

# **NON-INVASIVE CANCER CHARACTERISATION USING AUTOFLUORESCENCE IMAGING**

**Abbas Habibalahi**

Under supervision of Prof. Subhas Mukhopadhyay

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

School of Engineering  
Macquarie University, Sydney

2019



# Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signature:      Abbas Habibalahi

Date:            9/01/2019

# ACKNOWLEDGMENT

After a PhD period, today is one of my best days, as I am writing this appreciation, which finalises the writing of my thesis. This intense period of learning has improved my personal and scientific abilities to the extent that I couldn't imagine at the very beginning of my PhD. This period has had a great impact on me and my life, which was not possible without the people who kindly have supported me.

First and foremost, I wish to thank my advisors, Professor Subhas Mukhopadhyay and Professor Ewa Goldys, for giving me moral support and also intellectual freedom over my PhD studies. They showed me how to make and develop new ideas. Being under their supervision greatly improved my understanding of professionalism. Their guidance and patience over this period are really appreciated. Undoubtedly, they are and will be the most influential individuals in my professional life.

I would like to express my sincere gratitude to my co-advisor, clinical A/Prof. Chandra Bala, an eye surgeon who really admires research, for all his scientific contributions and financial support to make this PhD experience productive. Although having a hectic schedule, he always tried to squeeze his time to support me.

I would like to extend my special thanks to Dr. Alexandra Allende for her great contribution to this thesis as a pathologist. This project could not have progressed successfully without her substantial collaboration.

My deepest gratitude to Professor Helen Rizos who was the melanoma cancer specialist in this study for all of her time and guidance. To Ayad Anwer, Martin Gosnel, Bernadette Pedersen for their technical support.

I would like to thank people in our group who became my close friends rapidly and made my good times much better and my hard times much easier over my studies at MACUARIE. I am extremely grateful to the centre for nanoscale biophotonics and also MACUARIE university, for providing me with the funding sources to pursue my graduate studies.

Finally, I would like to express my very special appreciation to my family for all the love and support they gave me: first and foremost to my great parents, Maryam and Hassan, who backed me up in every step of my life; to my brother and sister, Habib and Masoumeh, for being so encouraging and inspiring.

Abbas Habibalahi 9/10/2019

# List of publications

## Manuscripts, revised papers for Peer review journals

1. **Abbas Habibalahi**, Chandra Bala , Alexandra Allende, Ayad G. Anwer ,and Ewa M.Goldys, “Novel Automated Non-invasive Detection of Ocular Surface Squamous Neoplasia Using Multispectral Autofluorescence Imaging”, published online, ‘Ocular Surface’, (2019) Medical journal Impact factor: 5.53
2. Jared M. Campbell, **Abbas Habibalahi**, Saabah Mahbub, Martin Gosnell, Sharon Paton, Stan Gronthos, Ewa Goldys, “Multispectral Microscopy for the Non-destructive, Label-free Identification of Cell Cycle Phase”, to be submitted to ‘Molecular Cancer, Impact factor: 7.5, submission date 15/7/2019
3. **Abbas Habibalahi**, Chandra Bala , Subhas Mukhopadhyay, Alexandra Allende, Ayad G. Anwer ,and Ewa M.Goldys, “ Optimising multi spectral imaging to detect ocular surface squamous neoplasia using swarm intelligence”, final draft is circulated to the co-authors for final confirmation and to submit to “ Investigative Ophthalmology & Visual Science”
4. **Abbas Habibalahi**, Helen Rizo , Subhas Mukhopadhyay, Bernadette Pedersen, Ayad G. Anwer,and Ewa M.Goldys, “Discrimination of Melanoma and Fibroblast by Multispectral Autofluorescence Imaging”, To be submitted to “ Journal of Biophotonics”, Impact factor:3.76

## Conferences

5. Ewa M. Goldys, **Abbas Habibalahi**, Jared Campbell, Ayad Anwer, Martin Gosnell, Saabah Mahbub, “ Label-free multispectral imaging for biological research and medical diagnostics”, SPIE photonics west 2019, 5-7 Feb 2019, Moscone Center, San Francisco, USA

6. **Abbas Habibalahi**, Chandra Bala , Ayad G. Anwer ,and Ewa M.Goldys  
“Autofluorescence application in cancer characterisation ”, CNBP  
Conference 2018, 4-7 Dec 2018, Melbourne, Australia
  
7. **Abbas Habibalahi**, Chandra Bala , Ayad G. Anwer ,and Ewa M.Goldys  
“Hyperspectral technology applied to detect eye cancer”, CNBP Conference  
2018, CNBP conference 2017, 28 Nov-1 Dec 2017, Melbourne, Australia
  
8. **Abbas Habibalahi**, Ayad G. Anwer ,and Ewa M.Goldys, “Hyperspectral  
imaging used to distinguish various cell types”, SPIE Biophotonics  
Australasia, 16- 19 October 2016, Adelaide Convention Center, Adelaide,  
Australia

# Keywords

Autofluorescence, characterisation, cancer, early detection, Porphyrins, Nicotinamide adenine dinucleotide, Flavin, fluorescence lifetime imaging, excitation /emission wavelength, Ocular surface squamous neoplasia, intra- and inter-patient classification, machine learning, intraoperative assessment, swarm intelligence, particle swarm intelligence, differential evolution, ant colony optimisation, melanoma, fibroblast, single cell resolution.



# Abstract

Autofluorescence imaging plays a special role in cancer detection, as it is capable of recognising aspects of the chemical composition of tissue based on spectral signatures of naturally fluorescent compounds. Various components naturally present in cells and tissue have auto-fluorescence properties, including Porphyrins (PPIX), Nicotinamide adenine dinucleotide (NADH), and Flavins, whose contents are modified in cancer due to a transformation in cell metabolism. In particular, a lack of iron or ferrochelatase in tumours results in a change of PPIX concentration relative to the normal host tissue. Thus, quantification of these components and studying their variation provides valuable insights into the diagnosis and characterisation of cancer cells and tissue.

Conventionally, auto-fluorescence imaging technology has been limited to a few costly channels ( $n < 4$ ) employed in some microscope methodologies such as fluorescence lifetime imaging (FLIM). Consequently, these technologies can monitor only a limited number of fluorophores. However, in this study, a newly designed spectral imaging microscope that employs tens of different channels ( $n = 38$ ) was used. This system uses light excitation with a number of narrowband ranges of wavelength and collects the native fluorescence emission of the sample at specific wavelengths. A combination of excitation /emission wavelength bands forms a spectral channel, and a number of such channels ( $n = 38$ ) were used in this work. The sample is imaged in each of these channels, to acquire separate spectral images. This represents an advance over traditional auto-fluorescence imaging systems. The availability of multiple channels makes it possible to survey the overall biochemical composition of the tissue, in addition to detecting specific markers to identify the tissue state.

First, the newly-designed non-invasive auto-fluorescence multispectral imaging methodology was applied to detect Ocular surface squamous neoplasia (OSSN), with a view to a future clinical ophthalmological application. The aim was to distinguish between normal and neoplastic tissue in fixed human samples and sophisticated data analysis was applied to meticulously extract the spectral signature. Two different classification frameworks were deployed, namely intra- and inter-patient classification, to consider aspects of patients' variability and quantify the spectral

signature of OSSN. Using machine learning methods, an approach was also introduced for objective assessment of boundary detection. This technique creates a false colour map which can be rapidly generated in quasi-real time and used for intraoperative assessment. The neoplastic boundaries predicted by employing machine learning methods were validated and assessed by an anatomical pathologist. The approach introduced in this study has the potential to reduce the incidence of eye biopsies, prevent therapy delays and make treatment more effective.

Using such cutting-edge technology in auto-fluorescence imaging led us to employ a number of channels simultaneously to diagnose or monitor diseases with high accuracy. However, generating a large dataset based on tens of spectral images may also increase the possibility of having irrelevant channels which carry very little discriminatory information for a specific diagnostic application. Consequently, such multispectral imaging needs to be optimised in terms of the number of channels. Different known and unknown factors may influence the usefulness of the channels for a specific application, which cannot be determined by prejudgment. Hence, the best way for channel selection is to employ all of the channels for the detection and then determine which ones are the most relevant channels. In this study, an advanced methodology using a combination of swarm intelligence and cluster analysis was developed to discover rich and informative spectral channels for differentiating normal and diseased (OSSN) tissue. First, discrimination analysis was applied to find normal and diseased clusters and then a criterion function was defined to minimise the within-cluster variance while maximising the between-cluster variance. Such a criterion function was optimised using three different swarm intelligence methodologies including particle swarm intelligence (PSO), differential evolution (DE) and ant colony optimisation (ACO). Finally, depending on the required accuracy and criticality of the application, the richest subsets with a few channels were proposed (5 channels). Optimising the number of channels resulted in more efficient instrumentation in terms of equipment (5 out of 38 channels), acquisition time (80% less acquisition time) and computation complexity.

Moreover, the metabolic heterogeneity of melanoma cancer cells and fibroblast were considered in this study. Multi-spectral auto-fluorescence imaging was employed for evaluation of melanoma cells and fibroblast in order to produce discriminative information. Unlike typical auto-fluorescence imaging techniques that consider only a

few features, a variety of biologically relevant quantitative information was extracted from spectral images. Such a powerful analysis helped capture different aspects of the spectrum in a single cell resolution. Different features including intensity, first order and second order features, textural features, and different statistical measures of pixel values were quantitatively analysed. After selection of the most indicative features, a discriminative analysis was undertaken to distinguish melanoma cells from fibroblast efficiently. This was then followed by an examination of melanoma cells derived from a patient under treatment. Using unsupervised data processing, spectral features from the channels were de-correlated based on the principal component analysis (PCA). Then the data were quantitatively assessed using hierarchical clustering. Consequently, this study also successfully demonstrates ( $AUC > 0.9$ ) the possibility of obtaining information about melanoma cells and their environment to monitor their behaviour and discriminate them from normal skin cell types. Such a methodology may open a new way for cell therapy, regenerative medicine, personalized immunotherapy and cancer treatment.

# Table of Contents

Statement of Original Authorship .....	i
ACKNOWLEDGMENT .....	ii
List of publications .....	iv
Keywords .....	vi
Abstract .....	vii
Table of Contents .....	x
List of Figures .....	xiii
List of Tables.....	xvi
<b>Chapter 1: THESIS OVERVIEW .....</b>	<b>1</b>
1.1 Background .....	1
1.2 Statement of the problem .....	2
1.3 Significance AND novElty .....	3
1.4 Thesis Outline .....	4
<b>Chapter 2: Autofluorescence in cancer characterisation .....</b>	<b>7</b>
2.1 Fluorescence principles.....	7
2.2 Quantum Yield.....	10
2.3 Factors affecting autofluorEscence properties .....	11
2.4 Labelled and label-free detection by fluorescence .....	11
2.5 NATIVE fluorophores .....	12
2.6 Native fluorophores used as cancer biomarker .....	13
2.7 NADH AND FLAVINS.....	16
2.8 NAD(P)H and FAD in Cellular Metabolism.....	18
2.9 Porphyrins .....	21
2.10 Screening.....	24
2.11 Conclusion .....	29
<b>Chapter 3: Research Design.....</b>	<b>31</b>
3.1 Introduction .....	31
3.2 CustomiSed flUorescence microscope.....	32
3.3 Excitation Sources.....	33
3.4 Emission Bands.....	35
3.5 Camera .....	37
3.6 System control.....	39
3.7 Dishes and slides for imaging .....	40
3.8 Imaging protocols .....	41

3.9	Image preparation and processing .....	42
3.10	Calibrating the images .....	43
3.11	Removing the background.....	44
3.12	Image equalization.....	44
3.13	Image noises .....	45
3.14	Cell and tissue segmentation .....	48
3.15	Data analysis .....	48
3.16	Ethics .....	49
<b>Chapter 4:</b>	<b>OSSN DETECTION AND DELINEATION .....</b>	<b>51</b>
4.1	Introduction .....	52
4.2	Methods .....	54
4.3	Data analysis.....	56
4.4	Results .....	60
4.5	Discussion.....	69
4.6	conclusion.....	72
4.7	Author contribution .....	72
4.8	Supplementary Notes associated with Method section.....	74
4.9	Supplementary Notes associated to Result section.....	80
<b>Chapter 5:</b>	<b>CHANNEL SELECTION AND SYSTEM OPTIMISATION ...</b>	<b>84</b>
5.1	Introduction .....	85
5.2	Method.....	87
5.3	results.....	91
5.4	Discussion.....	98
5.5	Conclusion .....	99
5.6	Author contribution .....	99
5.7	APPENDIX A.....	101
<b>Chapter 6:</b>	<b>MELANOMA CELL CLASSIFICATION .....</b>	<b>103</b>
6.1	Introduction .....	104
6.2	Method.....	105
6.3	Data analysis.....	106
6.4	Results .....	113
6.5	discussion.....	125
6.6	Author contribution .....	127
<b>Chapter 7:</b>	<b>Conclusions .....</b>	<b>129</b>
7.1	Summary of the methodology adopted by this study .....	129
7.2	Summary of the experimental results .....	130
7.3	Limitations.....	132

7.4	Future study.....	133
	<b>REFERENCES.....</b>	<b>137</b>

# List of Figures

Figure 1. De-excitation pathway diagram.....	8
Figure 2. Fluorescence principle.].....	10
Figure 3. Excitation and emission spectra for native fluorophores: (a)excitation spectra (b) emission spectra [20]. ....	13
Figure 4. Microscopic images of (a) breast cancer and (b) associated normal tissue samples from the same patient. The fluorescence spectra of (a) breast cancer and (b) normal tissues were acquired with an excitation of 340 nm[40]. ....	14
Figure 5. Mean fluorescence spectra for normal tissue and cancerous breast tissue gained using a 340nm excitation wavelength[40]. ....	15
Figure 6. Autofluorescence spectra associated with normal colonic tissue and adenomatous colonic tissue (a),(b) spectra associated with normal colonic tissue in 2D and 3D, respectively. (c),(d) adenomatous colonic tissue in 2D and 3D, respectively[42]. ....	16
Figure 7. NADH fluorescence excitation and emission spectra[43]. ....	17
Figure 8. Flavin fluorescence excitation and emission spectra[9]. ....	18
Figure 9. NADH and Flavins roles in cell metabolism[51]. ....	20
Figure 10. PPIX fluorescence emission and absorption [58]. ....	22
Figure 11. The excitation-emission matrix of undiluted urine samples: (a) normal, (b) cancer, wavelength ranges at 250–450 nm for excitation and 270–750 nm for emission[66]. ....	23
Figure 12. Fluorescence properties of normal and abnormal (cancerous) oral mucosal tissues with 410 nm excitation wavelength [63]. ....	26
Figure 13. Available auto fluorescence imaging instrumentation .....	27
Figure 14. Autofluorescence image of normal and cancerous gastric mucosa. ....	28
Figure 15. (a) Light-induced fluorescence endoscopy system diagram.....	29
Figure 16. Customised microscope employed in this study: (a) the microscope[90], (b) a sample filter cube .....	33
Figure 17. Excitation wavelengths used in this study .....	34
Figure 18. Centre wavelength of LEDs used on various excitation fluorophore graph .....	34
Figure 19. Filter cube specifications used in this study .....	37
Figure 20. ORCA-flash4.0lt performance.....	38
Figure 21. ORCA-flash4.0lt performance in terms of quantum efficiency and field of view .....	38
Figure 22. The graphical user interface (GUI) used in this study to control the camera and LED bank[90]. ....	40

Figure 23. Slides and dishes used in this study: (a) A sample, (b) a sample petri dish, (c) a sample 96 well dish. ....	40
Figure 24. A spectral image under processing to remove different noises: (a) Before processing (b) After processing .....	48
Figure 25 Sample preparation and characterization. (a) Collection of eye biopsies from patient subjected to eye surgery. (b) Preparation of histology samples. (c) Cutting of two adjacent tissue sections. (d) Pathology assessment of a stained slide of the sample used as a reference. (e) Multispectral imaging of an unstained slide from adjacent tissue section. ....	54
Figure 26. Inter patient classification workflow and analysis. ....	61
Figure 27. False colour map superimposed on the DIC image of testing patients in the inter-patient classification for different patients in comparison with the corresponding histology images. Data blocks are coloured in red or green if they are respectively predicted to be neoplastic or normal. ....	63
Figure 28. Intra-patient analysis workflow and data analysis. ....	66
Figure 29) False colour map generated to locate the neoplastic boundary on the testing tissue in the intra patient classification framework in comparison with the corresponding histology images. ....	68
Figure 30. Customized auto-fluorescence imaging system setup. One LED with specific wavelength excites the tissue at a time. The scattered auto-fluorescence light pass a channel pass filter and captured by a CCD camera. This process repeats until all the spectral images are taken. ....	74
Figure 31. DIC image registration for a sample patient. ....	76
Figure 32. a spectral image after being pre-proceed and registered. (EX: $378 \pm 10$ nm, EM: 454-496).....	77
Figure 33. Extracting sectors from multi spectral images.....	77
Figure 34. Testing and training tissue for a sample patient in intra-patient framework. ....	78
Figure 35. False colour map superimposed on the DIC image of testing patients in the inter-patient classification for different patients with corresponding H&E images with dash lines to highlight boundary of normal and neoplasm tissue with green and red color, respectively. ....	82
Figure 36) False colour map generated .....	83
Figure 37. Channel selection strategy .....	89
Figure 38. Standard deviation and centre distance of the clusters as the criterion function .....	90
Figure 39. Criterion function improvement using PSO, ANT, DE after iteration .....	93
Figure 40. Cluster improvement due to selecting more effective .....	94
Figure 41. Validating the selected channels on blind test patients. (a) patient no.1 (b) patient no.2 (c) patient no.3 (d) patient no.4. ....	95



Figure 42. Criterion function vs the number of selected channels.....	96
Figure 43. The criterion value of channel subset selection with a curve fitted to the associated criterion.....	97
<b>Figure 44. Performance rate vs increasing number of channels. ....</b>	<b>97</b>
Figure 45. ROC curve of SVM classifier for channel subset selection of DE: (a) after 10 iterations, (b) after 30 iterations, (c) after 100 iterations, (d) after 10k iterations. ....	102
Figure 46. Univariate data analysis to differentiate melanoma cells from normal cells (fibroblast) .....	115
Figure 47. Data analysis framework .....	116
Figure 48. Univariate distance of normal and melanoma cell data points in univariate analysis.....	117
Figure 49. Features selected for multivariate analysis .....	117
Figure 50. Clustering melanoma and normal (fibroblast) cells: (a) training set (Separability distance = 3.56) (b) testing set .....	119
Figure 51. ROC curve for linear and quadratic classifier .....	120
Figure 52. Proportion of variance captured by each principal component .....	122
Figure 53. A discriminative space created by the two top principal components (separability distance ~6.5).....	123
Figure 54. Unsupervised agglomerative hierarchical classification for monitoring the treatment.....	124

# List of Tables

Table 1. Summary of hyperspectral imaging to detect cancer types.....	12
Table 2. Detailed ORCA-flash4.0lt performance and specifications[94]. .....	39
Table 3. Spectral channels used in this study .....	41
Table 4. Summary of patient information and surgical specimens .....	55
Table 5. The excitation/emission wavelength ranges (spectral channels) used in this study .....	55
Table 6. Schedule for tumour cell extraction from the patient.....	121

# Chapter 1: THESIS OVERVIEW

---

This chapter outlines the background (section 1.1) and context of the research, and its purposes (section 1.2). Section 1.3. describes the significance and scope of this research. Finally, section 1.4. includes an outline of the remaining chapters of the thesis.

## 1.1 BACKGROUND

When cells (or tissues) are illuminated by specific wavelengths, they become excited and produce emission light which is called autofluorescence. Autofluorescence is emitted by endogenous (native) fluorophores which are present within cells or tissues naturally [1]. Cells and tissues contain many native fluorophores such as Protoporphyrin IX (PPIX), reduced nicotinamide adenine dinucleotide (NADH), Flavins, cologones. Fortunately, these components significantly contribute to cell metabolism and can, therefore, deliver an autofluorescence signature to evaluate biological conditions [2, 3]. Such a concept has been utilised in autofluorescence microscopy, which has helped detect the complex biological interactions and visualise cellular structures[4, 5].

Autofluorescence imaging (AFI) is an imaging technique that can capture the autofluorescence spectra of native fluorophores[6, 7]. Mainly, autofluorescence concept is used in biology as a non-invasive diagnostic and evaluation tool [8], which can be utilised as a real-world technology and also is an appropriate research tool to assist in learning about molecular interactions and mechanism [5, 8, 9]. Such technologies can be employed with no staining or preparation of the sample and can be used to study intact living cells or tissues with no possible artefactual side effects made by introduced chemicals.

Autofluorescence application in health science is quite novel although autofluorescence phenomenon had been discovered for a long time [10]. Conventionally, auto-fluorescence imaging technology has been limited to a few costly channels employed in some microscope methodologies such as fluorescence lifetime imaging (FLIM). Consequently, these technologies can monitor only a limited number of fluorophores. .Recently, new equipment such as advanced cameras and

computational power have facilitated auto fluorescence applications in the medical environment with advanced technology[9].

The current thesis employed a cutting-edge technology called multispectral autofluorescence imaging developed in our group recently[9, 11]. Multispectral autofluorescence imaging is cost-effective instrumentation as it uses LED to excite the sample rather than expensive lasers and employs tens of distinctive spectral channels to study the biological condition of cells or tissues[11]. This represents an advance over traditional auto-fluorescence imaging systems. The availability of multiple channels makes it possible to survey the overall biochemical composition of the tissue, in addition to detecting specific markers to identify the tissue state.

Such technology advancement and previous promising results obtained in our group such as finding different cell subpopulations[9, 11] and ability to detect various compartments of the cells[9] inspired us to modify the technology further and apply it for more critical applications such as cancer characterisation and detection. In this study, we investigated two cancer types including eye surface neoplasia[12, 13] and melanoma (skin cancer)[14-16] where our technology can be simply translated for real-world applications as these diseases grow on the surface of the eye and skin where taking spectral images are straightforward compared to some organs such as stomach. In this study, the modified multispectral autofluorescence technology is demonstrated by imaging ex vivo human ocular tissue (normal and neoplasm) and also cells derived from human melanoma tumours.

## **1.2 STATEMENT OF THE PROBLEM**

Early cancer detection is highly significant to improve patients' chance of survival and their life quality[17, 18]. Generally, signs of cancer in the early stages are conventional[19, 20]. Therefore, clinicians need to make a reliable decision based on histopathology analysis, followed by biopsies. Unfortunately, biopsy is an imperfect approach due to several reasons: depending on skills and experience of the clinician to sample the tissue, biopsy can be subjective[20]. Biopsy tissue may fail to collect the diseased tissue, especially for sampling small tumours, and, therefore, the sample may show to be free of cancer incorrectly[21]. In addition, Histopathology assessment is time-consuming and cannot be performed easily for cancer margin assessment[20, 22].

Therefore, an autofluorescence imaging diagnostic technique which is cost-effective, potentially free of staining, and able to be performed in the clinic for real-

time cancer detection and margin delineation is highly desirable and can have a significant impact, especially in ophthalmology field. In addition, such a technique is desirable in the cancer research field. The opportunity of gaining information about living cancer cells in a label-free manner can be significant for precise early detection, refining current therapies and also discovering innovative treatment procedures [23-27].

In addition, the autofluorescence multispectral imaging is quite new, so there are several gaps which are needed to be explored in terms of the instrumentation and data analysis. Although tens of spectral channels can be employed in multispectral imaging, the number of channels need to be wisely selected and optimised depending on the application to facilitate translating the technology for a real-world application. In addition, data analysis in multispectral images is challenging. Multispectral imaging provides a big volume of data which may be interfered by different noises. Therefore, a comprehensive data analysis methodology needs to be employed to extract rich and valuable information from spectral images for cancer characterisation. Moreover, sophisticated data analysis method has been applied throughout this thesis to extract the informative features from the spectral images.

### **1.3 SIGNIFICANCE AND NOVELTY**

This thesis modified the multispectral imaging further in terms of illumination and filter arrangement to make the technology specialised for cancer detection and evaluation. This system uses light excitation with a number of narrowband ranges of wavelength and collects the native fluorescence emission of the sample at specific wavelengths. A combination of excitation /emission wavelength bands forms a spectral channel. In this study, 38 channels were arranged, which capture various fluorophores, especially the cancer biomarkers. To acquire separate spectral images and document the colour of sample, samples are imaged in each of these channels.

This research study has focused on three main contributions after establishing the setup and modify the newly-designed auto-fluorescence multispectral imaging methodology. First, the customised non-invasive auto-fluorescence multispectral imaging methodology is applied to detect and delineate Ocular surface squamous neoplasia (OSSN). There is a gap for non-invasive OSSN detection using

autofluorescence. Such technology can have a significant impact in the ophthalmology field as described in Chapter 4:.

Secondly, this study improves the technology one step further for real world translation. In fact, the technology needs to be optimised further for OSSN detection in terms of the effective channels to facilitate industrial translation. Using such cutting-edge technology in auto-fluorescence imaging led us to employ some channels simultaneously to diagnose or monitor diseases with high accuracy. However, the effectiveness of a lower number of channels needs to be evaluated to find an opportunity for optimising the system as described in chapter Chapter 5:.

Third, the modified technology is evaluated on melanoma cells. So metabolic heterogeneity of melanoma cancer cells and fibroblast were considered in this study. Multi-spectral auto-fluorescence imaging was employed for evaluation of melanoma cells and fibroblast in order to produce discriminative information including a variety of biologically relevant quantitative information. This study also demonstrates the possibility of obtaining information about melanoma cells when they are excised from a patient under treatment as described in Chapter 6:.

Multispectral imaging produces a big volume of information about a sample. Consequently, this information needs to be best analysed to characterise cancer cells or tissues accurately. This study gives information about the imaging protocols and the way to process images and minimise the random or systematic errors which affect the spectral images. Data analysis were performed in this thesis by univariate analysis, multivariate analysis, and artificial intelligence. In fact, in-depth data analysis makes multispectral auto-fluorescence imaging a more powerful tool for the detection and evaluation of cancer cells or tissues.

## **1.4 THESIS OUTLINE**

This thesis has been written in a thesis by manuscript format and includes 7 chapters. Chapter 1 outlines a general introduction and the motivation of this study. Chapter 2 investigates the concept of autofluorescence and describes the basic biology related to the autofluorescence. The various native fluorophores which have been mostly used as cancer biomarker are described and finally reviews various cancer types which have been characterised by autofluorescence concept.

Chapter 3 provides the details of the multispectral system employed in this study. The excitation and emission channels which are specified with a full description in the first part of the chapter. Then, various random or systematic noise which may interfere with the spectral channel and the techniques to treat them are outlined. The pre-processing strategy to prepare the spectral channels is elaborated in this chapter.

Chapter 4 presents the application of technology in the detection of ocular surface squamous neoplasia using multispectral autofluorescence imaging. This chapter contains the revised manuscript and answers to reviewers. Firstly, section 4.1 describes the current methodology for OSSN detection and the challenges leading us to apply the technology in the field of ophthalmology. Section 4.2 and 4.3 provides the details related to the patients, a brief description of the instrumentation and details of frameworks, data analysis and artificial intelligence used to delineate the boundary of OSSN. Section 4.4 presents the results of two frameworks used to detect OSSN and provides the statistical measures of the performance of the technology in detecting and delineating OSSN. Finally, section 4.5 discusses all of the results, challenges and future perspective of the study. This chapter also contains the supplementary material for the manuscript in section 4.8 and 4.9.

Chapter 5: presents a novel methodology to optimise the multispectral system to facilitate the technology for industrial translation. Section 5.1 describes the problem and the available methodology for optimising a system and the challenges to optimise the multispectral imaging, especially for cancer detection and also the novelty of the proposed technology. Section 5.2 presents the technique developed in this thesis to optimise the system after describing the samples and methodology briefly. Section 5.3 presents the results of the proposed technique. Finally, section 5.4 discusses the results and the future prospective of the technique.

Chapter 6: demonstrates the application of the technology for melanoma cell characterisation. Section 6.1 presents the opportunities of characterisation of the melanoma cells in a label-free manner. Section 6.2 describes the cell lines used in this research study, a brief description of the image acquisition system and the univariate and multivariate approaches for data analysis. Section 6.4 presents the results obtained from the spectral channels to distinguish melanoma cells from fibroblast followed by characterisation of a cell line related for a patient under treatment. Finally, section 6.5 discussed the results and future prospective of the research. In Appendix B, the

application of the artificial neural network to evaluate the data extracted from the spectral information of the channels were also presented.

Chapter 7: gives the summary of this work, the limitations and discusses the possible extension of this study.



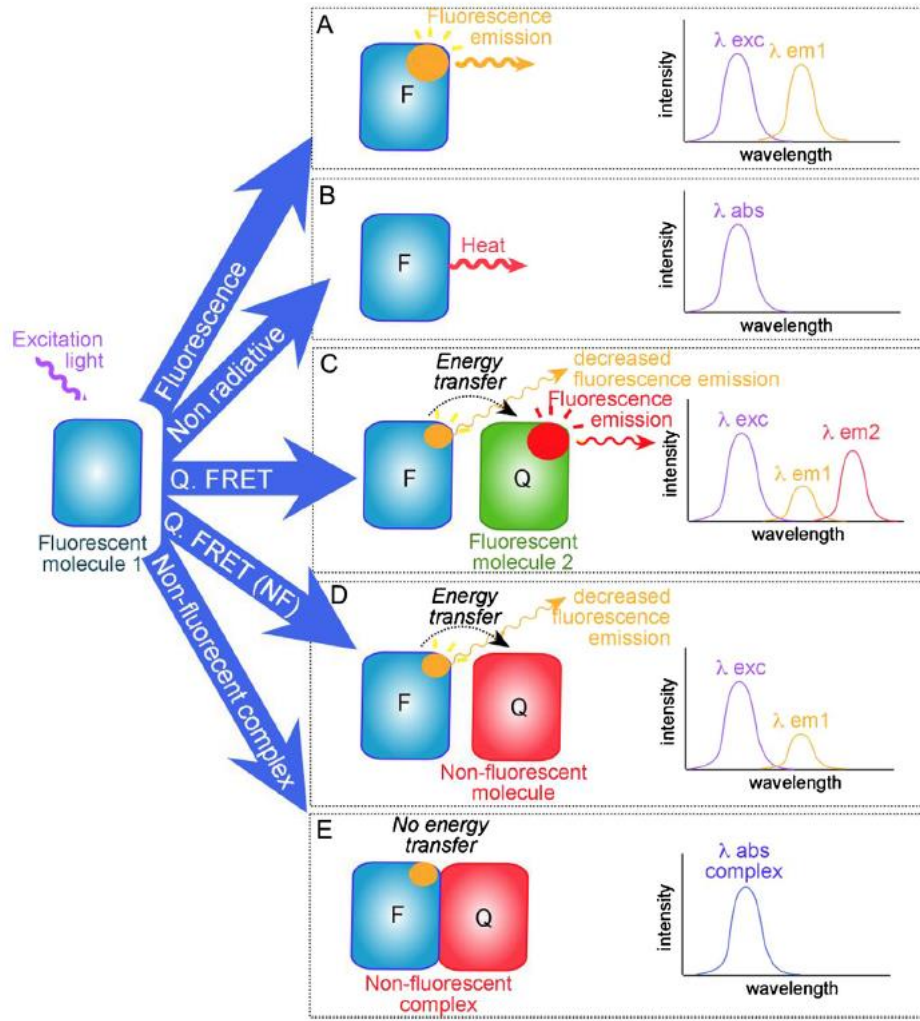
## Chapter 2: Autofluorescence in cancer characterisation

---

This chapter outlines the basic principles of fluorescence (section 2.1 and 2.2) and factors which may affect the fluorescence properties (section 2.3). Next, labeled and label-free detection using fluorescence concept was described in section 2.4. After describing various native fluorophores present within cells and tissues (section 2.5), the most important native fluorophores for cancer characterisation are described in section (2.6 , 2.7 and 2.9). Finally, screening several cancer types using autofluorescence is outlined in section 2.10.

### 2.1 FLUORESCENCE PRINCIPLES

Molecules in cells and tissue, which naturally emit light when irradiated by light at suitable wavelengths are called fluorophores [9]. When illuminated, fluorophores can absorb energy, and an electron of the molecule may move to an excited state. For a short time, the fluorophore stays in an excited electronic energy level and then emits fluorescent light to get back to its ground state. However, it is possible that the excited electron returns to the ground state without fluorescence emission via different non-radiative processes[28]. Figure 1 shows different de-excitation pathways, including fluorescent emission of a fluorophore, radiation-less deactivation, energy transmission to another fluorophore within the cells, energy transmission to a non-fluorescent molecule and static quenching[28].



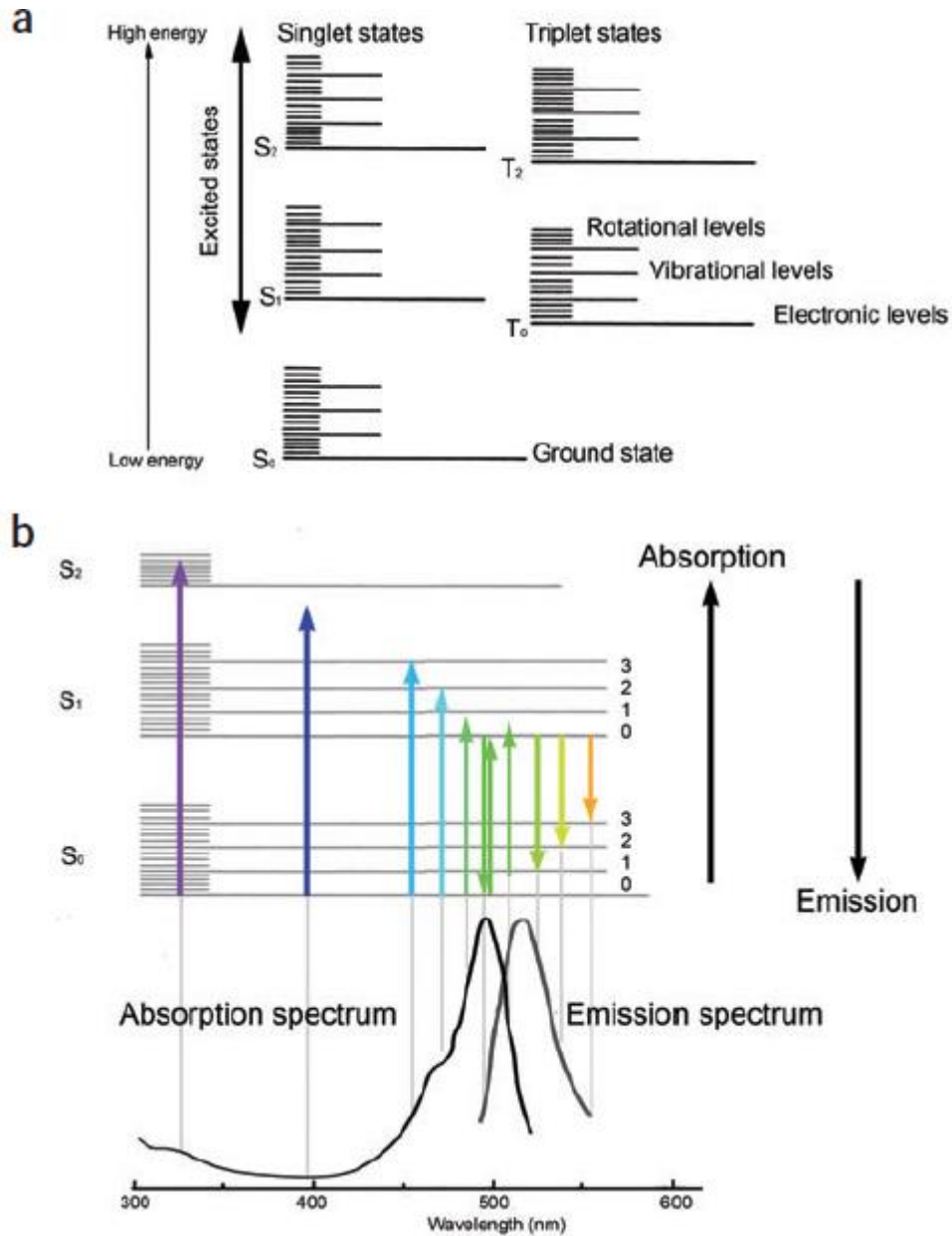
**Figure 1. De-excitation pathway diagram. (A) fluorescent emission of a fluorophore, (B) Radiation-less deactivation, (C) Energy transmission to another fluorophore within the cells, (D) energy transmission to a non-fluorescent molecule and (E) static quenching[28].**

In the fluorescent emission pathway, when emitting a photon, the energy of the fluorophore drops by a value equal to the energy of the photon. Due to some energy loss during the process, the fluorescent photon emitted by a fluorophore has a longer wavelength than the photon absorbed by the fluorophore. This process can be shown using the following equations:

$$S_0 + h\nu_{ex} \rightarrow S_1 \quad \text{Eq. 1}$$

$$S_1 \rightarrow S_0 + h\nu_{em} + \text{heat} \quad \text{Eq. 2}$$

Equation 1 demonstrates the excitation process in which  $S_0$  is the ground state,  $h$  is Planck's constant,  $\nu_{ex}$  the frequency of the excitation light and  $S_1$  is the excited state of the electron. Equation 2 shows the emission process in which  $\nu_{em}$  is the emission frequency [29]. The principle of fluorescence emission is illustrated by a classical Jablonski diagram schematically as shown in Figure 2 (a,b,c).



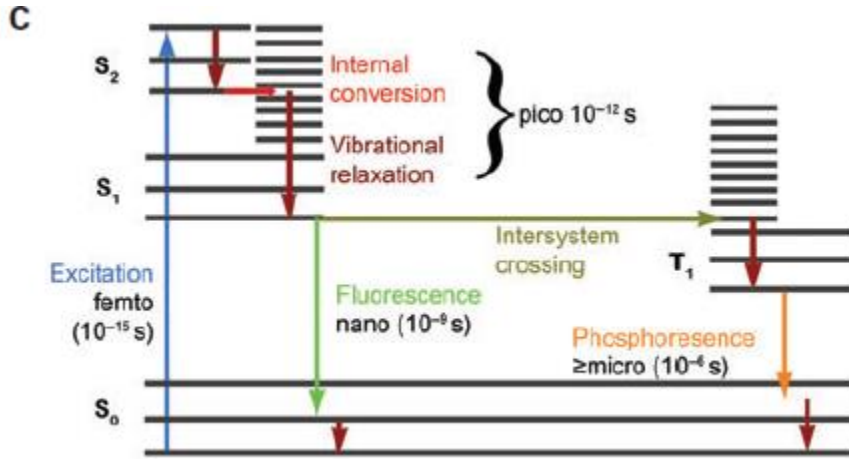


Figure 2. Fluorescence principle. (a) Jablonski diagram illustrates the different energy levels of a molecule. Before absorbing energy, molecules are on the ground state with the least vibrational states  $S_0$ . (b) The spectral properties associated with energy absorption and emission by a molecule could be linked to the energy values required to take a molecule from one energy state to another. Here, the absorption and emission spectra related to a sample fluorophore (FITC) are illustrated in a Jablonski diagram. Individual perpendicular gray line shows the absorption or emission spectra. Depending on the wavelength of the photons, the colours of the arrows are defined. As a case in point, purple colour represents a photon with ultraviolet energy and orange colour demonstrates the least energy that a photon can carry as it returns to the ground state  $S_0$ . (c) The period related to different stages including excitation, emission, and phosphorescence that the change can take [30].

## 2.2 QUANTUM YIELD

Quantum yield ( $\Phi$ ) is a parameter which characterises how effectively a fluorophore can produce fluorescence emission and is described by the following equation:

$$\Phi = \frac{NP_{emitted}}{NP_{absorbed}} \quad \text{Eq. 3}$$

which is the ratio of the number of emitted photons ( $NP_{emitted}$ ) to the number of absorbed photons ( $NP_{absorbed}$ ). Quantum yield is a molecular characteristic and is not related to instrumentation [31].

## **2.3 FACTORS AFFECTING AUTOFLUORESCENCE PROPERTIES**

The complexity of the absorption and emission peaks associated with biological components can be much more than the fluorescence properties of primary fluorophores [32]. The local environmental factors of biological fluorophores can influence their spectral properties. Such factors include temperature, specific ion concentrations, the degree of humidity, pH value and the relative interaction with other molecules[33]. As a case in point, different tryptophan or tyrosine molecules in a protein may show different fluorescence properties depending on the other amino acids present in their approximation[33, 34]. Besides, fluorescence intensity of a specific fluorophore may decrease due to the presence of another molecule which can absorb the emitted photons. Such a process is called fluorescence quenching [28].

Another factor which may change the emission fluorescence spectra and the quantum yield of a fluorophore is the pH value. The chemical specification of a molecule can be altered due to a different value of pH. Consequently, such a chemical specification variation can influence autofluorescence in intracellular locations[28]. Temperature also affects fluorescence properties based on various mechanisms. Temperature can change the medium viscosity and subsequently the interaction between different fluorophores. Also, some metabolic processes are dependent on the temperature, which can change the availability of some components[28].

## **2.4 LABELLED AND LABEL-FREE DETECTION BY FLUORESCENCE**

Methods which evaluate the tissues or cells by fluorescence imaging can be classified depending on the fluorophore types investigated to detect disease and the principles employed in instrumentation. There are two major categories for fluorophores: endogenous and exogenous fluorophores. Endogenous fluorophores are native tissue fluorescence and are responsible for autofluorescence, which is the focus of this study. On the contrary, Exogenous fluorophores have external origins and mostly are used for labeling.

Auto-fluorescence is a term used to differentiate the native fluorescence of cells and tissues from the fluorescence acquired by treating samples with exogenous fluorescent markers such as indocyanin green (ICG), that binds the structure of cells or tissues[20, 32]. The technique using exogenous fluorescent markers for diagnostics is based on variances in uptake between healthy and unhealthy tissues. However,

autofluorescence detection relies on variations of the concentration and distribution of endogenous fluorophores within the tissues or cells[20, 32].

A similar method to multispectral autofluorescence imaging is hyperspectral reflectance imaging, which has dominantly used to detect difference cancer types, which is reviewed in ref [35] as shown in Table 1. The advantage of the hyperspectral reflectance is its simplicity; however, this method is limited to specify specific excitation band, which affect its accuracy. The gap available in the literature in the spectral images is that there is no system with flexible excitation and emission bands.

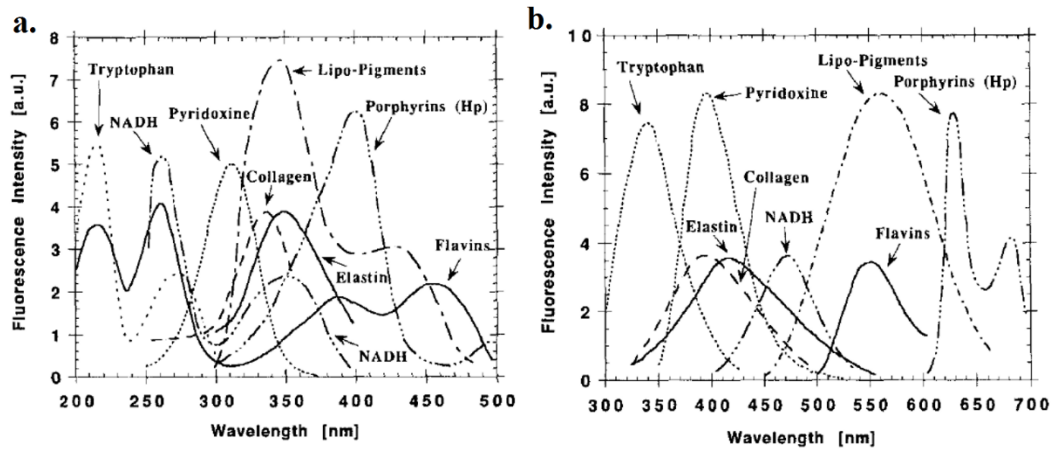
**Table 1. Summary of hyperspectral imaging to detect cancer types**

Spectral range (nm)	Spectral resolution ( $\mu\text{m}/\text{pixel}$ )	Detector	Dispersive device	Acquisition mode	Measurement mode	Application
200 to 700	$\sim 5$	CCD	Filter wheel	Staring	Fluorescence and reflectance	Cervical neoplasia
330 to 480	5	CCD	Filter wheel	Staring	Fluorescence and reflectance	Cervical cancer
5000 to 10,526	11	HgCdTe	—	FTIR	Reflectance	Cervical pathology
400 to 720	—	CCD	LCTF	Staring	Fluorescence	Tumor hypoxia and microvasculature
440 to 640	1 to 2	CCD; ICCD	AOTF	Staring	Fluorescence and reflectance	Skin cancer
365 to 800	$\sim 1$	CCD	Prism	Pushbroom	Transmission	Melanoma
450 to 700	—	CCD	LCTF	Staring	Reflectance	Breast cancer
1000 to 2500	6.29	HgCdTe	PGP	Pushbroom	Reflectance	Gastric cancer
450 to 650	4 to 10	CCD	Prism	Snapshot	Reflectance	Endoscope
450 to 950	2	CCD	LCTF	Staring	Reflectance	Prostate cancer
400 to 640	—	CCD	Filter wheel	Staring	Fluorescence and reflectance	Ovarian cancer
2500 to 11,111	—	HgCdTe	—	FTIR	—	Breast cancer

## 2.5 NATIVE FLUOROPHORES

Most native fluorophores are related to the tissue's structural matrix or are associated with the metabolic processes of cells. The key native fluorophores of the tissue structural matrix are elastin and collagen. Such fluorescence is due to cross-linking among amino acids. Fluorophores related to metabolism of cells are flavins and reduced nicotinamide adenine dinucleotide (NADH). In addition, there are other native fluorophores such as tryptophan, tyrosine and phenylalanine, which categorized as the aromatic amino acids. Porphyrins and lipopigments are also among the main fluorescence components which are the end-products of lipid metabolism [36].

