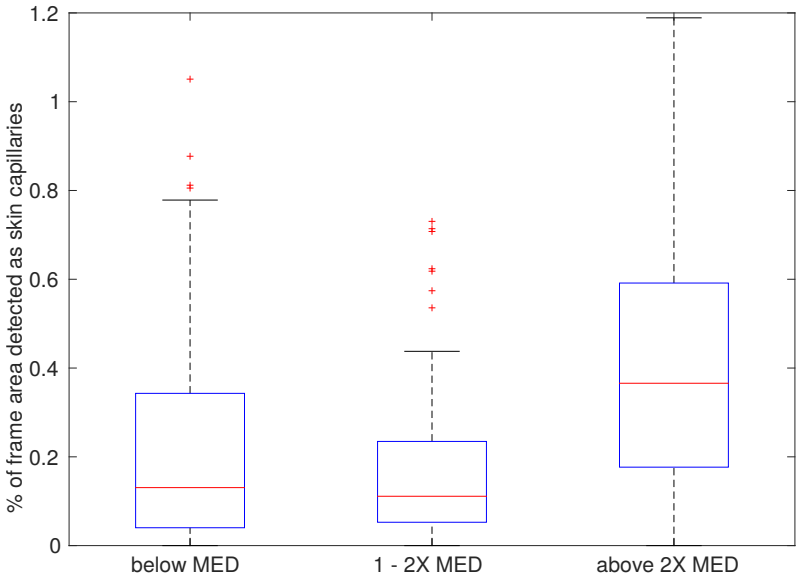


Figure 7.24: Dark skin: The percentage of total capillary area per field of view of the video ($0.5 \times 0.5mm^2$) is given in % with respect to the MED. A comparison is done between doses below MED; doses representing 1 to 2 $\times$ MED and doses greater than 2$\times$ MED.

tion 2.4.2. Figures 7.25 and 7.26 show the total area detected as skin capillaries as a percentage of the frame area available ($0.5 \times 0.5mm^2$) versus the a^* normalised by the volunteer's a^*_{MED}. As opposed to the exposed dose, the a^* value is dependent upon the volunteer's personal provocation level at the skin surface and different a^* levels cannot be easily compared across different volunteers.

By observing the x-axis of plots (1) to (7) in Figure 7.25, it can be noted that the a^* increases by a factor of 3 in some cases (volunteer # 4), but up to 1.6 in others (volunteer # 2) for an exposure of 11.3 SED. In 4 (out of 7) fair cases the a^* increased by more than a factor of 2 of a^*_{MED} at the 11.3 SED exposure site. For each volunteer a positive correlation between total capillary area and normalised a^* is observed. Spearman's rank correlation was calculated for each volunteer resulting in values between 0.54 and 0.73 for fair skinned volunteers.

The box plot in Figure 7.25 shows the capillary area results for all fair skinned volunteers combined. The range of possible normalised a^* values ($\bar{a^*} = \frac{a^*}{a^*_{MED}}$) were collated into four groups: $1 \geq \bar{a^*}$; $1 < \bar{a^*} \leq 1.5$; $1.5 < \times\bar{a^*} \leq 2$ and $\bar{a^*} > 2$. The median percentage value of detected skin capillaries increased from 0.28 to 0.63 to 1.03 and reduced to 1.01% at these four defined bands. There is a

statistically significant increase between the first and last group of values. The Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$).

The above calculations were repeated for volunteers having a darker skin tone (volunteers # 8-11) in Figure 7.26. By analysing the x-axis of plots (8) to (11) in Figure 7.26 it was noted that the maximum observed increase in a^* was 0.3 for an exposure of 120 SED (volunteer #10). The increase in a^* for dark skinned volunteers was less than the increase observed in fair skinned volunteers, which increased by a maximum factor of 3. For each volunteer a positive correlation between detected area and $\bar{a^*}$ is observed. The Spearman's rank correlation was calculated for each volunteer and the values achieved ranged between 0.32 and 0.64.

The box plot in Figure 7.26 shows the capillary area results for all dark skinned volunteers combined. The range of possible $\bar{a^*}$ values were collated into four groups: $1 \geq \bar{a^*}$; $1 < \bar{a^*} \leq 1.1$; $1.1 < \times\bar{a^*} \leq 1.2$ and $> \bar{a^*} > 1.2$. The median percentage value of detected skin capillaries increased from 0.13 to 0.17 to 0.34 to 0.38% at these four defined bands. There is a statistically significant increase between the first and last group of values. The Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$).

7.5.3 Calculating parameters from individual capillaries

Figures 7.27 and 7.28 illustrate the individual capillary areas detected in μm^2 and the length of the minor axis detected in μm versus the dose in SED for fair volunteers. These plots incorporate all the measured parameters for each detected individual capillary. The area detected increased from a median value of $138\mu m^2$ to $249\mu m^2$ from a dose of 0 to 11.3 SED. The minor axis length increased from a median value of $13.2\mu m$ to $17.8\mu m$ across the same doses. Both the area and minor axis length increments are statistically significant. The Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$) between the 0 and 11.3 SED irradiation doses.

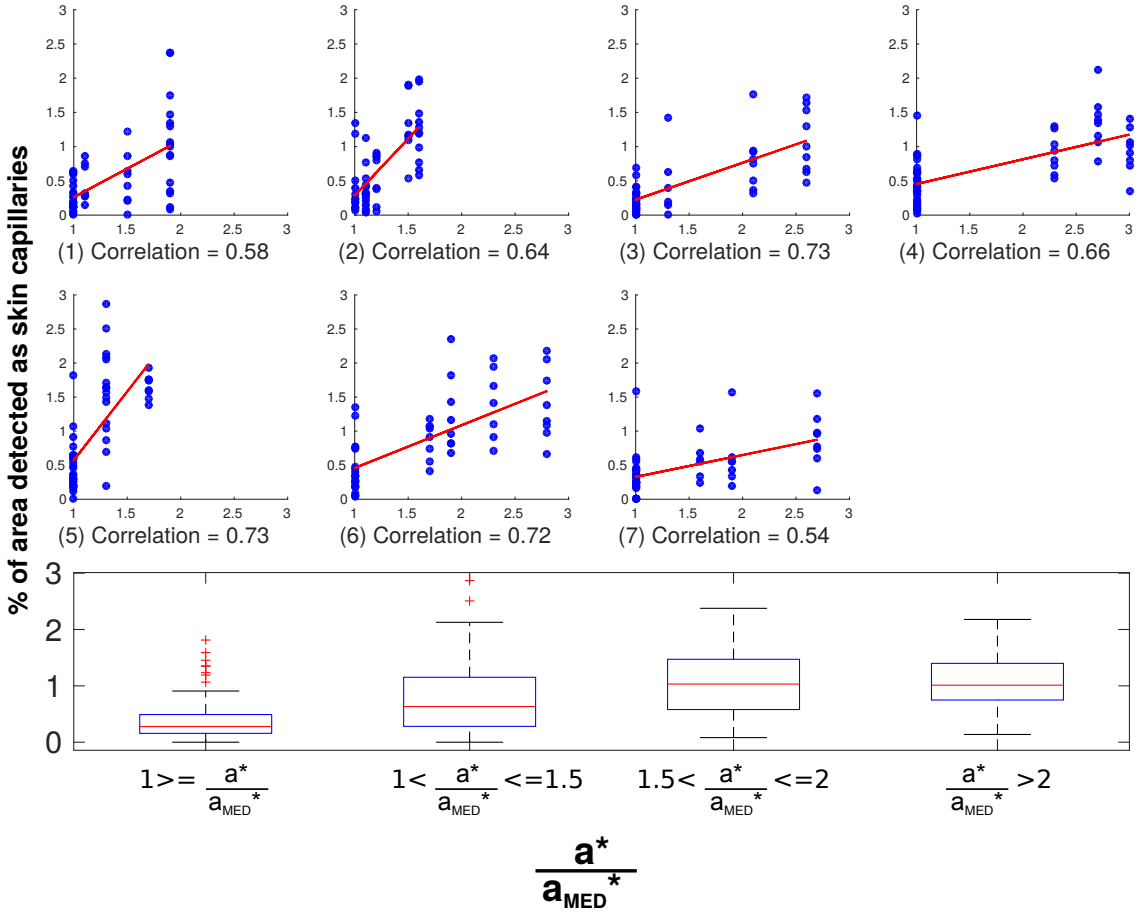


Figure 7.25: Fair skin: Plots (1)-(7) represent the results achieved for volunteers # 1-7, whereas the box plot combines all the volunteers' results together. The percentage of frame area detected as skin capillaries using the proposed algorithm is plotted against $\bar{a^*}$. The value of a^*_{MED} is 4.45, 3.92, 3.82, 1.67, 4.31, 3.62 and 2.69 for volunteers # 1-7.

Figure 7.29 illustrates the increase in individual capillary area from 146.5 to 166.0 to 243.3μm^2 for dose ranges below a volunteer's MED; above MED, but below 2$\times$ MED; and greater than 2$\times$ MED respectively. In addition, the length of the minor axis increased from a median value of 13.6 to 14.5 to 17.6μm for the afore mentioned dose bands. All increments are statistically significant for both the individual capillary area and the minor axis length. The Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$) for all cases.

Figures 7.30 and 7.31 show these results for volunteers with dark skin type (volunteers #8-11) for doses between 30-120 SED. The area detected increased from a median value of 130μm^2 to 162μm^2 from a dose of 30 to 120 SED. The

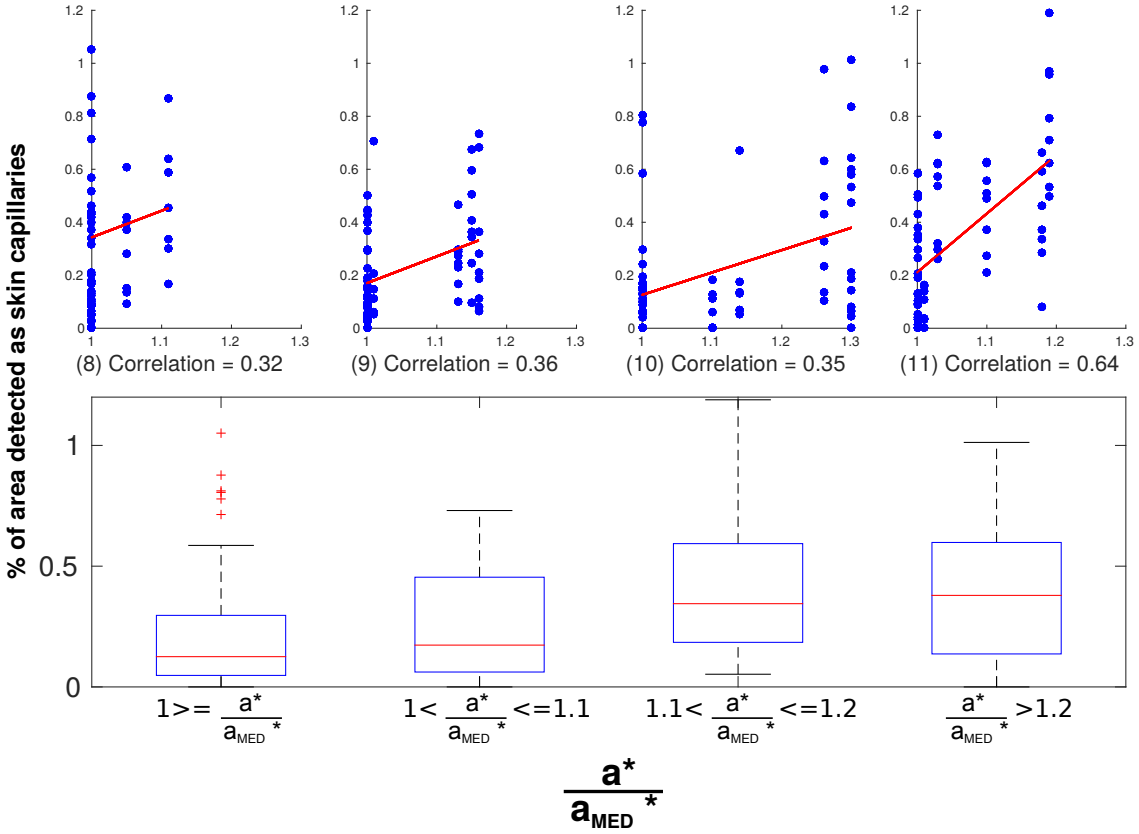


Figure 7.26: Dark skin: Plots (8)-(11) represent the results achieved for volunteers # 1-7, whereas the box plot combines all the volunteers' results together. The percentage of frame area detected as skin capillaries using the proposed algorithm is plotted against $\bar{a^*}$. The value of a^*_{MED} is 6.46, 7.41, 6.15 and 6.15 for volunteers # 8-11.

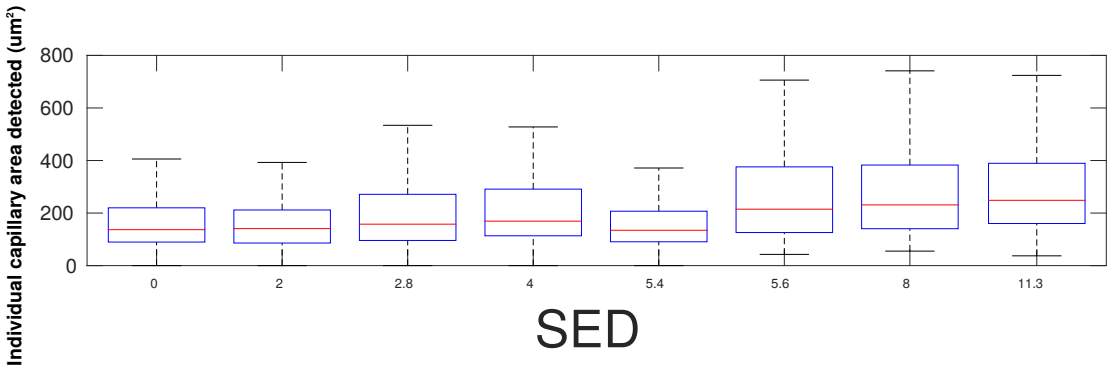


Figure 7.27: Fair skin: Plots (1)-(7) represent the individual capillary area (μm^2) versus dose (SED) for volunteers # 1-7. The box plot represents the combined results of all the volunteers.

minor axis length increased from a median value of 12.9μm to 14.4μm across the same doses. Both the area and minor axis length increments are statistically significant. The Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$).

