

A Hyperspectral Optical Surgical Tool for Circumferential Resection Margin Assessment in Gastrointestinal Cancer Surgery

Ioannis Gkouzionis

This thesis is submitted to fulfil the requirements for the
degree of Clinical Medicine Research Doctor of Philosophy
(PhD).

Imperial College London
Department of Surgery and Cancer
29/09/2023

Declaration of Originality

I hereby declare that this thesis is my own work, except for where indicated, and has not been submitted in any form for another degree or diploma at any university. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given in the bibliography.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Gastrointestinal (GI) cancers represent a significant contributor to the global cancer burden. Surgery typically constitutes the primary therapeutic approach for GI malignancies, with the primary objective of maximising tumour tissue excision while maintaining an adequate margin of healthy tissue (clear margin) surrounding the tumour. The presence of a clear margin, devoid of cancer cells, is essential in minimising the risk of cancer recurrence. Accurate mapping of tumour margins is crucial for achieving a negative resection margin. Local recurrence following surgical resection significantly impacts the overall survival and quality of life for patients with GI cancer.

In this thesis, we have investigated diffuse reflectance spectroscopy (DRS) technology, combined with artificial intelligence, to develop a pioneering optical surgical tool aimed at improving the identification of resection margins in GI cancer surgery.

The primary goal is to enhance the precision and reliability of tumour margin demarcation, thereby increasing the likelihood of complete tumour removal with clear margins, which is critical for reducing the risk of cancer recurrence and improving patient survival and quality of life. This work presents a two-fold unique contribution towards the optimal resection margins assessment in GI surgery. Firstly, by introducing a novel exploratory data analysis and classification routine employing deep neural networks to distinguish healthy and malignant tissues based on spectral measurements, and secondly, by developing a deep learning model to facilitate the detection and tracking of the DRS fibre probe during surgery. This approach represents a significant advancement over previous studies by providing a comprehensive analysis using machine learning techniques for GI cancer discrimination and addressing the challenge of probe detection and tracking for the first time.

Through validation in clinical settings, this work aims to advance the standard of care in GI cancer surgery, demonstrating the potential of DRS technology and AI to improve surgical outcomes.

Table of Contents

Chapter 1: Introduction	21
1.1 Introduction & Motivation	21
1.2 Thesis Objectives.....	22
1.3 Thesis Outline.....	24
Chapter 2: Understanding the Clinical and Technical Background	26
2.1 Oesophagogastric Cancer	26
2.1.1 Epidemiology	26
2.1.2 Pathophysiology.....	29
2.1.3 Staging System & Definitions	31
2.1.4 Treatment	35
2.2 Colorectal Cancer	37
2.2.1 Epidemiology	37
2.2.2 Pathophysiology.....	39
2.3 Surgical Resection Margins	41
2.4 Methods for Resection Margin Assessment	43
2.4.1 Current Methods for Margin Analysis.....	43
2.4.2 Methods for Intraoperative Margin Assessment (IMA)	44
2.5 Diagnosis & Histopathology.....	46
2.5.1 Benign vs Malignant Tumours	46
2.5.2 Histopathology Examination.....	47
2.5.3 Tumour Grading & Classification	48
2.5.4 Diagnosis & Treatment	48
2.6 Challenges & Opportunities	48
2.6.1 The Clinical Challenge	48
2.6.2 The Opportunity.....	50
2.7 Multispectral Methods	51
2.7.1 Optical Biopsy	51
2.7.2 Light – Tissue Interaction	51
2.7.3 Optical Spectroscopy in Tissue Analysis.....	54
2.7.4 Diffuse Reflectance Spectroscopy.....	54
2.7.5 Tissue Fluorescence Spectroscopy.....	57
2.7.6 Raman Spectroscopy.....	59
2.7.7 Fibre-Optic Spectroscopy.....	61
2.7.8 Advantages of Multispectral Cancer Screening Methods	64
2.8 Clinical Application of Optical Spectroscopy in GI Cancers	66
Chapter 3: Integrated DRS Probe Tracking Methods to Aid Tumour Margin Assessment ..	70
3.1 Background and Need for DRS Probe Detection	70
3.2 Visual Tracking	72
3.3 Object Representation.....	74
3.4 Evaluation Metrics.....	75
3.4.1 Reference Point	75
3.5 A Marker-based Probe Detection Methodology.....	76
3.5.1 Optical Tracking using 2D Markers	76

3.5.2	Optical DRS Probe Tracking using 2D Markers.....	77
3.5.3	Related Works.....	78
3.5.4	Marker Design.....	79
3.5.5	Camera Calibration	82
3.5.6	Marker Detection & Segmentation.....	83
3.5.7	DRS Probe Pose Estimation using PCA.....	86
3.5.8	DRS Probe Tip Tracking using a Kalman Filter.....	89
3.5.9	Spectroscopic Augmented Reality	95
3.5.10	Experimental Evaluation & Results	96
3.6	A Marker-less Neural Network-based Probe Detection Methodology	99
3.6.1	Related Work	101
3.6.2	You Only Look Once (YOLO) v5 Network Architecture	101
3.6.3	In-vivo Video Data Collection and Labelling.....	106
3.6.4	DRS Probe Detection Model Performance Evaluation Metrics.....	108
3.6.5	Experimental Evaluation and Results.....	111
3.7	Summary.....	119
<i>Chapter 4: DRS as a Surgical Guidance Tool in Gastrointestinal Cancer Surgery: Insights from Ex-vivo Investigations</i>		<i>121</i>
4.1	Introduction	121
4.2	Spectroscopy System and User Interface	122
4.3	Histopathology Correlation.....	124
4.4	Study Design.....	127
4.4.1	Differentiating Healthy Upper GI Tissue from Cancer: Insights from Ex-Vivo Study.....	127
4.4.2	Distinguishing Healthy Colorectal Tissue from Cancer: Ex-Vivo Study Findings	127
4.5	Data Acquisition Protocol	127
4.6	Data Analysis.....	130
4.7	Results	141
4.7.1	Differentiating Healthy Upper GI Tissue from Cancer: Insights from Ex-Vivo Study.....	141
4.7.2	Distinguishing Healthy Colorectal Tissue from Cancer: Ex-Vivo Study Findings	157
4.8	Summary.....	167
<i>Chapter 5: Bridging the Gap: In-Vivo DRS for Tissue Discrimination in Upper Gastrointestinal Surgery</i>		<i>169</i>
5.1	Introduction	169
5.2	In-vivo Spectroscopy System	170
5.3	Histopathology Correlation.....	172
5.4	Study Design.....	173
5.5	Data Acquisition Protocol	174
5.6	Data Analysis.....	176
5.7	Results	177
5.8	Summary.....	187
<i>Chapter 6: Conclusions and Future Work.....</i>		<i>190</i>
6.1	Conclusions	190

6.2	Future Work	194
	<i>Bibliography</i>	<i>198</i>

List of Tables

Table 1: T, N, and M status and histologic grade definitions for oesophagus and oesophagogastric junction cancer in the 8 th edition of the American Join Committee on Cancer (AJCC) Cancer Staging Manual.	32
Table 2: Projection error of DRS probe's tip location.	99
Table 3: Rotation angle and maximum detectable distance from the camera to track the DRS probe.	99
Table 4: Experimental settings.....	111
Table 5: Training parameters for the modified YOLOv5 network.....	113
Table 6: DRS probe QR600-7-SR-125F specifications.....	123
Table 7: Cohort characteristics of n=51 patients included in the analysis. Characteristics are presented for all patients in no. (%) as well as for gastric cancer and oesophageal cancer patients.	142
Table 8: Mean increase in overall accuracy, sensitivity, specificity, and AUC for all normalisation techniques compared to no normalisation for all classifiers for the stomach dataset.	145
Table 9: Mean increase in overall accuracy, sensitivity, specificity, and AUC for all normalisation techniques compared to no normalisation for all classifiers for the oesophagus dataset.	146
Table 10: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the stomach dataset. Only non-normalised and SNV normalised data was used.	147
Table 11: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the oesophagus dataset. Only non-normalised and SNV normalised data was used.	148
Table 12: Mean increase in classification performance metrics for all feature selection methods compared to the use of the entire spectrum for the stomach dataset. Only SNV normalised data was used.....	149
Table 13: Mean increase in classification performance metrics for all feature selection methods compared to the use of the entire spectrum for the stomach dataset. Only SNV normalised data was used.....	149
Table 14: Performance metrics for the spectral data classification using Approach 1 (reflectance values).	151
Table 15: Performance metrics for the spectral data classification using Approach 2 (reflectance values and shape-based extracted features).	151
Table 16: Performance metrics for the spectral data classification using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature).	152
Table 17: Performance metrics for the spectral data classification using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature).	152
Table 18: Performance metrics for the multiclass spectral data classification using SNV normalisation and Boruta feature selection for both stomach and oesophagus datasets.	157
Table 19: Cohort characteristics of n=23 patients included in the analysis. Characteristics are presented for all patients in no. (%).	158
Table 20: Mean increase in overall accuracy, sensitivity, specificity, and AUC for SNV normalisation compared to no normalisation for all classifiers for the colorectal dataset.....	161

Table 21: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the colorectal dataset. Only non-normalised and SNV normalised data was used.	162
Table 22: Mean increase in classification performance metrics for all feature selection methods compared to the use of the entire spectrum for the stomach dataset. Only SNV normalised data was used.	163
Table 23: Performance metrics for the spectral data classification using Approach 1 (reflectance values).	164
Table 24: Performance metrics for the spectral data classification using Approach 2 (reflectance values and shape-based extracted features).	164
Table 25: Performance metrics for the spectral data classification using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature).	165
Table 26: Performance metrics for the spectral data classification using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature).	165
Table 27: Cohort characteristics of n=10 patients included in the analysis. Characteristics are presented for all patients in no. (%) as well as for gastric cancer and oesophageal cancer patients.	177
Table 28: Mean increase in overall accuracy, sensitivity, specificity, and AUC for SNV normalisation technique compared to no normalisation for all classifiers for the in-vivo stomach dataset.	180
Table 29: Mean increase in overall accuracy, sensitivity, specificity, and AUC for SNV normalisation technique compared to no normalisation for all classifiers for the in-vivo oesophagus dataset.	181
Table 30: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the in-vivo stomach dataset. Only SNV normalised data is used.	182
Table 31: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the in-vivo oesophagus dataset. Only SNV normalised data is used.	182
Table 32: Mean increase in classification performance metrics for Boruta feature selection method compared to the use of the entire spectrum for the in-vivo stomach dataset. Only SNV normalised data is used.	183
Table 33: Mean increase in classification performance metrics for Boruta feature selection method compared to the use of the entire spectrum for the in-vivo oesophagus dataset. Only SNV normalised data is used.	184
Table 34: Performance metrics for the spectral data classification using Approach 1 (reflectance values) using SNV normalisation and Boruta feature selection.	184
Table 35: Performance metrics for the spectral data classification using Approach 2 (reflectance values and shape-based extracted features) using SNV normalisation and Boruta feature selection.	185
Table 36: Performance metrics for the spectral data classification using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) using SNV normalisation and Boruta feature selection.	185
Table 37: Performance metrics for the spectral data classification using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) using SNV normalisation and Boruta feature selection.	186

List of Figures

Figure 1: (a) Estimated number of new cancer cases in 2020 (world). (b) Estimated number of cancer-related deaths in 2020 (world). (c) Estimated number of prevalent cancer cases (5-year) in 2020 (world) [Source: GLOBOCAN 2020, https://gco.iarc.fr/today/home].	22
Figure 2: Estimated age-standardised incidence rates (world) in 2020, oesophageal cancer, both sexes, all ages [Source: GLOBOCAN 2020, https://gco.iarc.fr/today/home].	27
Figure 3: Estimated age-standardised incidence rates (world) in 2020, gastric cancer, both sexes, all ages [Source: GLOBOCAN 2020, https://gco.iarc.fr/today/home].	28
Figure 4: Estimated new cases (a) and deaths (b) of oesophageal cancer worldwide in 2020 and projections to 2030 and 2040. Numbers are shown in the scale of x10,000 [Source: GLOBOCAN 2020, https://gco.iarc.fr/today/home].	28
Figure 5: Survival of oesophageal cancer patients per TNM (7 th edition) stage at diagnosis [Source: http://www.cijfersoverkanker.nl].	30
Figure 6: Survival of gastric cancer patients per TNM stage at diagnosis (TNM 7 th edition) [Source: http://www.cijfersoverkanker.nl].	31
Figure 7: Estimated age-standardised incidence rates (world) in 2020, colorectal cancer, both sexes, all ages [Source: GLOBOCAN 2020, https://gco.iarc.fr/today/home].	39
Figure 8: Colorectal cancer stages and development (adopted by Hossain et al. ⁶⁸).	40
Figure 9: (Left) A surgically excised specimen with an inadequate or positive surgical margin; the cancerous tissue is at the edge of the specimen. (Right) A surgically excised specimen with an adequate margin along the resected surface; the cancerous tissue is in the centre of the surgically excised specimen, with a rim of healthy tissue surrounding it. (adopted by Tringale et al. ⁸⁴).	42
Figure 10: Absorption spectra of oxy- and deoxyhaemoglobin and water in the spectral range 400-1400 nm (adopted by Regensburger et al. ¹⁴⁵).	53
Figure 11: Basic principle of DRS demonstrating absorption and scattering inside tissue (created with BioRender.com).	54
Figure 12: Principle of Raman scattering. (A) Raman scattering and Rayleigh scattering. (B) Energy level diagram showing the states involved in Raman signal (Raman scattering, Rayleigh scattering, and infrared absorption) (adopted by Liu et al. ¹⁶⁴).	60
Figure 13: Fibre-optic spectroscopy technologies. (i) Diffuse light spectroscopy involves the measurement of scattered light across a range of wavelengths providing information about tissue composition and structure. (ii) Depth-selective spectroscopy utilises fibre-optic probes with controlled light penetration and selectively investigates specific tissue depths, such as the mucosa layer, to target particular regions of interest. (iii) Elastic scattering spectroscopy examines the redirection of incident light without altering its energy, providing insights into tissue morphology and cellular structure. (iv) Absorption spectroscopy focuses on the absorption of specific wavelengths of light by tissue constituents, revealing details about molecular content, including chromophores like haemoglobin. (v) Fluorescence spectroscopy examines the emission of fluorescent light after tissue is excited with high-energy photons, enabling the detection of naturally occurring fluorophores and exogenous probes. (vi) Raman scattering analyses the inelastic scattering of light, unveiling molecular vibrational and rotational information, facilitating molecular identification (adopted by Roy et al. ¹⁶⁸).	62
Figure 14: (a) Average colour distribution in the H – S plane for 3,000 laparoscopic images, together with the H – S occupation of the potential colour marker in the highlighted oval ²²⁵ . Additionally, it's evidence that green hues could also be used as potential marker. (b) DRS fibre probe with the green colour marker attached.	80
Figure 15: DRS fibre probe detection and tracking pipeline.	82

Figure 16: DRS fibre probe detection workflow. (a) Example input video frame. (b) Blurred frame. (c) Hue channel of the blurred input frame. (d) Binary mask result of HSV segmentation. (e) Detected edge lines of the probe. (f) Detected tip point of the probe.	88
Figure 17: Flowchart of a Kalman Filter Matrix process (https://medium.com/@jaems33/understanding-kalman-filters-with-python-2310e87b8f48)	90
Figure 18: Illustration of DRS probe tip detection during sampling ex-vivo gastric specimen. Blue point: detected probe's tip location using the marker-based methodology. Green points: optical biopsy sites.	97
Figure 19: (a) and (b) Illustration of probe tracking error with overlaid ground truth (blue line) and estimated (yellow line) probe tip position for two video sequences (500 and 450 frames each). (c) – (f) The ground truth and estimated projected DRS probe tip pixel position for x and y axis for each of the video sequences. Blue and red lines indicate the estimated and ground truth probe tip position, respectively.	98
Figure 20: Modified YOLOv5-based network architecture ²⁵¹	104
Figure 21: Example frames extracted from two different in-vivo videos.....	107
Figure 22: Video data processing and labelling pipeline ²⁵¹	107
Figure 23: Frame annotation example showing (a) the bounding box and (b) the probe tip point.	108
Figure 24: Intersection over Union explained [Source: https://www.v7labs.com/blog/yolo-object-detection].	109
Figure 25: Model's performance evaluation using precision, recall, mAP@0.5, and mAP@0.5:95.	113
Figure 26: Comparison of the models' accuracy between Zhong's modified YOLOv5 network that was trained using only ex-vivo imaging data, and the current thesis's work where Zhong's pretrained network was fine-tuned using in-vivo data.	115
Figure 27: Examples of DRS probe tip prediction location using the fine-tuned modified YOLOv5 model. The green point is the ground truth annotated tip location, and the blue point is the predicted tip.	116
Figure 28: Integrated video analysis workflow.	117
Figure 29: Illustration of video output featuring trajectory tracking. Historical detection points of the DRS probe tip are marked in green and current tip location highlighted in blue. (a) Full probe tip locations trajectory. (b) Most recent 100 probe tip locations trajectory....	118
Figure 30: A comprehensive representation of the configuration workflow and versatile video output.	119
Figure 31: (a) Schematic illustration of the DRS instrumentation. (b) Photograph of the setup. The DRS probe was connected to both the light source and the spectrometer to allow for data acquisition and sample illumination. All electronic devices communicated with proprietary software designed in-house with Python v.3.9 (Python Software Foundation). VNIR indicates visible and near-infrared.	123
Figure 32: Python proprietary software for hardware DRS system control, data acquisition and real-time spectral classification. The Control Panel allows for execution of the data acquisition protocol, DRS probe detection and tracking, hardware calibration and light source control. The Augmented Live Video View hosted the live video feed from the RGB camera. Reflectance data in the 400-1000 nm spectral range displayed in real-time under the corresponding view of the UI.	124
Figure 33: Histopathology correlation pipeline for ground truth labels generation.	125
Figure 34: Histopathological correlation workflow is shown on an example of oesophageal adenocarcinoma. The correlation was performed using yellow tissue paint (marked with a yellow arrow across panels (a) – (e)). (a) The ex-vivo specimen with tracked optical biopsy sites of the suspected tumour area. (b) Ex-vivo specimen following painting of the suspected	

tumour tissue with yellow paint. (c) Specimen, having been placed in formalin, with yellow painted area visible. (d) Macroscopic slice of oesophageal tissue with the yellow painted area. (e) Microscopic H&E-stained slice with yellow paint (yellow arrow) and confirmed oesophageal adenocarcinoma (red outline).....	126
Figure 35: Histopathology correlation for a colorectal specimen. (a) Photo of scanned colon tissue with tracked DRS locations over suspected tumour area indicated. (b) Photo of colon tissue with painted yellow dye over the suspected tumour tissue. (c) Photo of colon tissue after being in formalin with yellow tissue dye visible. (d) Photo of sliced colon tissue with yellow dye visible over the suspected tumour. (e) Scanned H&E slide showing yellow tissue dye over confirmed tumour tissue (marked in red).....	126
Figure 36: Photograph of the DRS system during ex-vivo data acquisition of an oesophageal specimen.	128
Figure 37: Workflow for ex-vivo data acquisition.	130
Figure 38: The process of stratified group five-fold cross-validation. When fitting a model, each row represents the split between training (blue) and validation datasets (orange). We typically estimate the final performance metrics through an average of the runs.	135
Figure 39: The architecture of the proposed one-dimensional convolutional neural network for colorectal tissue classification. The SpecNet is a five-layer convolutional neural network (CNN) for learning visual representations in diffuse reflectance spectroscopy (DRS) data. It uses 20 kernels to filter the input data, and the output layer classifies spectra based on the spectral band size and number of output classes in the dataset.	138
Figure 40: Data analysis framework for ex-vivo gastrointestinal tissue classification.	140
Figure 41: Mean spectra with standard deviation for (a) normal stomach and gastric cancer, and (b) normal oesophagus and oesophageal cancer.	143
Figure 42: Illustration of shape-based features extracted for (a) stomach and (b) oesophagus datasets. (1) Depicts the secondary peak intensity value and corresponding wavelength in the 480 – 550 nm spectral range. (2) Depicts the primary peak intensity value and corresponding wavelength in the 580 – 700 nm spectral range.....	143
Figure 43: Selected features for the stomach dataset for each normalisation method. (a) Boruta and Permutation Feature Importance after SNV normalisation. (b) Boruta and Permutation Feature Importance after AUC normalisation. (c) Illustration of the 8 selected spectral bands. Light grey and light blue areas represent individual spectral bands. (d) Boruta and Permutation Feature Importance with no normalisation.	144
Figure 44: Selected features for the oesophagus dataset for each normalisation method. (a) Boruta and Permutation Feature Importance after SNV normalisation. (b) Boruta and Permutation Feature Importance after AUC normalisation. (c) Illustration of the 6 selected spectral bands. Light grey and light blue areas represent individual spectral bands. (d) Boruta and Permutation Feature Importance with no normalisation.	144
Figure 45: ROC curves with AUC values for all classifiers using Approach 1 (entire spectrum) for (a) stomach dataset (SNV normalisation, no feature selection) and (b) oesophagus dataset (no normalisation, Boruta feature selection).	151
Figure 46: ROC curves with AUC values for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for (a) stomach dataset (SNV normalisation, Boruta feature selection) and (b) oesophagus dataset (SNV normalisation, Boruta feature selection).	151
Figure 47: ROC curves with AUC values for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for (a) stomach dataset (SNV normalisation, Boruta feature selection) and (b) oesophagus dataset (SNV normalisation, Boruta feature selection).	152

Figure 48: ROC curves with AUC values for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for (a) stomach dataset (SNV normalisation, Boruta feature selection) and (b) oesophagus dataset (SNV normalisation, Boruta feature selection).	152
Figure 49: Illustration of real-time DRS probe tracking at each optical biopsy site coupled with the binary classification of each site on a gastric and oesophageal tissue specimen. (a) – (c) Real-time visual correlation of scanned optical biopsy sites on tissue (green dots on each frame) using the DRS probe and marker-based tracking. (d) – (f) Real-time tracking at each optical biopsy site coupled with binary classification of each site as either normal (100% green) or tumour tissue (100% pink) via a graduated colour map. Tissue type highlighted on screen in real-time (bottom left of each frame).	154
Figure 50: Labels distribution for (a) gastric and (b) oesophageal cancer pathological (TNM) staging.	155
Figure 51: Mean spectra of the different pathological tumour stages for (a) gastric cancer, and (b) oesophageal cancer.	155
Figure 52: Selected features for the tumour pathological staging of the stomach dataset for each normalisation method. (a) Boruta and Permutation Feature Importance with no normalisation. (b) Boruta and Permutation Feature Importance after SNV normalisation. (c) Boruta and Permutation Feature Importance after AUC normalisation.	156
Figure 53: Selected features for the tumour pathological staging of the oesophagus dataset for each normalisation method. (a) Boruta and Permutation Feature Importance with no normalisation. (b) Boruta and Permutation Feature Importance after SNV normalisation. (c) Boruta and Permutation Feature Importance after AUC normalisation.	156
Figure 54: ROC curves for one-vs-rest multiclass classification for (a) stomach and (b) oesophagus pathological staging.	157
Figure 55: Mean spectra with standard deviation for normal and tumour colorectal tissue.	159
Figure 56: Illustration of shape-based features extracted for the colorectal dataset. (1) Depicts the secondary peak intensity value and corresponding wavelength in the 480 – 550 nm spectral range. (2) Depicts the primary peak intensity value and corresponding wavelength in the 580 – 700 nm spectral range.	159
Figure 57: Selected features for the colorectal dataset for each normalisation method. (a) Boruta and Permutation Feature Importance with no normalisation. (b) Boruta and Permutation Feature Importance after SNV normalisation. (c) Boruta and Permutation Feature Importance with AUC normalisation.	160
Figure 58: ROC curves with AUC values for all classifiers using Approach 1 (entire spectrum) for colorectal dataset (SNV normalisation, Boruta feature selection).	164
Figure 59: ROC curves with AUC values for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for colorectal dataset (SNV normalisation, Boruta feature selection).	164
Figure 60: ROC curves with AUC values for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for colorectal dataset (SNV normalisation, Boruta feature selection).	165
Figure 61: ROC curves with AUC values for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for colorectal dataset (SNV normalisation, Boruta feature selection).	166
Figure 62: (a) – (c) Real-time tracking at each optical biopsy site coupled with binary classification of each site as either normal (100% green) or tumour tissue (100% pink) via a graduated colour map.	166
Figure 63: Schematic illustration of the sterilisable in-vivo DRS probe.	171

Figure 64: A gastrectomy specimen prior to fixation in formalin and processing by the pathologist. (The greater omentum is present and attached to the greater curvature on the inferior border of the stomach). The yellow-dyed area corresponds to the suspected tumour area that was sampled ex-vivo. Three measurements are annotated 1 to 3 on the figure. Line (1) is the distance from the proximal resection margin to the proximal border of the yellow-dyed area. Line (2) is the length of the yellow-dyed area in the proximal to distal plane. Line (3) is the 3 cm line drawn.	172
Figure 65: Photograph of (a) the DRS system setup and (b) the DRS probe during in-vivo data acquisition.	175
Figure 66: Mean absorption spectra with standard deviation for (a) in-vivo normal stomach and gastric cancer, and (b) in-vivo normal oesophagus and oesophageal cancer.	178
Figure 67: Illustration of shape-based features extracted for in-vivo (a) stomach and (b) oesophagus datasets. (1) Depicts the secondary peak intensity value and corresponding wavelength in the 480 – 550 nm spectral range. (2) Depicts the primary peak intensity value and corresponding wavelength in the 580 – 700 nm spectral range.	178
Figure 68: Selected features for the in-vivo stomach dataset for (a) SNV normalisation and (b) no normalisation using Boruta feature selection method.	179
Figure 69: Selected features for the in-vivo oesophagus dataset for (a) SNV normalisation and (b) no normalisation using Boruta feature selection method.	179
Figure 70: ROC curves with AUC values for all classifiers using Approach 1 (entire spectrum) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.	185
Figure 71: ROC curves with AUC values for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.	185
Figure 72: ROC curves with AUC values for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.	186
Figure 73: ROC curves with AUC values for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.	186
Figure 74: Illustration of real-time DRS probe marker-less tracking at each optical biopsy site during an oesophagectomy. (a) – (c) Real-time visual correlation of scanned optical biopsy sites on tissue (green dots on each frame) using the DRS probe and marker-less tracking.	187

List of Equations

(Equation 1): $\mu s' = \mu s * 1 - g$	56
(Equation 2): $\mu t = \mu a + \mu s$	57
(Equation 3): $I = I_{oe} - \mu_{eff} Z$	57
(Equation 4): $\mu_{eff} = 1\delta$	57
(Equation 5): $\delta = [3\mu a \mu a + \mu s' 1/2$	57
(Equation 6): $ j - j < T$	75
(Equation 7): $ j - j < a \cdot max_{h,w}$	76
(Equation 8): $\mathbf{Y} = KLT\{\mathbf{X}\}$	87
(Equation 9): $\mathbf{u}_j = 1n_i = 1n\mathbf{X}_{i,j}$	87
(Equation 10): $\mathbf{B} = \mathbf{X} - \mathbf{h}uT$	87
(Equation 11): $\mathbf{C} = 1n - 1\mathbf{B} * \mathbf{B}$	87
(Equation 12): $\mathbf{V} - 1\mathbf{C}\mathbf{V} = \mathbf{D}$	87
(Equation 13): $\mathbf{D}k,l = \lambda k, k = l0, k \neq l$	87
(Equation 14): $angle = arctan_{yeig1xeig1}$	88
(Equation 15): $K = E_{est}E_{est} + E_{mea}$	91
(Equation 16): $x_t = x_t - 1 + Kp - x_t - 1$	91
(Equation 17): $E_{stt} = E_{stt} - 1 + KM - E_{stt} - 1$	92
(Equation 18): $E_t = 1 - KE_t - 1$	92
(Equation 19): $X_{tp} = AX_t - 1 + But + wt$	92
(Equation 20): $X = xyxy$	92
(Equation 21): $x = x_0 + x_t + 12x_t^2$	92
(Equation 22): $d = d_0 + vt + 12at^2$	92
(Equation 23): $AX = 1\Delta T01xx = x + \Delta Txx$	93
(Equation 24): $\mathbf{B}u = \mathbf{12}\Delta T^2\Delta T\mathbf{a} = \mathbf{a12}\Delta T^2\mathbf{a}\Delta T$	93
(Equation 25): $Y_t = CY_{tm} + Z_t$	93
(Equation 26): $P = \sigma x^2 \sigma x x \sigma x x \sigma x^2$	94
(Equation 27): $Pk = APk - 1AT + Q$	94
(Equation 28): $Kk = PkHTHPkHT + R$	94
(Equation 29): $\sigma x \sigma y = i = 1nx - xiy - yiN$	95
(Equation 30): $P = TP + FP$	110
(Equation 31): $R = TP + FN$	110
(Equation 32): $AP = 01P(r)dr$	110
(Equation 33): $mAP = i = 1KAPik$	110
(Equation 34): $\lambda DRS = \lambda RAW - \lambda DN \lambda WST - \lambda DN$	129
(Equation 35): <i>Haemoglobin Saturation Coefficient</i> = $R585nmR557nm$	133

List of Abbreviations (with alphabetical order)

1D: one-dimensional	GBDT: gradient boosting decision tree	PCA: principal component analysis
2D: two-dimensional	GC: gastric cancer	PET: positron emission tomography
3D: three-dimensional	GI: gastrointestinal	PLS-DA: partial least-squares discriminant analysis
AJCC: American Joint Committee on Cancer	GORD: gastro-oesophageal reflux disease	PTD: probe-to-target distance
AP: average precision	GOSS: gradient-based one-side sampling	REIMS: rapid evaporative ionisation mass spectrometry
APER: abdominoperineal excision of rectum	GPU: graphics processing unit	RF: random forest
AR: augmented reality	H&E: haematoxylin and eosin	RGB: red-green-blue
AUC: area under the curve	HbO ₂ : oxygenated haemoglobin	RMSE: root mean squared error
CCRT: concurrent chemoradiotherapy	HGD: high-grade dysplasia	ROC: receiver operator characteristic
CE: collection efficiency	HHb: deoxygenated haemoglobin	ROI: region of interest
CIN: chromosomal instability	HSV: hue-saturation-value	RS: Raman spectroscopy
CNN: convolutional neural network	IC: imprint cytology	SD: standard deviation
CRC: colorectal cancer	IMA: intra-operative margin assessment	SCC: squamous cell carcinoma
CRM: circumferential resection margin	IoU: intersection over union	SDS: source-to-detector distance
CT: computed tomography	IR: infrared	SKCV: stratified k-fold cross-validation
CUDA: compute unified device architecture	KBB: keypoint threshold bounding box	SNV: standard normal variate
DL: deep learning	KF: Kalman filter	SPP: spatial pyramid pooling
DRS: diffuse reflectance spectroscopy	KT: keypoint threshold	SVM: support vector machine
EAC: oesophageal adenocarcinoma	LDA: linear discriminant analysis	TNM: tumour-node-metastasis
EC: oesophageal cancer	LGBM: light gradient boosting machine	UI: user interface
EFB: exclusive feature bundling	LGD: low-grade dysplasia	UK: United Kingdom
ESCC: oesophageal squamous cell carcinoma	mAP: mean average precision	US/ USA: United States of America
EUS: endoscopic ultrasound	MRI: magnetic resonance imaging	VIS: visible
FAD: flavin adenine dinucleotide	MSI: microsatellite instability	VR: virtual reality
FDG: fluorodeoxyglucose	NA: numerical aperture	WHO: World Health Organisation
FoV: field-of-view	NADH: nicotinamide adenine dinucleotide	XGB: extreme gradient boosting
FPS: frames-per-second	NCCN: National Comprehensive Cancer Network	YOLO: you only look once
FS: fluorescence spectroscopy	NIR: near-infrared	βc: beta-carotene
FSA: frozen section analysis	OS: optical spectroscopy	

Acknowledgements

I would like to express my sincere gratitude to the Cancer Research United Kingdom (CRUK) Convergence Science Centre for funding my PhD studies. Without their support, this work would not have been possible. I am truly grateful for the opportunity they have given me to contribute to cancer research.

I am immensely thankful to my supervisors and colleagues at the Hamlyn Centre, especially Prof Daniel Elson, Mr Christopher Peters, Ms Scarlet Nazarian, Dr Maria Leiloglou, and Dr Vadzim Chalau. Their guidance, expertise, assistance, and unwavering support have been invaluable throughout my research journey. I sincerely hope that our collective efforts and contributions will prove beneficial to the wider research community and contribute towards advancements in healthcare.

I am particularly indebted to Prof Daniel Elson for his trust in my abilities, allowing me to make independent research decisions and design my own methodology. His constant encouragement and mentorship during times of uncertainty have been instrumental in my growth as a researcher. I am also grateful to Mr Christopher Peters for his valuable insights and teachings in the field of clinical medicine research.

Lastly, I extend my heartfelt appreciation to my family and friends for their unwavering love, inspiration, and support throughout my journey. Their encouragement and belief in me have been a constant source of strength.

Once again, I am grateful to all those who have contributed to my academic and research endeavours. I sincerely hope that our collective efforts will have a positive impact on the broader research community and ultimately contribute to advancements in healthcare.

Contribution

1. Peer-reviewed articles

- a) **Ioannis Gkouzionis**, Scarlet Nazarian, Michal Kawka, Ara W. Darzi, Nisha Patel, Christopher J. Peters, Daniel S. Elson, "Real-time tracking of a diffuse reflectance spectroscopy probe used to aid histological validation of margin assessment in upper gastrointestinal cancer resection surgery," J. Biomed. Opt. 27(2) 025001 (1 February 2022) <https://doi.org/10.1117/1.JBO.27.2.025001>
- b) Nazarian S*, **Gkouzionis I***, Kawka M, et al. Real-time Tracking and Classification of Tumor and Nontumor Tissue in Upper Gastrointestinal Cancers Using Diffuse Reflectance Spectroscopy for Resection Margin Assessment. JAMA Surg. 2022;157(11):e223899. doi:10.1001/jamasurg.2022.3899
- c) **Gkouzionis, I.**, Zhong, Y., Nazarian, S. et al. A YOLOv5-based network for the detection of a diffuse reflectance spectroscopy probe to aid surgical guidance in gastrointestinal cancer surgery. Int J CARS (2023). <https://doi.org/10.1007/s11548-023-02944-9>
- d) Nazarian, Scarlet*; **Gkouzionis, Ioannis***; Murphy, Jamie; Darzi, Ara; Patel, Nisha; Peters, Christopher J.; Elson, Daniel S. Real-time classification of tumour and non-tumour tissue in colorectal cancer using diffuse reflectance spectroscopy and neural networks to aid margin assessment. International Journal of Surgery, 10.1097/JS9.0000000000001102, January 22, 2024.

2. Conference papers

- a) **Ioannis Gkouzionis**, Fernando Avila-Rencoret, Jamie Murphy, Christopher Peters, Daniel Elson, "Real-time optical tracking of a diffuse reflectance spectroscopy probe for gastrointestinal tissue analysis," Proc. SPIE 11631, Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XIX, 1163108 (5 March 2021); <https://doi.org/10.1117/12.2575889>

* Co-first authors

- b) Scarlet Nazarian, **Ioannis Gkouzionis**, Arun Anandakumar, Nisha Patel, Daniel Elson, Christopher Peters, SP7.1.6, "Using Diffuse Reflectance Spectroscopy (DRS) to Identify Tumour and Non-tumour Tissue in Upper Gastrointestinal Specimens", British Journal of Surgery, Volume 108, Issue Supplement_7, October 2021, znab361.142, <https://doi.org/10.1093/bjs/znab361.142>
- c) **I. Gkouzionis**, S. Nazarian, A. Anandakumar, A. Darzi, N. Patel, C. Peters, and D. S. Elson, "Using Diffuse Reflectance Spectroscopy Probe Tracking to Identify Non-tumour and Tumour Tissue in Upper Gastrointestinal Specimens," in European Conferences on Biomedical Optics 2021 (ECBO), OSA Technical Digest (Optica Publishing Group, 2021), paper ETu1A.1. Available [here](#).
- d) S. Nazarian, **I. Gkouzionis**, M. Kawka, A. Darzi, N. Patel, D.S. Elson, C.J. Peters, "The use of a diffuse reflectance spectroscopy probe and tracking system to classify tumour and non-tumour tissue in upper gastrointestinal cancer specimens to aid margin assessment", London Surgical Symposium, September 2021.
- e) Scarlet Nazarian, **Ioannis Gkouzionis**, Michal Kawka, Nisha Patel, Ara Darzi, Daniel Elson, Christopher Peters, O-OGC03 Real-time tracking and classification of tumour and non-tumour tissue in upper gastrointestinal cancer specimens using diffuse reflectance spectroscopy, British Journal of Surgery, Volume 108, Issue Supplement_9, 15 December 2021, znab429.035, <https://doi.org/10.1093/bjs/znab429.035>
- f) **I. Gkouzionis**, M. Kawka, S. Nazarian, C.J. Peters, D.S. Elson, "Real-time tracking and classification of tumour and non-tumour tissue in upper gastrointestinal cancer specimens using diffuse reflectance spectroscopy", Joint ESDE and IGCAEuC Congress, November 2021.
- g) **Ioannis Gkouzionis**, Scarlet Nazarian, Michal Kawka, Ara Darzi, Nisha Patel, Christopher Peters, Daniel Elson, "Real-time tissue classification in stomach and oesophageal cancer based on optical tracking of a diffuse reflectance spectroscopy probe," Proc. SPIE PC11949, Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XX, PC119490B (9 March 2022); <https://doi.org/10.1117/12.2608431>

- h) **Ioannis Gkouzionis**, Scarlet Nazarian, Ara Darzi, Nisha Patel, Christopher Peters, Daniel Elson, "Three-dimensional tissue reconstruction and tracking of a diffuse reflectance spectroscopy probe for real-time tissue classification in upper gastrointestinal cancer surgery," Proc. SPIE PC12146, Clinical Biophotonics II, PC121460D (24 May 2022); <https://doi.org/10.1117/12.2620518>
- i) **I. Gkouzionis**, S. Nazarian, N. Patel, C. Peters, and D. S. Elson, "Towards real-time upper gastrointestinal resection margin assessment using a diffuse reflectance spectroscopy probe," in Biophotonics Congress: Biomedical Optics 2022 (Translational, Microscopy, OCT, OTS, BRAIN), Technical Digest Series (Optica Publishing Group, 2022), paper TS2B.3. Available [here](#).
- j) Huang, B., Zheng, J.Q., Nguyen, A., Xu, C., **Gkouzionis, I.**, Vyas, K., Tuch, D., Giannarou, S., Elson, D.S. (2022). Self-supervised Depth Estimation in Laparoscopic Image Using 3D Geometric Consistency. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science, vol 13437. Springer, Cham. https://doi.org/10.1007/978-3-031-16449-1_2
- k) I. Gkouzionis, S. Nazarian, A. Darzi, N. Patel, C.J. Peters, D.S. Elson, "Using diffuse reflectance spectroscopy for real-time tissue assessment during upper gastrointestinal cancer surgery", IEEE BHI-BSN 2022, Ioannina, Greece.
- l) **I. Gkouzionis**, S. Nazarian, A. Darzi, N. Patel, D.S. Elson, C.J. Peters, "Discriminating between cancer and healthy tissue in upper gastrointestinal cancer surgery using deep learning and diffuse reflectance spectroscopy", 17th London Surgical Symposium, 10 November 2022, London, UK.
- m) **Ioannis Gkouzionis**, Scarlet Nazarian, Nisha Patel, Ara Darzi, Christopher J. Peters, Daniel S. Elson, "Diffuse reflectance spectroscopy for in-vivo stomach and oesophageal tissue classification during upper gastrointestinal cancer surgery," Proc. SPIE PC12368, Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XXI, PC123680M (6 March 2023); <https://doi.org/10.1117/12.2649625>
- n) S. Nazarian, **I. Gkouzionis**, A. Darzi, N. Patel, D.S. Elson, C.J. Peters, "Real-time Classification of Tumour and Non-Tumour Colorectal Tissue using Diffuse

Reflectance Spectroscopy to Aid Resection Margin Assessment”, Digestive Disease Week, May 6-9 2023, Chicago, IL.