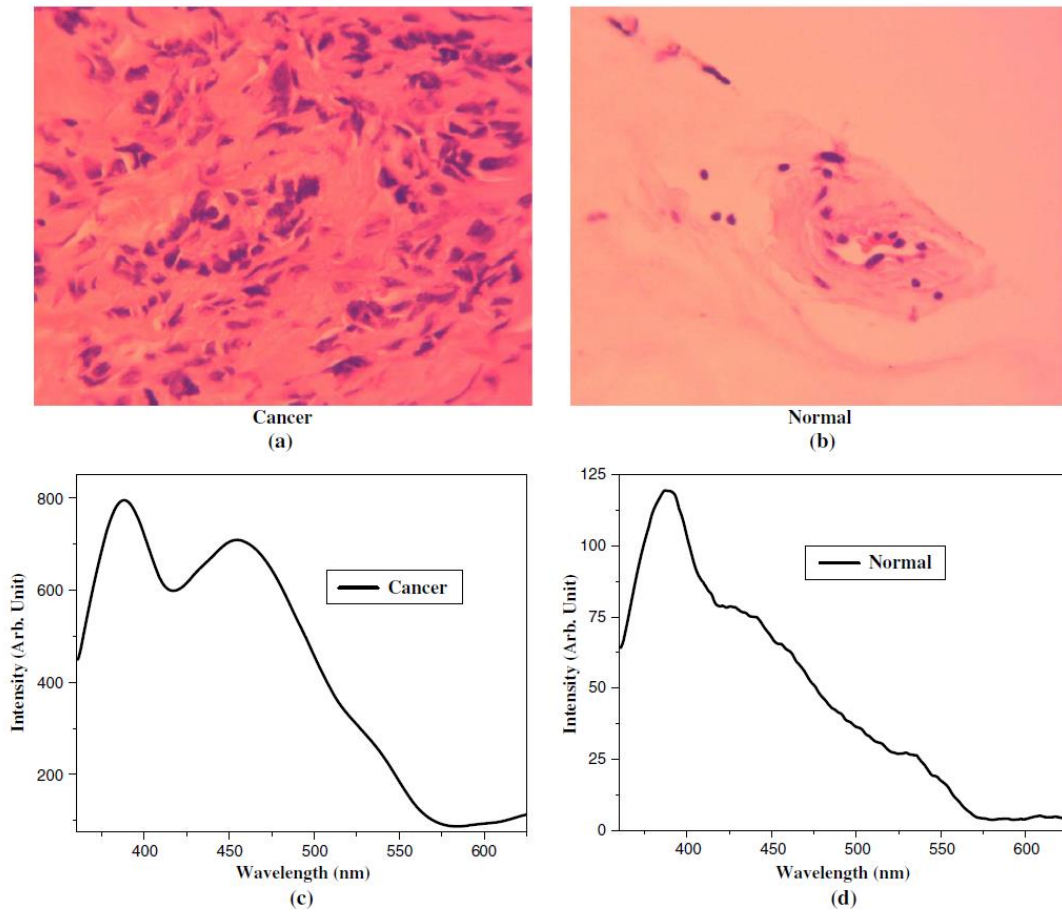
To interpret clinical and biological studies based on autofluorescence, the following key points need to be considered. First, each fluorophore possesses a unique excitation and emission spectra (see Figure 3). Second, biological tissues or cells may have a blend of several fluorophores with various concentrations. Consequently, to diagnose cancerous lesions based on autofluorescence, the concentration variation of one or more fluorophore(s) needs to be considered [20].



**Figure 3. Excitation and emission spectra for native fluorophores: (a)excitation spectra (b) emission spectra [20].**

## 2.6 NATIVE FLUOROPHORES USED AS CANCER BIOMARKER

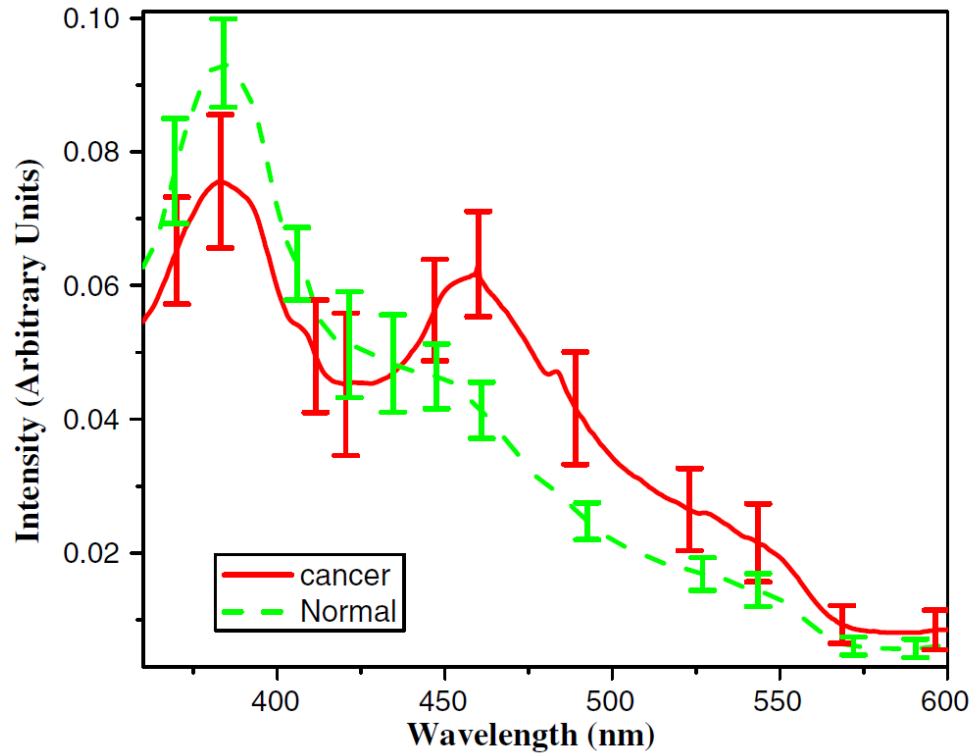
The fluorescence spectra of cancerous and normal tissue can be different, due to metabolism changes and morphologic alterations of the epithelial surface and underlying stroma [37, 38]. In addition, absorption and scattering of light changes due to the increased numbers of nuclei and increased microvasculature present in cancerous tissue[39]. Figure 4 shows microscopy images of a breast cancer tissue obtained from a patient who is an 85-year-old female. The disease is ductal carcinoma in situ (DCIS) with grade 3. Figure 4 (a) and (b) demonstrate microscopy images with a 40X magnification of the cancerous and normal breast tissues, respectively. Figure 4 (c) and (d) represent fluorescence spectra of the cancerous and normal breast tissues obtained from the same patient [40]. Figure 4 (c) and (d) demonstrate a clear difference between cancerous and normal fluorescence spectra.



**Figure 4. Microscopic images of (a) breast cancer and (b) associated normal tissue samples from the same patient. The fluorescence spectra of (a) breast cancer and (b) normal tissues were acquired with an excitation of 340 nm[40].**

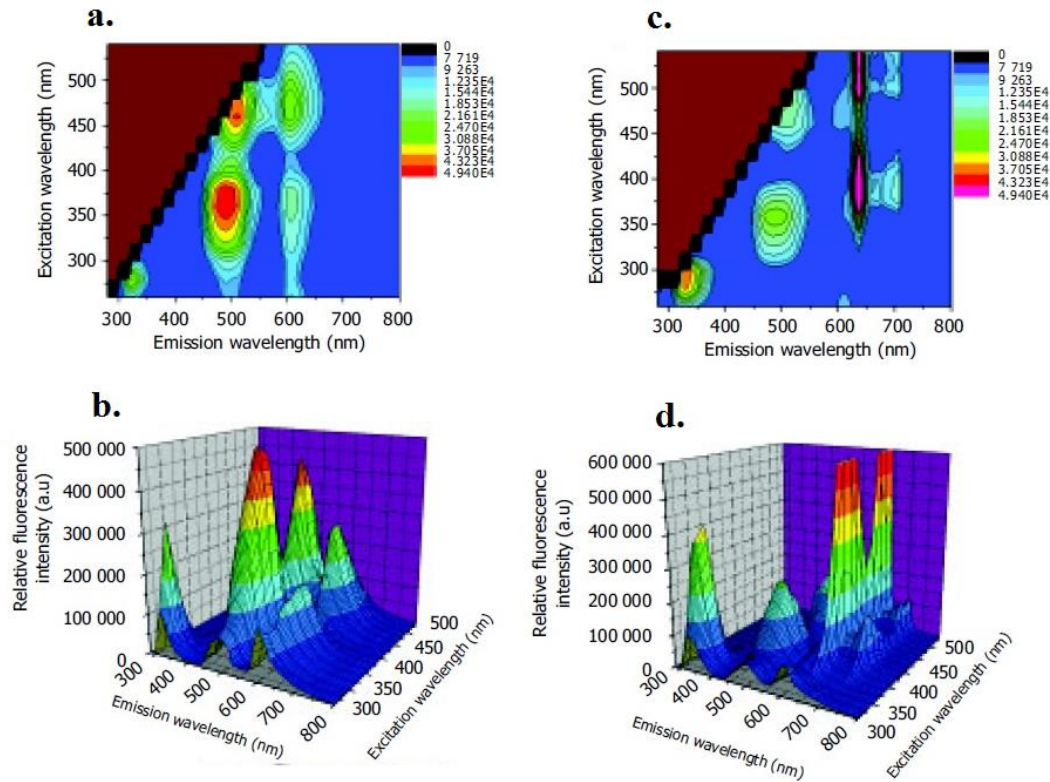
Figure 5. demonstrates the average fluorescence spectra associated with cancerous and normal breast tissue from 38 patients, which was obtained at a 340 nm excitation wavelength[40].





**Figure 5. Mean fluorescence spectra for normal tissue and cancerous breast tissue gained using a 340nm excitation wavelength[40].**

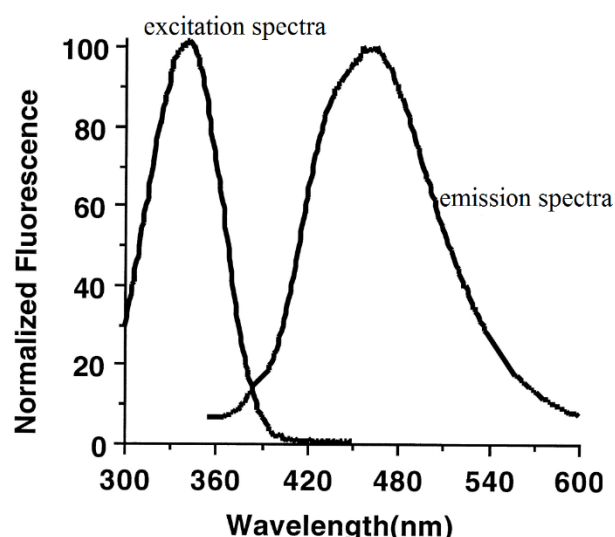
Although several native fluorophores have been used to detect various cancers, most of the researchers have tracked NADH, FAD, and PPIX[41]. In addition, it is reported that the autofluorescence spectra of normal and cancerous tissues have no significant difference regarding the fluorescence types, but a great difference can be observed in the fluorescence peaks and associated intensities between normal and adenomatous tissue from the same patient[42]. As shown in Figure 6, autofluorescence spectroscopic differences in normal and adenomatous colonic tissues are investigated based on three native fluorophores, NADH, FAD, and PPIX. Although the same spectra can be observed in terms of the emitted light wavelength (fluorophore types), substantial differences in autofluorescence values can be seen in normal and adenomatous colonic tissues [42].



**Figure 6.** Autofluorescence spectra associated with normal colonic tissue and adenomatous colonic tissue (a),(b) spectra associated with normal colonic tissue in 2D and 3D, respectively. (c),(d) adenomatous colonic tissue in 2D and 3D, respectively[42].

## 2.7 NADH AND FLAVINS

Native NADH exists in the free or bound form in cells and tissue. The excitation and the emission peaks of NADH are shown in Figure 7. Because more than 80% of mitochondrial NADH is bound to proteins, its fluorescence properties can be different in different cases [9].



**Figure 7. NADH fluorescence excitation and emission spectra[43].**

Flavin can be excited at visible light wavelengths and is a significant source of cell autofluorescence. Oxidised Flavin can be produced from riboflavin and flavo-proteins, which are specifically concentrated in mitochondria. The free excitation spectrum of Flavin adenine dinucleotide (FAD), riboflavin and Flavin mononucleotide (FMN) were obtained, and it is reported that the spectra are very similar to each other as shown in Figure 8 [9].

Riboflavin is the precursor associated with the main cofactor of flavoproteins. It is a biologically common form including FAD and FNM. Their similar emission spectra range from 490-570nm, and their excitation spectra range from 320-480nm. Such co-factor side chains play a significant part in protein binding, so most flavoproteins contain bound FAD or FMN[44, 45].

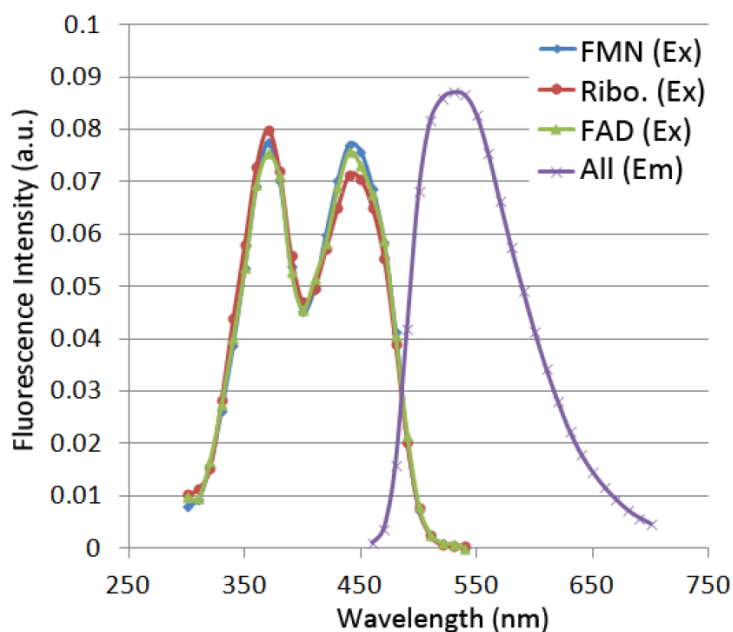


Figure 8. Flavin fluorescence excitation and emission spectra[9].

## 2.8 NAD(P)H AND FAD IN CELLULAR METABOLISM

Metabolism is described as all of the biochemical interactions in the cells [46]. Metabolism includes degrading diet components, synthesising macromolecules that are required by the cell, and producing small precursor molecules such as some amino acids for the needs of the cells. Metabolism also includes those reactions which involve electron transfers, such as oxidation reduction[47]. Metabolism serves three major objectives including the food/fuel transformation to energy for cellular processes, converting food/fuel to building blocks, and the eliminating nitrogenous wastes. Based on such processes organisms can grow and replicate, keep their structures, and interact with their environments[47].

Metabolism can be defined with two categories, catabolism and anabolism[48]. Catabolism is the breaking down of organic components, such as the conversion of glucose to pyruvate. On the other hand, the building up of cell components, including proteins and nucleic acids is defined as anabolism. Typically, catabolism releases energy and anabolism consumes energy. The chemical reactions taking place in metabolism are controlled into metabolic pathways, where one chemical is obtained from another chemical through a series of steps using a sequence of enzymes. Enzymes play a pivotal role in metabolism as they allow organisms to handle desirable reactions

which need energy and will not happen by themselves. Enzymes are defined as catalysts, which help reactions to progress more rapidly. [9, 46].

The speed of metabolism, which is defined as the metabolic rate, can affect the amount of food an organism will need, and also influences the way it can gain that food. A prominent aspect of metabolism is the metabolic pathways similarity of various species[49]. As a case in point, the set of carboxylic acids which are the intermediates in the citric acid cycle exist in all recognised organisms [50]. Normally, catabolism takes electron carriers to their associated reduced form. However, oxidative phosphorylation takes them to their oxidised level. Glycolysis generates a net two pyruvate molecules, two ATP, and also reduces two components of NAD<sup>+</sup> to NADH as shown in Figure 9. Next, Pyruvate may move into the mitochondria, in which it undertakes further broken down through pyruvate dehydrogenase complex (PDHC), generating acetyl CoA. Throughout this progression, initially, electrons are transported to the FAD component, which is an autofluorescent molecule and bound to LipDH within the PDHC, to make FADH<sub>2</sub>[51].

LipDH is defined as an in-between electron transporter and hands off the electrons to the NAD<sup>+</sup> present in cells to create NADH inside the mitochondria. Next, the acetyl CoA component can go into the tricarboxylic acid (TCA) cycle. This process breaks down carbon-based molecules to CO<sub>2</sub> and generates three NADH for each acetyl CoA. In this process, dehydrogenase complexes, which contain autofluorescent LipDH, carries electron. In addition, NADH dehydrogenase inside the ETC can oxidise the NADH which is generated by catabolism[51].

Electrons are transferred by oxidative phosphorylation to several enzyme complexes which consume energy to deliver protons to the mitochondria intermembrane space. This process terminates with oxygen which is the final electron acceptor. The proton gradient generated between the inner mitochondrial membrane controls ATP construction by ATP synthase. This process, including NADH and FAD electron transfer, delivers an effective energy conversion from carbohydrate catabolism to ATP. Well-organized ATP generation is critical in mortally differentiated cells. Subsequently, such cells mainly depend on oxidative phosphorylation, which generates a great baseline optical redox ratio of FAD/ [NAD(P)H+FAD][51].

Conditions leading to cell hypoxia drop the redox ratio due to decreasing oxidative phosphorylation and increasing glycolysis to generate enough ATP. Through the

anaerobic glycolysis process, pyruvate and NADH are transformed to NAD<sup>+</sup> and lactate via LDH to keep break more glucose down and generate ATP. Consequently, free cytosolic NADH increases due to anaerobic glycolysis and results in dropping the optical redox ratio. In addition, Glucose catabolism increases relative to oxidative phosphorylation through biosynthesis. Glycolysis increase despite oxygen availability is called the Warburg effect[52], which leads to decreases in the optical redox ratio[51].

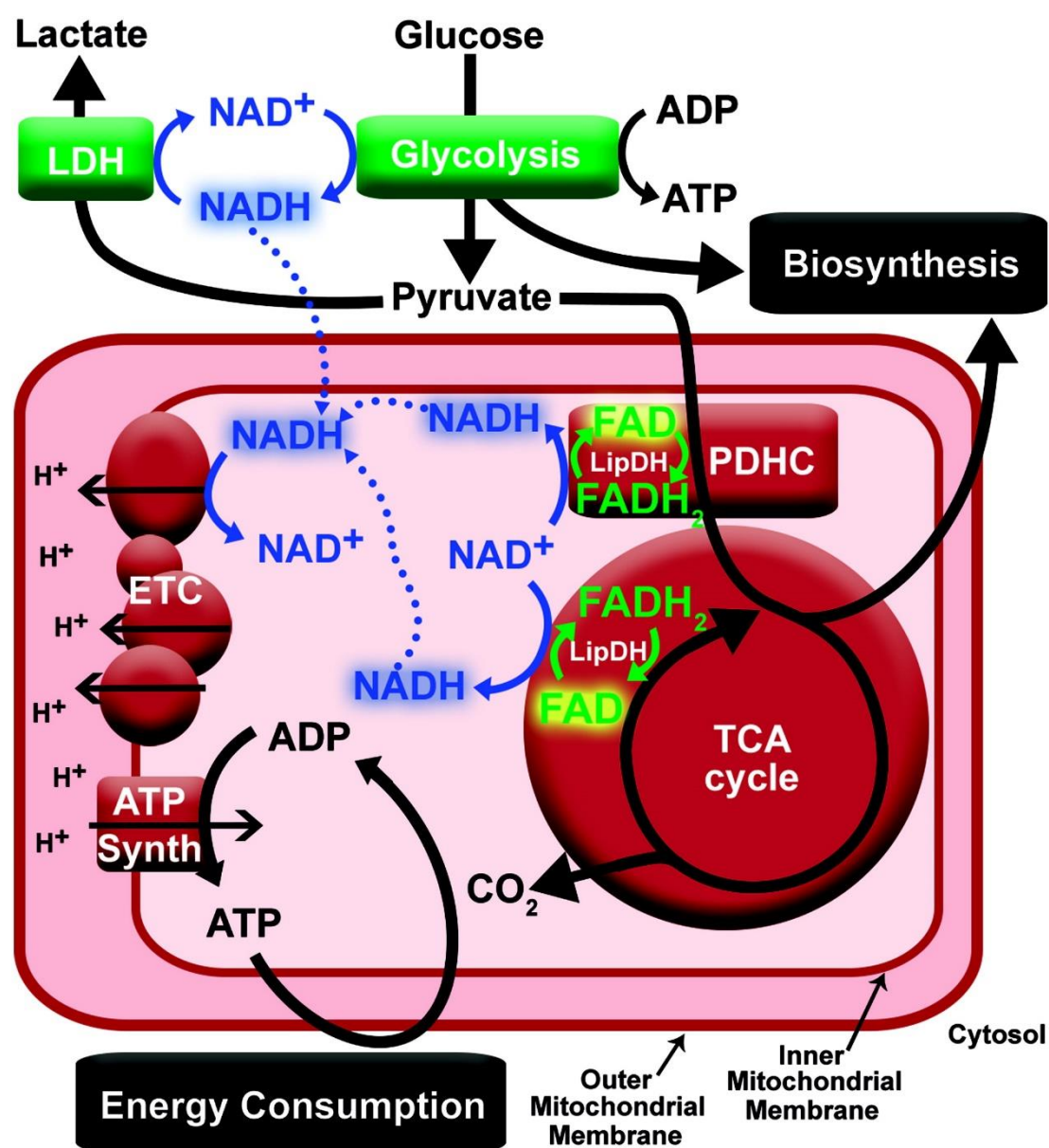


Figure 9. NADH and Flavins roles in cell metabolism[51].

In fact, Cancer cells can ferment glucose into lactate irrespective to the presence of oxygen. Such effect helps the efficient synthesis of macromolecular components which is necessary for fast dividing cells [52-54]. During proliferation, the great increase in glycolytic flux quickly produces cytosolic ATP, which results in variation in the number of high energy molecules such as NADH [55]. Moreover, the FAD compound concentration changes when cells become cancerous, as FAD was contributed to the oxidation of metabolic molecules [3, 55, 56].

To give more information about metabolism and glycolysis, metabolism is converting food into energy. Metabolism is a complex biochemical process, which describes how food calories are combined with oxygen to produce the energy our body requires to function. Glycolysis is a sequence of reactions that produce energy from glucose via splitting it into two three carbon molecules with no need of oxygen presence.

## **2.9 PORPHYRINS**

Porphyrins play a significant role in disease detection, clinical fluorescence spectroscopy and photodynamic diagnostics. Typically, porphyrins are synthesised from 5-aminolevulinic acid (5-ALA). While heme, the precursor of the red blood chromophore haemoglobin, is produced inside the mitochondria, a key porphyrin named Protoporphyrin IX (PpIX) is generated[57]. As shown in Figure 10, the maximum emission of fluorescent porphyrin components is at 630 nm and maximum excitation is at 400 nm[58]. It is also reported that a distinct 650nm emission peak measured in skin autofluorescence is due to Porphyrins[59].

Cells contain a small amount of Porphyrins naturally. Under normal conditions, Protoporphyrin synthesis depends on feedback control. Cells generate protoporphyrin at a rate just sufficient to match their heme levels. Excessive cellular proliferation, however, makes the feedback mechanism lose control. Consequently, extra porphyrin creation appears in the tissues and blood [60, 61]. Various forms of native porphyrins exist in organisms, including Protoporphyrin IX, uroporphyrin, chlorins, phthalocyanines, coproporphyrin and hematoporphyrin. Depending on particular forms and links to other molecules, Porphyrins may show a complex spectral and optical behaviour[62].

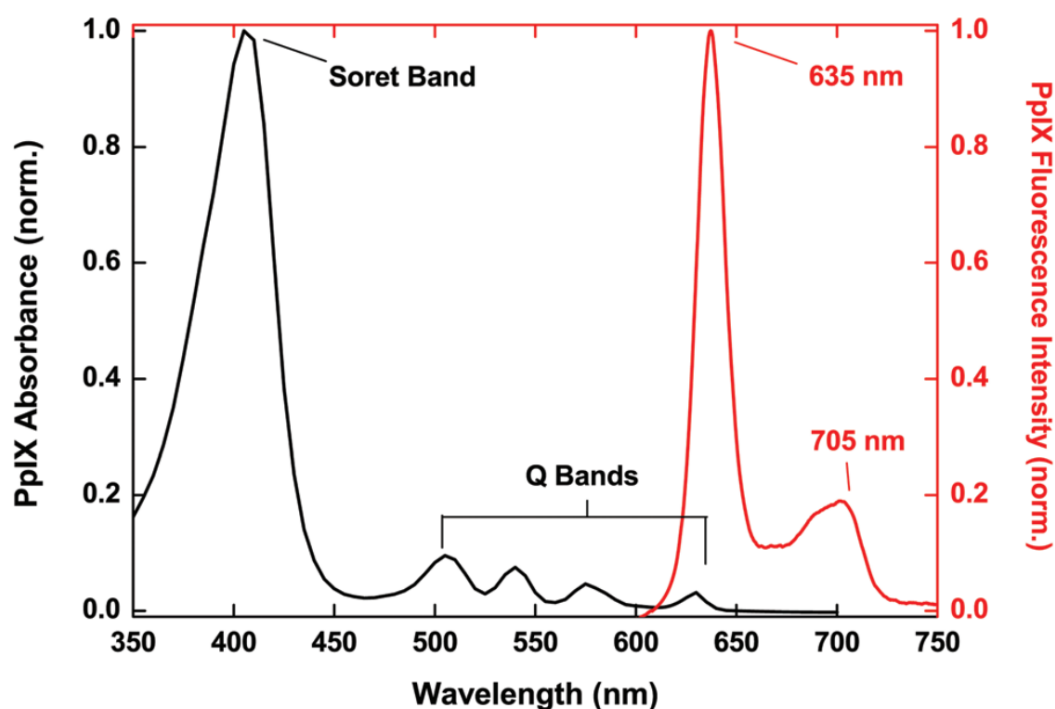


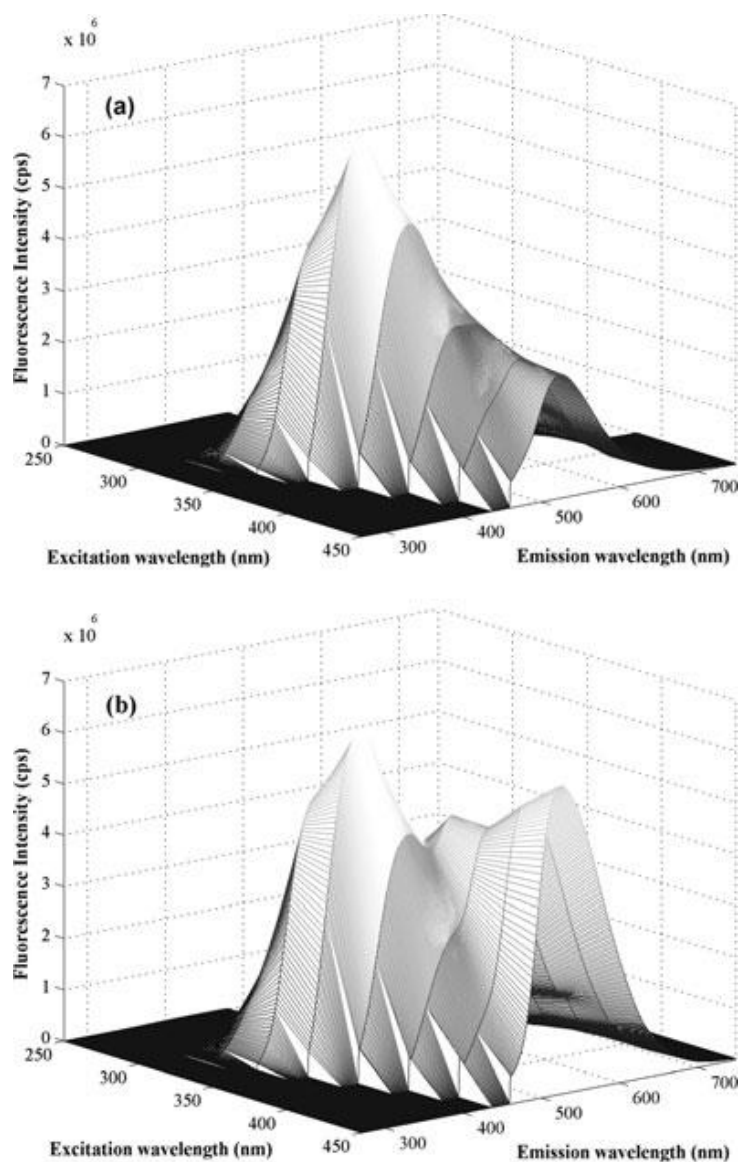
Figure 10. PPIX fluorescence emission and absorption [58].

Protoporphyrin IX (PpIX) is a key porphyrin in eukaryotic cells among various forms of Porphyrins. PpIX is a part of heme synthesis and is a robust biomarker to detect cancerous tissue [62]. The fluorescence variation of PpIX between normal and cancerous tissue and cells might be affected by various factors. For example, pharmacological and physiological aspects may change from various positions of organs and tissues and also from different stages of disease [20]. Change in the amount of PpIX in cancerous cells may also be associated with lack of iron or ferrochelatase [63] in tumours, which results in a change of the PpIX concentration relative to normal host tissue[20, 64].

The application of native PpIX as a cancer marker has been reviewed in[42, 64] for colonic cancer and in [61] for breast cancer. It is reported that red fluorescence at 635 nm due to porphyrin compounds in sera of cancer patients has been seen. In addition, it has been demonstrated that there is a relationship between porphyrin (with fluorescence at 630 nm) and cancer cell proliferation in an animal tumour model and it was found that higher concentrations of porphyrins in the blood can be used as a measure of cancer severity [61, 65]. In addition, Ramu Rajasekaran [66] reported that fluorescence emission spectral properties of urine at 405 nm excitation, which is responsible for PpIX, provide a statistical difference between patients with cancer



(Head and Neck, Cervix and Breast cancer) and normal subjects. He studied native fluorescence properties of human urine specimens based on excitation-emission matrices (EEMs) using a range of excitation and emission wavelengths. As can be seen from Figure 11, the fluorescence spectra of urine samples of cancer patients demonstrate significant spectral differences in both EEMs and emission spectra from normal subjects.



**Figure 11. The excitation-emission matrix of undiluted urine samples: (a) normal, (b) cancer, wavelength ranges at 250–450 nm for excitation and 270–750 nm for emission[66].**

## 2.10 SCREENING

Over a few past decades, autofluorescence has been evaluated in the detection of various cancer and neoplastic tissue. In the following, the application of autofluorescence in characterising several cancers and neoplastic tissue is described.

Autofluorescence has been widely used to characterise breast cancer which is a major cause of cancer deaths among women [67].

In 2003 Paul J Tadrous et al [68] used FLIM to characterise human breast tissue at a histological resolution in unstained tissue samples. They used alcohol-fixed tissue samples from 13 patients stimulated by laser pulses at 415 nm. They measured the decay rate and compared average lifetimes of different tissue regions. They reported a significant difference between benign stroma, malignancy-associated stroma, blood vessels, and malignant epithelium ( $p < 0.05$ ) [68].

In 2004 Tara M Berslin [69] performed autofluorescence measurements on 56 samples of tumours or benign breast tissue. They used a support vector machine algorithm and compared classification results with the histological diagnosis (gold standard). In their report, several excitation wavelengths and diffuse reflectance spectra demonstrated significant differences between tumour and benign tissues. They got a sensitivity of 70.0% and specificity of 91.7%. In 2010 Matthew D. Keller et al. [70] employed spectral measurements to evaluate the surface of the tissue mass from 12 patients. They could achieve 85% sensitivity and 96% specificity after comparing their results to histopathology assessment. In 2012 Vikrant Sharma [71] investigated the use of auto-fluorescence lifetime measurement (FLIM) for detecting invasive ductal carcinoma (IDC) in human ex vivo breast specimens. They utilised excitation at  $447 \pm 60$  nm with several emission wavelengths ( $532 \pm 10$ ,  $562 \pm 40$ ,  $632 \pm 22$ , and  $644 \pm 24$  nm). They evaluated auto-fluorescence lifetimes data for six specimens and could achieve  $92.3 \pm 0.8\%$  accuracy.

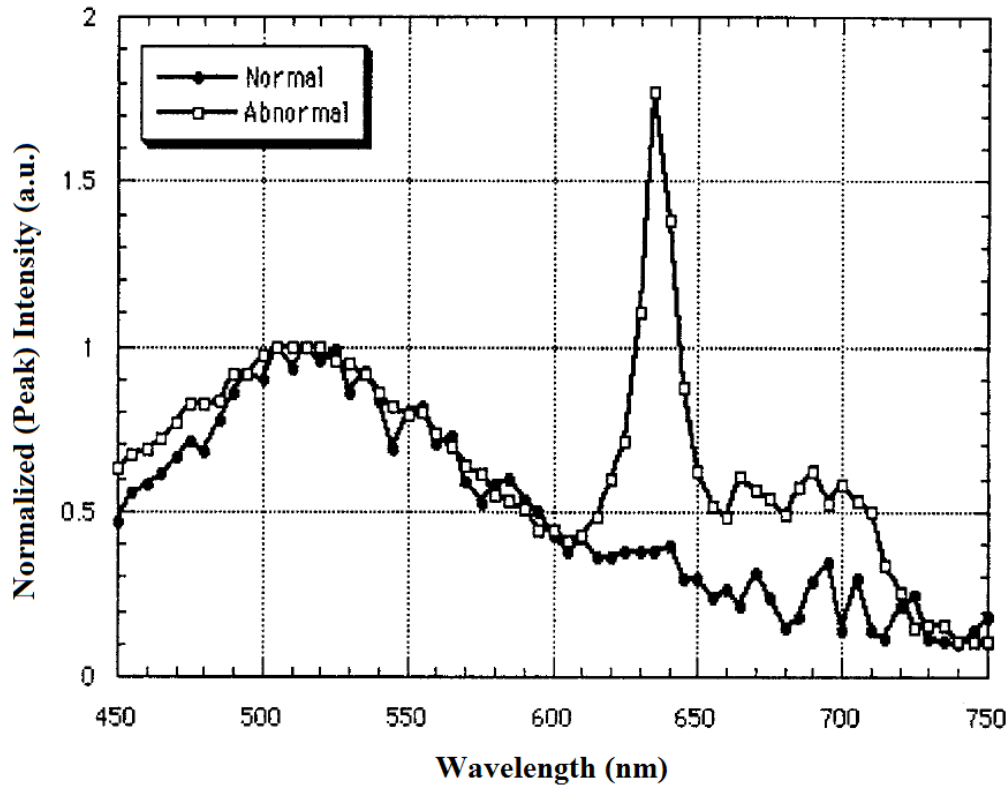
In 2015, Rodrigo de Andrade Natal[72] analysed sections obtained from 14 patients with invasive ductal breast carcinoma. They used FLIM to monitor NAD(P)H and FAD fluorescence lifetime using 405nm excitation and with two emissions,  $445 \pm 45$  and  $535 \pm 22$ . They demonstrated that breast cancer cells in contact with desmoplastic reaction showed a significantly lower NAD(P)H and FAD fluorescence lifetime. Furthermore, the optical redox ratio was also lower in these tumour cells[72]. In 2017, Jennifer E Phipps [67] obtained FLIM data from breast cancer specimens (*N*

= 20) to detect breast specimens. They used excitation at 355 nm and collect the emission using 4 channels at 390nm, 466 nm, 542nm and 629/53 nm. They achieved about 97% accuracy for automated detection of cancerous, fibrous, and adipose tissue from breast cancer specimens. Ivana Pakova , et al. [73] employed autofluorescence technique to characterise the structure of blood plasma and mammary gland homogenates. They recruited patients at three different breast cancer stages with respect to healthy specimens. They demonstrated that the blood plasma and homogenate of patients with breast cancer have a significant difference in autofluorescence values in comparison with normal cases. Such results may serve as fast preliminary markers of cancer detection[73].

Colonic cancer has also been evaluated based on autofluorescence concept in several reports [74]. Bu-Hong Li et al.[42] obtained excitation-emission spectra of normal and adenomatous colonic tissues to explore auto fluorescence spectroscopic differences. Five patients were recruited in their study. They used excitation wavelengths ranging from 260 to 540 nm and collected emission spectra from a range starting 20 nm above the excitation wavelength and ended to 800 nm. In comparison to normal tissues, lower NAD (P)H and FAD, higher amino acids and protoporphyrin IX were seen in their results[42]. Li et al. used in ex vivo tissue and demonstrated that longer auto fluorescence lifetimes were presented in cancerous colonic tissues than in normal colon. They used excitation wavelength at 397nm and collected emission at 635nm, which was related to the contribution of protoporphyrin IX (PpIX) [75]. Moreover, statistically significant FLIM ( excitation at 355 nm) results were seen between cancerous and healthy colon tissue from 16 unstained surgical specimens of colon cancer[76].

The hypothesis that autofluorescence is capable of differentiating normal from diseased oral tissue has been evaluated by several studies[37, 77]. Duncan et al. [63] employed biopsy specimens from patients to investigate the differences between normal and dysplastic tissues using spectroscopic measurements. They acquired fluorescence spectra from twelve histologically normal, including healthy mucosa or benign lesions, and ten abnormal, including dysplastic or malignant, tissue samples. They observed a significant spectral difference between normal and abnormal groups. Such differences were noticeable at the excitation wavelength of 410 nm and emission of 600 nm to 750 nm as shown in Figure 12. They demonstrated the capability of









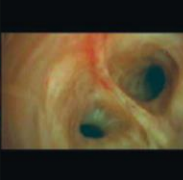
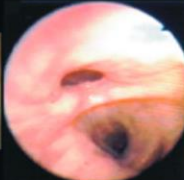

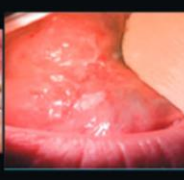
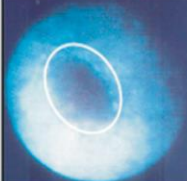
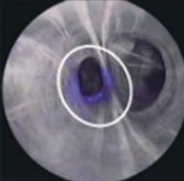

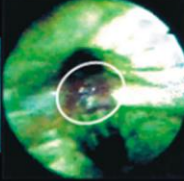
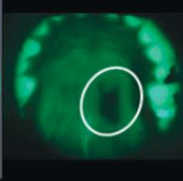
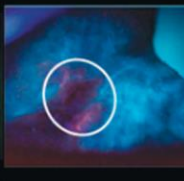
autofluorescence to distinguish normal and dysplastic mucosa in vitro with a high degree of accuracy.



**Figure 12. Fluorescence properties of normal and abnormal (cancerous) oral mucosal tissues with 410 nm excitation wavelength [63].**

Recently, Sun et al. used a FLIM system consisting of a rigid fibre-bundle endoscope providing a 4mm field of view. 10 patients were recruited for imaging of head and neck squamous cell carcinoma. The imaging was performed at 337-nm excitation, and fluorescence was collected in the 435- to 485-nm. They showed that head and neck squamous cell carcinoma exhibited a shorter average lifetime ( $1210 \pm 40$  picoseconds) than the surrounding normal tissue ( $1490 \pm 60$  picoseconds) [78].

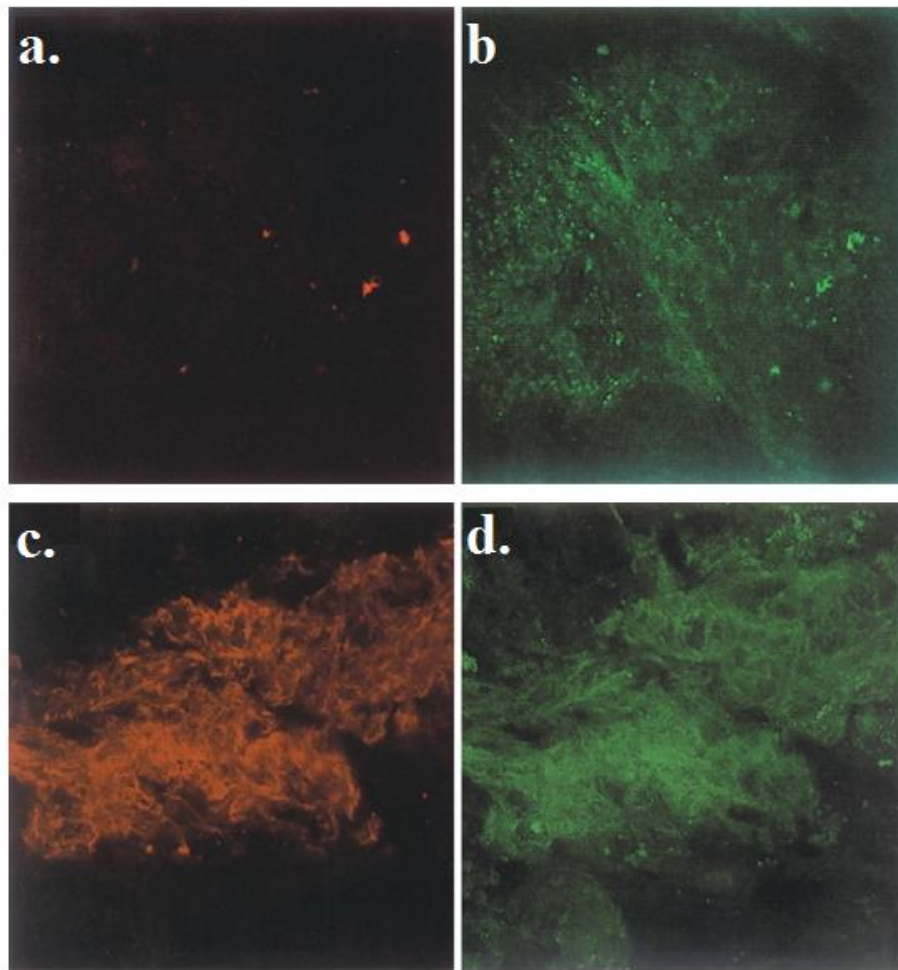
To detect the neoplastic transformations related to head and neck, various autofluorescence imaging modalities are available (see Figure 13) although they have not been particularly designed for head and neck region.

	LIFE system	DAFE system	SAFE system	D-Light-AF system	VELscope	IdentaFi
Instrument						
Dx	Colonic adenoma	Bronchial carcinoma	Bronchial carcinoma	Bronchial carcinoma	Salivary gland tumor	Tongue SCC
Brightfield						
Autofluorescence						

**Figure 13. Available auto fluorescence imaging instrumentation to detect the neoplastic transformation and associated images illustrating the tumour detections[1].**

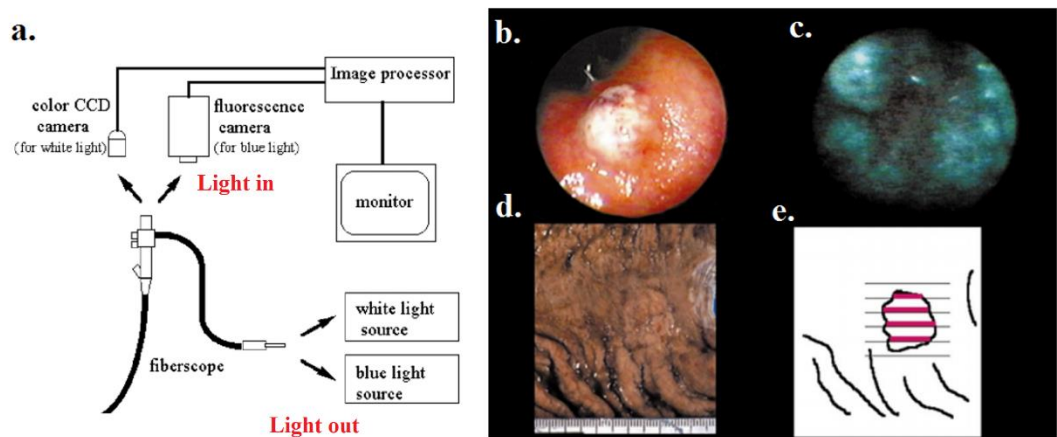
Cancer characterisation using autofluorescence was not limited to the above-mentioned cancer types. Mengyan Wang et al. employed a two-channel laser scanning confocal microscope to obtain autofluorescence images of lung tissue samples. They evaluated autofluorescence images to characterise the cellular morphology and tissue structure of lung cancer. They demonstrated that most of the patients showed discriminating fluorescence in tumour tissues compared to normal tissues. They also reported that most of the samples have lower NADH/FAD values and higher 630 nm/590 nm values in cancerous tissues than in normal tissues[79].

Autofluorescence imaging was also evaluated to analyse gastric cancer[80]. Shu Dong et al. employed a double-channel laser scanning confocal microscope with two excitation wavelengths, 488 nm and 543 nm, to acquire autofluorescence from 16 gastric cancerous tissue samples and corresponding normal gastric tissue [80]. They could catch a discriminative autofluorescence signal in gastric cancer tissue which appeared as a reddish-brown image (see Figure 14).



**Figure 14. Autofluorescence image of normal and cancerous gastric mucosa. (a) Red autofluorescence image of normal gastric mucosa. (b) Green autofluorescence image of normal gastric mucosa. (c) Red autofluorescence image of cancerous gastric mucosa. (d) green autofluorescence image of cancerous gastric mucosa[80].**

Autofluorescence technology was utilised in a light-induced fluorescence endoscopy system (see Figure 15 (a)) to detect early gastric cancer in a real-time manner. Using this technology, 85% of the cancer lesions could be detected with 94% and 86% sensitivity and specificity, respectively. Figure 15 (b-e) demonstrates the comparison between a white light endoscopic image, an autofluorescence image and a corresponding histology assessment[81].