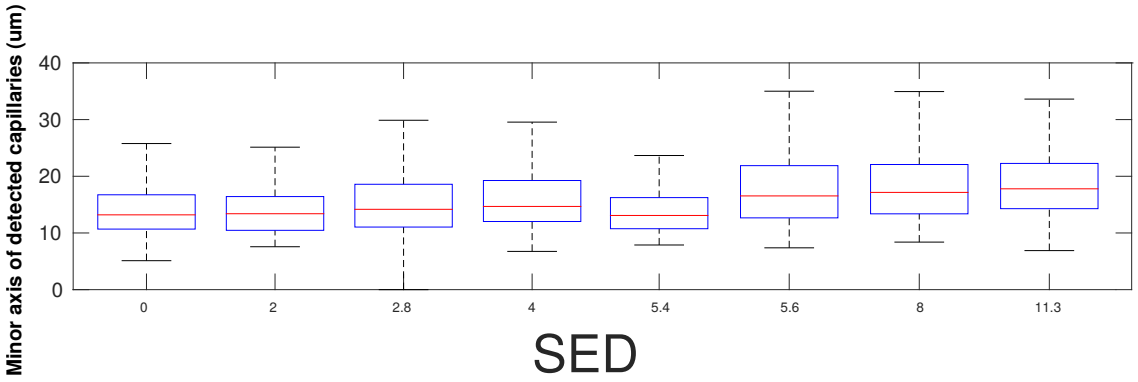


Figure 7.28: Fair skin: Plots (1)-(7) represent the minor axis length (μm) versus dose (SED) for volunteers # 1-7. The box plot represents the combined results of all the volunteers.

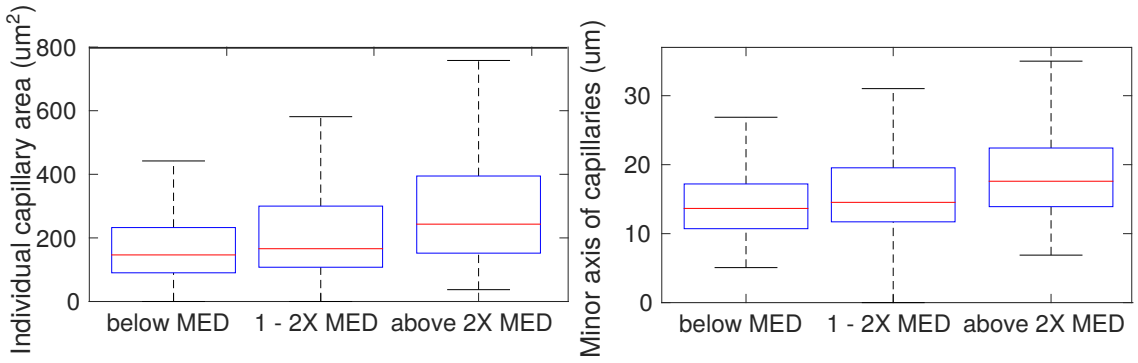


Figure 7.29: Fair skin: Plots of the individual capillary area in μm^2 and length of minor axis in μm, given with respect to the MED.

Figure 7.32 illustrates the incremental individual capillary area from 90.4 to 133 to 162μm^2 for dose ranges below a volunteer's MED; above MED, but below 2$\times$ MED; and greater than 2$\times$ MED respectively. In addition, the length of the minor axis increased from a median value of 10.7 to 13.0 to 14.3μm for the afore mentioned dose bands. All increments are statistically significant for both the individual capillary area and the minor axis length. The Mann-Whitney U-test rejected the null hypothesis of equal medians ($p < 0.01$) for all cases.

The correlations obtained between the individual capillary area (or minor axis length) and SED irradiation are shown in Table 7.2. The tabulated correlation values are new values and calculated per volunteer from the data used to generate Figures 7.27-7.32. The correlation between the individual capillary area and SED is the same as that between the minor axis length and SED. For

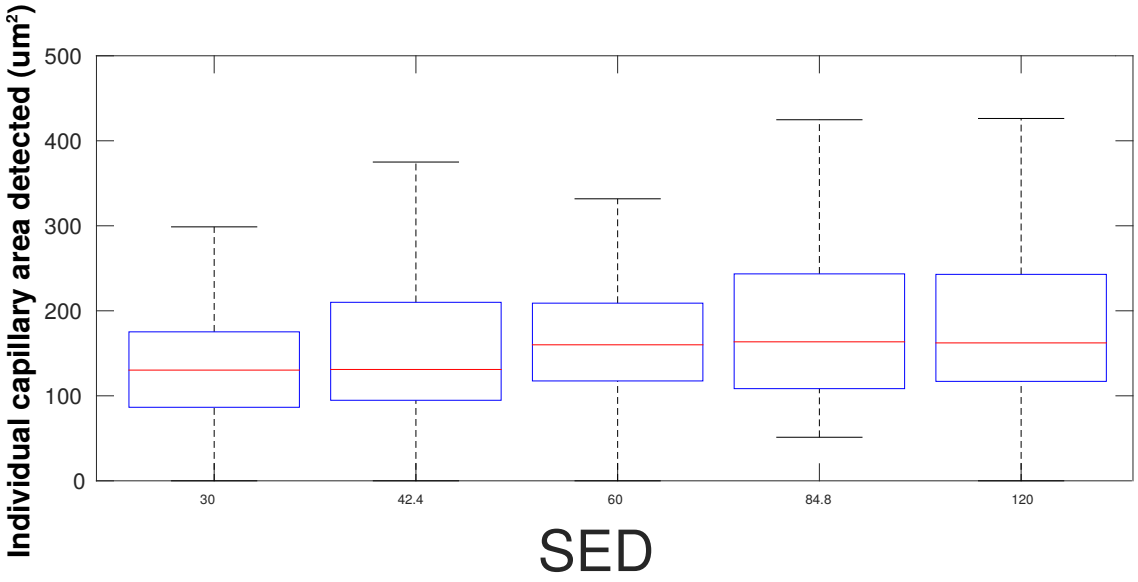


Figure 7.30: Dark skin: Plots (8)-(11) represent the individual capillary area (μm^2) versus dose (SED) for volunteers # 8-11. The box plot represents the combined results of all the volunteers.

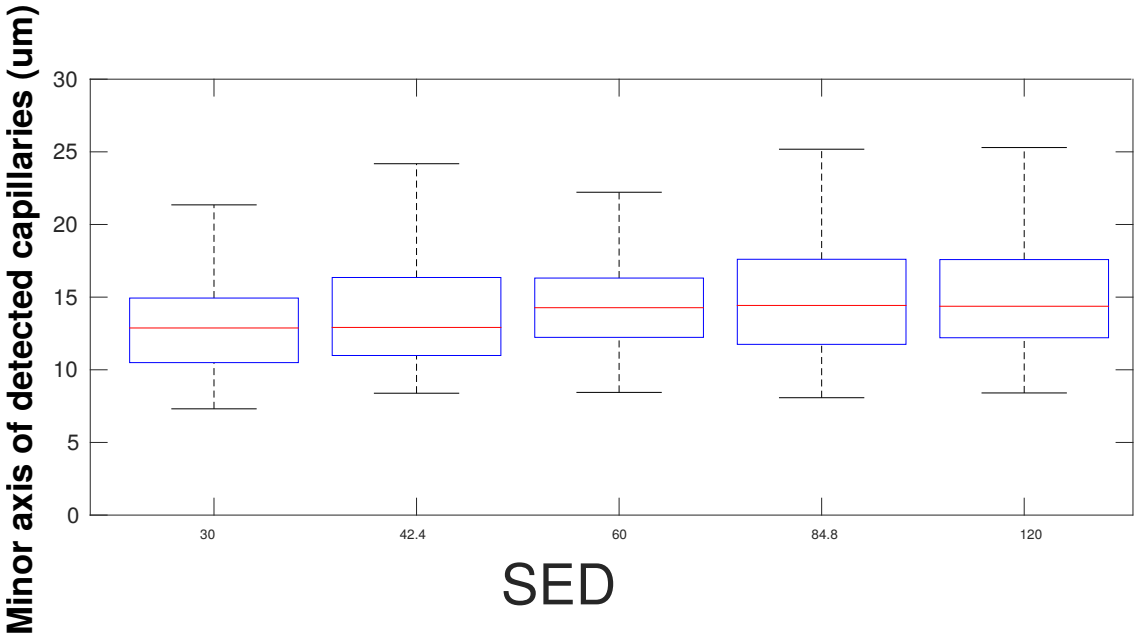


Figure 7.31: Dark skin: Plots (8)-(11) represent the minor axis length (μm) versus dose (SED) for volunteers # 8-11. The box plot represents the combined results of all the volunteers.

the fair skinned volunteers (volunteers #1 $-$ 7) the correlation value ranged from 0.22 $-$ 0.38, whereas for dark skinned volunteers (volunteers #8 $-$ 11) the value ranged between 0.07 $-$ 0.30.

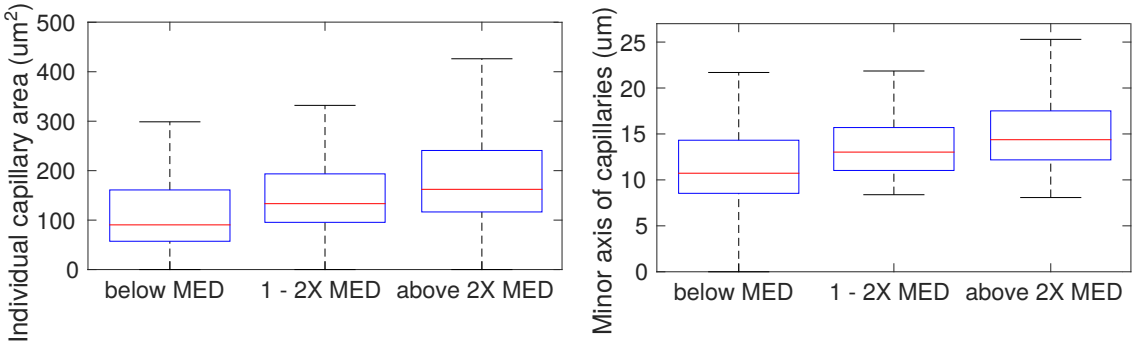


Figure 7.32: Dark skin: Plots of the individual capillary area in μm^2 and length of minor axis in μm, given with respect to the MED.

Volunteer	#1	#2	#3	#4	#5	#6	#7
Correlation	0.38	0.36	0.35	0.22	0.25	0.36	0.31
Volunteer	#8	#9	#10	#11			
Correlation	0.30	0.11	0.07	0.28			

Table 7.2: The Spearman's correlation values between the individual capillary area (or minor axis length) and SED. For each volunteer, the Spearman's correlation between the individual capillary area and SED is the same as that between the minor axis length and SED.

7.5.4 RBCs velocity in cross-sectionally imaged capillaries

Figure 7.33 illustrates the time varying mean pixel value within the arterial and venous capillary limbs for UVR doses 0-11.3 SED. These capillaries were imaged from the same fair skinned volunteer (volunteer # 1). For each dose two time sequences are plotted ($S_A(t)$ and $S_B(t)$), where these sequences represent the arterial and venous limbs of the same capillary imaged cross-sectionally. In the example shown in Figure 7.33 as the dose increases, most noticeably from doses equal to or greater than 4 SED, it becomes harder to detect the corresponding peaks visually. When comparing the plots below 4 SED to those equal or greater than 4 SED one notices a larger number of peaks per unit time, this corresponds to more flashes of RBCs in the RCM video sequences at larger doses. For low doses, most particularly for the 2.8 SED example one can differentiate between the arterial and venous signal by visual inspection. For example, in the 2.8 SED signals as the peaks in the orange signal lead those in the blue signal, the orange signal represents the arterial limb. However as the dose increases, it is not possible to make the distinction of which signal is leading by visual inspection. Furthermore, neither the number of peaks nor the mean intensity value can be

used to differentiate between the arterial and venous signal.

Figure 7.34 looks at the 2.8 SED plot in Figure 7.33 in more depth over a time interval of 9-15 s. This SED and time period have been specifically selected as the graphs shows a good visual correspondence between the two signals. In addition, as illustrated in Figure 7.34, the range of intensity values cannot be used to differentiate between the arterial and venous limb.

Figure 7.35 shows a box plot of the time (i.e. lag) in seconds it takes for a RBC to travel from the arterial to the venous capillary limb for a given SED. The plot incorporates all the results of capillaries captured cross-sectionally for fair skin. The capillaries taken into consideration are those that are represented as two circles of similar area and their centroids are at a distance $20 - 35\mu m$ apart. The Spearman's rank correlation between the UVR exposed dose and the cross-correlation lag is -0.26 with a p-value of 0.023.

The median lag (s) for each dose, extracted from Figure 7.35, and the corresponding velocity (mm/s) is presented in Table 7.3. The velocity was calculated using the methodology defined in Section 7.3.4. The estimated distance between the two limbs is $50\mu m$, as calculated in Section 7.3.4.

Dose (SED)	Δt (s)	Velocity (mm/s)
0	0.254	0.20
2	0.220	0.23
2.8	0.174	0.29
4	0.148	0.34
5.6	0.193	0.26
8	0.157	0.32
11.3	0.090	0.56

Table 7.3: Calculated velocity of blood from a cross-correlation analysis of the time varying signals ($S_A(t)$ and $S_B(t)$)Cross-sectionally imaged capillaries were searched for from all the fair skinned volunteers. A total of 5, 4, 6, 7, 18, 18 and 21 capillaries for doses 0, 2, 2.8, 4, 5.6, 8 and 11.3 SED respectively. All the possible capillaries that fell within the cross-sectional description given in Section 7.3.4 were used.