- o) Y. Zhong, **I. Gkouzionis**, S. Nazarian, A. Darzi, N. Patel, C.J. Peters, D.S. Elson, “A YOLOv5 based network for the detection of a diffuse reflectance spectroscopy probe to aid surgical guidance in gastrointestinal cancer surgery”, The 14th International Conference on Information Processing in Computer-Assisted Interventions, 20-21 June 2023, Munich, Germany.
- p) **I. Gkouzionis**, S. Nazarian, A. Darzi, N. Patel, C.J. Peters, D.S. Elson, “Diffuse reflectance spectroscopy for tissue discrimination during in vivo upper gastrointestinal cancer surgery”, European Conferences on Biomedical Optics (ECBO), 25-29 June 2023, Munich, Germany.
- q) **I. Gkouzionis**, S. Nazarian, M. Giot, A. Darzi, N. Patel, C.J. Peters, D.S. Elson, “In-Vivo Classification of Stomach and Oesophageal Tissues During Upper Gastrointestinal Cancer Surgery Using Diffuse Reflectance Spectroscopy”, SPIE Photonics Europe 2024, 7-11 April 2024, Strasbourg, France.

Chapter 1: Introduction

This dissertation focuses on the critical issue of evaluating circumferential resection margins in gastrointestinal cancer surgery. This first chapter presents an overview of the clinical challenge involved in accurately assessing these margins, highlighting the significant rates of margin positivity and the subsequent impact on both patients and healthcare systems. The methods that are currently used in clinical practice to improve tumour resection are discussed along with their advantages and limitations. Furthermore, the rationale behind undertaking this research is outlined. Finally, an overview of the entire thesis is provided in Section 1.3.

1.1 Introduction & Motivation

Although modern surgical advancements have improved surgical oncology, adequate tumour visualisation remains a limitation preventing total removal of cancerous tissue. In cancer surgery, complete removal of the tumour is pathologically assessed by determining the circumferential resection margin (CRM), the shortest distance between tumour and resected surface, measured over the entire circumferential resection plane.

Surgeons rely primarily on white light reflectance, which has limited differentiation between healthy tissue and tumour and can lead to residual cancer cells inadvertently being left behind at the resection border. Defined as positive circumferential resection margins, these remaining cells generally have poor prognostic implications across different tumour types and typically require adjuvant treatments that confer significant increased costs and burden to the patient and healthcare system (National Comprehensive Cancer Network, 2017).

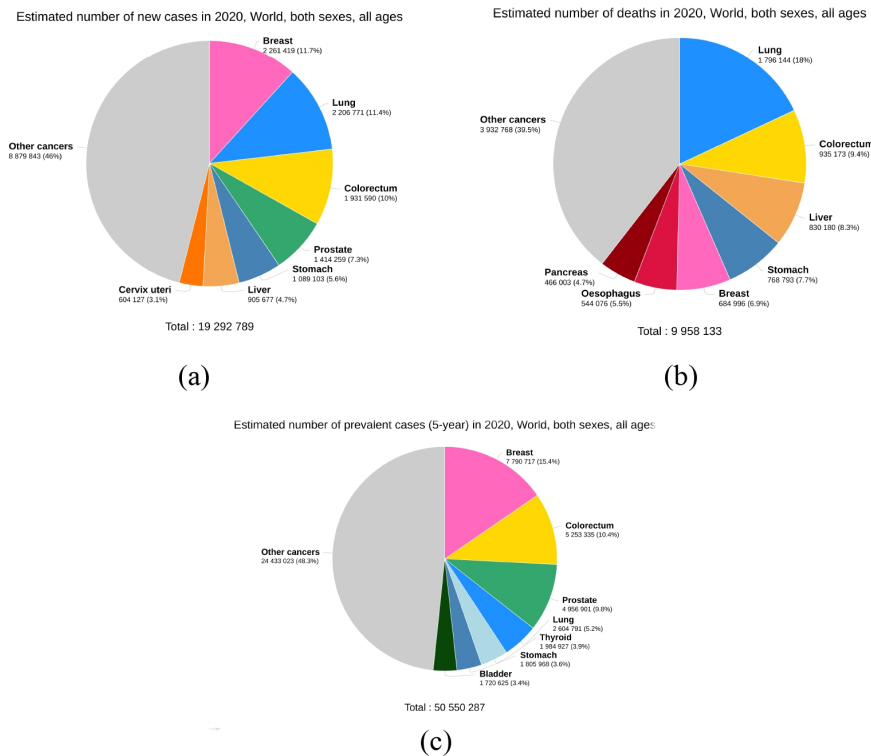


Figure 1: (a) Estimated number of new cancer cases in 2020 (world). (b) Estimated number of cancer-related deaths in 2020 (world). (c) Estimated number of prevalent cancer cases (5-year) in 2020 (world) [Source: GLOBOCAN 2020, <https://gco.iarc.fr/today/home>].

Advances in optical spectroscopy and imaging offer pseudo-colour distinction between tumour and adjacent healthy tissue to improve accuracy of resection and potentially reduce positive CRMs. Improved real-time, intraoperative visualisation can decrease the rate of positive CRMs across tumour types to facilitate cancer treatment while decreasing adjuvant therapy costs and improving prognosis.

1.2 Thesis Objectives

Diffuse reflectance spectroscopy (DRS) is a technique that involves acquiring a full spectrum of light for each scanned pixel of an image. Unlike traditional imaging methods that capture colour or grayscale intensity based on a limited number of wavelength bands, hyperspectral DRS captures a much wider range of wavelengths.

By acquiring a full spectrum for each pixel, hyperspectral DRS allows for a more detailed and comprehensive characterization of the optical properties of the tissue or material being analysed. Each spectrum represents the reflectance or absorption of light at different wavelengths, providing information about the molecular composition, structure, and optical properties of the sample.

With hyperspectral DRS, these spectra can be processed and analysed to create an image where each pixel is represented by its spectral characteristics, typically displayed as a colour or grayscale intensity. This allows for the visualization of spatial variations in optical properties or molecular composition within the imaged area. By mapping the spectral information onto an image, hyperspectral DRS enables the identification and differentiation of different tissue types, detection of abnormalities, or characterization of specific features within the sample.

Diffuse reflectance spectroscopy (DRS) has the potential to play a valuable role in surgical procedures for gastrointestinal cancer by providing margin maps to help surgeons effectively delineate tumours. By scanning large areas of tissue using hyperspectral DRS, detailed spectral information can be obtained, allowing for the characterization of different tissue types and the identification of tumour margins. This margin map can guide surgeons in real-time during the procedure, highlighting areas of potential tumour involvement or distinguishing tumour margins from healthy tissue. By using this map as a visual guide, surgeons can more effectively remove the tumour while minimizing the risk of leaving cancerous tissue behind.

The use of hyperspectral DRS in imaging applications provides researchers and clinicians with a powerful tool for non-invasive characterization and analysis of biological tissues, monitoring disease progression, and guiding diagnostic and treatment decisions. It offers a more comprehensive and detailed understanding of tissue properties compared to conventional imaging methods.

Aims:

- Investigate the translation of DRS technology from ex-vivo to in-vivo application.
- Investigate the ability of DRS technology to assess intraoperative CRMs by manually scanning GI tissue with a hyperspectral optical fibre probe for automated tissue classification.

Hypothesis: DRS technology can assess intraoperative CRMs by distinguishing healthy and tumourous gastrointestinal (GI) tissue.

Objectives:

- Demonstrate the feasibility of a low cost, compact optical spectroscopy system for quantitative sensing of tissue optical properties in the visible and near-infrared spectral range in order to aid resection margin assessment.

- Acquire DRS data from fresh ex-vivo and in-vivo GI specimens.
- Design and develop an image-based framework for the detection and tracking of the DRS fibre probe in order to deal with the non-invasive nature of DRS of not leaving any marks on the tissue during sampling and aid histopathological correlation.
- Build a spectral database from oesophageal, gastric, and colorectal tissue.
- Design, develop, and clinically validate a machine learning pipeline for spectral data processing, feature engineering, and classification between healthy and tumour tissue.

The proposed DRS system alongside the detection and tracking module of the DRS probe as well as the machine learning pipeline were validated through two clinical pilot studies, ex-vivo and in-vivo.

1.3 Thesis Outline

The research described in this thesis is focused on the use of DRS in oesophagogastric and colorectal cancer surgery to develop a technique that can be used real-time during surgery to distinguish healthy tissue from tumour tissue. The ultimate goal is to incorporate DRS into a surgical device. In this way, an enhanced augmented view of the surgical scene will be available to the surgeon to aid complete tumour resection. The remainder of this document is structured as follows.

Chapter 2 delves into the clinical and technical foundations, providing a comprehensive understanding of the clinical problem and challenges, as well as the potential of optical spectroscopy techniques in tissue differentiation.

Chapter 3 of the thesis focuses on two image-based techniques for detecting and tracking the DRS probe. These techniques are designed to enhance the accuracy and reliability of DRS measurements during surgery. The chapter presents the details of these techniques and provides the corresponding results obtained from the experiments conducted.

In **Chapter 4** the research is carried out in a controlled setting using entire ex-vivo tissue specimens from oesophagogastric and colorectal cancer patients. The DRS measurements are performed on these tissue specimens after surgical resection. By using whole tissue specimens instead of small tissue samples, this chapter takes a step closer to simulating the conditions during actual surgery. The chapter focuses on the acquisition of DRS

spectra from the tissue specimens and analyses the obtained spectra to classify the tissue as healthy or tumorous.

Chapter 5 presents an in-vivo study involving the use of DRS during oesophagogastric cancer surgery. Real-time measurements are taken using DRS on patients undergoing surgery for oesophagogastric cancer. Similar to Chapter 4, the acquired DRS spectra are analysed and correlated with the pathology results to classify the tissue. This chapter also includes a comparison between the ex-vivo and in-vivo measurements to evaluate the performance and reliability of DRS in both settings.

In **Chapter 6**, a comprehensive conclusion is provided, summarizing all the results obtained throughout the thesis. The chapter explores the implications of the findings and their potential clinical significance. It also includes a future perspective, discussing the possible directions for further research and the potential applications of DRS in oesophagogastric and colorectal cancer surgery.

Overall, the thesis aims to develop a technique using DRS that can be used in real-time during surgery to distinguish healthy tissue from tumorous tissue. The outlined chapters progressively investigate and validate the application of DRS in different settings, from ex-vivo tissue specimens to in-vivo surgeries, with the goal of advancing towards the incorporation of DRS into a surgical device.

Chapter 2: Understanding the Clinical and Technical Background

2.1 Oesophagogastric Cancer

The escalating incidence and unfavourable prognosis associated with oesophageal cancer necessitate the development of enhanced detection and prediction methods crucial for surgical assessment. Oesophageal cancer ranks among the most lethal malignancies globally, with a notable surge in its occurrence witnessed in Western nations over recent decades. Despite advancements in the management and treatment of oesophageal cancer patients, the overall prognosis remains exceedingly poor, with 5-year survival rates typically below 10% and 5-year post-esophagectomy survival rates ranging between 15% and 40%. The late-stage diagnosis of oesophageal cancer predominantly stems from the absence of early clinical symptoms. To improve patient outcomes following surgery, neoadjuvant concurrent chemoradiotherapy (CCRT) or chemotherapy is often administered to reduce tumour size. However, CCRT can intensify toxicity levels and potentially result in surgical delays for patients exhibiting poor response to this treatment approach. Consequently, the identification of precise biomarkers capable of predicting or discerning patients likely to respond favourably or unfavourably to CCRT can aid physicians in selecting appropriate therapeutic strategies. The identification of vulnerable genes and biomarkers holds promise for predicting treatment response and enhancing survival rates among oesophageal cancer patients.

2.1.1 Epidemiology

Globally, gastric cancer (GC) and oesophageal cancer are amongst the most commonly diagnosed malignancies worldwide^{1,2}. In 2020, gastric cancer was one of the most commonly diagnosed cancers, with over 1 million reported new cases and approximately 770,000 deaths³. In 2020, oesophageal carcinoma was responsible for about 604,000 new cases and 544,000 deaths worldwide. The combined annual incidence of gastric and oesophageal cancer makes them the second most commonly diagnosed cancers worldwide⁴. Due to their close anatomical proximity, both oesophageal cancer and gastric cancer share several risk factors

and epidemiological characteristics. However, they also exhibit distinct geographical and temporal patterns in terms of their incidence rates. Variations in the distribution of tumour subtypes among different populations have been identified as a contributing factor to the differences observed in the burden of GC and oesophageal cancer. These variations help explain the contrasting impact of these cancers in different regions or populations^{5,6}.

The majority of oesophageal cancers are either adenocarcinomas, where the cells grow in a glandular pattern and produce mucin, or squamous cell carcinomas (SCC) with intercellular bridges and keratin production^{7–9}. SCC is more common worldwide, but the incidence of adenocarcinomas is rising in the West⁷. Oesophageal adenocarcinomas are generally late-presenting, resulting in an overall 5-year survival of less than 25% in most cases^{9,10}. In contrast, when detected early (carcinoma limited to mucosa or submucosa), the 5-year survival is closer to 80%⁷.

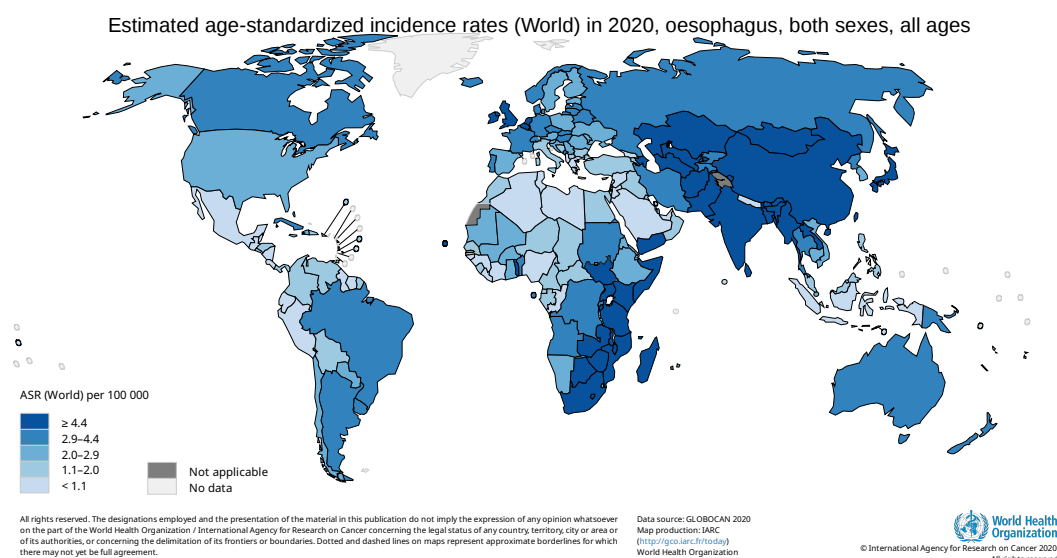


Figure 2: Estimated age-standardised incidence rates (world) in 2020, oesophageal cancer, both sexes, all ages [Source: GLOBOCAN 2020, <https://gco.iarc.fr/today/home>].

Gastric cancer is one of the leading causes of death worldwide^{7,9,11,12}. In the UK, gastric cancer is the fourth commonest cause of death from malignancy¹², with the worldwide 5-year overall survival rate being 10–30% and 35–55% after surgery^{9,11}.

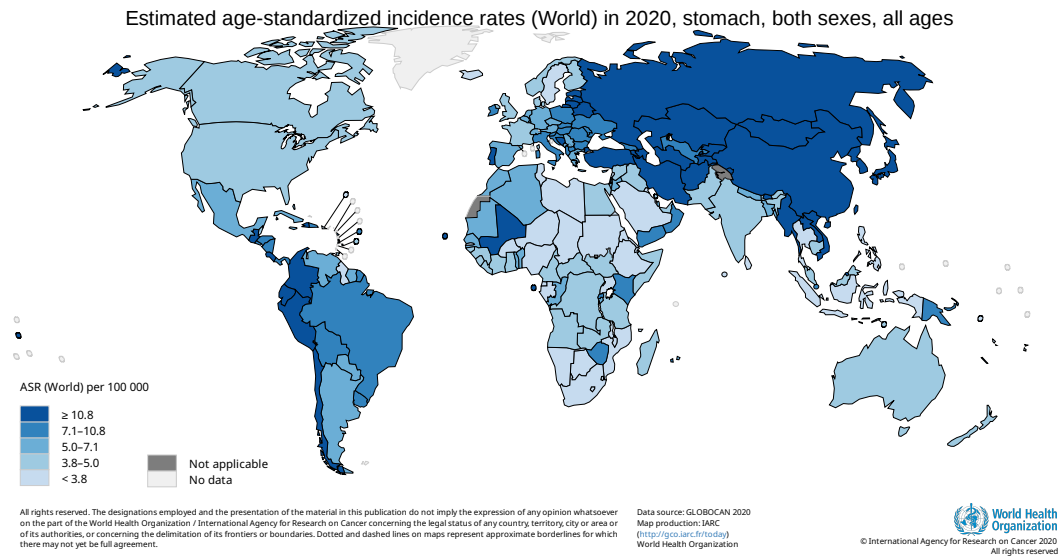


Figure 3: Estimated age-standardised incidence rates (world) in 2020, gastric cancer, both sexes, all ages [Source: GLOBOCAN 2020, <https://gco.iarc.fr/today/home>].

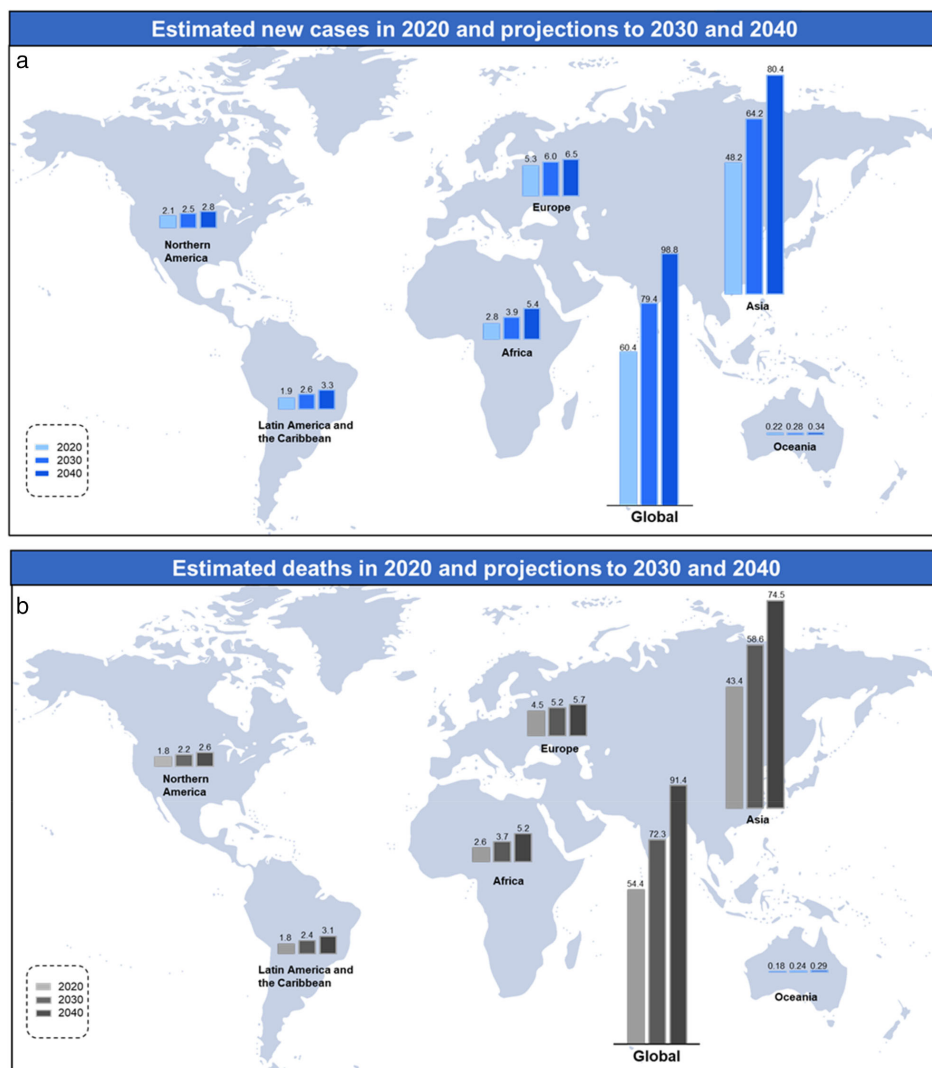


Figure 4: Estimated new cases (a) and deaths (b) of oesophageal cancer worldwide in 2020 and projections to 2030 and 2040. Numbers are shown in the scale of $\times 10,000$ [Source: GLOBOCAN 2020, <https://gco.iarc.fr/today/home>].

The global prevalence of oesophageal cancer has shown a significant increase in recent years. According to projections based on [GLOBOCAN 2020](#) data, if current trends continue, the number of new cases and deaths from oesophageal cancer is expected to rise in 2030 and 2040. In 2030, it is estimated that there will be 730,000 new cases and 720,000 deaths from oesophageal cancer, while in 2040, these numbers are projected to reach 980,000 new cases and 900,000 deaths. This represents an increase of 31% and 33% in new cases and deaths, respectively, compared to 2020.

The highest change in the number of new cases is expected in Africa, with a rise from 27,546 cases in 2020 to 38,656 cases (an increase of 40.3%) in 2030, and further to 54,279 cases (an increase of 97.9%) in 2040. Europe, on the other hand, is projected to have the lowest increase, with cases rising from 52,993 in 2020 to 59,626 cases (an increase of 12.5%) in 2030, and 64,720 cases (an increase of 22.1%) in 2040. Asia will continue to have the highest number of new cases, with an estimated increase from 481,552 cases in 2020 to 642,148 cases (an increase of 33.3%) in 2030, and further to 803,911 cases (an increase of 66.9%) in 2040. Similar patterns can be observed for projected oesophageal cancer deaths in 2030 and 2040 across six continents, showing an increase in line with the number of new cases. These projections highlight the growing burden of oesophageal cancer globally and emphasize the need for effective prevention, early detection, and improved treatment strategies to mitigate its impact.

2.1.2 Pathophysiology

Oesophageal adenocarcinoma is a type of cancer that typically occurs in patients with obesity and chronic gastro-oesophageal reflux disease (GORD)¹³. These factors are related to the occurrence of Barrett's oesophagus, a condition in which the squamous epithelium is replaced by columnar epithelium as a response to acid and bile reflux. Barrett's mucosa is thought to be a precursor of oesophageal adenocarcinoma¹³. Risk factors for oesophageal squamous cell carcinoma are smoking, alcohol consumption, nutritional deficiencies and possibly other dietary components¹⁴. The presence of dysplasia further increases the risk of cancer development⁷⁻⁹. Tumours may start as flat patches which can grow to large masses of 5 cm or more in diameter, or infiltrate diffusely or ulcerate and invade deeply. They most

commonly form glands and produce mucin, often exhibiting intestinal-type morphology. Less frequently, diffusely infiltrative signet-ring cells can be seen.

Squamous cell carcinoma of the oesophagus occur in the middle third of the oesophagus in approximately 50% of the cases. It typically originates as squamous dysplasia, characterised by a small, grey-white, plaque-like thickening. Over time, the tumour undergoes progressive growth, forming a polypoid or exophytic mass that may obstruct the oesophageal lumen and lead to dysplasia⁷. It can also spread in an ulcerative or diffusely infiltrative manner, resulting in thickening, rigidity, and luminal narrowing of the oesophageal wall. In advanced stages, SCC can invade neighbouring structures such as the mediastinum, lungs, heart, and major blood vessels like the aorta.

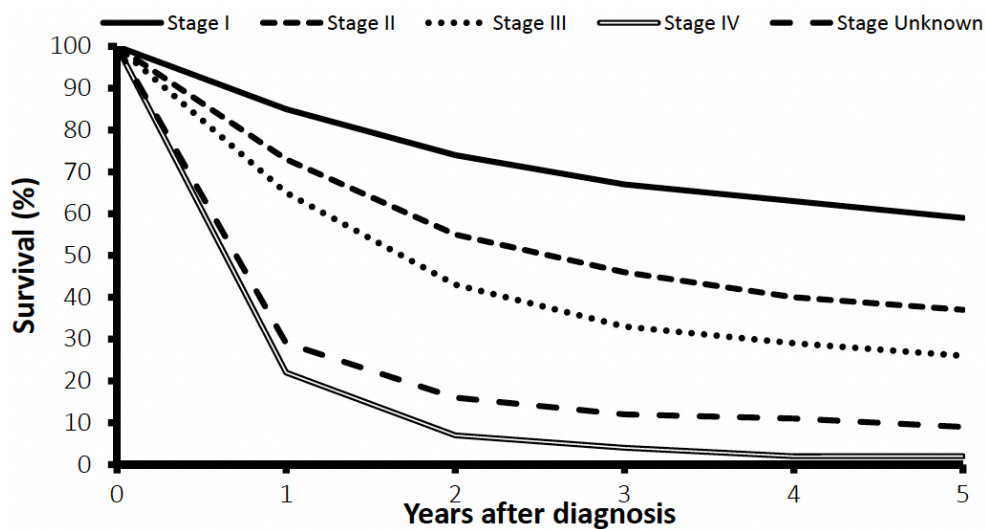


Figure 5: Survival of oesophageal cancer patients per TNM (7th edition) stage at diagnosis [Source: <http://www.cijfersoverkanker.nl>].

Over 90% of gastric cancers are adenocarcinomas, which are classified according to their location, macroscopic appearance, and histology. According to the World Health Organisation (WHO) and Lauren's classification system, the two distinct histological types of gastric cancer are intestinal and diffuse^{7,8,11}. Intestinal-type adenocarcinomas tend to form bulky tumours composed of glandular structures, while diffuse infiltrative tumours are composed of discohesive cells with no visible glands and tend to infiltrate the stomach wall instead of forming discrete masses. There are a few risk factors contributing to a high incidence of gastric cancer. *Helicobacter Pylori* is a gastric bacteria infecting over 50% of the world's population, and is the strongest risk factor for gastric non-cardia cancer¹⁵.

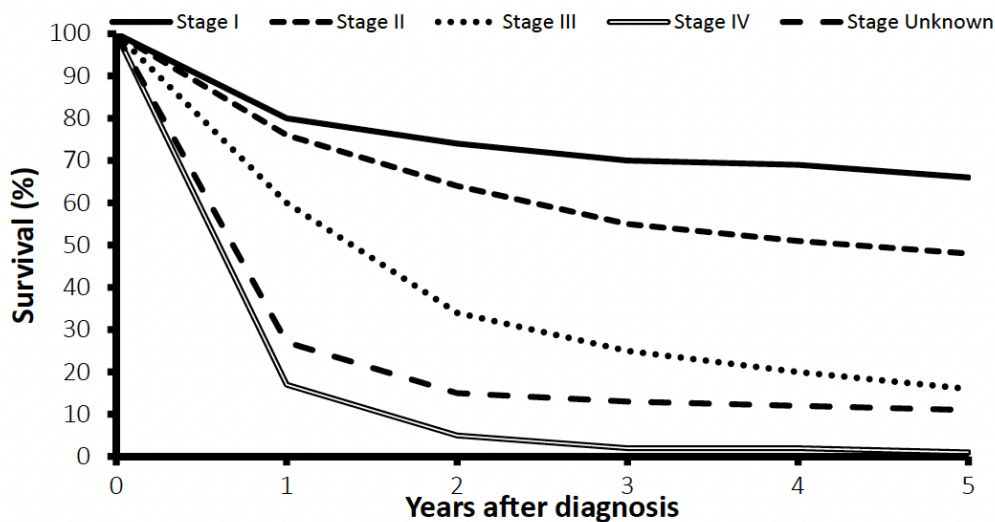


Figure 6: Survival of gastric cancer patients per TNM stage at diagnosis (TNM 7th edition) [Source: <http://www.cijfersoverkanker.nl>].

2.1.3 Staging System & Definitions

T Stage of Oesophagogastric Cancer

The staging of oesophageal cancer is determined by the American Joint Committee on Cancer (AJCC) Staging System, which utilizes tumour-node-metastasis (TNM) sub-classifications based on the depth of tumour invasion (T), lymph node involvement (N), and presence of metastatic disease (M). When using computed tomography (CT) imaging to assess the oesophagus, a key starting point is evaluating the thickness of the oesophageal wall. A wall thickness greater than 5 mm is considered abnormally thick, as the distended wall of the oesophagus is typically less than 3 mm^{16,17}. However, it is important to note that oesophageal wall thickness asymmetry, although a classic finding, is nonspecific and should be interpreted alongside the consideration of wall thickness symmetry. Compared to other assessment modalities like endoscopic ultrasound (EUS), CT has been shown to be less accurate¹⁸.

CT imaging of the oesophagus is limited in accurately differentiating between T1, T2, and T3 stages of primary tumour invasion, which is crucial for determining stage-specific treatment protocols. The most valuable aspect of CT imaging in assessing T status is evaluating whether the primary tumour invades adjacent structures. If the fat planes between the primary tumour and adjacent structures are obliterated on CT, it indicates a T4 stage cancer. CT has a sensitivity and specificity ranging between 85% to 100% for detecting mediastinal invasion^{19,20}.

However, it should be noted that obliteration of fat planes on CT, while reliable in establishing a T4 stage tumour, can also occur in patients with prior radiation therapy or cachexia. EUS is now considered the most accurate imaging modality for determining T staging of oesophageal cancer. EUS outperforms CT in differentiating between T1, T2, and T3 tumours, with EUS accurately determining preoperative T stages in 76% to 89% of cases compared to 49% to 59% with CT imaging^{21–24}.

Table 1: T, N, and M status and histologic grade definitions for oesophagus and oesophagogastric junction cancer in the 8th edition of the American Joint Committee on Cancer (AJCC) Cancer Staging Manual.

T status – primary tumour	
T _x	Primary tumours cannot be evaluated
T ₀	No evidence of primary tumour
T _{is}	High-grade dysplasia
T ₁	Invasion into the lamina propria, muscularis mucosae, or submucosa
T ₂	Invasion into muscularis propria
T ₃	Invasion into adventitia
T _{4a}	Invades resectable adjacent structures (pleura, pericardium, diaphragm)
T _{4b}	Invades unresectable adjacent structures (aorta, vertebral body, trachea)
N status – regional lymph nodes (adenocarcinoma, squamous cell carcinoma)	
N _x	Regional lymph nodes cannot be assessed
N ₀	No regional lymph node metastases
N ₁	1 to 2 positive regional lymph nodes
N ₂	3 to 6 positive regional lymph nodes
N ₃	7 or more positive regional lymph nodes
M status – distal metastasis (adenocarcinoma, squamous cell carcinoma)	
M ₀	No distant metastases
M ₁	Distant metastases
Histologic grade – degree of differentiation (adenocarcinoma, squamous cell carcinoma)	
G ₁	Well differentiated
G ₂	Moderately differentiated
G ₃	Poorly differentiated
G ₄	Undifferentiated

A study by Rösch demonstrated that EUS correctly staged oesophageal cancer 84% of the time, with accuracy increasing as the T stage of the primary tumour increased²⁵. Accuracy ranged from 75% to 82% for T1 disease and 88% to 100% for T4 disease²⁶. EUS is a valuable tool for assessing the extent of disease and monitoring response to chemotherapy, particularly when analysing tumour dimensions. However, EUS is unreliable for staging oesophageal cancer after neoadjuvant chemoradiation²⁷. Some limitations of EUS include operator dependence for accurate interpretation and difficulty visualizing the entire tumour when the oesophageal lumen is narrowed due to strictures or stenosis²⁸.

N Stage of Oesophagogastric Cancer

In oesophageal cancer, N staging is determined by the involvement (N1, N2, N3) or absence of involvement (N0) of periesophageal lymph nodes. The sensitivity and specificity of CT scans in detecting periesophageal lymph node involvement depend on the size of the lymph nodes. Typically, a size criterion of 1 cm is used to define an enlarged lymph node. The reported sensitivity ranges from 30% to 60%, while specificity ranges from 60% to 80%²⁹. However, a limitation of CT imaging is its inability to detect nodal involvement in cases where a normal-sized lymph node may contain metastatic foci without a noticeable increase in size. Additionally, an enlarged lymph node does not always indicate metastasis, as benign enlargement and inflammation can also occur. The accuracy of CT imaging in detecting N stage has been reported as 46% to 58%²⁴.

Endoscopic ultrasound has shown greater accuracy in determining nodal involvement in oesophageal cancer, with an accuracy ranging from 72% to 80%³⁰. The combination of EUS with guided fine-needle aspiration has further improved the accuracy of nodal staging by allowing for sampling and evaluation of lymph nodes for metastasis.

Fluorodeoxyglucose-positron emission tomography (FDG-PET) has also been used to assess nodal involvement in oesophageal cancer. Evaluating the uptake of FDG in local and regional lymph nodes is challenging due to the intense uptake of FDG by the primary oesophageal tumour. However, FDG-PET is useful in detecting distant metastases, including metastasis to abdominal and cervical lymph nodes. Sensitivities as high as 90% have been reported for detecting distant lymph node metastasis using FDG-PET³¹.

Overall, while CT imaging has limitations in accurately detecting nodal involvement, EUS and FDG-PET offer improved sensitivity and specificity in assessing nodal status in oesophageal cancer.

M Stage of Oesophagogastric Cancer

Oesophageal cancer is known for its aggressive and invasive nature, with 20% to 30% of patients already having distant metastasis at the time of initial diagnosis³². The presence or absence of distant metastasis plays a crucial role in guiding treatment options and determining operability. Common sites of distant metastasis in oesophageal cancer include the liver, lungs, and bones²⁸.

In the AJCC classification system, distant metastasis is divided into M_{1a} and M_{1b} classifications, each of which is significant in determining appropriate treatment approaches. M_{1a} includes metastasis to the celiac and cervical lymph node groups and is associated with a relatively better prognosis compared to M_{1b}. Patients classified as M_{1a} often undergo neoadjuvant therapy followed by surgical resection. On the other hand, patients with M_{1b} classification have distant site metastasis, indicating a worse prognosis as surgical resection with curative intent is not recommended in these cases³³.

CT imaging is commonly used to assess distant metastasis in oesophageal cancer patients. Contrast-enhanced CT scans can quickly evaluate common areas of distant metastasis. Spiral CT has reported sensitivities as high as 90% for detecting masses ≥ 1 cm³⁴.

Endoscopic ultrasound has limitations in assessing distant metastasis and is generally not preferred for M staging of oesophageal cancer. Instead, CT or FDG-PET is favoured. FDG-PET plays a significant role in detecting distant metastasis in oesophageal cancer staging. Compared to CT, PET has been shown to be more accurate in identifying distant metastasis³⁵. One study demonstrated that PET detected distant metastasis in 15% of patients who were initially believed to have only primary oesophageal cancer according to other imaging modalities³⁶. The presence of distant metastasis places the patient in the M_{1b} category, and surgery with curative intent is no longer recommended.

2.1.4 Treatment

The survival rate for patients with oesophageal cancer is poor. Surgical treatment of resectable oesophageal cancer results in 5-year survival rates of 5% to 30%, with higher survival rates in patients with early-stage cancer³⁷. In early-stage disease, more emphasis is placed on tumour resection rather than on systemic chemotherapy³⁸. Patients with early-stage cancer (Stage II or less) undergo surgical resection to remove the malignancy. The preferred surgical approach is selected based on tumour location and invasion depth. Common techniques include endoscopic mucosal resection, distal oesophagectomy, subtotal or total gastrectomy. Although surgery remains the only curative treatment for gastric cancer, the use of chemotherapy, whether administered pre- (neoadjuvant), post- (adjuvant), or peri-operatively, has demonstrated survival benefits. A study involving 206 gastric cancer patients found that those with Stage II and III cancer had higher survival rates when adjuvant therapy was added to surgery compared to surgery alone³⁹.

The National Comprehensive Cancer Network (NCCN) offers treatment guidelines for oesophageal cancer, which take into account factors such as stage, tumour location, and the patient's overall health⁴⁰. Treatment options outlined by the NCCN include local mucosal resection or ablation therapies, oesophagectomy, chemotherapy, and radiation therapy. However, there is a lack of definitive data from randomized trials to guide treatment decisions for many clinical situations in oesophageal cancer. Moreover, outcomes are generally poor with several treatment strategies, leading to ongoing research to establish optimal treatments for different scenarios⁴¹.

Due to the absence of definitive evidence, the NCCN guidelines often allow for a range of potential treatments for various clinical situations. Given the overall poor prognosis and potential treatment-related morbidity, a multidisciplinary evaluation involving surgery, medical oncology, and radiation oncology is recommended for all patients before initiating a treatment strategy. Non-guideline treatments should typically only be considered within the context of clinical trials⁴². Treatment decisions for oesophageal and gastric cancer patients can be simplified based on three main categories: early-stage superficial cancers, locally advanced cancers with locoregional disease but no distant metastases, and cancers with distant disease.

For patients presenting with clinically advanced disease, including T2-4 tumours or the presence of lymph node involvement, the NCCN guidelines advocate the use of either preoperative chemoradiotherapy or perioperative chemotherapy⁴³. It is worth noting that preoperative chemotherapy has seen a rise in use since 2005. Insights from the MAGIC study, which investigated perioperative chemotherapy using the ECF regimen (consisting of epirubicin, cisplatin, and a continuous infusion of 5-fluorouracil), and the SWOG INT-0116 study, which examined the MacDonald regimen of postoperative chemoradiotherapy with 5FU as radiosensitiser, have highlighted the benefits of a combined therapeutic approach in advanced gastric cancer, showing improved overall survival rates compared with surgery alone⁴⁴.