**Figure 15. (a) Light-induced fluorescence endoscopy system diagram. (b) White-light endoscopic image of early gastric cancer. (c) Corresponding autofluorescence image which shows a dark red pattern. (d) Resected picture of the same site. (e) Sectioning schema; red lines demonstrate the tumour invasion [81].**

Multi spectral imaging has been also employed to detect skin lesions[82]. Uncontrolled development of abnormal cells in skin cancer may change the spectral features to identify skin cancer lesions. Multispectral imaging systems can produce precise quantification of spectral and spatial features, which have been used over the last several years to extract spectral and colorimetric properties of the skin reliably and non-invasively [83]. Jakovels et al. [84] employed wavelength range from 450 nm to 950 nm to recognize melanoma. Delpueyo et al. [85] also introduced a light-emitting diodes (LEDs)-based multispectral imaging system utilising eight different wavelengths (414–995 nm) and has reported they could successfully improve the detection of skin cancer lesions.

## 2.11 CONCLUSION

Autofluorescence concept has terrific capabilities of cancer characterisation, diagnostic and screening. Technologies employing autofluorescence are able to observe chemical composition change of the cells and tissues due to diseases onset and

development such as neoplastic or cancer metabolism. The main advantage of the autofluorescence techniques is that they are non-invasive. So, such techniques potentially can minimise biopsy incidence. Autofluorescence imaging is fast and can inspect a large area of tissue so it can be used for margin assessment and cancer boundary delineation in vivo and on a real-time basis. Therefore, autofluorescence imaging can be utilised for both cancer detection and cancer boundary delineation.

Autofluorescence has capabilities to be merged with existing and new technologies such as robotic surgery in which low-light environment gives an optimal prospect for autofluorescence application. Finally, collaboration in a multi-disciplinary area, including engineers, physicist, radiologist surgeons and the pathologist can lead to discovering more capabilities of autofluorescence technology for the more advanced application. The gap in the literature regarding autofluorescence imaging is absence of a user-friendly system with flexible excitation and emission band selection which is considered in this study.



# Chapter 3: Research Design

---

This chapter describes the design adopted by this research to achieve informative autofluorescence spectral images to characterise cancerous tissue or cells. Section 3.2 discusses the customised microscope used in this study and its associated details, including spectral channels, camera and the control system, described in section 3.3, 3.4, 3.5 and 3.6. Dishes and slides used in this study are described in section 3.7. Details related to the imaging protocol is described in 3.8. Image preparation importance in spectral images is described in section 3.9, and also the specific systematic noise involved with spectral images are described in 3.10, 3.11 and 3.12. In addition, the random noise which may corrupt the spectral images and relevant techniques to remove them are described in section 3.13. After pre-processing the spectral images, image segmentation is described in 3.14. Finally, section 3.16 discusses the ethical considerations of the research and its problems and limitations.

## 3.1 INTRODUCTION

Auto-fluorescence microscopy can play a pivotal role in the biology and biochemical fields such as cancer science by providing information about micro-organism, biostructure and cell environment in a non-invasive manner[86]. Various scientists have progressed auto-fluorescence application to biology, but lots of different aspects have still not been explored[87, 88]. In addition, there is still much room to make progress in the field concerning new imaging technology, new image processing and different ways to extract valuable information from auto-fluorescence related signals, especially in cancer science.

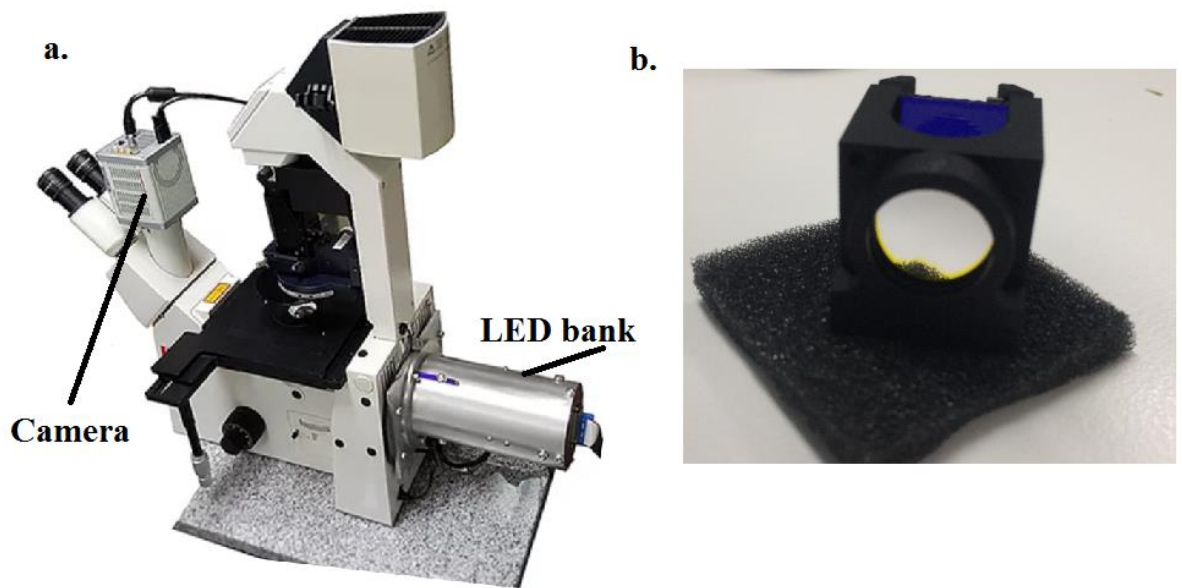
Extracting valuable information from autofluorescence spectral images requires to develop a particular procedure due to the presence of various types of noises[89]. Such noises may hide the informative part of the spectral signal and corrupt the extracted information. Besides, autofluorescence in single-cell resolution may not be a solid signal, so a meticulous methodology should be employed to acquire reliable information[9]. Thanks to the modern technology in the camera industry, and customising current microscopic technology, a new methodology is used in this research study to extract information from the autofluorescence properties of cells and

tissues[9]. In addition, sophisticated image pre-processing techniques were deployed to prepare spectral images for various computational analysis and decision making.

### **3.2 CUSTOMISED FLUORESCENCE MICROSCOPE**

The central part of the multi-spectral system used in this study is a fluorescence microscope. Such a microscope uses the fluorescence concept along with typical optical devices including scattering, reflection lights and proper light sources and sensors. These kinds of microscopes have capabilities to extract fluorescence information to form different spectral images. In this project, a customised system previously developed in our group was used [9], however with new modifications regarding new protocol, filters and light source arrangement.

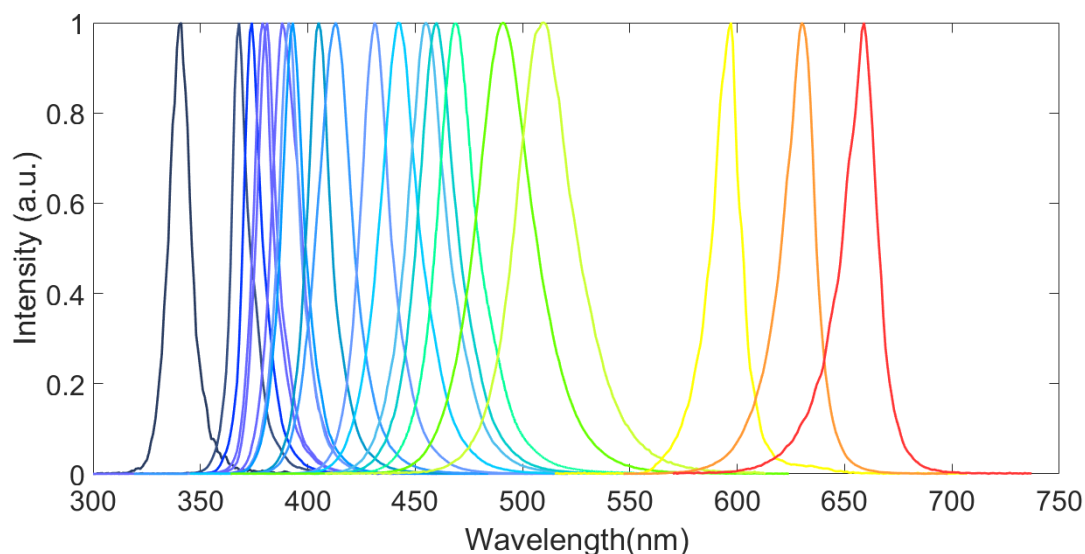
Our microscope technology includes a Leica microscope, which has some specific and distinctive optical features. First, an EMCCD camera can be simply fixed on the microscope, as there are multiple input and output ports in the microscope. Several rear ports were available which enable designing a proper illumination system. The microscope has an external power supply to facilitate better thermal stability. The microscope has a mechanical wheel which can house four different filter cubes. Compared to confocal systems with several channels provided by high-power lasers, the price of our customized system was more cost-effective as the system uses the LED source to light the sample instead of high-power lasers. Figure 16 (a) shows the system used in this study.



**Figure 16. Customised microscope employed in this study: (a) the microscope[90], (b) a sample filter cube**

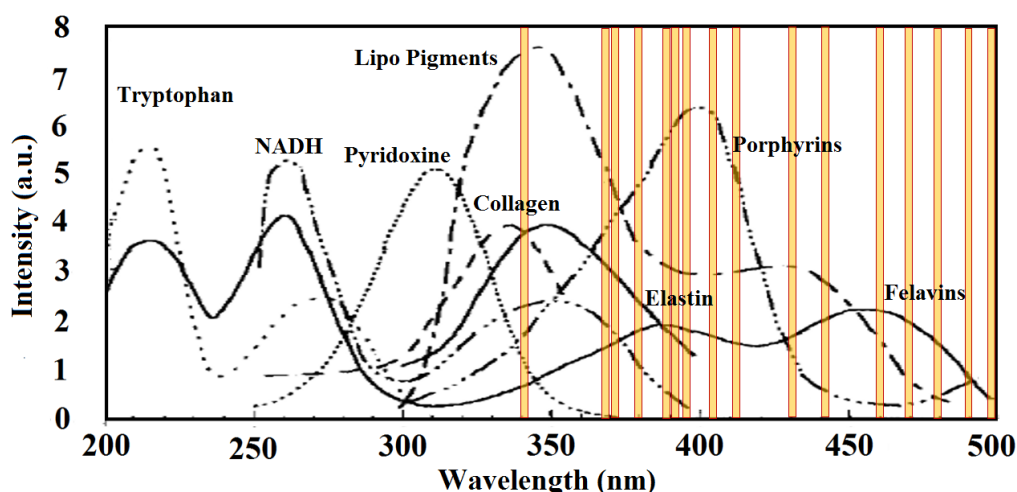
### **3.3 EXCITATION SOURCES**

Lamps using light emitting diodes are called LEDs. New technology in the LED industries results in advanced LED lamps[91]. These lights have a more abundant lifespan, up to 25 times than of typical lights, and are cost effective. These lights are much more efficient in consuming energy with a very limited time required for warming up and stabilisation. LEDs with advanced technologies have found their ways for more professional applications like medical instrumentation [92]. LEDs can be manufactured with a very small bandwidth and can be manufactured with various centre wavelength from UV range to close to red. So, they can give a big wavelength choice to users and also can reduce the need for costly filters to get a specific wavelength. The output power of the LEDs are usually in tens of milliwatts (mW) which makes them proper choices for critical medical applications such as ophthalmological equipment. This research has used multiple LEDs as excitation sources, each of which has a specific wavelength. A typical microscope was advanced by adding more excitation light to make the system capable of extracting more information from biological materials [90]. Figure 17. shows the spectra of the LEDs used in this study.



**Figure 17. Excitation wavelengths used in this study**

To show which fluorophores potentially get excited by wavelengths used in this study, the centre frequency of each LED is shown on various excitation fluorophore graphs (see Figure 18). This graph blends the LEDs and cell fluorophore and can show which excitation wavelength can excite what fluorophores at the same time, and also it represents the possible fluorophores that can be excited using all of the available LEDs.

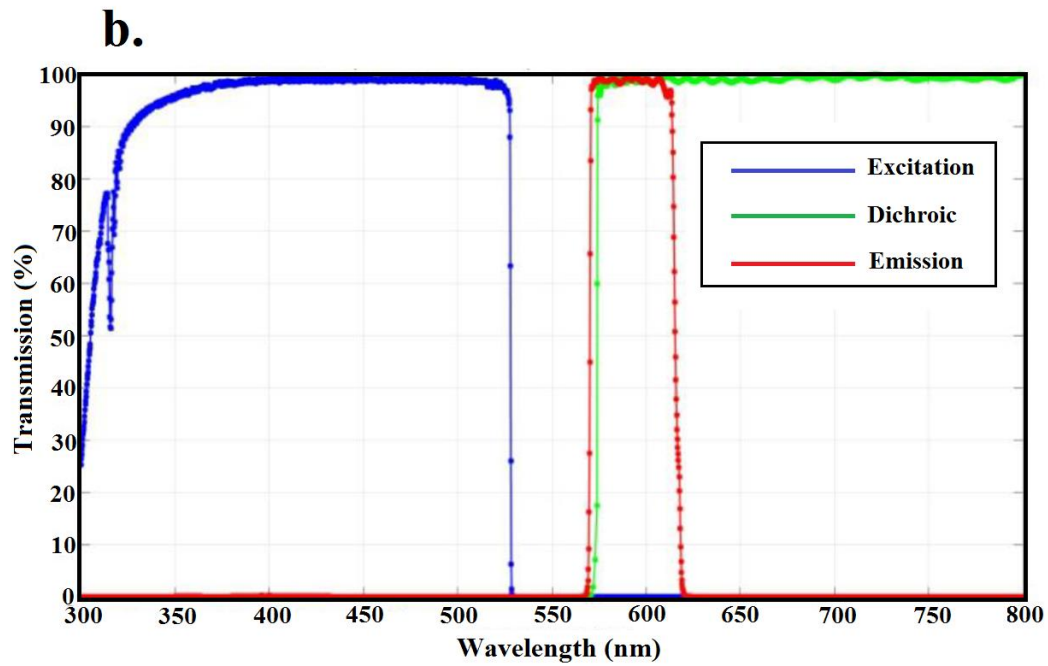
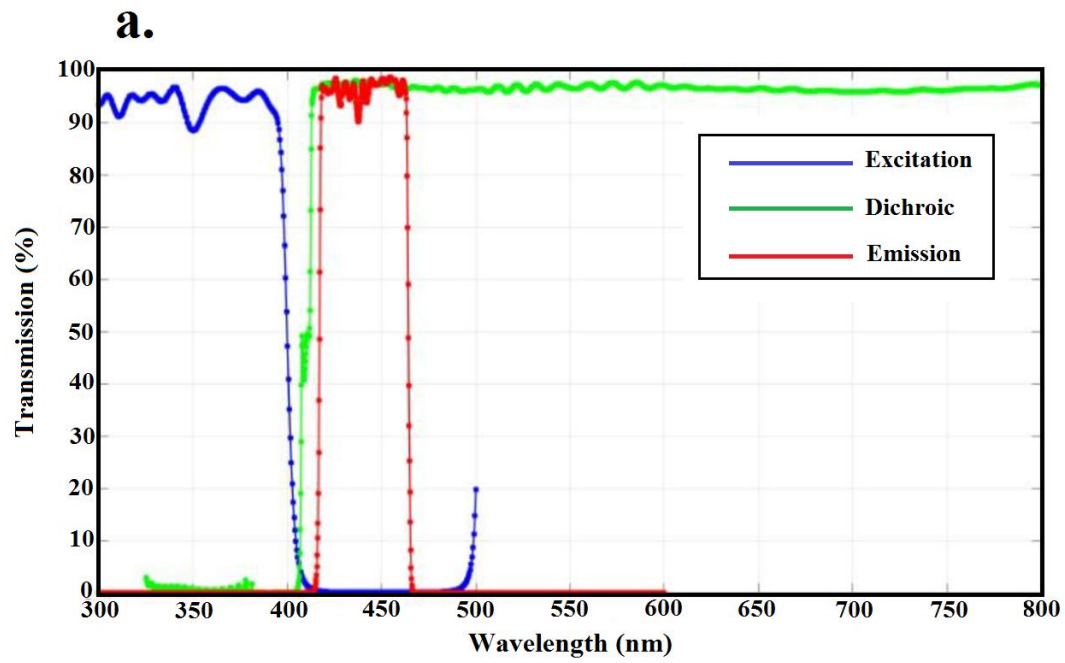


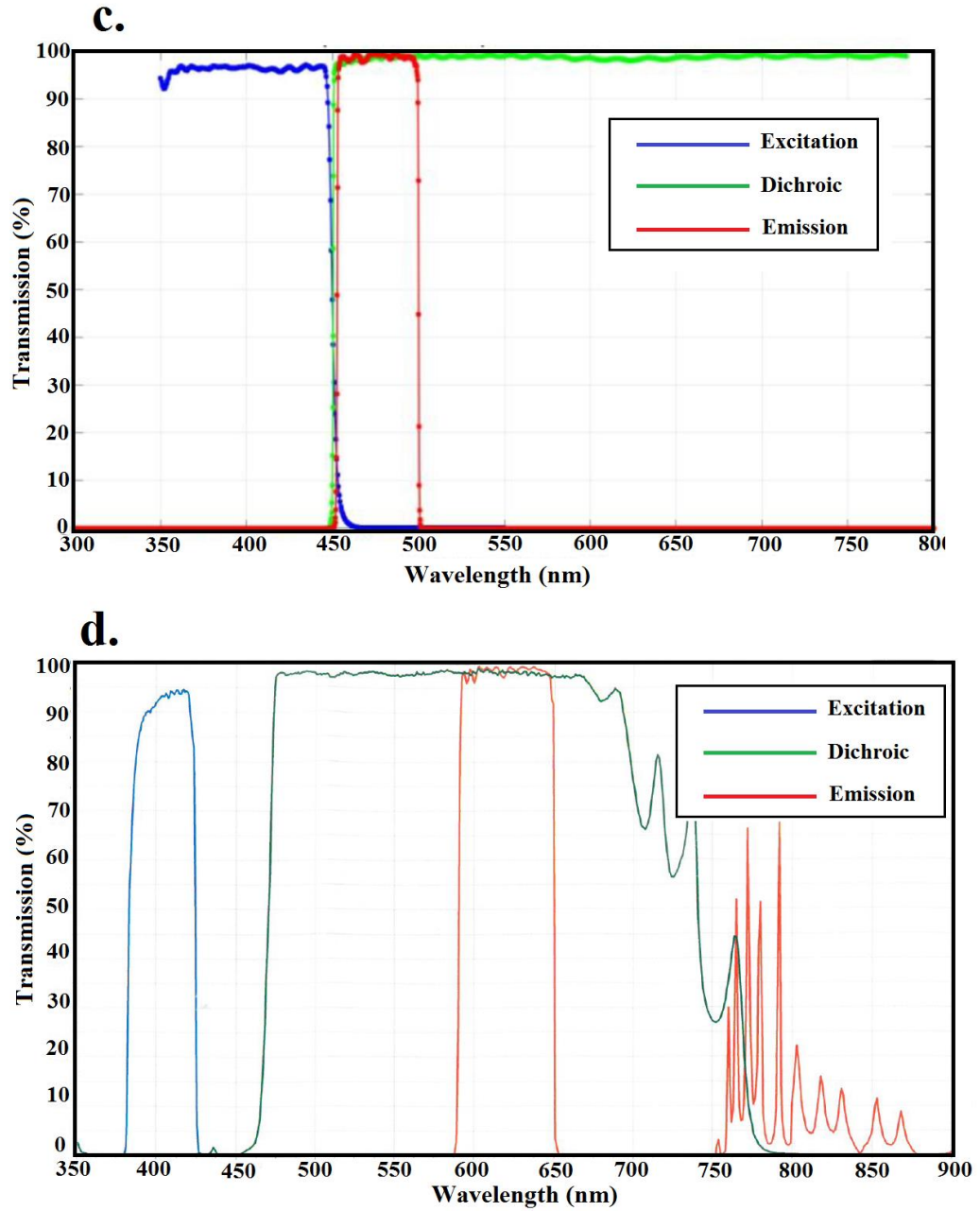
**Figure 18. Centre wavelength of LEDs used on various excitation fluorophore graph. (the original figure was borrowed from [20] ). Further information about the channels created by excitation wavelengths described here is listed in table 2.**

It is worth mentioning that Beer's law applies in the samples with the finite thickness. The Beer-Lambert law is describing the relation between the attenuation of light and the properties of the sample through which the light is traveling[93]. In this study, the samples have almost the same thickness and the emission light is collected from the same side where it was excited. Therefore, the effect of the Beer's law is the same for all of the samples used in this study. Moreover, the autofluorescence signals are collected while the sample is illuminated by the LEDs, so there is no effect of fluorescence decay interfering our results.

### **3.4 EMISSION BANDS**

The customised microscope uses four filter cubes to manage the excitation and emission wavelengths. A sample filter cube used in this study is shown in Figure 16 (b). Each filter cube contains three optical filters including an excitation filter, a dichroic beam splitter and an emission filter. Although the excitation wavelength associated with each LED is narrow, the excitation filter is used to make the excitation light more restricted. The dichroic is responsible to split the light, reflect the excitation wavelength to the sample and block the excitation light from reaching the detector. The emission filter confines the light coming from the sample to a specific wavelength band. Different microscopes can hold a different number of the filter cubes. In the current research, four filter cubes were used to generate the spectral images. These four filter cubes were mounted on a mechanical wheel which could be turned manually. A spectral image formed by the combination of a specific excitation wavelength with a specific filter is called a channel in this study. This control provides a specific colour to be captured in each channel with defined excitation and emission wavelength. Figure 19 shows the cut-off wavelength for the excitation, dichroic and emission for each filter cube. The band pass region of the emission filters is in the range of ~50-100 nm to collect enough signal. Narrowing further down the region may have adverse effect on the subsequent analysis due to lack of informative signal.



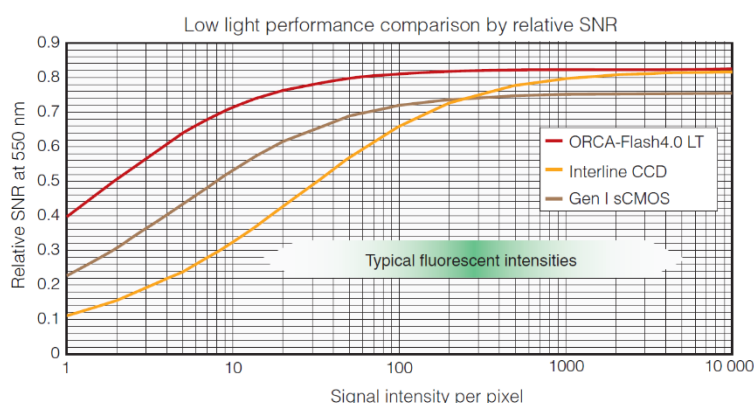


**Figure 19. Filter cube specifications used in this study: (a) filter cube 1 (b) filter cube 2 (c) filter cube 3 (d) filter cube 4. Further information about the channels created by excitation wavelengths described here is listed in table 2.**

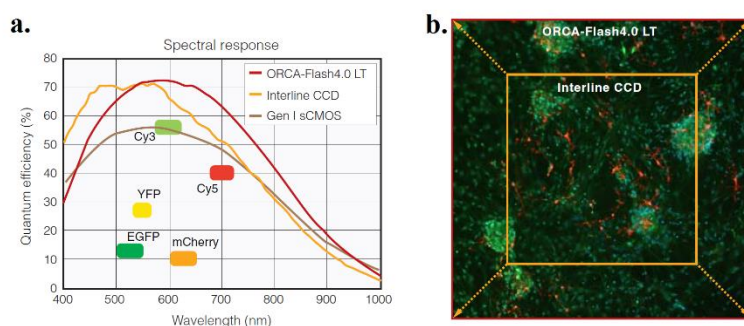
### 3.5 CAMERA

Another key part of the microscope used in this study is the camera which captures the signal coming from the samples. Its performance directly influences the quality of the images and subsequently, the information extracted from the images. In this research, a digital camera manufactured by Hamamatsu Company named ORCA-

flash4.0lt was used. It was a low noise camera with high quantum efficiency, which could produce high-quality images valuable for quantitative analysis. ORCA-flash4.0lt performance surpassed the performance of Gen I sCMOS and interline CCD cameras at different input levels. Such attribute helps the camera to have a strong signal to noise ratio even with a limited exposure time compared to other cameras. Figure 20. illustrates the relative SNR of the ORCA-flash4.0lt compared to other cameras at 550 nm [94].



**Figure 20. ORCA-flash4.0lt performance**



**Figure 21. ORCA-flash4.0lt performance in terms of quantum efficiency and field of view (a) quantum efficiency of the camera for different wavelength range (b) comparing the field of view of the ORCA-flash4.0lt and interline CCD [94]**

Figure 21(a) shows the quantum efficiency of the camera for various wavelength range. Quantum efficiency is the effectiveness of the camera to capture photons and convert them to signals. The ORCA-flash4.0lt shows overall better performance than other cameras. The maximum quantum efficiency of ORCA-flash4.0lt is about 600 nm and is about 70%. The camera shows acceptable performance for all of the filters



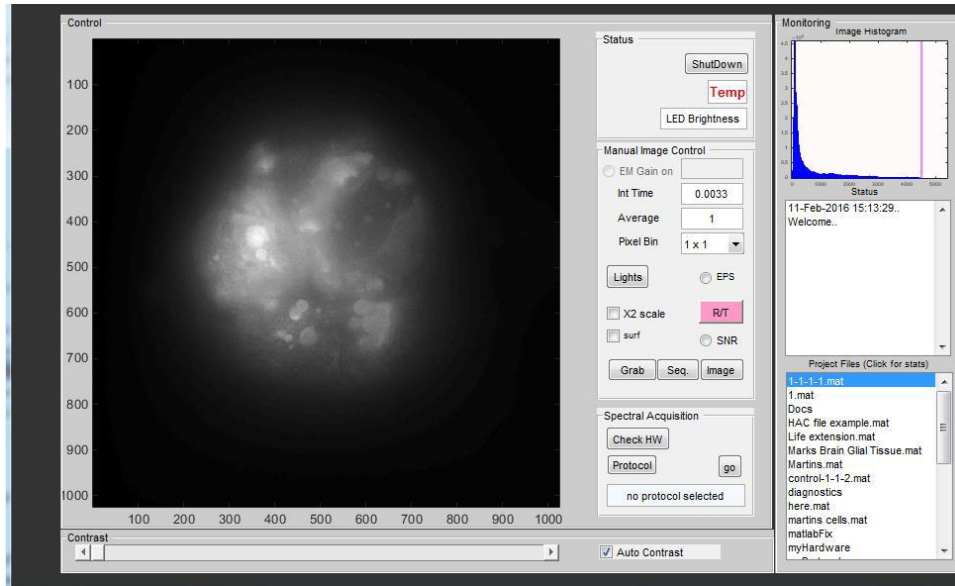
used for this research project. Figure 21(b) compares the field of view for the ORCA-flash4.0lt to that of a CCD camera. It shows that ORCA-flash4.0lt can capture almost three times as large an area as the CCD camera. The larger field of view helps to capture more cells or image a larger area of tissue in one shot. Other important specifications of the camera are presented in Table 2.

**Table 2. Detailed ORCA-flash4.0lt performance and specifications[94].**

Imaging device	Scientific CMOS sensor FL-400
Effective number of pixels	2048(H) × 2048(V)
Cell size	6.5 $\mu\text{m}$ × 6.5 $\mu\text{m}$
Effective area	13.312 mm × 13.312 mm
Full well capacity (typ.)	30 000 electrons
Readout noise (typ.)	0.9 electrons (median), 1.5 electrons (rms)
Readout noise (Rapid rolling mode) (typ.)	1.3 electrons (median), 1.9 electrons (rms)
Dynamic range (typ.) <sup>*1</sup>	33 000:1
Quantum efficiency	Over 70 % at 600 nm
Cooling temperature	+10 °C (Ambient +25 °C)
Dark current (typ.)	0.6 electrons/pixel/s
PRNU (Column fixed pattern noise <sup>*2</sup> )	≤ 0.1 %
MTF (Theoretical)	76.9 lp/mm
Frame rate <sup>*3</sup>	
Full resolution	30 frames/s
2048 × 1024 <sup>*4</sup>	60 frames/s
2048 × 8 <sup>*4</sup>	7696 frames/s
512 × 8 <sup>*4</sup>	25 000 frames/s (Rapid rolling mode only)
AD conversion <sup>*5</sup>	16 bit

### 3.6 SYSTEM CONTROL

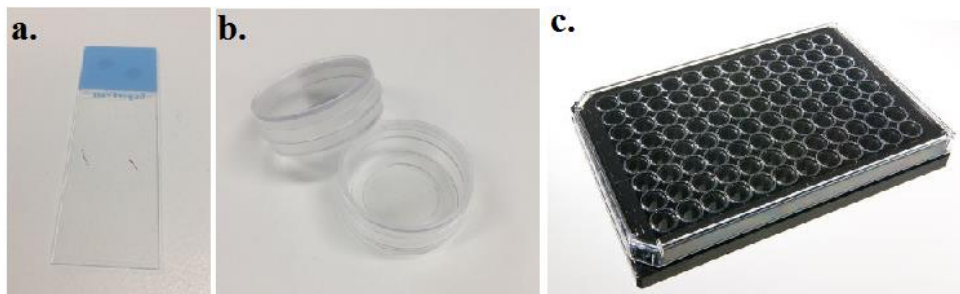
Synchronising the multispectral system was done via custom-made software written in MATLAB [90]. Library functions provided by the camera provider were used to control the camera. To control the LED bank, a digital I/O card with 24 channels was employed. Using MATLAB software, a graphical user interface (GUI) was developed which provided complete control of the camera and LED bank to activate them one by one. After defining a protocol to set operational parameters such as exposure time, excitation wavelength, and emission filter, the user can run the GUI to acquire the images. Figure 22 shows the GUI interface used in this study.



**Figure 22.** The graphical user interface (GUI) used in this study to control the camera and LED bank[90].

### 3.7 DISHES AND SLIDES FOR IMAGING

This research project involved both tissue and cell characterisation. Tissues were processed at pathological laboratories and were placed on optical microscope slides (Figure 23 (a)). About the cells, after being cultured, cells were transferred into appropriate microscope dishes. Dishes used in this study were petri dishes (Figure 23 (b)) and 96 well dishes (Figure 23 (c)). A specialised petri dish was also used to find the focus point of the microscope. To prepare such dishes, 10×10 grids were etched at the bottom of the dishes. Each grid was square with a 50  $\mu\text{m}$  edge.



**Figure 23.** Slides and dishes used in this study: (a) A sample, (b) a sample petri dish, (c) a sample 96 well dish.

### 3.8 IMAGING PROTOCOLS

Combining the excitation wavelengths to the emission filters led to define a number of channels, each of which forms a spectral image. However, to define a channel effectively, we need to consider the “Stokes shift” of the native fluorophores. As a case in point, defining a channel with an excitation less than 360nm and emission more than 700 nm is almost useless as there are no natural fluorophores which can glow with those specific excitation and emission wavelengths. In this study 38 spectral channels have been used as shown in Table 3.

**Table 3. Spectral channels used in this study**

Channel No.	Excitation wavelength	Emission wavelength	Channel No.	Excitation wavelength	Emission wavelength
1	340	420-460	20	382	573-613
2	368	420-460	21	388	573-613
3	373	420-460	22	391	573-613
4	378	420-460	23	394	573-613
5	382	420-460	24	405	573-613
6	388	420-460	25	413	573-613
7	340	454-495	26	432	573-613
8	368	454-495	27	441	573-613
9	373	454-495	28	455	573-613
10	378	454-495	29	460	573-613
11	382	420-460	30	470	573-613
12	388	454-495	31	491	573-613
13	391	454-495	32	510	573-613
14	394	454-495	33	382	575-650
15	405	454-495	34	388	575-650
16	340	573-613	35	391	575-650

17	368	573-613	36	394	575-650
18	373	573-613	37	405	575-650
19	378	573-613	38	413	575-650

### 3.9 IMAGE PREPARATION AND PROCESSING

The imaging system employed in this study provides both spatial and spectral information simultaneously and generates a data file called a data block. The data block is a rich three-dimensional data file, so it may potentially encompass more diagnostic information. On the other hand, such a significant volume information also causes complexities in implementing analysis and processing of spectral data[89]. Therefore, to extract valuable spectral and spatial information, sophisticated image processing and precise methods are required. In this section, the pre-processing steps leading to images ready for extracting diagnostically valuable information are described.

To quantify an image, it is a must to pay attention to the presence of erroneous data values such as dead or saturated pixel and/or non-informative background [9, 89]. Otherwise, further processing is misleading, as such measurement distorts the conclusion and so leads to an unreliable identification. The abnormal observation may have various sources; however, the most prominent bases resulting in anomalies in spectral images are instrument and radiation. Furthermore, the camera, which is based on diode array detectors, may be the source of the different potential errors. In fact, the malfunction of one of the diodes in the detector array will result in dead pixels (missing or zero values) or saturated values. Moreover, the illumination of the sample ideally should be uniform, but in the real world this is impossible and the illumination profile has an uneven curvature [95].

Each of the mentioned examples can be a source of error distorting the final results [89]. Consequently, pre-processing the image is a mandatory step toward obtaining reliable results. In this study, pre-process the medical multi-spectral imaging carefully is investigated and an appropriate algorithm is developed. To prepare the spectral images, it is required to acquire some supporting images. Two reference images were taken at the beginning of each experiment, including calibration and

background images. Using these images different steps were taken to prepare all of the spectral images called pre-processing steps. This algorithm involves cosmetic ray removal, smoothing, background subtraction and microscope calibration, image equalisation and region of interest segregation. Depending on the nature of the analysis, it is possible that some steps were omitted without affecting the outcome.

### **3.10 CALIBRATING THE IMAGES**

To illuminate a sample, the fluorescence microscope utilises various LEDs, each of which possesses a specific wavelength and illumination profile. Ideally, the output light of the LEDs should be identical at all wavelengths, irrespective of the sample under examination. However, the imaging system often displays significant signal variations. Such non-uniformity of the illumination can be affected by various factors and so needs to be corrected.

To calibrate the microscope, a calibration fluid and a precise Fluorolog spectrometer were needed. The calibration fluid was produced which was a mixture of chemicals which together had fluorescence properties in all the spectral channels. This fluid was a mixture of several chemicals, including NADH, PPIX and riboflavin, depending on the channels which might be used. On the other hand, the mixture was chosen such that there is no zero response for all of the channels used in the microscope. The spectrum of calibration fluid was measured first by a Fluorolog spectrometer as the reference instrument. The Fluorolog spectrometer is a Cary Eclipse Fluorescence Spectrophotometer™, which has been calibrated periodically in our laboratory by certified technicians and produces reliable results. Then, the same fluid was used to take spectral images using the microscope.

As it was mentioned above, these two spectra should be ideally identical as they are taken from the same fluid with two different instruments. However, they were not due to potential different illumination profile of each LED used in our customised system. Consequently, the microscope was calibrated based on Fluorolog spectrometer values. To achieve that, firstly, the Fluorolog spectrum was normalised such that the values add to one. These values were then used to form a set of correction calibration images named calibration vector. To calibrate all of the spectral images, the sample image was divided by the image acquired from the calibration fluid. Next, the

correction vector at each pixel position is multiplied the image value to correct and calibrate the output of the hardware.

### 3.11 REMOVING THE BACKGROUND

In the multispectral imaging, a non-informative observation that is inconsistent with the whole data set is called background. These corruptive signals are mostly caused by ambient lights, and residual autofluorescence from the microscope slide, petri dish, objective or internal microscope elements[89]. To remove background superimposed on a spectral image, an image from water (or medium) which has no fluorescence signal in calibration sample petri dishes were taken and subtracted from the HIS image. Careful attention is needed while getting the focus point to take the water image, as it directly affects the accuracy of the results. Previous researchers proposed that the focal point with five minor units of the microscope scale upper than inside dish surface is a proper point to take a background image [9].

### 3.12 IMAGE EQUALIZATION

The multi-spectral cube has various spectral channels generated at various wavelengths. Depending on the wavelength, the pixel value which the camera catches is very different[9]. In addition, taking the images from different systems may result in different pixel values due to different specifications of the cameras. To minimise such challenges, an image equalisation formula was used to convert the intensity unit to photons per pixel per second (PPS). To achieve this, firstly, a quantum efficiency graph associated with the camera should be obtained. As Figure 21 (a) illustrates, quantum efficiency shows the relation between the number of photons and the pixel value recorded for each pixel. To convert intensity to PPS the following equation was used:

$$Y_{PPS}(i, j) = \frac{(Y_{intensity}(i, j) - Bo) \times Se}{Qe \times Ex} \quad \text{Eq. ( 4)}$$

where  $Y_{PPS}(i, j)$  is the pixel value converted to PPS in position  $(i, j)$  in the image,  $Y_{intensity}(i, j)$  is the image intensity in the in position  $(i, j)$  in the image. 'Bo' is the "dark count" which is the amount of electrons that are randomly captured into the corresponding wells of CCD pixels without the presence of any photons. 'Se' is the sensitivity which describes how many electrons correspond to the smallest graduation

of a pixel value. ‘ $Ex$ ’ is the exposure time to minimise the exposure time’s influence on the associated number of photons.

### **3.13 IMAGE NOISES**

It is crystal clear that experimental signals in the real world come with noise. This noise can be negligible if the signal to noise ratio is high enough, but generally speaking, noise corrupts the signals and must be reduced in order to proceed further with data analysis. The processing of signals to remove noise mostly is called denoising or smoothing [89].

Auto-fluorescence signals are quite weak. Consequently, spectral images are susceptible to the presence of noise. In addition, autofluorescence equipment utilises a small channel width, so the image sensor cannot capture much energy, and self-generated noise inside the sensor happens. Moreover, light variation and environment aberrations decrease the amount of spectral signal which is captured. Consequently, denoising is an essential part of autofluorescence image processing before any further investigation [11, 95].

#### **3.13.1 Removing spikes (cosmetic ray)**

A rapid rise followed by a sharp drop in a small area of a spectral image is named a cosmetic ray (see Figure 24 (a)). Such abnormalities can be due to the irregular behaviour of the detector, failures of electronic circuits or environmental circumstances. These cosmetic rays may mask the informative details of the spectral channel and may cause us to misinterpret the spectral information[9, 89]. In order to identify cosmetic rays, a typical solution is manual supervision, yet this method needs human care which requires lots of time. A solution is even tougher when multispectral data are considered since a multispectral data cube encloses many spectral channels and manual observation is not a possible approach. In this study, an automatic method for cosmetic ray removal was utilised which has two steps, cosmetic ray localisation and replacement.

Cosmetic rays are formed due to deposition of high amounts of energy deposit in a few pixels, leading to saturated pixels with impulsive variations. This effect can be inspected in the histogram distribution of a spectral channel. Points positioned in the high count interval, and somewhat compact, in an image histogram can be

considered as Cosmetic rays[46]. To detect cosmetic rays, an algorithm developed consisting of the following steps:

- Select small sub-frames which cover the whole frame with some degree of overlap.
- Calculate the standard deviation of the pixel distribution.
- Generate a histogram of the distribution of counts.
- Catch the mode of the histogram, which is its peak.
- Catch gaps in the histogram, which means bins with zero values.
- Catch the first gap which is broader than a predefined threshold, typically three standard deviations.
- Identify pixels with values more than the gap which is influenced by cosmetic rays[46].

After the cosmetic rays were localised to a spectral channel, they need to be replaced by a proper value in order to minimise the risk of information loss. There are various ways to estimate the value of signals, yet the two most typical ones are interpolating and averaging. In this study, we decided to substitute the cosmetic ray values by the average of the counts in the nearby pixels.

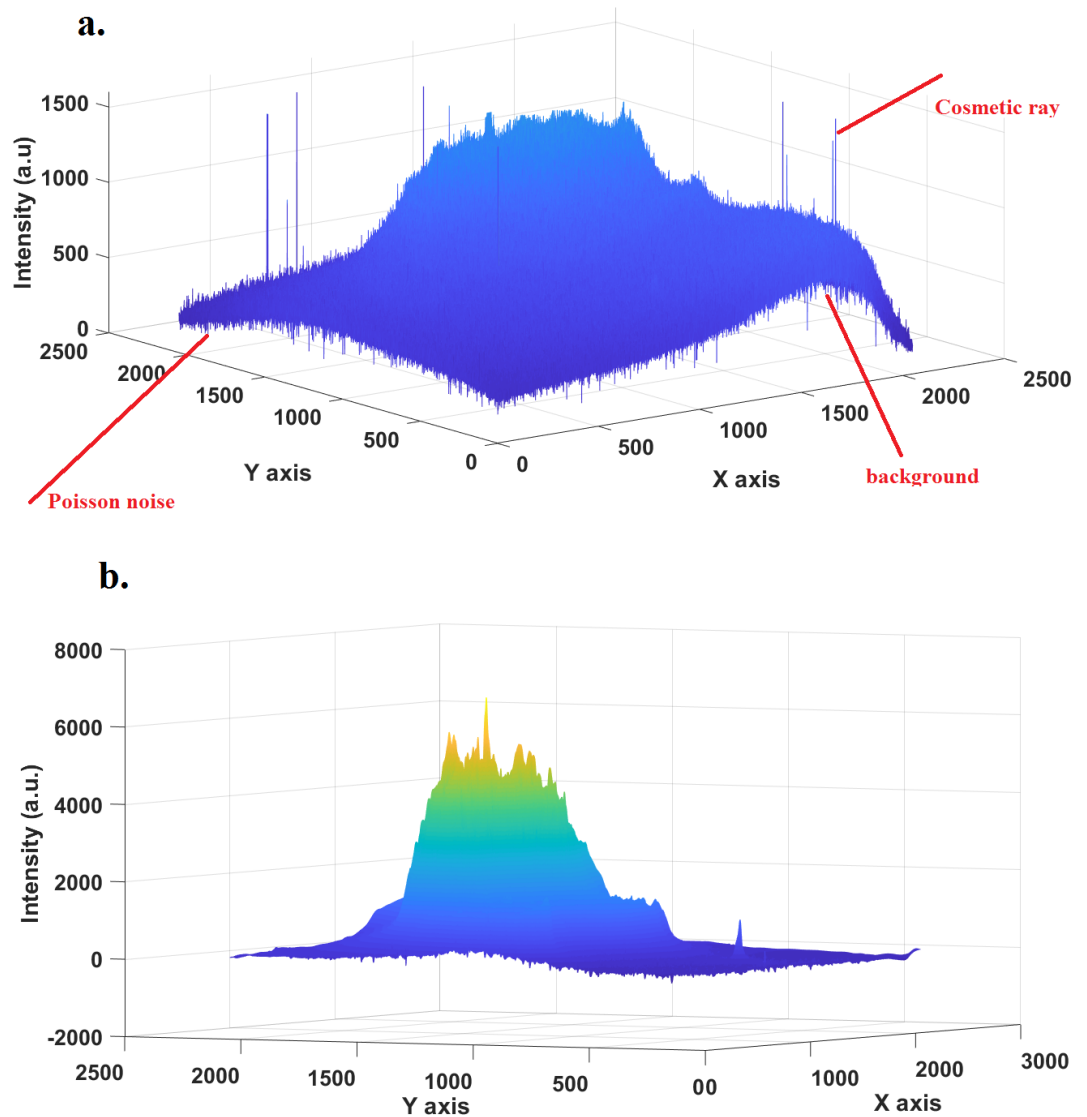
### **3.13.2 Image smoothing**

Spectral channels are also prone to get stochastic noise which can be illumination based or non-illumination based. Dark current shot noise and readout noise are considered as illumination independent noise [9]. Such noise can be minimized by using a sensor whose working temperature is low. In this study, we used a camera working below  $-65^{\circ}\text{C}$ . In addition, the camera was resistant to temperature related noise as it was working at practically the same temperature all the imaging time[9]. Another type of random noise which occurs due to illumination, such as photon shot noise, can be considered as Poisson noise due to a high degree of similarity[9].

Different methods can be applied to remove stochastic noise. Conventionally, a moving spatial average was a typical method. With the advent of the Fourier transform, de-noising mostly performed in the frequency domain [96]. In this study, the image



demonising by using wavelet transformation was achieved. This kind of smoothing removes Poisson noise by dividing the spectral image into different images with a range of scales and then zeros the coefficients below a particular threshold value[96]. Zeroing the coefficients with lower amplitude helps because image components with high frequency and low amplitude mostly do not convey valuable information. After removing the image components with low amplitude, a reverse wavelet was applied to reconstruct the spectral image. Figure 24 (a) shows a spectral image before image processing, and de-noising and Figure 24 (b) demonstrate the same image after processing in this study. Observing Figure 24 reveals that the stochastic noise along with background and cosmetic rays were optimally removed thanks to the image processing approach applied in this study.



**Figure 24. A spectral image under processing to remove different noises: (a) Before processing  
(b) After processing**

### **3.14 CELL AND TISSUE SEGMENTATION**

The spectral images acquired in this study have a square field of view in which cells are distributed. In such an image, cells do not cover all the scanned area, and consequently, some areas are left outside the cells or tissue. Also, this study was conducted statistical analysis in a single cell resolution, so the cells need to be isolated from the background. In addition to single cell analysis, this study also involved with tissue characterising to delineate OSSN which required tissue segmentation from the spectral images. Particular parts of the tissue were needed to be considered for classification purpose which was delineated by a trained pathologist. Therefore, spectral images acquired in this study were required to be segmented to obtain a region of interests (ROI). In this study, the ROIs were segmented by producing a manual binary mask in which the pixels corresponding to the ROI were 1, and all other pixels were 0. Then, such a mask was applied to associated spectral images.

### **3.15 DATA ANALYSIS**

In this study a range of different analysis such as one variate analysis, multivariate analysis and machine learning techniques. Depending on the target of the analysis, a specific analysis has been selected and justified in the associated section. To validate the method, different level of cross validation was used which is described in the appropriate section. A range of different data clustering has been used in this study, such as principal component analysis[97]. The separability of the clusters was quantified using inter-cluster distances, and various classifiers were also developed to predict the label of the cells or tissue. Such methods have been also used hyperspectral remote sensing, which is a close field to the auto fluorescence spectral imaging. Although the image formation process is different, the concept of the data analysis can be similar[98].