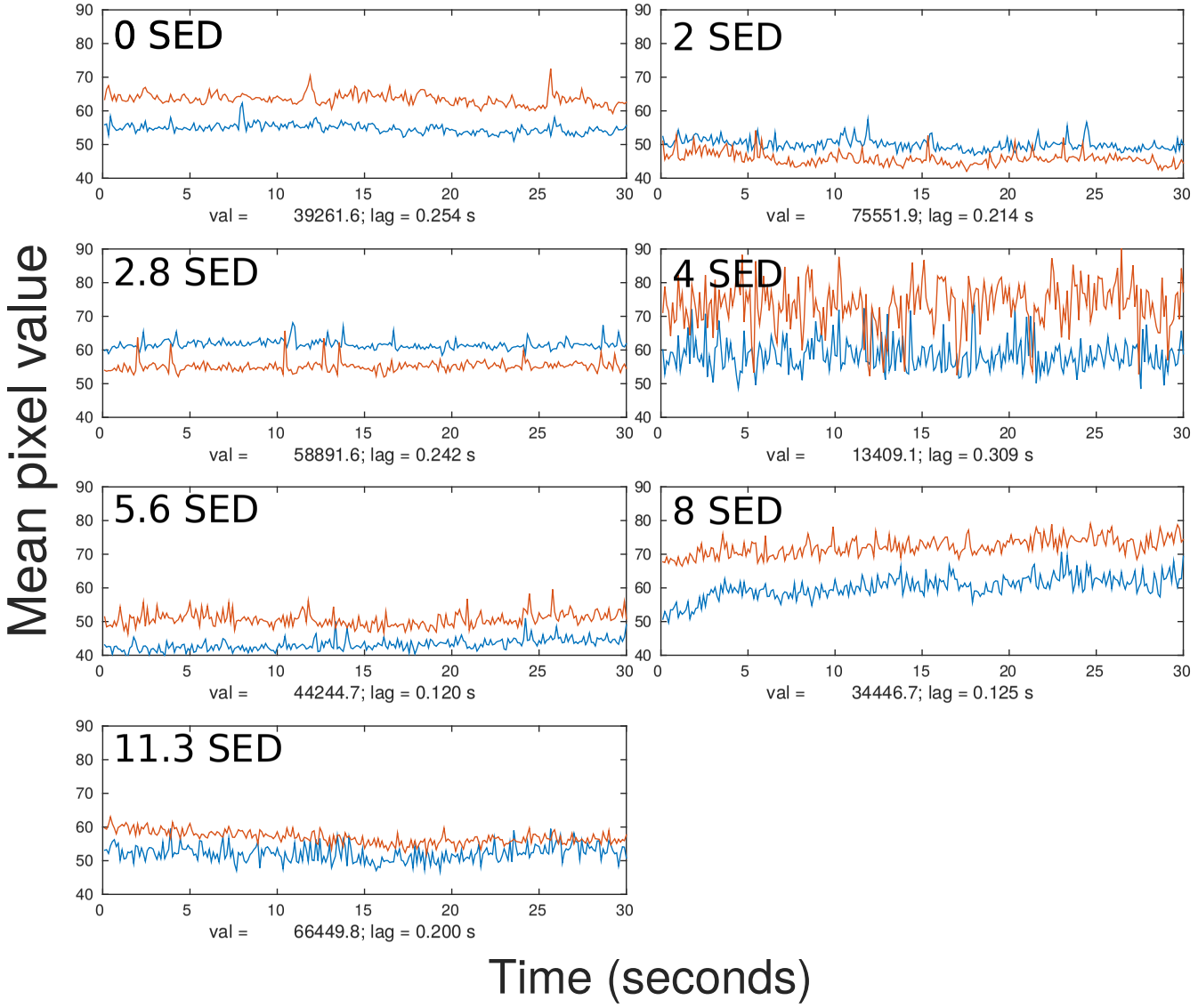


Figure 7.33: Time varying mean pixel values within the cross-sectional area of capillaries. The two sequences represent the arterial and venous limbs for volunteer #1 for doses 0 (unexposed), 2, 2.8, 4, 5.6, 8 and 11.3 SED. The value represents the cross-correlation maximum value and the time represents the lag between the arterial and venous limb (corresponding to the maximum value). The blue and orange signals represent the venous and arterial limbs respectively. The high intensity range found in the 4 SED example is due to the video sequence having a large amount of background noise.

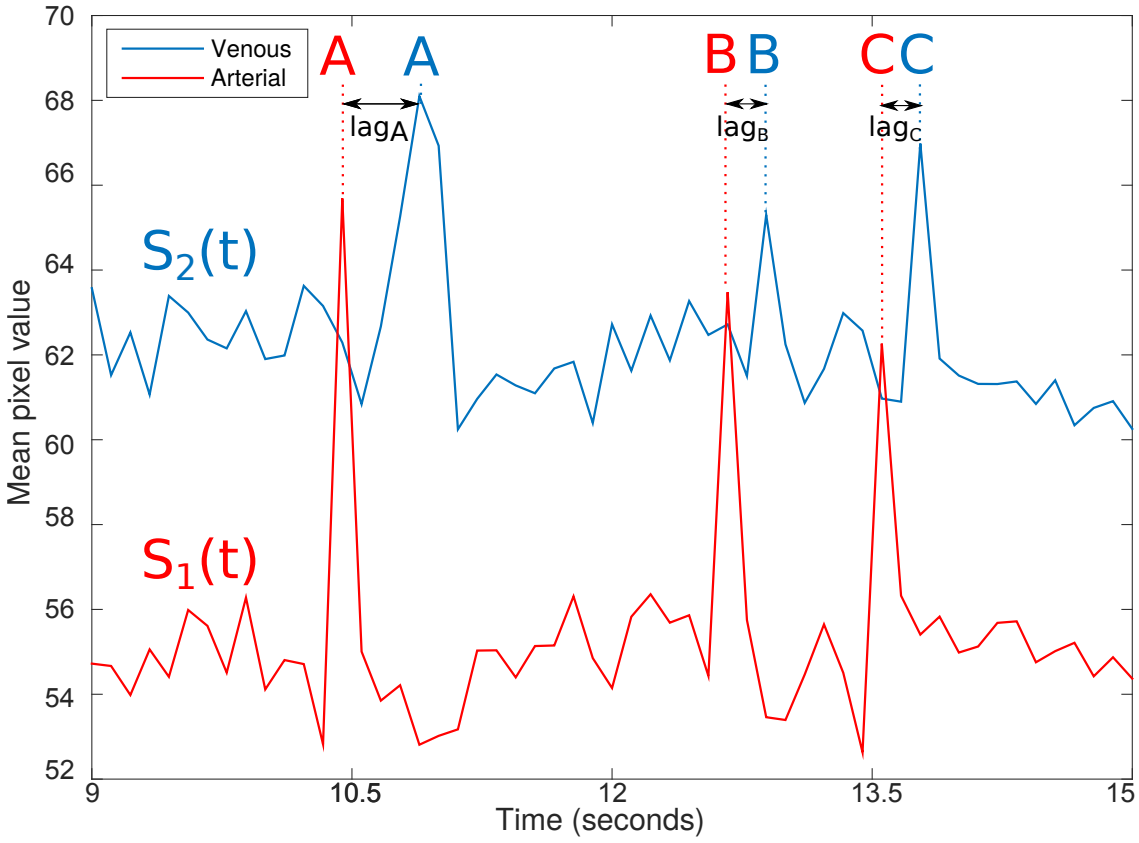


Figure 7.34: The 2.8 SED example illustrated in Figure 7.33 analysed in depth. The arterial ($S_A(t)$) and venous ($S_B(t)$) time sequences are represented by the red and blue colours respectively. As $S_A(t)$ is leading before $S_B(t)$, represented by a negative cross-correlation lag, then $S_A(t)$ is the arterial sequence and $S_B(t)$ is the venous sequence. Three dominant peaks, labelled as A, B and C, are found in both the arterial and venous signals. The lags between the arterial and the venous signals are represented by lag_A, lag_B and lag_C for the three distinct peaks.

7.6 Discussion

7.6.1 Introduction

In this chapter an algorithm that automatically detects skin capillaries in RCM video sequences has been developed. The flow in detected capillaries give a measure of the underlying micro-circulation at the DEJ. As opposed to redness measures at the skin surface, the algorithm analyses the anatomical changes present at vasodialtion (or at vasocontraction). RBCs moving along capillaries appear as bright flashes, and therefore capillaries are characterised as structures having flashes (fluctuations of bright and dim intensity values) across time. The developed algorithm makes use of these fluctuations in intensity to detect capillaries, calculate capillary parameters and also to estimate RBC velocity.

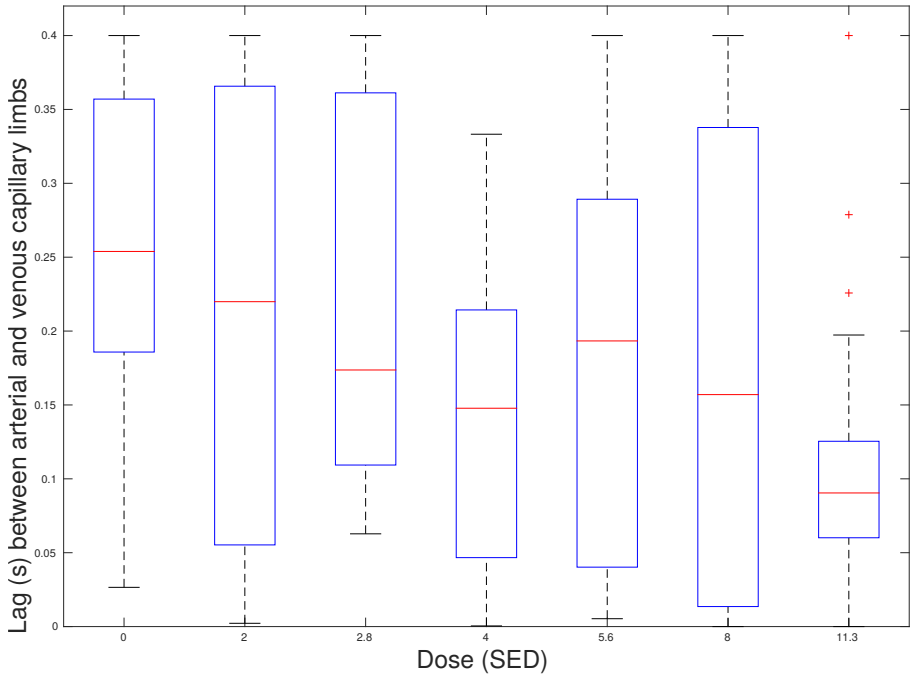


Figure 7.35: A box plot of lag (s) between RBCs appearing in the arterial and venous areas of a single capillary against UVR dose (SED) in fair skinned subjects.

This Section discusses (i) the main findings of the algorithm that detects skin capillary area, (ii) the calculations of individual capillary parameters, (iii) the main findings of the algorithm that quantifies RBC velocity across cross-sectionally imaged capillaries and (iv) the limitations in algorithm and methodology.

7.6.2 Discussion of main findings of algorithm to detect skin capillary area

The total capillary area was compared with the UVR irradiated dose in SED units and in factors of the MED. This comparison contrasts with that found in the literature, which in general focus on changes in size of individual capillaries rather than total capillary area (discussed in Section 7.6.3). From the Experiment identified in Section 7.4.3, it was found that the total skin capillary area increased as the UVR dose increased. This increase is attributed to two main factors: a larger number of capillaries are visible in RCM video sequences as the UVR increases and the size of individual capillaries also increases due to vasodilation. Figures 7.19 and 7.20 illustrates these two factors.

The total capillary area was compared with the a^* redness value measured using a colorimeter. An increase in capillary area was noted before a^* started to increase. This observation strongly backs the argument that changes in a^* cannot be used solely to extract changes in underlying circulation at skin capillaries located at the DEJ. The parameters achieved from the proposed algorithm are due to anatomical changes in the skin capillaries rather than to a change in skin colour at the skin surface.

There was good correspondence between the a^* values measured for fair skinned volunteers to the a^* values reported in the literature. Gambichler et al. [108] reported a $\bar{a}^*$ ($\frac{a^*}{a^*_{MED}}$) of 1.9 at a site irradiated with $3 \times MED$ 24 hours after irradiation. This value corresponds to those achieved from the Experiment identified in Section 7.4.4, where the calculated $\bar{a}^*$ ranged between 1.6-3.

A well documented disadvantage of using the a^* value is that it does not change significantly in volunteers with dark skin. Therefore, it is harder to use the colorimeter's reading as a guide to estimate the MED in dark skinned volunteers than in fair skinned volunteers. This is illustrated by the small range of a^* values achieved as the UVR exposure dose was incremented in dark skinned volunteers. An alternative method to using the colorimeter to measure the level of vasodilation is to quantify the capillary size, discussed in Section 7.6.3. This is particularly useful for dark skinned volunteers.

7.6.3 Discussion of main calculations of individual capillary parameters

From the results achieved from the Experiment in Section 7.4.5 the median capillary diameter (minor axis length) increased by 35% for fair skinned volunteers as the dose was increased to 11.3 SED (from unexposed). In contrast, the capillary diameter increased by 12% in dark skinned volunteers from unexposed sites to sites irradiated with 120 SED. The relevant literature results that this Experiment can be compared to are shown in Figure 7.36.