Multimodality neoadjuvant CCRT has witnessed increased utilization in the treatment of oesophageal cancer (EC) over the past years⁴⁵. Recent randomized trials have demonstrated a substantial survival rate benefit when neoadjuvant CCRT is followed by surgery for patients with locally advanced EC⁴⁶. The evaluation of pathological responses to CCRT in clinical specimens is a crucial prognostic indicator associated with survival rate outcomes⁴⁷. The efficacy of chemoradiotherapy as an adjunctive therapy to surgery in patients diagnosed with resectable oesophageal cancer has been validated in terms of overall survival benefits⁴⁸. However, CCRT is currently not recommended as first-line treatment option for metastatic EC⁴⁹.

Most recently, the preferred perioperative regimen became the FLOT (docetaxel, 5-FU, leucovorin, and oxaliplatin) following the positive results from the phase III FLOT4-AIO trial⁵⁰. In clinical practice, the FLOT regimen has emerged as the standard perioperative treatment for patients with resectable gastric and oesophagogastric cancers that are at T2 stage or higher, and/or exhibit positive lymph node involvement. This approach is particularly preferred for those patients demonstrating a good performance status and without competing comorbidities.

Nearly 40% who are diagnosed with gastric cancer will initially present with metastatic disease, with peritoneal metastases being observed in approximately one third of these cases⁵¹. Following gastrectomy, up to 46% of patients may present with peritoneal lesions, indicating a lack of response to treatment⁵². Traditional systemic agents, often used in the adjuvant therapy, have limitations in accurately targeting peritoneal lesions due to the

presence of the peritoneal-blood barrier, prompting the need for more direct therapeutic techniques to address this issue.

Postoperative adjuvant therapy has been found to enhance disease-free survival rates and decrease the duration of hospitalization for non-responder patients with oesophageal squamous cell carcinoma (ESCC)⁵³. A study evaluating the efficacy of chemotherapy in nonsurgical local ESCC patients employed nimotuzumab (200 mg) administered weekly, paclitaxel (175 mg/m²) on day 1, and cisplatin (30 mg/m²) on days 1 and 2, with a treatment cycle repeating every 3 weeks over six cycles. The overall response rate was determined to be 51.8%, while the median duration of disease control and overall survival for all patients was 10.8 months (95% CI, 5.9-15.7 months) and 20.2 months (95% CI, 11.5-28.9 months), respectively⁵⁴.

In the case of oesophageal adenocarcinoma (EAC) patients, neoadjuvant chemotherapy has been extensively employed to treat individuals with locally advanced and/or lymph node positive tumours⁵⁵. A cohort of 205 patients with T2-T4 or lymph node positive EAC received neoadjuvant chemoradiotherapy followed by surgery. Among patients who achieved a pathologic complete response, the 3-year overall survival and recurrence-free survival rates were 86% and 80%, respectively, accounting for 19% of the treated population. This finding emphasizes the specificity of neoadjuvant chemoradiation therapy for oesophageal adenocarcinoma⁵⁶.

Despite the wide range of treatment options available for the management of oesophagogastric cancer, a considerable number of patients experience rapid disease progression. Such progression can be attributed to two key factors: (1) significant inter-tumoural and intra-tumoural heterogeneity, and (2) frequent diagnosis of the disease at an advanced stage, which often results in increased resistance to chemotherapy⁵⁷.

2.2 Colorectal Cancer

2.2.1 Epidemiology

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths in the United States. The American Cancer Society provides periodic updates on CRC statistics based on data from population-based cancer registries for incidence and the National Center for Health Statistics for mortality. In 2023, it is estimated that around 153,020 individuals will be

diagnosed with CRC, and 52,550 individuals will die from the disease. Among those diagnosed, approximately 19,550 cases and 3,750 deaths will occur in individuals younger than 50 years old. The rate of decline in CRC incidence has slowed over the years, dropping by 3%-4% annually during the 2000s but only 1% annually from 2011 to 2019. This slowdown in decline can be attributed, in part, to an increase in the incidence of CRC in individuals younger than 55 years old by 1%-2% annually since the mid-1990s. As a result, the proportion of CRC cases among those younger than 55 years old has increased from 11% in 1995 to 20% in 2019. Since around 2010, there has been an increase in CRC incidence among individuals younger than 65 years old, particularly for regional-stage disease (increasing by about 2%-3% annually) and distant-stage disease (increasing by 0.5%-3% annually). This reversal in the trend of earlier stage diagnosis seen during 1995-2005 has led to a higher proportion of advanced cases, with 60% of new cases being advanced in 2019 compared to 52% in the mid-2000s and 57% in 1995, prior to widespread screening efforts⁵⁸⁻⁶².

Colorectal cancer is the third most common cancer among men and women in the UK⁶³ and second leading cause of cancer-related death in Western countries, with more than 880,000 deaths in 2018^{1,64,65}. In the past decade, the incidence has decreased by 1.4-4% annually⁶⁶. Five-year survival rates are 64-66%⁶⁶. In patients with non-metastatic colorectal cancer, surgery is the cornerstone of curative therapy. It is estimated that by the year 2035, the total number of deaths from rectal and colon cancer will increase by 60% and 71.5%, respectively⁶⁷.

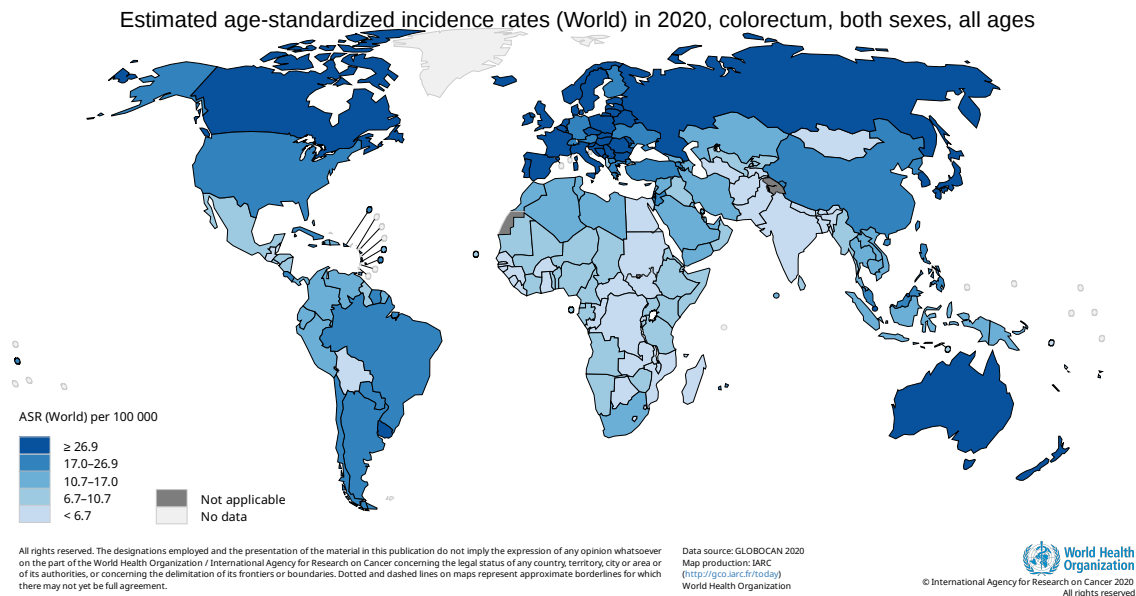


Figure 7: Estimated age-standardised incidence rates (world) in 2020, colorectal cancer, both sexes, all ages [Source: GLOBOCAN 2020, <https://gco.iarc.fr/today/home>].

2.2.2 Pathophysiology

Colorectal cancer development occurs through four stages: initiation, promotion, progression, and metastasis (Figure 8). During the initiation stage, genetic alterations or mutations occur in normal colon cells, leading to the formation of pre-cancerous cells. These mutations can be triggered by various factors such as environmental exposures or inherited genetic predispositions.

In the promotion stage, additional genetic and epigenetic changes accumulate, promoting the growth and survival of pre-cancerous cells. This stage involves the expansion of the mutated cell population and the development of early adenomas or polyps in the colon lining.

As the progression stage begins, further genetic alterations and genomic instability drive the transition from benign adenomas to malignant carcinomas. These changes lead to the invasion of nearby tissues and the potential spread of cancer cells to lymph nodes. Metastasis is the final stage of CRC development, where cancer cells acquire the ability to invade blood vessels or lymphatic vessels and establish secondary tumours in distant organs. The liver is the most common site of metastasis for CRC, followed by the lungs and bones.

The duration required for each stage of CRC development can vary significantly between individuals, and it is challenging to precisely determine the time taken to progress

through each stage. However, it is generally understood that CRC development is a complex and gradual process that may span several decades from the initiation of genetic alterations to the formation of clinically detectable cancer.

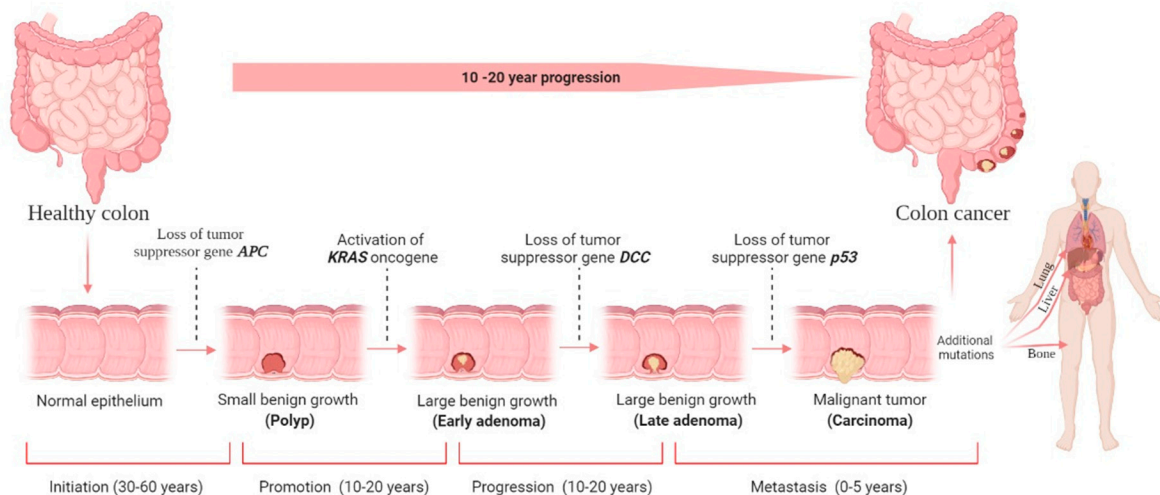


Figure 8: Colorectal cancer stages and development (adopted by Hossain et al.⁶⁸).

Colorectal cancer develops when epithelial cells undergo a series of genetic or epigenetic changes that promote excessive cell proliferation⁶⁹. These rapidly dividing cells give rise to benign adenomas, which can progress to cancer and potentially metastasize through various pathways, including microsatellite instability (MSI), chromosomal instability (CIN), and serrated neoplasia^{70–72}.

The adenoma-carcinoma sequence describes the progression of cancer in the traditional pathway, which accounts for the majority of sporadic CRC cases. It begins with the transformation of normal cells into small adenomas, which then grow into larger adenomas and eventually develop into cancer⁷³. This pathway is strongly associated with the development of the CIN-positive subtype. Approximately 10-15% of sporadic CRC cases follow this pathway.

The adenocarcinoma accounts for up to 80% of colorectal cancers, which affects the entire length of the colon equally, with similar histological properties. However, tumours in the proximal length of the colon tend to grow along one wall as polypoid, exophytic masses, which rarely cause obstruction, whereas in the distal colon lesions tend to be annular and their growth causes constrictions^{7,63,74,75}. Ten to fifteen percent of adenocarcinomas are mucinous and they have an increased tendency to develop peritoneal metastases, invade adjacent organs, and spread extensively to lymph nodes⁷⁴.

The development of colorectal cancer typically involves the progression from non-cancerous tissue to dysplastic tissue formation, eventually leading to the development of CRC as cells undergo multiple abnormal DNA changes. Initially, a non-cancerous soft tissue tumour may form, which is benign and does not spread to other parts of the body (metastasize). This tumour is referred to as a polyp or adenoma (stage 0). Approximately 10% of adenomatous polyps have the potential to become malignant, transforming into an adenocarcinoma that invades the muscularis propria (stage I).

As the disease progresses, the tumour grows in size and further invades surrounding tissues, including the serosa (stage II) and visceral peritoneum (stage III). At this stage, there is a possibility of developing lymphatic or blood vessel metastasis, where cancer cells can spread to distant sites in the body (stage IV)⁷⁶. The stage of the disease determines its severity and helps guide treatment options⁷⁷.

For stages 0 to II CRC, surgery is the standard treatment choice. Stage III CRC requires both surgery and adjuvant chemotherapy. In the case of stage IV or recurrent CRC, treatment typically involves surgery, chemotherapy, and targeted therapy. It is important to note that, cytoreductive surgery (CRS), whether performed as an open or laparoscopic procedure, when combined with hyperthermic intraperitoneal chemotherapy (HIPEC), is being increasingly recognised as a potentially effective therapeutic approach, particularly for younger patients (aged <50 years) diagnosed with locally advanced (pT4) colorectal cancer accompanied by peritoneal metastases. This treatment combination has the potential to increase overall survival rates, improve quality of life, and reduce the chances of tumour recurrence⁷⁸.

While surgery remains a cornerstone of treatment for early-stage CRC, advanced stages of the disease often require a combination of therapies, including chemotherapy and targeted therapy, to manage and control the cancer. Research and advancements in treatment continue to be pursued in the ongoing effort to improve outcomes for patients with CRC.

2.3 Surgical Resection Margins

Surgical excision of the primary tumour remains the gold-standard treatment method for most solid organ tumours⁷⁹. The aim of GI cancer surgery is the complete resection of cancerous region, both on macroscopic and microscopic levels⁷⁹. This requires the affected tissue to be excised with a rim of healthy tissue with an adequate margin. The amount of

margin, or distance from the cancerous tissue to the edge of the specimen, required to be termed “clear”, or negative, is different for each pathology. A positive surgical margin and an inadequate resection occurs when the surgical removal does not achieve the desired thorough removal, leaving residual cancer cells within the defined distance from the edge of the resected tissue surface^{80–83}.

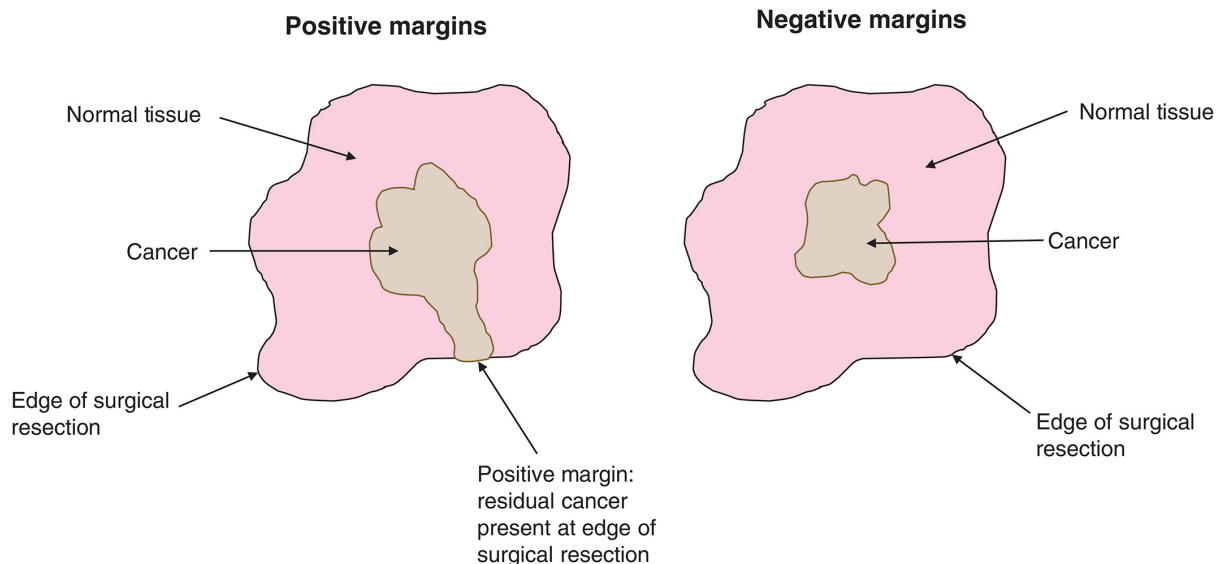


Figure 9: (Left) A surgically excised specimen with an inadequate or positive surgical margin; the cancerous tissue is at the edge of the specimen. (Right) A surgically excised specimen with an adequate margin along the resected surface; the cancerous tissue is in the centre of the surgically excised specimen, with a rim of healthy tissue surrounding it. (adopted by Tringale et al.⁸⁴).

A positive CRM not only affects prognosis, but also future management, meaning the need for adjuvant therapies with significant side effects⁸⁵. The cost of a positive surgical margin to the patient, in terms of increased treatment burden, further anxiety, reduced survival, and additional side effects is difficult to quantify, but the effect on the financial resources of the healthcare system is undoubtedly significant^{86,87}.

Obtaining negative surgical margins is critical. Comprehensive research emphasises the significance of preserving resection margins of 10 cm proximal and 5 cm distal to the visible tumour when the oesophagus is in its natural state. It is commonly noted that adenocarcinomas located in the lower oesophagus frequently expand into areas such as the gastric cardia, fundus, and lesser curve. Therefore, some degree of gastric excision becomes necessary not only for thorough tumour resection but also for abdominal lymphadenectomy. It is imperative to guarantee adequate radial margins, taking into consideration the excision of the diaphragm and crura which is frequently required, especially for junctional tumours. Longer resection lengths are necessary for surgical safety, particularly for primary tumours

that have multicentric SCC lesions. Despite these efforts, achieving negative margins remains challenging. For example, a resection margin of 4 cm, which is occasionally employed based on tumour characteristics and location, has been associated with positive margins in about 40% of cases. Even when a more conservative 10 cm margin is adopted, there is still a 17% chance of finding residual tumour cells at the margins⁸⁸.

The Standards Practice Task Force of the American Society of Colon and Rectal Surgeons recommends, depending on anatomic factors, a 1–5 cm surgical margin (strong recommendation, level 1A evidence)⁶⁵. The National Comprehensive Cancer Network states that for an adequate excision for tumour in the upper rectum, a distal rectal margin of at least 5 cm below the distal edge of the tumour is necessary. For tumours in the distal rectum, a distal margin of 1–2 cm may be acceptable (NCCN, 2017). Despite these best practices, positive CRMs still occur in 6–13% of colorectal cases^{89,90}. Local recurrence rates with a negative margin range from 6 to 13% but are 16–78% with a positive circumferential margin^{91–95}. Although adjuvant radiation is indicated for CRMs, it does not compensate in terms of recurrence^{96,97}.

Tramontano et al. used the Surveillance, Epidemiology, and End Results (SEER) of Medicare database to determine the costs of oesophageal treatment between 1998 and 2013. The results were adjusted for 2018 levels of inflation and stratified according to age, diagnosis stage, histology, and treatment strategy. The highest terminal phase costs were in Stages II (\$20,533) and III (\$20,599) due to high rates of aggressive care, such as hospitalisations and intensive care admissions among cancer patients during the last months of life. Stage IV had the most expensive initial phase cost (\$9,263). Among the staging costs, Stage III emerged as the most expensive with \$9,249, while Stage II incurred almost identical expenses, with an average of \$9,171⁹⁸. For colorectal cancer, chemotherapy costs \$11,800–\$13,000⁹⁹, and radiation therapy costs range from \$1,400 to \$3,800¹⁰⁰. Surgery costs range from \$14,000 to \$16,000¹⁰¹.

2.4 Methods for Resection Margin Assessment

2.4.1 Current Methods for Margin Analysis

The gold-standard method for analysis of resection margins of surgically excised specimens is currently histopathology. After the excised specimen is removed during surgery, it is sent to

the histopathology lab for further examination. In the lab, the specimen undergoes a series of processes including formalin fixation, inking to encode its different views, radial slicing, staining, and microscopic examination. These steps aim to identify any microscopic disease present on the specimen margins.

Despite being the gold-standard, histopathology is prone to errors, as studies have reported an error rate of up to 11% in cancer diagnosis¹⁰². One source of variability in histopathology lies in the interpretation and reporting of tumour grading. Different pathologists may assign different grades to the same tumour, leading to inconsistency in the assessment of its aggressiveness or malignancy^{103,104}. Moreover, there can be variability in the final diagnosis reached by different pathologists examining the same specimen. This indicates that even with standardised protocols and guidelines, there can be differences in the interpretation of histopathological findings, potentially impacting the accuracy and reliability of the diagnosis¹⁰⁵. This is even more apparent when diagnosing early cancers, where histopathology can have around 50% inter-observer agreement^{105,106}.

Ideally, the entire processing of the specimen takes approximately 72 hours according to guidelines provided by professional organisations such as the American Society of Clinical Oncology and the College of American Pathologists. However, practical limitations such as human and infrastructure resources can significantly impact the turnaround time for histopathological examination. In practice, the examination process may take several days or even longer depending on the availability of resources.

2.4.2 Methods for Intraoperative Margin Assessment (IMA)

Since the histopathological assessment of resection margins currently requires up to a few days, no direct feedback can be given to the surgeon during the initial procedure. Therefore, multiple other techniques have been explored for resection margin assessment during surgery. Current and prospective methods of intra-operative margin assessment are generally based on ex-vivo analysis of the excised specimen, as it is the most practical way of avoiding surgical artefacts such as blood contamination and the space constraints of in-vivo analysis.

Established methods that are currently available in the clinic are frozen section analysis (FSA)^{107–112}, imprint cytology (IC) or touch preparation cytology^{107–112}, specimen radiography^{108–110,113,114}, the iKnife^{115,116}, and ultrasound^{108–111}.

FSA is an intraoperative pathological examination, where parts of the excised specimen are frozen, sectioned, stained, and analysed under a microscope. The technique has shown to reduce re-operation rates. However, it is labour-intensive, time-consuming (mean: 27.8 min)¹¹⁷ and its sensitivity at detecting tumour varies (65%-78%)¹¹⁸. Moreover, it is not feasible nor economically viable in the vast majority of hospitals, as it introduces workflow delays and requires specialised staff.

IC is another intra-operative pathological examination, whereby malignant cells from the specimen borders are adhered onto a glass slide, fixated, stained, and examined under a microscope. This technique is faster (mean: 16.6 min)¹¹⁷ and more sensitive (96%-100%) than FSA. However, it is affected by cautery and is limited to examining only the superficial aspect of the specimen¹¹⁸.

Specimen radiography is the intra-operative X-ray imaging of the freshly excised specimen and is universally used to confirm tumour position, as well as to guide intraoperative decision-making if further resection margins are needed. The technique can be helpful in visualising the solid tumour's extent¹¹⁹. However, the overall sensitivity has been shown to be very low (53%)¹¹⁷.

The iKnife, also known as the Intelligent Knife or the Rapid Evaporative Ionisation Mass Spectrometry (REIMS) system, is a novel intraoperative technology used for real-time tissue characterization during surgical procedures. The iKnife works by using electrosurgical dissection to create smoke from the tissue being operated on. The smoke generated contains vaporized molecules from the tissue, including lipids, which carry unique metabolic signatures. This smoke is then captured and analysed by mass spectrometry, which provides information about the lipidomic profile of the tissue¹²⁰. The iKnife has demonstrated impressive performance across various studies, with sensitivity and specificity generally exceeding 90%. Specifically, in ovarian cancer classification, the iKnife reached a sensitivity of 97.4% and specificity of 100%¹²¹, while in oral cavity cancer surgery, figures were slightly lower at 83% and 82%¹²², respectively. For colorectal and hepatocellular carcinoma, the sensitivity and specificity ranged between 92.4% to 100%¹²³ and 90.5% to 100%¹²⁴, respectively.

While the iKnife offers real-time characterization during surgical procedures, it does have certain limitations. One of the disadvantages of the iKnife is its complexity and technical requirements. The generation and capture of smoke, through electrosurgical dissection, for

analysis by mass spectrometry require specialized equipment and expertise, making the implementation challenging. Additionally, the iKnife relies on the use of mass spectrometry, which is an expensive and technically demanding technique. The instruments and infrastructure required for mass spectrometry analysis can be costly, limiting its accessibility and practicality in many clinical settings. Another drawback of the iKnife is its destructive nature. The tissue being operated on is vaporized and consumed during the generation of smoke, preventing any further analysis or histopathological examination of the sample. This limitation can be significant, as preserving tissue samples for additional tests or analysis may be crucial for comprehensive diagnosis and follow-up.

Intra-operative ultrasound has also been used to guide lumpectomy in breast conserving surgery, but margin assessment was associated with a high false positive and false negative rate¹²⁵ and poor sensitivity¹²⁶.

The extensive range of methods suggested for IMA demonstrates that no single method has proved to be effective enough to be accepted into routine clinical practice. Despite advancements in intra-operative tools, their reliability in identifying positive resection margins during surgery remains limited, making post-operative histopathology the current gold-standard procedure for determining the need for a re-operation.

2.5 Diagnosis & Histopathology

2.5.1 Benign vs Malignant Tumours

The differentiation between benign and malignant tumours is typically based on morphological characteristics, allowing for confident classification in most cases. Histological diagnosis of tumours, on the other hand, relies on factors such as differentiation, local invasion, and metastasis⁷.

A well differentiated tumour is one in which the neoplastic parenchymal cells closely resemble normal cells from the same tissue, both morphologically and functionally. Benign tumours are generally well differentiated, whereas malignant tumours exhibit a wide range of differentiation, ranging from undifferentiated (anaplastic) to well differentiated.

Poor differentiation is associated with a worse prognosis and an increased risk of lymphatic spread, particularly in early-stage tumours¹²⁷. Lack of differentiation is a characteristic feature of malignancy and is often accompanied by other morphological

changes, including pleomorphism, abnormal nuclear morphology, abnormal mitoses, and loss of polarity.

Metastasis, the distant spread of tumour cells, is a definitive hallmark of malignancy. This can occur through various routes, including the circulatory system (haematogenous spread), lymphatic system, surgery, and body cavities^{7,74}. Haematogenous spread predominantly occurs through the veins due to their thinner walls, leading to frequent liver and lung metastases, as these organs receive a significant proportion of venous blood. The lymphatic system serves as the primary pathway for initial tumour spread, and the pattern of lymph node involvement reflects the natural lymphatic drainage of the affected area.

Although local lymph nodes can be bypassed in certain cases, the spread of tumours typically follows predictable patterns, enabling minimally invasive monitoring of tumour dissemination through the sampling of sentinel lymph nodes. These sentinel nodes, being the first nodes in the region to receive lymph flow from the tumour, have a high likelihood of containing tumour cells if metastasis has occurred. Biopsies or frozen sections obtained from sentinel nodes play a crucial role in guiding therapy for GI cancer.

Additionally, the seeding of body cavities occurs when carcinomas invade open spaces, most commonly the peritoneum. Peritoneal spread is commonly diagnosed through laparoscopy, which allows for direct visualisation and histological examination of biopsied suspicious peritoneal lesions¹²⁸.

2.5.2 Histopathology Examination

Histological reporting confirms the diagnosis and provides information on tumour morphology and staging. This is important, as it guides further treatment (i.e., surgery, chemotherapy, radiotherapy) and provides information on the patient's prognosis, feedback on the quality of surgical resection and monitoring of the effects of neoadjuvant chemotherapy^{74,127}. The main role of histological examination of a biopsy from a potentially malignant lesion is to confirm the diagnosis of adenocarcinoma or other diseases such as dysplasia. The histological type of tumour (i.e., adenocarcinoma, SCC, etc.), grade of differentiation, depth of invasion (including serosal involvement) and involvement of proximal, distal, and circumferential resection margins are reported on. Amongst others, the

overall dimensions of the specimen and the position, distance from resection margins, and macroscopic appearance of the tumour (e.g., ulcerated, polypoid) are recorded^{8,127}.

2.5.3 Tumour Grading & Classification

Cancer grading involves assessing the differentiation, number of mitoses, and architectural characteristics of the tumour. This grading system varies depending on the specific type of cancer and may not always provide clinically meaningful information, as histological features may not consistently correlate with the tumour's behaviour in-vivo. Staging, on the other hand, pertains to the extent of tumour spread and is determined by evaluating the size of the primary lesion, lymph node involvement, and the presence of distant metastases. The widely used staging system is the TNM classification detailed in Section 2.1.3. In the case of colorectal tumours, the Dukes classification is also employed as an additional staging system.

2.5.4 Diagnosis & Treatment

Endoscopy followed by histological examination of biopsies is the gold-standard diagnostic process for gastrointestinal cancer¹⁰. Surgery is the standard treatment for oesophagogastric and colorectal cancers. In advanced stage GI cancer, surgery is often combined with chemotherapy. Neoadjuvant chemotherapy is given to downstage the tumour and adjuvant chemotherapy to reduce the risk of recurrence. For advanced stage GI cancer, surgery is mostly combined with neoadjuvant (chemo)radiotherapy¹²⁹.

2.6 Challenges & Opportunities

2.6.1 The Clinical Challenge

A number of studies have shown that early and even sometimes advanced malignancies can be missed in conventional endoscopy¹³⁰. The rate of missed cancers in the West has been reported to be between 6.7–14% on initial endoscopy¹³¹. The sensitivity and specificity of diagnosing a malignant lesion with endoscopy alone is 81% and 91.7% respectively, rising to 97% and 99.4% when combined with histology results¹³². This demonstrates the inability of endoscopy alone to adequately detect areas of disease, which is why non-targeted biopsies are essential.

Additionally, around 40% of patients present with metastatic disease at diagnosis, which is a major reason for the dismal overall survival of gastric adenocarcinoma patients¹³³. Moreover, around 14% of the newly diagnosed gastric adenocarcinoma patients present with peritoneal metastases¹³⁴.

In surgery, a delicate balance should be achieved by the surgeon, considering both complete removal of the tumour and preservation of adjacent healthy tissue. While extending surgical resection margins can help reduce the risk of positive CRMs, it may also lead to damage of vital and functional surrounding structures, such as the bladder and hypogastric plexus. Currently, after rectal cancer surgery, bladder dysfunction occurs in 20–30% of the patients and sexual disorders in approximately 30%.

Erroneous surgery that leaves a positive CRM in gastrointestinal cancer can have a detrimental impact on patient survival. Achieving clear resection margins is crucial to ensure the complete removal of the cancerous tissue, reducing the risk of tumour recurrence and improving patient outcomes. Surgeons require intraoperative technologies that can accurately identify the precise cancer resection margin, allowing for a clear R0 margin resection.

The main challenge in GI cancer surgery lies in identifying the optimal resection plane that encompasses a three-dimensional clearance around the tumour. This resection plane should ensure an adequate CRM that is clear of cancer cells in all dimensions. This clearance is crucial not only for the thorough removal of the cancerous tissue but also for preservation of healthy surrounding tissue. Striking this balance is essential to achieve successful outcomes and minimise potential adverse effects on patients' quality of life.

To determine the optimal resection plane during GI cancer surgery, real-time tissue discrimination is crucial. However, in such surgical procedures, the surgeon often faces challenges in visually discriminating tissue due to the limited space within the pelvic cavity. Furthermore, the increasing use of laparoscopic techniques has diminished the tactile feedback that surgeons traditionally rely on for tissue classification, further complicating the task of defining the ideal resection plane.

By taking into account the prolonged histopathology examination for resection margin assessment, there is a clear need for more efficient and reliable intraoperative tools to accurately assess resection margins in real-time to facilitate timely decision-making in patient management.

2.6.2 *The Opportunity*

Introducing intraoperative real-time technology capable of tissue classification during GI cancer surgery would be highly advantageous for the surgeon. Such technology can significantly aid in identifying the appropriate resection plane, leading to improved surgical outcomes. By providing reliable and immediate tissue classification information, such technology can enhance the surgeon's decision-making process and contribute to more precise tumour resection while preserving healthy surrounding tissue.

To address the aforementioned challenges as well as the gap in surgical care and IMA techniques, novel optical methods show promise as they can provide sensitive and specific biochemical information at a molecular level in a rapid and non-destructive manner.

The advancement of light-based systems in the field of medicine and surgery has sparked considerable interest in modelling the propagation of light through tissues and characterising their optical properties^{135–140}. This research direction holds great potential for enhancing diagnostic applications. One particularly attractive concept is that of an “optical biopsy”, which aims to enable quick and accurate identification of diseased and malignant tissue regions, particularly on the mucosal surfaces, in a non-invasive manner^{138,139,141–145}.

Numerous research efforts are currently focused on realising this concept through various methods, including diffuse reflectance spectroscopy. The importance of achieving clear resection margins extends beyond patient outcomes. There are also economic implications associated with positive margins. A study cited the costs of positive margins alone to be \$228.4 million for a population of 140,000¹⁴⁶. These costs include expenses related to additional treatments, extended hospital stays, and potential complications resulting from incomplete tumour removal.

Therefore, the development and implementation of intraoperative technologies that can reliably and accurately identify cancer resection margins are of great importance. These technologies can help improve surgical precision, reduce the risk of positive margins, decrease morbidity and mortality, and have potential economic benefits by reducing the associated costs of positive margins.

2.7 *Multispectral Methods*

2.7.1 *Optical Biopsy*

Optical biopsy is an emerging medical diagnostic technique that uses light-based technologies to obtain information about tissue morphology without the need to remove tissue. Optical biopsy enables diagnostic information to be obtained less invasively. Instead of removing tissue for subsequent analysis in a laboratory, optical probes are used to make measurements on tissue in situ. There are several modalities being evaluated for optical biopsy, including laser confocal microscopy, optical coherence tomography, elastic scattering and reflection spectroscopy, Raman spectroscopy, and angle-resolved low-coherence interferometry.

These novel optical techniques have predominantly been tested in diseases where superficial epithelial layer changes can be detected, such as oral cancer, oesophageal cancer, adenomatous polyps, and ulcerative colitis. Since these diseases often present cellular changes in their early stages, optical biopsy techniques hold significant potential for early dysplasia detection. This could lead to a reduction in the number of biopsies needed to achieve a diagnosis.

The primary advantage of optical biopsy lies in its potential for concurrent implementation with a screening procedure and surgical applications, thus establishing a closed cycle of secondary prevention: simultaneous detection and diagnosis within a single procedure.

2.7.2 *Light – Tissue Interaction*

Optical techniques offer a unique and non-destructive approach to investigate the biological composition of tissue. These techniques exploit the specific interactions between optical radiation and tissue. One such interaction is optical absorption, where the tissue absorbs the incident optical radiation. Another interaction is scattering, which occurs when the tissue redirects the incident optical radiation. Additionally, certain optical techniques leverage fluorescence, where specific molecules in the tissue absorb the light and subsequently emit light of longer wavelength. These molecules, known as fluorophores, can be either naturally occurring within the tissue (endogenous) or externally introduced as contrast agents to enhance the tissue's fluorescent properties (exogenous). Optical technologies are particularly

advantageous as the optical radiation used (within the range of 400–1300 nm) is non-ionising and generally considered safe for biological applications.

In tissue of the GI tract, the primary endogenous absorbers within the visible spectrum include oxygenated haemoglobin (HbO_2), deoxygenated haemoglobin (HHb), and β -carotene (βc). These absorbers are directly associated with blood content, oxygenation levels, and the presence of adipocytes. Each of these absorbers demonstrates distinct spectral characteristics that are dependent on their concentration and can be observed across different wavelengths. Besides absorption, various biological factors contribute to optical scattering within the tissue. These factors encompass the size and density of cells as well as the arrangement of collagens, which serves as a structural component providing support to the tissue. Cellular structures and collagen in GI tissue cause light to scatter. The optical absorbers in GI tissue, HbO_2 and HHb, present in blood vessels and βc , typically stored in adipocytes (fat cells). The absorption of light in the mucosa epithelium of the oesophagus and colon is primarily influenced by the presence of haemoglobin. Haemoglobin concentration and oxygen saturation play crucial roles in determining the absorption characteristics of these tissues. The total absorption of light in these organs is expected to be proportional to the size and number of blood vessels and capillaries present in the connective tissue. These factors collectively contribute to the overall optical absorption properties of the mucosa epithelium in the oesophagus and colon¹⁴⁷.

Haemoglobin, a protein found in the cytoplasm of red blood cells, plays a crucial role in the transportation of oxygen from the lungs to the body tissues. It possesses the ability to bind with oxygen, enabling efficient oxygen delivery¹⁴⁸. Structurally, haemoglobin consists of four amino acid chains, each containing a haem group. Within the haem group, an atom of ferrous iron is present, which can bind with a molecule of oxygen. When all four heme groups of a haemoglobin molecule are bound to oxygen, it forms oxyhaemoglobin, characterized by a bright red colour. As oxygen is released from the haemoglobin molecule, it transitions to deoxyhaemoglobin, resulting in a deep purple colouration¹⁴⁸.

The binding of oxygen to haemoglobin is primarily governed by the partial pressure of oxygen in the blood. The saturation of haemoglobin, which represents the ratio of oxyhaemoglobin concentration to total haemoglobin concentration, increases as the partial pressure of oxygen rises, following a sigmoidal relationship as depicted in Figure 10. Under normal conditions, arterial haemoglobin saturation is approximately 97%. This pattern is

attributed to the preferential binding of oxygen to haemoglobin molecules that already have at least one oxygen molecule bound to them. Temperature and pH are additional factors that can influence the binding of oxygen to haemoglobin¹⁴⁸.

In the visible (VIS) spectral range, both oxyhaemoglobin and deoxyhaemoglobin exhibit strong absorption characteristics (Figure 10). However, the absorption of deoxyhaemoglobin becomes slightly stronger beyond approximately 590 nm. Consequently, venous blood appears darker red compared to arterial blood. It is important to note that the absorption of HbO₂ experiences a sharp decline and remains low, while the absorption of HHb does not exhibit a sharp decline and maintains relatively high levels. The two absorption spectra intersect at approximately 800 nm, which is referred to as an isosbestic point¹⁴⁹.