### **3.16 ETHICS**

This study involved with human samples and was undertaken in full agreement with human ethics at Macquarie University( reference numbers: 5201600708 and 5201400458).



# Chapter 4: OSSN DETECTION AND DELINEATION

---

This chapter had been submitted to a medical journal named “ocular surface” and recently has got revision recently. In this chapter, the revised paper, supplementary materials and response to reviewers and editor are provided. The paper describes the background, method, results and discussion of OSSN detection in sections: 4.1, 4.2, 4.4 and 4.5. The revised supplementary material is provided in section 4.8 and 4.9. Finally, comments with corresponding answers are provided in section **Error! Reference source not found..**

## NOVEL AUTOMATED NON-INVASIVE DETECTION OF OCULAR SURFACE SQUAMOUS NEOPLASIA USING MULTISPECTRAL AUTOFLUORESCENCE IMAGING.

Abbas Habibalahi<sup>1</sup>, Chandra Bala<sup>2\*</sup>, Alexandra Allende<sup>3,4</sup>, Ayad G. Anwer<sup>5</sup>, Ewa M. Goldys<sup>5\*</sup>

<sup>1</sup>ARC Centre of Excellence for Nanoscale Biophotonics, Macquarie University, North Ryde 2109, NSW Australia

<sup>2</sup>Department of Ophthalmology, Faculty of Medicine and Health Sciences, Macquarie University, Sydney NSW 2109

<sup>3</sup>Douglass Hanly Moir Pathology, Macquarie Park NSW 2113

<sup>4</sup>Faculty of Medicine and Health Sciences, Macquarie University, Sydney NSW 2109

<sup>5</sup>University of New South Wales, Sydney 2032, NSW

\*correspondence to Chandra Bala or Ewa M. Goldys

### Abstract

**Purpose:** Diagnosing Ocular surface squamous neoplasia (OSSN) using newly designed multispectral imaging technique.

**Methods:** Eighteen patients with histopathological diagnosis of Ocular Surface Squamous Neoplasia (OSSN) were recruited. Their previously collected biopsy specimens of OSSN were reprocessed without staining to obtain auto fluorescence multispectral microscopy images. This technique involved a custom-built spectral imaging system with 38 spectral channels. Inter and intra patient frameworks were deployed to automatically detect and delineate OSSN using machine learning methods. Different machine learning methods were evaluated, with K nearest neighbour and Support Vector Machine chosen as preferred classifiers for intra and interpatient frameworks, respectively. The performance of the technique was evaluated against a pathological assessment.

**Results:** Quantitative analysis of the spectral images provided a strong multispectral signature of a relative difference between neoplastic and normal tissue both within each patient (at  $p < 0.0005$ ) and between patients (at  $p < 0.001$ ). Our fully automated diagnostic method based on machine learning produces maps of the relatively well circumscribed neoplastic-non neoplastic interface. Such maps can be rapidly generated in quasi-real time and used for intraoperative assessment. Generally, OSSN could be detected using multispectral analysis in all patients investigated here. The cancer margins detected by multispectral analysis were in close and reasonable agreement with the margins observed in the H&E sections in intra and inter patient classification, respectively.

**Conclusions:** This study shows the feasibility of using multispectral auto-fluorescence imaging to detect and find the boundary of human OSSN. Fully automated analysis of multispectral images based on machine learning methods provides a promising diagnostic for OSSN which can be translated to future clinical applications.

**Keywords:** Ocular surface squamous neoplasia, Boundary detection, Auto fluorescence

## 4.1 INTRODUCTION

Ocular surface squamous neoplasia (OSSN) [99] is considered the most common neoplastic eye surface disease [100, 101] and constitutes approximately 10% of ocular surface lesions [102]. OSSN is a broad term which encompasses noninvasive conjunctival intraepithelial neoplasia (CIN), dysplasia and squamous cell carcinoma in situ (SCCIS), and invasive squamous cell carcinoma (SCC) where tumor cells invade the stroma by breaching the basement membrane [99, 103, 104]. Early detection and accurate boundary delineation of OSSN results in simple and more curative treatments such as topical therapies[105, 106], whereas advanced invasive lesions may require eye exenteration [107] and also has the risk of mortality[108]. The clinical symptoms of OSSN are variable [109], so patients may experience delay in treatment or be diagnosed inappropriately due to lack of suitable methods [107, 110].

Currently, accurate diagnosis relies on clinical suspicion which can be confirmed by impression cytology or biopsy [111, 112]. Ocular biopsy represents the gold standard for diagnosis of OSSN which relies on a histological assessment [111]. Biopsy is invasive and, theoretically, there is risk of seeding [113, 114]. Impression cytology (IC) is another OSSN diagnostic method which examines tumor cells collected from the superficial layers of the ocular surface [112]. Although it is less invasive than a biopsy, its reliability remains disputed [115]. Biopsies and IC are only carried out in the case of clinically visible disease, so small lesions may be missed which results in a false negative outcome [21]. Intraoperatively, it is extremely difficult to distinguish tissue margins using both biopsy and IC methods. The corneal end can be friable and split away from tissue samples, with resulting fragmentation of tissue sometimes hampering histological assessment of margin clearance. Moreover, histology and IC processing require long sample preparation time [22]. More recently, new imaging modalities have been used to detect OSSN, such as ultrasound biomicroscopy [116], *in vivo* confocal microscopy [117] and optical coherence tomography (OCT) [118]. These imaging modalities [118] do not provide a very sharp distinction between OSSN and benign lesions, which makes clinical applications of these technologies challenging [119]. Moreover, some technologies such as Raman spectroscopy[120] need sophisticated technology to be implemented although showed promising results in detection of cancer. Therefore, their application for routine OSSN detection is also challenging.

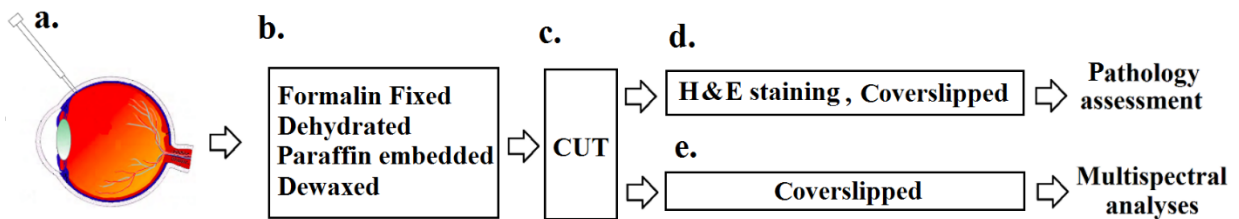
To overcome these limitations, an imaging diagnostic technique for OSSN which is cost-effective and potentially free of staining, and able to be performed in the clinic is highly desirable. Recently, auto-fluorescence imaging techniques have been applied for cancer diagnostics, using advanced modalities, e.g. fluorescence lifetime imaging (FLIM) [121, 122]. These were applied to various types of cancer such as neck cancer, lung cancer, cervical cancer, as reviewed in [22, 123, 124]. Auto-fluorescence imaging plays a special role in cancer detection, as it is capable of recognizing aspects of chemical composition of tissue based on spectral signatures [125, 126] of naturally fluorescent compounds such as, Protoporphyrin IX (PPIX), reduced nicotinamide adenine dinucleotide (NADH) and Flavin adenine dinucleotide (FAD) [42, 52] whose content is modified in cancer due to a transformation in cell metabolism. In particular, lack of iron or ferrochelatase in tumors results in a change of PPIX concentration relative to the normal host tissue [20, 64]. The utilisation of native PPIX as a cancer marker has been explored in colon and breast cancer [42, 61, 64]. Furthermore, cancerous cells transform glucose into lactate regardless of the presence of oxygen, resulting in concentration change of NADH and FAD and their ratio [3, 55]. Consequently tumor cells can be differentiated from healthy cells by auto-fluorescence imaging [20].

In this study, we used a newly-designed non-invasive auto-fluorescence multispectral imaging methodology of OSSN, with a view to future clinical ophthalmological application. This study aims to distinguish normal and neoplastic tissue in fixed human samples. Two different classification frameworks were deployed namely intra- and inter-patient classification [22] to consider aspects of patients' variability and quantify spectral signature of OSSN. We also introduced an approach for objective assessment of boundary detection using machine learning methods. This technique creates a false colour map which can be rapidly generated in quasi-real time and used for intraoperative assessment. The neoplastic boundaries predicted by employing our machine learning methods were validated and assessed by an anatomical pathologist. To the best of our knowledge, this is the first study where multispectral analysis of auto-fluorescence in conjunction with artificial intelligence was applied to identify and find boundaries of human OSSN.

## 4.2 METHODS

### 4.2.1 Patient recruitment

Our analysis of patients' samples was performed under the permission from Macquarie University Human Research Ethics Committee, reference No: 5201600708 and informed consents were obtained. The 4 $\mu$ m tissue sections were prepared in pairs; these included a standard FFPE H&E (formalin fixed paraffin embedded haematoxylin and eosin) section used as reference, and adjacent formalin fixed, dehydrated, paraffin embedded dewaxed and cover slipped section without staining used for multispectral analysis. The region of interest (ROI) which served as our gold standard, was located on the H&E section and correlated with the unstained slide where multispectral imaging was performed. Figure 25 illustrates the sample preparation and characterisation in this study. To categorize lesions as normal or neoplastic, classical grading of dysplasia based on thickness of epithelial involvement by abnormal cells was used. Involvement of basal third and basal two thirds correspond to low grade (CIN1) and moderate grade dysplasia (CIN2) respectively, and involvement of all thirds and full thickness involvement correspond to high grade dysplasia (CIN3) and SCCIS respectively. All lesions with varying grades of dysplasia, SCCIS or SCC were considered neoplastic. A total of 18 samples from 18 eyes of 18 patients (40 to 82 years old) were included in this study, as described in Table 4.



**Figure 25 Sample preparation and characterization. (a) Collection of eye biopsies from patient subjected to eye surgery. (b) Preparation of histology samples. (c) Cutting of two adjacent tissue sections. (d) Pathology assessment of a stained slide of the sample used as a reference. (e) Multispectral imaging of an unstained slide from adjacent tissue section.**



**Table 4. Summary of patient information and surgical specimens**

Patient ID	Age	Gender	eye	Histology analysis
PID 1	63	Male	Right	SCCIS
PID 2	78	Male	Right	SCCIS
PID 3	69	Female	Right	High grade dysplasia
PID 4	59	Male	left	Low grade dysplasia
PID 5	80	Male	Right	High grade dysplasia
PID 6	72	Male	Right	SCCIS
PID 7	78	Male	Right	SCC
PID 8	41	Male	Right	Low grade dysplasia
PID 9	79	Male	left	Low grade dysplasia
PID 10	62	Female	Right	SCCIS
PID 11	64	Female	left	SCCIS
PID 12	74	Male	left	High grade dysplasia
PID 13	60	Male	Right	SCCIS
PID 14	58	Male	left	Moderate grade dysplasia
PID 15	72	Male	left	SCCIS
PID 16	82	Male	Right	Moderate grade dysplasia
PID 17	47	Male	Right	SCCIS
PID 18	59	Male	Right	SCCIS

#### 4.2.2 Fluorescence multispectral imaging system and imaging procedure

We used a novel custom-made wide-field fluorescence microscopy system [11, 95, 127] incorporating a number of light emitting diodes (LEDs) for multispectral excitation. These LEDs determine the excitation wavelength ranges as specified in Table 5. This Table also lists optical filters used to determine the excitation/emission wavelength ranges used in this work. Taken together, these excitation and emission wavelength ranges produce a number of distinctive spectral channels ( $N_{ch}=38$ ). These channels can capture various fluorophores, including NADH, PPIX, Flavins, lipopigments, collagen, Elastin. Multispectral imaging involves generating a sequence of fluorescence images of the same sample area in each of these spectral channels, collectively referred to as a “data block”. Each data block is accompanied by a differential contrast microscopy (DIC) image of the sample, used as broad reference. A CCD camera with high quantum efficiency was used to acquire the fluorescence images in the spectral channels. This camera has a 16 bit A/D converter and 2048x2048 pixel sensor. The images we captured using an adapted IX83 Leica microscope, at 40 times magnification (40x).

**Table 5. The excitation/emission wavelength ranges (spectral channels) used in this study (green squares). The excitation was supplied by a number of LED light sources whose spectra had listed**

peak wavelengths and approximately 10 nm channelwidth, covering a spectral range from 340 nm to 510 nm. The wavelength range of the fluorescence emission light produced by the sample was determined by four channel pass emission filters which span the electromagnetic spectrum from 420 to 650 nm. Each channel pass filter is combined with an appropriate dichroic mirror with cut off wavelengths of 405nm, 450nm, 573nm and 460 nm for emission filters with 420-460 nm channel pass, 454-496 nm channel pass, 573-613 nm and 575-650 nm, respectively. This arrangement of 17 LED excitation wavelength ranges with 4 emission channel pass filters results in 38 channels (our microscope setup is shown in Supplementary Figure 30).

Em. band pass filter (nm)	LED Ex.wavelength $\pm$ 5 (nm)																
	340	368	373	378	382	388	391	394	405	413	432	441	455	460	470	491	510
420-460																	
454-496																	
573-613																	
575-650																	

## 4.3 DATA ANALYSIS

### 4.3.1 Image preprocessing

The image preprocessing is performed to remove random and systematic errors which influence the individual spectral channel images [11, 89, 95, 128]. Random errors include Poisson's noise, dead pixels, spikes due to cosmic rays, analogue-to-digital converter errors, bit errors in transmission, and similar [89, 95]. Spikes, saturated or dead pixels were located by a 'threshold limiting window' and then replaced by values interpolated from the immediately adjacent 8 pixels [129] ( for more information see supplementary material section 4.8.2). To remove Poisson's noise, a wavelet filter with symmetric Mother function was used [128, 130, 131] (for more information see supplementary material section 4.8.3).

Apart from random errors, two systematic errors make contributions to the spectral channel images namely the fluorescent background (from the microscope optics and the sample slide) and uneven illumination of the field of view [11]. To remove this background, a reference background was subtracted from each data block, channel by channel (for more information see supplementary material section 4.8.4). To correct for uneven illumination, the spectral images were divided by the images

taken from a reference calibration fluid for each spectral channel separately [95] ( for more information see supplementary material section4.8.6).

#### **4.3.2 Image stitching**

The use of 40 times magnification objective used to acquire fluorescence images has a limited field of view ( $\sim 0.20 - 0.25 \text{ mm}^2$  of a sample covered in each image). To enable the inspection of larger fields of view, a series of overlapping microscopy images were taken. Then, the affine algorithm [132] was applied to co-register the overlapping images based on an image taken by 5x objective with a bright field microscope. Finally, a composite image was generated (Supplementary Figure 31 and Figure 32.).

#### **4.3.3 Establishing initial normal and neoplastic signatures**

In order to extract specific quantitative information from images in each spectral channel to differentiate between neoplastic and normal tissue, the images were first divided into corresponding sectors (squares,  $10 \times 10$  pixels each) and average intensity in each spectral channel was calculated over each sector at XY plane (Supplementary Figure 33). This averaging decreases the computational time and acts as an additional low pass filter which removes some of the remaining random variations in tissue and/or fluorescent signals. The average fluorescence intensity in each sector forms an  $N_{ch}$  ( $N_{ch} = 38$ ) dimensional feature vector, referred to as the spectral signature of that sector. Hence, many sectors ( $> 10000$ ) were extracted from each tissue, each of which produces a vector of features. Depending on the pathology assessment, each sector was labelled as normal or neoplasm. Subsequently, these sectors were fed into machine learning techniques as each dimension of such sector provide an input feature.

#### **4.3.4 Classification frameworks**

In this study, we used two different standard frameworks for classification: inter-patient classification and intra-patient classification [22].

We used the inter-patient classification framework [133] to detect patient's OSSN considering between patient variability and heterogeneity. Due to relatively

small size of the patient group (eighteen) we applied the standard “leave one [patient] out” approach [134]. A single patient was chosen and the data associated with the patient (named the “testing” patient) were put aside. Further, the multispectral signatures of the neoplastic and normal tissue were recognized based on the data from the remaining cohort of patients (named “training” patients). Hence, the tissue from the “testing” patient has no contribution to the multispectral neoplastic signature identified in this step. Accordingly, the recognized signature is applied to predict block labels of the “testing” patient in order to generate the false colour map and statistical metrics. This is then repeated until all patients in the cohort considered as testing patients have been processed. This workflow is illustrated in Figure 26 (a).

In the intra-patient framework, the multispectral signature of neoplastic/normal tissue was recognized based on a single patient independently of other patients and then tested on another part of tissue from the same patient. This framework was proposed in the literature to minimize interpatient variability and heterogeneity[22]. Figure 28(a) illustrates the intra-patient workflow. The unstained tissue is first divided into two separate sections called the training and testing tissue. The training tissue has both normal and neoplastic areas, the availability of which makes it possible to extract the multispectral neoplastic signature. The trained classifiers were then applied to the data blocks of the testing area of the same patient for the label prediction (“neoplastic” vs “non-neoplastic”), which resulted in a false colour map (Figure 29) and related statistical metrics. The Supplementary Figure 34 shows the testing and training tissue for a sample patient. This process was repeated for all patients (n=6) whose samples were appropriate for intra-patient classification. Such samples must have sufficiently large neoplastic and normal area in their training tissue, in order to properly identify the multispectral neoplastic signature. Moreover, their testing tissue must incorporate the intersection between neoplastic and normal tissue, which is required for drawing an accurate false colour map.

#### **4.3.5 Multivariate analysis and artificial intelligence methods**

A range of different univariate, multivariate and artificial intelligence methods were used in this study including ANOVA[135], PCA[97], statistical learning and machine learning analyses [136, 137]. ANOVA was used to evaluate the within and between group variance and verify the statistical significance of our tissue

identification. In addition, t- test was applied on the elements of spectral signature from neoplastic and normal tissue to prove that they are not from a normal distribution with equal means[138](more information can be found at supplementary material section 4.8.9). A significant difference can be regarded as indicative of a strong multispectral signature of a relative difference between neoplastic and normal tissue from an individual patient. To evaluate the robustness of the multispectral content hidden in the training data, the unsupervised PCA analysis may also be implemented (more information can be found at supplementary material section 4.8.9.4.8.10).

Machine learning methods as artificial intelligence approaches were used to generate smart statistical models relating the spectral signatures to tissue pathological condition. To select a robust prediction model, different classifiers were evaluated and chosen based on their properties and performance, including the artificial neural network classifier (ANN)[139], quadratic discriminant analysis classifiers (QDA) [140], Decision Trees [141], Support vector machines (SVM) [142] with linear and quadric kernels, and K nearest neighbour (KNN) [143]. Evaluation of the classifier training process was performed by K fold (K=10) cross validation [144], which also kept the classifiers secure from overtraining[145] (more information can be found at supplementary material section 4.8.9.4.8.11).

After evaluating the predictive model using ROC and their performance [144], the KNN and SVM were chosen as the preferred classifiers in this study for the intra patient and inter patient frameworks, respectively. KNN was selected as it showed high performance in intra patient classification with the area under curve (AUC) close to 1. The KNN learns the data construction to predict class labels by defining the nearest neighbours and assessing the similarity to a query example [143]. In fact within a dataset, the data points will usually be in adjacent proximity to other data points that have comparable properties. Providing the data points are recognized with a classification label, then an unclassified data point can be labeled from its nearest neighbours [136]. KNN rapidly learns the data structure and it can be quickly re-trained, which is beneficial for intra patient classification. In the intra patient framework, a specialized classifier is generated for each patient, which can then be stored in the doctors' office computer. SVM is a strong supervised method which forms a hyperplane with maximum margins in the high dimensional spectral feature space to separate data points into neoplastic and normal classes. In addition to strong

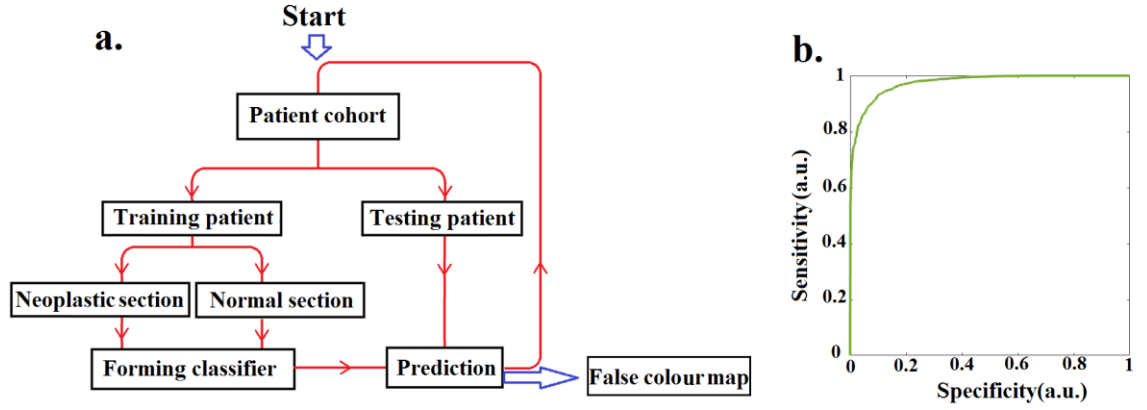
performance ( $AUC > 0.98$ ), SVM with quadric kernel is chosen for inter patient classification as it has robust performance in dealing with sparse and noisy data [142] caused from between patient variability. These machine learning techniques were trained on sectors extracted from neoplastic and normal area of patients' samples to find the multispectral signature of neoplasia. Then, the trained model was applied on an unknown area to determine the existence of similar pathological conditions [22]. Finally, the predicted outcomes were compared with the pathology gold standard to evaluate the classifier performance.

## 4.4 RESULTS

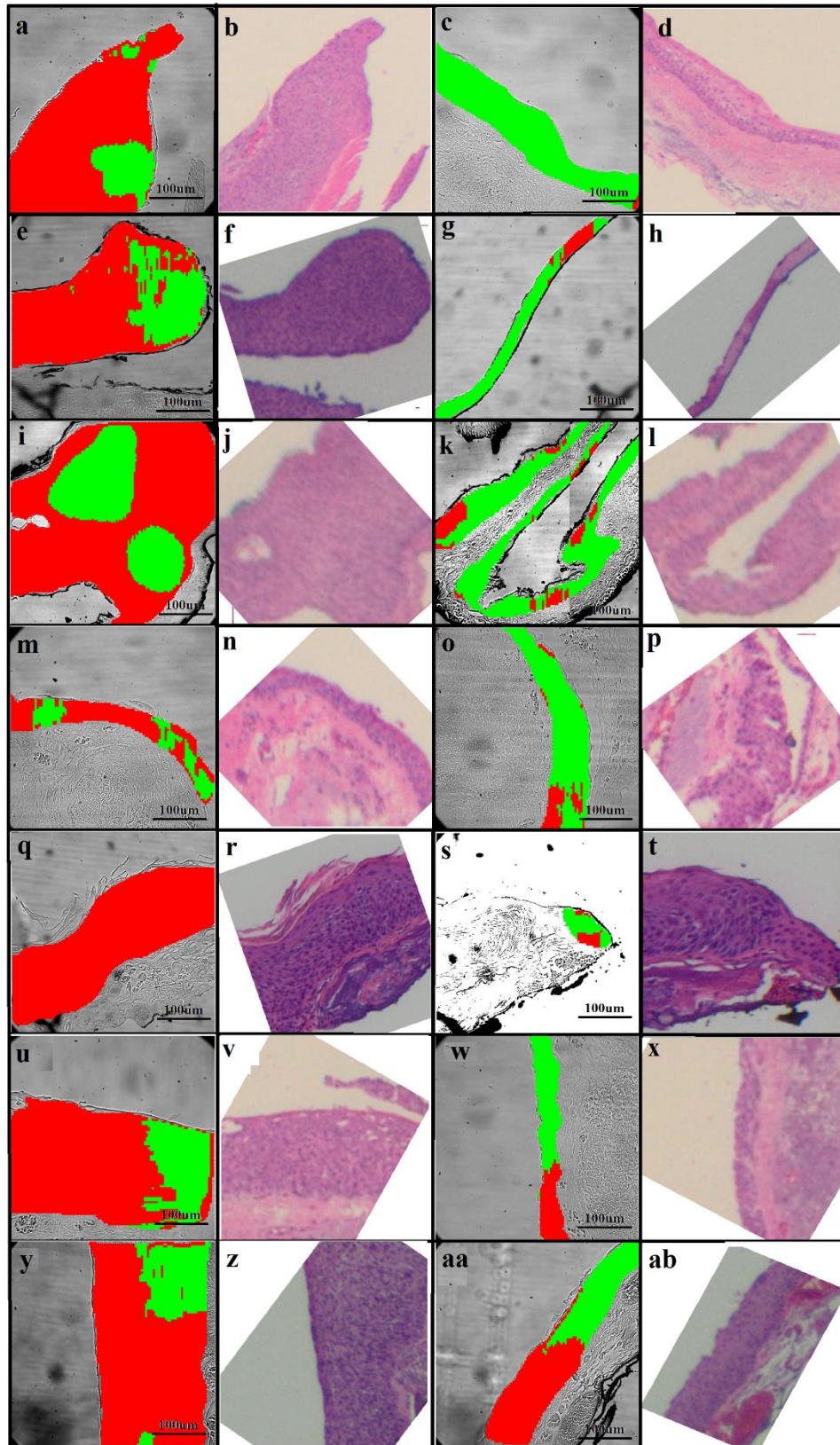
In order to effectively analyze the multispectral data and consider the “between-patients” and “within-patient” variability of the spectral neoplastic signatures, we developed two different frameworks named intra- and inter-patient classification [22]. The neoplastic boundary outlined by the pathologist served as the gold standard for our analysis of neoplastic prediction.

### 4.4.1 Inter-patient classification results

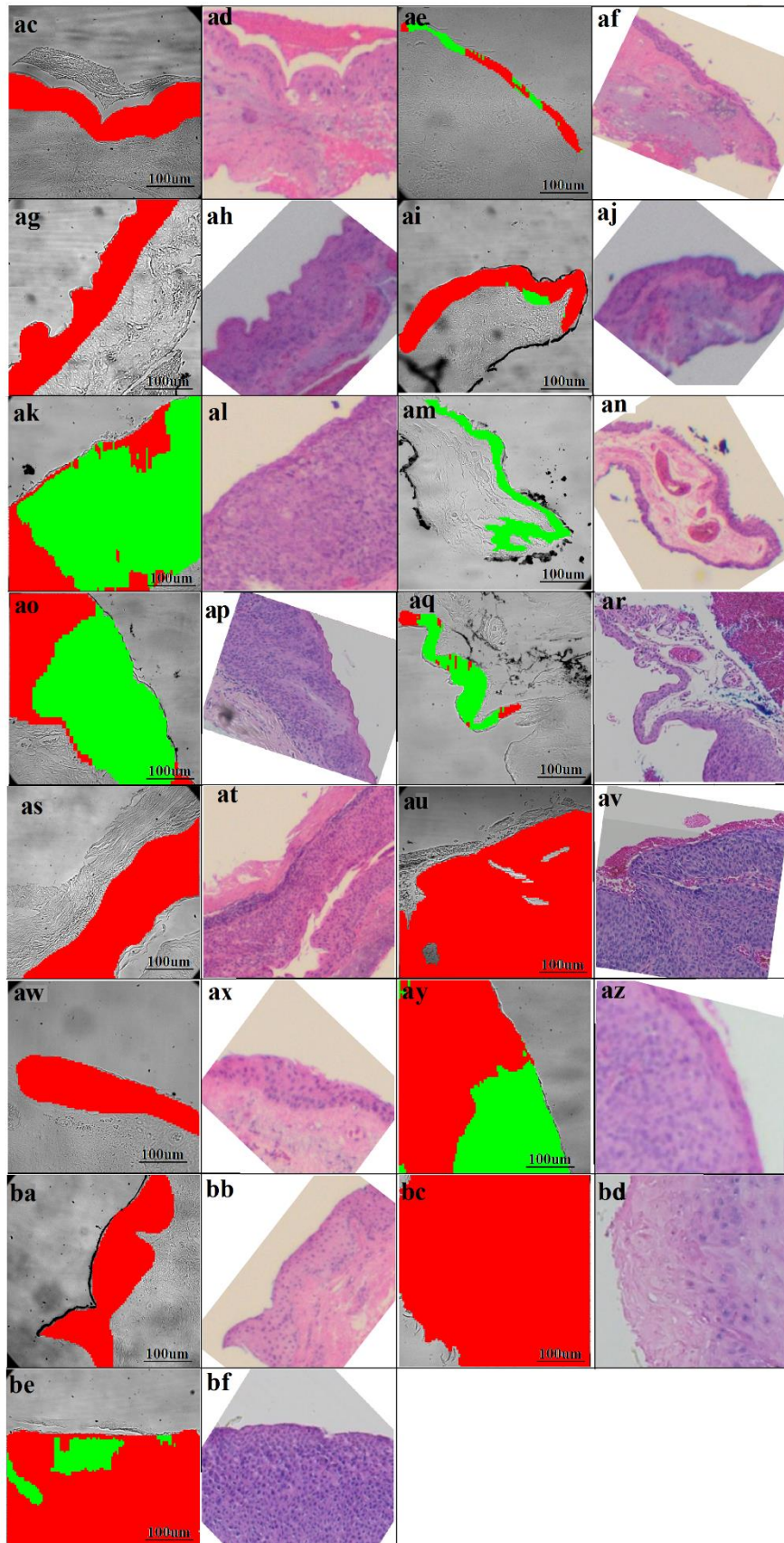
In the inter-patient classification we evaluate a relative difference between neoplastic and normal tissue from all patients in the training cohort. The tissue sectors extracted from the training area were grouped into normal and neoplastic sections and passed the univariate t-test in each iteration; this has verified that significant differences exist (on average at  $p < 0.001$ ). As described above, SVM classifier was selected and trained as the predictive model for inter-patient classification. Subsequently, the image sectors of the testing patient were fed into the classifier to predict their labels. The final ROC curve shown in Figure 26(b) is an average of the ROC curves generated in each iteration of the loop. The ROC with  $AUC > 0.98$  demonstrated that the SVM classifier was well trained and could successfully generate the multispectral signature of the training sectors in the inter-patient framework.



**Figure 26. Inter patient classification workflow and analysis. (a) The interpatient work flow incorporates a loop beginning from the 'START' point. The cohort of patients is divided into two groups: the testing group comprising one patient and the training group with 19 patients. Neoplastic and normal area of training patients are segmented and classified as normal or neoplastic sectors. After SVM the classifier is shaped, and it is then applied to the testing patient to predict sector labels in that patient. Then, another iteration commences from the START point with a consecutive testing patient. The loop continues until the data blocks from all of the eligible testing patients are examined. (b) ROC curve of SVM classifier used for the prediction of data block labels in the inter-patient classification (AUC >0.9 which demonstrate a grate separation between classes).**







**Figure 27.** False colour map superimposed on the DIC image of testing patients in the inter-patient classification for different patients in comparison with the corresponding histology

images. Data blocks are coloured in red or green if they are respectively predicted to be neoplastic or normal. First/third column is the multispectral false colour map and the second/fourth column is the corresponding H&E image. (a),(b) and (c),(d) neoplastic and normal section for PID 1, respectively. (e),(f) and (g),(h) neoplastic and normal section for PID 2, respectively. (i),(j) and (k),(l) neoplastic and normal section for PID 3, respectively. (m),(n) and (o),(p) neoplastic and normal section for PID 4, respectively. (q),(r) and (s),(t) neoplastic and normal section for PID5, respectively. (u),(v) and (w),(x) neoplastic and normal section for PID 6, respectively. (y),(z) and (aa),(ab) neoplastic and normal section for PID 7, respectively. (ac),(ad) and (ae),(af) neoplastic and normal section for PID 8, respectively. (ag),(ah) and (ai),(aj) neoplastic and normal section for PID 9, respectively. (ak),(al) and (am),(an) neoplastic and normal section for PID 10, respectively. (ao),(ap) and (aq),(ar) neoplastic and normal section for PID 11, respectively. (as),(at) neoplastic section for PID 12. (au),(av) neoplastic section for PID 13. (aw),(ax) neoplastic section for PID 14. (ay),(az) neoplastic section for PID 15. (ba),(bb) neoplastic section for PID 16. (bc),(bd) neoplastic section for PID 17. (be),(bf) neoplastic section for PID 18. (The borders of the lesions are highlighted by dash line in the Supplementary Material Figure 35 )

The representative false colour maps for neoplastic boundary evaluations are shown in Figure 27(a-be). In this Figure, the neoplastic tissue is coloured in red and normal in green on the DIC image depending on the classifier predictions. Generally, neoplastic areas determined by multispectral imaging are reasonably similar to neoplastic areas determined by histology. Figure 27 c,p,ab,af,ar,ar,at,av,az,al,bb show the identical results om multispectral analysis and sector classification to give identical results with histopathology assessment. In several samples, the neoplastic areas determined by multispectral imaging were smaller than those determined by histology (Figure 27. a,e,l,m,t,x,aj,an,bd). We note that in some cases (Figure 27 i,bd) the areas classified by multispectral imaging as healthy were surrounded by neoplastic areas which is unlikely. Such samples can be post-processed and reclassified based on supervised corrections to remove those areas. Figure 27 g,k,o,r,v,z,ad,ah , ap depict normal samples which have shown false positive regions. We did not correlate the disease grade with the multispectral signature and there is no general correlation concerning the grade of disease and agreement between histological images and multispectral approach in this study. Overall, Figure 27 demonstrates that ~ 40% of patients show strong correlation (>85%) with histopathology assessment, ~50% reasonable(>65%) and ~10% poor agreement (<65%) with the histopathology assessment. The poor agreement between multi spectral and histopathology assessment led to 11%, 18%, 88% and 80% false negative, false positive, sensitivity

and specificity rate, respectively, which can be considered reasonably high for the system. In addition, the correlation between histology assessment and multi spectral analysis pixel-wise is  $\sim 78\%$ .

Figure 27 indicates that it is possible to form clinically useful judgments based on multispectral analysis. It confirms that regions identified as neoplasm in pathology can be classified as neoplastic tissue using our multispectral imaging approach with reasonable agreement. We emphasize that all patients investigated here who were diagnosed to have OSSN by pathology, have been identified as having neoplastic tissue by our approach. This means that, in principle, this system has the capability to eliminate the histological assessment for confirmation of OSSN in patients. In addition, if biopsy was prescribed to validate the positive multispectral outcome, a physician can use the multispectral map to locate the area most likely to have OSSN and avoid blind area selection. Localizing biopsy sample also minimizes size of excised tissue.

#### **4.4.2 Intra patient classification results**

In the intra patient classification, we evaluate a relative difference between neoplastic and normal tissue from one patient at a time. In this classification, patient variability in detecting OSSN is minimized. The KNN classifier was selected for reasons explained in the data analysis section and was trained as the predictive model for intra patient classification. A significant difference (at  $p < 0.0005$ ) was demonstrated in the t-test and regarded as indicative of a strong multispectral signature of a relative difference between neoplastic and normal tissue from an individual patient. Figure 28(b) shows the three highest PCA scores in the discriminant space, demonstrating that the neoplastic sectors are very clearly separated from the normal sectors. Subsequently, these training data were used to shape the classifiers. The ROC curve for the KNN classifier obtained from the example sample is shown in Figure 28(c) which demonstrates the area under the curve (AUC) of selected classifier is approaching 1, proving that the predictive models could efficiently recognize the multispectral signature of the neoplastic versus normal tissues. To generate the false colour map highlighting the neoplastic areas, firstly, the average spectral signature of the sectors in the testing area were extracted and fed into the KNN classifier to predict

if they are neoplastic or normal. Then, these sectors were coloured red (neoplastic) or green (normal) and positioned on the composite DIC image. Figure 29 (a,e,i,c,g,k) show the false colour maps for samples from patients used in the intra-patient framework.

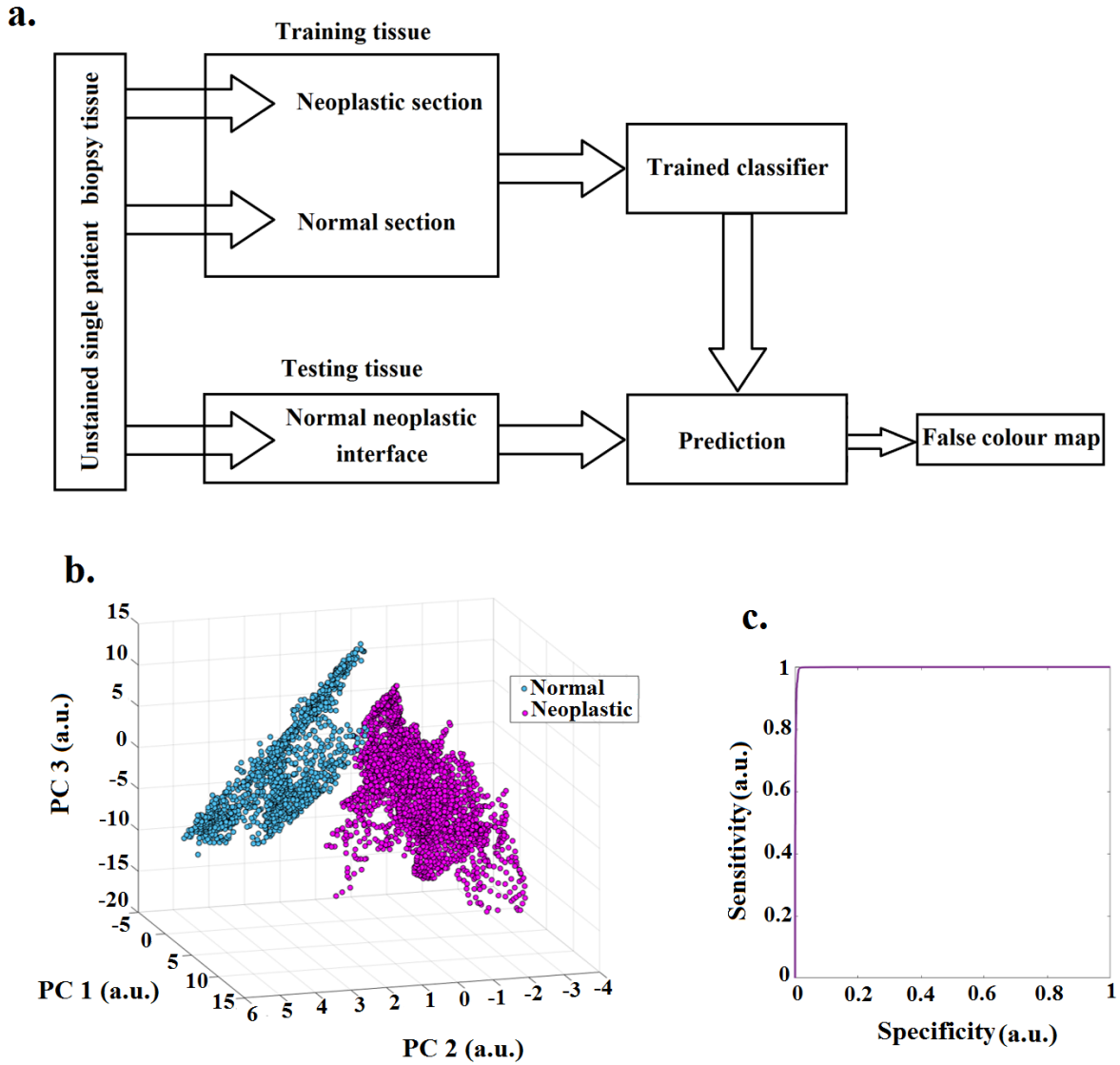


Figure 28. Intra-patient analysis workflow and data analysis. (a) Neoplastic section, normal section and normal neoplastic interface from an unstained tissue of a single patient are collected. Neoplastic and normal sectors are used as training tissue and the normal-neoplastic interface serves as the testing tissue. Data blocks extracted from neoplastic and normal areas are used to train classifiers. The classifiers are then applied to the sectors of the testing tissue to predict their labels. Finally, the predicted labels are compared to the gold standard pathology recommendations to draw metrics (Figure 34 demonstrate

an example framework to show how samples from a single patient was used for training and testing). (b) Data blocks extracted from the training area are projected onto a discriminative space created by unsupervised PCA for the example patient. The neoplastic or normal data points for the sample patient are shown by purple or blue colour, respectively (separability~5). The arbitrary units (a.u.) has been used as it is combination of different channels . (c) ROC curve associated with KNN classifier applied on the sample patient (AUC~1 which demonstrates the separation between the classes).

In the intra patient classification, the testing tissue incorporates the normal/neoplasm interface. Hence, to evaluate the concordance between the multi spectral false colour map and H&E sections in Figure 29, H&E images were orientated with the neoplastic /normal areas located on the left/right side of the tissue for all sections with their corresponding colour maps. Figure 29 shows a remarkably high correlation between the multispectral prediction and histopathological evaluation at the normal/neoplasm interface for all of the 6 patients tested in the intra patient framework . Figure 29 a,e,i,c,g and k confirms that the classifier was able to correctly detect the tumour and normal areas in 100% of patients investigated. In Figure 29 i j and k, we note that a small normal area is developing in the neoplastic areas, which is highly unlikely. In this framework, the correlation between the histology assessment and multispectral analysis is ~94%. The agreement between H&E sections and multispectral analysis can be improved by a supervised correction to reclassify these areas. In addition, smoothing the colour map images can remove slight misclassified spots. Generally, in all patients the tumour was detected within less than 200 microns from a pathological boundary, suggesting that such a small excision margin would be safe and adequate.

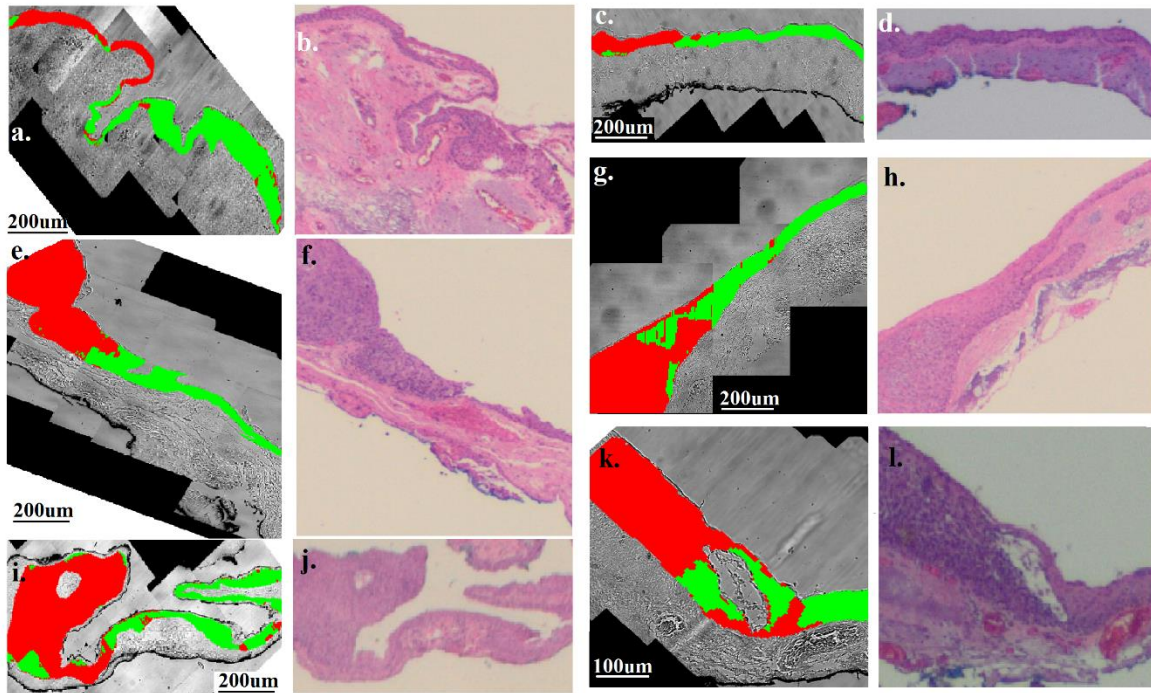


Figure 29) False colour map generated to locate the neoplastic boundary on the testing tissue in the intra patient classification framework in comparison with the corresponding histology images. The position of the block data on the DIC image is coloured in red or green if they are predicted to be neoplastic or normal, respectively. First/third column is the multispectral false colour map and second/third column is corresponding H&E section. (a) and (b) associated to PID4, (b) and (c) associated to PID9, (e) and (f) associated to PID 10. (g) and (h) associated to PID 1. (i) and(j) associated to PID 3, (k) and(l) associated to PID7 (borders of the lesions were highlighted by dash line in the Supplementary Material Figure 36).



## 4.5 DISCUSSION

Imaging modalities capable of non-invasively detecting OSSN at early stages and precise localization of tumour margins are highly desirable in the clinic. This approach has the potential to reduce the incidence of eye biopsies, prevent therapy delays and make treatment more effective. In this study, we propose a non-invasive auto-fluorescence multispectral imaging procedure for quantitative analysis of cancer-related spectral signatures of the tissue retaining the relevant spatial information. This approach is applicable both to early diagnosis and the delineation of neoplastic boundary on a real time basis, employing cost-effective wide-field microscopy instrumentation. Our system uses light excitation with a number of narrow band wavelength ranges and it collects native fluorescence emission of the tissue also at specific wavelengths ranges. Each excitation /emission wavelength band combination forms a spectral channel, and we used ( $n=38$ ) such channels in this work (Table 5). The sample is imaged in each of these channels, to acquire separate fluorescent channel images. This represents an advance over the current, ophthalmological auto-fluorescence imaging systems. Current auto fluorescence technologies employed in fundus camera typically use only one blue light excitation to evaluate retina diseases such as macular dystrophies, age related macular degeneration, white dot syndromes, retinitis pigmentosa as reviewed at [146, 147]. In fact, the availability of multiple channels makes it possible to survey the overall biochemical composition in the tissue, in addition to detecting specific markers to identify tissue state.