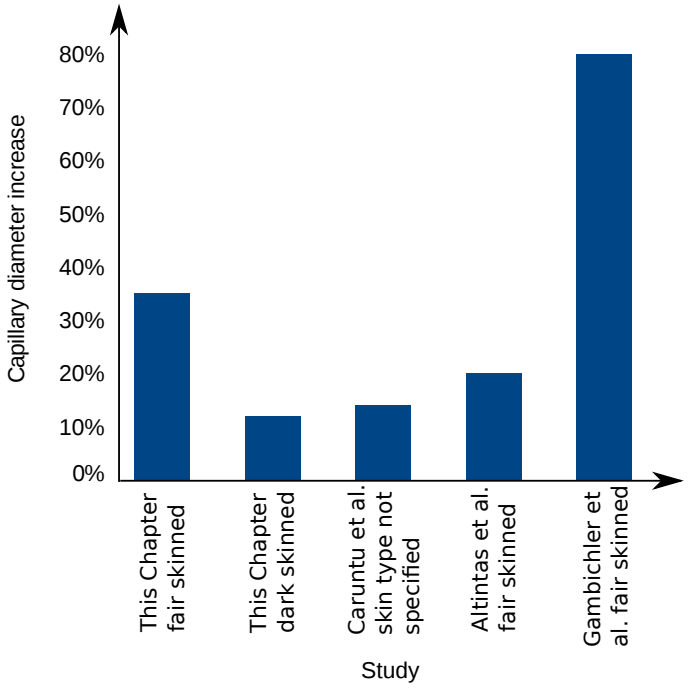


Figure 7.36: A graph of the percentage difference in the diameter of skin capillaries achieved in this study and in prior literature. The results achieved in this Chapter for fair and dark skinned volunteers are compared against those reported in the literature by Căruntu et al. [39], Altintas et al. [107] and Gambichler et al. [108].

Căruntu et al. [39] who reported an increase of 14% in diameter after 40 minutes of applying capsaicin and Altintas et al. [107] who reported an increase of 20% in the capillary diameter after 5 minutes of heating the skin surface. Although both capsaicin and local-heating achieve vasodilation, one cannot compare the results to a particular UVR dose value.

On the other hand, Gambichler et al. [108] made use of UVR irradiation and reported an increase of 80% in the capillary diameter at a site exposed at $3 \times MED$ after 24 hours of exposure of fair skinned volunteers. $3 \times MED$ approximately corresponds to the 11.3 SED site in fair skin. The increase reported by Gambichler et al. [108] is significantly larger than that achieved in the Experiment reported in this Chapter. From the paper [108] there is no obvious reason for this discrepancy as they capture the video sequences using the same VivaScope® 1500 RCM system. However, the paper does not identify the depth at which the video sequences are being recorded at and makes use of UVB irradiation (in this Chapter a wider UVR wavelength range was used). These two

observations might be the reason for this discrepancy.

7.6.4 Discussion of main findings of algorithm to quantify RBC velocity across cross-sectionally imaged capillaries

A cross-correlation methodology, identified in Section 7.4.6, was used to estimate blood velocity in capillaries captured cross-sectionally. From the preliminary results achieved, an increase in blood velocity was observed with an increase in UVR exposure dose. The median blood velocity of fair skinned volunteers increased from 0.20 to 0.56 mm/s as the UVR dose was increased from 0 to 11.3 SED. Similarly, Winchester et al. [148] reported an increase in sublingual blood velocity in response to both hot and cold stimulation. In [148] the velocity ranged from 0.15 to 0.70 mm/s and was calculated using the auto-correlation function on laser speckle measurements. The calculated velocities also fall within the nailfold capillary velocities reported by Stücker [111], which range from 0.14-0.93 mm/s.

Altintas et al. [107] reported an increase in blood flow in response to a hot stimulus. The units of the tabulated blood flow are in quantitative blood cell flow per minute (qbf/min). This measure corresponds to the manual count of peaks inside the capillary loops of RCM video sequences. Assuming this measure is proportional to m^3/s, one can derive the velocity by dividing the the quantitative blood cell flow per minute value by capillary area. Using this methodology, the derived velocity decreased by a 8.4% and increased by 3.2% directly after and 5 minutes after local heating respectively. This trend [107] suggests that the estimated flow could contain errors, or that the reported qbf/min is not necessary proportional to the standard blood flood units (m^3/s). As depicted in Table 3.2 (Section 3.4), the calculated flow in RCM literature is in cells/minute. This limits direct comparison to the derived blood velocity estimated in this Experiment.

When estimating the cross correlation lag between the arterial and venous signals, one cannot guarantee that the cross-correlation lag is correct, mainly due to the limited RCM frame rate (of 9fps). As higher blood velocities are expected at higher UVR exposures, the limited frame rate affects lag estimations from

videos captured at sites exposed with higher doses more than those captured at low doses.

7.6.5 Limitations in algorithm and methodology

This Section is subdivided into: (i) variability found across intra- and inter-volunteer sites, (ii) confounding features within the data that hinder the developed algorithm, and (iii) limitations in the development of the algorithm.

7.6.5.1 Intra- and inter-volunteer variability

As shown in the results section, the higher the UVR dose the easier it is to see RBCs. At higher doses, the capillaries dilate and thus appear larger both if imaged cross-sectionally or in-plane. In addition, a higher dose is complemented with a larger blood velocity. These two factors make skin capillaries more visible and detectable at higher doses.

Another observation that can be deduced from these results is that the deeper within tissue the capillary is visualised the the easier it can be detected. In all of the volunteers, the capillaries were captured inside the DEJ region. However, in some volunteers the DEJ is characterised more with epidermal structures, whilst in other volunteers the DEJ appears to contain more dermal structures. In the case that the capillaries are surrounded by epidermal like structures, it is more difficult to detect, as the texture of the epidermal structures are easier to mistake as blood flow than dermal-like structures.

7.6.5.2 Confounding features in data

The presence of artefacts typically result from poor optical contact between the microscope and the skin (e.g. wrinkles) hinder the developed automated capillary detection algorithm, especially for non-exposed and low dose sites (below the volunteer's MED). Non-contact artefacts tend to result in large time varying changes in pixel value and can be mistaken as blood flow in capillaries by the developed algorithm. Even small pressure changes between the microscope and skin surface can result in rapid pixel value variation within these artefacts.

Due to these factors, artefacts can cause varying intensity values across different frames, which cannot be adequately corrected for using non-rigid registration algorithm.

Although efforts were made to capture videos in areas free from artefacts, in some volunteers (especially skin types I-II) it is not possible to find regions exhibiting no artefacts, as the exposed region being scanned is thoroughly characterised with wrinkles. To illustrate this, I will focus on the results achieved for volunteer # 7 which has the lowest correlation value (0.48) for the fair skin cases, and also has a large presence of wrinkles. As depicted in Figure 7.37, wrinkles are mistaken to be blood flow mainly for plots A-D which correspond to doses 0-4 SED. The MED for volunteer # 7 is 4 SED. As the RBC signal is lower in videos scanned on sites with lower doses (doses below the volunteer's MED) than in sites exposed with higher doses, the wrinkles have a higher chance of being mistaken as RBC signal in lower dose sites. Despite this limitation, neither wrinkles nor other artefact regions were manually masked out.

In general, the videos collected from one unexposed or exposed site have the same level of confounding features across the $8 \times 8mm^2$ scanning view. However, for the case of fair skin, such features have a higher chance of being mistaken to be blood in unexposed or low dose sites by the algorithm. As the dose increases, the signal from blood increases, resulting in an increase in the signal to noise ratio. Therefore, confounding features have a greater chance of biasing results for doses on the lower end than for those at the higher end of the dose scale.

The presence of a high concentration of melanin is a confounding feature in darker skin types (skin type IV-VI). Its presence hinders the automated detection algorithm for the case of darker skin types. As opposed to confounding features in fair skin, melanin affects videos across all doses as the signal coming from melanin is brighter than that of RBCs. This effect is shown in Figure 7.38, where the worst case for each site for volunteer # 9 is depicted.

To summarise, confounding features are more likely to be mistaken to be

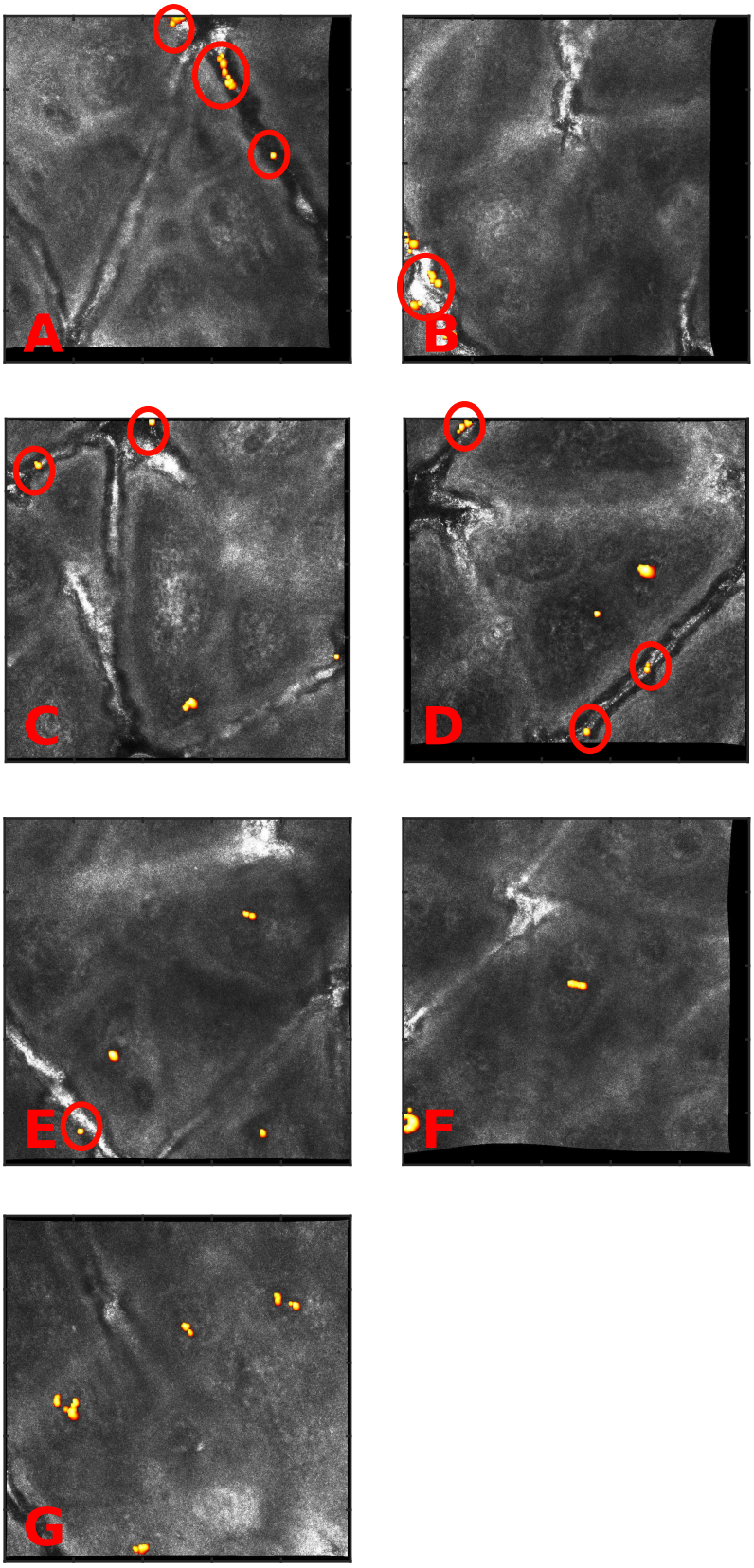


Figure 7.37: Fair skin: volunteer # 7: (A) 0 SED (B) 2 SED (C) 2.8 SED (D) 4 SED (E) 5.6 SED (F) 8 SED (G) 11.3 SED. The circled regions are wrinkles mistaken to be blood flow - a higher presence of mistaken regions is present in doses below or equal to the volunteer's MED (4 SED)

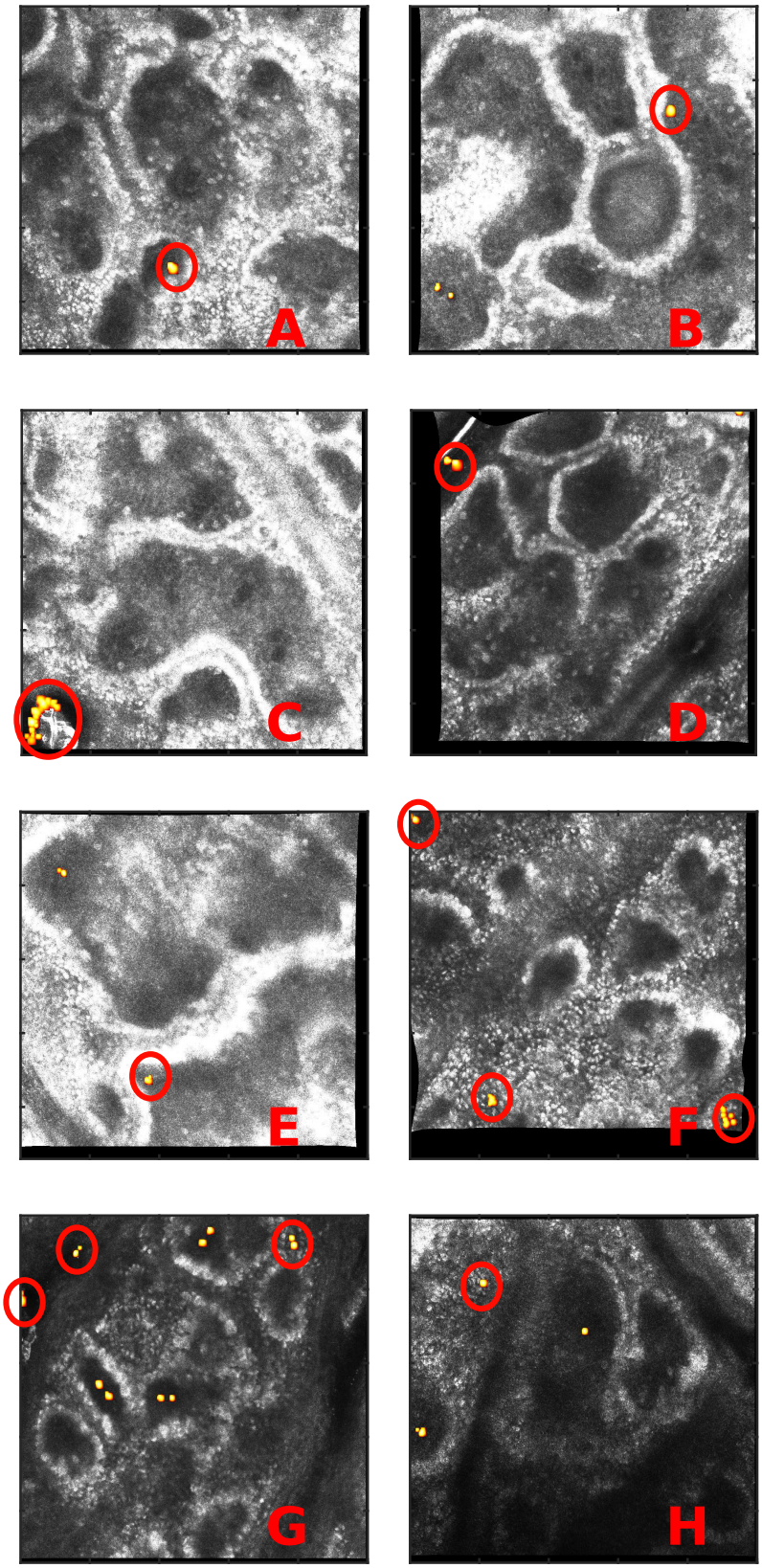


Figure 7.38: Dark skin: volunteer # 9: (A) 0 SED (B) 15 SED (C) 21.1 SED (D) 30 SED (E) 42.4 SED (F) 60 SED (G) 82.4 SED. The circled regions are wrinkles, artefacts and melanin mistaken to be blood flow.