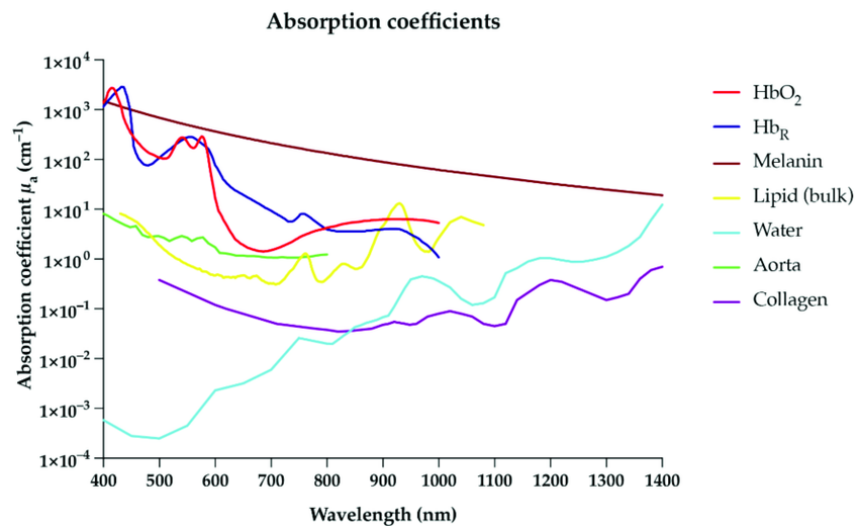


Figure 10: Absorption spectra of oxy- and deoxyhaemoglobin and water in the spectral range 400-1400 nm (adopted by Regensburger et al.¹⁵⁰).

Water is also considered to be one of the most significant chromophores in tissue spectroscopy measurements due to its high concentration in biological tissue. The absorption spectrum of water, as depicted in Figure 10, demonstrates its absorption characteristics in the range 300–1000 nm (shown in light blue). Notably, the absorption of water exhibits a sharp increase after 900 nm as the wavelength increases, with a discernible peak observed at 970 nm¹⁵¹. This region of elevated absorption signifies the presence of water as an absorptive component. However, the range of low absorption (wavelength < 900 nm) serves as a “water transmission window”, enabling near-infrared (NIR) spectroscopic measurements to penetrate several centimetres of tissue. This transmission window facilitates the acquisition of spectroscopic data through the tissue, offering insights into its optical properties and composition.

2.7.3 Optical Spectroscopy in Tissue Analysis

At present, numerous optoelectronic screening techniques for GI cancer are under investigation. Many of these techniques rely on spectroscopy principles. One approach involves the use of multispectral methods to screen for neoplasia. This approach is based on the understanding that normal and neoplastic tissues exhibit distinct optical characteristics, resulting in unique spectral signatures. These signatures are determined by the tissue's biochemistry, as well as its macro and ultrastructural morphology. By analysing the intensity and wavelength of the light emitted or reflected by the tissues when illuminated in a specific manner, it becomes possible to optically discriminate between healthy and neoplastic tissues.

2.7.4 Diffuse Reflectance Spectroscopy

One notable technique in the field of optical biopsy is DRS, which, when combined with appropriate light transport models, enables the extraction of quantitative information regarding tissue absorption and scattering. These parameters provide insights into the underlying composition of the tissue^{152–157}. Zhu et al. conducted a study to identify the optical contrast between malignant and non-malignant breast tissues, which can be utilised for intra-operative margin assessment. The findings of this study align with similar research reported in the literature^{153,158–160}. The study revealed a statistically significant decrease in β -carotene and total haemoglobin concentration, as well as a statistically significant increase in the wavelength-averaged reduced scattering coefficient in malignant tissues compared to non-malignant tissues.

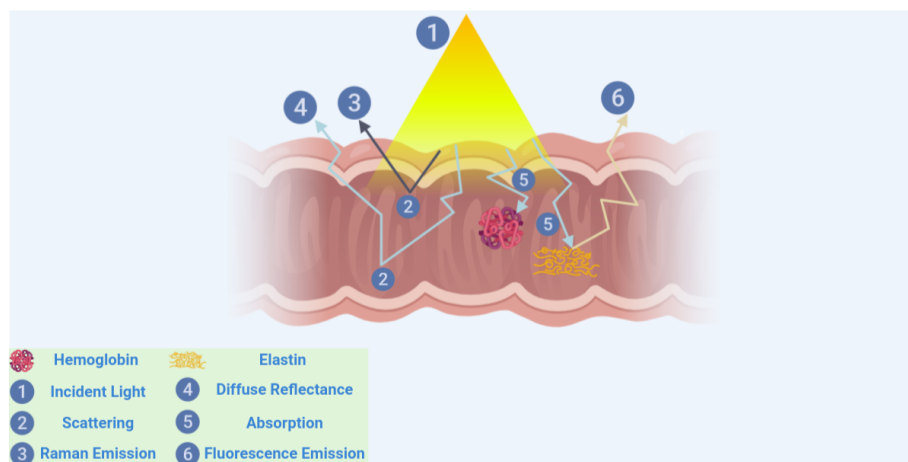


Figure 11: Basic principle of DRS demonstrating absorption and scattering inside tissue (created with [BioRender.com](https://www.biorender.com/)).

Diffuse reflectance spectroscopy is an optical technique that utilises the absorption and scattering properties of a tissue sample and can be used for intra-operative real-time tissue classification. DRS relies on the interaction between light with tissue to gather information about the tissue composition. This technique involves emitting light into the tissue and capturing the light that is reflected back from the tissue surface. By analysing the characteristics of the reflected light, an optical fingerprint or signature of the examined tissue can be obtained. This optical fingerprint provides valuable insights into the biochemical and structural properties of the tissue, allowing for non-invasive assessment of tissue composition and potential abnormalities. Figure 11 illustrates the process of interrogating a tissue specimen using DRS. Typically, light is delivered to the sample through an optical fibre. Upon entering the tissue, the light undergoes several interactions with the tissue, including absorption and scattering. Scattering refers to the change in direction of individual light particles as they interact with various structures within the tissue. This scattering can occur due to interactions with cellular components, extracellular matrix elements, and other tissue structures. On the other hand, absorption refers to the process by which light energy is absorbed by chromophores, such as haemoglobin or other molecules present in the tissue. Absorption prevents a portion of the incident light from being reemitted back from the tissue surface. A portion of the scattered light is directed towards the tissue surface and collected by an adjacent optical fibre. This collected light, known as diffuse reflectance, is then analysed.

In DRS, the distance between the emitting and receiving fibres plays a crucial role in determining the depth at which the tissue is primarily assessed. By increasing the distance between the fibres, the average path of the detected light penetrates deeper into the tissue. The specific measurement depth is primarily determined by the fibre distance, but it is also influenced by the composition of the tissue and the wavelength of the light used. As a general guideline, the measurement depth in DRS is approximately equal to the distance between the emitting and receiving fibres.

The intensity of the received signal is influenced by the interaction of light with tissue components. When the wavelength of the incident light corresponds to a range that is primarily absorbed by the tissue, the intensity of the received signal tends to be low. This is because a significant portion of the incident light is absorbed by chromophores or other absorbing molecules present in the tissue.

On the other hand, when the incident light wavelength corresponds to a range that is predominantly scattered by the tissue, the intensity of the received signal tends to be higher. This is because the scattering events cause the light to deviate from its original path, resulting in a larger portion of the scattered light reaching the detector.

The absorption and scattering properties of tissue exhibit wavelength dependency. Therefore, the diffuse reflectance is measured over a range of wavelengths to generate a complete diffuse reflectance spectrum. To analyse the spectrum, a predictive photon migration model could be employed. This model enables the extraction and quantification of the absorption coefficient (μ_a) and scattering coefficient (μ_s). The absorption coefficient represents the inverse of the mean free path of a photon before being absorbed. Similarly, μ_s represents the inverse of the mean free path before undergoing scattering. The reduced scattering coefficient is defined as:

$$(Equation 1): \mu'_s = \mu_s * (1 - g)$$

where the anisotropy coefficient, g , denotes the average cosine of the scattering angle. These parameters characterise the average scattering properties of the tissue volume under examination through the measurement of the reflected light. These coefficients are influenced by the wavelength of the incident light and the physical attributes of the scattering particles present in the tissue, including their number, size, shape, structure, and refractive index.

In columnar epithelial tissues, characterised by a single layer of cells on the surface, the measured diffuse reflectance primarily consists of photons that have undergone scattering predominantly by extracellular matrix components, specifically the collagen and elastin fibres present in the connective tissue. The diffuse reflectance spectrum obtained from such tissues is commonly described using mathematical models that are functions of the reduced scattering coefficient and the absorption coefficient¹⁴⁷.

In tissue, which is a non-transparent and turbid medium with multiple scattering events, the attenuation of light over distance is influenced by both scattering and absorption phenomena. As light propagates through the tissue, it interacts with scattering particles and is redirected multiple times, resulting in a scattered distribution of photons. Additionally, some photons may undergo absorption events, where they are absorbed by chromophores or other absorptive components within the tissue. The interplay between scattering and

absorption contributes to the distance-dependent attenuation of light within the tissue¹⁶¹. The total attenuation coefficient (μ_t) is the sum of absorption and scattering parameters:

$$(Equation\ 2): \mu_t = \mu_a + \mu_s$$

With this attenuation, the intensity of the incident light (I) diminishes according to a modified Lambert-Beer law:

$$(Equation\ 3): I = I_0 e^{-\mu_{eff} z}$$

where the I_0 is the intensity of the incident light at the surface (before any absorption or scattering), z is the depth (distance from the surface), and μ_{eff} is the effective attenuation coefficient:

$$(Equation\ 4): \mu_{eff} = \frac{1}{\delta}$$

where μ_{eff} is the inverse of the mean penetration depth (δ):

$$(Equation\ 5): \delta = [3\mu_a(\mu_a + \mu'_s)]^{1/2}$$

The intensity of reflected light at a specific point within a tissue is influenced by various factors, including the initial intensity of the incident light, the depth to which the light penetrates the tissue, and the optical attenuation experienced along its path. Optical attenuation is directly related to the scattering events that occur within the tissue and is highly influenced by the total absorbance encountered during the light's propagation through the tissue. These complex interplays between scattering, absorbance, and attenuation contribute to the observed diffuse reflectance characteristics in tissues.

By harnessing DRS, the goal is to develop a technology that can facilitate efficient biopsies by guiding mucosal sampling during surgery to determine tumour margins, assisting surgeons in achieving precise and effective tumour resection^{138,142}. DRS, specifically, holds the potential to revolutionise cancer surgery and become the next gold-standard for IMA.

2.7.5 Tissue Fluorescence Spectroscopy

Fluorescence and absorption are interconnected phenomena in which an illuminated molecule undergoes excitation by absorbing light energy and subsequently emits stored energy. The extent of light absorption by the molecule, as well as the intensity of fluorescent light emitted, are determined by the molecular structure of the molecule and the wavelength of the incident light. These characteristics are represented by the molecule's absorption spectrum and emission fluorescence spectrum, respectively. Jablonsky diagrams are

employed to depict the electronic energy levels of molecules, offering a visual representation of the possible outcomes when a molecule is excited by light. By utilizing these diagrams, researchers can gain insights into the energy transitions and fluorescence properties of molecules upon excitation¹⁶².

Jablonsky diagrams visualise the process wherein a molecule transitions from its ground state to an excited electronic state following absorption of light. This transition occurs when the energy of the excitation photon matches the energy difference between two specific energy levels. The resulting excited state energy level depends on the wavelength, and thus the energy, of the incident light. Subsequently, the molecule can undergo vibrational relaxation, moving from the initial excited state to the lowest excited state energy level. From there, the molecule can return to its ground state by releasing its energy through two possible pathways: (1) emitting a photon in the form of light, known as fluorescence or phosphorescence, or (2) undergoing non-radiative relaxation, such as heat dissipation¹⁶². The wavelength of the emitted photon aligns with the energy level difference between the excited state and one of the ground state energy levels, in accordance with Bohr's frequency condition.

The average duration that a molecule remains in its excited state before returning to the ground state is referred to as its lifetime. For fluorescence, the lifetime is typically on the order of one nanosecond, while for phosphorescence, it can extend to several milliseconds or even longer¹⁶³. The absence of fluorescence can result from either non-radiative relaxation, such as thermal dissipation, or a slower radiative relaxation process, as observed in phosphorescence¹⁶⁴.

The quantum yield of a molecule's fluorescence is a parameter that quantifies the efficiency of photon emission in relation to photon absorption. It represents the ratio of the number of emitted photons to the number of absorbed photons. Both the fluorescence lifetime and quantum yield can be influenced by factors such as pH and temperature¹⁶². Changes in pH and temperature can alter the molecular environment and affect the relaxation processes, thereby influencing the fluorescence characteristics of the molecule.

The absorption and fluorescence spectra of a fluorophore exhibit a phenomenon known as the Stokes's shift. This shift refers to the red-shifted position of the fluorescence spectrum compared to the absorption spectrum. The Stokes's shift arises from the partial loss

of energy through intermolecular collisions, also known as vibrational relaxations, which occur prior to the emission of fluorescence¹⁶⁵.

In summary, the fluorescence spectrum of a fluorophore is determined by its intrinsic properties, including the extinction (or absorption) coefficient at the excitation wavelength and the quantum yield at the emission wavelength. These properties give each fluorophore a unique spectral signature. However, the fluorescence spectrum can also be influenced by factors such as the concentration of the fluorophore and the intensity of the excitation light^{165,166}.

The detection of fluorescent signals in medical diagnostic instruments offers the advantage of relying on endogenous fluorophores present in the tissue, eliminating the need for exogenous fluorophores that require considerations of toxicity, dose optimisation, and injection timing¹⁶⁴. The main endogenous fluorophores can be broadly categorised as follows:

- Connective tissue proteins: collagen and elastin
- Enzymes: reduced nicotinamide adenine dinucleotide (NADH) and flavins (such as flavin adenine dinucleotide (FAD)) involved in cellular energy metabolism
- Aromatic amino acids: tryptophan, tyrosine, phenylalanine
- Lipopigments: products resulting from lipid peroxidation (e.g., lipofuscin)
- Porphyrins and vitamins

2.7.6 Raman Spectroscopy

The Raman scattering process, also known as the Raman Effect, was discovered by Sir Chandrasekhara Venkata Raman and Kariamanickam Srinivasa Krishnan in liquids¹⁶⁷ and by Grigory Landsberg and Leonid Mandelstam in crystals^{162,168}. In recognition of his work on the scattering of light, Raman was awarded the Nobel Prize in 1930.

When light interacts with an atom or molecule and undergoes scattering, the majority of photons are elastically scattered (Rayleigh scattering), resulting in scattered photons with the same energy (frequency) and wavelength as the incident photons. However, a small fraction of the scattered light, approximately 1 in 10 million photons, undergoes an inelastic scattering process. In this process, the scattered photons have a different frequency from, and usually lower than, the frequency of the incident photons.

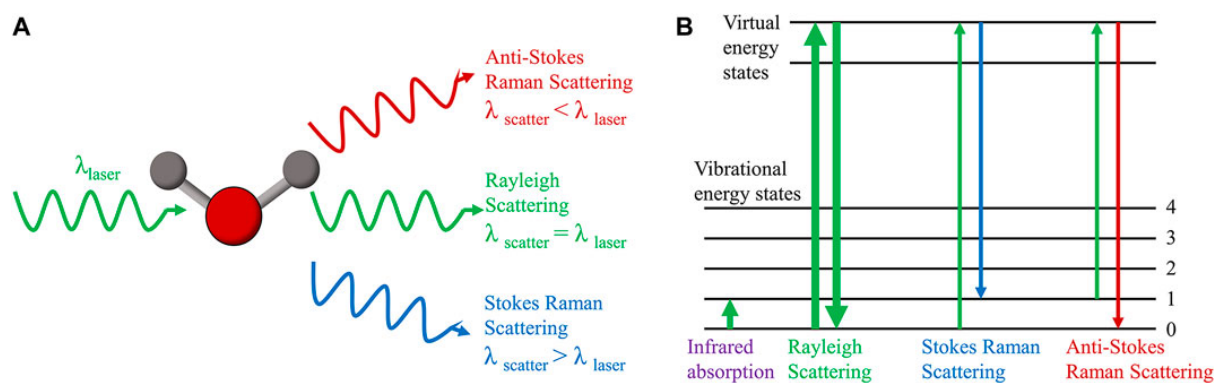


Figure 12: Principle of Raman scattering. (A) Raman scattering and Rayleigh scattering. (B) Energy level diagram showing the states involved in Raman signal (Raman scattering, Rayleigh scattering, and infrared absorption) (adopted by Liu et al.¹⁶⁹).

Conventional Raman spectroscopy primarily focuses on Stokes Raman scattering. The Raman spectrum is generated by measuring the Raman intensity as a function of frequency shift, known as the Raman shift, which is quantified in wavenumber ($1/\lambda_{\text{excitation}} - 1/\lambda_{\text{Raman}}$). The wavenumber in Raman spectra is commonly expressed in inverse centimetres (cm^{-1}). As a result, the Raman spectrum exhibits distinct bands that correspond to specific groups of vibrational bonds within the molecules present in the sample.

It is worth noting that the frequency shift for a specific vibration band of the same molecule remains the same in both Raman scattering and infrared (IR) absorption, as depicted in Figure 12. However, the selection rules governing Raman scattering differ from those of IR absorption. In IR absorption, a molecule absorbs light only when there is a change in dipole moment during molecular vibration. In contrast, the Raman effect occurs due to an oscillation-induced dipole moment, where the molecular interaction with light is through the polarizability of the molecule. Consequently, not all molecules exhibit both Raman and IR activity, making Raman spectroscopy and IR spectroscopy complementary techniques.

In Raman spectroscopy, a laser is typically used as the excitation light source. This is because lasers provide a single, narrow band of monochromatic light that is well-suited for probing molecules. The use of lasers is advantageous in Raman spectroscopy due to the weak nature of Raman scattering, where a large amount of light is required to elicit a detectable signal. Furthermore, the unique specificity of Raman spectra relies on the use of a single wavelength, as a broader light source would result in overlapping spectra that would be difficult to interpret.

The most commonly used type of laser in Raman spectroscopy of biological materials is the diode laser. Diode lasers are particularly suitable as they efficiently convert electrical

energy into light energy, providing a reliable and convenient light source for Raman experiments.

It is worth noting that there is a trade-off when selecting the laser wavelength, especially when working with biological samples. Decreasing the laser wavelength can increase the Raman signal intensity. However, shorter wavelengths can also lead to heating of the specimens, potentially causing molecular changes or damage. Additionally, shorter wavelengths are more susceptible to the interference of autofluorescence, which can complicate the interpretation of Raman spectra in biological samples¹⁷⁰. Therefore, careful consideration is required to strike a balance between maximizing Raman signal and minimizing potential sample effects and interference.

In Raman spectroscopy, the NIR spectrum is commonly employed, particularly when analysing biological tissue samples. The NIR spectrum offers several advantages, including achieving an optimal balance between signal intensity, tissue penetration depth, and minimizing confounding factors such as autofluorescence¹⁷¹. Moreover, it is considered to be the safest wavelength range for biological applications.

2.7.7 Fibre-Optic Spectroscopy

Optical Fibre Probes

Fibre probes used in optical spectroscopy can be categorised into two types: mono-fibre probes and bifurcated probes. Mono-fibre probes consist of a single fibre that serves both for illumination and detection. In contrast, bifurcated probes have separate fibres for excitation and detection. The term “bifurcated” refers to the two arms of the probe, with one arm connected to the light source for excitation and the other arm connected to the disperser/detector system for detection. Mono-fibre probes are relatively simple to construct and allow for a smaller diameter, resulting in a higher light collection efficiency compared to bifurcated probes. However, they require additional optics, such as beamsplitter, to separate the excitation and emission light. One limitation of the mono-fibre approach is the collection of a large amount of specular light, as well as internal reflections at each surface of the fibre, which can induce autofluorescence. This autofluorescence arises from the interaction between the illumination light and the optical fibre itself¹⁷².

The optical fibres employed in these applications are typically multimode fibres with diameters in the range of hundreds of microns and a numerical aperture of approximately 0.22. To minimize parasitic autofluorescence originating from the fibre itself, high-grade silica is preferred over regular silica as the material for the fibre core. Regular silica can generate fluorescence when excited by UV light, which is the case in this work. Additionally, to prevent light from exiting the fibre core and reaching the jacket due to fibre flexion or defects, polyimide is chosen as the material for the fibre coating instead of nylon, as nylon exhibits fluorescence under UV light.

The selection of materials for the probe components is not only focused on minimizing autofluorescence but also on ensuring UV/visible light transmission. Furthermore, as the probe is the only part of the optical spectroscopy setup that comes into contact with tissue, the materials must be biocompatible.

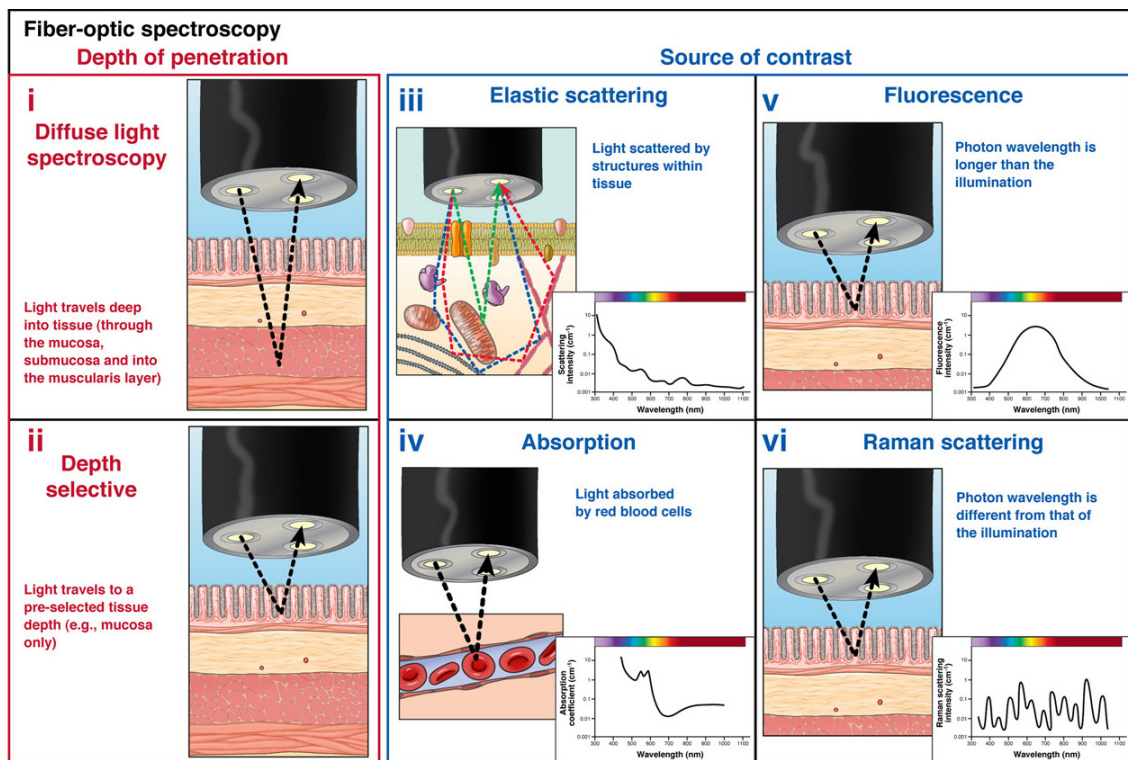


Figure 13: Fibre-optic spectroscopy technologies. (i) Diffuse light spectroscopy involves the measurement of scattered light across a range of wavelengths providing information about tissue composition and structure. (ii) Depth-selective spectroscopy utilises fibre-optic probes with controlled light penetration and selectively investigates specific tissue depths, such as the mucosa layer, to target particular regions of interest. (iii) Elastic scattering spectroscopy examines the redirection of incident light without altering its energy, providing insights into tissue morphology and cellular structure. (iv) Absorption spectroscopy focuses on the absorption of specific wavelengths of light by tissue constituents, revealing details about molecular content, including chromophores like haemoglobin. (v) Fluorescence spectroscopy examines the emission of fluorescent light after tissue is excited with high-energy photons, enabling the detection of naturally occurring fluorophores and exogenous probes. (vi) Raman scattering analyses the inelastic scattering of light, unveiling molecular vibrational and rotational information, facilitating molecular identification (adopted by Roy et al.¹⁷³).

Probe Geometry

The geometric parameters of a probe, including probe-to-target distance (PTD), source-to-detector separation (SDS), fibre numerical aperture (NA), and diameter, have a significant influence on the generation and detection of light during optical measurements. Inverse Monte Carlo simulations can be employed to estimate tissue optical properties, while forward Monte Carlo simulations are valuable for understanding the extent and characteristics of the tissue volume being excited and probed, including the impact of probe geometry on these factors.

The probe geometry used in optical spectroscopy plays a crucial role in defining the sensing volume and depth of the measurement. There are two main categories of probe geometries: contact and non-contact.

Contact geometries involve direct physical contact between the probe and the sample being measured. These contact probes are more commonly used in quantitative optical spectroscopy because they allow for well-defined sensing depths and facilitate accurate and quantitative measurements. By placing the probe in direct contact with the sample, the light interaction can be controlled, and the measurement can be targeted to a specific depth within the sample.

On the other hand, non-contact methods do not require physical contact between the probe and the sample. These methods are more suitable for imaging applications, as they allow for the collection of spatial information over a larger area. However, non-contact methods present challenges in terms of modelling and quantitation, as the light interaction with the sample becomes more complex. Specialized approaches have been developed to address these challenges and enable quantitative measurements using non-contact methods. According to Bargo et al. study¹⁷⁴, the diffuse reflectance collection efficiency (CE) of a single fibre probe and a bifurcated probe, consisting of one excitation fibre surrounded by a ring of collection fibres, was calculated using a semi-infinite medium with varying optical properties. The main finding of the study was that the CE was proportional to $\sin^2(\text{NA})$ for highly scattering samples, regardless of whether a single optical fibre or a multiple fibre probe was used. However, the CE of the multiple fibre probe was more influenced by the sample's absorption coefficient. This was attributed to the longer mean pathlength of collected photons in the multiple fibre probe, as they travelled from the source fibre to the detector fibre, allowing the absorption to have a larger distance over which it could exert its effect.

The probing volume and depth in optical spectroscopy are influenced by several factors, including the optical properties of the medium (such as absorption and scattering coefficients) and the probe geometry, particularly the source-detector separation¹⁷⁵.

Increasing the source-detector separation leads to an increase in the probing depth and volume. This means that the probe becomes more sensitive to deeper layers of tissue¹⁷⁵. By increasing the distance between the source and detector, the light has a greater chance to penetrate deeper into the tissue before being detected, allowing for measurements from deeper regions.

However, the extent to which the source and detector can be separated is limited by the tissue's ability to transmit light. Tissue has inherent absorption and scattering properties that can attenuate and scatter the light as it propagates through the medium. As the separation distance increases, the amount of light reaching the detector decreases, and the signal-to-noise ratio may be reduced.

In the NIR range, the source-detector separations are typically larger (on the order of centimetres) compared to the UV/visible range (on the order of millimetres). This is because near-infrared light has better tissue penetration capabilities due to lower absorption and scattering in this wavelength range. Therefore, NIR spectroscopy is often used for probing deeper tissue layers and organs.

2.7.8 Advantages of Multispectral Cancer Screening Methods

The emergence of optical spectroscopy offers great promise for wide use in medical fields, especially for early detection of GI cancer. The field of optical spectroscopy (OS) offers valuable insights into disease processes by detecting structural alterations and biochemical transformations in various biomolecules. These changes, including modifications in proteins, nucleic acids, lipids, and carbohydrates, contribute to the development of diseases such as cancer^{176,177}.

OS utilizes the interaction of photons with different chemical elements and compounds within biological samples, leading to the absorption and scattering of light. By analysing the optical spectra generated from these interactions, a unique optical fingerprint or signature can be obtained. This optical fingerprint provides a wealth of information about the composition and properties of the sample, including disease-related alterations^{178,179}.

One of the key advantages of OS in clinical diagnostics is its ability to analyse small sample sizes with minimal sample preparation, preserving the biochemistry of the sample¹⁸⁰. This non-destructive nature allows for real-time, non-invasive analysis of complex biological substrates and biofluids¹⁸¹. Moreover, OS is a cost-effective tool, making it accessible for diagnostic applications^{182–184}.

The information obtained from optical spectra can be used for a wide range of clinical purposes, such as disease diagnosis, monitoring treatment response, assessing disease progression, and guiding therapeutic interventions. For example, OS has been explored in the diagnosis of various cancers, including breast, prostate, colorectal, and lung cancers. It has also shown promise in the detection of infectious diseases, metabolic disorders, and neurodegenerative conditions.

In summary, optical spectroscopy is a powerful and versatile tool for clinical diagnostics. Its ability to generate unique optical fingerprints from biological samples provides valuable information about disease-related changes. With its non-invasive, real-time, and cost-effective nature, OS holds great potential for improving disease diagnosis and patient care. Continued research and advancements in this field are expected to further enhance the clinical utility of optical spectroscopy in various areas of medicine.

Multispectral cancer screening offers several potential advantages when compared to other techniques (Section 2.4.2). Firstly, it is a simple and low-cost method that does not require complex equipment or expensive procedures. Additionally, multispectral screening is non-invasive, minimising patient discomfort and risks associated with invasive procedures.

Another advantage is the potential for real-time results. By analysing the spectral characteristics of tissues, multispectral screening can provide immediate feedback during the procedure, allowing for prompt decision-making by the surgeon. Furthermore, this technique can be easily performed via the working channels of endoscopes, making it convenient and feasible for integration into existing endoscopic procedures.

An important benefit is that multispectral screening does not rely on the use of pigments, dyes, or external fluorophores. While the addition of these substances can potentially enhance the imaging process when used in conjunction with multispectral approaches, their use is not essential. This reduces the need for additional agents and simplifies the screening procedure.

Moreover, multispectral screening provides objective and quantitative data, although it typically does not generate traditional visual images. The obtained data can be analysed objectively, enabling more accurate assessments and comparisons.

The instrumentation needed to conduct DRS measurements is characterised by its simplicity and affordability. It typically comprises a broadband light source, often a high-power xenon arc lamp, which emits light over a broad range of wavelengths. Optical probes, usually utilising fibre optics, are employed to deliver and collect light to and from the target tissue area. The collected light is then dispersed into its different wavelengths of interest using a spectrograph. Finally, a detector captures the dispersed light for analysis^{147,172}.

Additionally, leveraging the field effect carcinogenesis model, multispectral methods have the potential to visualise tissue and cellular changes at a micro scale that are otherwise invisible to the naked eye. This capability could enable the detection of early-stage dysplasia without the need for expensive microscopic imaging devices. These nanoscale changes occurring in the early stages of dysplasia can be examined using multispectral techniques, which may significantly enhance screening efforts by enabling extremely early detection of pre-neoplastic transformations. In this way, the utilisation of DRS holds promise for advancing early detection and risk assessment in the context of GI cancers.

Overall, multispectral cancer screening shows promise as a cost-effective, non-invasive, real-time, and objective screening technique that can be seamlessly integrated into clinical practice.

2.8 Clinical Application of Optical Spectroscopy in GI Cancers

In recent years, several published reviews have been focused on tissue differentiation using optical spectroscopy.

Oesophageal cancer is a significant cause of cancer-related mortality, often diagnosed at an advanced stage with poor survival rates¹⁸⁵. Early detection of oesophageal cancer is crucial for improving patient outcomes. Optical spectroscopy has emerged as a promising technique for the early diagnosis of cancer. In a study by Ishigaki et al.¹⁸⁶, ex-vivo Raman spectroscopy was conducted on early-stage oesophageal cancer tissues (stages 0 and I). The analysis demonstrated a discriminative sensitivity of 81.0% and a specificity of 94.0% in

distinguishing between cancerous and normal samples. These findings provide increasing evidence for the efficacy of OS, specifically Raman spectroscopy, in the early diagnosis of oesophageal cancer. Further research is warranted to validate these results and explore the clinical utility of OS in routine diagnostic practices.

Barrett's oesophagus is recognized as a precursor lesion to oesophageal adenocarcinoma, with a sequential progression from low-grade dysplasia (LGD) to high-grade dysplasia (HGD) and ultimately adenocarcinoma^{187,188}. Individuals with Barrett's have a significantly higher risk of developing adenocarcinoma compared to those without this condition¹⁸⁹. The annual risk of malignant progression is particularly elevated in Barrett's patients with HGD compared to those with nondysplastic Barrett's¹⁹⁰. Early detection of dysplasia and early neoplasia in Barrett's can play a crucial role in improving patient outcomes and preventing the progression to invasive malignancy¹⁹¹. In a study involving 673 samples of Barrett's acquired from 62 patients, endoscopic Raman spectroscopy was employed in an ex-vivo setting. The analysis demonstrated a sensitivity of 86% and a specificity of 88% in distinguishing advanced neoplasia (HGD/adenocarcinoma) from LGD, nondysplastic Barrett's, and normal squamous oesophagus tissue¹⁹².

Endoscopic procedures in the GI tract have been targeted for improvement through OS studies. One study by Georgakoudi et al. utilized a combination of DRS and fluorescence spectroscopy (FS) to enhance the recognition of high-grade Barrett's oesophagus in 16 patients¹⁹³. By analysing the spectroscopic data using visible light illumination, they were able to distinguish normal oesophageal mucosa from low- and high-grade dysplasia with a sensitivity of 79% and specificity of 88%.

In an in-vivo study conducted by Lovat et al., elastic scattering spectroscopy (ESS) was evaluated for its ability to discriminate between normal mucosa and low-grade dysplasia from high-grade dysplasia and oesophageal cancer within Barrett's oesophagus in 81 patients¹⁹⁴. The study reported a sensitivity of 92% and a specificity of 60% in identifying high-grade dysplasia and oesophageal cancer, demonstrating the potential of ESS as a valuable tool for targeted biopsy in the surveillance of Barrett's oesophagus.

Teh et al. conducted two recent studies using Raman spectroscopy (RS) on ex-vivo stomach tissue samples. In the first study, they evaluated the capability of their Raman system to differentiate between normal gastric mucosa and dysplastic gastric tissues¹⁹⁵. By applying

principal component analysis, they achieved a maximum sensitivity of 95% and specificity of 91% in distinguishing between these tissue types.

In a subsequent study, the authors compared Raman spectra obtained from normal gastric tissue with those from intestinal-type and diffuse-type gastric adenocarcinomas in 62 patients during gastroscopy¹⁹⁶. They were able to discriminate between these two specific cancer types based on differences in collagen content, as well as specific variations in lipid and protein content at 1450 cm^{-1} . The predictive accuracies for differentiating these tissues ranged from 88% to 94%.

Dhar et al. and Wang et al. conducted studies introducing DRS to colonoscopy procedures. The Dhar group focused on differentiating normal colon mucosa from various (pre)malignant lesions¹⁹⁷. They achieved sensitivities and specificities ranging from 75% to 85% in distinguishing between different colonic tissues.

In a study by Wang et al., increased levels of total haemoglobin and decreased levels of oxygen saturation were found in (pre)malignant lesions¹⁹⁸, which aligns with DRS results obtained in malignant tissue from other organs.

Furthermore, Chowdary et al. performed an ex-vivo analysis of colonic tissue using RS¹⁹⁹. They demonstrated a discriminative accuracy of 95% between normal and malignant tissue, highlighting the feasibility of RS for future in-vivo studies in combination with endoscopic procedures.

Accurate differentiation between malignant and benign gastric ulcers is crucial in clinical practice, as some gastric ulcers can be ulcerated gastric carcinomas. Li et al. conducted a study in which they measured diffuse reflectance spectra of four types of stomach tissues: normal mucosa, chronic gastric ulcer, poorly differentiated gastric adenocarcinoma, and undifferentiated gastric adenocarcinoma²⁰⁰. Using wavelengths ranging from 400 to 650 nm, they observed that malignant tissues exhibited an increase in optical absorption at the HbO₂ band. The DR spectral intensity followed the pattern of undifferentiated adenocarcinoma < differentiated adenocarcinomas < gastric ulcer < normal mucosa. Additionally, there was a significant decrease in the DR ratio of R540/R575 in malignant tissues compared to benign ulcers.

Bergholt et al. investigated the diagnostic potential of Raman endoscopy for in-vivo discrimination between benign and malignant gastric ulcers²⁰¹. They employed a partial least-squares discriminant analysis algorithm (PLS-DA) to analyse a large dataset of 1102 Raman

spectra obtained from 71 patients. Their results demonstrated a high sensitivity of 90.8% and a specificity of 93.8% for identifying normal mucosa, 84.7% and 94.5% for identifying benign ulcers, and 82.1% and 95.3% for identifying malignant ulcerous lesions.

Considerable efforts have been dedicated to optimizing the detection of polyps during colonoscopy. Rodriguez-Diaz et al. conducted an in-vivo study in which they applied elastic-scattering spectroscopy for real-time classification of polyps as neoplastic or non-neoplastic lesions²⁰². Their approach achieved a sensitivity of 91.5%, specificity of 92.2%, and accuracy of 91.9%. Notably, a high-confidence rate of approximately 90% was associated with accuracies of 90.6% for polyps smaller than 1 cm and 90.1% for polyps equal to or smaller than 5 mm. Inelastic-scattering spectroscopy has also been investigated as a method for differentiating adenoma from hyperplastic polyps. This technique demonstrated a diagnostic sensitivity of 90.9% and a specificity of 83.3% in distinguishing between these two types of polyps^{203–205}.

In light of the established efficacy of optical spectroscopy techniques in distinguishing healthy from cancerous tissue, this chapter underscores the necessity for technological evolution despite the already promising accuracy. The motivation to innovate further is anchored in the objective to transition DRS from ex-vivo to in-vivo applications seamlessly, enhancing intraoperative CRM assessment accuracy. By developing a compact, cost-effective optical spectroscopy system, augmented with a machine learning pipeline for real-time, automated tissue classification able to guide surgeons with unparalleled precision through augmented surgical maps, this research aims to set a new benchmark in surgical precision marking a significant leap beyond current capabilities.

Chapter 3: Integrated DRS Probe Tracking

Methods to Aid Tumour Margin Assessment

3.1 Background and Need for DRS Probe Detection

The accurate assessment of tumour margins is a pivotal aspect in the field of oncology, directly influencing patient outcomes and treatment strategies. Achieving precise and reliable tumour margin evaluation during surgical interventions is critical for minimizing the risk of cancer recurrence and maximizing healthy tissue preservation. Optical techniques, such as DRS, have emerged as promising tools for real-time, non-invasive tumour margin assessment. By probing the tissue with light and measuring the reflected spectra, DRS enables the identification of molecular and structural characteristics associated with tumour cells.

In the realm of in-vivo clinical diagnostic applications, DRS has proven valuable for providing detailed biochemical information about tissues. However, its traditional use with handheld fibre-optic probes has been limited to collecting point spectra, offering diagnostic information at discrete locations rather than generating a diagnostic image for a large area. While this pointwise application can yield highly accurate diagnoses at specific points, it falls short in enabling comprehensive tumour margin delineation unless clinicians can simultaneously visualise and record diagnoses for numerous points across the tissue.