Artificial intelligence is applied in this study to eliminate the need of data interpretation by a highly skilled operator. We established two independent classifications, in the intra- and inter-patient frameworks to provide maps of neoplastic boundaries. In both cases, the analyzed data were validated by comparing with the pathology standard. In the intra-patient classification, the neoplastic boundary was detected using a model formed by the tumour and normal tissue from the same patient. This approach showed reliable results with high agreement between false color multispectral map and corresponding H&E sections. The intra-patient method is highly relevant clinically as corresponding healthy areas in the same patient (such as the other eye) are usually available. The intra patient framework can be applied for accurate

boundary delineation, and also to detect possible OSSN spots in another eye or the OSSN recurrence at the early stages.

Our intra patient classification may help to monitor OSSN treatment. Surgical excision and topical interferon with retinoic acid are the mainstay of treatment where our technology can offer an advantage. Surgical excision [148] needs to be complete [149-151] to avoid subsequent recurrence [152, 153]. Currently, OSSN recurrence rate after surgery can be high ( $>30\%$ ) [148, 154]. An accurate intra-patient image could assist the surgeon to precisely locate the neoplastic boundaries and to completely remove the residual neoplastic tissue. On the other hand, accurate boundary detection may help the surgeon to avoid redundant resection of healthy tissue. In the case of tumour regression by topical treatment, premature treatment termination led by a false clinical impression may also increase the risk of recurrence. The technology could be used to monitor treatment efficacy and detect recurrences non-invasively.

In this study, the inter-patient classification framework shows acceptable accuracy with reasonable qualitative performance. Lower accuracy of interpatient classification compared to intra patient framework is attributed to biological heterogeneity of patients and neoplasms [18]. The effects unrelated to neoplastic disease such as different sunlight exposure, different ethnicities, eye colours etc. may mask the multispectral signature and influence the results. Furthermore, metabolic activity within the tumour may not be uniform which could also influence the classifier. In addition, sample preparation such as uneven cut, or folded tissue may affect the classification. Although such diversity casts a shadow on the statistical power, the significance tests demonstrate that interpatient classification is still acceptably accurate with the p-value of  $< 0.001$  but with room for improvement. In this feasibility study, our inter-patient models were constructed based on a limited number of patients. A larger patient's database is likely to refine the spectral signatures and limit random side effects. The interpatient classifier can be directly employed to diagnose the incidence of neoplastic tissue in the eye or to assist in decision making for clinicians.

The technique discussed here has the potential to be used in a range of applications. It can be simply integrated with a slit lamp in the clinics or be integrated with the equipment in surgical suites. The number of the channels can be optimized bearing in mind the misclassification error, critical nature of the application and the



consequences of misclassification error; with a smaller number of channels likely to make the system more user-friendly. To potentially apply our technology in a real clinical setting safe excitation lights were employed based on the power intensity. In fact, the highest UV power used in this study is less than half of the solar UV power which would reach the pupil on an average sunny day at earth sea level.

The technology described here offers distinctive advantages in ophthalmology applications. It is an inexpensive noninvasive technique with capability to be used on an outpatient basis. The imaging system used in this study doesn't require any contact with the ocular surface in contrast to both ultrasound biomicroscopy and some confocal microscopy devices [21]. The contactless nature of our approach decreases both the length of time and technical expertise required for employing the system for real clinical applications. To use the time more effectively, a relatively small yet informative number of spectral images were collected in specific spectral channels using our technology. Consequently, the data for a single field of view can be acquired in less than 3 minutes and data analysis required to predict tissue labels using pre-trained classifiers is on a real time basis. This time factor is critical when the technology is used in the operating theatre.

This study can be considered as the first step towards clinical application, using the concept that proposed technology with artificial intelligence is capable of detecting and discriminating between normal and neoplastic eye tissue with a reasonable degree of accuracy. The surgeon could easily identify normal areas before photographing the suspect area at the slit lamp. As this study was conducted on paraffin embedded ex vivo tissue, the results are valid only in this scenario and evaluating in vivo application of the proposed technology is a potential extension of this study. Investigation of the effect of tissue fixation was beyond of the ethics and the scope of this study and was not considered in this thesis. At the current stage, the system cannot be applied directly for live patients and it needs some further measure such as optimizing the system. Although the same classifiers may not be applicable for the live tissue, the technique and approach to train the classifiers are the same for live tissue and patients. Another possible extension of this study could focus on discrimination between eye lesions of different grades including aggressive lesions, requiring action and resection, and benign cases. In this study, benign entities such as pterygia [155] were not evaluated. As a future study, pterygia could be investigated based on the technique described

here, as it is likely that pterygia will have a different chemical composition when compared to normal cells. In fact, this study could be extended to classify tissue in three different groups, comparing normal tissue, pterygia and neoplastic tissue.

#### **4.6 CONCLUSION**

In this study quantitative analysis of the spectral images extracted a strong multispectral signature of a relative difference between neoplastic and normal tissue both within each patient (at  $p < 0.0005$ ) and between patients (at  $p < 0.001$ ). The fully automated diagnostic method developed in this study based on machine learning produces maps of the relatively well circumscribed neoplastic-non neoplastic interface. Generally, OSSN could be detected using multispectral analysis in all patients investigated here. we got ~94% correlation in intra patient classification and ~78% in inter patient classification. Such percentage shows the pixels with the same label in both histology and multi spectral analysis This study shows the feasibility of using multispectral auto-florescence imaging to detect and find the boundary of human OSSN. One of the limitations of this study is the limited number of the available patients which is considered for the future study.

#### **4.7 AUTHOR CONTRIBUTION**

A.H, C.B, and E. G. designed the study, developed the methodology and drafted the manuscript. A. H. acquired the data, processed images, performed image registration, biostatistics, computational and machine learning analysis. A.H, C.B, A. Allende, A. Anwer and E.G interpreted and discussed the data. A. Allende performed histological analysis and interpretation. C.B, A. Allende, A. Anwar, A.H and E.G were responsible for technical, material and administrative support. C.B and E.G supervised the study. All authors critically reviewed the paper.

- Disclosure of funding

This work was supported by Personal Eyes Pty Ltd [Ref: 5201600708]; the ARC Centre of Excellence for Nanoscale Biophotonics [award CE140100003] and the iMQRES scholarship.

- Acknowledgment

The authors would also like to thank the Human Ethics Faculty Sub-Committee who reviewed the ethics application (No. 5201600708). We thank Dr Con Petsoglou from the University of Sydney for his kind material support of this work.

- Disclosures/conflict of interest

Nil

## 4.8 SUPPLEMENTARY NOTES ASSOCIATED WITH METHOD SECTION

### 4.8.1 A schematic for customized auto-fluorescence imaging system setup

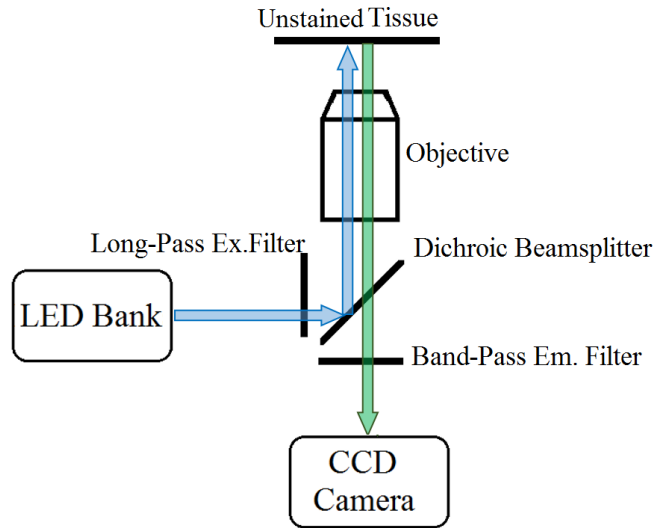


Figure 30. Customized auto-fluorescence imaging system setup. One LED with specific wavelength excites the tissue at a time. The scattered auto-fluorescence light pass a channel pass filter and captured by a CCD camera. This process repeats until all the spectral images are taken.

### 4.8.2 Treating spikes

To remove the spikes, saturated or dead pixels, a 'threshold limiting window' is scanned over the spectral channel images to locate them; these spikes are then replaced by values interpolated from the immediately adjacent pixels [129]. To provide more details, removing spikes has two stages: identification and replacement. This effect can be inspected via histogram distribution of a spectral channel as cosmic rays do not follow a typical Gaussian distribution of the pixel and have high values compared to other pixel. A typical way to detect cosmic rays is to find them through image histogram globally and just consider the pixel values which are much higher than the most frequently occurring pixel value (mode) of all pixels.

In this study to advance the analysis and make sure to find all of the cosmic rays, we find the cosmetic rays locally and applied a 'threshold limiting window'

scanned over the spectral channel images. After evaluation different dimension, a scanning window with 20x20 was selected, which had an acceptable performance to find all of the cosmetic rays for different spectral images. After finding the cosmic rays, they needed to be replaced by a close value to the adjacent pixels to avoid over smoothing the pixel. Hence, only 8 immediate pixel were chosen to interpolate. More information has now been provided in supplementary material.

#### **4.8.3 Treating Poisson's noise**

The aim for de- noising was to treat Poisson's noise for improving signal to noise ratio with minimal loss of the informative part of signal. To do this aim, wavelet transformation was used. Wavelets with different mother functions, including Haar, Daubechies, Symmetric, Biorthogonal, were tried and finally a customized wavelet filter with symmetric Mother function was used [128, 130] which fitted well to spectral channels rather and could successfully improve the signal to noise ratio. Each spectral channel image was divided into different scale components with specific coefficients generated by the wavelet function. To ensure that the minimum amount of information loss happens, wavelet function with different levels was tried and finally 3 levels was set. Finally, by applying an inverse wavelet transform, the de-noised image was reconstructed [131].

#### **4.8.4 Treating background**

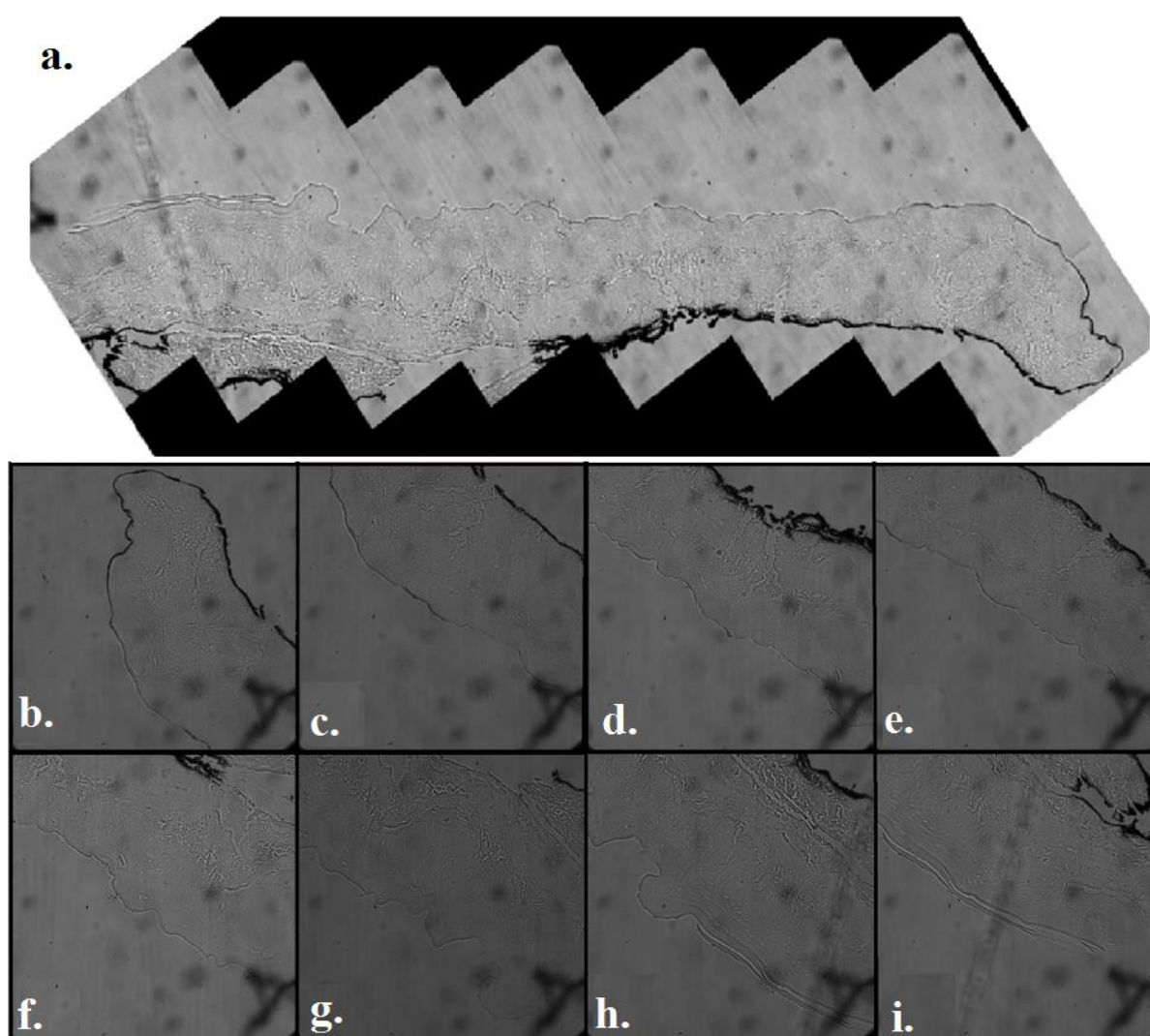
The background signal is a non-informative portion of the spectral images, which is unrelated to the auto fluorescence emission of the tissue sample. To capture the background reference signal, a series of spectral images acquired from an empty sample. These reference background images were de- noised and then subtracted from images taken from samples, channel by channel, to correct the back ground from tissue images.

#### **4.8.5 Treating uneven illumination**

To capture the illumination curve of each spectral channel, a series of spectral images acquired from a reference calibration fluid which has a fluorescence signal in all of the channel. Then, the calibration image were de - noised and its background

were corrected. Finally, the images from samples were divided by the calibration image, channel by channel. Such technique removes spectral distortion created by uneven illumination of the field of view [95].

#### 4.8.6 Image stitching



**Figure 31. DIC image registration for a sample patient.** To enable the inspection of larger fields of view, a series of overlapping microscopy images were taken (b-I). Then, the affine algorithm was applied to co-register the overlapping images based on an image taken by 5x objective with a bright field microscope. Finally, a composite image was generated (a).

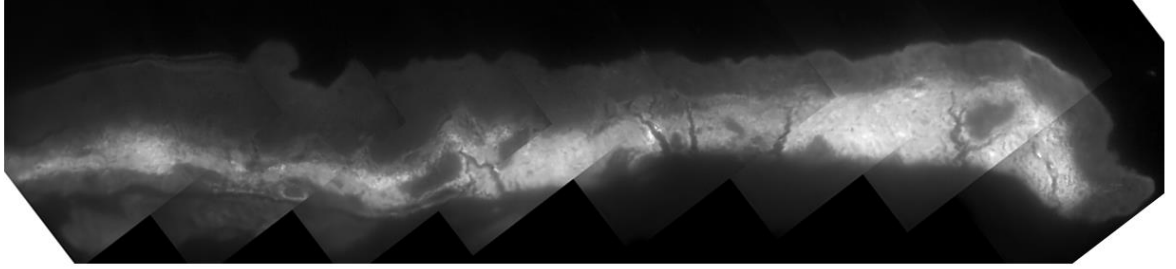


Figure 32. a spectral image after being pre-proceed and registered. (EX:  $378 \pm 10$  nm, EM: 454-496)

#### 4.8.7 Generating sectors and spectral signature

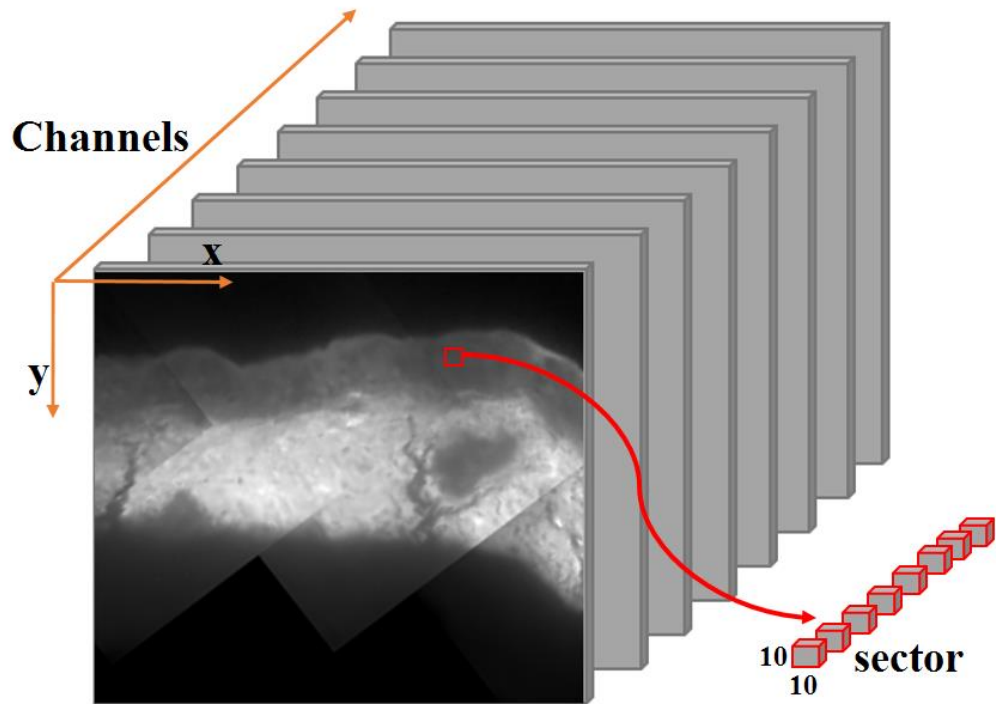
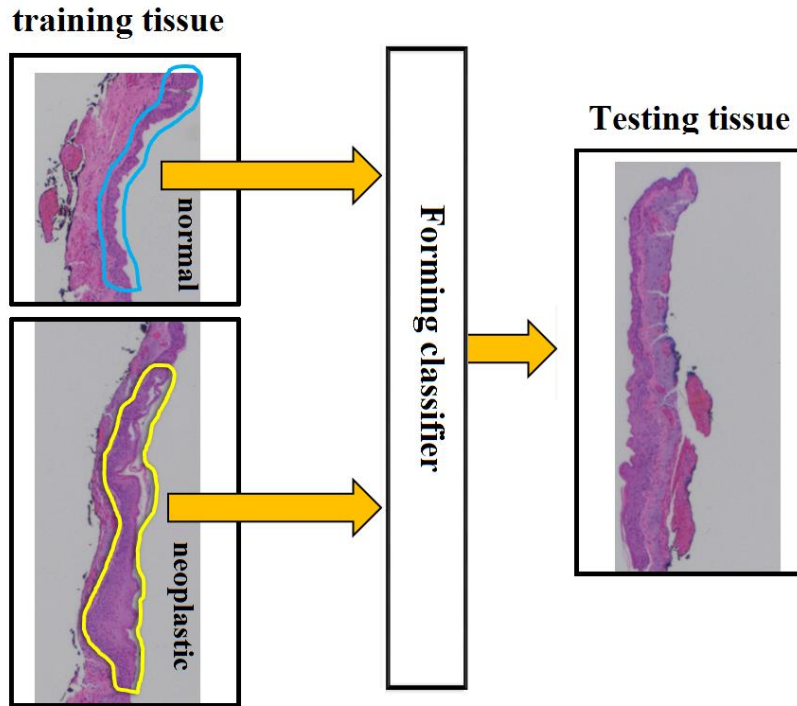


Figure 33. Extracting sectors from multi spectral images.

#### 4.8.8 Testing and training tissue for a sample patient in intra-patient framework.



**Figure 34.** Testing and training tissue for a sample patient in intra-patient framework. Tissue from one single patient is first divided into two separate sections called the training and testing tissue. The training tissue has both normal and neoplastic areas, the availability of which makes it possible to extract the multispectral neoplastic signature. The trained classifiers were then applied to the data blocks of the testing area of the same patient for the label prediction (“neoplastic” vs “non-neoplastic”), which resulted in a false colour map and related statistical metrics.

#### 4.8.9 t-test

To apply t- test in this study, first the sectors extracted from the training area and then spectral signature of the sector were obtained. The spectral signature for each sector is a feature vector with 38 elements corresponding to 38 spectral channels. Then the spectral signature of sectors were grouped to “neoplastic” and “normal” sets depending to the each sector. Finally, t- test was applied on the spectral signature of sectors, element by element. Therefore, for each feature (each element of the spectral signature), t test was applied once and to verify if there is a statistically significant difference between neoplastic and normal groups. In the better words, T test proves that the features extracted from the spectral images are significant for two groups. A significant difference can be regarded as indicative of a strong multispectral signature



of a relative difference between neoplastic and normal tissue from an individual patient.

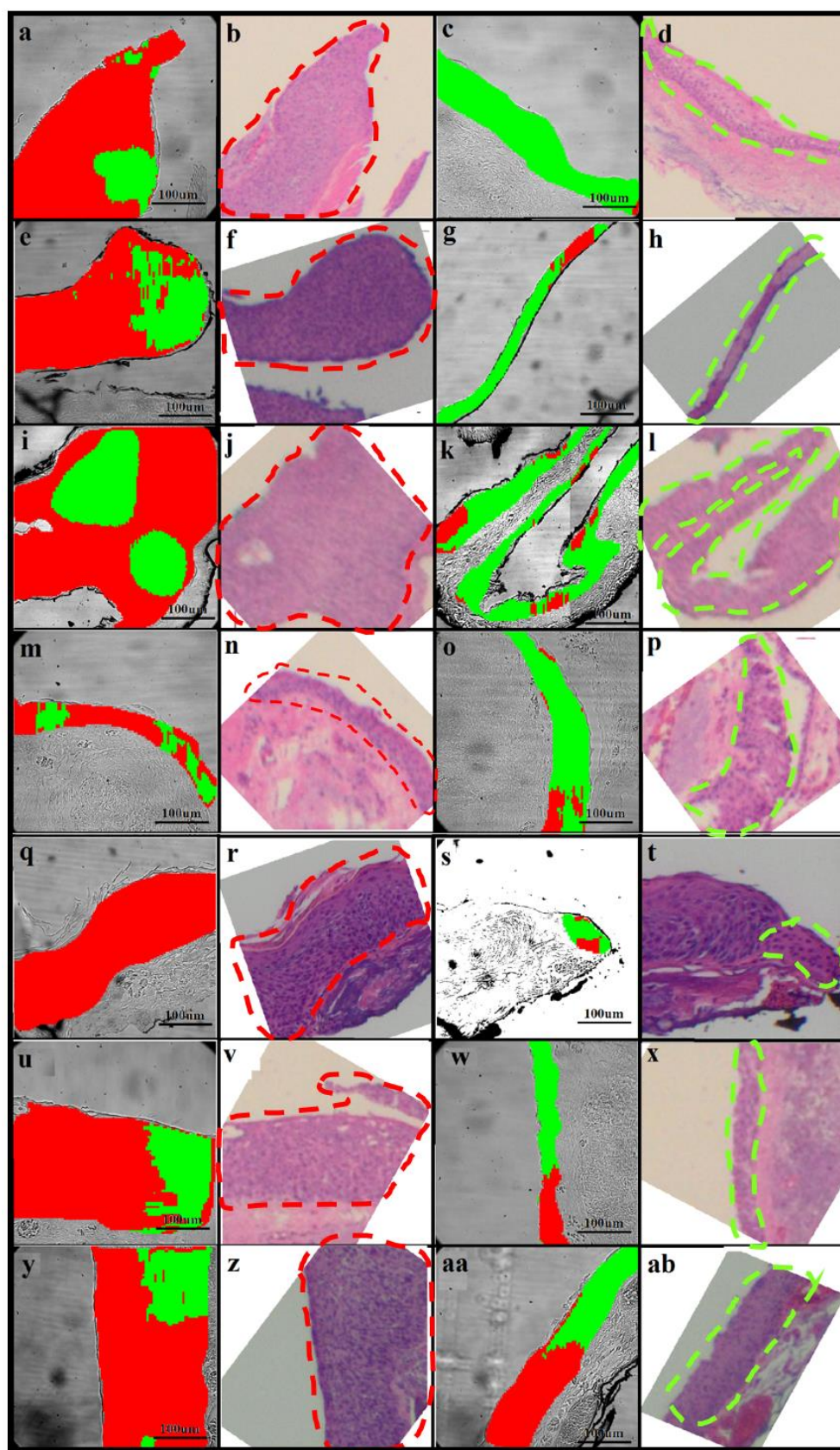
#### **4.8.10 Principal component analysis (PCA)**

PCA is a linear transformation which reflects the feature vectors obtained from sectors into a new uncorrelated orthogonal space. The training data were projected onto the discriminative space created by the PCA analysis and the uncorrelated scores were calculated. PCA was used to de-correlate the image data across spectral channels. PCA projects the feature vectors onto the eigenvectors of the covariance matrix, illustrating informative variations of uncorrelated data.

#### **4.8.11 Machine learning performance evaluation**

After verifying the robustness of the spectral signatures using t test, machine learning methodology were utilized. In the training process, K fold (K=10) cross validation was employed to monitor the performance of machine learning methods. To this aim, the training data was randomly disaggregated into 10 equal size folds. Each fold was used to evaluate the classifier algorithm trained by the 9 remaining folds. The process ran for 10 iterations, and all the accuracies obtained were averaged to calculate the overall performance. This was then repeated for the classifier algorithm using different thresholds to generate the 'receiver operating characteristic' (ROC) curve, which leads to an optimized training process [145]. The learning methods were also statistically evaluated based on area under ROC curve (AUC). AUC values close to 1 shows the excellent performance of the classifier. For intra patient classification where a classifier is designed particularly for a patient, the training time efficiency was also considered.

#### 4.9 SUPPLEMENTARY NOTES ASSOCIATED TO RESULT SECTION



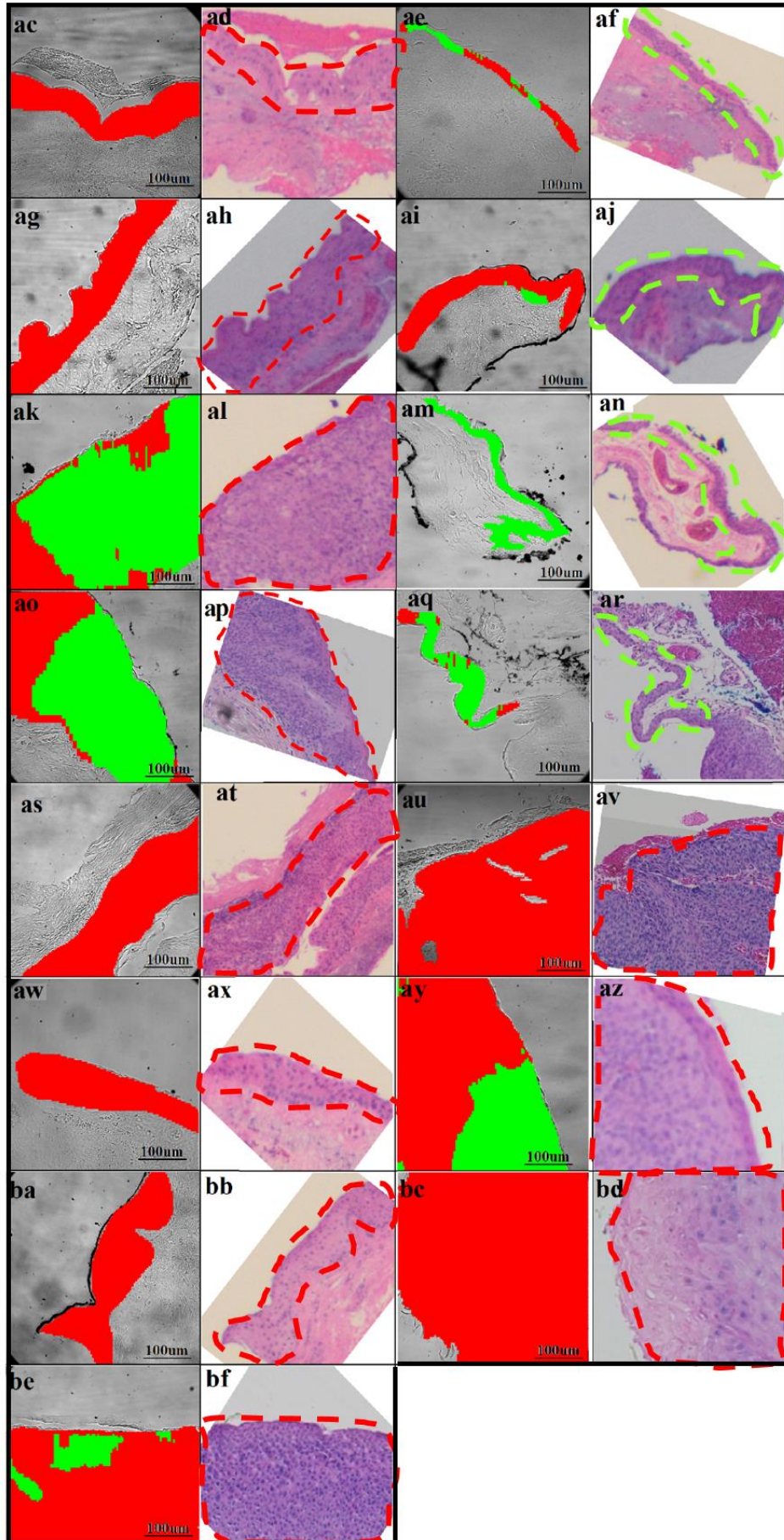




Figure 35. False colour map superimposed on the DIC image of testing patients in the inter-patient classification for different patients with corresponding H&E images with dash lines to highlight boundary of normal and neoplasm tissue with green and red color, respectively. Data blocks are coloured in red or green if they are respectively predicted to be neoplastic or normal. First/third column is the multispectral false colour map and the second/fourth column is the corresponding H&E image. (a),(b) and (c),(d) neoplastic and normal section for PID 1, respectively. (e),(f) and (g),(h) neoplastic and normal section for PID 2, respectively. (i),(j) and (k),(l) neoplastic and normal section for PID 3, respectively. (m),(n) and (o),(p) neoplastic and normal section for PID 4, respectively. (q),(r) and (s),(t) neoplastic and normal section for PID5, respectively. (u),(v) and (w),(x) neoplastic and normal section for PID 6, respectively. (y),(z) and (aa),(ab) neoplastic and normal section for PID 7, respectively. (ac),(ad) and (ae),(af) neoplastic and normal section for PID 8, respectively. (ag),(ah) and (ai),(aj) neoplastic and normal section for PID 9, respectively. (ak),(al) and (am),(an) neoplastic and normal section for PID 10, respectively. (ao),(ap) and (aq),(ar) neoplastic and normal section for PID 11, respectively. (as),(at) neoplastic section for PID 12. (au),(av) neoplastic section for PID 13. (aw),(ax) neoplastic section for PID 14. (ay),(az) neoplastic section for PID 15. (ba),(bb) neoplastic section for PID 16. (bc),(bd) neoplastic section for PID 17. (be),(bf) neoplastic section for PID 18.

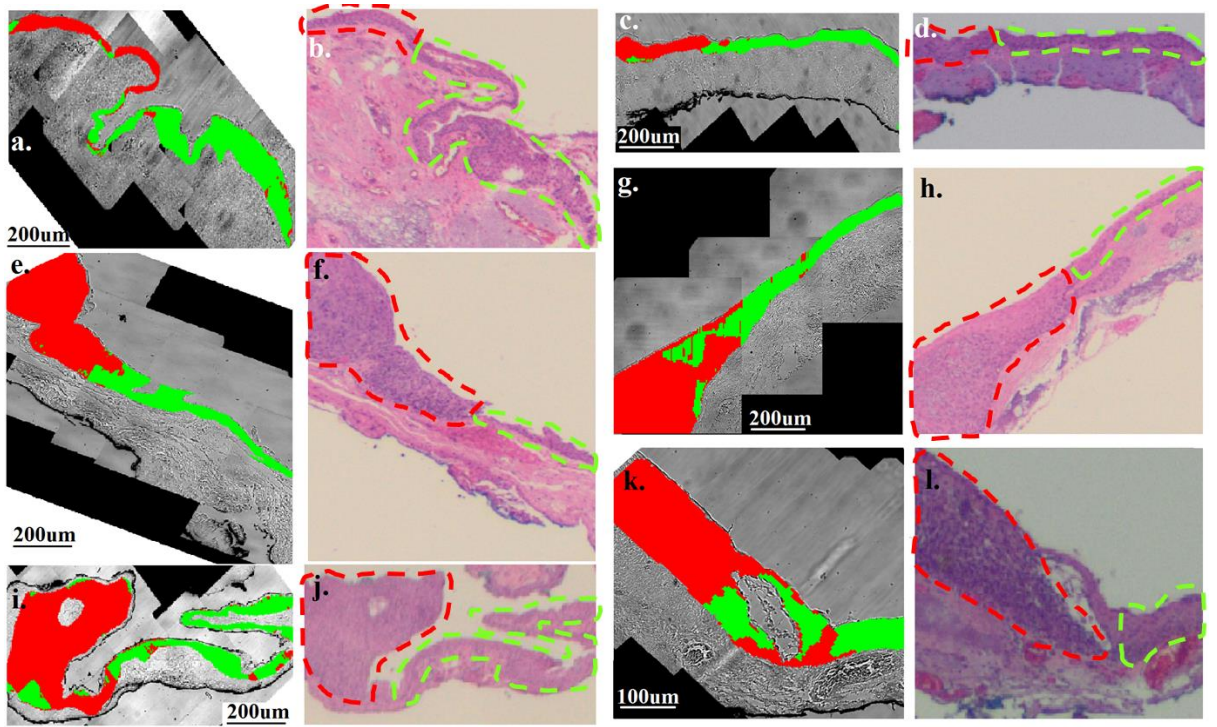


Figure 36) False colour map generated to locate the neoplastic boundary on the testing tissue in the intra patient classification framework in comparison with the corresponding histology images with dash lines to highlight boundary of normal and neoplasm tissue with green and red color, respectively. . The position of the block data on the DIC image is coloured in red or green if they are predicted to be neoplastic or normal, respectively. First/third column is the multispectral false colour map and second/third column is corresponding H&E section. (a) and (b) associated to PID4, (b) and (c) associated to PID9, (e) and (f) associated to PID 10. (g) and (h) associated to PID 1. (i) and(j) associated to PID 3, (k) and(l) associated to PID7.

# Chapter 5: CHANNEL SELECTION AND SYSTEM OPTIMISATION

---

This chapter describes the methodology adopted by this research to optimise the multispectral imaging for detection of OSSN in terms of the number of channels and progress the technology one step further for real-world clinical application. This chapter has been written in a manuscript format to be submitted in ‘Investigative Ophthalmology & Visual Science’ which outlines the background and the purpose of the chapter in section 5.1. The methodology of the research is described in 5.2. Section 5.3 and 5.4 represent the results and discussion, respectively.

## Abstract

Clinical OSSN diagnostics by non-invasive spectral imaging of eye autofluorescence must be rapid enough to be comfortable for patients – without sacrificing accuracy. This requires identifying optimised spectral signatures of OSSN. Here, we identified such signatures using a data-driven methodology of swarm intelligence. Ten patients with histopathological diagnosis of ocular surface squamous neoplasia (OSSN) were recruited. Their unstained biopsy OSSN specimens were investigated using a custom-built autofluorescence multispectral microscopy imaging system. The images were taken in 38 spectral channels spanning specific excitation (340nm-510nm) and emission (420nm-650nm) wavelength ranges. To identify optimised OSSN spectral signatures, swarm intelligence was combined with discriminative cluster analysis. Three evolutionary strategies including particle swarm intelligence, differential evaluation and ant colony optimization were employed to discover a group of interacting channels yielding the optimised OSSN spectral signature. To validate the results and assess the generality power of our OSSN spectral signature, external and internal cross-validations were employed. Our study found optimized spectral signature of OSSN allowing rapid diagnostic imaging in clinical settings and showed the feasibility of using optimized multispectral auto-florescence spectral signature to detect and find the boundaries of human OSSN. This advance will facilitate translating the multispectral imaging technology for non-invasive OSSN diagnostics.

Keywords: Ocular surface squamous neoplasia (OSSN), Autofluorescence spectral signature, Swarm intelligence

## 5.1 INTRODUCTION

Conventionally, autofluorescence imaging technologies such as fluorescence lifetime imaging microscopy (FLIM) have been limited to employ only a few costly channels ( $N < 3$ ) [156, 157]. New cutting-edge advancement of autofluorescence sensing technology has introduced an unlimited number of spectral channels, which could track several native fluorophores simultaneously with denser spectral resolution. Such innovative technology employs a number of channels to cover a broad spectrum ranging from deep UV to infrared wavelengths [11, 95]. Recently, such multispectral autofluorescence imaging with 38 spectral channels showed successful results in detection and boundary delineation of Ocular surface squamous neoplasia (OSSN), [99] which is referred to the most common neoplastic eye surface disorder [100, 101].

Although multispectral imaging provides the opportunity of using tens of channels with unrestricted spectral specifications to detect OSSN, the number of channels and associated multispectral signature need to be optimised to facilitate translating the technology for a real-world ophthalmology application. Each channel employed to the multispectral imaging imposes additional cost and extra acquisition time. Irrespective to the cost, time plays a crucial factor for precise OSSN detection for a cautious patient whose eyes may move. Such an unintentional movement can be minimised by reducing the acquisition time.

In fact, We have been able to identify the OSSN signature using 38 spectral channels detecting fluorescent signals from native fluorophores such as protoporphyrin IX (PPIX), reduced nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide (FAD)[158]. The relative content of these native fluorophores were reported to be modified in cancer due to cancer-induced transformation in cell metabolism[20, 42, 52, 64]. However. This technology is required for an efficient and patient-friendly diagnostic, because imaging in each channel at the maximum eye-safe excitation limit requires several seconds. As a result, employing tens of channels leads to a long imaging time, during which unintentional eye movement and blinking cause unavoidable interference such as image shifts and defocusing [159]. Therefore, to enable clinical translation of the technology and its

deployment for ophthalmologic application, a robust methodology must be developed to identify the smallest number of imaging channels with optimised spectral specifications which can accurately determine the spectral signature of OSSN.

Prejudgment [160] is a simple way to optimise multispectral imaging technology. Prejudgment suggests tracking specific biomarkers previously discovered and design a channel arrangement based on their excitation-emission spectra. However, such a methodology has two limitations. First, the multispectral imaging becomes restricted to known biomarkers, and there is no opportunity to find new and potentially more indicative channels. Secondly, excitation-emission spectra of a fluorophore are not a reliable source to select channels, as they may change due to various factors such as change in PH [32] as described in section 2.3. Therefore, prejudgment cannot be a proper technique to select the indicative channels, as there are various known and unknown factors that influence the usefulness of the channels for a specific diagnostic application. Hence, the best option is to investigate the complex aspects of the disease using all channels and then choose the effective ones directly from some patients before translating the technology to a real clinical application.

As another option, a channel subset can be selected by scoring the robustness of each channel to detect diseases independently of other channels[161, 162]. Although beneficial for removing very irrelevant channels, this methodology cannot be useful to select an efficient subset of channels. In fact, due to possible mutual information between close channels[11], grouping individually good channels cannot necessarily result in an efficient classification performance[163].

Channel selection for OSSN classification using multi-spectra autofluorescence imaging can be more challenging than other classification problems, as the inter-patient variability may have a considerable influence on cancer spectral signature. It is likely that some channels may be affected by factors unrelated to cancer transformation such as age, gender, race, smoking status, and more especially, eye colour and various amount of sunlight exposure[164]. Such side effects can potentially touch the spectral signature of the tissue and subsequently interferes in the channel selection process.

In this study, we investigate the problem of channel selection to detect OSSN to propose a few indicative channels which together can extract the spectral signature of the OSSN effectively. This method showed promising results in similar application



[165]; however, its application in biomedical multi spectral imaging needs to be investigated further. This technique for channel selection is proposed which addresses the above-mentioned challenges as follows:

- 1- From the entire set of channels, we search for an indicative combination of good channels and consider their performance as a whole. Channels are selected based on cluster analysis by considering channel's contribution to the quality of clusters found by discrimination analysis.
- 2- A creative approach is used to minimise the influence of the side effects unrelated to cancer transformation for channel selection.

In this study, the channel selection process involves a selection criterion and searching algorithms. First, diseased (OSSN) and normal clusters are formed based on discrimination analysis [166], and then a standard criterion function to evaluate the quality of clusters is defined as a selection measure. To minimise side effects, the criterion function is calculated on intra-patient basis and then averaged over several patients. Such methodology fuses the advantage of both interpatient and intra-patient frameworks and keeps the criterion function almost safe from being affected by inter-patient variability. To optimise the system, swarm intelligence[167-169], which is referred to the collective intelligence behaviour of a group of naïve agents [170, 171], is used. The performance of various searching algorithm based on swarm intelligence, including particle swarm intelligence (PSO) [172-174], differential evolution (DE) [175] and ant colony optimisation (ACO)[176, 177] are evaluated to find the good channels optimising the criterion function. To investigate the generality power of the proposed method, testing patients which did not have any contribution to developing the model were evaluated.

## **5.2 METHOD**

### **5.2.1 Sample collection**

Our analysis of patients' samples was performed under the permission from Macquarie University Human Research Ethics Committee, reference No: 5201600708 and informed consents were obtained. Eye biopsies from OSSN positive patients were collected as part of the routine medical procedure, and they were analysed at the Histopathology department/ Douglas Hanly Moir Pathology, Sydney. Each biopsy

diced into two sections, one of which was used for multispectral analysis with no staining. Another cut was used to produce a matched hematoxylin-eosin (HE)-stained slide as a reference to locate the diseased and normal areas of the tissue and served as our gold standard. As the stained and non-stained section were close matches, the associated region of interest was also found on the unstained tissue. In this work ten patients were required with both normal and OSSN samples.

### **5.2.2 Image acquisition and preparation**

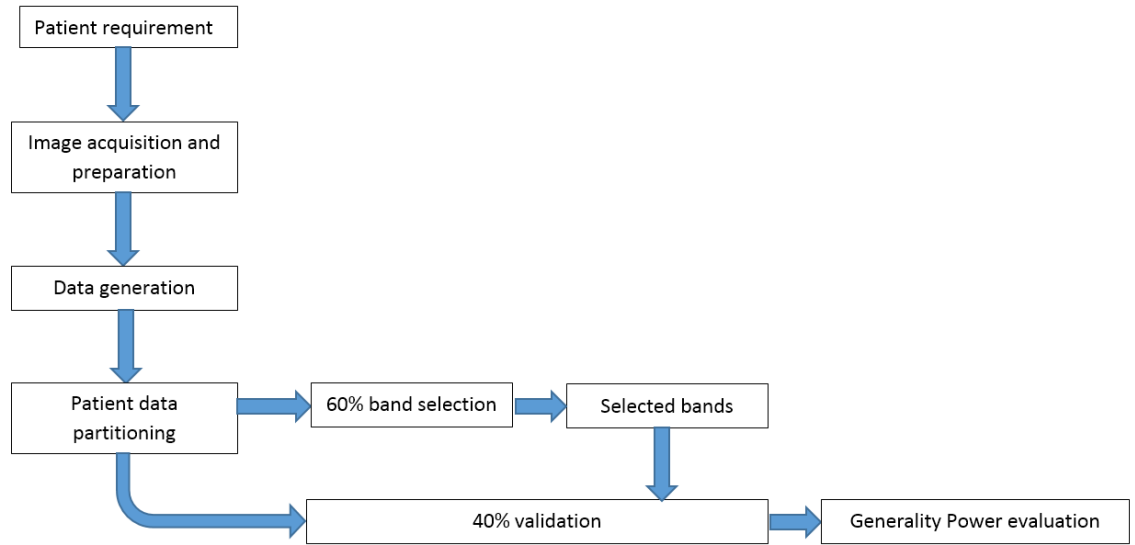
The images were captured using an adapted IX83 Leica microscope, at 40 x magnification. A camera with high quantum efficiency is used to acquire fluorescence images. Multispectral imaging involves generating a sequence of fluorescence images of the same sample area in each of these channels. The normal and cancerous sections for each patient were imaged with 38 channels covering a range of UV to red. Then the images were pre-processed to remove the background and light curvature. Random errors include Poisson noise, dead pixels, spikes due to cosmic rays, analogue-to-digital converter errors, bit errors in transmission, and similar sources [89, 95]. To remove spikes, saturated or dead pixels, a 'threshold limiting window' is scanned over the spectral images to locate them; these spikes are then replaced by values interpolated from the immediately adjacent 8 pixels [129]. To remove the Poisson noise, a customised wavelet filter with symmetric Mather function was used [128, 130]. Apart from random errors, two systematic errors make a contribution to the spectral images, namely the fluorescent background (from the microscope optics and the sample slide) and uneven illumination of the field of view [11].

### **5.2.3 Data generation**

To generate data points, images acquired from non-stained tissue for each patient were binned with 10×10 pixels for all the channels. As such, many data points (>10k) from each patient were obtained. Such binned pixels were extracted and served data points, each of which has 38 dimensions corresponding to the number of channels. Each data point was labelled as diseased or normal depending on the pathology assessment. Therefore, each patient had two groups of data points including diseased and normal sets.

### 5.2.4 Channel selection algorithm

In this study, the patients were divided into two groups randomly: 60% of patients ( $N=6$ ) used for developing channel selection algorithm, and 40% of the patients, called testing patients, were put aside for blind testing. The testing patients do not contribute to developing the channel selection algorithm. The procedure is shown in Figure 37.