RBCs in lower doses than in higher doses for fair skin cases. This is due to the fact that in fair skin the signal generated from these confounding features is lower than that generated from RBCs, which is more visible at higher doses. However, for the case of dark skin, the signal generated from the melanin is higher than that of the RBCs, and thus can be mistaken to be RBCs at all doses.

7.6.5.3 Limitations in data collection

The field of view of the VivaScope®1500 is $8 \times 8mm^2$ and a video sequence is captured from a sub-site of area $0.5 \times 0.5mm^2$. This leads to the need for the confocalist to search the field of view to locate eight sub-sites which have visible capillaries. For unexposed and low dose sites (generally below the volunteer's MED), the user needs to search a larger area to find eight sub-sites containing capillaries from which to record video sequences. Thus, the captured videos in higher doses are more representative of the $8 \times 8mm^2$ field of view than the videos captured from unexposed or low dose sites. For the latter sites, the selected video locations do not represent what is happening across the whole $256 \times (0.5 \times 0.5mm^2)$ possible scan areas, but give a representation of the maximum capillary response in the site. Therefore, this approach will tend to cause a possible bias in results of the total capillary area for sites with no or low irradiation dose.

For volunteers with a dark skin tone, the power that was used to capture the video sequences using the VivaScope® 1500 confocal microscope was manually set to $2 - 2.5mW$ to cater for the lack of signal present from blood flow at lower power levels. However, the signal generated from the capillaries was still low when contrasted to those captured for fair skin. This results from the melanin at the DEJ boundary attenuating the laser beam. This can be compensated by manually setting a high laser power. Whilst this makes capillaries more visible, there remains a higher noise level in those cases.

7.6.5.4 Limitations in the development of the algorithm

The algorithm has been designed to detect capillaries by detecting the passage of multiple RBCs in the same spatial location. Therefore it is easier for the algo-

rithm to detect blood in cross-sectionally imaged capillaries than those imaged in-plane, as in the cross-sectional case RBCs are present in the same circular area, whilst in the in-plane case RBCs appear along different points of the capillary.

A further limitation, is that the algorithm searches for RBCs that appears as bright circles, the rationale being that RBCs are highly refractive and appear circular in shape in confocal microscopy. RBCs in cross-sectionally imaged capillaries are easily identified (Figure 7.1). RBCs in capillaries imaged in-plane may not be individually resolved, therefore they may appear as non-circular structures of high intensity (Figure 7.2). For example, in the case of a chain of cells travelling through an elongated capillary together, they will appear more cylindrical-like than circular and may not be detected by the Hough circular transform.

7.7 Future Work

The results achieved from the preliminary RBC velocity calculations are promising. Further experiments can be done in this area to test the robustness of this method. The experiment can be extended by initially analysing more cross-sectional capillaries. Capturing videos of a cross-sectional capillary at various depths, for example at $2.8\mu m$ depth increment, can provide multiple velocity values of the same capillary. Theoretically the velocity values from each depth should be equal.

More work needs to be done for this algorithm to be applied to a clinical setting. Initially, a semi or fully-automated algorithm that masks out confounding features such as wrinkles and melanin is needed. An algorithm that detects confounding features has also been proposed in Section 5.6.3. For example, wrinkles which are linear in shape can be detected using the linear Hough transform. Another limitation is the operational time needed; at present it takes approximately 4 hours to collect and process one video sequence. The majority of this time ($\sim$ 2 hours) is needed to motion correct the video sequence. Registration algorithms that require less operation time can be evaluated. Other methods to speed up processing, such as parallel computing could be investigated.

Once these limitations have been addressed, a direct application of the extended automated algorithm is to phototherapy, where the skin is exposed to UVR doses to treat a number of skin conditions. This algorithm could provide phototherapists with quantitative parameters of skin capillary changes. Furthermore, the extended automated algorithm can be used to measure a patient's MED during phototherapy.

Although a provocation method was used in this study, the proposed algorithm can be applied to detect capillaries in non-provoked areas. Thus dermatologists can use this algorithm to quantify vasculature diseases, that affect skin capillary populations and sizes.

7.8 Conclusion

This Chapter has described the processes that were involved in designing an algorithm that automatically detects and quantifies skin capillary area and the velocity at which RBCs travel at within the capillaries in RCM video sequences. Throughout this chapter the potential of RCM to analyse skin capillaries quantitatively is shown.

The algorithm developed here was applied to datasets collected from a number of volunteers exposed with incremental UVR doses. The algorithm detected an increase in the total capillary area, individual capillary area and individual minor axis length with an increase in UVR dose. These increments confirm that the algorithm is capable of quantifying changes attributed to vasodilation.

The developed algorithm can also be used to quantify blood velocity inside capillaries imaged cross-sectionally. Prior to this work no studies had used RCM video sequences to measure the velocity at which RBCs travel at inside the capillaries located at the DEJ.

Although there are limitations with the developed algorithm with respect to

operation time and confounding features, the proposed algorithm is the first step in being able to quantify changes in circulation at the DEJ due to a provocation methods or due to skin conditions and diseases in the clinical setting.

Chapter 8

Conclusion

In this thesis automated algorithms have been designed and developed to aid dermatologists when analysing reflectance confocal microscopy (RCM) images of the skin. This chapter summarises the main achievements and limitations of the developed algorithms, additionally the potential of these algorithms in the clinical setting is defined.

8.1 Summary and Context of the Contributions

Throughout this thesis three main algorithms were developed. In this Section, a summary of each algorithm, together with the clinical impact and the limitations are highlighted.

8.1.1 Detecting the dermal-epidermal junction in RCM cubes of images

A classification algorithm that identifies and quantifies the dermal-epidermal junction (DEJ) from cubes of images of the skin was designed and developed. The aspect that distinguishes this algorithm from those developed in the prior literature [1] [4] [5] is that the designed algorithm makes use of clinically representative datasets, without needing to remove tiles or regions exhibiting artefacts. A total of 11 datasets, each having approximately 1400 images, were used for this Experiment. To cater for the large number of tiles the features extracted were from tiles rather than sub-tiles and thus such features quantify larger scale

structures rather than from cell information.

The mean percentage of correctly classified images was of 63.2% and 62.1% for intra and inter-volunteer layer classifiers when using both texture and anatomical based features. This value is better than the 58.6% correctly classified images reported in [3] for the same application (that of differentiating between epidermal, DEJ and dermal images).

The mean error in detecting the DEJ thickness was of $10.8\mu m$ and $20.1\mu m$ using the classification results achieved from the intra and inter-volunteer models using both texture and anatomical based features. The diameter of a basal cell, which defines the boundary of the DEJ region, is of $10 - 15\mu m$ [1]. Kurugol et al. [3] state that an error smaller than $15\mu m$ is negligible when locating the DEJ region. The DEJ thickness errors cannot be compared directly to those reported in prior literature as no paper quantifies the thickness.

Potential Clinical Impact An algorithm which is capable of differentiating between the skin layers and of extracting quantitative information, can aid clinicians in visualising and analysing RCM images faster [5]. The algorithm can provide dermatologists with quantification parameters of layer thicknesses and locations. A potential application for this algorithm would be to serve as a platform for subsequent automated algorithms. For example, if a new algorithm aims to quantify the size and concentration of papillae at the DEJ, it can initially process the images with this algorithm to locate the tiles classed as DEJ. Another possible application of this algorithm is to aid clinicians who are relatively new to RCM imaging in familiarising themselves with the different skin layers. Thus, such an algorithm can serve as a training platform.

Limitations At present the mean errors in locating the epidermal-DEJ and DEJ-dermal boundaries are of $12.8\mu m$ ($27.0\mu m$) and $15.8\mu m$ ($6.6\mu m$) for the intra (inter) volunteer models using both texture and anatomical based features. Assuming that the errors should be less than $15\mu m$ [1], more work needs to be done for the classifier to be capable to differentiate between tiles of class epidermis and DEJ,

in particular for the inter-volunteer models. This error has the potential of being reduced by introducing more training datasets from different volunteers.

Currently this experiment was only performed on healthy skin sites and not tested on diseased data. For this experiment to be adopted in the clinical setting, this algorithm needs to be capable of detecting the DEJ in datasets captured both from healthy and diseased sites. Furthermore, this algorithm is not currently designed to provide real-time identification and quantification of the DEJ. Thus, further work needs to be done to implement this algorithm in real-time at the clinical setting.

8.1.2 Detecting cancerous regions in RCM images of excised tissue

A classification algorithm that identifies regions of disease in images of excised skin tissue acquired during Mohs surgery was developed to differentiate between cancerous, inflammatory, and non-diseased regions. This study is the first to automatically analyse RCM images of excised tissue. To improve algorithmic performance nuclei concentration based features were automatically extracted. The introduction of nuclei concentration based features increased the number of correctly classified regions classed as *tumour* and *inflammation* on average by 8.6% and 62.5% respectively. Thus, nuclei concentration based features are particularly beneficial for *inflammation*.

Potential Clinical Impact At present the developed algorithm only correctly classifies tumour regions 58.3% of the time when using both texture and nuclei concentration based features. For this method to substitute histopathology in Mohs surgery, the number of correctly classified cases should be greater than 99% [26]. Possible future work that can be done to improve classification results is to provide the classifier with neighbouring knowledge when making a classification of a region and also to sub-divide class *other* into further sub-classes. This algorithm provides a step towards the acceptance of applying RCM to image excised tissue in the clinical setting and of computer-aided diagnosis.

A possible application of RCM imaging of excised tissue in Mohs surgery is to RCM image excised tissue (instead of performing a histology examination) for a subset of the excisions performed during Mohs surgery. In the proposed methodology the first and last excisions would be analysed using histology, whilst all the intermediate excisions would be analysed using RCM imaging. Once the image is analysed by the clinician it can be fed into the algorithm which will serve as a second reader to achieve a computer aided diagnosis. Using this methodology the duration and cost of Mohs surgery can be reduced significantly, resulting in greater accessibility to a wider set of patients.

Limitations At present the procedure used to prepare excised tissue for RCM imaging does not always translate to images that can be used to make a diagnosis. This is mainly due to the sample not being evenly thick and the RCM technology not being able to image all parts of a sample which does not have the same thickness throughout, resulting in the epidermis (or the thinnest part of the tissue) being missed from the image. Image quality is also reduced when the fluorescent dye is not evenly distributed and results in over-saturated images. Due to this limitation, in this thesis I have focused on BCC cases classed as micronodular, nodular and infiltrative. Superficial BCC was not included in this study as the imaging procedure used missed the epidermis layer in some cases resulting in only a few images covering superficial BCC. Possible future experiments include preparing the tissue on an unlevelled microscopic slide to compensate for unevenness in sample thickness.

8.1.3 Detecting skin capillaries in RCM video sequences

An algorithm was developed to detect and quantify the presence of skin capillaries at the DEJ from RCM video sequences. An additional algorithm was developed to quantify the velocity at which red blood cells are travelling at inside capillaries captured cross-sectionally. According to the author's knowledge, these are the first automated algorithms that detects skin capillaries and which quantifies the velocity of RBCs using confocal microscopy.

This algorithm was applied to video sequences captured from forearm sites of fair (skin type I-III) and dark skinned (skin type VI) volunteers exposed with incremental UVR doses. In the case of fair skinned volunteers, a dose range of 0 - 11.3SED was used, whereas for the dark skinned a dose range of 0 - 120SED was used. Results show that the RCM video sequences can be used to analyse vasodilation and the developed algorithm is capable of quantifying the different levels of vasodilation found across sites irradiated with incremental UVR doses.

The median capillary diameter increased by 35% for fair skinned volunteers as the dose increased from 0 SED to 11.3SED, and by 12% for dark skinned volunteers as the dose increased from 0 SED to 120 SED. The increase in capillary diameter linked to vasodilation reported in prior literature ranged between 14% and 80% [39] [107] [108]. The median RBC velocity in fair skinned volunteers increased from 0.20 to $0.56 mm/s$ as the UVR dose increased from 0 to 11.3SED. This compares well to the range of RBC velocities of 0.15 to $0.70 mm/s$ reported in the literature [148].