The main limitation of point-based acquisition and tissue characterisation methods, such as the DRS, is the difficulty of tracking probed sites. The absence of visible marks on the tissue when scanning with the DRS probe indeed presents a challenge for tracking the probed sites during tumour margin assessment. In situations where physical marks cannot be relied upon, an alternative approach involves tracking the tip of the DRS probe itself. As mentioned, the tip of the probe must be in direct contact with the tissue during spectra acquisition, making it a potential reference point for localisation.

To address these challenges, the integration of image guidance with DRS probe tracking has emerged as a promising solution. Image guidance involves the utilization of

advanced imaging modalities, such as ultrasound, magnetic resonance imaging (MRI), or intraoperative optical imaging, to provide real-time visualization and navigation during surgery²⁰⁶. By combining DRS with image guidance, the spatial relationship between the DRS probe and the tumour region can be continuously monitored, enhancing the accuracy and reliability of margin assessment.

By continuously monitoring the position and orientation of the DRS probe's tip, the spatial relationship between the probe and the tissue can be established and tracked. This approach leverages the principles of surgical navigation, wherein advanced imaging modalities and sensor technologies are integrated to provide feedback and guidance during the surgical procedure.

The process of localising the biopsy site through DRS probe tip tracking involves several key steps:

- In marker-based tracking, the DRS probe is equipped with optical markers which enable real-time localisation of the probe tip in relation to the patient's anatomy. These markers are not used in markerless probe tracking, where deep neural networks are often utilised to process optical images of the surgical field and infer the DRS probe's position within it.
- Optical imaging is utilized to visualize the surgical field and the DRS probe's position within it. This imaging data serves as a reference for tracking the probe's movement.
- The information from the DRS probe and the image guidance system is fused, allowing for a comprehensive understanding of the probe's position relative to the tissue.
- The localized position of the DRS probe tip is displayed to the surgical team on a navigation system, providing real-time feedback during the scanning process. This feedback enables the surgeon to assess the coverage of the tumour region and identify areas that require further investigation.

By employing DRS probe tip tracking in combination with image guidance, surgeons can enhance the accuracy and reliability of tumour margin assessment without relying on visible marks on the tissue. Such a combined system would enable clinicians to obtain real-time, high-resolution spectroscopic information while simultaneously visualizing the tissue's overall appearance. As a result, they would have a more comprehensive understanding of the tumour boundaries and the surrounding healthy tissue. This spatial co-registration of

diagnostic information could facilitate precise tumour margin delineation, aiding surgeons in making informed decisions during surgical interventions and enhancing the overall treatment planning process.

Incorporating DRS into the surgical workflow can greatly aid in histopathology correlation as well, which refers to the ability to link real-time spectroscopic information with the corresponding histopathological features in the tissue. This correlation is essential for accurate interpretation of the spectral data and validation of its diagnostic potential. By precisely tracking the position and orientation of the DRS probe during tissue scanning, the spectroscopic data can be spatially mapped to the tissue's histological features. This spatial co-registration enables the identification of specific regions of interest, such as tumour boundaries, areas of inflammation, or healthy tissue, within the spectroscopic measurements.

3.2 *Visual Tracking*

The aim of object tracking in the context of medical interventions is to continuously locate and monitor a predefined target over a period of time. Several approaches exist for achieving this task, each with its own advantages and limitations^{207,208}.

While electromagnetic tracking²⁰⁹ and inertial sensors^{210,211} are common approaches to object tracking, their use in certain medical applications is not ideal. Electromagnetic tracking raises concerns about interference with medical devices and may lack the required accuracy. Inertial sensors, while suitable for less accurate scenarios, may not meet the high precision requirements of our DRS system.

An alternative approach is mechanical tracking²¹², which utilizes the mechanical links of a surgical robot or a passive arm to track the object's position. If the tracked object is physically connected to the surgical device, and the kinematic chain between the object and the device is well-defined, the spatial position of the object can be derived. This method can be effective for surgeries that have mechanical support, such as robot-assisted surgeries, where the surgical robot's precise movement can be used to track the object's position. However, mechanical tracking requires an exact calibration process, and it is limited to situations where the surgical device can be directly linked to the object being tracked.

Visual tracking offers a promising and straightforward approach for object tracking in medical interventions, particularly when using image sequences as input. Integrating visual tracking does not require significant effort and allows for measurements directly in the reference frame of the observing camera, such as an endoscope²¹³. Unlike other tracking techniques that might necessitate complex extrinsic calibrations for visualization applications, visual tracking offers a more convenient and flexible solution. In the context of this dissertation, we define visual tracking as "the problem of generating an inference about the motion of an object given a sequence of images", following the definition proposed by Forsyth and Ponce²¹⁴. This approach is widely used in various application domains, including autonomous car driving, surveillance, and sports. In our specific case, visual tracking pertains to the ability to locate and monitor the surgical instrument in a sequence of surgical images.

The advantage of using visual tracking in our medical applications lies in its simplicity and compatibility with standard imaging equipment, such as endoscopes. By employing image sequences captured by these devices, we can track the motion and position of the surgical instrument without the need for complex external calibrations. This direct, camera-centric approach streamlines the tracking process and facilitates the integration of visual tracking into the surgical workflow. With visual tracking, we can continuously monitor the movement of the surgical instrument during the procedure, providing real-time feedback to the surgical team. This dynamic tracking capability is crucial in surgeries where precise instrument localization is paramount, ensuring accurate and efficient surgical interventions. By leveraging visual tracking for locating the surgical instrument, we enhance the overall control and precision of medical procedures. The ability to track the instrument in the reference frame of the camera allows for seamless visualization and natural interaction between the surgeon and the tracking system. Additionally, visual tracking complements the existing imaging modalities used in medical interventions, facilitating a comprehensive understanding of the surgical field.

Visual tracking approaches can be broadly categorized into two main types: temporal tracking and tracking by detection. Temporal tracking, also known as tracking by association, involves associating information from previous frames with the information obtained from the current frame. The primary objective of temporal trackers is to simplify the tracking problem by focusing on updating the object's location based on its previous known positions. By utilizing the temporal coherence of the object's motion, these trackers can provide smooth

and continuous trajectories of the tracked object. Temporal tracking is particularly useful when the object being tracked exhibits consistent and predictable motion patterns.

On the other hand, tracking by detection, as the name suggests, relies solely on the information available in the current frame to infer the location of the object. Instead of leveraging temporal information, detection-based trackers utilise sophisticated object detection algorithms to identify the presence and location of the object in each frame independently. These methods are generally well-suited for situations where the object's motion is less predictable or when dealing with occlusions and abrupt changes in the object's appearance.

For a more comprehensive understanding of the general concepts and techniques employed in visual tracking, Yilmaz et al.²¹⁵ provide an in-depth survey that covers various approaches, algorithms, and applications in the field of visual tracking.

3.3 Object Representation

In surgical instrument tracking, the representation of the instrument's shape and appearance in the image plays a crucial role and can vary depending on the specific application requirements. The typical surgical instrument is a metallic and relatively rigid object, which allows for its representation in 2D space. For applications that necessitate only a basic understanding of the instrument's movement, simpler geometric representations such as a centroid or a bounding box around the instrument are often sufficient. These geometric representations indicate the presence of the instrument and enable the calculation of a rough trajectory. The bounding box can also facilitate subsequent instrument classification and surgical action recognition.

However, in cases where a higher level of detail is required, such as graphical overlay for augmented reality applications, a coarse localization with a centroid or bounding box may not be adequate. In such scenarios, a more comprehensive representation that divides the image into two disjoint sets is more helpful. This representation includes the 2D boundary of the instrument, known as the contour of the object, and the region inside the contour, known as the silhouette or segmentation. By utilizing this more detailed representation, every pixel of the image can be assigned to either the instrument or the background, providing precise and fine-grained information about the instrument's position and shape.

3.4 Evaluation Metrics

The evaluation of an instrument tracker's performance depends on the chosen object representation and the underlying machine learning-based approach. In medical applications, where the available dataset is often limited, careful evaluation strategies are necessary to ensure the model's generalizability.

A common approach for evaluating machine learning-based trackers is to split the available dataset into independent subsets. These subsets are usually organized in a sequence- or surgery-wise manner to ensure that the model does not simply memorize specific scenarios but instead learns a generally applicable representation.

3.4.1 Reference Point

The following metrics can be used to evaluate accuracy for the reference points. **Keypoint Threshold (KT)**, also known as Threshold Score, is a metric used to assess the quality of an estimated reference point location in object tracking tasks, specifically in the context of pixel distance measurement²¹⁶. The goal of KT is to determine how accurately an estimated reference point matches the ground truth annotation. Given an estimated reference point location $j \in R^2$ and its corresponding ground truth annotation $\hat{j} \in R^2$, KT evaluates the correctness of the estimation by computing the Euclidean distance between the estimated and ground truth positions. The estimation j is considered correct if the Euclidean distance between j and $\hat{j}$ is lower than a predetermined pixel threshold $T \in R$:

$$(Equation 6): ||j - \hat{j}|| < T$$

In other words, if the distance between the estimated and ground truth positions is within the threshold T , the estimation is considered accurate. Otherwise, it is considered inaccurate. It is important to note that KT provides a separate evaluation for each reference point $j \in J$. In object tracking scenarios, multiple reference points may be tracked simultaneously, each corresponding to different objects or features of interest. KT evaluates the accuracy of each reference point's estimation independently, providing a comprehensive assessment of the tracking performance for each individual reference point. The choice of the pixel threshold T depends on the specific tracking task and the desired level of precision. A

smaller threshold will require a higher level of accuracy in the estimation, while a larger threshold may allow for some degree of error tolerance.

Keypoint Threshold Bounding Box (KBB) is a metric used to assess the quality of estimated reference point locations in object tracking tasks²¹⁷. Unlike the standard Keypoint Threshold metric, KBB adapts the pixel-distance threshold to the pixel size of the instrument manipulator. This adaptation is crucial to provide reliable and meaningful evaluations, especially when averaging the results over several images with different resolutions and instrument sizes. The KBB metric uses a tightly cropped, axis-aligned bounding box that contains all reference points of the instrument in the respective frame. This bounding box is determined based on the ground truth annotations of the instrument's reference points. The goal is to ensure that the threshold is appropriately scaled according to the size of the instrument manipulator in the image. For a given reference point j , located at position $j \in R^2$, and its corresponding ground truth annotation $\hat{j}$, KBB evaluates the correctness of the estimation by computing the pixel distance between the estimated and ground truth positions. The estimation j is considered correct if the pixel distance between j and $\hat{j}$ is smaller than a scaled threshold, which is calculated as:

$$(Equation 7): ||j - \hat{j}|| < \alpha \cdot \max(h, w)$$

where h and w are the width and height of the bounding box around the instrument, given by the ground truth. The scaling factor $\alpha \in R$ is used to adapt the threshold to the instrument's size and pixel resolution of the image. It is important to note that KBB requires the ground truth annotations of all reference points of the instrument. However, the evaluation is still applicable even if only one reference point is estimated. By using KBB, the metric accounts for variations in instrument sizes and pixel resolutions, providing a more accurate assessment of the tracking performance. It offers a reliable means of comparing results across different images and instrument manipulator sizes.

3.5 A Marker-based Probe Detection Methodology

3.5.1 Optical Tracking using 2D Markers

Optical tracking using 2D markers is a powerful and widely employed method for accurately and non-invasively tracking the position and orientation of objects in various fields, ranging from medical interventions to virtual reality and motion capture applications. The technique

leverages optical sensors and computer vision algorithms to precisely locate and monitor 2D markers, providing real-time spatial information with high accuracy and minimal latency.

In this method, small, distinct markers with known spatial coordinates are placed on the object of interest. These markers can take the form of simple geometric shapes, such as squares or circles, or feature-specific patterns that are easily distinguishable from the surrounding environment. The optical tracking system captures images or video of the markers using one or more cameras, and specialized software then processes the acquired data to calculate the position and orientation of each marker in 2D space.

Optical tracking using 2D markers offers several key advantages. First, it enables a cost-effective and efficient solution for tracking objects, as the markers are relatively easy to fabricate and attach. Second, the method allows for a wide range of applications, as it can be adapted to track objects of various sizes and shapes. Moreover, optical tracking excels in scenarios where high precision and low latency are crucial, making it an ideal choice for real-time applications, such as surgical navigation. However, challenges remain, particularly in environments with occlusions, complex backgrounds, or marker occlusion due to the object's movement or occlusion from blood and body fluids during in-vivo applications. Moreover, ensuring the sterilisation of these markers is paramount to maintain a sterile surgical environment and avoid contamination²¹⁸.

3.5.2 *Optical DRS Probe Tracking using 2D Markers*

In the context of surgical instrument tracking, the choice of optical tracking sub-category is a crucial decision that directly impacts the tracking system's performance and usability. Optical tracking using 2D fiducial markers presents itself as a compelling and efficient option, offering several advantages over other alternatives^{219–223}. In this section, we explore why 2D fiducial markers prove to be the optimal choice from the surgeon's perspective.

In this approach, small 2D fiducial markers are securely attached to the exterior of the surgical probe. These markers can take the form of simple geometric shapes or unique patterns, which are easily detectable by the optical tracking system. The attachment of these markers has minimal impact on the surgical procedure, and from the surgeon's perspective, nothing substantial changes other than the colour of the probe.

One of the significant advantages of using 2D fiducial markers is the ease of implementation and convenience. Compared to 3D markers, which might require complex and intrusive modifications to the probe, 2D markers can be affixed externally without disrupting the probe's form and function. As a result, the surgeon's workflow remains largely unaffected, simplifying the integration of the tracking system into surgical procedures.

Furthermore, the appearance of 2D markers remains consistent under both in-vivo and ex-vivo settings. This stability eliminates the challenges associated with marker-less systems that might struggle with drastic changes in appearance due to different imaging conditions. With 2D fiducial markers, the tracking system can accurately and consistently detect and follow the markers, ensuring reliable and robust instrument tracking throughout the surgical process.

3.5.3 *Related Works*

In their research, Zhang et al.²²⁴ propose a novel method for tracking a handheld ultrasound scanner using a laparoscope and a specially designed 2D marker. The marker is designed with rows of chessboard-vertices and circular dots, allowing it to be fitted onto a cylindrical probe. This unique marker design facilitates accurate tracking of the ultrasound scanner's rotation along its axis (the roll). To effectively track the marker's features, the authors describe a segmentation method that involves analysing multiple binarized images using different thresholds. Binarization is the process of converting grayscale images into binary images, where pixels are classified as either black or white based on predefined thresholds.

Jayarathne et al.²²⁵ build upon the methodology proposed by Zhang et al.²²⁴, aiming to enhance the tracking results for handheld ultrasound scanners in scenarios where the probe may be occluded. They introduce novel techniques that leverage temporal information and a stereo imaging system to address the challenges posed by occlusion during in-vivo surgeries.

In their work, Hulke and Gupta²²⁶ propose a system for tracking laparoscopic tools using coloured markers. Their approach involves a calibration process that employs a video of a moving chessboard grid pattern, along with the algorithm presented by Zhang²²⁷. By utilizing the HSV colour-space, they successfully detect the markers and leverage the

calibration matrix, along with the position and size of the markers, to determine the pose of the laparoscopic tool.

In their research, Zhang and Payandeh²²⁸ introduce a method for tracking a laparoscopic tool, with a specific focus on its position and orientation, excluding the rotation along the tool axis. To achieve this, they affix a marker to the cylindrical tool, creating three stripes that form a ring around the tool. While only one ring was technically required for tracking, the researchers use multiple rings to account for potential occlusion that may occur during the laparoscopic procedure.

3.5.4 *Marker Design*

In the context of tracking an optical surgical tool, choosing a distinct and easily distinguishable marker colour is essential, as the non-uniqueness of the probe's natural colour can lead to tracking inaccuracies. To address this, an artificial colour is employed as a coding method for the probe, allowing for effective and reliable tracking during surgical procedures^{222,223}. While pattern markers, such as a checkerboard-like pattern, can also be used for tracking, colour markers are generally preferred over pattern markers²²⁹ for several reasons.

Firstly, a unique colour marker is highly effective in minimising tracking inaccuracies. Unlike a pattern marker, which relies on the recognition and interpretation of complex patterns, a unique colour marker is easier for tracking systems to identify quickly and with high precision. In addition, a colour marker offers superior robustness to variations in lighting conditions. Surgical environments can be dynamic, with regular changes in ambient lighting. A pattern marker, such as a checkerboard, may be susceptible to lighting changes that obscure its distinctive pattern, leading to tracking errors. In contrast, a colour marker remains consistent and recognisable in varying lighting conditions, increasing its reliability during surgical procedures. Finally, colour markers are more robust to occlusion and partial visibility than pattern markers. In a surgical environment, the probe may occasionally be partially blocked or obscured by surgical instruments, tissue, or other objects. Pattern markers require the entire pattern to be visible for detection and tracking, which can be a challenge in cluttered and complex environments²²¹. In contrast, colour markers can be detected and tracked even if only a small portion of the marker is visible, making them more robust to occlusion.

To determine the optimal colour for the marker, an analysis of the colour distribution in real laparoscopic images is conducted^{230,231}. The HSV (hue-saturation-value) colour space is utilised, as it directly relates to the colour signature of the images. The hue (H) and saturation (S) values are particularly important in this analysis as they represent the hue and intensity of the colours present in the images, respectively. Figure 14 illustrates a typical colour distribution in the H-S plane of laparoscopic images captured during a surgical procedure. As the background in these images mainly consists of biological tissue, the hue ranges from yellow to purple. The highlighted area in the plot represents the distribution of pixels corresponding to colours that do not occur naturally in the images. This range of colours can be utilized as the artificial colour for the marker²²². By selecting pixels within the highlighted range of hue and saturation values, the system can automatically segment the coloured marker. This approach is effective because the colours within this range are exclusively related to the surgical instrument (i.e., the DRS probe) and do not include any anatomical features.

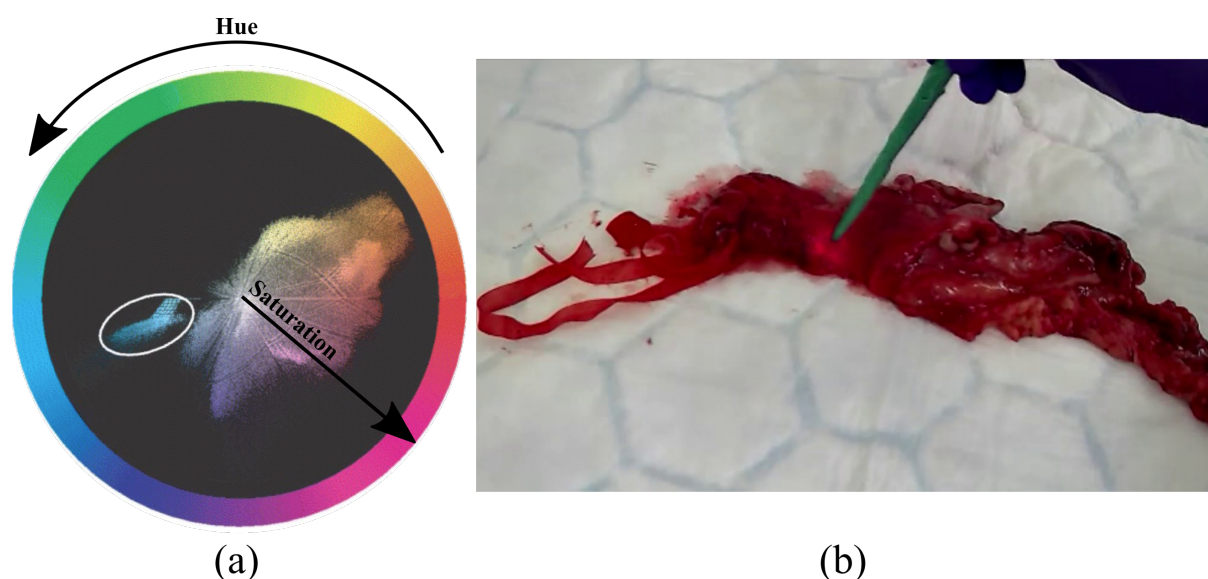


Figure 14: (a) Average colour distribution in the $H-S$ plane for 3,000 laparoscopic images, together with the $H-S$ occupation of the potential colour marker in the highlighted oval²³⁰. Additionally, it's evidence that green hues could also be used as potential marker. (b) DRS fibre probe with the green colour marker attached.

The use of the HSV colour space in the colour analysis is particularly advantageous. The HSV representation separates colour information from intensity, making it more robust to variations in lighting conditions. As surgeries often involve varying lighting conditions within the surgical environment, the HSV colour space helps maintain the marker's distinctive colour signature regardless of changes in illumination.

To ensure the coded colour is distinguishable from the colours present in the image, two alternative colours – cyan and green – were selected. Selecting pixels with H-S values within these colour areas ensures that the coloured marker can be effectively segmented, as these colours belong exclusively to the probe and not the surrounding tissue.

The colour marker, attached to the distal end of the shaft of the DRS probe, can take the form of a printed colour strip or a surgical instrument identification marking tape. The marker wraps around the cylindrical probe, with the long side aligned parallel to the probe's axis. The short side of the marker should be at least the circumference of the probe, ensuring that the marker can be tracked accurately regardless of the probe's rotation. By wrapping the marker around the cylindrical probe with the long side parallel to the probe's axis, the tracking system gains the ability to monitor the tool's position and orientation effectively. The location of the marker close to the tip of the cylindrical probe is strategically chosen to enhance tracking precision. The tip of the tool is often the most critical region during surgery, as it is responsible for interacting with tissues and organs. By positioning the marker at this critical point, the tracking system can precisely monitor the tool's movements, enabling surgeons to navigate with confidence and perform delicate procedures with greater accuracy. The use of printed colour strips or surgical instrument identification marking tape as the marker ensures ease of application and durability during the surgical procedure.

One crucial aspect of the chosen artificial colour is its ability to be automatically segmented from the optical images. This segmentation process is vital for accurately identifying the marker's position in the image frames and determining the tool's pose. The unique colour areas chosen, cyan and green, allows for straightforward and reliable marker segmentation, simplifying the tracking process and contributing to real-time feedback for the surgeon. Moreover, the two alternative marker colours, cyan and green, offer flexibility in adapting to different surgical settings. Depending on the surgical environment's lighting conditions or potential colour interference, surgeons can choose between the two colours to ensure optimal marker visibility and tracking performance.

Incorporating the artificial colour marker into the tracking system enables precise and continuous tracking of the optical surgical tool, even in the presence of occlusion. During surgery, parts of the tool may become obscured by organs or tissue, causing temporary occlusion. However, the use of multiple markers and the distinct artificial colour facilitate consistent tracking, ensuring uninterrupted monitoring of the tool's movements.

3.5.5 Camera Calibration

The overall methodology for using a monocular red/green/blue (RGB) camera (BFS-U3-50S5C-C, Teledyne FLIR) in a 2D colour-based optical marker tracking system is summarized in Figure 15.

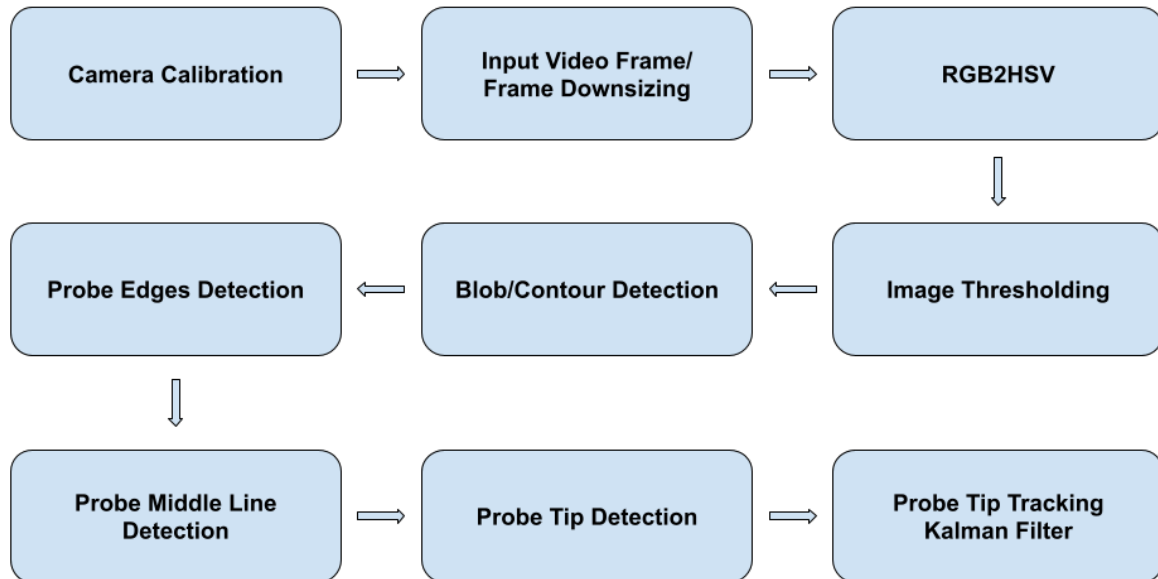


Figure 15: DRS fibre probe detection and tracking pipeline.

The calibration process for the 2D colour-based optical marker tracking system employs the 2D calibration object method. This method requires capturing observations of a planar pattern, such as a chessboard, from different orientations and positions using the monocular RGB camera²³². To begin the calibration process, a set of pre-captured images containing a chessboard pattern is analysed. These images are captured by the camera and serve as the input for the calibration function.

The calibration function then performs an estimation of the intrinsic parameters of the camera, represented by the 3×3 intrinsic matrix K . These intrinsic parameters include essential details such as the focal length of the camera, which is crucial for accurately converting 2D image points into 3D world coordinates.

Additionally, the calibration function determines the extrinsic matrix, which comprises the 3×3 rotation matrix R and the 3×1 translation vector t . The extrinsic matrix represents the camera's position and orientation with respect to the world coordinate system. By estimating the rotation and translation, the calibration process establishes the camera's

relationship to the calibration object, which is essential for accurate pose estimation during the tracking process.

By calibrating the camera using the 2D calibration object method, the tracking system gains the ability to accurately project 2D image points of the artificial colour marker onto 3D world coordinates. This calibration step ensures that the marker's position and orientation in the real world can be precisely determined based on its appearance in the camera's view. As a result, the optical marker tracking system can provide accurate and reliable tracking information, enabling effective surgical navigation and improved surgical interventions.

3.5.6 *Marker Detection & Segmentation*

To calculate the 2D probe tip position in the image plane, the location of the DRS probe needs to be determined. This can be achieved by identifying the midline of the DRS probe from the two edges visible in the image. The midline represents the centreline of the probe, and the tip point of the probe is expected to lie somewhere along this line²³³.

The initial step of the algorithm involves the resizing of the input video frames to achieve a higher frames-per-second (FPS) rate. Resizing the frames from their native resolution of 1920×1080 to a downsized resolution of 640×480 allows for faster processing and real-time performance. The significance of this optimization lies in the ability to handle a higher frame rate, providing surgeons with smoother and more responsive tracking feedback during the surgical procedure. The impact of frame resizing on the computational efficiency and tracking accuracy is thoroughly evaluated, shedding light on the trade-offs involved in achieving real-time tracking capabilities.

To address the challenge of high-frequency noise present in surgical video frames, blurring techniques are employed as a noise reduction strategy. High-frequency noise can adversely affect the accuracy of the tracking algorithm, leading to suboptimal tracking results. By applying blurring operations on the downsized video frames, the noise is effectively reduced, while preserving essential features and structures of the image. A comprehensive analysis of different blurring techniques, such as Gaussian and bilateral filtering, is conducted to identify the most suitable approach for maintaining tracking precision while mitigating noise artifacts.

By resizing and blurring the video frames, the algorithm achieves a more efficient and noise-resistant processing pipeline, enabling real-time tracking of the artificial colour marker attached to the DRS probe. The downsizing of the video frame not only improves the algorithm's computational efficiency but also ensures that the system can handle a higher frame rate, providing smoother and more responsive tracking feedback to the surgeon during the surgical procedure.

The next step in the optical marker tracking methodology involves converting the video frame from the RGB colour space to the HSV colour space. The HSV colour space is chosen because it provides a more intuitive representation of colour, separating the colour information into three components: hue, saturation, and value. In the HSV colour space, the hue (h) component represents the dominant colour information, ranging from 0 to 360 degrees, with red at 0 degrees, green at 120 degrees, and blue at 240 degrees. The saturation (s) component measures the intensity or purity of the colour, ranging from 0 (completely desaturated) to 1 (fully saturated). The value (v) component represents the brightness or lightness of the colour, ranging from 0 (black) to 1 (white).

The HSV frame is then thresholded based on pre-defined lower and upper boundaries specific to the cyan or green colour of the marker. These boundaries are established based on prior analysis of the colour distribution of the marker in real laparoscopic images. The thresholding process involves keeping only those pixels in the HSV frame whose hue (h) and saturation (s) values fall within the specified range ($h_{min} < h < h_{max}$, $s_{min} < s < s_{max}$), while discarding the rest.

By applying this thresholding technique, the algorithm isolates and retains the principal desired colour of the marker in the frame. This step helps in effectively segmenting the marker from the background and other colours present in the image, improving the accuracy and robustness of the tracking process. The thresholded HSV frame now contains only the pixels representing the colour of the marker, which allows for subsequent marker detection and tracking with greater precision. However, to ensure a more robust and accurate segmentation of the marker and to eliminate spot noise and fill any gaps in the mask, morphological operations are employed.

The first morphological operation applied is erosion. Erosion involves the removal of pixels at the boundaries of regions in the binary mask, effectively shrinking the regions. This operation helps to eliminate small, isolated spots or noise present in the segmented frame,

reducing false positives and enhancing the overall quality of the mask. Following erosion, the next morphological operation is dilation. Dilation, on the other hand, expands the boundaries of regions in the binary mask. This operation helps to fill in gaps or holes in the mask caused by imperfections or irregularities in the marker's appearance. By dilating the mask, the segmented region becomes more connected and complete, improving the accuracy and completeness of the marker detection.

After applying both erosion and dilation, the segmented frame applied to the RGB image. The resulting frame now contains the marker in its original colour, with improved noise reduction and filled gaps in the mask. The combination of thresholding and morphological operations ensures a robust and precise segmentation of the artificial colour marker, allowing for more reliable tracking of the DRS probe throughout the surgical procedure.

The next stage of the optical marker tracking process involves identifying and analysing the various contours in the segmented mask, which represent the segmented objects in the frame. Among these contours, the largest one is retained as it is likely to correspond to the detected colour marker attached to the optical probe.

Contour detection refers to identifying the curve that joins all continuous points with the same colour or intensity in the segmented mask. The largest contour in terms of area corresponds to the detected colour marker of the probe. By finding and retaining this contour, the system successfully localizes the position of the marker within the frame. Following contour detection, a bounding box is created around the detected contour. This bounding box acts as the smallest rectangular area that completely encloses the marker's contour. The corner points of the bounding box surrounding the largest contour are then extracted and utilized in conjunction with the Hough transform to detect the edges of the probe. The Hough transform is a popular technique for detecting simple shapes, such as lines, in an image. In this case, it helps identify the edges of the DRS probe from the corner points obtained from the bounding box.

Once the edges of the probe are extracted, the next step is to calculate the midline of the probe. The midline represents the centreline of the probe and is crucial for accurately determining the position and orientation of the DRS probe. The standard description of a line in the x-y plane ($y = mx + b$) is employed to calculate the midline, where m represents the slope and b represents the y-intercept of the line. The slope and y-intercept of the line are calculated using the corner points of the bounding box surrounding the largest contour.

The midline is then calculated as the line passing through the centre of the probe and perpendicular to the edges of the probe.

By using these techniques, the system can precisely identify the location of the artificial colour marker on the DRS probe and calculate the midline of the probe. This information is essential for real-time tracking of the probe's position and orientation, enabling effective surgical navigation and providing valuable feedback to the surgical team during minimally invasive surgeries.

3.5.7 *DRS Probe Pose Estimation using PCA*

To accurately locate the tip of the DRS probe, a 1D edge detection process is conducted along the segmented midline of the probe in the binary mask obtained during the segmentation phase. The segmented midline represents the centreline of the probe, and the edge detection aims to identify the exact point where the probe's tip is located along this line. The binary mask obtained during segmentation consists of pixels with binary values, where the probe is represented by white pixels and the background by black pixels. Since the probe is uniformly white, the boundary between the probe and the background is where there is a sudden change in intensity from black to white or vice versa. This sudden change in intensity corresponds to the tip of the probe.

To search for the tip location, both the x and y directions in the image plane are considered. In the x -direction, the 1D edge detection algorithm starts by selecting a region of interest (ROI) along the segmented midline of the probe. The ROI is a small segment of the midline where the tip is likely to be located. The algorithm then scans this ROI along the horizontal axis of the image plane, pixel by pixel. It examines the intensity values of adjacent pixels in the ROI and identifies the point where there is a significant change in intensity from black to white or vice versa. When the algorithm encounters the transition from black to white, it marks this position as a potential candidate for the tip location. Similarly, when the transition is from white to black, it also considers this as a potential tip candidate. This is because the probe's tip might appear as a bright spot on a dark background or vice versa, depending on the lighting conditions and the orientation of the probe.

In the y -direction, the process is similar but performed along the vertical axis of the image plane. The algorithm selects another ROI, perpendicular to the midline, and scans it

pixel by pixel along the vertical direction. It again looks for the abrupt change in intensity that indicates the presence of the tip.

In order to find in which direction we should search for the location of the probe's tip, we employ the principal component analysis (PCA) method. PCA is a statistical procedure that is commonly used for dimensionality reduction²³⁴. The goal of PCA is to transform a dataset $\mathbf{X}$ of dimension p to another dataset $\mathbf{Y}$ of smaller dimension L . This can be done by finding the matrix Karhunen-Loeve transform (KLT) of matrix $\mathbf{X}$.

$$(Equation 8): \mathbf{Y} = KLT\{\mathbf{X}\}$$

As mentioned above, our aim is to reduce the initial dataset of observations of p variables to a dataset of observations of L variables, with $L < p$. Suppose that the initial dataset is arranged as a set of n data vectors $x_1, \dots, x_n$. After having organised our dataset, we subtract the mean for each point. We find the mean along each dimension $j = 1, \dots, p$ and then store the mean values into a vector $\mathbf{u}$ of dimensions $p \times 1$.

$$(Equation 9): \mathbf{u}[j] = \frac{1}{n} \sum_{i=1}^n \mathbf{X}[i, j]$$

Mean subtraction is an integral part of the solution towards finding a principal component (PC) basis that minimises the mean square error of approximating the data. Hence, we proceed by centering the data. In this step, the mean vector $\mathbf{u}$ is subtracted from each row of the data matrix $\mathbf{X}$ and stored in a $n \times p$ matrix $\mathbf{B}$.

$$(Equation 10): \mathbf{B} = \mathbf{X} - \mathbf{h}\mathbf{u}^T$$

where $\mathbf{h}[i] = 1, i = 1, \dots, n$.

The next step of the procedure consists of the calculation of the covariance matrix $\mathbf{C}$ with dimensions $p \times p$.

$$(Equation 11): \mathbf{C} = \frac{1}{n-1} \mathbf{B}^* \cdot \mathbf{B}$$

where $*$ is the conjugate transpose operator.

By calculating the covariance matrix in the previous step, we can get the eigenvectors and eigenvalues of it. We compute the matrix $\mathbf{V}$ of eigenvectors which diagonalises the covariance matrix $\mathbf{C}$.

$$(Equation 12): \mathbf{V}^{-1} \mathbf{C} \mathbf{V} = \mathbf{D}$$

where $\mathbf{D}$ is the diagonal $p \times p$ matrix of eigenvalues of the covariance matrix $\mathbf{C}$.

$$(Equation 13): \mathbf{D}[k, l] = \begin{cases} \lambda_k, & k = l \\ 0, & k \neq l \end{cases}$$

where λ_j is the j -th eigenvalue of the covariance matrix $\mathbf{C}$.

Then, the eigenvalues and eigenvectors are ordered and paired such that the j -th eigenvalue corresponds to the j -th eigenvector. The eigenvector with the largest eigenvalue gives us the direction along which the data has the largest variance.

The input data to the PCA algorithm is the area of the segmented colour marker. After the calculation of the mean, covariance matrix, eigenvectors, and eigenvalues the angle of the first eigenvector, the one with the largest eigenvalue, is calculated. Let us denote the first eigenvector as $eig1$, then the angle of the probe can be computed as follows:

$$(Equation\ 14): angle = \arctan \frac{y_{eig1}}{x_{eig1}}$$

Given the orientation of the probe we decide in which direction to search for the location of the tip at each input frame.

After performing edge detection, the algorithm obtains the coordinates (x, y) of the tip. These coordinates represent the exact position of the tip in the image plane. The algorithm can then use this information for further calculations, such as tracking the movement of the probe during the surgical procedure.

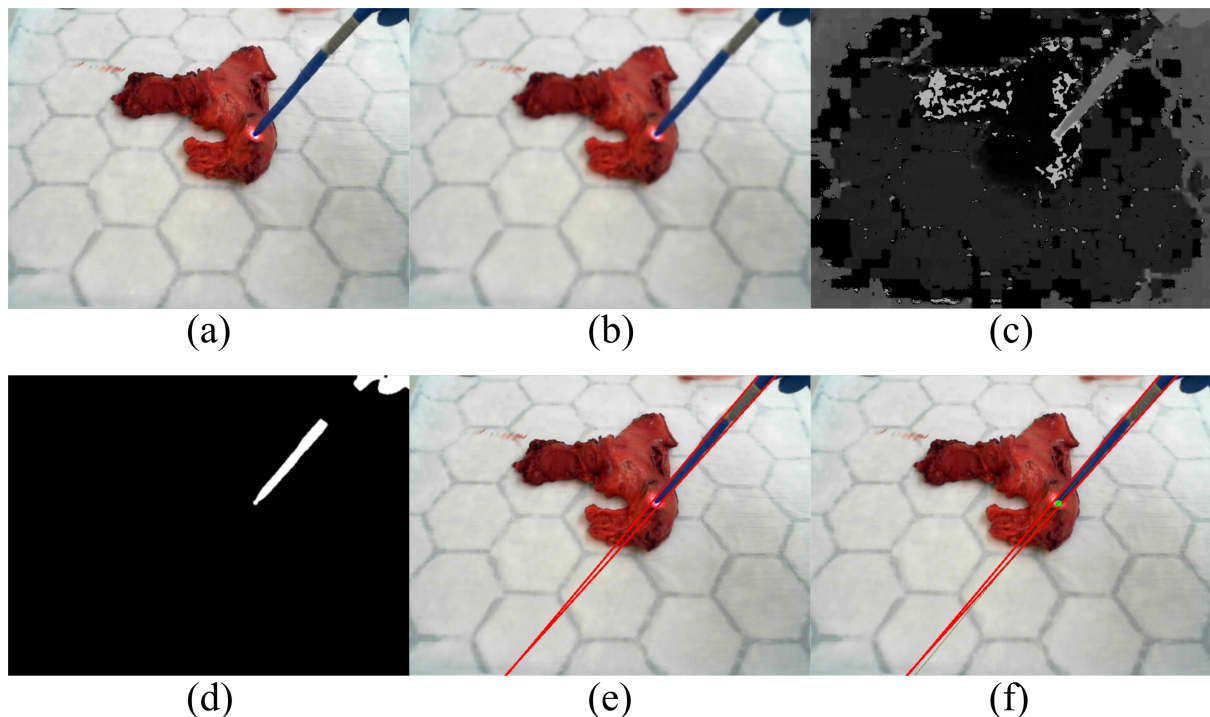


Figure 16: DRS fibre probe detection workflow. (a) Example input video frame. (b) Blurred frame. (c) Hue channel of the blurred input frame. (d) Binary mask result of HSV segmentation. (e) Detected edge lines of the probe. (f) Detected tip point of the probe.