**Figure 37. Channel selection strategy**

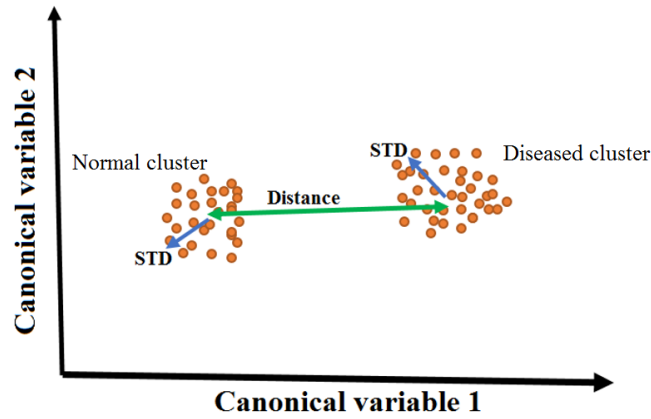
The proposed channel selection algorithm was performed by making use of cluster analysis and swarm intelligence. First, Diseased and normal clusters were found using linear discrimination analysis[166]. Then, different search studies based on swarm intelligence [171] were employed in the space of possible channel subsets, guided by the quality improvement of the clusters. Hence, the channel selection methodology involves a criterion function and search algorithms to optimise the criterion function. In the following, initially, the criterion function is described, and then the way of its optimisation is explained.

### 5.2.5 Criterion function

This study employed an effective criterion function based on cluster analysis. First, the data points are reflected into a discriminative space to find the diseased and normal clusters. Such space was created based on linear discrimination analysis using two canonical variables[166]. The canonical variables are a linear combination of the

selected channels. Figure 38 represents the parameters used to define the discrimination power of clusters. The blue arrows are the standard deviation, and the green arrow is the centroid distance.  $QC$  (see Eq.5) is a measure which evaluates the ratio of within-cluster variance and between cluster variance. To improve the discrimination power of clusters, the  $QC$  value needs to be minimised.

$$QC = \frac{STD_1^2 + STD_2^2}{Dis^2} \quad \text{Eq. 5}$$



**Figure 38. Standard deviation and centre distance of the clusters as the criterion function**

In Eq.5,  $STD_1$  and  $STD_2$  are the standard deviations of diseased and normal clusters, respectively, and  $Dis$  is the centroid distance of the two clusters. To have quality clusters,  $QC$  should be minimised.

In this study, to deal with side effects, we fused inter and intra-patient frameworks to get the advantage of both frameworks while minimising their limitations. We calculated the  $QC$ s for patients, one by one. In this case, the criterion is mostly focused on cancer retransformation, as other side effects such as age, eye colour, etc. are almost constant for one patient. Then, we calculate a global criterion function using the bellow equation over a number of patients to consider different aspects of disease and develop a powerful criterion function.

$$Criterion = \frac{Std(QC_1, QC_2, \dots, QC_n)}{\sum_{i=1}^n \left( \frac{QC_i}{MQC_i} \right)^2} \quad \text{Eq. 6}$$

where  $QC_i$  is calculated for patient number  $i$  using a subset of channels. To normalise  $QC_i$ , it was divided by  $MQC_i$  which is corresponding  $QC_i$  when all of the channels are used. Then, an indicative subset of channels would be selected by minimising the criterion function

### 5.2.6 Search algorithms

After defining a criterion function, swarm intelligence [171] is applied for optimisation. There are various swarm intelligence techniques; therefore, several methods including ant colony optimization (ACO)[176], particle swarm intelligence (PSO) [172] and differential evaluation(DE) [175] were used in this study and their performance were evaluated.

Ant Colony Optimization (ACO) is a metaheuristic tactic suggested by the foraging behaviour of real ants. This approach has four main contributors including ant, pheromone, daemon action, and decentralised control. In this approach, a simple agent called an ant simulates the survey and exploitation in search space created by channels and follows a direction marked by most other ants. This simple concept leads to ants' collaboration to find the shortest and optimised paths[176, 178]. Particle swarm intelligence (PSO) was inspired by birds' behaviour, consisting of separation, alignment, and cohesion. The PSO algorithm starts with a number of simple agents, and the fitness value of each agent is calculated followed by determining individual and overall bests. Next, the velocity and the position of each agent is updated until the system gets optimised through attracting all agents towards the best answer[178, 179]. The Differential Evolution (DE) algorithm is a population-based algorithm which starts with initialising a population and then the fittest members are evaluated. Next, a new channel set is generated based on mutation. Such a channel set is mixed through the crossover, and the algorithm runs until we get the optimised solution [178].

## 5.3 RESULTS

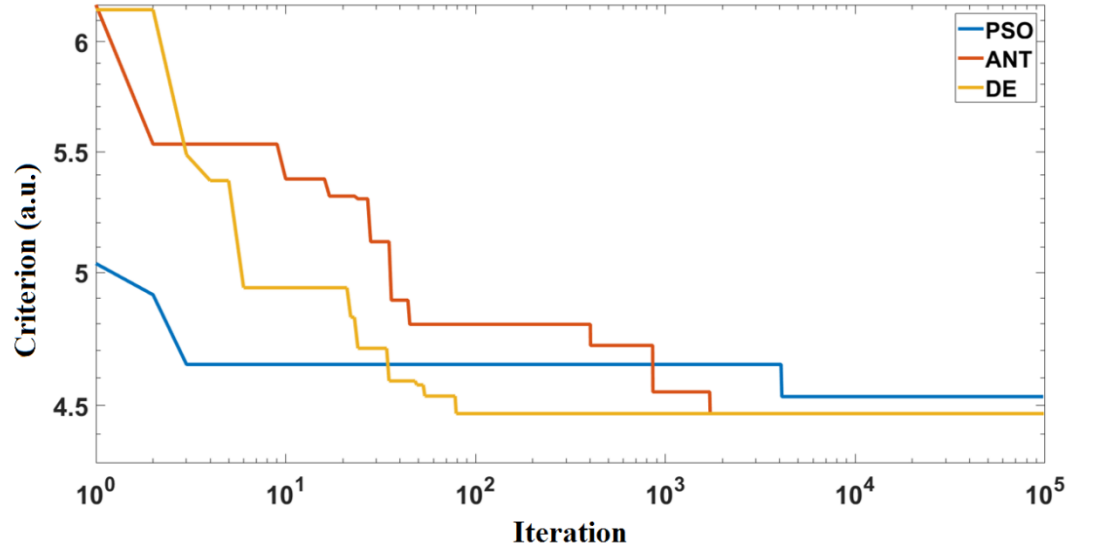
The objective of this study is to identify a channel subset that distinguishes diseased (OSSN) and normal tissue effectively. Initially, the number of channels is restricted to a specific number, and the minimal criterion value and the corresponding

channels will be found (section 5.3.1). Then, the same procedure will be repeated for a various number of channels to discover an optimal channel subset (section 5.3.2).

### **5.3.1 Finding a subset of indicative channels with a particular number of elements**

After selecting 60% of patients ( $N=6$ ) randomly from the available cohort to develop the methodology, the data points for each patient were extracted and grouped into normal and diseased sets. To find the clusters and calculate the criterion, the data points for each patient needed to be projected into a discrimination space created by canonical variables. To provide the methodology with generality power, a cross-validation approach was used[144]. The data points for each patient were divided into 50% for designing the discriminative space spanned by canonical variables and 50% to calculate the criterion value. So, the best discrimination space which can separate normal and neoplastic data points were found using 50% of the data and then the remaining data points were reflected into the generated space to calculate  $QC$ . This process was repeated for six patients, and finally, the criterion function value was obtained over all six patients.

Assuming that the required number of selected channels is  $K$ , which is less than the entire channels, so the  $K$  channels were selected to produce a discrimination space for each patient. These  $K$  channels were combined linearly to form canonical variables, and clusters were generated that best classify normal and diseased groups. In this section, we present the results in detail for  $K$  ( $K=6$ ) channels and show how the clusters and criterion function are changing with a different combination of 6 channels. PSO, ANC and DE were applied to minimise the criterion function and find the subset of channels. Algorithms were run until leading to a convergence, that no variation can be identified even after many iterations (number of iteration  $>10^5$ ).

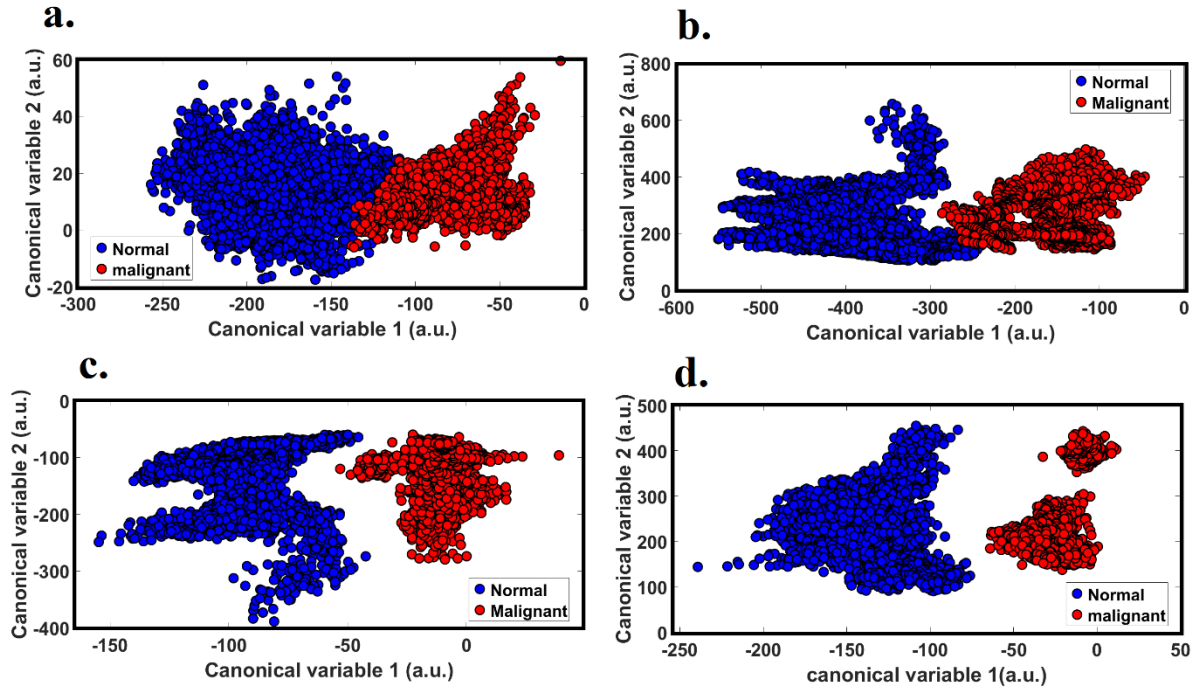


**Figure 39. Criterion function improvement using PSO, ANT, DE after iteration**

Figure 39 represents the performance of the three optimisation algorithm used in this study to find the best channels over 100k iterations. It shows that PSO could minimise the criterion to 4.53 and both DE and ANT colony minimised the error to 4.47. It is worth mentioning that the channels selected by DE and ACO were exactly the same after  $10^4$  iterations with the same criterion value. Figure 39 also illustrates that PSO rapidly gets a reasonably good outcome. In fact, after only 20 iterations a good subset was found with a criterion value of 4.649, just 4% above the minimum value which is 4.472. Results show that PSO is a perfect candidate when we want to get the results after a limited number of iterations. However, Figure 39 shows that PSO was not able to find an optimum value after  $10^5$  iterations compared to DE and ACO. Looking at DE and ACO reveals that both could successfully find an optimum subset of channels. Comparing DE and ACO colony shows that DE could come to a conclusion much faster than ANT colony, as DE could reach 4.47 after 82 iterations and kept that value for all 100k iteration.

To show how the quality of clusters with various iterations and corresponding channels changes, the clusters of a sample patient used in developing the algorithm are shown in Figure 40. This figure represents the discrimination space generated by channels suggested based on DE after 10, 30, 100, 10k iterations. The proposed channels were taken, a discrimination space was generated, and then the data points associated with a sample patient were reflected on the space. As is shown, the quality

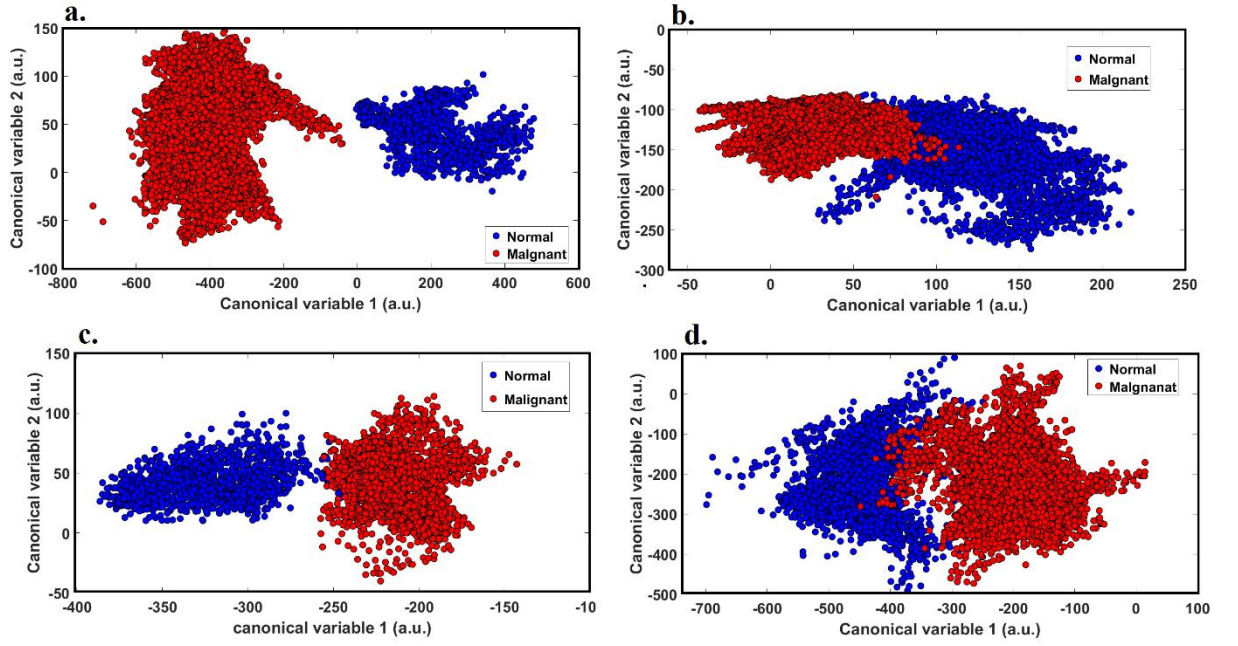
of the clusters improved as the number of iterations increased and the criterion function is minimised.



**Figure 40. Cluster improvement due to selecting more effective channel subsets: (a) after 10 iterations, (b) after 30 iterations, (c) after 100 iterations, (d) after 10k iterations.**

After finding the candidate channels, they were evaluated based on the testing patients, put aside for the blind test. The selected channels were extracted from the data points associated with the blind-test patients, and corresponding data points were reflected on the discriminative space. Figure 41 demonstrates the performance of the optimised channels for four testing patients. It is clear that those channels are effective for testing patients, as we have separated clusters for malignant and normal data points. Figure 41 validates the generality power of the methodology.

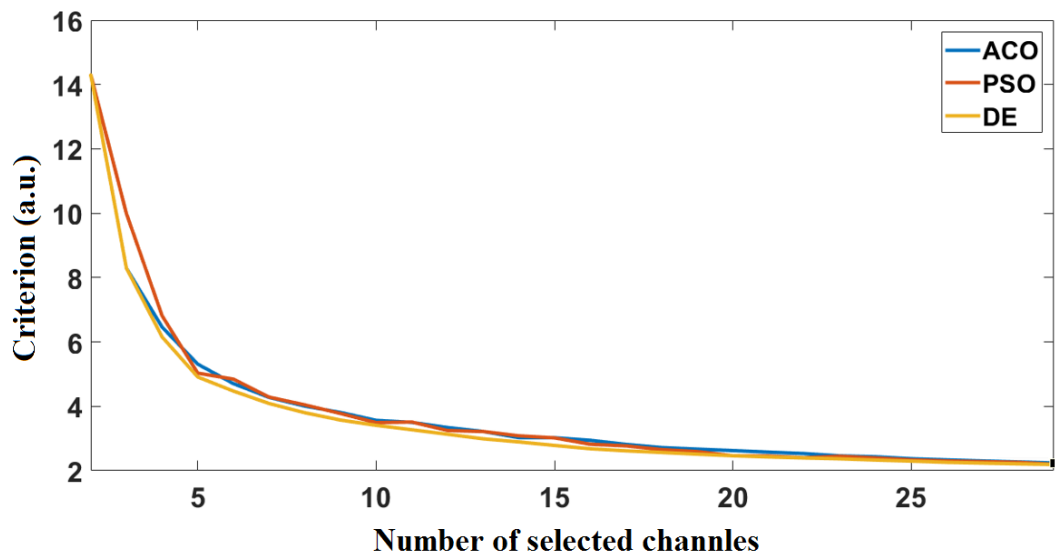




**Figure 41. Validating the selected channels on blind test patients. (a) patient no.1 (b) patient no.2 (c) patient no.3 (d) patient no.4.**

### 5.3.2 The optimum number of channels for OSSN detection

After developing the methodology to select a particular number of channels, we applied it to a range of a different number of channels to find the optimum number. Therefore, the minimal criterion value for the optimum combination of channels with a different number ( $N \in \{2, 3, \dots, 30\}$ ) was calculated using three searching algorithms as shown in Figure 42.

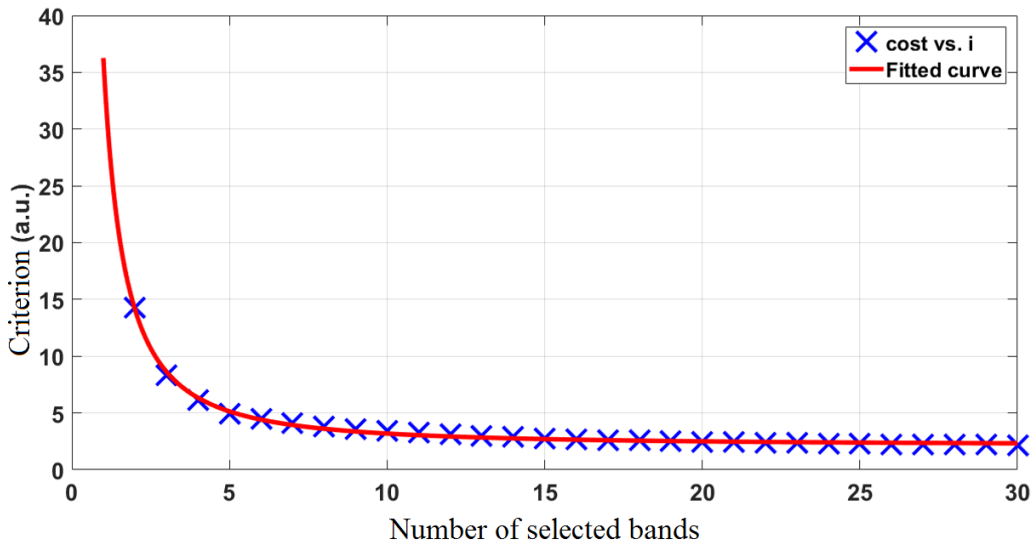


**Figure 42. Criterion function vs the number of selected channels**

Figure 42 shows that at different numbers of channels, ACO, PSO and DE show the same trend yet slightly different performance to minimise the criterion. Generally, DE shows better performance rather than ACO and PSO. For less than five channels ACO has better performance than PSO. To synergise the performance of different algorithms, the outcome of three searching algorithms were fused to generate one combined algorithm which demonstrates the best subset of the channels for all of the numbers. In fact, for each number of selected channels, the outcome of the lowest criterion function is considered using the function:

$$S_i = \min\{ACO_i, PSO_i, DE_i\} \text{ Eq. 7}$$

Consequently, the fused curve, which gives the lowest criterion of the three curves, was generated as shown with blue crosses in Figure 43. In this case, DE was demonstrated the best performance mostly in different number of channels and data fusion could select its outcome. As is demonstrated in Figure 43, with an increasing number of channels, the criterion decreases. However, the rate of improvement drops after selection of 10 channels. To quantify how the performance changes with an increasing number of channels, the best curve fitting the data was found as shown in Figure 43 with a red line.

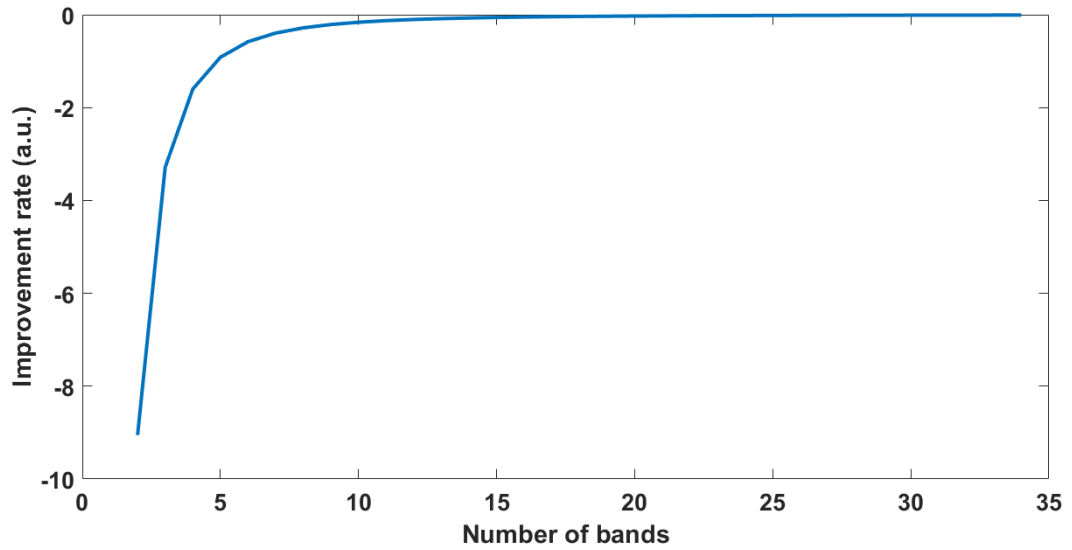


**Figure 43. The criterion value of channel subset selection with a curve fitted to the associated criterion**

The most appropriate curve for the points is a power curve which successfully fits the points with  $R\_square=0.9934$  and  $RMSE=0.2228$  with the following equation:

$$Criterion\_Function(N) = 34.14N^{-1.50} + 2.125 \quad \text{Eq. 8}$$

where  $N$  is the number of channels. Criterion function equation demonstrates that with an increasing number of the channels, the criterion decreases. To evaluate how the improvement rate increases with an increase in the number of channels, the derivative of the criterion function was obtained as shown in Figure 44.



**Figure 44. Performance rate vs increasing number of channels.**

Looking at how Figure 44 represents the first three channels, the rate increases very fast and then decreases. In fact, after selecting ten channels, the improvement is close to zero, and after six channels (Ch.16, Ch.36, Ch.37, Ch.3, Ch.5, Ch.11) the improvement rate is less than 0.5 in criterion function value. This graph proves that for more than ten channels (Ch.36, Ch.20, Ch.37, Ch.1, Ch.11, Ch.16, Ch.19, Ch.2, Ch.6, Ch.24 ) the performance is constant with no improvement or deterioration.

## 5.4 DISCUSSION

This study developed a clever methodology for discovering richly informative spectral channels with respect to differentiating normal and diseased (OSSN) tissue through a combination of swarm intelligence and cluster analysis. employing tens of channels leads to a long imaging time( each spectral image to acquire need 4 seconds), during which unintentional eye movement and blinking cause unavoidable interference such as image shifts and defocusing. A criterion function is defined to minimise the within-cluster variance while maximising the between-cluster variance and was optimised using three swarm intelligence methodologies, including PSO, DE and ACO, to find the richest subset of channels. The criterion function developed in this study has been defined to get minimally influenced by side-effects, which leads to improve generality power of the methodology. Subsets with various numbers of channels were evaluated to find the optimised number of possible channel.

This study proposes different subsets of channels, which can be used for various applications. Decision on the number of required channels depends on various factors including cost limitation, the criticality of the application, acquisition time, etc. As a case in point, if the methodology is supposed to be used for cancer boundary detection in an operating room, the application criticality would be high, so a higher number of channels is suggested.

If it is required to find only a specific number of channels, the number of selected channels can be restricted to a predefined number. For example, the number of channels may be restricted to 3 and the proposed methodology can then be performed. Sometimes, even some specific biomarker, such as NADH and PpIX, are defined in advance; however, the specific spectral channel must be defined to strongly catch those biomarkers in the tissue. In this case, a couple of adjacent channels can be designed about that biomarkers' spectra and, the best channel can be selected directly from the samples.

The indicative channels may also assist the clinical practice in learning disease-specific patterns. Correlating such channels with the diseases may give an in depth understanding into the underlying processes that produced the data. Such investigation

can extend the application of multispectral imaging to the discovery of unknown biomarkers and knowledge. The prospect of this study is providing one more step toward translating the multispectral imaging technology for real medical application in ophthalmology. So, optimising the number of channels results in more efficient instrumentation regarding equipment, acquisition time and computation complexity. This methodology was tried on OSSN detection, but it is not limited to that. Different diseases can be evaluated using this methodology, and also it is possible to define mores customised criterion function depending on the goal of the study.

## 5.5 CONCLUSION

The recently introduced auto fluorescence multi spectral imaging opens the opportunity of spectral imaging employing a large number of spectral channels with defined excitation and emission wavelength ranges, making it possible to extract spectral signatures of OSSN. However, in clinical ophthalmic applications such as diagnostics of OSSN or identification of cancer boundaries, such signature needs to be optimized in terms of spectral content and the number of channels. In particular, it is significant to reduce the imaging time without sacrificing classification accuracy. This study discovered a richly informative spectral signature able to to rapidly differentiate normal and diseased (OSSN) tissue through a combination of cluster analysis and swarm intelligence. In this study, discovering the spectral signature of OSSN was formulated as a two-objective optimization task with a fitness function reflecting the discriminative performance of the spectral signature and the reduction in the number of elements. Three swarm intelligence methods including PSO, DE, and ACO well established in the domain of machine learning were assessed, and DE was found to have the best capability to search and optimize the OSSN spectral signature. Depending on the criticality of the application, this study found spectral signatures based on 5 and 10 channels which could successfully detect OSSN while reducing the scanning time by 87% and 73%, respectively, compared to all 38 channels in our original work [158].

## 5.6 AUTHOR CONTRIBUTION

Authors:

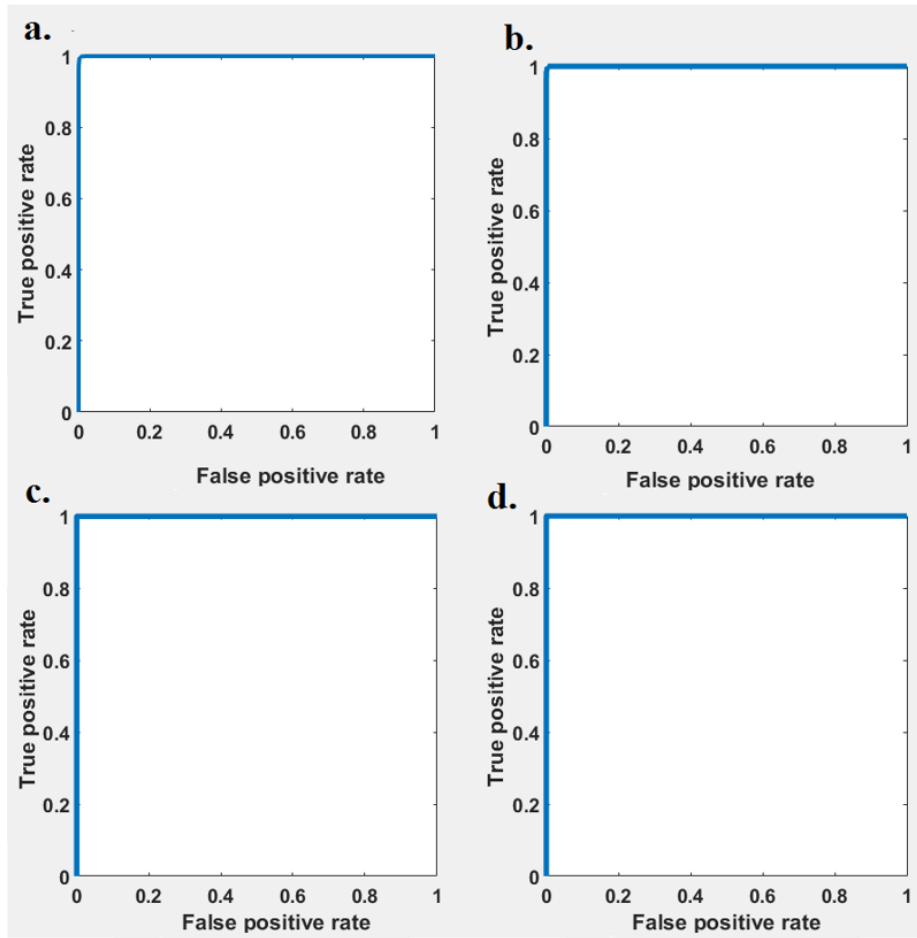
**Abbas Habibalahi**, Chandra Bala, Subhas Mukhopadhyay, Alexandra Allende, Ayad G. Anwer ,and Ewa M.Goldys

A.H, C.B, S.M and E. G. designed the study. A. H. acquired the data, processed images, performed image registration, biostatistics, computational and machine learning analysis and swarm intelligence. A.H, C.B, A. Allende, A. Anwer ,S.M and E.G interpreted and discussed the data. A. Allende performed histological analysis and interpretation. C.B, A. Allende, A. Anwar, A.H and E.G were responsible for technical, material and administrative support. C.B, S.M, and E.G supervised the study.

## 5.7 APPENDIX A

In this study, the criterion function was developed based on the cluster analysis. Such a criterion function was easy to be modified, and also its range was from 0 to infinity. Another criterion function which can also be defined is based on the performance of a classifier such as Support Vector Machine (SVM). SVM is a strong supervised method which forms a hyperplane with maximum margins in the high dimensional spectral feature space to separate data points into neoplastic and normal classes, whose theoretical background is well described in ref: [142]. To compare the effectiveness of the proposed methodology, we evaluate SVM to find the optimum channels, and its performance was compared against our methodology.

We evaluate AUC [180] associated with SVM classifier. Therefore, an SVM classifier was developed based on the channels suggested by DE and cluster analysis after 10, 30, 100, 10k iterations, and ROC was produced. As Figure 45 represents, AUC gets to saturation even based on the channel subset adopted after ten iterations, and no change happens to that for channels adopted after 30, 100, 10k. However, Figure 40 demonstrates that we have some change by increasing the number of iterations. Investigation of Figure 40 and Figure 45 proves the limitation of AUC as a criterion function for multispectral channel selection. In this case, such a criterion cannot be a proper criterion function for channel selection in multispectral imaging.



**Figure 45. ROC curve of SVM classifier for channel subset selection of DE: (a) after 10 iterations, (b) after 30 iterations, (c) after 100 iterations, (d) after 10k iterations.**



# Chapter 6: MELANOMA CELL CLASSIFICATION

---

This chapter describes the application of multispectral autofluorescence imaging to classify melanoma cells. This chapter has been written as a draft of a manuscript to be submitted in ‘Journal of Biophotonics’, and outlines the background and the purpose of the chapter in section 6.1. The methodology of the research, including cell lines used in this study, a brief description of the system and data analysis techniques, is described in section 6.2. Section 6.4 represent the results, including melanoma and normal cells (fibroblast) classification and monitoring spectral change due to treatment. Finally, section 6.5 discusses the results. After successful results on applying the multispectral imaging for detection of OSSN, this research were extended to evaluate skin cancer cells. This chapter demonstrate the application of the multispectral autofluorescence imaging to detect cancer cells (Melanoma) which is according to the target of thesis.

## **Melanoma cell classification using autofluorescence multispectral imaging**

### Abstract

This study presents a novel multispectral autofluorescence imaging microscopy method to differentiate melanoma and fibroblast. A customized instrumental approach based on autofluorescence imaging and advanced data analysis discriminate skin cell types based on a non-invasive, label-free technique. The customized technique involves a sequence of autofluorescence images with slightly different spectral specification. After preparing the images and diminish unwanted random and systematic noises, many features were extracted from the spectral images. Then, the dimension of the features was reduced, and linear discriminative analysis applied to classify cell types which could successfully reach to high accuracy with a single cell resolution. The same methodology was also applied to differentiate melanoma cells

before and after treatment. This approach may have potential to be used for more advanced applications such as melanoma early detection and treatment monitoring.

## 6.1 INTRODUCTION

‘Cutaneous malignant melanoma’ is the most aggressive skin cancer, responsible for 74% of skin cancer-related deaths, and its incidence is constantly growing worldwide [19, 23, 24, 181-183]. The opportunity of gaining information about melanoma cells in a label-free manner can be significant for precise early detection, refining current therapies and also discovering innovative treatment procedures [23-27]. Recently, autofluorescence spectrum analysis as a non-invasive tool has demonstrated a promising tool in monitoring biological substances in the absence of any perturbation prompted by exogenous fluorophores [184], as reviewed in several reports [20, 41, 185].

Common methods used for characterising the auto-fluorescence spectra of cells are Fluorescence spectroscopy[186, 187], Multiphoton microscopy and fluorescence lifetime imaging microscopy (FLIM) [55]. Fluorescence spectroscopy measures the average auto-fluorescence emission of cells, which is disconnected from single cell properties [27, 188, 189]. Cell populations are heterogeneous and analysing single-cell properties provides the opportunity to discover mechanisms which may not be identified when analysing a bulk population of cells[189]. Multiphoton and fluorescence lifetime imaging microscopy (FLIM) can perform autofluorescence imaging in a single cell resolution [11]. However, these technologies are mostly limited to two costly auto fluorescence channels. In fact, employing only a few channels may not provide sufficient information for accurate decision making [190, 191].

In this study, a new single-photon excited fluorescence imaging microscope, recently designed in our group [95], was used to discriminate melanoma cells. Unlike common autofluorescence microscopy, this system employs a number of channels covering a broad range of spectrum using economical instrumentation. The system splits the excitation light into tens of narrow channel wavelength series and collects emissions at a number of wavelengths. A combination of each excitation and emission wavelength results in a distinct spectral channel. In this work, 38 channels were used, which covers a number of native fluorophores, including NADH, PpIX, flavins. Cells

were imaged in each spectral channel, to acquire a separate fluorescence channel spectrum, different in each pixel. Employing tens of spectral channels enables this technology to monitor overall biochemical dynamic equilibrium and also to seek specific markers, which is an advance on the current autofluorescence imaging.

In this work, firstly, classification of melanoma cells and fibroblasts as malignant and normal skin cells, respectively, were considered. Multispectral autofluorescence imaging was evaluated for assessment of melanoma cells and fibroblasts to produce discriminative information. In contrast to typical autofluorescence imaging which considers a few single features, a variety of biologically relevant quantitative information extracted from the spectral images to capture different aspects of the spectrum with a single cell resolution. This quantitative information was in terms of different features including intensity, first order, second order, textural features and various statistical measures of pixel values [95, 192, 193]. A few most indicative features were selected and then discriminative analysis [180] was undertaken to classify melanoma from normal cells (fibroblast). This was then followed by examination of melanoma cells derived from a patient under treatment. Based on unsupervised data processing, we applied PCA decorrelation on the data[97] to monitor any possible spectral change in cells, before and after treatment. Then the data were quantitatively assessed based on hierarchy clustering[194]. To the best of our knowledge, the introduced technology is novel in skin cells and is a first step toward applying the technology to more advanced cases.

## **6.2 METHOD**

### **6.2.1 Patient recruitment and Cell isolation**

Our analysis of patients' samples was performed under the permission from Macquarie University Human Research Ethics Committee, reference No: 5201400458. Cells used in this study originated from 10 patients. All of the cells used in this study were cultured in a biologically similar manner. Cell lines were cultured in Dulbecco's Modified Eagle Medium supplemented with 10 heat inactivated fetal bovine serum (FBS; Sigma-Aldrich, St. Louis, MO, USA), 11.25 mM glutamine (Gibco, Thermo

Fisher Scientific, Waltham, MA, USA), and 10 mM HEPES (Gibco) and were maintained at 37°C with 5% CO<sub>2</sub>.

### **6.2.2 Fluorescence multispectral system and image acquisition**

We used a custom-made multispectral microscopy system to generate a sequence of fluorescence spectral images of the one sample area. Each spectral image has a specific excitation and emission wavelength. The excitation wavelength range is determined by a number of light emitting diodes (LED) covering a range of wavelengths from 340 to 510 nm. Four filter cubes with particular optical filters and a dichroic mirror are used to confine the light scattered from the samples and span the electromagnetic spectrum from 420 to 650 nm. These excitation-emission wavelengths form 38 distinct spectral channels packed in a hypercube.

## **6.3 DATA ANALYSIS**

In this study, powerful data analysis techniques were performed to evaluate the spectral information of the cells. After imaging, the spectral images were pre-processed to treat any possible artefacts and noises. Then, cells were segmented and various spectral information was extracted in terms of quantitative features. Several techniques were employed to analysis the features in this study. Analysis of variance (ANOVA) was employed to evaluate the robustness of the features. Also, Pearson correlation measure was used to calculate the correlation among the features. Principal component analysis (PCA) was performed to de-correlate the features and linear discrimination analysis (LDA) was performed to find the clusters. In addition, particle swarm optimization (PSO) was applied for feature selection due to its high speed to get convergence. These analyses are described in detail in the following sections.

### **6.3.1 Image pre-processing and enhancement**

Each channel spectrum is influenced by possible various errors, including Poisson noise, dead pixels and spikes. To treat spikes and dead pixels, a 'threshold limiting window' was slid over the spectral channels to find and replace them by values interpolated from the nearly adjacent pixels. In addition, the spectral images were smoothed by a wavelet filter with a symmetric mother function to treat Poisson noise.

In addition, uneven LED illumination of the field of view and a fluorescence background, coming from the microscope optics and the sample slide, also influence the spectral images. The background is removed by subtracting an image of the medium in a culture dish from each spectral image. Uneven illumination is treated via dividing spectral channels by images taken from a reference calibration fluid which has a fluorescence signal in all channels. The error removal procedure is performed on each channel separately.

### 6.3.2 Segmentation

After reducing the noise signal from the spectral images, cells need to be isolated [128] from the background to be ready for single cell analysis. In this study, cells were segmented from the DIC image manually to create a binary mask. Then, the mask was applied to all channels to isolate the cells in channels.

### 6.3.3 Spectral feature extraction

In contrast to typical autofluorescence image analysis, which mostly uses only a few features, in this study various types of features were used[9]. The features include cell colours, first order features, second-order features, textural features and various statistical measures of cell intensities.

Cell colours and their associated features extracted from autofluorescence spectral images are directly linked to the chemical composition of the cells [95]. They include the mean intensity ( $E_i$ ), second order moment ( $\sigma_i$ ), third order moment ( $S_i$ ) as follows:

$$E_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad \text{Eq. 9}$$

$$\sigma_i = \sqrt{\frac{1}{(N-1)} \sum_{j=1}^N (P_{ij} - E_i)^2} \quad \text{Eq. 10}$$

$$S_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^3} \quad \text{Eq. 11}$$

where  $P_{ij}$  is the pixel value of each pixel. In addition, Haralick's texture features [195] were extracted and used to differentiate the cell types. Textural features represent some biological characteristics such as fluorophore distribution within a cell to accomplish

broad information of the location of fluorescent intracellular components [9].

Haralick's texture features are based on the grey level co-occurrence matrix:

$$G = \begin{bmatrix} p(1,1) & P(1,2) & \dots & p(1,N_g) \\ p(2,1) & p(2,2) & \dots & p(2,N_g) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ p(N_g,1) & p(N_g,2) & \dots & p(N_g,N_g) \end{bmatrix} \quad \text{Eq. 12}$$

$G$  is a square matrix whose dimension is the grey levels in the image  $N_g$ .  $p(i, j)$  is the probability that a pixel with value  $i$  can be found adjacent to a pixel of value  $j$  [196]. After generating the co-occurrence matrix, Haralick's texture features are calculated for each cell of the spectral images as follows [9, 95, 193]:

Contrast:

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)(i - j)^2 \quad \text{Eq. 13}$$

Correlation:

$$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_i)(j - \mu_j)p(i, j)}{\sqrt{\nu_i \nu_j}} \quad \text{Eq. 14}$$

where  $\mu$  and  $\nu$  are the mean and variance, respectively.

Entropy:

$$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log p(i, j) \quad \text{Eq. 15}$$

Homogeneity

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{1 + (i - j)^2} \quad \text{Eq. 16}$$

Such features were extracted for each cell in all 38 channels and also statistical measures of cell intensities such as channel ratio were calculated, so each cell had

many (>2000) quantitative features. Consequently, spectral channels of each cell provide a feature vector with a label defining cell types. As a result of this stage, many features were obtained for subsequent cell classification into normal and cancerous groups. Extracting many features helps evaluate different aspects of the spectral properties of the cells. However, all of the features may not be indicative of cell groups or they may have some degree of correlations to each other.

#### 6.3.4 Analysis of variance

ANOVA is a statistical test frequently used for experimental data to prove that the difference between groups did not happen by chance, via rejecting a null hypothesis. Typically, it is followed by statistically significant results when a probability is less than a specific threshold (mostly  $p < 0.05$ ) [197, 198].

ANOVA partitions the data variance into two components:

- Variation between groups, which is  $y_j - Y$ , where  $y_j$  is the mean of the sample and 'Y' is the total variance of the data.
- Variation within a group, which is  $y_{ij} - y_j$  where  $y_{ij}$  is the observations.

In fact, ANOVA divides the total sum of squares (SST) between the group sum of squared (SSR) and the sum of squared errors (SSE) as follows:

$$\sum_i \sum_j (y_{ij} - Y)^2 = \sum_j n_j (y_j - Y)^2 + \sum_i \sum_j (y_{ij} - y_j)^2 \quad \text{Eq. 17}$$

where  $\sum_i \sum_j (y_{ij} - Y)^2$ ,  $\sum_j n_j (y_j - Y)^2$  and  $\sum_i \sum_j (y_{ij} - y_j)^2$  are SST, SSR and

SSE, respectively and  $n_j$  represents the sample size for different groups as  $j$  varies.

Then, the between groups variation is compared to the within-group variation. Hence, a ratio of between and within-group variation is calculated. If the ratio is significantly high, it can be concluded that the groups have different means and they are different groups. It can be measured by an  $F$  value as follows:

$$F = \frac{SSR / k - 1}{SSE / N - k} = \frac{MSR}{MSE} \quad \text{Eq. 18}$$

where  $MSR$  is the mean squared value,  $MSE$  is the mean squared error,  $k$  is the group number, and  $N$  is the number of all observations [199, 200].

### 6.3.5 Pearson correlation

To find valuable information from spectral channels, many features from different aspects are extracted, which may lead to some degree of correlation among the features. Hence, to develop a valuable model, it is necessary to find correlated features. A statistical measure to evaluate the correlation is the correlation coefficient, representing a putative linear relation between features. In this study, Pearson product moment correlation coefficients [201] were used, with the following formula:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

where ‘n’ is the sample size,  $x_i$  and  $\bar{x}$  are sample points from different observations and the associated mean value, respectively, and analogously for  $y_i$  and  $\bar{y}$ . The correlation value varies from -1 (showing a high negative correlation) through 0 (representing no correlation) to +1 (as the perfect positive correlation). Based on a Rule of Thumb [202], a correlation in the range of -0.3 to 0.3 is negligible.

### 6.3.6 Principal component analysis

Principal component analysis (PCA) is used in this study to transform to a feature space consisting of linear combinations of the original features with zero correlation between the new features. PCA is a transformation to reflect the feature vectors into a new linearly uncorrelated orthogonal space. The largest variance present in the data lays on the top principal components known as the most informative axes [9]. Transforming the original features into a new basis found by the PCA transformation provides another informative feature which is a linear combination of the original data.



### 6.3.7 Linear discrimination analysis

To generate a discrimination space and categorize the groups, we used linear discrimination analysis (LDA). LDA is a supervised learning which determines optimal basis axes to find different clusters[203, 204].

To generate the discrimination space, imagine there are  $C$  pattern classes,  $C_1, C_2, C_3, \dots, C_c$  in an  $N$ -dimensional space. The number of observations is  $l_i$  and  $x_{ij}$  is the  $j$ -th sample.  $\mu_i$  denotes the mean vector of the observations in the  $i^{th}$  class whose population mean value is assumed to be  $\mu_i = E(x_{ij} | C_i)$ . The expected mean value of the entire data set is  $\mu_0$ . Hence, the inter-class scatter is as follows:

$$S_b = \frac{1}{M} \sum_{i=1}^C l_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T \quad \text{Eq. 19}$$

and the intra-class scatter is :

$$S_w = \frac{1}{M} \sum_{i=1}^C \sum_{j=1}^{l_i} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T \quad \text{Eq. 20}$$

An arbitrary choice can be made to select groups of observations. The class mean  $\mu_i$  can be estimated from the class sample average and the similarity as follows:

$$m_i = \frac{1}{l_i} \sum_{j=1}^{l_i} x_{ij} \quad \text{Eq. 21}$$

To estimate the expected mean value of the whole data set, the average of the observations is used:

$$m_0 = \frac{1}{M} \sum_{i=1}^C \sum_{j=1}^{l_i} x_{ij} \quad \text{Eq. 22}$$

We need to find a transformation matrix producing new variables which are able to maximize the ratio of inter to intra-scatter ( $J_f(W)$ ).

$$J_f(W) = \frac{w^T S_b w}{w^T S_w w} \quad \text{Eq. 23}$$

So, a new coordinate system can be formed by the eigenvectors  $w_1, w_2, \dots, w_d$  concerning:

$$S_b w = \lambda S_w w \text{ Eq. 24}$$

Therefore, the projection of the input data to the new space created by the eigenvectors leads to approximate optimal class discrimination. Each axis of the new discrimination space indicates a canonical variable generated by an eigenvector [9, 95].