Potential Clinical Impact This algorithm points to a new use for RCM as a means of providing a more quantitative method of assessing erythema and vasodilation.

In addition to providing quantification parameters of skin capillaries and measures of RBC velocity, this algorithm can also provide clinicians with a measure of vasodilation. Thus, this algorithm can be used to track the response of certain skin conditions (that affect skin capillary population) to treatment, by checking whether the capillary population is regulating.

Additionally, the ability to measure vasodilation and RBC velocity in superficial capillaries in skin with RCM provides a new and objective means of quantifying the physiological effect that underpins erythema. It therefore may serve as a more valuable end-point than colorimeter measurements in photobiology experiments and treatments.

Limitations Although experiments were performed on volunteers with fair (I/II) and dark (VI) skin tones, they did not cover all six skin types. Further experimental work involves the algorithm being tested on video sequences captured from all skin types. In addition, further experiments are needed to capture video sequences with clearer skin capillaries for dark skinned volunteers and/or to develop algorithms that enhance the visibility of RBCs inside skin capillaries of video sequences captured from dark skinned volunteers. The current method adopted to enhance the visibility of RBC signals in video sequences captured from dark skinned volunteers is to increase the power of the laser. However, this method besides enhancing the RBC signal enhances the melanin signal, resulting in poor detection of skin capillaries in video sequences captured from dark skinned volunteers.

8.2 Future Work

Of the three algorithms that have been designed and developed in this thesis, the one that is closest to clinical application involves detecting and quantifying skin capillaries in RCM video sequences. For future translational projects, this is the algorithm I propose which has the most novelty and applicability. In this section methods to make the measurements more robust and accurate are identified. Additionally, potential applications of such values in the clinical setting are identified.

8.2.1 Enhancing the RBC signal

Overall, the developed algorithm works very well in datasets of fair skinned volunteers, however not so well in datasets captured from dark skinned volunteers. A correlation value ranging from 0.48 to 0.74 between total capillary area detected and SED dose for datasets captured from fair skinned volunteers. Whereas, the correlation value between the total capillary area and SED dose reduced to a range from 0.37 to 0.64 for dark skinned volunteers. Similarly, the correlation values between total capillary area and $\frac{a^*}{a^*_{MED}}$ ranges from 0.54 to 0.73 for fair skinned volunteers and between 0.32 and 0.64 for dark skinned volun-

teers.

For this algorithm to be used in a clinical setting in conjunction with a colorimeter the results achieved need to be improved, especially for dark skinned volunteers. As the colorimeter a^* value is not representative of the level of vasodilation in dark skinned volunteers, there is a bigger clinical need to measure the level of vasodilation using an alternative method for dark skinned volunteers. Future work that enhances the RBC signal will enable better detection of skin capillaries. This would encourage future translational projects that would make use of the algorithm to measure vasodilation.

The first methodology in attempting to enhance the RBC signal is to capture longer video sequences than the current $30s$ sequences used in this thesis. Longer video sequences have the potential of having a higher signal to noise ratio. These sequences then can be fed to the developed algorithm and the results achieved can be analysed to check if this improves capillary detection. It is suggested that the developed algorithms run in parallel (using parallel computing) to cope with the larger file sizes associated with longer video sequences.

Another way of enhancing the RBC signal and of improving the robustness of the algorithm is to develop an algorithm that masks out confounding features, such as wrinkles and melanin in darker skin. At present, confounding features exhibit large changes in intensity values over a short period of time, resulting in the algorithm mistaking them as RBC signal, especially in datasets captured from dark skinned volunteers and from sites irradiated by low (or no) UVR dose. The need of algorithms that detect and segment confounding structures (like wrinkles and melanin regions) has been identified in Section 5.6.3.

8.2.2 Repeatability of RBC velocity measures

The results achieved in this thesis indicate that RCM technology has great potential in visualising RBCs and that it can be used to quantify the velocity at which they travel at inside capillaries. However, as identified in Section 7.5.4 the cross-

correlation lag (s), the time for an RBC to travel from the arterial limb to the venous limb, varies considerably across different capillaries imaged from sites exposed with the same UVR dose. Here, a methodology that achieves more accurate values of RBC velocity by repeated measures is defined. A cross-sectional capillary can be imaged at different skin depths using RCM video sequences. The RBC velocity can be calculated at each video sequence (using an estimated distance for each depth) and the results achieved can be assessed. This experiment could illustrate whether repeatability and more accurate velocity measures can be obtained from the same skin capillary.

8.2.3 Apply algorithms to diseased datasets

In this thesis I have made use of the developed algorithms to detect and quantify the differences in capillary size and RBC velocity attributed to vasodilation provoked by UVR irradiation. However, these algorithms can be also applied to quantify diseases characterised by abnormal capillary populations, such as Raynaud's phenomenon [52]. The developed algorithms can be applied to video sequences captured from diseased and non-diseased (control) sites. A study could be done to analyse and quantify how the skin condition is affecting the capillary population and capillary size. Furthermore, the effect that the condition has on RBC velocity could be studied.

8.3 Summary

This thesis has attempted to demonstrate the value of RCM in dermatology and how a more detailed analysis of the images can provide valuable clinical information that is not easily or quickly extracted by visual analysis alone.

RCM has a near histology lateral resolution and an ability to obtain multiple images from a wide field of view and from different depths. A typical RCM examination provides the user with a large amount of images. Automation of algorithms to extract all clinical information is unlikely ever to be possible and, indeed, the role of the dermatologist in interpreting the images will remain critical.

However, the work here has shown that, in carefully selected areas, automation can assist the user in (i) locating a key anatomical feature, the DEJ, with useful accuracy, (ii) highlighting regions of tissue samples to direct the clinicians attention to the possible presence of cancer and inflammation and (iii) highlighting some of the smallest capillaries in the human body and providing a measure of the velocity of RBCs as they pass through a capillary loop.

RCM is gradually gaining acceptance in dermatology clinics in Europe and the USA. Current barriers to its acceptance include the traditional reliance on visual examination of skin and the time it takes up in analysing the images obtained from a typical RCM scan. Any software tools that can help speed examination, aid in training clinicians who are new to RCM and which provide objective quantitative analyses will be valuable. This thesis presents some early novel efforts in this direction.

List of Presentations and Meetings

2013	Training	In-vivo VivaScope® 1500 and 3000 training organised by Mavig GmbH, Modena, Italy (April 2013)
2015	Poster presentation	Assessment of UVR induced erythema by detection of changes in skin capillary area in reflectance confocal microscopy video sequences, St. Thomas' Hospital, London, UK (November, 2015)
2016	Presentation	Automated detection of skin capillaries in reflectance confocal microscopy video sequences, Joint Medial Physics Seminar, St. Thomas' Hospital, London, UK (January 2016)
2016	Meeting	Ex-vivo VivaScope® 2500 meeting organised by Mavig GmbH, Barcelona, Spain (November 2016)

Bibliography

[1] Sila Kurugol, Jennifer G Dy, Dana H Brooks, and Milind Rajadhyaksha. Pilot study of semiautomated localization of the dermal/epidermal junction in reflectance confocal microscopy images of skin. *Journal of biomedical optics*, 16(3):036005–036005, 2011.

[2] Sila Kurugol, Milind Rajadhyaksha, Jennifer G Dy, and Dana H Brooks. Validation study of automated dermal/epidermal junction localization algorithm in reflectance confocal microscopy images of skin. In *SPIE BiOS*, pages 820702–820702. International Society for Optics and Photonics, 2012.

[3] Sila Kurugol, Kivanc Kose, Brian Park, Jennifer G Dy, Dana H Brooks, and Milind Rajadhyaksha. Automated delineation of dermal–epidermal junction in reflectance confocal microscopy image stacks of human skin. *Journal of Investigative Dermatology*, 135(3):710–717, 2015.

[4] Eduardo Somoza, Gabriela Oana Cula, Catherine Correa, and Julie B Hirsch. Automatic localization of skin layers in reflectance confocal microscopy. In *Image Analysis and Recognition*, pages 141–150. Springer, 2014.

[5] Samuel C Hames, Marco Ardigò, H Peter Soyer, Andrew P Bradley, and Tarl W Prow. Automated segmentation of skin strata in reflectance confocal microscopy depth stacks. *PLOS ONE*, 11(4):e0153208, 2016.

[6] M Wiltgen, A Gerger, C Wagner, J Smolle, et al. Automatic identification of diagnostic significant regions in confocal laser scanning microscopy of melanocytic skin tumors. *Methods of Information in Medicine*, 47(1):14–25, 2008.

[7] Marco Wiltgen, Marcus Bloice, Silvia Koller, Rainer Hoffmann-Wellenhof, Josef Smolle, and Armin Gerger. Computer-aided diagnosis of melanocytic skin tumors by use of confocal laser scanning microscopy images. *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology*, 33(2):85–100, 2011.

[8] Silvia Koller, Armin Gerger, Verena Ahlgrimm-Siess, Wolfgang Weger, Josef Smolle, and Rainer Hofmann-Wellenhof. In vivo reflectance confocal microscopy of erythematosquamous skin diseases. *Experimental dermatology*, 18(6):536–540, 2009.

[9] Jessica L Dobbs, Jenna L Mueller, Savitri Krishnamurthy, Dongsuk Shin, Henry Kuerer, Wei Yang, Nirmala Ramanujam, and Rebecca Richards-Kortum. Micro-anatomical quantitative optical imaging: toward automated assessment of breast tissues. *Breast Cancer Research*, 17(1):1–14, 2015.

[10] Fisseha Bekele Legesse, Anna Medyukhina, Sandro Heuke, and Jürgen Popp. Texture analysis and classification in coherent anti-stokes raman scattering (CARS) microscopy images for automated detection of skin cancer. *Computerized Medical Imaging and Graphics*, 43:36–43, 2015.

[11] Bryan Selkin, Milind Rajadhyaksha, Salvador Gonzalez, and Richard G Langley. In vivo confocal microscopy in dermatology. *Dermatologic clinics*, 19(2):369–377, 2001.

[12] Jamshid Sourati, Kivanc Kose, Milind Rajadhyaksha, Jennifer G Dy, Deniz Erdogmus, and Dana H Brooks. Automated localization of wrinkles and the dermo-epidermal junction in obliquely oriented reflectance confocal microscopic images of human skin. In *SPIE BiOS*, pages 856503–856503. International Society for Optics and Photonics, 2013.

[13] Marek Elbaum, Alfred W Kopf, Harold S Rabinovitz, Richard GB Langley, Hideko Kamino, Martin C Mihm, Arthur J Sober, Gary L Peck, Alexandru Bogdan, Dina Gutkowicz-Krusin, et al. Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study. *Journal of the American Academy of Dermatology*, 44(2):207–218, 2001.

[14] Milind Rajadhyaksha, Melanie Grossman, Dina Esterowitz, Robert H Webb, R Rox Anderson, et al. In vivo confocal scanning laser microscopy of human skin: melanin provides strong contrast. *Journal of Investigative Dermatology*, 104(6):946–952, 1995.

[15] Rainer Hofmann-Wellenhof, Giovanni Pellacani, Joseph Malvehy, and Hans Peter Soyer. *Reflectance confocal microscopy for skin diseases*. Springer, 2012.

[16] Salvador González, Kirsty Swindells, Milind Rajadhyaksha, and Abel Torres. Changing paradigms in dermatology: confocal microscopy in clinical and surgical dermatology. *Clinics in dermatology*, 21(5):359–369, 2003.

[17] National Academy of Engineering. *Frontiers of Engineering: Reports on Leading Edge Engineering From the 1998 NAE Symposium on Frontiers of Engineering*. Compass series. National Academies Press, 1999.

[18] Salvador Gonzalez and Yolanda Gilaberte-Calzada. In vivo reflectance-mode confocal microscopy in clinical dermatology and cosmetology. *International journal of cosmetic science*, 30(1):1–17, 2008.

[19] Anca L Branzan, Michael Landthaler, and Rolf-Markus Szeimies. In vivo confocal scanning laser microscopy in dermatology. *Lasers in medical science*, 22(2):73–82, 2007.

[20] Marvin Minsky. Memoir on inventing the confocal scanning microscope. *Scanning*, 10(4):128–138, 1988.

[21] Estee L Psaty and Allan C Halpern. Current and emerging technologies in melanoma diagnosis: the state of the art. *Clinics in Dermatology*, 27(1):35–45, 2009.

[22] S Gonzalez. Confocal reflectance microscopy in dermatology: promise and reality of non-invasive diagnosis and monitoring. *Actas Dermo-Sifiliograficas*, 100:59–69, 2009.

[23] Babar K Rao, Rosa Mateus, Cindy Wassef, and Giovanni Pellacani. In vivo confocal microscopy in clinical practice: comparison of bedside diagnostic

accuracy of a trained physician and distant diagnosis of an expert reader. *Journal of the American Academy of Dermatology*, 69(6):e295–e300, 2013.

[24] Carlos A Charles, Ashfaq A Marghoob, Klaus J Busam, Lesley Clark-Loeser, and Allan C Halpern. Melanoma or pigmented basal cell carcinoma: a clinical-pathologic correlation with dermoscopy, in vivo confocal scanning laser microscopy, and routine histology. *Skin Research and Technology*, 8(4):282–287, 2002.

[25] Zeina S Tannous, Martin C Mihm, Thomas J Flotte, Salvador González, et al. In vivo examination of lentigo maligna and malignant melanoma in situ, lentigo maligna type by near-infrared reflectance confocal microscopy: comparison of in vivo confocal images with histologic sections. *American Academy of Dermatology*, 46(2):260–263, 2002.

[26] Daniel S Gareau. Feasibility of digitally stained multimodal confocal mosaics to simulate histopathology. *Journal of biomedical optics*, 14(3):034050–034050, 2009.

[27] National Institute for Clinical Excellence. Skin cancer: the vivascope 1500 and 3000 systems for detecting and monitoring skin lesions. *NICE diagnostics guidance [DG19]*, November 2015.