3.5.8 *DRS Probe Tip Tracking using a Kalman Filter*

In the context of tracking probe's motion using the previously described methodology, a simple approach involves extracting edges from the images and calculating the tool's tip position frame by frame. While this method may provide some initial results, it has certain limitations and drawbacks that make it less efficient and susceptible to the effects of random noise in the image, such as variations in illumination.

Firstly, the edge extraction process relies on detecting sudden changes in pixel intensities, which are indicative of object boundaries. However, edge detection alone may not always accurately represent the true shape and orientation of the tool, especially in the presence of noise or partial occlusion. Secondly, frame-by-frame tip calculation solely based on edge information can result in jittery and erratic motion estimation. Due to noise and variations in the imaging conditions, the edge extraction may yield inconsistent or noisy results from one frame to another, leading to unstable tracking. Furthermore, this approach may suffer from false detections or missed detections of the tool's tip, particularly in challenging lighting conditions or when the tool's appearance changes significantly.

To address these noise effects and improve probe tracking accuracy, the proposed methodology employs a more sophisticated approach using a Kalman filter (KF) to estimate the location of the probe's tip²³⁵. The Kalman filter is a recursive estimation technique used to predict and update the state of an object based on a series of measurements. In our case, the object being tracked is the DRS probe, and the state corresponds to the location of the probe's tip. The measurements are the coordinates of the tip of the probe obtained from the optical marker tracking system.

The Kalman filter operates in a continuous loop, continuously updating its estimate of the probe's tip location as new measurements become available. It takes into account the uncertainty in the measurements and the dynamics of the probe's movement to make predictions for the current state.

At each time frame, the Kalman filter is provided with the location of the probe's tip as a measurement. It combines this measurement with the previous estimated state and the system's dynamic model to generate a prediction for the current state. The Kalman filter effectively fuses the noisy measurements and the dynamic model of the probe's movement to produce a more accurate and robust estimation of the tip's location. By incorporating past

estimates and current measurements, the Kalman filter helps mitigate the impact of measurement noise and provides a smoother trajectory for the probe's tip.

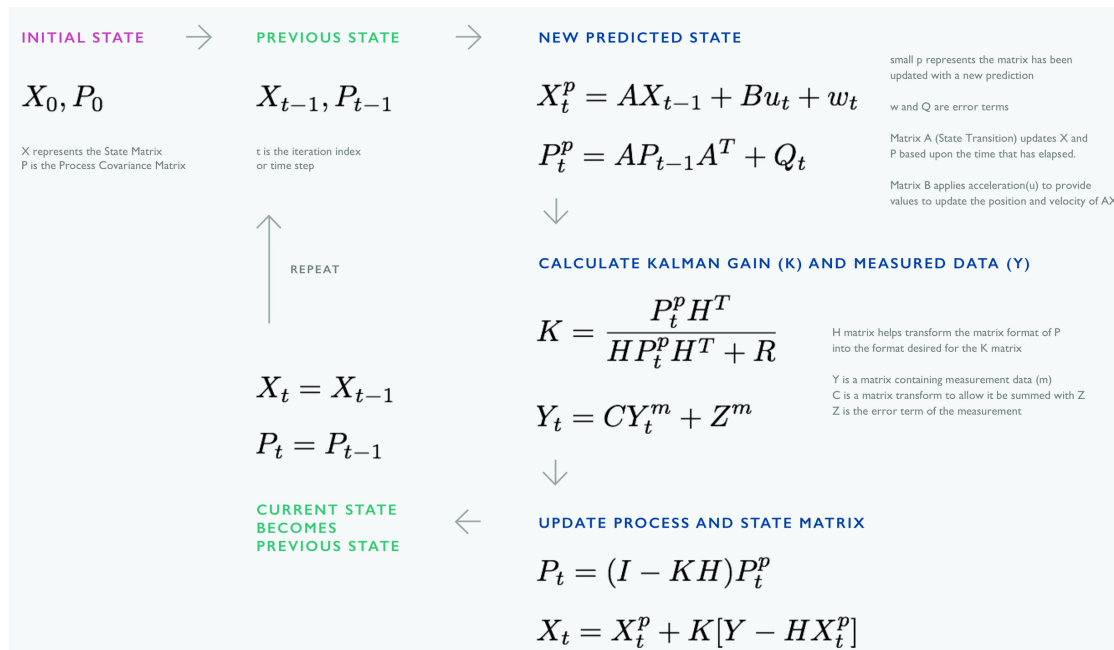


Figure 17: Flowchart of a Kalman Filter Matrix process (<https://medium.com/@jaems33/understanding-kalman-filters-with-python-2310e87b8f48>)

The flowchart above shows the recursive process. The Kalman filter performs two essential steps: prediction and correction. The primary objective of the KF is to keep track of the state matrix **X**, which contains the current value of the measurements being tracked, such as the probe's tip location. Additionally, the process covariance matrix **P** represents the predictive error associated with these measurements. In the prediction step, the KF utilizes the state transition matrix to project the current state matrix **X** and process covariance matrix **P** to their new values based on the system dynamics and the current position. This prediction step estimates a new position for the tracked object (i.e., the probe's tip) as well as a new covariance to account for uncertainties in the object's motion. The Kalman gain is then computed in the correction step, which combines the predicted state and covariance with the observed measurements from the optical marker tracking system. The Kalman gain represents a weighting factor that determines how much the predicted state, and the new measurements contribute to the updated state estimation. In the correction step, the KF incorporates the observed measurements by adjusting the predicted state and covariance matrix based on the Kalman gain. This update of the current state using the Kalman gain, in conjunction with the measured data at that time frame, allows the KF to adaptively fuse the predictions with real-time measurements, resulting in an improved and more accurate

estimation of the object's state. By iteratively performing the prediction and correction steps for each time frame, the KF continuously refines its estimates of the probe's tip location, effectively compensating for measurement noise and uncertainties in the system. This recursive process ensures that the estimated state remains updated and aligns with the real-time measurements, providing a smooth and reliable tracking of the probe's motion during the surgical procedure.

Kalman Gain

The Kalman gain plays a critical role in the Kalman filtering process and is used to determine the balance between the new measurements and the current estimate when updating the state. To calculate the Kalman gain, two pieces of information are required:

- The error in the estimate: This refers to the discrepancy between the predicted state of the system (based on the previous estimate) and the actual measurements obtained from the optical marker tracking system. It represents the uncertainty in the current estimate.
- The error in the measurement/data: This represents the uncertainty or noise in the real-time measurements obtained from the optical marker tracking system. It accounts for variations or inaccuracies in the acquired data.

The Kalman gain is calculated based on these two sources of error, and it essentially determines how much weight should be assigned to the new measurements when updating the current estimate. The gain serves as a relative importance factor that balances the contributions of the estimate's error and the measurement's error in the updating process.

$$(Equation 15): K = \frac{E_{est}}{E_{est} + E_{mea}}$$

When the error in the estimate is smaller compared to the error in the measurement, the Kalman gain will emphasize the estimate more, as it is considered more reliable and trustworthy. Conversely, when the error in the data is smaller, the gain will give greater weight to the new measurements, as they are deemed more accurate.

$$(Equation 16): x_t = x_{t-1} + K[p - x_{t-1}]$$

The Kalman gain is a number between 0 and 1 and acts as a weight in the update equation for the current estimate. In the update step, the Kalman gain is multiplied by the

difference between the new measurement and the previous estimate, and this value is then added to the previous estimate to obtain the updated estimate.

$$(Equation 17): Est_t = Est_{t-1} + K[M - Est_{t-1}]$$

As the Kalman filter iterates and refines its estimates over time, the Kalman gain adapts based on the available data. Initially, the gain might be relatively high to allow for rapid updates based on new data. However, as the filter converges and the model makes more accurate estimates, the Kalman gain will tend to decrease, resulting in more stable and precise estimates.

$$(Equation 18): E_t = [1 - K]E_{t-1}$$

The Kalman gain effectively controls the speed at which the estimated value converges toward the true value. When the gain is high, new data has a strong impact on the model, leading to faster convergence. Conversely, a small Kalman gain implies that the model is more confident in its current estimate and is less affected by new data, resulting in slower updates.

State Matrix

In the Kalman filter, the state matrix represents the object being tracked and contains information about its current state, such as position and velocity.

$$(Equation 19): X_t^p = AX_{t-1} + Bu_t + w_t$$

The state matrix is updated based on the previous state, the control variable matrix (**u**), and a noise term (**w**) that accounts for any uncertainties or errors in the process.

$$(Equation 20): X = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix}$$

The update of the state matrix is performed using the state transition matrix (**A**). Matrix **A** captures the relationship between the current state and the next state over a time step (delta t). When matrix **A** is multiplied by the state matrix (**x**), it predicts the next state of the object, including its position and velocity at the next time step.

$$(Equation 21): x = x_0 + \dot{x}t + \frac{1}{2}\ddot{x}t^2$$

$$(Equation 22): d = d_0 + vt + \frac{1}{2}at^2$$

To update the position of the object, a time step is taken, and the velocity is added to the previous position. The velocity remains the same during this prediction step. However, it's important to note that if there was any acceleration affecting the object's motion during the time step, the velocity would change.

$$(Equation\ 23): AX = \begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} = \begin{bmatrix} x + \Delta T \dot{x} \\ \dot{x} \end{bmatrix}$$

To account for acceleration, the Kalman filter incorporates the control variable matrix (**B**) in the update equation. The **B** matrix models part of the kinematics equation, where velocity and acceleration are multiplied by time (delta t). When the control variable matrix (**u**), representing the acceleration, is multiplied by matrix **B** and added to the product of matrix **A** and **x**, it results in a change to both the position and velocity of the object due to acceleration.

$$(Equation\ 24): Bu = \begin{bmatrix} \frac{1}{2} \Delta T^2 \\ \Delta T \end{bmatrix} [a] = \begin{bmatrix} a \frac{1}{2} \Delta T^2 \\ a \Delta T \end{bmatrix}$$

Observation Data

In the Kalman filter process, the observation data is combined with the state matrix, which contains the most recent prediction, to update and correct the estimate of the object's state. The observation is represented by matrix **C** times the variables that are observed, and it includes the measurement noise (**Z**) that may be present in the data. Matrix **C** is a key component of the observation model and is designed to map the variables being observed (**Y**) to the state matrix (**X**). The shape and entries of matrix **C** depend on the specific variables that we want to observe from the state matrix. If we wish to examine all the variables in **Y**, then matrix **C** would largely be an identity matrix, indicating a direct observation of all state variables.

$$(Equation\ 25): Y_t = CY_t^m + Z_t$$

State Covariance and Measurement Covariance Matrix

The state covariance matrix is a crucial component of the Kalman filter and represents the uncertainty or error associated with the estimate of the object's state. As the Kalman filter iterates through the prediction and correction steps, the state covariance matrix is continually updated to account for the inherent noise, errors, and uncertainties in the estimation process.

The state matrices in the Kalman filter are estimates of the subsequent states of the object being tracked. However, since all measurements and predictions have some level of noise and uncertainty, the state covariance matrix (**P**) is introduced to quantify and capture this uncertainty. The elements of the state covariance matrix represent the variances and covariances between different state variables (e.g., position) in the estimation.

$$(Equation\ 26): P = \begin{bmatrix} \sigma_x^2 & \sigma_{x\dot{x}} \\ \sigma_{\dot{x}x} & \sigma_{\dot{x}}^2 \end{bmatrix}$$

To improve the accuracy of the Kalman filter, a process noise covariance matrix (**Q**) is introduced to adjust the state covariance matrix during the prediction step. The process noise covariance matrix accounts for external factors or disturbances that affect the system being tracked but are not explicitly modelled. These factors can include environmental changes, sensor inaccuracies, or other unknown sources of noise.

The measurement covariance matrix (**R**) represents the error or uncertainty in the measurements obtained from the sensor or observation system. It quantifies the accuracy and reliability of the measurements taken for the object being tracked. Since real-world measurements are never perfectly precise, the measurement covariance matrix accounts for the inevitable errors and uncertainties in the measurement process. By introducing the measurement covariance matrix (**R**) into the Kalman filter, we explicitly account for the uncertainty in the measurements. This ensures that the filter adapts and incorporates the observed measurements appropriately, even if they have some degree of error. It prevents the state covariance matrix (**P**) from becoming zero, which would imply that the filter completely ignores the observation data and relies solely on the predictions.

$$(Equation\ 27): P_k = AP_{k-1}A^T + Q$$

$$(Equation\ 28): K_k = \frac{P_k H^T}{H P_k H^T + R}$$

The **A** and **H** matrices in the Kalman filter are used to define the state transition and observation models, respectively. The state transition matrix (**A**) helps update the state estimate based on the previous state and any control inputs (such as acceleration). The observation matrix (**H**) maps the state vector to the observation space to predict the measurements that would be obtained based on the current state estimate. These matrices are crucial for formulating the Kalman filter equations and enabling the integration of both prediction and measurement data.

Covariance is a measure of the joint variability or co-variability of two random variables. It quantifies the degree to which two variables change together or show similar behaviour. When the values of one variable tend to increase or decrease along with the values of the other variable, the covariance is positive. Conversely, if the values of one variable tend to increase when the values of the other variable decrease, the covariance is negative. The sign of the covariance is crucial because it indicates the direction of the linear relationship between the two variables.

$$(Equation\ 29): \sigma_x \sigma_y = \sum_{i=1}^n \frac{(\bar{x} - x_i)(\bar{y} - y_i)}{N}$$

A positive covariance suggests that when one variable has higher values, the other variable tends to have higher values as well. This indicates a positive correlation between the two variables. On the other hand, a negative covariance indicates an inverse relationship between the variables, where higher values of one variable are associated with lower values of the other variable.

Overall, the use of a Kalman filter in the tracking system improves the overall performance and reliability of the optical marker tracking. It ensures that the probe's tip location is continuously updated in real time, providing surgeons with accurate and reliable information for precise instrument control during the surgical procedure.

3.5.9 *Spectroscopic Augmented Reality*

In the rapidly advancing field of medical imaging technology, there is a growing emphasis on real-time data acquisition and the integration of data visualization for faster and more reliable diagnoses. Augmented reality (AR) is emerging as a powerful tool in the operating theatre, offering the potential to fuse computer-processed images with the real surgical environment, thereby assisting surgeons in operation planning and achieving complete tumour resection in cancer surgeries.

By accurately tracking the position of the optical surgical probe in real-time, our AR system can offer various benefits to the surgical team. For example, during cancer surgery, AR can help display important information about the location of tumours and critical structures, allowing surgeons to navigate complex anatomy with greater precision and confidence.

The image-based tracking framework described in section 3.5 plays a crucial role in enabling the mapping of all spatio-temporally tracked optical biopsy sites. This mapping allows for the creation of a live augmented enhanced view that can be projected onto the image plane, providing valuable information to the surgeon during the procedure. The augmented view allows for re-targeting and re-examination of potential tumour areas, enhancing the surgeon's ability to make informed decisions during the surgery.

A key feature of this framework is the use of different colours to represent each optical biopsy site based on the probability that the acquired spectrum is classified as cancer or healthy tissue. This colour coding system provides real-time feedback to the surgical team regarding the nature of the tissue at each biopsy site. It allows for quick identification and differentiation of cancerous and healthy tissue regions, aiding in precise tumour localization and facilitating targeted resection.

The biopsy site's area is defined by combining the tracked probe's tip location, which is known from the tracking procedure, with the geometrical characteristics of the probe, such as its width. This information is used to accurately represent the spatial distribution of the optical biopsy sites on the augmented view, ensuring their alignment with the real-world surgical scene.

3.5.10 Experimental Evaluation & Results

To assess the precision and fidelity of the DRS probe tracking system, a meticulous evaluation was undertaken, involving the analysis of two distinct video recordings for each tissue specimen collected (Section 4.5). The probe's 2D tip coordinates (x; y) were subjected to a meticulous process of manual annotation by two separate individuals employing a publicly available data labelling tool (<https://labelbox.com>, Labelbox Inc., California). Subsequently, these manually annotated coordinates were juxtaposed with the corresponding tip positions estimated by the proposed algorithm (Figure 19). The evaluation process was guided by the root mean squared error (RMSE) metric, which effectively quantified the disparity between the algorithm's inferred tip location and the manually identified position.

The acquired tip positions were subsequently juxtaposed against the algorithmically inferred tip locations, as depicted in Figure 19. The RMSE between the estimated and manually registered tip coordinates exhibited values of 1.18 ± 0.58 mm (corresponding to

4.21 ± 2.07 pixels) along the x-axis and 1.05 ± 0.28 mm (equating to 3.75 ± 0.97 pixels) along the y-axis, as delineated in Table 2. The utmost observed discrepancy encompassing all 1,050 frames from both video sequences stood at 1.76 mm (or 6.28 pixels) at a working distance of 30 cm. Illustrated in Figure 19 are individual frames extracted from each video sequence, where the manually annotated and algorithm-derived tip positions are visually superimposed, each represented by distinct colours.

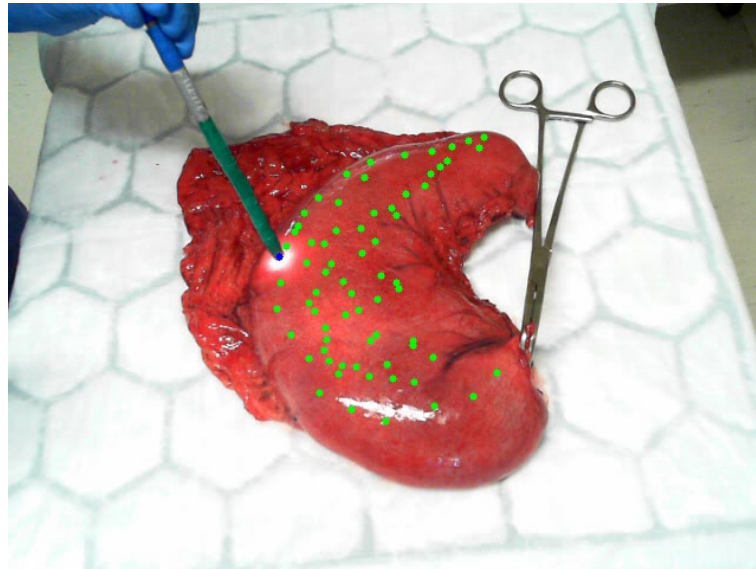
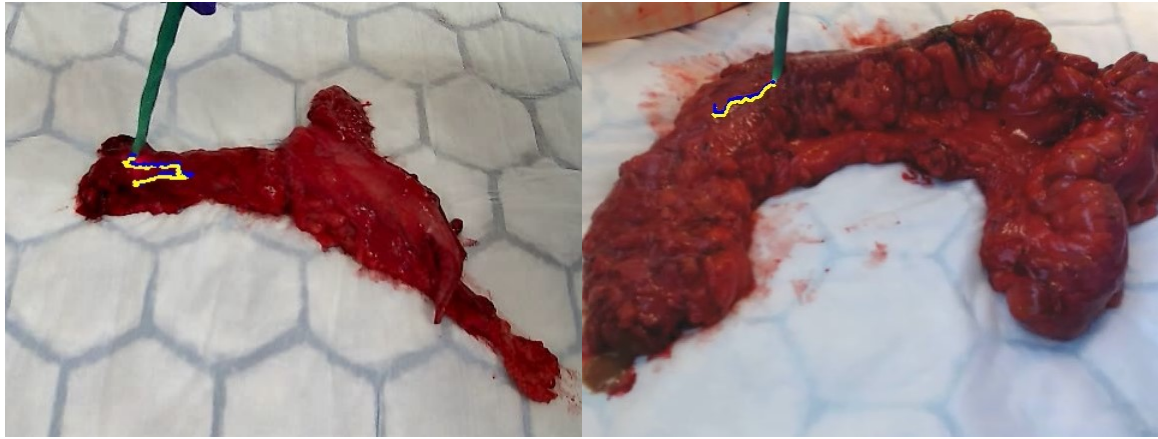
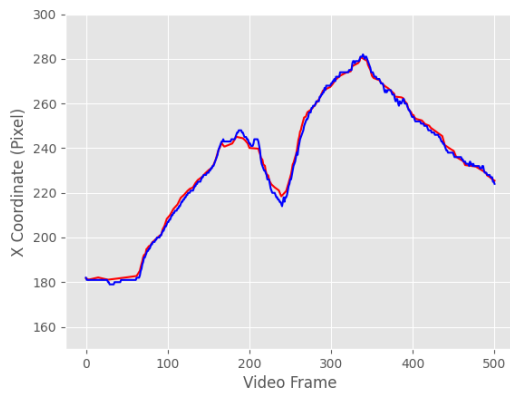


Figure 18: Illustration of DRS probe tip detection during sampling ex-vivo gastric specimen. Blue point: detected probe's tip location using the marker-based methodology. Green points: optical biopsy sites.

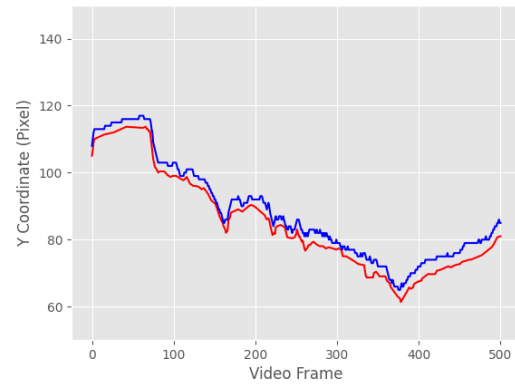


(a)

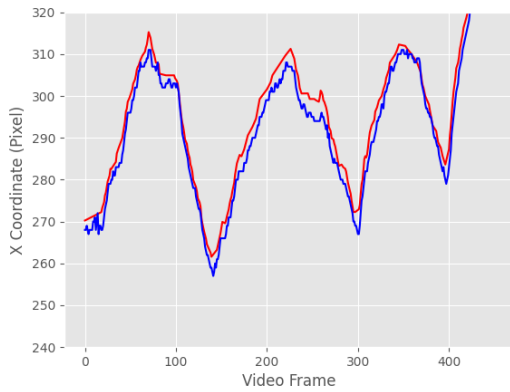
(b)



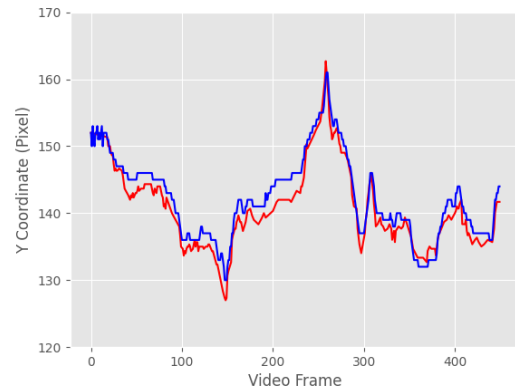
(c)



(d)



(e)



(f)

Figure 19: (a) and (b) Illustration of probe tracking error with overlaid ground truth (blue line) and estimated (yellow line) probe tip position for two video sequences (500 and 450 frames each). (c) – (f) The ground truth and estimated projected DRS probe tip pixel position for x and y axis for each of the video sequences. Blue and red lines indicate the estimated and ground truth probe tip position, respectively.

To enhance the validation process, the assessment of detection capabilities encompassed the computation of detection limits and detection rates. This involved the meticulous recording of maximal experimentally determined distances and rotation angles of the probe, as outlined in Table 3. The distance measurement was established by gauging the

separation between the camera and the probe. Meanwhile, the parameters for rotational motion limits were established about the probe's local coordinate axes, specifically encompassing roll, pitch, and yaw. During the investigation of distance limits, the probe was incrementally translated along the camera's axis until the point of detection failure was reached. Regarding the determination of rotational motion limits, the probe was situated at a standardized distance of 30 cm from the camera—a pragmatic choice commonly observed during ex-vivo tissue scanning.

Table 2: Projection error of DRS probe's tip location.

Dimension	RMSE (mm)	Maximum error (mm)	Minimum error (mm)
x	1.18 ± 0.58	1.76	0.00
y	1.05 ± 0.28	1.39	0.00

Table 3: Rotation angle and maximum detectable distance from the camera to track the DRS probe.

Rotation axis	Rotation angel and distance
Roll (deg)	360 deg
Pitch (deg)	360 deg
Yaw (deg)	0 deg to 180 deg
Distance to camera (cm)	10 to 100 cm

The image-based tracking framework was adeptly employed to facilitate the comprehensive mapping of all spatio-temporally tracked biopsy sites. These sites were subsequently re-projected onto the image plane, effectively generating an augmented view that dynamically portrayed the unfolding spectral acquisition process. The delineation of the biopsy site was meticulously accomplished by integrating the spatial coordinates of the probe tip, as acquired through the tracking procedure, and by considering the intricate geometric attributes intrinsic to the probe itself.

3.6 A Marker-less Neural Network-based Probe Detection Methodology

While the colour marker-based DRS probe detection and tracking approach described in the previous sections provides valuable information and has its merits, it also comes with several

limitations that can impact its effectiveness in certain scenarios. These limitations highlight the need for more advanced methodologies, such as Deep Learning (DL), to overcome these challenges:

- **Sensitivity to Lighting Conditions:** The colour marker-based approach heavily relies on the specific colour of the marker for detection and tracking. Variations in lighting conditions, shadows, and reflections can affect the colour appearance, leading to inaccuracies in marker detection and potentially causing tracking errors.
- **Occlusion:** In complex surgical environments, instruments and tissue structures can occlude the colour marker, hindering its visibility. When parts of the marker are obscured, the detection and tracking algorithms may struggle to accurately locate the probe's position, leading to incomplete or erroneous tracking results.
- **Marker Attachments:** The requirement of physically attaching the colour marker to the probe can be cumbersome and may interfere with the surgical workflow. It could also pose a risk of contamination if not applied properly or require additional sterilization steps. In addition to these operational concerns, it's important to consider regulatory approval when integrating such marker-based tracking systems into the surgical environment.

To address these limitations and enhance the robustness and accuracy of instrument detection and tracking, Deep Learning methodologies come into play. DL techniques, such as Convolutional Neural Networks (CNNs), can learn and extract relevant features directly from the image data, enabling the system to be more invariant to lighting conditions and occlusion. CNNs can also handle complex and diverse marker patterns or even be extended to marker-less approaches, removing the need for physical attachments. Moreover, DL methodologies can learn spatial and temporal dependencies within the image sequences, enabling more robust and continuous tracking of instruments, even during motion or challenging surgical conditions.

In the context of surgical instrument tracking without using physical markers, we focus on marker-less tracking methods. Marker-less tracking does not require any artificial markers or modifications to the surgical instrument. Instead, it relies on the natural, local features present on the instrument to characterize and track it within the image sequence. These natural features can include distinctive textures, shapes, or patterns that are unique to the

instrument. By opting for marker-less tracking, we ensure that the surgical instrument remains unaltered and maintains its original functionality, thereby minimizing any potential disruption to the surgical procedure.

3.6.1 *Related Work*

The advancement of deep learning techniques, particularly CNNs, has revolutionized the field of surgical instrument recognition and tracking. Several studies have proposed innovative algorithms that leverage the power of CNNs to achieve real-time and accurate detection and analysis of surgical instrument motion.

Jo et al. introduced a novel real-time detection algorithm based on CNNs for identifying surgical instruments²³⁶. Their approach utilized the capability of CNNs to automatically learn relevant features from the input images, enabling precise detection of instruments within the surgical scene. The CNN model was trained on a large dataset of annotated surgical images, allowing it to generalize and detect instruments in various situations effectively.

Similarly, Zhao et al. presented an algorithm that combined deep learning techniques with line detection and spatio-temporal context to track surgical instruments²³⁷. Their approach leveraged the capabilities of CNNs to recognize instruments and used line detection to estimate the motion of the instruments over time. The integration of spatio-temporal context helped in providing a more comprehensive understanding of instrument behaviour during the surgical procedure.

A recent literature review provided a comprehensive overview of various studies focused on surgical tool detection²³⁸. This review highlighted the diversity of approaches employed, including traditional computer vision techniques, machine learning algorithms, and more prominently, the growing trend of deep learning-based methods. The review shed light on the progress made in the field and the potential of deep learning for advancing surgical tool detection and analysis.

3.6.2 *You Only Look Once (YOLO) v5 Network Architecture*

The domain of deep learning-based object detection algorithms has demonstrated remarkable advancements in both precision and efficiency of detection. Within this domain,

two primary categories of deep learning target detection algorithms have been established: those characterized by a dual-stage process and those characterized by a singular stage. In the context of the two-stage approach, the initial phase involves the extraction of pertinent features, following which the detection algorithm generates region proposals encompassing potential target objects. Subsequently, the secondary phase entails the precise localization and classification of these proposed regions. This is achieved through the utilization of convolutional neural networks renowned for their heightened accuracy in detection, examples of which include R-CNN²³⁹, Fast R-CNN²⁴⁰, Faster R-CNN²⁴¹, and SPPNet²⁴².

In contrast, the one-stage detection algorithms diverge in their *modus operandi*. These algorithms execute both localization and classification operations in tandem, rendering the need for preliminary region proposal generation superfluous. Noteworthy examples of one-stage detection algorithms encompass YOLO²⁴³, SSD²⁴⁴, RetinaNet²⁴⁵, and FCOS²⁴⁶. Although these algorithms demonstrate an enhanced detection velocity relative to their two-stage counterparts, it is acknowledged that their accuracy levels are comparatively diminished.

YOLOv5 stands as a significant iteration within the YOLO lineage of object detection algorithms. Launched in June 2020, this iteration swiftly garnered substantial attention within the realm of computer vision due to its notable enhancements in terms of operational speed, precision, and model size²⁴⁷. Noteworthy is its remarkable ability to undertake object detection in real-time or near real-time scenarios, thereby enabling the processing of video streams frame-by-frame with nearly instantaneous object detection outcomes. This synergy of rapid processing and elevated accuracy has rendered YOLOv5 a prevalent selection for diverse object detection applications across domains such as surveillance²⁴⁸, robotics²⁴⁹, and autonomous driving²⁵⁰. In contexts where prompt and precise object detection profoundly influences system efficacy, YOLOv5 emerges as an invaluable asset.

In our study, the YOLOv5 object detection framework serves as the foundational model upon which we have conducted optimization for the specific task of probe detection. Our optimization strategy encompasses the implementation of tailored modifications that are specifically engineered to enhance the detection efficacy of both small-scale features and larger structures. These modifications were based on Qi et al.²⁵¹ work on employing YOLOv5 for face detection. Zhong has used this modified version of YOLOv5 for the detection of the DRS probe on ex-vivo tissue sampling videos²⁵². The architecture of the modified network is shown in Figure 20.

The pivotal modifications introduced to the architecture are succinctly delineated as follows:

- 1) **Landmark Regression Head Augmentation:** An integral augmentation to the YOLOv5 network involves the integration of a landmark regression head. For this purpose, the Wing loss function²⁵³ is employed, substantiating the augmentation's utility. The precision of landmark localization is significantly bolstered through this additional supervisory layer, thereby positively impacting the probe detector's overall accuracy.
- 2) **Stem Block Replacement:** A noteworthy alteration pertains to the substitution of YOLOv5's native Focus layer with a Stem block configuration²⁵⁴. This strategic replacement has yielded twofold benefits: an augmented capacity for network generalization and a notable reduction in computational complexity.
- 3) **Revised spatial pyramid pooling (SPP) Block Configuration:** A significant modification within the SPP block²⁵⁵ entails the adoption of a smaller kernel size, thereby optimizing YOLOv5 for enhanced probe detection. This augmentation has been instrumental in heightening detection precision and subsequently improving the overall accuracy of the detector.
- 4) **P6 Output Block Incorporation:** A notable addition is the integration of a P6 output block possessing a stride of 64. This augmentation substantially magnifies the detector's aptitude in detecting larger structures.
- 5) **Point Regression:** For point detection, a four-point probe tip regression is introduced within the framework. In the visual representation, it is important to acknowledge that the probe tip is not confined to a singular point but rather spans a range. The integration of four points, positioned in a top-to-average configuration, serves the purpose of mitigating the potential occurrence of extreme errors.

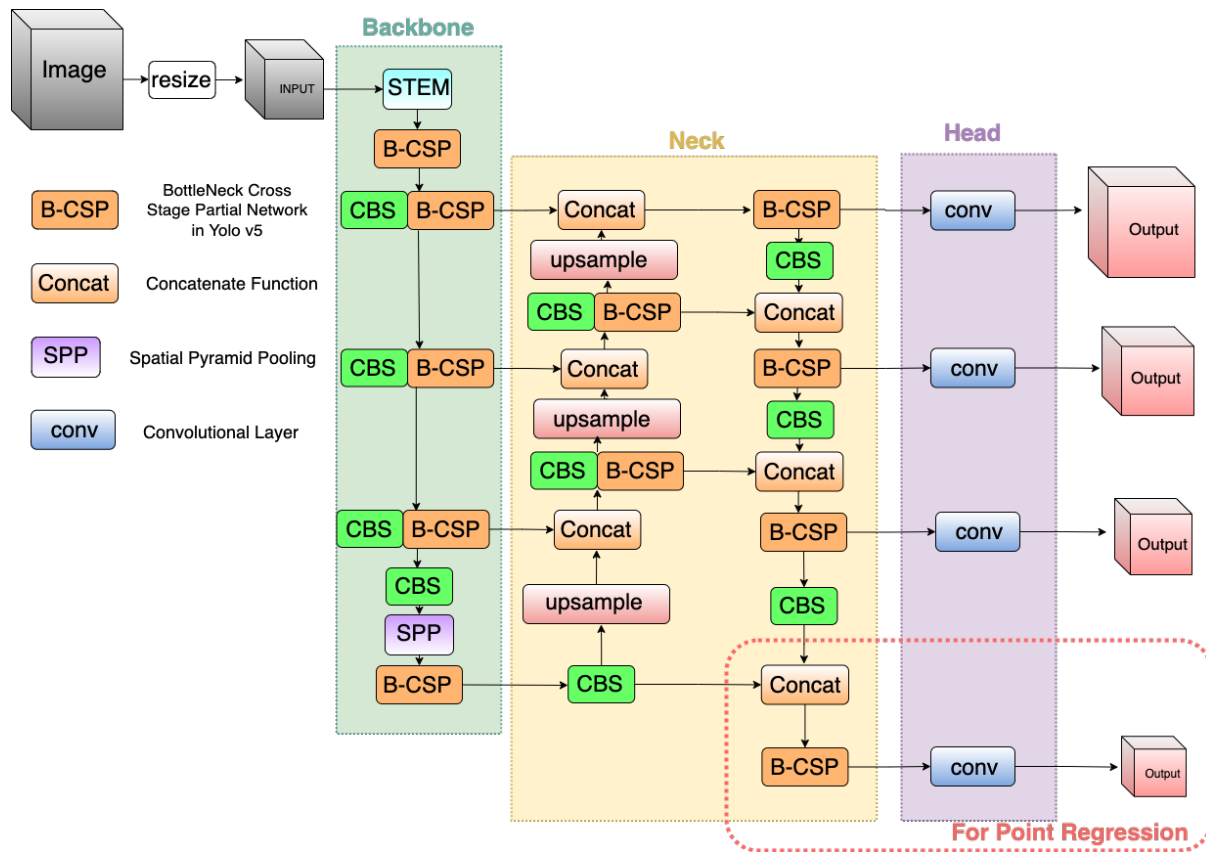


Figure 20: Modified YOLOv5-based network architecture²⁵⁶.

Zhong's work²⁵² centred on the adaptation of the YOLOv5 network through training and validation on two distinct video datasets. The primary dataset, designated as Group A, encompassed ex-vivo video data obtained from Imperial NHS Trust (ref. no. 08/H0719/37). This dataset comprised videos depicting the DRS probe's interaction with tissue specimens subsequent to their excision. The second dataset, Group B, relied on a solid tissue phantom to amass video data. In total, each dataset consisted of 11 videos. The network's performance manifested in a detection precision of 0.75, mAP0.5 of 0.99, mAP0.5:0.95 of 0.89, and an average Euclidean error measuring 7.13 pixels.

However, when applied to the context of in-vivo videos, the network exhibited challenges and limitations, thus prompting a need for elaborate consideration. The discrepancies between ex-vivo and in-vivo scenarios engender distinct hurdles that could potentially impact the network's efficacy. Some pertinent challenges include:

- 1) **Scale and perspective variability:** In-vivo settings introduce variability in the scale and perspective of the DRS fibre probe. Unlike ex-vivo scenarios where the probe's interaction with tissue is controlled, in-vivo conditions can lead to variations in probe

orientation, positioning, and depth of insertion. The YOLOv5 network might struggle to detect and track the probe accurately across varying scales and perspectives, as these scenarios were not extensively covered during training.

- 2) **Illumination and contrast changes:** In-vivo conditions often involve fluctuating lighting conditions and variable tissue types. These changes impact the probe's appearance, causing variations in illumination and contrast. The modified YOLOv5 model presented here, accustomed to the relatively controlled ex-vivo environment, might struggle to generalize to the diverse illumination settings encountered in real-time in-vivo scenarios.
- 3) **Occlusions and interactions:** In-vivo videos can exhibit occlusions and interactions that were not prevalent in ex-vivo data. Tissues, fluids, or other anatomical structures may occlude parts of the probe. The network's inability to handle occlusions effectively can result in fragmented or missed detections.
- 4) **Limited training data for in-vivo scenarios:** The YOLOv5 model was primarily trained only on ex-vivo data, which might not adequately capture the complexity of in-vivo scenarios. The absence of diverse in-vivo training data hampers the network's ability to adapt to the intricacies of real-time scenarios.
- 5) **Domain shift and generalisation:** The transition from ex-vivo to in-vivo constitutes a domain shift, leading to a discrepancy between the training and deployment domains. The modified YOLOv5 model might struggle to generalize well in the in-vivo domain due to the domain gap in data distribution.