### 6.3.8 Particle swarm optimisation

After removing correlated and insignificant features, the number of remaining features can still be too high to develop a productive model [205]. Consequently, a limited number of indicative features with maximum relevancy and minimum redundancy needs to be nominated as a subset of features. Here, Particle swarm optimisation (PSO) was employed for feature selection to maximise cluster quality as the fitting function. Such methodology also considers the interaction between the features. PSO is a swarm intelligence inspired by birds' behaviour and follows the following principal [179].

A particle is an individual in a PSO whose population forms a swarm. The  $i$ th particle has two aspects at an iteration  $k$  as follows:

The particle position in the space of the data as

$$X_i^k = (x_1^k, \dots, x_n^k, \dots, x_N^k) \text{ Eq. 25}$$

where  $x_n^k \in [l_n, u_n], 1 \leq n \leq N$ ,  $l_n$  and  $u_n$  are maximum and minimum of the channels, respectively, for the  $n^{th}$  dimension.

The particle speed in the space of the data is as follows:

$$V_i^k = (v_1^k, \dots, v_n^k, \dots, v_N^k) \text{ Eq. 26}$$

In which the maximum limit of the speed is  $V_{\max}^k = (v_{\max,1}^k, \dots, v_{\max,n}^k, \dots, v_{\max,N}^k)$  and the minimum limit is  $V_{\min}^k = (v_{\min,1}^k, \dots, v_{\min,n}^k, \dots, v_{\min,N}^k)$ .

As the algorithm runs, the swarm needs to be updated by the following equations:

$$V_i^{k+1} = \omega V_i^k + c_1 r_1 (P_b^k - X_i^k) + c_2 r_2 (P_g^k - X_i^k) \quad \text{Eq. 27}$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \quad \text{Eq. 28}$$

where the best particle location in the previous step is  $P_b$  and the best overall position is  $P_g$  found between all of the particles in the population.  $P_b$  and  $P_g$  can be calculated by the following equations:

$$P_b = \begin{cases} p_b : f(x_i) \geq f(P_b) \\ X_i : f(x_i) \leq f(P_b) \end{cases} \quad \text{Eq.29}$$

$$P_g \in \{P_0, P_1, \dots, P_m\} | f(P_g) = \min(f(P_0), f(P_1), \dots, f(P_m)) \quad \text{Eq. 30}$$

where  $f$  is the objective function,  $m$  is the number of individuals,  $r_1, r_2$  are generated from uniform random variable between 0 and 1 and  $\omega$  is an inertia coefficient initialised in the range of 0 to 1.  $c_1$  and  $c_2$  are constants to set to optimise the algorithm.

## 6.4 RESULTS

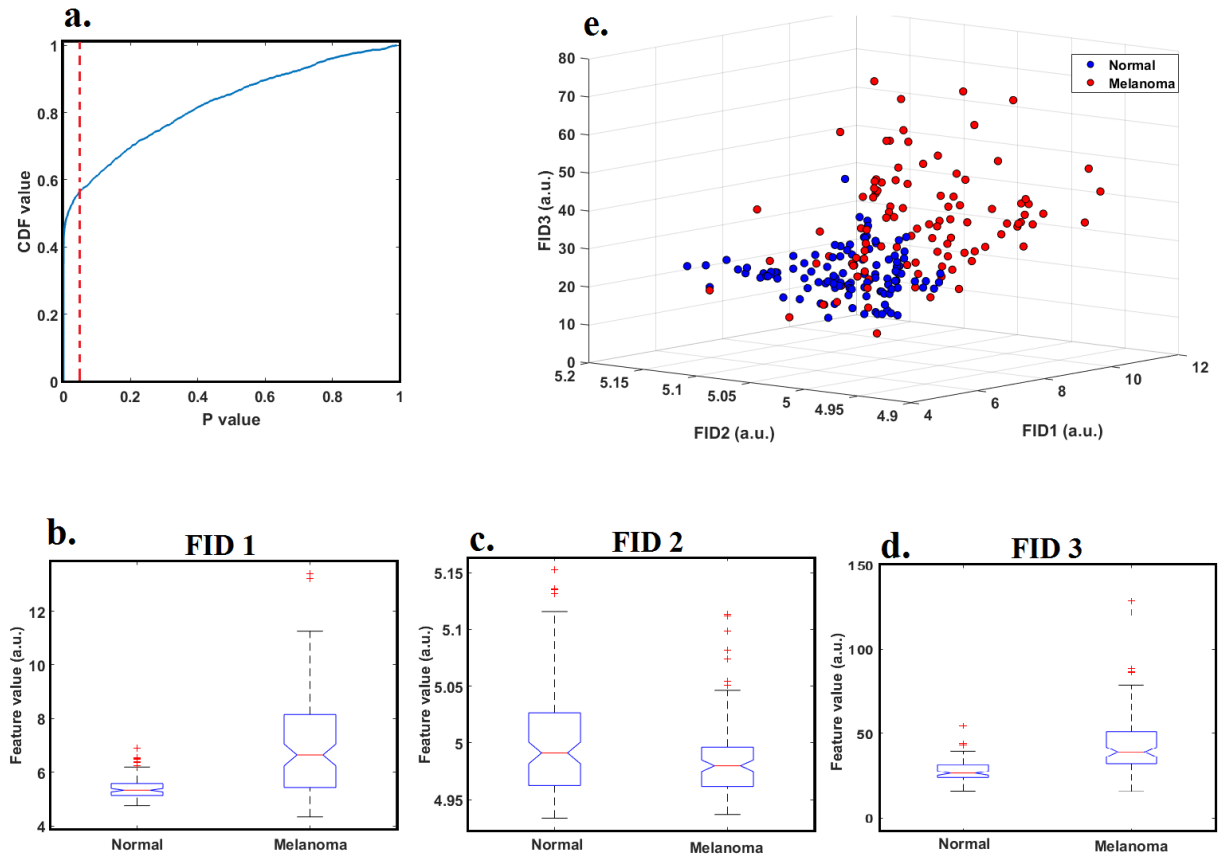
In this study, the newly designed multi-spectral autofluorescence imaging was applied on skin cells for two main purposes. First, the technology was employed to distinguish normal (fibroblast) and cancerous (melanoma) skin cell types. Second, the technology was used to detect possible spectral changes in melanoma due to treatment with a combination of BRAF and MEK.

### 6.4.1 Differentiating fibroblast and melanoma cells

After extracting spectral features on a single cell basis, each cell obtained a vector of features (dimension > 2000) with a normal or melanoma label. First of all, univariate analysis was applied to the data to discover any possible pattern among the spectral features, which informs the cell labels, and then advanced multivariate analysis were used to analyse the spectral data more efficiently.

As an initial demonstration of the features' robustness to distinguish the cell groups (melanoma and normal cells), the uni-variate ANOVA test was utilised on the features, one by one. Cells were grouped into normal and melanoma sets. Then, ANOVA was applied to obtain  $F$  factor and associated  $p$ -value. Figure 46 (a) plots the empirical cumulative distribution function (CDF) of  $p$ -values generated by ANOVA. The red dashed line in Figure 46 (a) shows the number of features whose  $p$ -values are less than 0.05. It demonstrates that more than 50% of the features can represent a significant difference between normal and melanoma cells. Although some features may show mutual correlation, this check shows that the technology used in this study could generate many indicative features to distinguish normal and melanoma cells.

Figure 46 (b,c,d) illustrates box plots of the three selected indicative features with the highest  $F$  factor ( $p$ -value  $< 0.005$ ) and negligible correlation ( $R < 0.1$ ). These features include FID1, which is the mean value of the cells in channel 13, FID2, which is the mean value of the cells in channel 17 divided by the mean value of those in channel 13, and FID 3, which is the mean value of channel 25 divided by the mean value of channel 9. Using these three features, a 3-dimensional space was created as shown in Figure 46 (e).

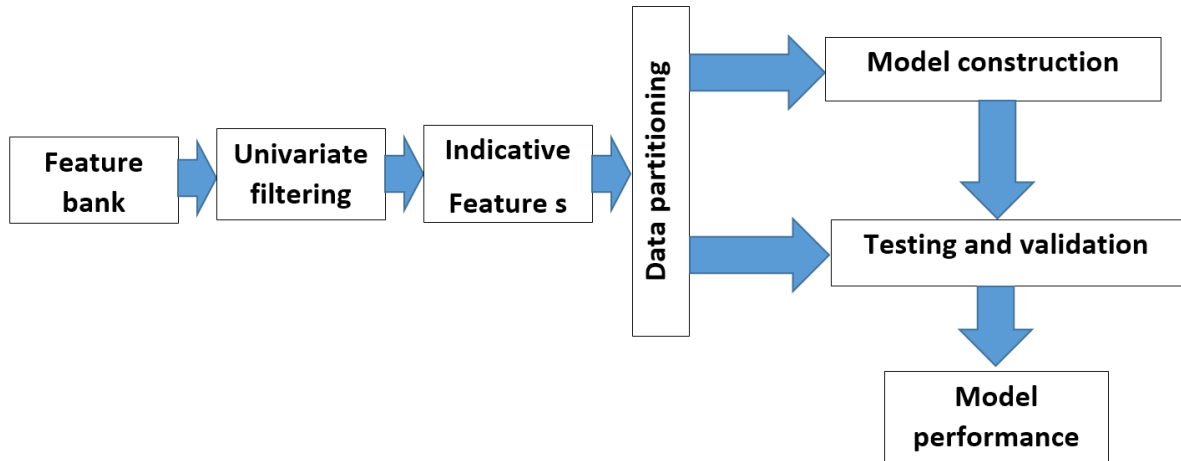


**Figure 46. Univariate data analysis to differentiate melanoma cells from normal cells (fibroblast):** (a) CDF value of features. (b) FID1, which is the mean value of the cells in channel 13 (c) FID2, which is the mean value of the cells in channel 17 divided by the mean value of those in channel 13. (d) FID3, which is the mean value of channel 25 divided by the mean value of channel 9. (e) 3D feature space created by FID1, FID2 and FID3 (Separability distance  $\sim 1.23$ )

Figure 46 (e) demonstrates that melanoma and normal data points form separate patterns although with some extent of confusion. Hence, to glean a more realistic picture of the data and to provide powerful data modelling able to extract a spectral signature more efficiently, multivariate analysis was applied in this study. In fact, as shown in Figure 46 (e), with univariate analyses, a reasonable separation between melanoma and normal cells can be obtained. However, such univariate analysis confines the spectral signature to use only 3 features, which may not be enough for an accurate decision making. So, a more advanced analysis may provide better separation by implementing spectral features more efficiently.

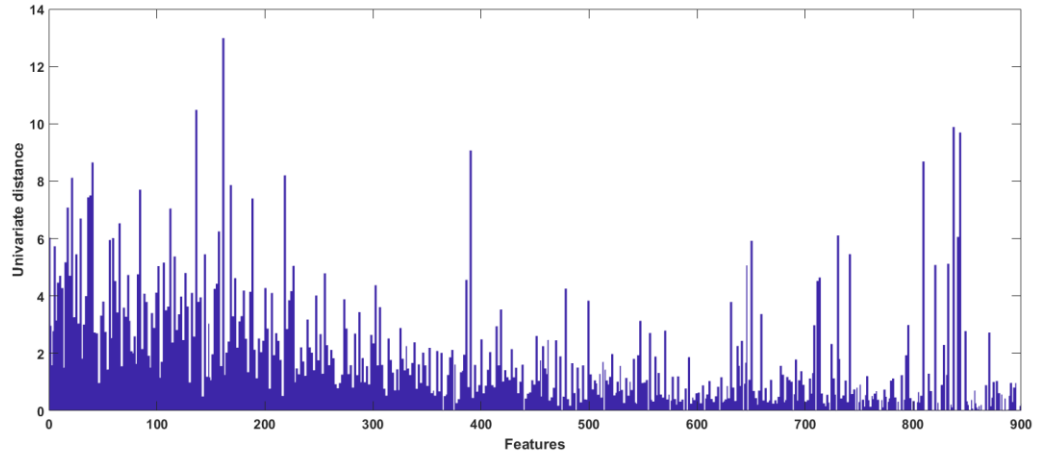
To advance the multivariate analysis and evaluate the generalisation competency of the methodology, a portion of the data was put aside for blind testing. Consequently, the data points involved partitioning into two separate subsets: (1) training data (60%),

which were used to develop a multivariate model, (2) testing data (40%), which were put aside to test the models. Testing data do not have any contribution to the model development. Developing a model has different stages in this study as shown in Figure 47.

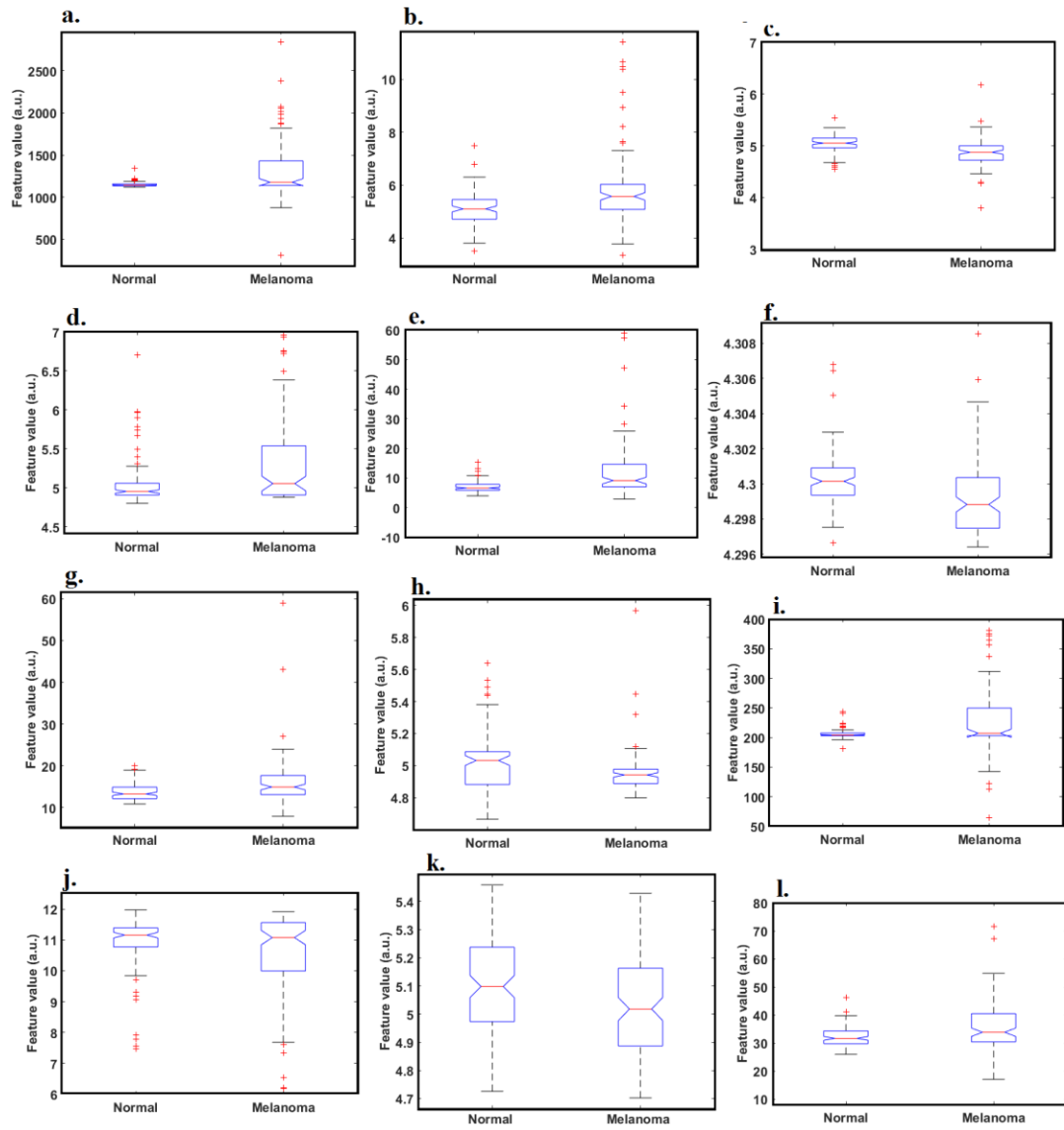


**Figure 47.**Data analysis framework

In supervised learning, the number of features (dimensionality) should be limited, as a high number of features may lead to overfitting the learning algorithm[206]. As a filter based feature selection using ANOVA, all of the features with a  $p$ -value of more than 0.05 were removed. Subsequently, about 900 features passed the filter and were considered significant. Figure 48 represents the univariate distance between normal and melanoma clusters made by each feature, which is a measure of their performance in distinguishing cell types. To select a collection of informative features, particle swarm optimisation (PSO) was employed. Finally, 12 indicative features were selected to distinguish between melanoma and normal cells as illustrated in Figure 49.



**Figure 48.** Univariate distance of normal and melanoma cell data points in univariate analysis



**Figure 49.** Features selected for multivariate analysis: (a) mean intensity of channel 7 divided by mean intensity of channel 12. (b) mean intensity of channel 7 multiplied by mean intensity of

channel 6. (c) mean intensity of channel 28 divided by mean intensity of channel 23. (d) contrast value of channel 16 (e) mean intensity of channel 14 multiplied by mean intensity of channel 22. (f) mean intensity of channel 17 divided by mean intensity of channel 9. (g) mean intensity of channel 19 multiplied by mean intensity of channel 2. (h) skewness value of channel 9. (i) mean intensity of channel 3 times by mean intensity of channel 8. (j) entropy of channel 3 (k) mean intensity of channel 2 divided by mean intensity of channel 13. (l) correlation of channel No. 23.

The selected features were fed to discrimination analysis which led us to generate an optimised space created by two canonical variables [11]. Canonical variables are linear combinations of the features which ensures excellent discrimination between melanoma and fibroblast. Consequently, the data points of the training were reflected into the discriminant space to visualise the structure and robustness of the newly selected feature space and the model as shown in Figure 50 (a). As can be seen from Figure 50 (a), the space generated by the selected features could clearly separate melanoma and normal cells, and reveals the presence of great and extremely significant differences between two cell types.

To evaluate the generality of the model developed by the training data sets, the testing data were assessed. In fact, the testing data did not have any contribution to find the indicative features or to develop the discriminative space and canonical variables. First, the training data were encircled in an ellipse which demonstrates the standard variation of the data across two data points' eigenvectors and accounts for more than 70% of the data population (see Figure 50 (a)). The ellipse shows the confines of the melanoma and normal clusters in the developed model. Then, to show how well the testing data points follow the separation power we achieved from the training data, the testing data points were reflected on the same discriminative space and ellipse, as shown in Figure 50 (b). It is evident that the testing data points were also clearly encircled via an ellipse. Comparing the results in Figure 46(e) and Figure 50 (a) demonstrates that multivariate analysis was able to extract spectral feature more efficiently than univariate analysis. Also, Figure 50 demonstrates that the spectral contents of channels are competent to discriminate biologically similar cultured cell groups of fibroblast (controls) and melanoma cells.



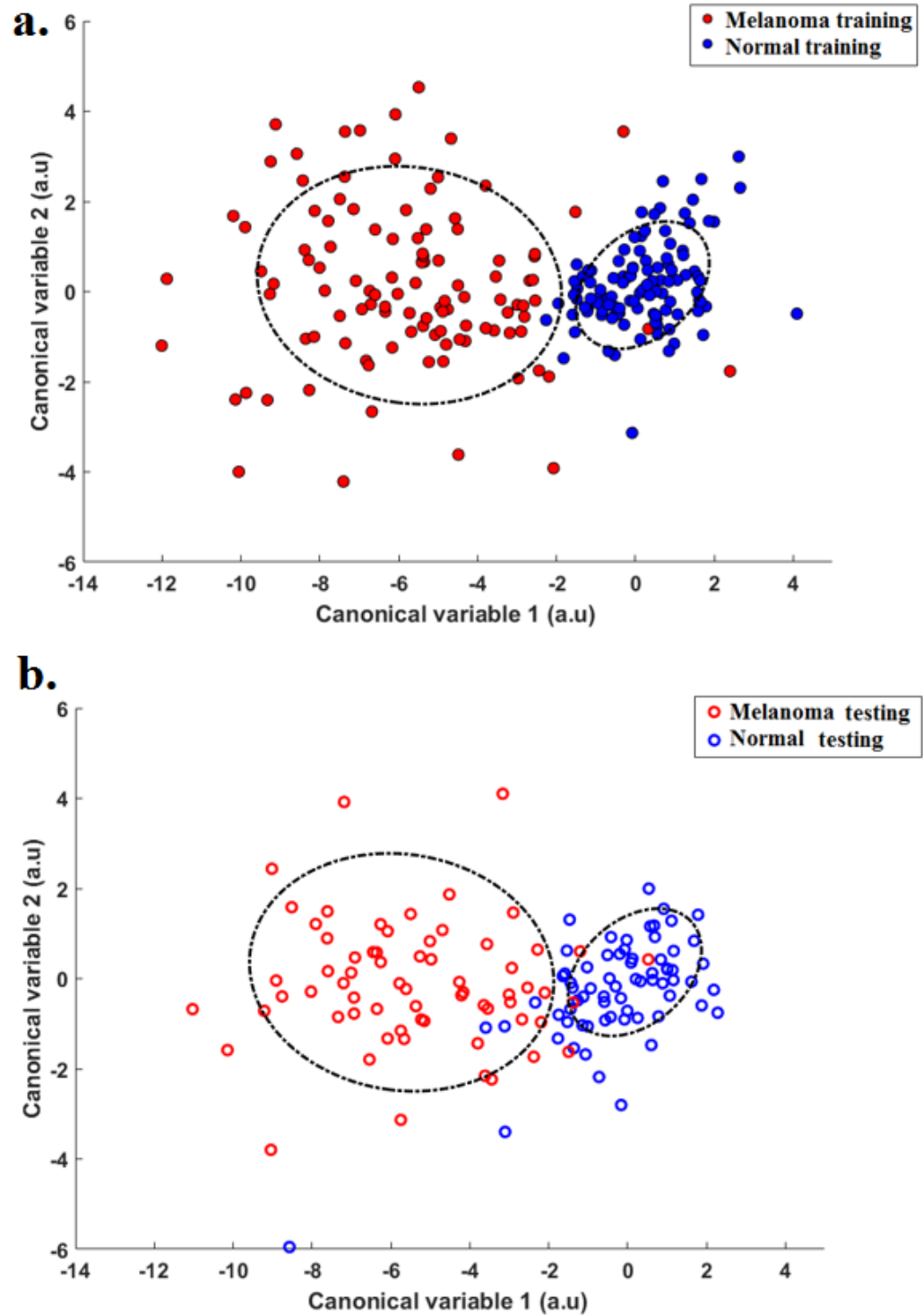
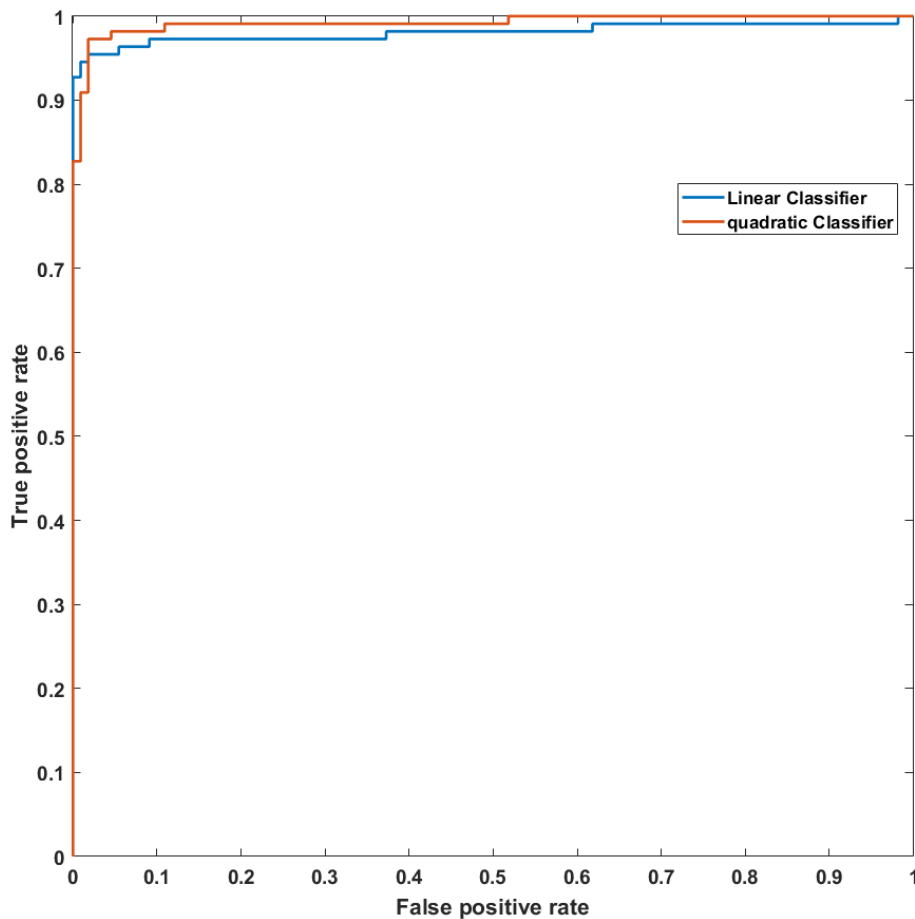


Figure 50. Clustering melanoma and normal (fibroblast) cells: (a) training set (Separability distance = 3.56) (b) testing set

The melanoma and normal clusters represented two separated clusters although there is a limited degree of the confusion, which is typical in cell clustering. The larger standard deviation of the melanoma cells may attribute to the variability of the cancer cells. For further analysis, linear and quadratic classifiers were designed, as they are effectively working with a limited volume of the training data [206] and also are resistant to the uncertainty of the data points which is unavoidable in biological data. Using K-fold cross-validation (K=10), ROCs were derived to evaluate the performance of linear and quadratic classifier systems as shown in Figure 51.



**Figure 51. ROC curve for linear and quadratic classifier**

Figure 51 demonstrates that the linear classifier has a 91.4% accuracy with AUC= 0.98, and the quadratic classifier shows 97.3% accuracy with AUC = 0.99.

Making a choice between these two classifiers, the linear classifier is the preferred candidate because it is less prone to overfitting the data.

#### 6.4.2 Monitoring spectral change due to treatment

As the second purpose of this study, we applied the technology to observe a possible spectral change in melanoma cells due to treatment. Tumours were obtained from a patient under treatment with a combination of BRAF and MEK inhibitors [207]. BRAF and MEK are typical genes in Melanoma. V600E, which is a BRAF mutation, is carried by almost half of all melanomas. This mutation is the BRAF kinase abnormality, stimulating cancer progression. Another typical gene mutation in Melanoma is V600K. To stop Melanoma growth, BRAF and MEK inhibitors can be used to hinder V600E and V600K mutations' activity, respectively [208-214].

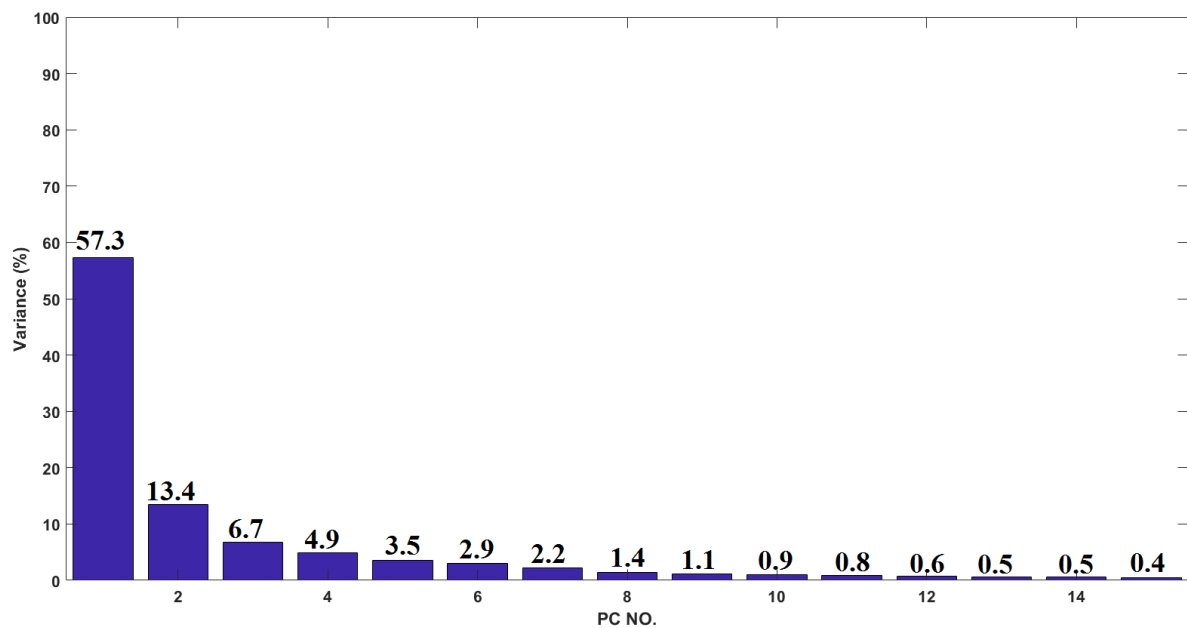
We applied the multispectral autofluorescence technology to observe if the cells excised before treatment are different to the cells excised after treatment in terms of the spectral information. Three cell lines were excised through a specific schedule from a melanoma patient who was under treatment. One cell line was obtained from a tumour excised before the treatment, and two cell lines were obtained from tumours during treatment as shown in Table 6. To analyse the data with no prior information, unsupervised analysis was performed in this section.

To track the spectral variation in the cells due to treatment, we used PCA analysis followed by hierarchical classification with a dendrogram. The multispectral autofluorescence colours of the cells were measured, then PCA was applied to the data to de-correlate the spectral information. After evaluating the principal components generated by PCA, a dendrogram analysis was employed to evaluate the multi-spectral signature of cell autofluorescence that reflects the level of correlation between similar cells.

**Table 6. Schedule for tumour cell extraction from the patient**

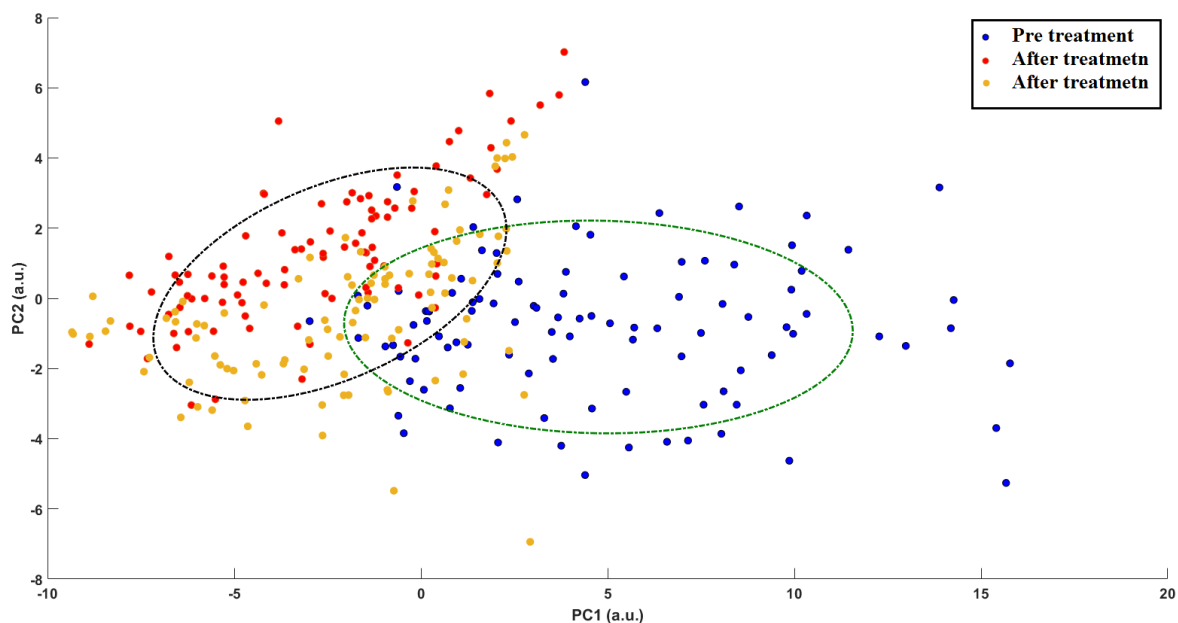
Label of cells	Time Schedule
<b>Pre</b>	Before the patient started therapy
<b>Prog 1</b>	~ four months after the patient started therapy

To de-correlate the spectral features, PCA analysis was applied and principal components were obtained. Principal components are orthogonal to each other, so there is no redundant information between them. Each principal component captures a proportion of the variance present in the spectral information. Figure 52 demonstrates the proportion of the variance associated with top 15 principal components which captured more than 97% variance of all spectral information.



**Figure 52. Proportion of variance captured by each principal component**

According to Figure 52, PC1 and PC2 hold the main variance proportion of the spectral information, which accounts for more than 70% of the total system variance. To visualise the robustness of the top two principal components in discrimination spectral variation associated with prior and during the treatment, the mean intensities of spectral cells were projected onto a two-dimensional space spanned by PC1 and PC2.



**Figure 53.A discriminative space created by the two top principal components (separability distance ~6.5)**

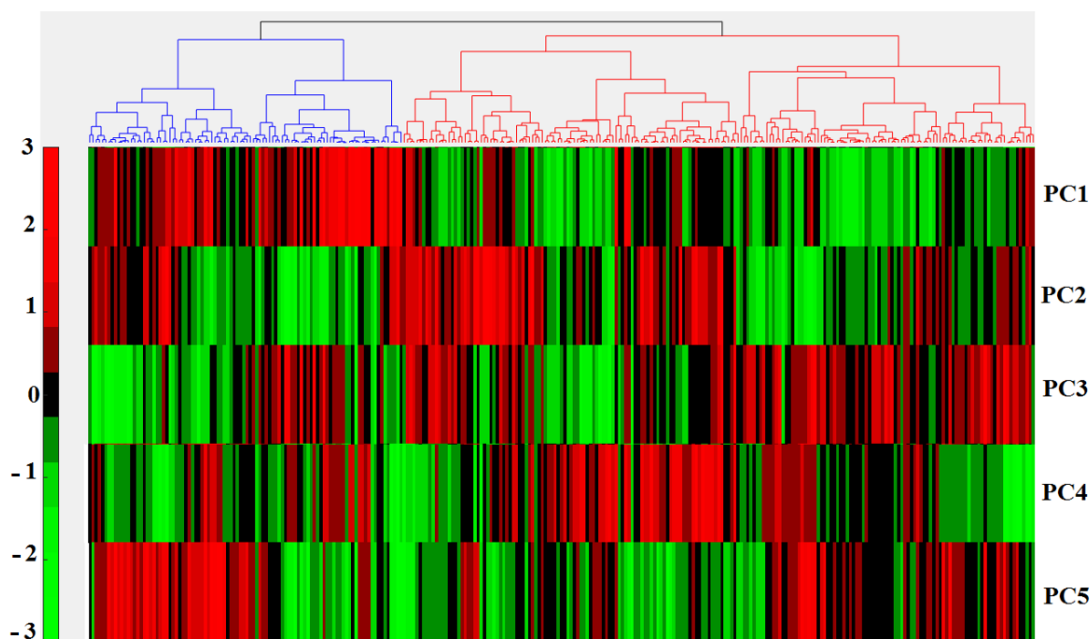
Figure 53 demonstrates that the cell groups after treatment move away from the cells prior to the treatment noticeably. In addition, two cell lines which are related to the after treatment have a correlation.

A theory which may explain such spectral change can be as follows: Physiochemical activities such as glucose uptake, and lactate creation increase in human cancer cells with oncogenes [210-212]. Consequently, their inhibition could lead to altering the amount of the physiochemical substances formed or uptaken by cancer cells [213]. It has been demonstrated that MEK inhibition in melanoma cells hosting mutant BRAF causes a significant decrease (to 36% of control) in glucose uptake [214]. Therefore, melanoma therapy with inhibitors may result in changing the chemical composition of the cells.

To capture spectral information in a higher dimension and include further indicative principal components to reveal the structure of the data quantitatively, an unsupervised agglomerative hierarchical classification with a dendrogram was used [194] as shown in Figure 54. The top 5 principal components were used as the input features of the dendrogram. These components could capture more than 85% of the total system variance as shown in Figure 52. Such an unsupervised classification is a bottom-up approach in which each observation is considered as a cluster initially.

Then, the pairs with a high value of similarity merge to form a new cluster and those with high dissimilarity move away from each other.

Consequently, a hierarchy will form, while observations merge progressively. This approach results in finding the groups and subgroups in the data without forcing the data to form a specific cluster. In fact, this methodology reveals the hidden structure within the data. In this study, the top 5 PCAs, which captured the most variance in the data (>85%), were used to form a hierarchical classification. Every single vertical bar represents a cell with 5 elements which are their PCAs. The values are normalised to -3 to 3. Finally, two major clusters can be recognised from the dendrogram as they are coloured in Figure 54 in blue and red. Considering these classifications revealed that the blue cluster was mostly occupied by pre-treatment cells and the red cluster was occupied by after treatment cells.



**Figure 54. Unsupervised agglomerative hierarchical classification for monitoring the treatment**

Figure 54 demonstrates that the dendrogram could find a spectral correlation of the cells prior to treatment and after treatment. Considering the two classes, which are shown with blue dendrogram and red dendrogram, the blue dendrogram is mostly occupied by cells before treatment (76%) and the red dendrogram mostly occupied by cells excised after treatment cells (84%).

## 6.5 DISCUSSION

Multispectral autofluorescence microscopy as demonstrated in this study provided rich spectral images which can be a source of valuable knowledge about the melanoma cells. Here, the spectral information in spectral images was obtained from inexpensive equipment which can be easily made-up on a standard fluorescence microscope. We improved the system by employing more spectral channels, unlike the two-channel autofluorescence microscope, and define tens of spectral channels ( $N=38$ ) to capture a number of fluorophores, especially cancer biomarkers. Employing 38 autofluorescence spectral channels enabled us to obtain a more powerful spectral signature to distinguish melanoma cells across a wide spectrum, which is an advancement to the current autofluorescence imaging systems employing only a few spectral channels. Advanced data processing methodology was employed to extract many informative, which led to finding distinct clusters of melanoma and normal cells. This study initially demonstrated the new technology's capability to extract significant features to distinguish Fibroblast and melanoma cells in a same-culture condition and with no external additive dyes. Simple univariate analysis showed a promising difference between the features, and then applying multi-variate statistics could extract an indicative signature which separated melanoma and normal clusters effectively. In fact, multispectral autofluorescence examination could show differences between melanoma and fibroblast and could successfully deal with a high cellular heterogeneity which normally is present in cell populations.

This study also shows the ability of the proposed technology to identify a potential change in cell chemical composition due to treatment, in a non-invasive manner with no additive dye. The correlation of the spectral values was removed using PCA analysis and the most indicative components extracted. The PCA space indicated that the cells after treatment moved away significantly from the cells before the treatment. In addition, hierarchical clustering without any prior knowledge and training could successfully detect the same spectral signature available in the cell from the same groups.

Understanding the spectral changes due to the treatment and learning the consequences of therapy with BRAF and MEK inhibitors on the chemical composition of the cells may allow the finding of further combinatorial treatment choices. It also may offer to refine the use of current medications and accomplish a better cancer

control procedure in the clinic. This non-invasive methodology may also recognise early pharmacodynamics responses more accurately. Predicting the therapy response early is crucial for cancer treatment, as tumour shrinkage may always appear in the early stages [212] which may result in a false negative indication for the treatment, and the physicians may terminate the therapy, and premature cessation of treatment may happen. Therefore, developing a technology which is capable to identify treatment response in a non-invasive manner to judge treatment usefulness may improve the survival rate of cancer patients.

Various label-free technologies may be available with a single cell resolution to discriminate different cell types such as Multi-spectral reflectance microscopy or FIIM. However, the novelty of the proposed technology is that we could look at the cells more accurately by defining specific excitation and emission spectral channels. This ability allows us to exclude or include some wavelengths due to specific circumstances. Besides, we have our excitation-emission wavelengths under control and can use as many channels as we need in a broad range of wavelengths, from very deep UV to IR regions, which cannot be achieved in reflectance methods. For example, we can avoid some UV range, which may be harmful to living organs, and only pick safe wavelengths or we may excite those LEDs with a controlled amount of exposure time and power intensity. In addition, our technology is very cost effective compared to advanced FLIM systems. All the cells have been images following the same fashion, so that the effect of the cell culture conditions, Beer's law and environment conditions are the same for all the cell groups.

The newly designed system described here has a vast potential to explore in the field of cancer diagnosis and treatment. This methodology may help to learn more about various ranges of abnormalities which may appear on the skin, including melanoma, carcinoma, etc. Usually, melanoma is diagnosed based on a clinical exam using morphological criteria qualitatively. Such an impression is challenging, as there is a wide range of different pigmented lesions on skin whose properties may overlap melanoma features[215]. Hence, developing a sensitive screening technology able to differentiate different types of skin cells is vital.

This technology may also enable finding particular cell subpopulations in order to support of drug screening processes. Consequently, an improved understanding of disease may lead to more effective diagnostics and therapeutics. This technology is a



non-invasive screening and selection of cells which is of vital significance in prospective clinical treatments. Using such technology may help to design a specific therapy for a particular patient. Our method may provide an appropriate and precise alternative to current invasive biopsy for monitoring the treatment response.

## 6.6 AUTHOR CONTRIBUTION

Authors:

**Abbas Habibalahi**, Helen Rizos , Subhas Mukhopadhyay, Bernadette Pedersen, Ayad G. Anwer ,and Ewa M.Goldys

A.H, H.R and E. G. designed the study. A. H. acquired the data, processed images, biostatistics, computational and machine learning analysis and swarm intelligence. A.H, H.R, B.P, A. A, S.M and E.G interpreted and discussed the data. H.R, B.P, A. A, A.H and E.G were responsible for technical, material and administrative support. H.R, S.M, and E.G supervised the study.



## Chapter 7: Conclusions

---

This thesis could successfully reach to its assigned targets. This study designed an illumination and filter arrangement to make the technology specialised for cancer detection and evaluation. In this study, 38 channels were arranged, which capture various fluorophores, especially the cancer biomarkers. This research study could effectively provide significant contributions in the field of cancer detection including detection and delineation of ocular surface squamous neoplasia (OSSN) and discrimination of melanoma cells. As a conclusion, this chapter covers a brief summary of the theory and methodology used in this study (7.1) and a brief summary of the results (section 7.2). The practical implication of the study is described in section 7.3 and finally the potential extension of the study is discussed in section 7.4.

### **7.1 SUMMARY OF THE METHODOLOGY ADOPTED BY THIS STUDY**

When cells or tissue are irradiated at a particular wavelength, some molecules, called fluorophores, would get excited and emit lights. Such a natural emission is called autofluorescence. Several fluorophores naturally available within cells and tissue have auto-fluorescence properties, including Porphyrins (PPIX), Nicotinamide adenine dinucleotide (NADH), and Flavins. Cancer transformation can lead to an alteration in the concentration of these native fluorophores, which results that emission and absorption properties of the tissues change. Due to these variations, cancerous cells or tissues can represent autofluorescent spectra which are different to associated normal cells or tissue. Therefore, observing the spectral pattern of normal and cancerous tissues or cells can be a tool to distinguish and characterise cancerous or abnormal conditions.

The first focus of this study was to establish a proper autofluorescence imaging setup for cancer characterisation. After conducting a literature review and investigation of previous research, this study arranged 38 distinctive spectral channels to monitor various fluorophores, especially cancer biomarkers, as presented in Chapter 3:. The availability of tens channels provided a rich source of information for univariate, multivariate analysis and artificial intelligence, and made it possible to

survey the overall biochemical composition of the tissue, in addition to detecting specific markers to identify the tissue or cell condition.

Due to various random and systematic noises caused by the camera sensor, illumination and microscope, the raw spectral images could not directly be fed to the data analyses strategies. Therefore, a comprehensive pre-processing strategy was developed in this study to prepare the spectral images for further analysis. Such preparation minimised various artefacts in spectral images including saturated or dead pixels, Poisson noise, and background and illumination curvatures. Subsequently, the system and a pre-processing algorithm were ready to be applied for cancer characterisation on different cell lines and tissues, as presented in chapter 4, 5 and 6.

## **7.2 SUMMARY OF THE EXPERIMENTAL RESULTS**

First, characterisation of eye surface neoplastic tissue based on the multispectral autofluorescence imaging was performed. 18 patients were employed and human OSSN samples were obtained. The target was to distinguish between normal and neoplastic tissue. We established two different classification frameworks: intra- and inter-patient classification. Therefore, we considered aspects of patients' variability and quantify the spectral signature of OSSN.

This study also used various data analysis techniques and machine learning methods to design a methodology for OSSN of boundary delineation. For an intraoperative assessment, machine learning could rapidly predict a false colour map to define the boundary of OSSN. The neoplastic boundaries defined by machine learning methods were validated and assessed by an anatomical pathologist.

The approach proposed to detect OSSN in this study provides several benefits: Potentially it can reduce the requirement of eye biopsies, prevent therapy delays and also make treatment more effective. In addition, it may assist a physician to monitor OSSN treatment. An accurate intra-patient image could assist the surgeon to precisely locate the neoplastic boundaries and to completely cut out the residual neoplastic tissue. The technology could be used to monitor treatment efficacy and detect recurrences non-invasively. Quantitative analysis of the spectral images proved a strong multispectral signature of a relative difference between neoplastic and normal tissue both within each patient (at  $p < 0.0005$ ) and between patients (at  $p < 0.001$ ). the

automated diagnostic method based on machine learning produces maps of the relatively well circumscribed neoplastic-non neoplastic interface. Such maps can be rapidly generated in quasi-real time and used for intraoperative assessment. Generally, OSSN could be detected using multispectral analysis in all patients investigated here.