[28] Caterina Longo, Iris Zalaudek, Giuseppe Argenziano, and Giovanni Pellacani. New directions in dermatopathology: in vivo confocal microscopy in clinical practice. *Dermatologic clinics*, 30(4):799–814, 2012.

[29] Thomas B Fitzpatrick. The validity and practicality of sun-reactive skin types I through VI. *Archives of dermatology*, 124(6):869–871, 1988.

[30] Jane Sandby-Moller, Thomas Poulsen, and Hans Christian Wulf. Epidermal thickness at different body sites: relationship to age, gender, pigmentation, blood content, skin type and smoking habits. *Acta Dermato Venereologica*, 83(6):410–413, 2003.

[31] Jean L Bolognia, Joseph L Jorizzo, and Ronald P Rapini. *Dermatology*. Elsevier Health Sciences UK, 2007.

[32] IM Braverman. Anatomy and physiology of the cutaneous microcirculation. *Bioengineering of the skin: cutaneous blood flow and erythema*, 2:3–22, 1995.

[33] Susan M Daly and Martin J Leahy. 'Go with the flow': a review of methods and advancements in blood flow imaging. *Journal of biophotonics*, 6(3):217–255, 2013.

[34] Enzo Berardesca, J-L Lévêque, and Philippe Masson. EEMCO guidance for the measurement of skin microcirculation. *Skin Pharmacology and Physiology*, 15(6):442–456, 2002.

[35] Bengt Fagrell. Microcirculation of the skin. *The physiology and pharmacology of the microcirculation*, 2:133–180, 1984.

[36] Alexander Birbrair, Tan Zhang, Zhong-Min Wang, Maria Laura Messi, John D Olson, Akiva Mintz, and Osvaldo Delbono. Type-2 pericytes participate in normal and tumoral angiogenesis. *American Journal of Physiology-Cell Physiology*, 307(1):C25–C38, 2014.

[37] Carolyn J Heckman, Rachel Chandler, Jacqueline D Kloss, Amy Benson, Deborah Rooney, Teja Munshi, Susan D Darlow, Clifford Perlis, Sharon L Manne, and David W Oslin. Minimal erythema dose (MED) testing. *JoVE (Journal of Visualized Experiments)*, (75):e50175–e50175, 2013.

[38] Barbara B Shih, Donald Allan, Frank R de Gruijl, and Lesley E Rhodes. Robust detection of minimal sunburn in pigmented skin by 785 nm laser speckle contrast imaging of blood flux. *The Journal of investigative dermatology*, 135(4):1197, 2015.

[39] Constantin Căruntu and Daniel Boda. Evaluation through in vivo reflectance confocal microscopy of the cutaneous neurogenic inflammatory reaction induced by capsaicin in human subjects. *Journal of biomedical optics*, 17(8):0850031–0850037, 2012.

[40] Toyonobu Yamashita, Hirotaka Akita, Susanne Astner, Michio Miyakawa, Ethan Arthur Lerner, and Salvador González. In vivo assessment of

pigmentary and vascular compartments changes in uva exposed skin by reflectance-mode confocal microscopy. *Experimental dermatology*, 16(11):905–911, 2007.

[41] T Gambichler, K Sauermann, MA Altintas, V Paech, A Kreuter, P Altmeyer, and K Hoffmann. Effects of repeated sunbed exposures on the human skin. in vivo measurements with confocal microscopy. *Photodermatology, photoimmunology & photomedicine*, 20(1):27–32, 2004.

[42] Pierre Corcuff, Céline Chaussepied, Geneviève Madry, and Christophe Hadjur. Skin optics revisited by in vivo confocal microscopy: melanin and sun exposure. *Journal of cosmetic science*, 52(2):91–102, 2001.

[43] Laura Dean. *Blood group antigens are surface markers on the red blood cell membrane*. National Center for Biotechnology Information (US), 2005.

[44] J Liam McWhirter, Hiroshi Noguchi, and Gerhard Gompper. Flow-induced clustering and alignment of vesicles and red blood cells in microcapillaries. *Proceedings of the National Academy of Sciences*, 106(15):6039–6043, 2009.

[45] E Cinotti, L Gergelé, JL Perrot, A Dominé, B Labeille, P Borelli, and F Cambazard. Quantification of capillary blood cell flow using reflectance confocal microscopy. *Skin Research and Technology*, 20(3):373–378, 2014.

[46] Mary Louise Turgeon. *Clinical hematology: theory and procedures*. Lippincott Williams & Wilkins, 2005.

[47] Christine E McLaren, Gary M Brittenham, and Victor Hasselblad. Statistical and graphical evaluation of erythrocyte volume distributions. *American Journal of Physiology-Heart and Circulatory Physiology*, 252(4):H857–H866, 1987.

[48] Prosenjit Bagchi. Mesoscale simulation of blood flow in small vessels. *Biophysical journal*, 92(6):1858–1877, 2007.

[49] Angela C Shore. Capillaroscopy and the measurement of capillary pressure. *British journal of clinical pharmacology*, 50(6):501–513, 2000.

[50] Rami Archid, Alexa Patzelt, Bernhard Lange-Asschenfeldt, Sufian S Ahmad, Martina Ulrich, Eggert Stockfleth, Sandra Philipp, Wolfram Sterry, and Juergen Lademann. Confocal laser-scanning microscopy of capillaries in normal and psoriatic skin. *Journal of biomedical optics*, 17(10):1015111–1015115, 2012.

[51] J Hegyi, V Hegyi, G Messer, P Arenberger, T Ruzicka, and C Berking. Confocal laser-scanning capillaroscopy: a novel approach to the analysis of skin capillaries in vivo. *Skin Research and Technology*, 15(4):476–481, 2009.

[52] Michael Berks, Phil Tresadern, Graham Dinsdale, Andrea Murray, Tonia Moore, Ariane Herrick, and Chris Taylor. An automated system for detecting and measuring nailfold capillaries. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 658–665. Springer, 2014.

[53] Maurizio Cutolo, Alberto Sulli, and Vanessa Smith. How to perform and interpret capillaroscopy. *Best practice & research Clinical rheumatology*, 27(2):237–248, 2013.

[54] G Moneta, M Brülisauer, K Jäger, and A Bollinger. Infrared fluorescence videomicroscopy of skin capillaries with indocyanine green. *International journal of microcirculation, clinical and experimental/sponsored by the European Society for Microcirculation*, 6(1):25–34, 1986.

[55] Mikael Norman, Peter Herin, and Bengt Fagrell. An evaluation of skin capillary blood flow determinations in neonates using a computerized videophotometric method. *Microvascular research*, 43(3):276–284, 1992.

[56] Alon Scope, Stephen W Dusza, Allan C Halpern, Harold Rabinovitz, Ralph P Braun, Iris Zalaudek, Giuseppe Argenziano, and Ashfaq A Marghoob. The ugly duckling sign: agreement between observers. *Archives of dermatology*, 144(1):58–64, 2008.

[57] Robert J Friedman, Darrell S Rigel, and Alfred W Kopf. Early detection of malignant melanoma: The role of physician examination and self-

examination of the skin. *CA: a cancer journal for clinicians*, 35(3):130–151, 1985.

[58] Franz Nachbar, Wilhelm Stolz, Tanja Merkle, Armand B Cognetta, Thomas Vogt, Michael Landthaler, Peter Bilek, Otto Braun-Falco, and Gerd Plewig. 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, 1994.

[59] Hubert Pehamberger, Michael Binder, Andreas Steiner, and Klaus Wolff. In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. *Journal of investigative dermatology*, 100:356S–362S, 1993.

[60] Cliff Rosendahl, Alan Cameron, Ian McColl, and David Wilkinson. Dermatoscopy in routine practice:'chaos and clues'. *Australian family physician*, 41(7):482, 2012.

[61] A Gola Isasi, B García Zapirain, and A Méndez Zorrilla. Melanomas non-invasive diagnosis application based on the abcd rule and pattern recognition image processing algorithms. *Computers in Biology and Medicine*, 41(9):742–755, 2011.

[62] Marie-Lise Bafounta, Alain Beauchet, Philippe Aegerter, and Philippe Saiag. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma?: Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. *Archives of dermatology*, 137(10):1343, 2001.

[63] Giuseppe Argenziano, Lorenzo Cerroni, Iris Zalaudek, Stefania Staibano, Rainer Hofmann-Wellenhof, Nicola Arpaia, Renato Marchiori Bakos, Brigitte Balme, Jadran Bandic, Roberto Bandelloni, et al. Accuracy in melanoma detection: a 10-year multicenter survey. *Journal of the American Academy of Dermatology*, 67(1):54–59, 2012.

[64] Julia Welzel. Optical coherence tomography in dermatology: a review. *Skin Research and Technology*, 7(1):1–9, 2001.

[65] Milind Rajadhyaksha, Salvador González, James M Zavislan, R Rox Anderson, and Robert H Webb. In vivo confocal scanning laser microscopy of

human skin II: Advances in instrumentation and comparison with histology. *Journal of Investigative Dermatology*, 113(3):293–303, 1999.

[66] Joseph A Izatt, Manish D Kulkarni, Hsing-Wen Wang, Kenji Kobayashi, and Michael V Sivak Jr. Optical coherence tomography and microscopy in gastrointestinal tissues. *Selected Topics in Quantum Electronics, IEEE Journal of*, 2(4):1017–1028, 1996.

[67] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.

[68] Cindy Wassef and Babar K Rao. Uses of non-invasive imaging in the diagnosis of skin cancer: an overview of the currently available modalities. *International journal of dermatology*, 52(12):1481–1489, 2013.

[69] Yang Zhou, Daiqiang Yin, Ping Xue, Naiyan Huang, Haixia Qiu, Ying Wang, Jing Zeng, Zhihua Ding, and Ying Gu. Imaging of skin microvessels with optical coherence tomography: potential uses in port wine stains. *Experimental and therapeutic medicine*, 4(6):1017–1021, 2012.

[70] Rainer A Leitgeb, René M Werkmeister, Cedric Blatter, and Leopold Schmetterer. Doppler optical coherence tomography. *Progress in retinal and eye research*, 41:26–43, 2014.

[71] BR Masters, PTC So, and E Gratton. Optical biopsy of in vivo human skin: multi-photon excitation microscopy. *Lasers in medical science*, 13(3):196–203, 1998.

[72] Barry R Masters and Peter TC So. Multi-photon excitation microscopy and confocal microscopy imaging of in vivo human skin: a comparison. *Microscopy and Microanalysis*, 5(04):282–289, 1999.

[73] Ulrike Tauer. Advantages and risks of multiphoton microscopy in physiology. *Experimental physiology*, 87(06):709–714, 2002.

[74] Ilona Kuzmina, Ilze Diebele, Dainis Jakovels, Janis Spigulis, Lauma Valeine, Janis Kapostinsh, and Anna Berzina. Towards noncontact skin melanoma selection by multispectral imaging analysis. *Journal of Biomedical optics*, 16(6):060502–060502, 2011.

[75] Janis Spigulis, Dainis Jakovels, and Uldis Rubins. Multi-spectral skin imaging by a consumer photo-camera. In *BiOS*, pages 75570M–75570M. International Society for Optics and Photonics, 2010.

[76] Peter Naglic. Raman spectroscopy for medical diagnostics. *University of Ljubljana*, 2012.

[77] Kenny Kong, Catherine Kendall, Nicholas Stone, and Ioan Notingher. Raman spectroscopy for medical diagnostics - from in-vitro biofluid assays to in-vivo cancer detection. *Advanced drug delivery reviews*, 89:121–134, 2015.

[78] Catherine Kendall, Martin Isabelle, Florian Bazant-Hegemark, Joanne Hutchings, Linda Orr, Jaspreet Babrah, Rebecca Baker, and Nicholas Stone. Vibrational spectroscopy: a clinical tool for cancer diagnostics. *Analyst*, 134(6):1029–1045, 2009.

[79] Conor L Evans, Eric O Potma, Mehron Puoris' haag, Daniel Côté, Charles P Lin, and X Sunney Xie. Chemical imaging of tissue in vivo with video-rate coherent anti-Stokes Raman scattering microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16807–16812, 2005.

[80] Dane M Drutis, Thomas M Hancewicz, Eugene Pashkovski, Lin Feng, Dawn Mihalov, Gary Holtom, Kavssery P Ananthapadmanabhan, X Sunney Xie, and Manoj Misra. Three-dimensional chemical imaging of skin using stimulated raman scattering microscopy. *Journal of biomedical optics*, 19(11):111604–111604, 2014.

[81] Chetan A Patil, Harish Kirshnamoorthi, Darrel L Ellis, Ton G van Leeuwen, and Anita Mahadevan-Jansen. A clinical instrument for combined raman spectroscopy-optical coherence tomography of skin cancers. *Lasers in surgery and medicine*, 43(2):143–151, 2011.

[82] Guillaume Mahé, Anne Humeau-Heurtier, Sylvain Durand, Georges Leftheriotis, and Pierre Abraham. Assessment of skin microvascular function and dysfunction with laser speckle contrast imaging. *Circulation: Cardiovascular Imaging*, 5(1):155–163, 2012.

[83] Mitchell A Davis, SM Shams Kazmi, and Andrew K Dunn. Imaging depth and multiple scattering in laser speckle contrast imaging. *Journal of biomedical optics*, 19(8):086001–086001, 2014.

[84] ID Swain and LJ Grant. Methods of measuring skin blood flow. *Physics in Medicine and Biology*, (34):151–75, 1989.

[85] Susan K Parsons, Jeffery A Chan, W Yu Winifred, Ndidiamaka Obadan, Sara J Ratichek, Jounghee Lee, Srila Sen, and Stanley Ip. Noninvasive diagnostic techniques for the detection of skin cancers. Technical Brief No. 11. (Prepared by the Tufts University Evidence-based Practice Center under Contract No. 290-2007-1055-1.) AHRQ Publication No. 11-EHC085-EF. *Agency for Healthcare Research and Quality (US)*, 2011.