Addressing these challenges necessitates fine-tuning the modified YOLOv5 network using in-vivo data. Fine-tuning would expose the model to the intricacies and complexities of real-time in-vivo scenarios, enabling it to learn the specific patterns, motion dynamics, and variations unique to these environments. This adaptation process would bridge the gap between the training and deployment domains, significantly enhancing the model's detection and tracking capabilities for the DRS probe in in-vivo videos.

The process of fine-tuning the modified YOLOv5 network for improved detection and tracking of the DRS probe within the intricate dynamics of in-vivo scenarios involves a systematic series of steps that underscore the network's adaptability to real-world

complexities. These steps are vital in bridging the gap between the controlled ex-vivo conditions and the dynamic challenges presented by in-vivo settings.

- 1) **Data Collection:** The foundation of this process lies in the acquisition of an extensive and representative in-vivo dataset. This dataset should encompass a diverse range of scenarios that the model is likely to encounter during actual usage. This diversity encompasses varying illumination conditions, dynamic motion patterns, an array of tissue types, and potential occlusions. This comprehensive dataset acts as the cornerstone for training the network to effectively navigate the complexities of in-vivo environments.
- 2) **Transfer Learning:** Leveraging transfer learning, the modified YOLOv5 network that has been trained on ex-vivo data serves as a foundational starting point. The existing knowledge regarding object detection from the ex-vivo domain is invaluable in speeding up the adaptation process to in-vivo scenarios, ultimately leading to quicker convergence during fine-tuning.

3.6.3 *In-vivo Video Data Collection and Labelling*

The foundation of enhancing the modified YOLOv5 network's adaptability for in-vivo scenarios lies in the acquisition of a curated dataset that encapsulates the intricacies of real-time tissue interactions. In this pursuit, a comprehensive collection of 14 new in-vivo videos was meticulously amassed, serving as the bedrock for the fine-tuning process. Each video within this dataset has been deliberately chosen to span a spectrum of scenarios representative of the challenges posed by in-vivo environments. Diverse tissue types, lighting conditions, and motion patterns have been thoughtfully included to ensure that the dataset encapsulates the complexity inherent to real-time dynamics. The dataset's design was driven by a commitment to accuracy and diversity, with the videos capturing the DRS fibre probe's interaction with living tissues as it encounters varying illumination levels, motion-induced deformations, and potential occlusions. These videos were captured under conditions akin to the intended usage, thereby encapsulating the specific nuances of in-vivo scenarios (Figure 21). It is worth noted that the camera settings used for capturing the new dataset were the same as the ones used in Zhong's work²⁵², and the videos were acquired at 980×820 pixels with 30 FPS.

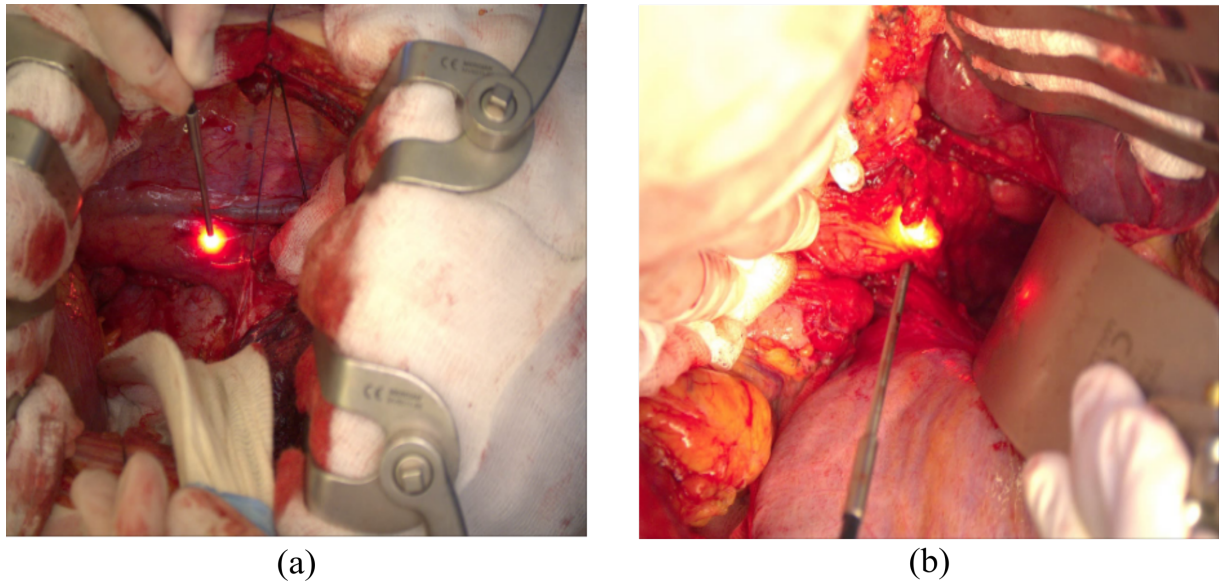


Figure 21: Example frames extracted from two different in-vivo videos.

Constructing a dataset for training and testing a model involves a meticulous annotation framework to prepare the data effectively. Figure 22 illustrates the comprehensive process employed for converting videos into images with annotations, facilitating subsequent model development and evaluation. This structured approach yielded a collection of 1,521 annotated frames, each serving as a valuable data point for the training and validation of the modified YOLOv5 network.

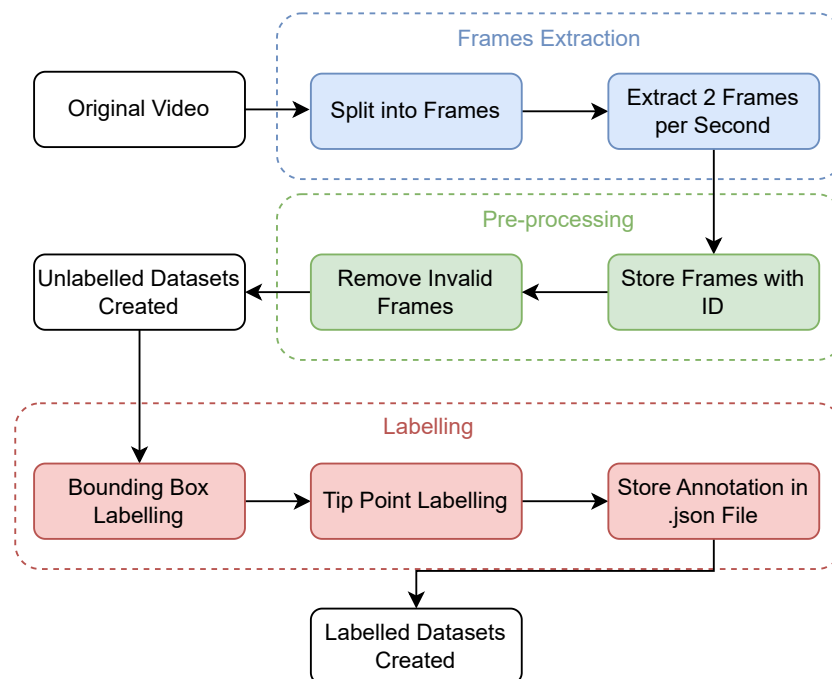


Figure 22: Video data processing and labelling pipeline²⁵⁶.

The dataset preparation process commenced with the extraction of frames from videos, a pivotal step in transitioning from video data to image-based annotations. This

extraction process was conducted programmatically, where two frames from each second of the video were selected for processing. Given a video's frame rate of 30 FPS, this approach yielded the 15th and 30th frames of every second, effectively capturing the temporal diversity of the video content.

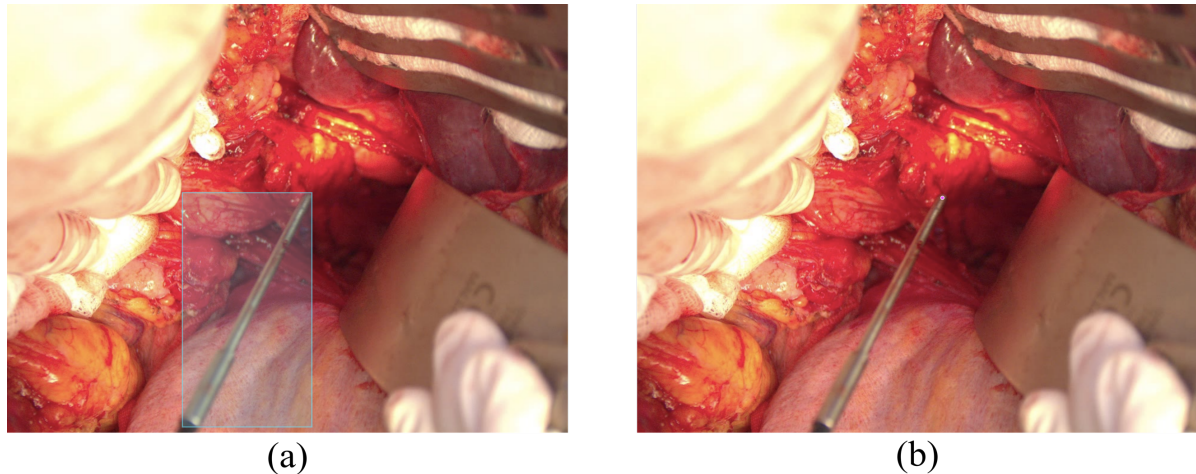


Figure 23: Frame annotation example showing (a) the bounding box and (b) the probe tip point.

The labelling process was executed using the online open-source labelling platform, Labelbox (<https://labelbox.com>, Labelbox Inc., California). Annotations were carried out with a dual purpose: the delineation of boundary boxes and the marking of DRS probe tip point (Figure 23). The dataset organization is delineated into 14 distinct folders, each corresponding to the individual videos from which the frames were extracted. Within each folder, a .json file is included that contains the annotations for each image frame as exported from Labelbox. The .json file constitutes a vital component of the dataset, storing essential information about the annotations. Specifically, the tip point annotation is denoted as 'Tip,' capturing the x and y coordinates of the point. Conversely, the bounding box annotation is designated as 'BB' and is defined by its top, left, height, and width values, effectively describing the dimensions of the bounding rectangle.

3.6.4 DRS Probe Detection Model Performance Evaluation Metrics

To determine the predictive performance of our modified YOLOv5 probe detection model, the Intersection over Union (IoU) and the Average Precision (AP) are employed.

Intersection over Union (IoU) stands as a widely adopted metric to gauge the accuracy of object localization and quantify the extent of localization discrepancies within object detection models. In the process of computing the IoU metric to juxtapose predicted and

ground truth bounding boxes, a systematic sequence is adhered to. Primarily, the area of overlap between the corresponding bounding boxes of the same object is determined. Subsequently, the collective extent of both bounding boxes is assessed, referred to as the "Union," encapsulating their joint spatial coverage. Crucially, the overlapping region between the bounding boxes, termed the "Intersection," is identified.

By dividing the Intersection by the Union, a ratio is derived that signifies the proportion of shared area in relation to the total spatial extent. This ratio furnishes an estimative indication of the proximity between the predicted bounding box and the original ground truth bounding box.

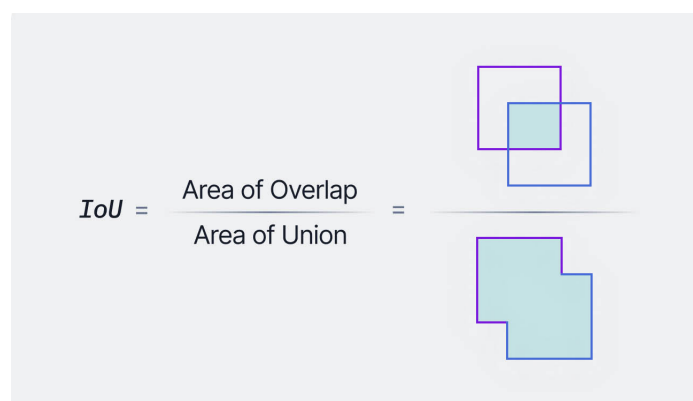


Figure 24: Intersection over Union explained [Source: <https://www.v7labs.com/blog/yolo-object-detection>].

Average Precision assumes its significance as the metric assessed through the interpretation of a precision-recall curve for a defined set of predictions.

Precision measures the proportion of true positives (correctly identified positive cases) in relation to all instances predicted by the model. Recall, on the other hand, expresses the ability of the model to find all relevant instances in a dataset, encapsulating the proportion of all instances from a specific class that the model successfully captures. A high precision means the model has fewer false positives, while a high recall means fewer false negatives.

The classification threshold is a value that separates the predicted positives from the predicted negatives. By varying this threshold, we can generate a trajectory that is graphically represented as a precision-recall curve. The curve shows how the precision and recall values change as the threshold varies. The area underneath the precision-recall curve corresponds to the average precision of a specific class for the model's predictions.

Of note, this Average Precision is further consolidated into a comprehensive measure, denoted as mean Average Precision (mAP), attained by averaging these class-specific Average Precision values across all distinct classes. It's imperative to underscore that, in the context of object detection, precision and recall find their purpose not in class predictions but in delineating bounding boxes to quantify decision performance. In this paradigm, an IoU value greater than 0.5 serves as a benchmark for positive predictions, while a value less than 0.5 corresponds to negative predictions.

Extending the evaluation to a nuanced spectrum, the mAP_{0.5:0.95} metric encapsulates the model's performance across a range of IoU thresholds from 0.5 to 0.95. This extended perspective provides insights into the model's versatility in detecting objects with varying degrees of overlap, enriching the evaluation of its localization accuracy.

$$(Equation\ 30): P = \frac{TP}{TP+FP}$$

$$(Equation\ 31): R = \frac{TP}{TP+FN}$$

$$(Equation\ 32): AP = \int_0^1 P(r)dr$$

$$(Equation\ 33): mAP = \frac{\sum_{i=1}^K AP_i}{k}$$

While AP provides valuable insights into a model's ability to precisely predict object locations, the mAP_{0.5:0.95} metric extends this evaluation to encompass a broader spectrum of object localization scenarios. This metric acknowledges the inherent variability in object poses, sizes, and orientations, which can lead to different levels of bounding box overlap between predictions and ground truth.

The IoU threshold of 0.5 has conventionally been used to determine whether a predicted bounding box is sufficiently aligned with the ground truth. However, a single IoU threshold may not adequately reflect the nuanced nature of detection scenarios. Objects with varying degrees of overlap require individual assessment to capture the full scope of the model's performance.

By considering an IoU range from 0.5 to 0.95, the mAP_{0.5:0.95} metric enables a more comprehensive evaluation of the model's ability to localize objects. This range encompasses instances where bounding boxes exhibit substantial overlap, as well as scenarios where the overlap is more nuanced. As a result, the model's adaptability to detecting objects with varying degrees of intersection is accurately captured.

3.6.5 Experimental Evaluation and Results

The dataset used in the transfer learning and fine-tuning process was composed by videos of in-vivo human tissue being scanned with the DRS probe. These videos were further split into frames as explained in section 3.6.3. The dataset consisted of 1,521 frames, with 80% used as the training set (1,258 frames), 10% used as the validation set (110 frames), and 10% used as the testing set (153 frames).

To address the requirement of deep learning methods for big data, the video dataset was further increased using image augmentation techniques. Specifically, the Mosaic²⁵⁷, mix-Up²⁵⁸, and non-max suppression²⁴³ methods were used. All images in the dataset were also preprocessed by resizing to a size of 640×640.

The experimental setup in this study entailed the integration of key components pivotal to the successful execution of the research objectives. Central to this setup were critical components, namely a Graphics Processing Unit (GPU) with CUDA (Compute Unified Device Architecture) support, and the PyTorch deep learning framework. These components collectively facilitated efficient learning processes and precise object detection within the study's context. Table 4 shows the configuration details for this experimental environment.

Table 4: Experimental settings.

Device	Configuration
System	Linux Ubuntu 20.04
CPU	13 th Gen Intel® Core™ i9-13900K 5.80GHz
GPU	NVIDIA® GeForce RTX™ 4090 24GB
Framework	PyTorch 2.0.1

In the realm of transfer learning, the pre-trained model from Zhong's work was utilised, which had been trained on ex-vivo video data. The transfer learning process commenced with the initialisation of our model's weights using Zhong's pre-trained modified YOLOv5 model. This step enabled our model to inherit the foundational knowledge and features gleaned from the ex-vivo domain. This initialisation provided our model with a head start, enabling it to already possess a rudimentary understanding of DRS probe detection. It further allowed our model to rapidly grasp salient features, textures, and patterns common to surgical imaging scenarios. This weight propagation mechanism encompassed the transfer

of learned parameters from source to target model layers. As a result, our model's convergence during fine-tuning was significantly accelerated, as it could swiftly adapt the acquired knowledge to in-vivo conditions.

Fine-tuning entailed the meticulous refinement of weights through iterations of backpropagation, optimising the model's response to in-vivo challenges. During this phase, the model underwent a dual transformation: retention of ex-vivo acquired insights and adaptation to the nuanced textures, lighting conditions, and motion dynamics unique to in-vivo data. Fine-tuning encouraged the model to recalibrate its learned features. While retaining its grasp on core detection principles from ex-vivo data, the model adapted its feature detectors to accommodate in-vivo intricacies, tuning itself for the accurate detection of the DRS fibre probe. This recalibration facilitated the development of contextual understanding, ensuring that the model could distinguish relevant object attributes amid the complexities presented by real-time in-vivo scenarios.

The parameters used for the training process of the modified YOLOv5 model on the in-vivo video data are summarised in Table 5. To avoid overfitting, the model was trained for 1,000 epochs with early stopping. Early stopping is a pivotal regularisation technique that operates as a safeguard against overfitting by monitoring a model's performance on the validation dataset. The core principle involves halting the training process when the model's performance ceases to improve, ultimately preventing further training that might lead to overfitting. A critical parameter within the early stopping framework is known as "patience". Patience refers to the predetermined number of epochs during which the model's performance on the validation set is allowed to stagnate without any substantial improvement. If the model's validation performance fails to show progress within this patience period, training is promptly terminated.

In the context of our work, the patience parameter was thoughtfully set to 100 epochs. This indicates that, during training, the model's performance on the validation set was observed over consecutive epochs. If, after 100 epochs, the model's validation performance did not exhibit notable enhancements, early stopping would be triggered. When the monitored validation performance of the model plateaued and no substantial improvements were observed after the designated patience period of 100 epochs, training was terminated. This proactive decision to halt training was taken to forestall the model from becoming too specialised to the training data, thereby averting overfitting – a phenomenon

where a model becomes excessively tailored to the training dataset but fails to generalise effectively to new, unseen data.

Table 5: Training parameters for the modified YOLOv5 network.

Parameter	Value
device	GPU (CUDA)
epochs	1000
batch size	32
image size	640
patience	100
learning rate	0.01

The results in Figure 25 provide insights into how the model learned over time and whether there were any particular trends or patterns in their performance. This visual representation enables an exploration of whether the model exhibited consistent improvement, encountered plateaus, or demonstrated erratic fluctuations in its learning trajectory.

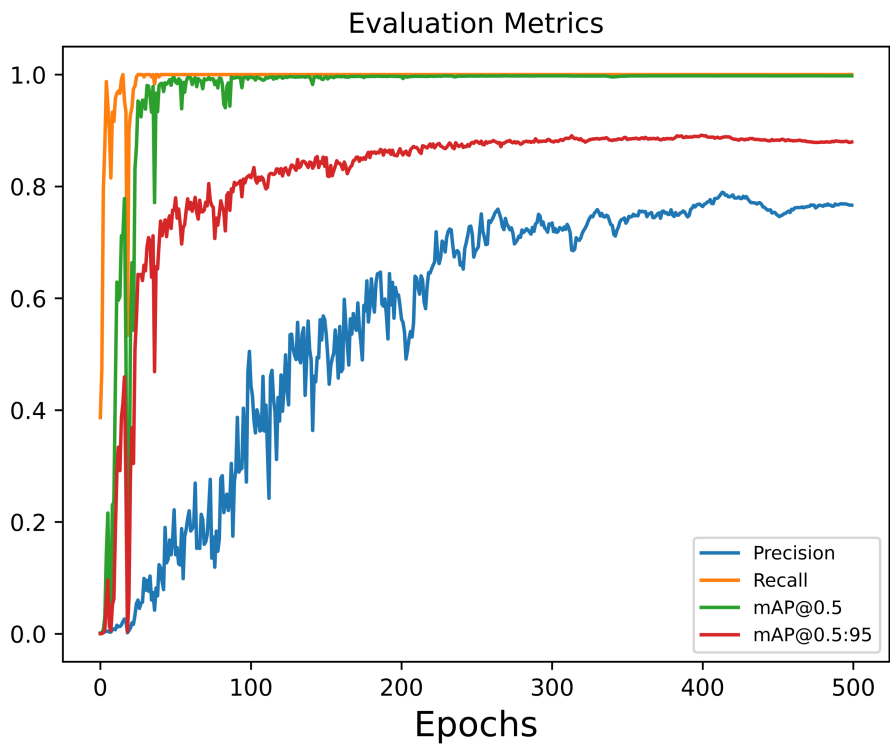


Figure 25: Model’s performance evaluation using precision, recall, mAP@0.5, and mAP@0.5:95.

Additionally, Figure 26 serves as an informative comparative illustration, contrasting the accuracy of the modified YOLOv5 model in two distinct contexts: ex-vivo only and the amalgamation of ex-vivo and in-vivo datasets. The precision of the fine-tuned modified YOLOv5 achieved a value of 0.7659, which signifies the model's accuracy in correctly identifying the DRS probe tip. On the other hand, the precision of the initial modified YOLOv5 model, which was exclusively trained on ex-vivo data, reached a value of 0.75. Both models had high recall, indicating that they correctly identified the DRS probe in the image. Specifically, the recall for both models was 1.00. Both model variants exhibited exceptional performance as reflected by their mAP@0.5 scores, signifying their remarkable capability to accurately detect the DRS probe. The fine-tuned modified YOLOv5 demonstrated an outstanding mAP@0.5 value of 0.9978, while the initial modified YOLOv5 attained an impressive value of 0.9975. These high mAP@0.5 values underscore the models' proficiency in accurately localising and identifying the DRS probe instances within the images. Furthermore, the evaluation extends beyond a single IoU threshold. The fine-tuned modified YOLOv5 also excelled in terms of mAP@0.5:95, recording a higher score of 0.8916, in comparison to the initial modified YOLOv5 with a value of 0.8901. The fine-tuned model's higher mAP@0.5:95 score signifies its superior accuracy in detecting the DRS probe across a broader range of IoU thresholds, especially at the higher end of the scale.

The introduction of in-vivo data into the model training scheme has a significant impact on model performance. The marginal increase in precision from the initial ex-vivo trained model to the fine-tuned version highlights the efficacy of adapting the model to in-vivo conditions. The modest yet meaningful enhancement in precision underscores the practical significance of the proposed approach, demonstrating that fine-tuning the model using in-vivo video data translates into refined and more accurate DRS probe detection.

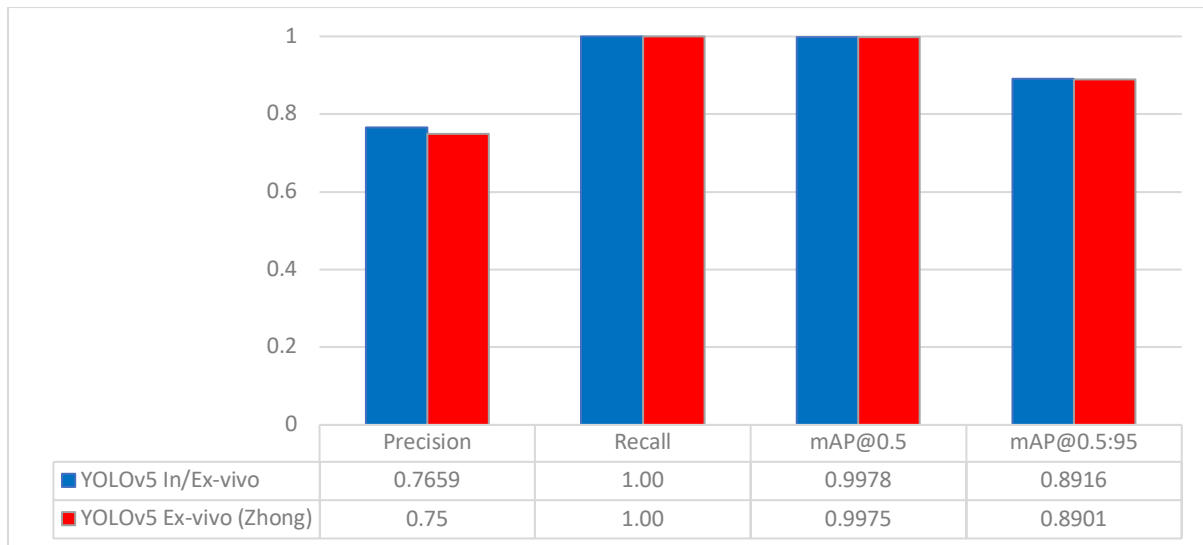


Figure 26: Comparison of the models' accuracy between Zhong's modified YOLOv5 network that was trained using only ex-vivo imaging data, and the current thesis's work where Zhong's pretrained network was fine-tuned using in-vivo data.

Upon subjecting the fine-tuned modified YOLOv5 model to rigorous evaluation on the testing set, a comprehensive set of performance metrics was acquired. One notable metric is the average loss, an indicator of the model's convergence and its alignment with the ground truth annotations. The computed average loss of 4.38 pixels in the prediction of the DRS probe tip location signifies the model's ability to minimise the disparity between its predictions and the actual annotations, encapsulating its learning progression during training.

Delving deeper into the evaluation, precision at various thresholds offers a nuanced perspective on the model's probe detection capability. Notably, when considering predictions with a loss less than 20 pixels, the model exhibited a precision of 0.9895. For predictions with a loss less than 15 pixels, the model's precision was 0.9860, demonstrating its ability to maintain robust precision even with slightly larger positional variations. Similarly, when the tolerance was tightened to predictions with a loss less than 10 pixels, the model's precision remained at 0.9266. In more challenging scenarios, where predictions are required to be within a mere 5-pixel margin, the model exhibited a precision of 0.6958.

The Figure 27 shows two prediction examples of the DRS probe tip using the fine-tuned modified YOLOv5 model in inference mode. In the first frame on the left, the prediction Euclidean distance error was 2.24 pixels, while the DRS probe detection accuracy was 85.30%. Conversely, the second frame on the right exhibits a slightly elevated Euclidean distance error of 6.08 pixels. While this value might seem relatively high on its own, it is imperative to contextualise it within the broader image dimensions. In the context of the expansive field-

of-view encompassing 980×820 pixels, the perceived error remains relatively small, reinforcing the model's capacity to discern probe tip positions with noteworthy precision. One additional dimension to consider is the inference time, a critical metric for real-time applications. Each image was processed in 13 milliseconds during inference. This swift inference time attests to the model's efficiency, making it amenable for surgical applications where real-time responsiveness is a key requirement.

For frames like the right frame in Figure 27, the impact of light-related distortions emerges as a prominent concern as it affects the probe tip point detection accuracy. This phenomenon, primarily attributed to the inherent characteristics of Diffuse Reflectance Spectroscopy, introduces distinctive challenges to the probe tip point detection process. This phenomenon is an outcome of the intricate interaction between incident light and the varying tissue properties, leading to the blurring of object boundaries and resulting in a discernible loss of clarity in the generated images.

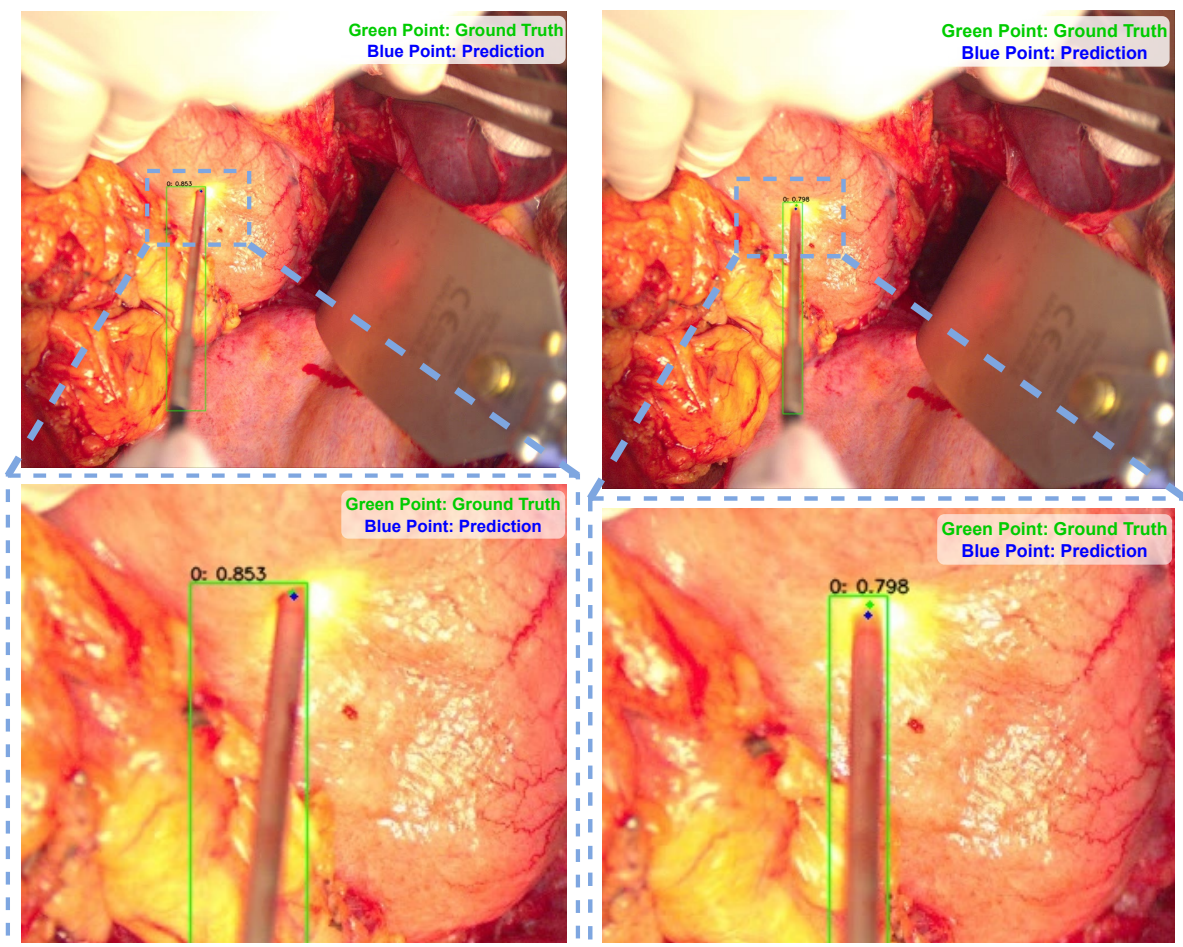


Figure 27: Examples of DRS probe tip prediction location using the fine-tuned modified YOLOv5 model. The green point is the ground truth annotated tip location, and the blue point is the predicted tip.

At the heart of the inference workflow lies an orchestrated sequence of processes, each contributing to a holistic analysis of the input video data. The initial step involves dissecting the input video into individual frames. This is achieved through a process known as video splitting, where each frame is isolated to serve as an independent unit for subsequent analysis. In the next phase, each isolated frame undergoes object detection using the trained fine-tuned modified YOLOv5 model. The model predicts two primary components: a bounding box that encapsulates the entire DRS probe, and a tip point that represents a specific location on the probe. This dual-prediction mechanism characterises the model's ability to discern both overall probe's presence and a specific reference point.

Once the probe detection has been performed on each individual frame, the detected bounding boxes and tip points are integrated back into the original frames. This integration forms the basis for generating a new video where each frame is enriched with the corresponding bounding box and tip point annotations. Finally, the x and y coordinates of the probe's tip point are recorded for each frame. This collection of coordinates establishes a trajectory that effectively traces the movement of the tip point throughout the duration of the video (Figure 28).

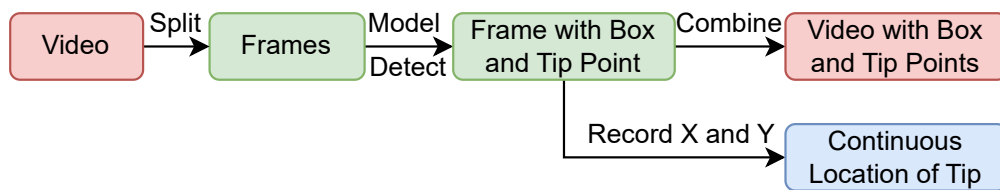


Figure 28: Integrated video analysis workflow.

The fundamental outcome of this process yields a dynamic video representation (Figure 29b). Within this video, a distinct green boundary box encapsulates the DRS probe, while a prominent blue tip point designates the current tip location of the probe. In addition, a complementary .txt file that contains the coordinates of the tip point is generated.

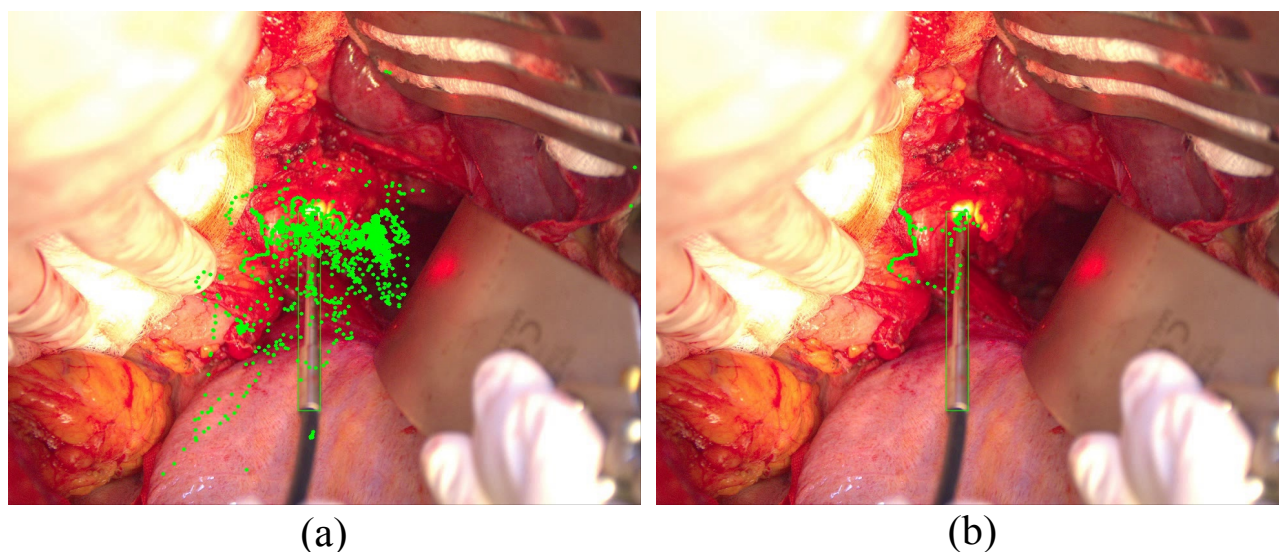


Figure 29: Illustration of video output featuring trajectory tracking. Historical detection points of the DRS probe tip are marked in green and current tip location highlighted in blue. (a) Full probe tip locations trajectory. (b) Most recent 100 probe tip locations trajectory.

Figure 29 provides two designed visual depictions of the DRS probe tip trajectory, each offering insights into the dynamics of the probe trajectory within the surgical scene. The first representation (Figure 29a) endeavours to provide a comprehensive depiction of the probe's trajectory, spanning its entire temporal continuum. This view facilitates an in-depth exploration of the probe's spatial dynamics, permitting an assessment of patterns, tendencies, and potential correlations that span the entirety of its path, providing to the surgeon the whole area that was already scanned. Conversely, the second visualisation (Figure 29b) exercises a strategic curation by spotlighting the probe's most recent 100 data points within its trajectory. As the video progresses, earlier data points undergo a systematic pruning, thus confining the display to the probe's most immediate movement dynamics. This dynamic representation provides real-time insights into the probe's contemporaneous spatial scanning of the underlying human tissue, furnishing an acute perspective on its current positional interactions. The whole process described under this section is also depicted in Figure 30.

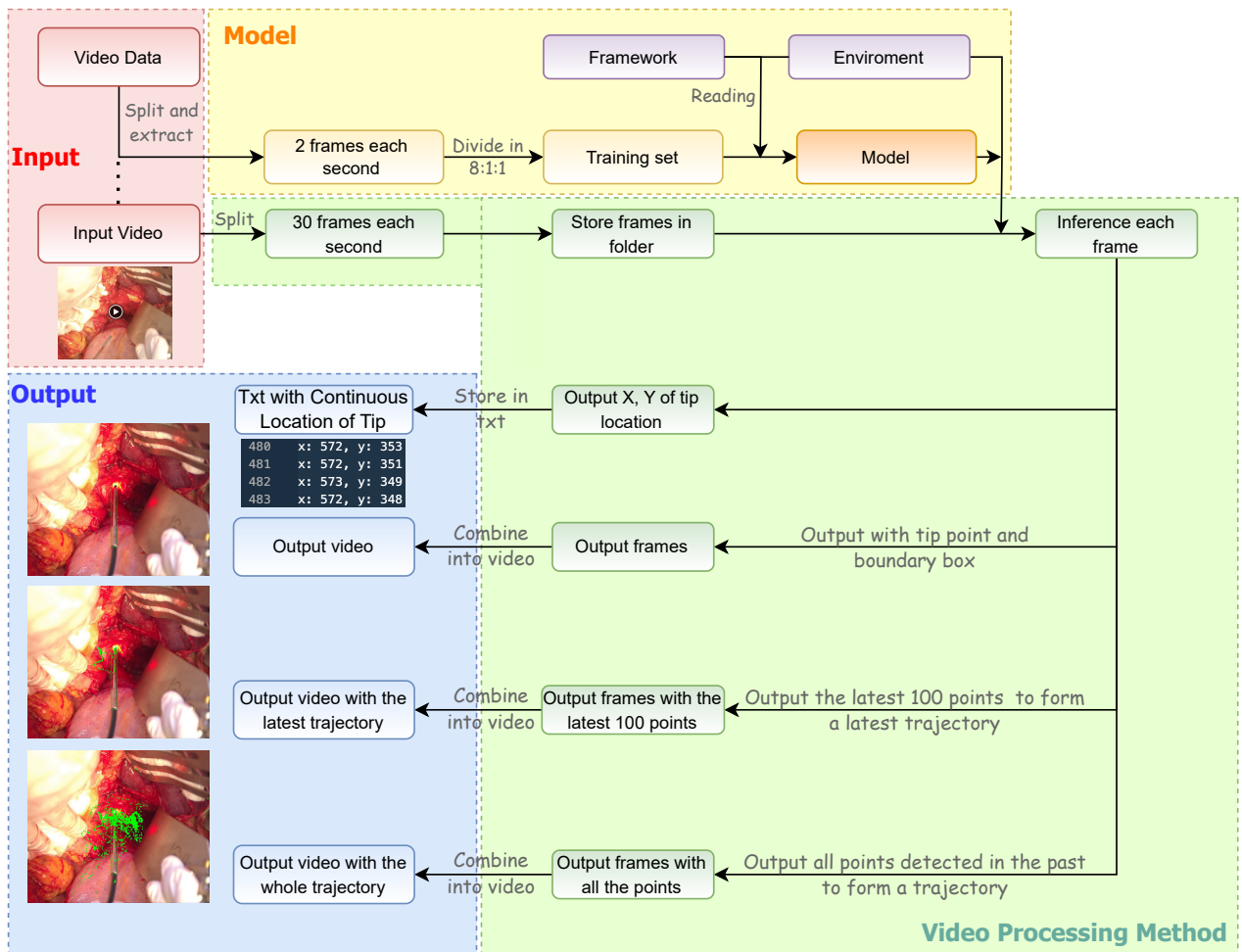


Figure 30: A comprehensive representation of the configuration workflow and versatile video output.