Secondly, the multispectral autofluorescence was optimised in terms of the number of channels to facilitate industrial translation. An advanced methodology using a combination of swarm intelligence and cluster analysis was developed to discover a subset of rich and informative spectral channels for differentiating normal and diseased (OSSN) tissue. First, discrimination analysis was employed to find normal and diseased clusters, and then a criterion function was defined to minimise the within-cluster variance while maximising the between-cluster variance. Such a criterion function was optimised using three different swarm intelligence methodologies including particle swarm intelligence (PSO), differential evolution (DE) and ant colony optimisation (ACO). Finally, depending on the required accuracy and circumstances of the application, the richest subsets with a few channels were proposed. Optimising the number of channels resulted in more efficient instrumentation in terms of equipment, acquisition time and computation complexity. Depending on the criticality of the application, this study found spectral signatures based on 5 and 10 channels which could successfully detect OSSN while reducing the scanning time by 87% and 73%, respectively, compared to all 38 channels in our original work [158].

Finally, melanoma and fibroblast were assessed in this study. Unlike typical auto-fluorescence imaging techniques that consider only a few features, a variety of biologically relevant quantitative information was extracted from spectral images. Such a powerful analysis helped evaluate different aspects of the spectrum in a single cell resolution. Different features including intensity, first order and second order features, textural features, and different statistical measures of pixel values were quantitatively analysed.

After selection of the most indicative features, a discriminative analysis was undertaken to distinguish melanoma cells from fibroblast efficiently ( $AUC > 0.9$ ). The results showed that melanoma and normal cells formed two separate clusters. The

generality power of the analysis was successfully validated by testing data which did not have contribution in developing the model.

This was then followed by an examination of melanoma cells derived from a patient under treatment. Using unsupervised data processing, spectral features from the channels were de-correlated based on the principal component analysis (PCA). The analysis showed that the data points associated with a tumour extracted after treatment moves away from the data points related to a tumour extracted before the treatment. Then the data were quantitatively assessed using hierarchy clustering. Therefore, this study also successfully demonstrates the possibility of obtaining information about melanoma cells and their environment to monitor their behaviour and discriminate them from normal skin cell types (accuracy>80%). Such a methodology may open a new way for cell therapy, regenerative medicine, personalised immunotherapy and cancer treatment. In general, the accuracy of the system employed in this thesis is greater than 80% which is quite comparable with the similar systems described in section 2.10 but with some distinctive features such as single cell resolution, simple design and flexibility in spectral specifications.

### **7.3 LIMITATIONS**

Autofluorescence cannot be a very strong signal, so the experiments need to be performed in a dark room to avoid ambient light interference.

This study involved the development of both hardware and software, each of which faced several limitations. As the system was customised, it was hard and time-consuming to find some parts to keep the system functional. Such a limitation could put a big delay in the research time to time. Moreover, this study was performed using only a 40X objective magnification. Therefore, to capture images from a large tissue, it was needed to take many images to make a final composite image. Consequently, acquisition and stitching the images were challenging.

As this study involved with human samples, ethics application was required which could put a delay in the research. OSSN is a relatively rare disease and almost 10% of all eye lesions were found to be OSSN. Therefore, a long period of sample collection was one big limitation of this study. Although we applied the standard “leave one [patient] out” (LOO) approach to deal with this limitation, the greater number of patients may help to improve the generality power of the proposed

technology, which can be considered for a possible extension of this study. Moreover, this study was multidisciplinary research which needed to collaborate with clinicians such as surgeons, cancer scientists and pathologists. Organising a meeting with such busy clinicians was another limitation of this study.

One further limitation which is related to the nature of the autofluorescence imaging is that autofluorescence signal is relatively weak, so it is needed to employ very sensitive camera, especially when the spectral signature is not strong.

## **7.4 FUTURE STUDY**

In combination with high content data processing, autofluorescence multispectral imaging has many potential areas in cancer science which are still novel and can be explored, some of which are as follow:

- As mentioned in 4.5, benign entities such as pterygia were not evaluated in this study. As a future study, pterygia could be investigated based on the technique described here, as it is likely that pterygia will have a different chemical composition when compared to normal tissue. In fact, this study could be extended to classify tissue in three different groups, comparing normal tissue, pterygia and neoplastic tissue with greater sample size.
- A possible extension of this study could focus on discrimination between eye lesions of different grades including aggressive lesions, requiring action and resection, and benign cases. OSSN is a broad term which encompasses noninvasive conjunctival intraepithelial neoplasia (CIN), dysplasia and squamous cell carcinoma in situ (SCCIS), and invasive squamous cell carcinoma (SCC) where tumour cells invade the stroma by breaching the basement membrane. Such discrimination can help clinicians to design the treatment.
- Autofluorescence imaging also can be applied to assess the patient diseases status and the effectiveness of the therapy. The multi-spectral analysis may help to understand better and broad range of diseases affecting the metabolism including cancer, neural degeneration.

- A possible extension of this study is to apply the technology to distinguish different types of skin cancer such as basal cell carcinoma, squamous cell carcinoma and melanoma.







# REFERENCES

---

1. Wu, C., et al., *In-vivo optical imaging in head and neck oncology: basic principles, clinical applications and future directions*. International journal of oral science, 2018. **10**(2): p. 10 % @ 2049-3169.
2. Quinn, K.P., et al., *Quantitative metabolic imaging using endogenous fluorescence to detect stem cell differentiation*. Scientific reports, 2013. **3**: p. 3432 % @ 2045-2322.
3. Skala, M.C., et al., *In vivo multiphoton microscopy of NADH and FAD redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia*. Proceedings of the National Academy of Sciences, 2007. **104**(49): p. 19494-19499 % @ 0027-8424.
4. Zimmermann, T., J. Rietdorf, and R. Pepperkok, *Spectral imaging and its applications in live cell microscopy*. FEBS letters, 2003. **546**(1): p. 87-92 % @ 1873-3468.
5. Goldys, E.M., *Fluorescence applications in biotechnology and life sciences*. 2009: John Wiley & Sons.
6. Luo, X., et al., *Accuracy of autofluorescence in diagnosing oral squamous cell carcinoma and oral potentially malignant disorders: a comparative study with aero-digestive lesions*. Scientific reports, 2016. **6**: p. 29943 % @ 2045-2322.
7. Herrero, L.A., B.L.A.M. Weusten, and J.J. Bergman, *Autofluorescence and narrow band imaging in Barrett's esophagus*. Gastroenterology Clinics of North America, 2010. **39**(4): p. 747-758 % @ 0889-8553.
8. Ghukasyan, V.V. and A.A. Heikal, *Natural Biomarkers for Cellular Metabolism: Biology, Techniques, and Applications*. 2014: CRC Press.
9. Gosnel, M., *Unlocking the potential of spectral imaging for the characterization of cells and stem cells population*. Thesis, Macquarie University, 2013.
10. Bottiroli, G. and A.C. Croce, *Autofluorescence spectroscopy of cells and tissues as a tool for biomedical diagnosis*. Photochemical & photobiological sciences: Official journal of the European Photochemistry Association and the European Society for Photobiology, 2004. **3**(11-12): p. 189-210 % @ 1474-905X.
11. Gosnell, M.E., et al., *Functional hyperspectral imaging captures subtle details of cell metabolism in olfactory neurosphere cells, disease-specific models of neurodegenerative disorders*. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2016. **1863**(1): p. 56-63 % @ 0167-4889.
12. Pe'er, J., *Ocular surface squamous neoplasia*. Ophthalmology clinics of North America, 2005. **18**(1): p. 1-13, vii % @ 0896-1549.
13. Stone, D.U., A.L. Butt, and J. Chodosh, *Ocular surface squamous neoplasia: a standard of care survey*. Cornea, 2005. **24**(3): p. 297-300 % @ 0277-3740.
14. Mayer, J., *Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma*. Medical journal of Australia, 1997. **167**(4): p. 206-210 % @ 0025-729X.
15. Reintgen, D.S., et al., *Malignant melanoma in black American and white American populations: a comparative review*. Jama, 1982. **248**(15): p. 1856-1859 % @ 0098-7484.

16. Kalialis, L.V., K.T. Drzewiecki, and H. Klyver, *Spontaneous regression of metastases from melanoma: review of the literature*. Melanoma research, 2009. **19**(5): p. 275-282 % @ 0960-8931.
17. Phallen, J., et al., *Direct detection of early-stage cancers using circulating tumor DNA*. Science translational medicine, 2017. **9**(403): p. eaan2415 % @ 1946-6234.
18. Etzioni, R., et al., *Early detection: The case for early detection*. Nature Reviews Cancer, 2003. **3**(4): p. 243 % @ 1474-1768.
19. Orzan, O.A., A. Şandru, and C.R. Jecan, *Controversies in the diagnosis and treatment of early cutaneous melanoma*. Journal of medicine and life, 2015. **8**(2): p. 132.
20. Wagnieres, G.A., W.M. Star, and B.C. Wilson, *In vivo fluorescence spectroscopy and imaging for oncological applications*. Photochemistry and photobiology, 1998. **68**(5): p. 603-632 % @ 1751-1097.
21. Calonge, M., et al., *Impression cytology of the ocular surface: a review*. Experimental eye research, 2004. **78**(3): p. 457-472 % @ 0014-4835.
22. Lu, G., et al., *Detection of Head and Neck Cancer in Surgical Specimens Using Quantitative Hyperspectral Imaging*. Clinical Cancer Research, 2017. **23**(18): p. 5426-5436 % @ 1078-0432.
23. Leiter, U. and C. Garbe, *Epidemiology of melanoma and nonmelanoma skin cancer—the role of sunlight*, in *Sunlight, vitamin D and skin cancer*. 2008, Springer. p. 89-103.
24. Boni, A., et al., *Selective BRAFV600E inhibition enhances T-cell recognition of melanoma without affecting lymphocyte function*. Cancer research, 2010. **70**(13): p. 5213-5219 % @ 0008-5472.
25. Ko, J.S., et al., *Diagnostic Distinction of Malignant Melanoma and Benign Nevi by a Gene Expression Signature and Correlation to Clinical Outcomes*. Cancer Epidemiology and Prevention Biomarkers, 2017. **26**(7): p. 1107-1113 % @ 1055-9965.
26. Minca, E.C., et al., *Comparison between melanoma gene expression score and fluorescence in situ hybridization for the classification of melanocytic lesions*. Modern Pathology, 2016. **29**(8): p. 832 % @ 1530-0285.
27. Wang, D. and S. Bodovitz, *Single cell analysis: the new frontier in 'omics'*. Trends in biotechnology, 2010. **28**(6): p. 281-290 % @ 0167-7799.
28. García-Plazaola, J.I., et al., *Autofluorescence: Biological functions and technical applications*. Plant Science, 2015. **236**: p. 136-145 % @ 0168-9452.
29. Masters, B.R., *Principles of fluorescence spectroscopy*. Journal of Biomedical Optics, 2008. **13**(2): p. 029901 % @ 1083-3668.
30. Lichtman, J.W. and J.-A. Conchello, *Fluorescence microscopy*. Nature methods, 2005. **2**(12): p. 910 % @ 1548-7105.
31. Sommer, M.E., et al., *Structure-based biophysical analysis of the interaction of rhodopsin with G protein and arrestin*, in *Methods in enzymology*. 2015, Elsevier. p. 563-608 % @ 0076-6879.
32. DaCosta, R.S., H. Andersson, and B.C. Wilson, *Molecular fluorescence excitation–emission matrices relevant to tissue spectroscopy*. Photochemistry and photobiology, 2003. **78**(4): p. 384-392 % @ 0031-8655.
33. Pan, Y.-L., *Detection and characterization of biological and other organic-carbon aerosol particles in atmosphere using fluorescence*. Journal of Quantitative Spectroscopy and Radiative Transfer, 2015. **150**: p. 12-35 % @ 0022-4073.

34. Hill, S.C., et al., *Fluorescence of bioaerosols: mathematical model including primary fluorescing and absorbing molecules in bacteria*. Optics Express, 2013. **21**(19): p. 22285-22313 %@ 1094-4087.
35. Lu, G. and B. Fei, *Medical hyperspectral imaging: a review*. Journal of biomedical optics, 2014. **19**(1): p. 010901 %@ 1083-3668.
36. DaCosta, R.S., H. Andersson, and B.C. Wilson, *Molecular Fluorescence Excitation–Emission Matrices Relevant to Tissue Spectroscopy*. Photochemistry and photobiology, 2003. **78**(4): p. 384-392 %@ 0031-8655.
37. Sweeny, L., et al., *Assessment of tissue autofluorescence and reflectance for oral cavity cancer screening*. Otolaryngology--Head and Neck Surgery, 2011. **145**(6): p. 956-960 %@ 0194-5998.
38. Mercadante, V., C. Paderni, and G. Campisi, *Novel non-invasive adjunctive techniques for early oral cancer diagnosis and oral lesions examination*. Current pharmaceutical design, 2012. **18**(34): p. 5442-5451 %@ 1389-2002.
39. De Veld, D.C.G., et al., *Autofluorescence and Raman microspectroscopy of tissue sections of oral lesions*. Lasers in medical science, 2005. **19**(4): p. 203-209 %@ 0268-8921.
40. Pu, Y., et al., *Native fluorescence spectra of human cancerous and normal breast tissues analyzed with non-negative constraint methods*. Applied optics, 2013. **52**(6): p. 1293-1301 %@ 2155-3165.
41. Monici, M., *Cell and tissue autofluorescence research and diagnostic applications*. Biotechnology annual review, 2005. **11**: p. 227-256 %@ 1387-2656.
42. Li, B.-H. and S.-S. Xie, *Autofluorescence excitation-emission matrices for diagnosis of colonic cancer*. World Journal of Gastroenterology: WJG, 2005. **11**(25): p. 3931.
43. Patterson, G.H., et al., *Separation of the glucose-stimulated cytoplasmic and mitochondrial NAD (P) H responses in pancreatic islet  $\beta$  cells*. Proceedings of the National Academy of Sciences, 2000. **97**(10): p. 5203-5207 %@ 0027-8424.
44. Chapman, S.K. and G.A. Reid, *Flavoprotein protocols*. Vol. 131 %@ 159259266X. 1999: Springer Science & Business Media.
45. Mewies, M., W.S. McIntire, and N.S. Scrutton, *Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: the current state of affairs*. Protein Science, 1998. **7**(1): p. 7-20 %@ 0961-8368.
46. Mahboub, S.B., *Unsupervised hyperspectral unmixing analysis with the modified dependent component analysis (DECA)*. 2017.
47. D.L. Nelson , M.M.C., *course smart international e book for principals of biochemistry*. palgrave Macmillan.
48. Kenealy, W.R. and J.G. Zeikus, *One-carbon metabolism in methanogens: evidence for synthesis of a two-carbon cellular intermediate and unification of catabolism and anabolism in Methanosarcina barkeri*. Journal of bacteriology, 1982. **151**(2): p. 932-941 %@ 0021-9193.
49. Pace, N.R., *The universal nature of biochemistry*. Proceedings of the National Academy of Sciences, 2001. **98**(3): p. 805-808 %@ 0027-8424.
50. Smith, E. and H.J. Morowitz, *Universality in intermediary metabolism*. Proceedings of the National Academy of Sciences, 2004. **101**(36): p. 13168-13173 %@ 0027-8424.

51. Kolenc, O.I. and K.P. Quinn, *Evaluating Cell Metabolism Through Autofluorescence Imaging of NAD (P) H and FAD*. Antioxidants & redox signaling % @ 1523-0864, 2018.
52. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation*. science, 2009. **324**(5930): p. 1029-1033 % @ 0036-8075.
53. Palmer, S., et al. *Discrimination of healthy and cancer cells of the bladder by metabolic state, based on autofluorescence*. 2015. International Society for Optics and Photonics.
54. Gatenby, R.A. and R.J. Gillies, *Why do cancers have high aerobic glycolysis?* Nature Reviews Cancer, 2004. **4**(11): p. 891 % @ 1474-1768.
55. Stringari, C., et al., *Metabolic trajectory of cellular differentiation in small intestine by Phasor Fluorescence Lifetime Microscopy of NADH*. Scientific reports, 2012. **2**: p. 568 % @ 2045-2322.
56. Conklin, M.W., et al., *Fluorescence lifetime imaging of endogenous fluorophores in histopathology sections reveals differences between normal and tumor epithelium in carcinoma in situ of the breast*. Cell biochemistry and biophysics, 2009. **53**(3): p. 145-157 % @ 1085-9195.
57. T, G., *Clinical fluorescence spectroscopy and imaging for the detection of early carcinoma by autofluorescence bronchoscopy and the study of the protoporphyrin IX pharmacokinetics in the endometrium*. PhD Thesis,, 2006.
58. Valentine, R.M., et al., *Modelling fluorescence in clinical photodynamic therapy*. Photochemical & Photobiological Sciences, 2013. **12**(1): p. 203-213.
59. Billinton, N. and A.W. Knight, *Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence*. Analytical biochemistry, 2001. **291**(2): p. 175-197 % @ 0003-2697.
60. Miah, M.I., *Fluorescence spectroscopy study of protoporphyrin IX metabolism level in cells*. Biopolymers: Original Research on Biomolecules, 2001. **62**(5): p. 237-240 % @ 0006-3525.
61. Kalaivani, R., et al., *Fluorescence spectra of blood components for breast cancer diagnosis*. Photomedicine and Laser Surgery, 2008. **26**(3): p. 251-256 % @ 1549-5418.
62. Kadish, K.M., K.M. Smith, and R. Guilard, *Handbook of porphyrin science with applications to chemistry, physics, materials science, engineering, biology and medicine: volume 23; synthesis*. 2012: World Scientific.
63. Ingrams, D.R., et al., *Autofluorescence characteristics of oral mucosa*. Head & Neck, 1997. **19**(1): p. 27-32 % @ 1043-3074.
64. Moesta, K.T., et al., *Protoporphyrin IX occurs naturally in colorectal cancers and their metastases*. Cancer research, 2001. **61**(3): p. 991-999 % @ 0008-5472.
65. Masilamani, V., et al., *Cancer diagnosis by autofluorescence of blood components*. Journal of Luminescence, 2004. **109**(3-4): p. 143-154 % @ 0022-2313.
66. Rajasekaran, R., et al., *Characterization and diagnosis of cancer by native fluorescence spectroscopy of human urine*. Photochemistry and photobiology, 2013. **89**(2): p. 483-491 % @ 0031-8655.
67. Phipps, J.E., et al., *Automated detection of breast cancer in resected specimens with fluorescence lifetime imaging*. Physics in Medicine & Biology, 2017. **63**(1): p. 015003 % @ 0031-9155.

68. Tadrous, P.J., et al., *Fluorescence lifetime imaging of unstained tissues: early results in human breast cancer*. The Journal of pathology, 2003. **199**(3): p. 309-317 %@ 1096-9896.
69. Breslin, T.M., et al., *Autofluorescence and diffuse reflectance properties of malignant and benign breast tissues*. Annals of Surgical Oncology, 2004. **11**(1): p. 65-70 %@ 1068-9265.
70. Keller, M.D., et al., *Autofluorescence and diffuse reflectance spectroscopy and spectral imaging for breast surgical margin analysis*. Lasers in surgery and medicine, 2010. **42**(1): p. 15-23 %@ 1096-9101.
71. Sharma, V., et al., *Auto-fluorescence lifetime and light reflectance spectroscopy for breast cancer diagnosis: potential tools for intraoperative margin detection*. Biomedical optics express, 2012. **3**(8): p. 1825-1840 %@ 2156-7085.
72. de Andrade Natal, R., et al. *Increased metabolic activity detected by FLIM in human breast cancer cells with desmoplastic reaction: a pilot study*. 2015. Optical Society of America.
73. Spakova, I., et al., *Autofluorescence of Breast Cancer Proteins*. Current Metabolomics, 2018. **6**(1): p. 2-9 %@ 2213-235X.
74. Coda, S., et al., *Fluorescence lifetime spectroscopy of tissue autofluorescence in normal and diseased colon measured ex vivo using a fiber-optic probe*. Biomedical optics express, 2014. **5**(2): p. 515-538 %@ 2156-7085.
75. Li, B., Z. Zhang, and S. Xie, *Steady state and time-resolved autofluorescence studies of human colonic tissues*. Chinese Optics Letters, 2006. **4**(6): p. 348-350.
76. Keller, R., et al., *Fluorescence endoscopy using a fluorescein-labeled monoclonal antibody against carcinoembryonic antigen in patients with colorectal carcinoma and adenoma*. Endoscopy, 2002. **34**(10): p. 801-807 %@ 0013-726X.
77. Lane, P.M., et al., *Simple device for the direct visualization of oral-cavity tissue fluorescence*. Journal of biomedical optics, 2006. **11**(2): p. 024006 %@ 1083-3668.
78. Sun, Y., et al., *Endoscopic fluorescence lifetime imaging for in vivo intraoperative diagnosis of oral carcinoma*. Microscopy and Microanalysis, 2013. **19**(4): p. 791-798 %@ 1431-9276.
79. Wang, M., et al., *Autofluorescence Imaging and Spectroscopy of Human Lung Cancer*. Applied Sciences, 2016. **7**(1): p. 32.
80. Xiao, S.D., et al., *Autofluorescence imaging analysis of gastric cancer*. Chinese Journal of Digestive Diseases, 2002. **3**(3): p. 95-98 %@ 1443-9611.
81. Kobayashi, M., et al., *Detection of early gastric cancer by a real-time autofluorescence imaging system*. Cancer letters, 2001. **165**(2): p. 155-159 %@ 0304-3835.
82. Rey-Barroso, L., et al., *Visible and extended near-infrared multispectral imaging for skin cancer diagnosis*. 2018. **18**(5): p. 1441.
83. Pladellorens, J., et al., *A device for the color measurement and detection of spots on the skin*. 2008. **14**(1): p. 65-70.
84. Jakovels, D., et al. *Evaluation of skin melanoma in spectral range 450-950 nm using principal component analysis*. in *European Conference on Biomedical Optics*. 2013. Optical Society of America.

85. Delpueyo, X., et al., *Multispectral imaging system based on light-emitting diodes for the detection of melanomas and basal cell carcinomas: a pilot study*. 2017. **22**(6): p. 065006.
86. Zeller, P., O. Ploux, and A. Méjean, *A simple protocol for attenuating the auto-fluorescence of cyanobacteria for optimized fluorescence in situ hybridization (FISH) imaging*. Journal of microbiological methods, 2016. **122**: p. 16-19 % @ 0167-7012.
87. Kubitscheck, U., *Fluorescence microscopy: from principles to biological applications*. 2017: John Wiley & Sons.
88. Rodrigues, R.M., et al., *Autofluorescence microscopy: a non-destructive tool to monitor mitochondrial toxicity*. Toxicology letters, 2011. **206**(3): p. 281-288 % @ 0378-4274.
89. Mahbub, S.B., et al., *Statistically strong label-free quantitative identification of native fluorophores in a biological sample*. Scientific reports, 2017. **7**(1): p. 15792 % @ 2045-2322.
90. gosnell, m., *Hyperspectral Imaging System Manual*. 2015.
91. Burroughes, J.H., et al., *Light-emitting diodes based on conjugated polymers*. nature, 1990. **347**(6293): p. 539 % @ 1476-4687.
92. Barolet, D., *Light-emitting diodes (LEDs) in dermatology*. , 27, 4, 2008. **27**(4): p. 227-238.
93. Swinehart, D.J.J.o.c.e., *The beer-lambert law*. 1962. **39**(7): p. 333.
94. HAMAMATSU, *ORCA-flash4.0lt*. 2016.
95. Gosnell, M.E., et al., *Quantitative non-invasive cell characterisation and discrimination based on multispectral autofluorescence features*. Scientific reports, 2016. **6**: p. 23453 % @ 2045-2322.
96. Buades, A., B. Coll, and J.-M. Morel, *A review of image denoising algorithms, with a new one*. Multiscale Modeling & Simulation, 2005. **4**(2): p. 490-530 % @ 1540-3459.
97. Maćkiewicz, A. and W. Ratajczak, *Principal components analysis (PCA)*. Computers & Geosciences, 1993. **19**(3): p. 303-342 % @ 0098-3004.
98. Zhong, Y., et al., *An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery*. 2006. **44**(2): p. 420-431.
99. Lee, G.A. and L.W. Hirst, *Ocular surface squamous neoplasia*. Survey of ophthalmology, 1995. **39**(6): p. 429-450 % @ 0039-6257.
100. Gichuhi, S., et al., *Epidemiology of ocular surface squamous neoplasia in Africa*. Tropical medicine & international health, 2013. **18**(12): p. 1424-1443 % @ 1365-3156.
101. Xu, Y., et al., *The clinical value of in vivo confocal microscopy for diagnosis of ocular surface squamous neoplasia*. Eye, 2012. **26**(6): p. 781 % @ 1476-5454.
102. Hirst, L.W., R.A. Axelsen, and I. Schwab, *Pterygium and associated ocular surface squamous neoplasia*. Archives of Ophthalmology, 2009. **127**(1): p. 31-32 % @ 0003-9950.
103. Basti, S. and M.S. Macsai, *Ocular surface squamous neoplasia: a review*. Cornea, 2003. **22**(7): p. 687-704 % @ 0277-3740.
104. Kim, Y.J., et al., *Conjunctival granuloma with necrosis associated with exposed suture in upper double lid masquerading as ocular surface squamous neoplasia: a case report*. BMC ophthalmology, 2017. **17**(1): p. 55 % @ 1471-2415.



105. Poothullil, A.M. and K.A. Colby. *Topical medical therapies for ocular surface tumors*. 2006. Taylor & Francis.
106. Zarei-Ghanavati, M., et al., *Changes in in vivo confocal microscopic findings of ocular surface squamous neoplasia during treatment with topical interferon alfa-2b*. The ocular surface % @ 1542-0124, 2018.
107. Radhakrishnan, A., *Squamous Neoplasia [OSSN]-A Brief Review*.
108. Ditta, L.C., Y. Shildkrot, and M.W. Wilson, *Outcomes in 15 patients with conjunctival melanoma treated with adjuvant topical mitomycin C: complications and recurrences*. Ophthalmology, 2011. **118**(9): p. 1754-1759 % @ 0161-6420.
109. Alomar, T.S., et al., *Corneal intraepithelial neoplasia: in vivo confocal microscopic study with histopathologic correlation*. American journal of ophthalmology, 2011. **151**(2): p. 238-247 % @ 0002-9394.
110. Nguena, M.B., et al., *Diagnosing Ocular Surface Squamous Neoplasia in East Africa*. Ophthalmology, 2014. **121**(2): p. 484-491 % @ 0161-6420.
111. Gichuhi, S., et al., *Clinical presentation of ocular surface squamous neoplasia in Kenya*. JAMA ophthalmology, 2015. **133**(11): p. 1305-1313 % @ 2168-6165.
112. Barros, J.N., et al., *Predictive index to differentiate invasive squamous cell carcinoma from preinvasive ocular surface lesions by impression cytology*. British Journal of Ophthalmology, 2009. **93**(2): p. 209-214 % @ 0007-1161.
113. Shyamala, K., H.C. Girish, and S. Murgod, *Risk of tumor cell seeding through biopsy and aspiration cytology*. Journal of International Society of Preventive & Community Dentistry, 2014. **4**(1): p. 5.
114. Galor, A., et al., *Predictors of ocular surface squamous neoplasia recurrence after excisional surgery*. Ophthalmology, 2012. **119**(10): p. 1974-1981 % @ 0161-6420.
115. Tananuvat, N., et al., *Role of impression cytology in diagnosis of ocular surface neoplasia*. Cornea, 2008. **27**(3): p. 269-274 % @ 0277-3740.
116. Char, D.H., et al., *20 MHz high frequency ultrasound assessment of scleral and intraocular conjunctival squamous cell carcinoma*. British journal of ophthalmology, 2002. **86**(6): p. 632-635 % @ 0007-1161.
117. Balestrazzi, A., et al., *Corneal invasion of ocular surface squamous neoplasia after clear corneal phacoemulsification: in vivo confocal microscopy analysis*. Journal of Cataract & Refractive Surgery, 2008. **34**(6): p. 1038-1043 % @ 0886-3350.
118. Vajzovic, L.M., et al., *Ultra high-resolution anterior segment optical coherence tomography in the evaluation of anterior corneal dystrophies and degenerations*. Ophthalmology, 2011. **118**(7): p. 1291-1296 % @ 0161-6420.
119. Ong, S.S., G.K. Vora, and P.K. Gupta, *Anterior segment imaging in ocular surface squamous neoplasia*. Journal of ophthalmology, 2016. **2016** % @ **2090-004X**.
120. Auner, G.W., et al., *Applications of Raman spectroscopy in cancer diagnosis*. 2018. **37**(4): p. 691-717.
121. Drozdowicz-Tomsia, K., et al., *Multiphoton fluorescence lifetime imaging microscopy reveals free-to-bound NADH ratio changes associated with metabolic inhibition*. Journal of biomedical optics, 2014. **19**(8): p. 086016 % @ 1083-3668.

122. Borst, J.W. and A.J.W.G. Visser, *Fluorescence lifetime imaging microscopy in life sciences*. Measurement Science and Technology, 2010. **21**(10): p. 102002 % @ 0957-0233.
123. Wang, M., et al., *Rapid diagnosis and intraoperative margin assessment of human lung cancer with fluorescence lifetime imaging microscopy*. BBA clinical, 2017. **8**: p. 7-13 % @ 2214-6474.
124. Wang, Y., et al., *Rapid, label-free, and highly sensitive detection of cervical cancer with fluorescence lifetime imaging microscopy*. IEEE Journal of Selected Topics in Quantum Electronics, 2016. **22**(3): p. 228-234 % @ 1077-260X.
125. Jahr, W., et al., *Hyperspectral light sheet microscopy*. Nature communications, 2015. **6**: p. 7990 % @ 2041-1723.
126. Luthman, A.S., et al., *Fluorescence hyperspectral imaging (fHSI) using a spectrally resolved detector array*. Journal of biophotonics, 2017. **10**(6-7): p. 840-853 % @ 1864-0648.
127. Rehman, A.U., et al., *Fluorescence quenching of free and bound NADH in HeLa cells determined by hyperspectral imaging and unmixing of cell autofluorescence*. Biomedical optics express, 2017. **8**(3): p. 1488-1498 % @ 2156-7085.
128. Vidal, M. and J.M. Amigo, *Pre-processing of hyperspectral images. Essential steps before image analysis*. Chemometrics and Intelligent Laboratory Systems, 2012. **117**: p. 138-148 % @ 0169-7439.
129. Pych, W., *A Fast Algorithm for Cosmic-Ray Removal from Single Images*. Publications of the Astronomical Society of the Pacific, 2003. **116**(816): p. 148 % @ 1538-3873.
130. Zhang, B., J.M. Fadili, and J.-L. Starck, *Wavelets, ridgelets, and curvelets for Poisson noise removal*. IEEE Transactions on Image Processing, 2008. **17**(7): p. 1093-1108 % @ 1057-7149.
131. Lin, E.-B. and Y. Ling, *Image compression and denoising via nonseparable wavelet approximation*. Journal of computational and applied mathematics, 2003. **155**(1): p. 131-152 % @ 0377-0427.
132. Zhang, Q., Y. Wang, and L. Wang, *Registration of images with affine geometric distortion based on maximally stable extremal regions and phase congruency*. Image and Vision Computing, 2015. **36**: p. 23-39 % @ 0262-8856.
133. de Lannoy, G., et al. *Feature Relevance Assessment in Automatic Inter-patient Heart Beat Classification*. 2010.
134. Drukker, K., et al., *Mammographic quantitative image analysis and biologic image composition for breast lesion characterization and classification*. Medical physics, 2014. **41**(3 % @ 2473-4209).
135. Anderson, M.J., *A new method for non-parametric multivariate analysis of variance*. Austral ecology, 2001. **26**(1): p. 32-46 % @ 1442-9993.
136. Michalski, R.S., J.G. Carbonell, and T.M.X. Mitchell, *Machine learning: An artificial intelligence approach*. 2013: Springer Science & Business Media.
137. Lee, A., et al., *Machine learning has arrived!* Ophthalmology, 2017. **124**(12): p. 1726-1728 % @ 0161-6420.
138. Markowski, C.A. and E.P. Markowski, *Conditions for the effectiveness of a preliminary test of variance*. The American Statistician, 1990. **44**(4): p. 322-326 % @ 0003-1305.

139. Khan, J., et al., *Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks*. Nature medicine, 2001. **7**(6): p. 673 % @ 1546-170X.
140. Bose, S., et al., *Generalized quadratic discriminant analysis*. Pattern Recognition, 2015. **48**(8): p. 2676-2684 % @ 0031-3203.
141. Loh, W.Y., *Classification and regression trees*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2011. **1**(1): p. 14-23 % @ 1942-4795.
142. Furey, T.S., et al., *Support vector machine classification and validation of cancer tissue samples using microarray expression data*. Bioinformatics, 2000. **16**(10): p. 906-914 % @ 1460-2059.
143. Cunningham, P. and S.J. Delany, *k-Nearest neighbour classifiers*. Multiple Classifier Systems, 2007. **34**: p. 1-17.
144. Refaeilzadeh, P., L. Tang, and H. Liu, *Cross-validation*, in *Encyclopedia of database systems*. 2009, Springer. p. 532-538.
145. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve*. Radiology, 1982. **143**(1): p. 29-36 % @ 0033-8419.
146. Yung, M., M.A. Klufas, and D. Sarraf, *Clinical applications of fundus autofluorescence in retinal disease*. International journal of retina and vitreous, 2016. **2**(1): p. 12 % @ 2056-9920.
147. Ly, A., et al., *Multispectral Pattern Recognition Reveals a Diversity of Clinical Signs in Intermediate Age-Related Macular Degeneration*. Investigative ophthalmology & visual science, 2018. **59**(5): p. 1790-1799 % @ 1552-5783.
148. Gichuhi, S., et al., *Topical fluorouracil after surgery for ocular surface squamous neoplasia in Kenya: a randomised, double-blind, placebo-controlled trial*. The Lancet Global Health, 2016. **4**(6): p. e378-e385 % @ 2214-109X.
149. Radcliffe, N.M. and P.T. Finger, *Eye cancer related glaucoma: current concepts*. Survey of ophthalmology, 2009. **54**(1): p. 47-73 % @ 0039-6257.
150. McCarthy, B.J., et al., *Factors associated with survival in patients with meningioma*. Journal of neurosurgery, 1998. **88**(5): p. 831-839 % @ 0022-3085.
151. Condra, K.S., et al., *Benign meningiomas: primary treatment selection affects survival*. International journal of radiation oncology, biology, physics, 1997. **39**(2): p. 427-436 % @ 0360-3016.
152. Backes, Y., et al., *Risk for incomplete resection after macroscopic radical endoscopic resection of T1 colorectal cancer: a multicenter cohort study*. The American journal of gastroenterology, 2017. **112**(5): p. 785 % @ 1572-0241.
153. Leung, F.W., *Incomplete resection after macroscopic radical endoscopic resection of T1 colorectal cancer—should a paradigm-changing approach to address the risk be considered?* Translational gastroenterology and hepatology, 2017. **2**.
154. Karp, C.L., et al., *Subconjunctival/perilesional recombinant interferon  $\alpha 2b$  for ocular surface squamous neoplasia: a 10-year review*. Ophthalmology, 2010. **117**(12): p. 2241-2246 % @ 0161-6420.
155. Zoroquiain, P., et al., *High frequency of squamous intraepithelial neoplasia in pterygium related to low ultraviolet light exposure*. Saudi Journal of Ophthalmology, 2016. **30**(2): p. 113-116 % @ 1319-4534.

156. Blacker, T.S., et al., *Separating NADH and NADPH fluorescence in live cells and tissues using FLIM*. Nature communications, 2014. **5**: p. 3936 %@ 2041-1723.
157. Wright, B.K., et al., *Phasor-flim analysis of NADH distribution and localization in the nucleus of live progenitor myoblast cells*. Microscopy research and technique, 2012. **75**(12): p. 1717-1722 %@ 1059-910X.
158. Habibalahi, A., et al., *Novel automated non invasive detection of ocular surface squamous neoplasia using multispectral autofluorescence imaging*. 2019.
159. Bentivoglio, A.R., et al., *Analysis of blink rate patterns in normal subjects*. 1997. **12**(6): p. 1028-1034.
160. Du, Q. and H. Yang, *Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis*. IEEE Geosci. Remote Sensing Lett., 2008. **5**(4): p. 564-568.
161. Wu, B., et al., *A comparative evaluation of filter-based feature selection methods for hyper-spectral band selection*. International journal of remote sensing, 2013. **34**(22): p. 7974-7990 %@ 0143-1161.
162. Ding, C. and H. Peng, *Minimum redundancy feature selection from microarray gene expression data*. Journal of bioinformatics and computational biology, 2005. **3**(02): p. 185-205 %@ 0219-7200.
163. Liu, J., S. Ranka, and T. Kahveci, *Classification and feature selection algorithms for multi-class CGH data*. Bioinformatics, 2008. **24**(13): p. i86-i95 %@ 1460-2059.
164. Ahmad, M.S., et al., *A new gender-specific model for skin autofluorescence risk stratification*. Scientific reports, 2015. **5**: p. 10198 %@ 2045-2322.
165. Nguyen, H.B., et al. *PSO and statistical clustering for feature selection: a new representation*. in *Asia-Pacific Conference on Simulated Evolution and Learning*. 2014. Springer.
166. Vapnik, V., *The nature of statistical learning theory*. 2013: Springer science & business media.
167. Beni, G., *Swarm intelligence*. Encyclopedia of Complexity and Systems Science, 2009: p. 1-32 %@ 3642277373.
168. Panigrahi, B.K., Y. Shi, and M.-H. Lim, *Handbook of swarm intelligence: concepts, principles and applications*. Vol. 8 %@ 364217390X. 2011: Springer Science & Business Media.
169. Hu, X., Y. Shi, and R. Eberhart. *Recent advances in particle swarm*. 2004. IEEE.
170. Garnier, S., J. Gautrais, and G. Theraulaz, *The biological principles of swarm intelligence*. Swarm Intelligence, 2007. **1**(1): p. 3-31 %@ 1935-3812.
171. Blum, C. and X. Li, *Swarm intelligence in optimization*, in *Swarm Intelligence*. 2008, Springer. p. 43-85.
172. Engelbrecht, A.P., *Fundamentals of computational swarm intelligence*. 2006: John Wiley & Sons.
173. Poli, R., J. Kennedy, and T. Blackwell, *Particle swarm optimization*. Swarm intelligence, 2007. **1**(1): p. 33-57 %@ 1935-3812.
174. Shi, Y. *Particle swarm optimization: developments, applications and resources*. 2001. IEEE.
175. Zhang, C., et al., *A novel hybrid differential evolution and particle swarm optimization algorithm for unconstrained optimization*. Operations Research Letters, 2009. **37**(2): p. 117-122 %@ 0167-6377.

176. Dorigo, M. and M. Birattari, *Ant colony optimization*, in *Encyclopedia of machine learning*. 2011, Springer. p. 36-39.
177. Shamsipur, M., et al., *Ant colony optimisation: a powerful tool for wavelength selection*. Journal of Chemometrics: A Journal of the Chemometrics Society, 2006. **20**(3-4): p. 146-157 % @ 0886-9383.
178. Resson, H.W., et al., *Peak selection from MALDI-TOF mass spectra using ant colony optimization*. Bioinformatics, 2007. **23**(5): p. 619-626 % @ 1460-2059.
179. Askari, E., P. Flores, and F. Silva, *A particle swarm-based algorithm for optimization of multi-layered and graded dental ceramics*. Journal of the mechanical behavior of biomedical materials, 2018. **77**: p. 461-469 % @ 1751-6161.
180. Abeel, T., et al., *Robust biomarker identification for cancer diagnosis with ensemble feature selection methods*. Bioinformatics, 2009. **26**(3): p. 392-398 % @ 1460-2059.
181. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015*. CA: a cancer journal for clinicians, 2015. **65**(1): p. 5-29 % @ 0007-9235.
182. Trotter, S.C., et al., *A global review of melanoma follow-up guidelines*. The Journal of clinical and aesthetic dermatology, 2013. **6**(9): p. 18.
183. Dicker, D.T., et al., *Differentiation of normal skin and melanoma using high resolution hyperspectral imaging*. Cancer biology & therapy, 2006. **5**(8): p. 1033-1038 % @ 1538-4047.
184. Croce, A.C., et al., *Dependence of fibroblast autofluorescence properties on normal and transformed conditions. Role of the metabolic activity*. Photochemistry and photobiology, 1999. **69**(3): p. 364-374 % @ 1751-1097.
185. Yuanlong, Y., et al., *Characteristic autofluorescence for cancer diagnosis and its origin*. Lasers in surgery and medicine, 1987. **7**(6): p. 528-532 % @ 0196-8092.
186. Lakowicz, J.R., *Principles of frequency-domain fluorescence spectroscopy and applications to cell membranes*, in *Fluorescence Studies on Biological Membranes*. 1988, Springer. p. 89-126.
187. Ramanujam, N., *Fluorescence spectroscopy of neoplastic and non-neoplastic tissues*. Neoplasia, 2000. **2**(1-2): p. 89-117 % @ 1476-5586.
188. Liu, X., et al., *Analysis of cell fate from single-cell gene expression profiles in C. elegans*. Cell, 2009. **139**(3): p. 623-633 % @ 0092-8674.
189. Altschuler, S.J. and L.F. Wu, *Cellular heterogeneity: do differences make a difference?* Cell, 2010. **141**(4): p. 559-563 % @ 0092-8674.
190. Coda, S., et al., *Biophotonic endoscopy: a review of clinical research techniques for optical imaging and sensing of early gastrointestinal cancer*. Endoscopy international open, 2015. **3**(5): p. E380.
191. Zhao, Y., et al., *Longitudinal label-free tracking of cell death dynamics in living engineered human skin tissue with a multimodal microscope*. Biomedical optics express, 2014. **5**(10): p. 3699-3716 % @ 2156-7085.
192. El Aziz, M.A., I.M. Selim, and S. Xiong, *Automatic detection of galaxy type from datasets of galaxies image based on image retrieval approach*. Scientific Reports, 2017. **7**(1): p. 4463 % @ 2045-2322.
193. Zhu, H., et al., *Hyperspectral imaging for presymptomatic detection of tobacco disease with successive projections algorithm and machine-learning classifiers*. Scientific Reports, 2017. **7**(1): p. 4125 % @ 2045-2322.

194. Fernández, A. and S. Gómez, *Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms*. Journal of Classification, 2008. **25**(1): p. 43-65 %@ 0176-4268.
195. Haralick, R.M., *Statistical and structural approaches to texture*. Proceedings of the IEEE, 1979. **67**(5): p. 786-804 %@ 0018-9219.
196. Haralick, R.M. and K. Shanmugam, *Textural features for image classification*. IEEE Transactions on systems, man, and cybernetics, 1973(6): p. 610-621 %@ 0018-9472.
197. Van Belle, G., *Statistical rules of thumb*. Vol. 699 %@ 1118210360. 2011: John Wiley & Sons.
198. Gelman, A., *Analysis of variance—why it is more important than ever*. The annals of statistics, 2005. **33**(1): p. 1-53 %@ 0090-5364.
199. Neter, J., et al., *Applied linear statistical models*. Vol. 4. 1996: Irwin Chicago.
200. MathWorks, *Matlab documentation*. 2018.
201. Mukaka, M.M., *A guide to appropriate use of correlation coefficient in medical research*. Malawi Medical Journal, 2012. **24**(3): p. 69-71 %@ 1995-7262.
202. Clarke, G.M. and D. Cooke, *A basic course in statistics*. 1978.
203. Balakrishnama, S. and A. Ganapathiraju, *Linear discriminant analysis-a brief tutorial*. Institute for Signal and information Processing, 1998. **18**: p. 1-8.
204. Ye, J., R. Janardan, and Q. Li. *Two-dimensional linear discriminant analysis*. 2005.
205. Kohavi, R. and G.H. John, *Wrappers for feature subset selection*. Artificial intelligence, 1997. **97**(1-2): p. 273-324 %@ 0004-3702.
206. Gosnell, M.E., et al. *Computer-assisted cystoscopy diagnosis of bladder cancer*. 2018. Elsevier.
207. Eroglu, Z. and A. Ribas, *Combination therapy with BRAF and MEK inhibitors for melanoma: latest evidence and place in therapy*. Therapeutic advances in medical oncology, 2016. **8**(1): p. 48-56 %@ 1758-8340.
208. Chapman, P.B., et al., *Improved survival with vemurafenib in melanoma with BRAF V600E mutation*. New England Journal of Medicine, 2011. **364**(26): p. 2507-2516 %@ 0028-4793.
209. Hauschild, A., et al., *Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial*. The Lancet, 2012. **380**(9839): p. 358-365 %@ 0140-6736.
210. Long, G.V., et al., *Adjuvant dabrafenib plus trametinib in stage III BRAF-mutated melanoma*. New England Journal of Medicine, 2017. **377**(19): p. 1813-1823 %@ 0028-4793.
211. Wang, J., et al., *Braf activation cooperates with Pten loss to regulate c-Myc activation in advanced prostate cancer*. Cancer research, 2012: p. canres. 0820.2012 %@ 0008-5472.
212. Miniatis, M.F., et al., *MEK1/2 inhibition decreases lactate in BRAF-driven human cancer cells*. Cancer research, 2013: p. canres. 1969.2012 %@ 0008-5472.
213. Yun, J., et al., *Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells*. Science, 2009. **325**(5947): p. 1555-1559 %@ 0036-8075.
214. Lodi, A., S.M. Woods, and S.M. Ronen, *Magnetic resonance-detectable metabolic consequences of MEK inhibition*. NMR in biomedicine, 2014. **27**(6): p. 700.

215. Chwirot, B.W., et al., *Fluorescence in situ detection of human cutaneous melanoma: study of diagnostic parameters of the method*. Journal of investigative dermatology, 2001. **117**(6): p. 1449-1451 %@ 0022-202X.