[86] Vinayakrishnan Rajan, Babu Varghese, Ton G van Leeuwen, and Wiendelt Steenbergen. Review of methodological developments in laser doppler flowmetry. *Lasers in medical science*, 24(2):269–283, 2009.

[87] S Del Bino, J Sok, E Bessac, and F Bernerd. Relationship between skin response to ultraviolet exposure and skin color type. *Pigment cell research*, 19(6):606–614, 2006.

[88] S Del Bino and F Bernerd. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *British Journal of Dermatology*, 169(s3):33–40, 2013.

[89] Vincent Chan and Anahi Perlas. Basics of ultrasound imaging. In *Atlas of Ultrasound-Guided Procedures in Interventional Pain Management*, pages 13–19. Springer, 2011.

[90] Alan JR Macfarlane, CH Cyrus, and Richard Brull. Essential knobology for ultrasound-guided regional anesthesia and interventional pain manage-

ment. In *Atlas of Ultrasound-Guided Procedures in Interventional Pain Management*, pages 21–33. Springer, 2011.

[91] AN Obeid, DM Boggett, NJ Barnett, G Dougherty, and P Rolfe. Depth discrimination in laser doppler skin blood flow measurement using different lasers. *Medical and Biological Engineering and Computing*, 26(4):415–424, 1988.

[92] Giovanni Pellacani, Pascale Guitera, Caterina Longo, Michelle Avramidis, Stefania Seidenari, and Scott Menzies. The impact of in vivo reflectance confocal microscopy for the diagnostic accuracy of melanoma and equivocal melanocytic lesions. *Journal of Investigative Dermatology*, 127(12):2759–2765, 2007.

[93] Caterina Longo, Alice Casari, Barbara Pace, Silvia Simonazzi, Giovanna Mazzaglia, and Giovanni Pellacani. Proposal for an in vivo histopathologic scoring system for skin aging by means of confocal microscopy. *Skin Research and Technology*, 19(1):e167–e173, 2013.

[94] Jamshid Sourati, Dana H Brooks, Jennifer G Dy, E Ataer-Cansizoglu, Deniz Erdogmus, and Milind Rajadhyaksha. Unsupervised wrinkle detection in reflectance confocal microscopy images of the human skin. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 705–708. IEEE, 2012.

[95] Dan Gareau, Ricky Hennessy, Eric Wan, Giovanni Pellacani, and Steven L Jacques. Automated detection of malignant features in confocal microscopy on superficial spreading melanoma versus nevi. *Journal of biomedical optics*, 15(6):061713–061713, 2010.

[96] Dan Gareau. Automated identification of epidermal keratinocytes in reflectance confocal microscopy. *Journal of biomedical optics*, 16(3):030502–030502, 2011.

[97] Anthony P Raphael, Timothy A Kelf, Elizabeth MT Wurm, Andrei V Zvyagin, Hans Peter Soyer, and Tarl W Prow. Computational characterization

of reflectance confocal microscopy features reveals potential for automated photoageing assessment. *Experimental dermatology*, 22(7):458–463, 2013.

[98] C Longo, M Rajadhyaksha, M Ragazzi, K Nehal, S Gardini, E Moscarella, A Lallas, I Zalaudek, S Piana, G Argenziano, et al. Evaluating ex vivo fluorescence confocal microscopy images of basal cell carcinomas in Mohs excised tissue. *British Journal of Dermatology*, 171(3):561–570, 2014.

[99] JK Karen, DS Gareau, SW Dusza, M Tudisco, M Rajadhyaksha, and KS Nehal. Detection of basal cell carcinomas in Mohs excisions with fluorescence confocal mosaicing microscopy. *British Journal of Dermatology*, 160(6):1242–1250, 2009.

[100] A Bennàssar, A Vilata, S Puig, and J Malvehy. Ex vivo fluorescence confocal microscopy for fast evaluation of tumour margins during mohs surgery. *British Journal of Dermatology*, 170(2):360–365, 2014.

[101] Arif Aslam, Anand Patel, and Sandeep Varma. What I tell my patients about Mohs micrographic surgery. *Dermatology in Practice*, 21(2):40–42, 2015.

[102] Neil A Swanson. Mohs surgery: technique, indications, applications, and the future. *Archives of dermatology*, 119(9):761–773, 1983.

[103] Vinh Q Chung, Peter J Dwyer, Kishwer S Nehal, Milind Rajadhyaksha, Gregg M Menaker, Carlos Charles, and S Brian Jiang. Use of ex vivo confocal scanning laser microscopy during Mohs surgery for nonmelanoma skin cancers. *Dermatologic surgery*, 30(12p1):1470–1478, 2004.

[104] Daniela Hartmann, Cristel Ruini, Leonie Mathemeier, Andreas Dietrich, Thomas Ruzicka, and Tanja von Braunmühl. Identification of ex-vivo confocal scanning microscopic features and their histological correlates in human skin. *Journal of Biophotonics*, 9(4):376–387, 2015.

[105] Antoni Bennàssar, Cristina Carrera, Susana Puig, Antoni Vilalta, and Josep Malvehy. Fast evaluation of 69 basal cell carcinomas with ex vivo fluorescence confocal microscopy: criteria description, histopathological cor-

relation, and interobserver agreement. *JAMA dermatology*, 149(7):839–847, 2013.

[106] Daniel S Gareau, Yogesh G Patel, Yongbiao Li, Kishwer S Nehal, Billy Huang, and Milind Rajadhyaksha. Multimodal confocal mosaicing of basal cell carcinomas in Mohs surgical skin excisions. In *Biomedical Optics (BiOS) 2007*, pages 64310U–64310U. International Society for Optics and Photonics, 2007.

[107] Mehmet Ali Altintas, AA Altintas, M Guggenheim, AE Steiert, MC Aust, AD Niederbichler, Christian Herold, and PM Vogt. Insight in human skin microcirculation using in vivo reflectance-mode confocal laser scanning microscopy. *Journal of digital imaging*, 23(4):475–481, 2010.

[108] Thilo Gambichler, Julia Huyn, Nordwig S Tomi, Georg Moussa, Cornelia Moll, Anna Sommer, Peter Altmeyer, and Klaus Hoffmann. A comparative pilot study on ultraviolet-induced skin changes assessed by noninvasive imaging techniques in vivo. *Photochemistry and photobiology*, 82(4):1103–1107, 2006.

[109] Ahmet Ali Altintas, Tilman Stasch, Arzu Oezcelik, and Mehmet Ali Altintas. In vivo comparison of microcirculation and histomorphology of two different flap donor sites. *Microscopy research and technique*, 74(4):308–313, 2011.

[110] Marina Venturini, Mariachiara Arisi, Arianna Zanca, Ilaria Cavazzana, Salvador Gonzàlez, Franco Franceschini, and Piergiacomo Calzavara-Pinton. In vivo reflectance confocal microscopy features of cutaneous microcirculation and epidermal and dermal changes in diffuse systemic sclerosis and correlation with histological and videocapillaroscopic findings. *European Journal of Dermatology*, 24(3):349–355, 2014.

[111] M Stücker, V Baier, T Reuther, K Hoffmann, K Kellam, and P Altmeyer. Capillary blood cell velocity in human skin capillaries located perpendicularly to the skin surface: measured by a new laser doppler anemometer. *Microvascular research*, 52(2):188–192, 1996.

[112] Joel Radford. Vivascope 1500 operations guide. *Caliber ID*, 2016.

[113] Joel Radford. Vivascope 2500 operations guide. *Caliber ID*, 2016.

[114] Catherine E Scarff. Reflectance confocal microscopy of cutaneous tumors: an atlas with clinical, dermoscopic and histological correlations. *Australasian Journal of Dermatology*, 50(2):152–152, 2009.

[115] Piergiacomo Calzavara-Pinton, Caterina Longo, Marina Venturini, Raffaella Sala, and Giovanni Pellacani. Reflectance confocal microscopy for in vivo skin imaging. *Photochemistry and photobiology*, 84(6):1421–1430, 2008.

[116] Klaus J Busam, Carlos Charles, Grace Lee, and Allan C Halpern. Morphologic features of melanocytes, pigmented keratinocytes, and melanophages by in vivo confocal scanning laser microscopy. *Modern Pathology*, 14(9):862–868, 2001.

[117] Heidy Sierra, Shadi Damanpour, Brian Hibler, Kishwer Nehal, Anthony Rossi, and Milind Rajadhyaksha. Confocal imaging of carbon dioxide laser-ablated basal cell carcinomas: An ex-vivo study on the uptake of contrast agent and ablation parameters. *Lasers in surgery and medicine*, 48(2):133–139, 2015.

[118] Michel Misiti, Yves Misiti, Georges Oppenheim, and Jean-Michel Poggi. Matlab wavelet toolbox user's guide. Version 3. 2004.

[119] Michael Felsberg and Gerald Sommer. A new extension of linear signal processing for estimating local properties and detecting features. In *Mustererkennung 2000*, pages 195–202. Springer, 2000.

[120] Michael Felsberg and Gerald Sommer. The monogenic signal. *Signal Processing, IEEE Transactions on*, 49(12):3136–3144, 2001.

[121] Peter Kovesi. Edges are not just steps. In *Proceedings of the Fifth Asian Conference on Computer Vision*, pages 822–827, 2002.

[122] Sylvia Rueda, Caroline L Knight, Aris T Papageorghiou, and J Alison Noble. Oriented feature-based coupled ellipse fitting for soft tissue quantifi-

cation in ultrasound images. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1014–1017. IEEE, 2013.

[123] Kashif Rajpoot, Vicente Grau, and J Alison Noble. Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3D echocardiography images. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 783–786. IEEE, 2009.

[124] Tao Peng, Arvind Balijepalli, Satyandra K Gupta, and Thomas W LeBrun. Algorithms for on-line monitoring of components in an optical tweezers-based assembly cell. In *ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 297–309. American Society of Mechanical Engineers, 2006.

[125] Ian H Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[126] Mark A Hall. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

[127] Brian D Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2008.

[128] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. Wiley-interscience, 2012.

[129] Seungyong Lee, George Wolberg, and Sung Yong Shin. Scattered data interpolation with multilevel B-splines. *Visualization and Computer Graphics, IEEE Transactions on*, 3(3):228–244, 1997.

[130] R Cuingnet, O Somphone, B Mory, R Prevost, M Yaqub, R Napolitano, A Papageorghiou, D Roundhill, JA Noble, and R Ardon. Where is my baby? A fast fetal head auto-alignment in 3D-ultrasound. pages 768–771, 2013.

[131] Matthew R Boutell, Jiebo Luo, Xipeng Shen, and Christopher M Brown. Learning multi-label scene classification. *Pattern recognition*, 37(9):1757–1771, 2004.

[132] Nikos Komodakis and Georgios Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Transactions on Image Processing*, 16(11):2649–2661, 2007.

[133] Haitao Xiong, Junjie Wu, and Lu Liu. Classification with class overlapping: A systematic study. In *The 2010 International Conference on E-Business Intelligence*, pages 491–497, 2010.

[134] Daniel S Gareau, Julie K Karen, Stephen W Dusza, Marie Tudisco, Kishwer S Nehal, and Milind Rajadhyaksha. Sensitivity and specificity for detecting basal cell carcinomas in Mohs excisions with confocal fluorescence mosaicing microscopy. *Journal of biomedical optics*, 14(3):034012–034012, 2009.

[135] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):898–916, 2011.

[136] Rolf Adams and Leanne Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6):641–647, 1994.

[137] David Nistér and Henrik Stewénius. Linear time maximally stable extremal regions. In *Computer Vision–ECCV 2008*, pages 183–196. Springer, 2008.

[138] Ron Kimmel, Cuiping Zhang, Alexander Bronstein, and Michael M Bronstein. Are MSER features really interesting? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(11):2316–2320, 2011.

[139] Kyle Lesack and Christopher Naugler. Morphometric characteristics of basal cell carcinoma peritumoral stroma varies among basal cell carcinoma subtypes. *BMC dermatology*, 12(1):1, 2012.

[140] Jerome Fan, Suneel Upadhye, and Andrew Worster. Understanding receiver operating characteristic (roc) curves. *Cjem*, 8(01):19–20, 2006.

[141] Linda J Warren Burhenne, Susan A Wood, Carl J D'Orsi, Stephen A Feig, Daniel B Kopans, Kathryn F O'Shaughnessy, Edward A Sickles, Laszlo

Tabar, Carl J Vyborny, and Ronald A Castellino. Potential contribution of computer-aided detection to the sensitivity of screening mammography 1. *Radiology*, 215(2):554–562, 2000.

[142] Julia A Schnabel, Daniel Rueckert, Marcel Quist, Jane M Blackall, Andy D Castellano-Smith, Thomas Hartkens, Graeme P Penney, Walter A Hall, Haiying Liu, Charles L Truwit, et al. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*, pages 573–581. Springer, 2001.

[143] Daniel Rueckert, Carmel Hayes, Colin Studholme, Paul Summers, M Leach, and David J Hawkes. Non-rigid registration of breast MR images using mutual information. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1144–1152. Springer, 1998.

[144] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990.

[145] Guido Gerig, Olaf Kübler, Ron Kikinis, Ferenc Jolesz, et al. Nonlinear anisotropic filtering of MRI data. *Medical Imaging, IEEE Transactions on*, 11(2):221–232, 1992.

[146] Kirsten Christensen-Jeffries, Richard J Browning, Meng-Xing Tang, Christopher Dunsby, and Robert J Eckersley. In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles. *Medical Imaging, IEEE Transactions on*, 34(2):433–440, 2015.

[147] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 2001.

[148] Leonard W Winchester and Nee-Yin Chou. Measurement of sublingual blood velocity as a tool for monitoring sepsis. In *Conference proceedings:*

Annual International Conference of the IEEE Engineering in Medicine and Biology Society., volume 2008, pages 3739–3742. NIH Public Access, 2008.