3.7 Summary

In this chapter, we introduced two novel approaches for acquiring spatial diffuse reflectance spectroscopic diagnostic information using computer vision tracking of a handheld spectroscopic probe. Our system enables simultaneous recording of the probe's position and orientation (pose) along with the diagnostic spectroscopic data for each measurement acquisition. These data are then overlaid onto the surgical field of view (FoV) to create an augmented reality display, providing clinicians with comprehensive control over diagnostic parameters and facilitating system tuning for diverse clinical contexts.

Through experiments on ex-vivo and in-vivo human tumour biopsies we demonstrate the system's capability for accurate lesion mapping. This feature grants surgeons greater clinical control, allowing them to make informed decisions during surgical interventions based on both biochemical and morphological information.

However, a key limitation of the current configuration is the challenge posed by in-vivo tissue motion. The dynamic nature of living tissues, such as pulsations, respiratory motion, and surgical manipulations, may compromise the precision of probe tracking and the quality of the augmented reality display overlay. During surgeries, especially those involving the abdomen or thorax, tissue can move from several millimetres to a few centimetres due to respiratory motion. These tissue movements can possibly cause inaccuracies in the alignment of diagnostic data, particularly during real-time surgical interventions.

By bridging the gap between point-based spectroscopic diagnoses and imaging information, our image-guided DRS probe-tracking system successfully overcomes the long-standing trade-off between diagnostic accuracy and the field of view that has limited conventional DRS diagnostic systems. The integration of computer vision tracking with DRS enables the acquisition of spatially resolved diagnostic data, addressing the limitations of conventional point spectroscopy. Within the domain of DRS diagnostics, the probe detection and tracking methods proposed in this chapter play a pivotal role by furnishing historical tissue scanning location data, thereby enhancing the diagnostic context. The augmented reality display provides surgeons with real-time information, enhancing their ability to precisely identify tumour margins and critical tissue regions during surgical procedures.

Chapter 4: DRS as a Surgical Guidance Tool in Gastrointestinal Cancer Surgery: Insights from Ex-vivo Investigations

4.1 Introduction

The clinical demand for real-time tumour detection of resection margins during gastrointestinal cancer surgery is yet to be met. Surgeons continue to rely on their visual and tactile perceptions to ascertain the optimal resection plane. The inadvertent retention of tumour tissue within the margins could significantly elevate the risk of local recurrence, prompting the need for additional interventions, such as re-excision or boost radiation therapy in patients with tumour-positive resection margins. Conversely, excessive removal of healthy tissue may negatively impact the outcome. Excessive removal of healthy tissue during surgery can lead to functional impairment, longer recovery times and increased post-operative complications for the patient. It can also result in reduced organ function or cosmetic issues, affecting the patient's quality of life and overall recovery experience. Therefore, achieving a precise balance in tissue removal is critical to optimising patient outcomes.

In this study, the use of diffuse reflectance spectroscopy as a potential surgical guidance tool for optimal resection margin detection is investigated. DRS, as an optical technology, rests on the fundamental principle of measuring optical responses, which are linked to the structural and compositional attributes of the underlying tissue (Figure 11). DRS demonstrates the capability to swiftly characterise tissue volumes of several cubic millimetres in near real-time, without the dependency on exogenous contrast agents. This distinctive feature positions DRS as a promising candidate for facilitating margin assessment during gastrointestinal cancer surgery.

The DRS modality has been applied to ex-vivo differentiation of tumours from healthy surrounding tissues across multiple malignancy types, such as colorectal, lung, and breast cancers^{259–262}. It is important to note that the term “healthy” in this context refers to the non-

cancerous tissue surrounding the tumour and tissue that retains its normal cellular and structural features and is not undergoing any malignant changes mainly close to the margin far from the tumour tissue. Nevertheless, to our knowledge, although DRS has been used in the context of premalignant upper gastrointestinal disease^{263,264}, its use in oesophageal and gastric cancer has not been fully explored. In addition, there has been significant interest in the use of DRS in colorectal cancer^{265–271}. These studies have been able to discriminate tumour from normal colon tissue with high overall accuracy of 0.95²⁶⁵, 0.94²⁶⁶, and 0.91²⁶⁷. However, the success thus far in the use of DRS technology for assessing colorectal specimens has lacked approaches which would make it clinically relevant and integrate smoothly into the operating field. For example, use of the DRS probe on the tissue was point-based, only a small number of spectra were acquired per sample point on the tissue and there was a disruption of the serosal layer in some studies, which hinders CRM measurement.

4.2 Spectroscopy System and User Interface

The DRS system consisted of a reflection/backscattering probe (Ocean Optics Inc., QR600-7-SR-125F) containing six peripheral illumination fibres around one light collection fibre (600 µm diameter, 0.22 numerical aperture) within a 0.125-inch ferrule. A tungsten halogen light source (360 – 2400 nm, typical output power 8.8 W, Ocean Optics Inc., HL-2000-HP) was connected to the six illumination fibres, while the central fibre was connected to a spectrometer (Ocean Optics Inc., USB4000) that resolved light in the spectral range 400 – 1000 nm (Figure 31). Both the light source and the spectrometer were controlled by a user interface (UI) developed with Python.

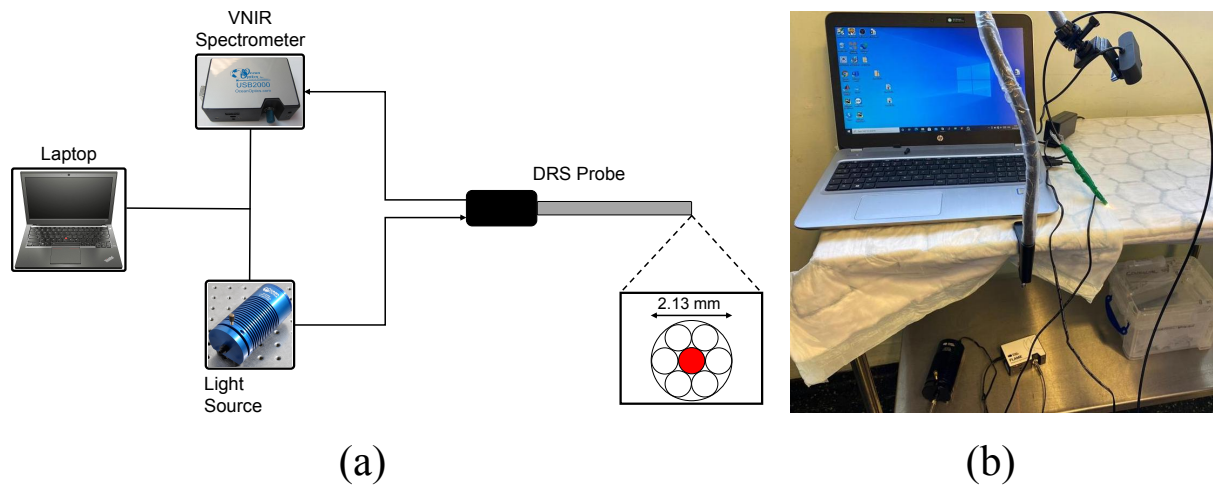


Figure 31: (a) Schematic illustration of the DRS instrumentation. (b) Photograph of the setup. The DRS probe was connected to both the light source and the spectrometer to allow for data acquisition and sample illumination. All electronic devices communicated with proprietary software designed in-house with Python v.3.9 (Python Software Foundation). VNIR indicates visible and near-infrared.

Table 6: DRS probe QR600-7-SR-125F specifications.

Specification	Value
Wavelength range	200 nm – 1.1 μ m
Fibre core size	600 μ m
Length	2 m
Jacket	Silicone-coated steel monocoil
Probe ferrule diameter	3.175 mm (1/8")
Probe ferrule length	76.2 mm (3")
Probe ferrule material	Stainless steel

The UI was developed using Python 3.9 (Python Software Foundation) and Qt5 toolkit to integrate all the functionalities on a standard PC (Intel i3 processor at 3.30 GHz with 8 GB RAM). The UI, as shown in Figure 32, consists of different functions such as the “start tracking” function for the initiation of the probe tracking and the “get spectrum” for the spectral data acquisition. Before starting the spectral data acquisition, the user entered the patient metadata, spectrometer’s integration time, spectral range, tissue type, and DRS calibration through a dialog box into the UI. A single frame of the tissue specimen being examined without the DRS probe was captured. This frame was annotated with all the optical biopsy locations at the end of the acquisition process. The live video feedback from the RGB camera was annotated in real-time with the probed positions, and the current spectrum at each probed location was displayed in the main UI window. Finally, to facilitate real-time spectral classification, a colourmap was utilised to visually indicate the probability of a data point

belonging to a class of tissue type (green for normal, pink for tumour). When the experiment was finished, all data was saved for further correlation with the pathology data.

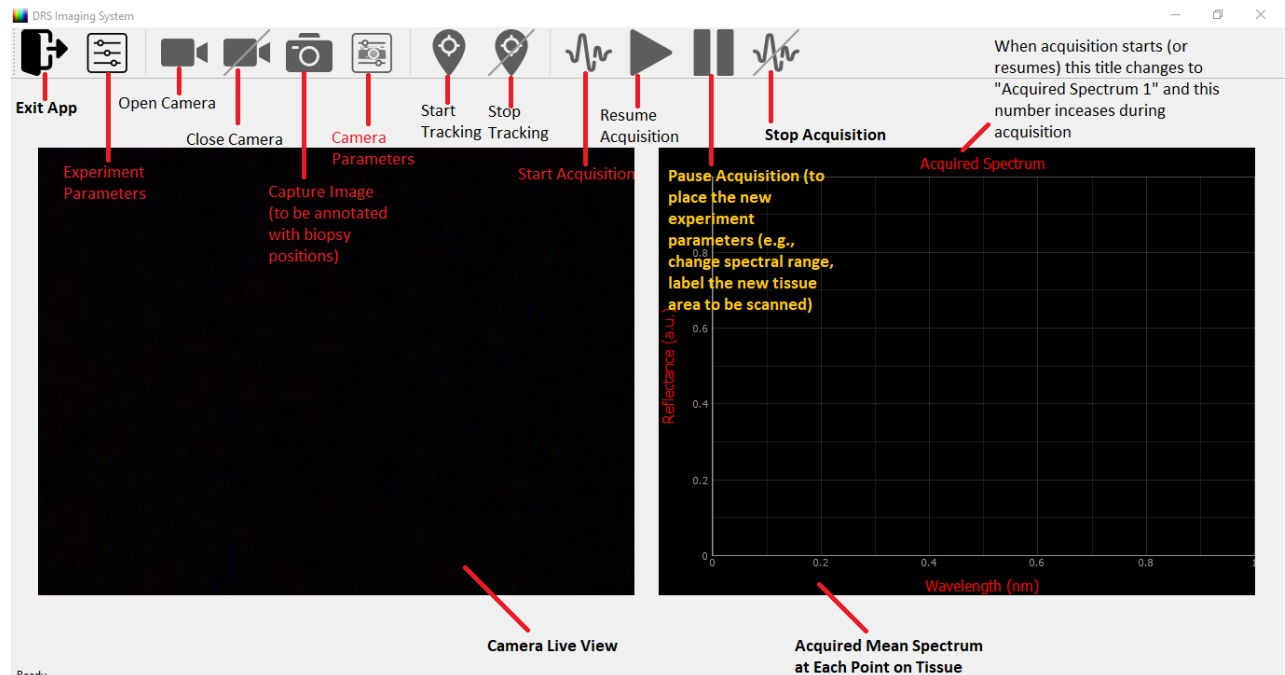


Figure 32: Python proprietary software for hardware DRS system control, data acquisition and real-time spectral classification. The Control Panel allows for execution of the data acquisition protocol, DRS probe detection and tracking, hardware calibration and light source control. The Augmented Live Video View hosted the live video feed from the RGB camera. Reflectance data in the 400-1000 nm spectral range displayed in real-time under the corresponding view of the UI.

4.3 Histopathology Correlation

Optical technologies hold the potential to revolutionise tissue characterisation, particularly in critical applications like the identification of positive resection margins during cancer surgery. The validation of the optical measurements involves a meticulous comparison between experimental outcomes and the established gold standard²⁷².

At present, the gold standard for tissue characterisation entails microscopic examination of tissue sections stained with haematoxylin and eosin (H&E) by a skilled pathologist. Nevertheless, a significant challenge arises due to the profound deformations endured by the tissue in H&E sections, in stark contrast to its appearance during optical measurements. The process of converting tissue into H&E sections inherently introduces these deformations. Consequently, superimposing a digitalised H&E section onto images capturing optical measurements becomes susceptible to registration errors^{273–276}.

The alignment discrepancies that can arise between a DRS measurement and histopathology data have the potential to detrimentally impact the accuracy assessment of this technology. This concern becomes even more salient during the development of

classification algorithms, where the precise assignment of the correct histopathology label to each measurement is pivotal. Achieving an accurate correlation between optical measurements and histopathological insights stands as a significant challenge in the clinical validation of optical technologies^{272,277,278}. It is noteworthy, however, that the literature dedicated to correlating optical measurements with histopathology remains relatively sparse^{279–281}. Furthermore, when methods are outlined to establish the connection between optical data and histopathology, the registration strategies described in existing research appear to lack the capability to address tissue deformation^{282–284}.

The histopathology correlation framework designed in our study is described below (Figure 33). After the optical measurements have been taken, the suspected tumour location was painted with yellow tissue dye (Cancer Diagnostics Inc.) to allow for correlation with histopathological analysis, which was used as a reference test. Samples were then placed into formalin and sent to the histopathology department, where they were processed according to standard protocols. Macroscopic photos of the whole specimen, as well as slices, were taken.

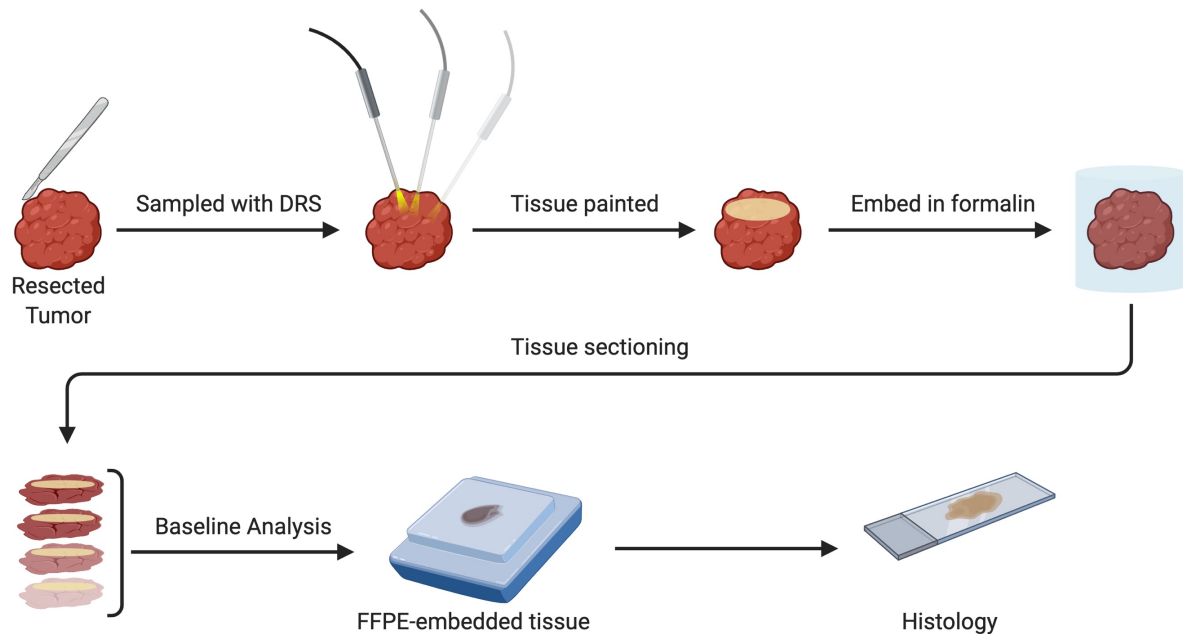


Figure 33: Histopathology correlation pipeline for ground truth labels generation.

Once it was confirmed that the yellow-painted area indeed contained tumour tissue, the acquired spectra and video frames were correlated with the histopathological information (Figure 34, Figure 35). These correlated regions were labelled as tumour, while the remaining spectra were labelled as healthy tissue. These labels were considered ground

truth labels for the training and validation of the classification algorithms. However, if the sampled suspected tumour area was identified as regressed tumour following neoadjuvant therapy (T0), only spectra representing healthy tissue were considered for further analysis.

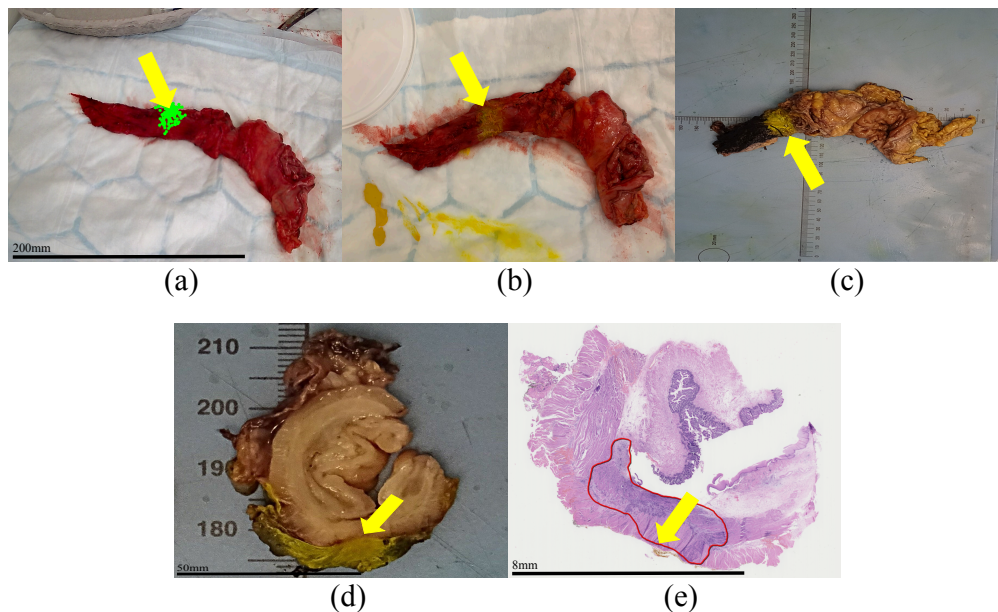


Figure 34: Histopathological correlation workflow is shown on an example of oesophageal adenocarcinoma. The correlation was performed using yellow tissue paint (marked with a yellow arrow across panels (a) – (e)). (a) The ex-vivo specimen with tracked optical biopsy sites of the suspected tumour area. (b) Ex-vivo specimen following painting of the suspected tumour tissue with yellow paint. (c) Specimen, having been placed in formalin, with yellow painted area visible. (d) Macroscopic slice of oesophageal tissue with the yellow painted area. (e) Microscopic H&E-stained slice with yellow paint (yellow arrow) and confirmed oesophageal adenocarcinoma (red outline).

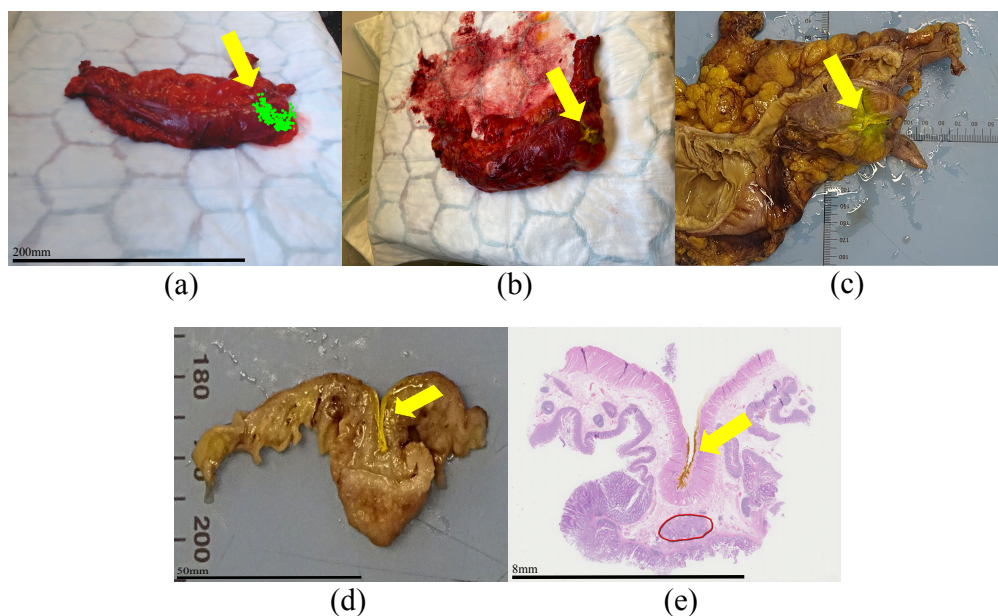


Figure 35: Histopathology correlation for a colorectal specimen. (a) Photo of scanned colon tissue with tracked DRS locations over suspected tumour area indicated. (b) Photo of colon tissue with painted yellow dye over the suspected tumour tissue. (c) Photo of colon tissue after being in formalin with yellow tissue dye visible. (d) Photo of sliced colon tissue with yellow dye visible over the suspected tumour. (e) Scanned H&E slide showing yellow tissue dye over confirmed tumour tissue (marked in red).

4.4 Study Design

4.4.1 *Differentiating Healthy Upper GI Tissue from Cancer: Insights from Ex-Vivo Study*

Patients undergoing oesophageal or gastric cancer resection were prospectively recruited into the study between July 2020 and July 2023 at Hammersmith Hospital in London, United Kingdom. Prior to patient enrolment, the study received ethical approval from the Harrow Research Ethics Committee under reference number 08/H0719/37. Furthermore, the principles of informed consent were rigorously upheld, with all participants providing comprehensive written consent prior to their involvement in the study.

Ex-vivo tissue specimens were included for patients undergoing elective surgery for either oesophageal carcinoma (adenocarcinoma or squamous cell carcinoma) or gastric adenocarcinoma. Healthy tissue data was included for analysis for cases confirmed as T0 on formal histology. It is worth noting that patients below the age of 18 and those subjected to non-elective surgical procedures were deliberately excluded from the study.

4.4.2 *Distinguishing Healthy Colorectal Tissue from Cancer: Ex-Vivo Study Findings*

This was a prospective validation study. Ethical approval was already in place from the Harrow Research Ethics Committee (ref. no. 08/H0719/37). Patients undergoing elective colorectal cancer resection surgery at Imperial College NHS Trust in London were prospectively recruited between April 2021 and July 2022. Abdominoperineal excision of rectum (APER) cases were excluded from this study, since we only collected colonic tissue data. This was due to the presence of mesorectal fat surrounding the rectum. Healthy tissue data was included for analysis for cases confirmed as T0 on formal histology.

4.5 Data Acquisition Protocol

Data acquisition was performed within 15 minutes of resection of the surgical specimen and lasted approximately 10 minutes. This timeframe was strategically chosen to mitigate the effects of tissue degradation and changes in haemoglobin absorption that might occur due to the lack of blood supply. This approach significantly minimises the potential alterations in

tissue characteristics that could impact the DRS measurements' accuracy. The decision was informed by existing literature that suggests the most pronounced physiological changes in excised tissues, including those affecting haemoglobin's optical absorption properties, tend to occur over more extended periods than our data acquisition timeframe²⁸⁵. An overview image of the specimen and a video recording during spectral scanning were captured by an RGB camera (Teledyne FLIR LLC, Blackfly S), which allowed for tracking of the tip of the DRS probe and mapping of optical biopsy locations in real-time (Figure 36). Spectral data were collected from oesophageal, stomach or colorectal specimens ex-vivo immediately after surgical resection. The hand-held DRS probe was used on healthy tissue within the sample and on macroscopically suspected cancerous or fibrosed tissue to obtain spectral information. DRS measurements of suspected macroscopic tumour site and healthy tissue were taken on the outer (serosal) layer of the tissue.

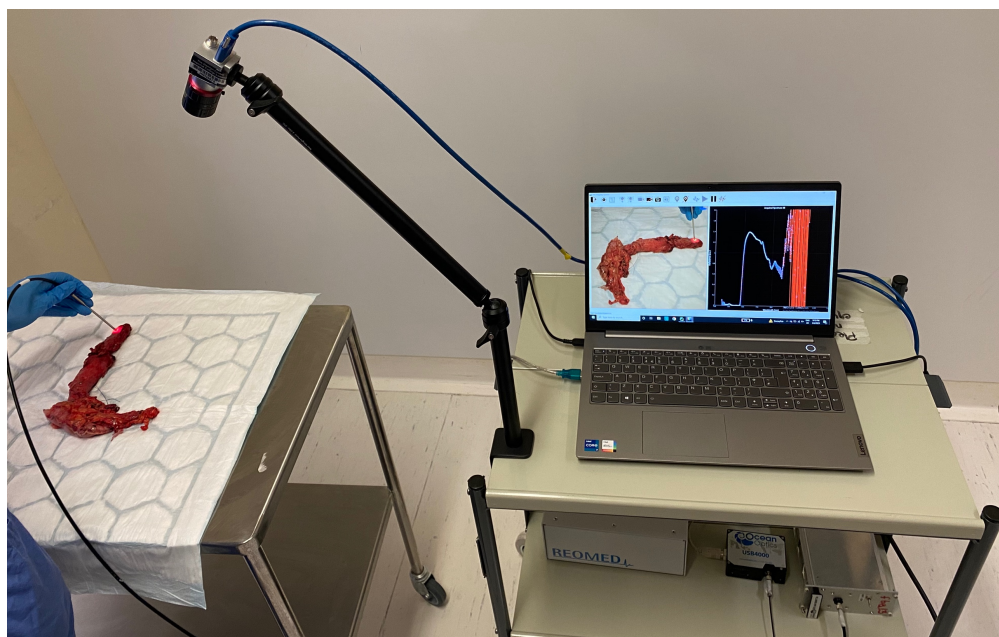


Figure 36: Photograph of the DRS system during ex-vivo data acquisition of an oesophageal specimen.

Healthy tissue measurements were taken as far from the suspected cancerous location as possible and close to the resection margin, based on visual and haptic inspection of the macroscopic tissue by a surgeon. Sampling at the resection margin itself was avoided to prevent any interaction with clips, sutures, or staples as they would be expected to affect the spectral data due to their reflective properties. The DRS probe recorded approximately 80 spectra per second. At each measurement location, twenty DRS measurements (1 spectrum per measurement) were acquired and averaged. A minimum of 200 different locations per tissue type (normal or tumour) were sampled, depending on the size of the

region. When possible, DRS recordings of stomach and oesophagus were taken from the same specimen. Data were acquired in the spectral range of 420 – 720 nm with a λ separation of 0.5 nm for the oesophageal and stomach specimens, and in the spectral range of 450 – 1100 nm with a λ separation of 0.34 nm for the colorectal specimens. Wavelength, or λ , separation denotes the interval between successive measurement points in a spectrum.

The DRS measurements were obtained in continuous mode starting in the healthy tissue, continuing through the transition zone of normal-tumour, and ending in the tumour. At each point measurement location, the acquired spectra were averaged to calculate a mean spectrum of each measurement location. To correct for the dark current, the spectrometer was calibrated prior to data acquisition²⁸⁶. A reference spectrum was acquired by direct contact measurement on a white reflectance phantom (optical-grade Spectralon® reference, Labsphere Inc.). Subsequently, to account for inter-patient comparability, background light, and signal quality variability, spectral normalisation, and noise reduction was performed for each acquired spectrum. This was achieved by correcting all raw spectra using the white reference reflectance spectrum and the dark current.

$$(Equation\ 34): \lambda_{DRS} = \frac{\lambda_{RAW} - \lambda_{DN}}{\lambda_{WST} - \lambda_{DN}}$$

where λ_{RAW} is the raw signal acquired from the spectrometer, λ_{DN} is the dark-field signal, λ_{WST} is the white reflectance standard signal and λ_{DRS} is the final reflectance value. DRS data were then processed using a Savitzky-Golay²⁸⁷ (or digital smoothing polynomial) filter to perform noise reduction and preserve higher order moments (spectral characteristics) of the original spectrum. As DRS data inherently consist of intricate spectral peaks and valleys corresponding to various tissue properties, the filter operates by consecutively fitting segments of adjacent spectral data points with a low-degree polynomial using a least-squares approach. By systematically replacing the central value of each segment (or window) with the value derived from the polynomial fit, the filter effectively smoothens the DRS data. The workflow of the ex-vivo data acquisition is shown in Figure 37.

The pivotal role in patient recruitment throughout the entire study duration was assumed by the author of this thesis. Specifically, this role involved consenting patients and acquiring data, as well as controlling the software system. This responsibility was carried out in close collaboration with Ms. Scarlet Nazarian and the proficient research team at Imperial College London.

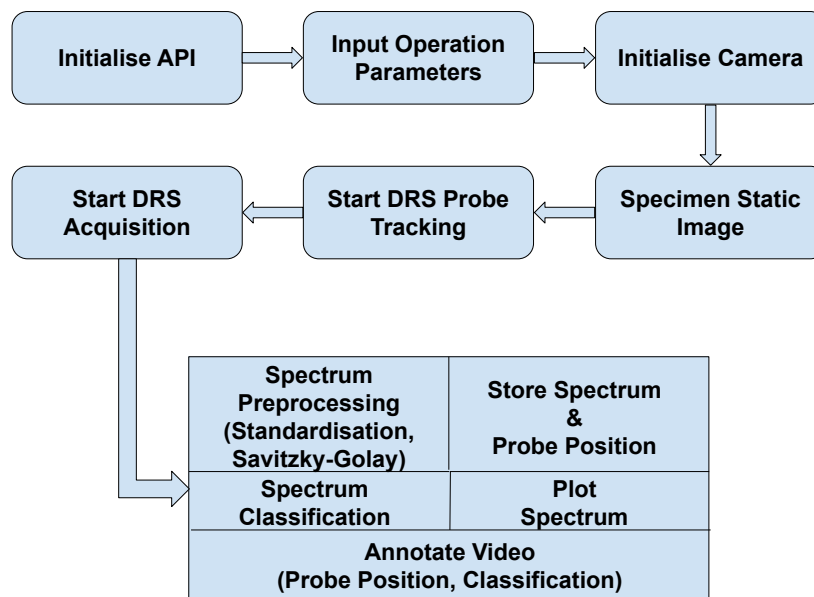


Figure 37: Workflow for ex-vivo data acquisition.

4.6 Data Analysis

The development of a diagnostic algorithm based on spectral information necessitates a couple of pivotal steps such as data normalisation, feature extraction and selection. The latter involves reducing the data used for classification, focusing solely on the data that encapsulate the clinically critical information. In the realm of tissue type classification, various methodologies have been employed. Firstly, direct comparison of single features or simple combinations of features, has been performed to distinguish different tissues^{261,288–292}. Additionally, several machine learning algorithms have been used for the tissue classification. Among these are classification and regression tree algorithms^{179,262,267,293,294}, linear discriminant analysis (LDA)^{197,295–297}, and support vector machines (SVM)^{179,265,295,298–302}. The current study provides a systematic exploration of strategies to optimise the classification of healthy upper GI tissue against tumour tissue in upper GI cancer. This optimisation entails an evaluation of various normalisation and feature extraction and selection techniques that will be used in combination with a variety of classifiers to determine the optimal classification strategy.

All data processing and analysis was performed by in-house developed software (using Python v3.9 and MATLAB v2022a). In the spectra obtained after the calibration performed during spectra acquisition, intensity differences can be observed. These differences might not

only stem from tissue specific difference, but due to factors such as the difference in the distance between the illumination source and the tissue specimen, compared to the distance between the illumination source and the Spectralon® sample used for the calibration. The distance between the illumination source and the sample is crucial. Specifically, the DRS probe must be in direct contact with the sample. This ensures that there is minimal scattering or loss of light due to air gaps or uneven surfaces between the probe and the sample. This direct contact should be consistent between calibration and tissue sample measurements. The quality of contact between the probe and the sample can dramatically affect the acquired spectra. Inconsistent or poor contact can lead to scattering, reflection, or absorption anomalies.

The calibration is performed to standardize all measurements as far as possible. When tissue thickness varied across the specimen, the distance between the illumination source and tissue were not similar for the entire specimen, leading to intensity differences. Moreover, the coupling between the fibres and spectrometer might cause a variation in detection efficiency of the acquired signal.

When light interacts with a surface, some of it undergoes a phenomenon known as specular reflection, where it bounces off the surface at the same angle as the incident light. However, if the surface is uneven or rough, this reflected light is scattered in different directions due to the irregularities. As a result, this scattered light takes on a diffuse character. It's important to note that this diffuse light does not penetrate the tissue and does not carry any information about the internal optical properties of the tissue, such as absorption and scattering. This type of reflection, often referred to as “surface diffuse reflectance”, is different from the diffuse reflectance we are trying to measure in this study.

To compensate for these intensity differences, spectra were normalised before feature extraction and classification. In this study, two different normalisation techniques were compared to each other, and to no normalisation. First, all spectra were normalised by setting the area under the curve (AUC) to one. As a second option, a standard normal variate (SNV) normalisation was used^{303,304}. SNV performs a normalisation of the spectra that consists in subtracting each spectrum by its own mean and dividing it by its own standard deviation. In this way, after SNV, each spectrum will have a mean of zero and a standard deviation of one³⁰⁵.

To deal with outliers, an automated method based on the inter-quartile range (IQR) was implemented for outlier detection and removal from the reflectance spectra dataset. For this method, the median of the residuals was calculated along with the 25th (Q1) and 75th (Q3) percentiles. The difference (i.e., inter-quartile range) between each historical value and the residual mean was then calculated ($Q3 - Q1$). If the historical value was below $Q1 - 1.5 * IQR$ or above $Q3 + 1.5 * IQR$, that value was flagged as an outlier. Due to the presence of noise in the range 420 – 468 nm, only data above 468 nm was used for further analysis. For the stomach and oesophagus datasets, data above 720 nm were also discarded due to the lack of clinically discriminative information. Spectral data above 900 nm were discarded for the colorectal dataset due to noise.

When the number of features in a dataset is very large relative to the number of samples, some classifiers struggle to train effectively, a phenomenon known as “curse of dimensionality”. One key reason is that as the number of features increases, the dataset’s dimensionality expands, leading to increased sparsity and potential difficulties in distinguishing relevant patterns from noise. To address the curse of dimensionality and at the same time reduce overfitting and improve classification accuracy, feature extraction and selection techniques can be used. Besides using the entire spectrum (505 features for the stomach and oesophagus datasets, 1922 features for the colorectal dataset) for classification, feature extraction on each of the stomach, oesophagus and colorectal datasets was performed. The two most prominent peaks of a spectrum were manually selected in the spectral ranges of 480 – 550 nm and 580 – 700 nm. The seven features that were identified and extracted to describe the coarse shape of the spectra while emphasising the differences between different tissue types were: i) maximum reflectance in the 480 – 550 nm range, ii) mean reflectance in the 480 – 550 nm range, iii) wavelength at which maximum reflectance in the 480 – 550 nm occurs, iv) maximum reflectance in the 580 – 700 nm range, v) mean reflectance in the 580 – 720 nm range, vi) wavelength at which maximum reflectance in the 580 – 700 nm occurs, vii) mean reflectance overall across 468 – 720 nm range. These features were hypothesised to be related to underlying tissue differences occurring due to carcinogenesis, such as haemoglobin saturation or tissue thickness changes. Moreover, a tissue optical feature derived following a Monte Carlo simulation that quantifies the effect of changing haemoglobin saturation on the curve features³⁰⁶. The so-called haemoglobin saturation coefficient was calculated for each spectrum.

$$(Equation\ 35): Haemoglobin\ Saturation\ Coefficient = \frac{R_{585nm}}{R_{557nm}}$$

where R_{585nm} is the reflectance at the isosbestic point at 585 nm and R_{557nm} is the reflectance at the second local maximum at 557 nm.

Furthermore, another three feature selection techniques were evaluated, namely the spectral bands clustering, Boruta, and Permutation Importance. First, spectral bands were selected using k-means clustering^{260,299}. Spectral data were divided into spectral bands based on wavelength. These bands were determined using k-means clustering. Values of k were used iteratively; mean silhouette values and classification accuracies were used to assess the best values of k for further analysis. Each spectral band is a region of continuous wavelengths associated with a k-means cluster, containing similar intensity values. Where a cluster contained non-continuous wavelength values, the values within this cluster were divided into separate spectral bands such that a continuous wavelength was maintained within each band. The mean intensity per spectral band was used as input for the further classification method. Additionally, the Boruta³⁰⁷ and Permutation Importance³⁰⁸ feature selection methods were evaluated.

Boruta is an algorithm designed to address the “all-relevant problem”, wherein its primary objective is to identify a subset of features within a given dataset which are relevant to a given classification task. This approach stands in contrast to the “minimal-optimal problem”, which resolves around identifying the minimal subset of features which are performant in a model. The algorithm’s methodology involves taking features of the original dataset and creating shuffled copies of them, referred to as shadow features, to attain randomness. The shadow features are then merged with the original features to obtain a new feature space whose dimension is twice the original dataset. Then, it trains a classifier – Extreme Gradient Boosting in our case – on this extended dataset (original features plus shadow features) which determines their importance using a statistical test, known as the Z-score. During each iteration, the algorithm computes the Z-score of all original and shadow features, and then examines whether an original feature has higher importance than the maximum importance of all shadow features. For this assessment, a statistical test is applied on all original features and keeps memory of its results. The null hypothesis posits that the importance of a particular feature is equal to the maximal importance observed among shadow features. The statistical test assesses the equality between the original and shadow

features. The null hypothesis is rejected when the importance of a feature is significantly higher or lower than one of those of shadow features. Features that fail to reject this hypothesis are removed from the dataset. The features that qualify in the first iteration create the dataset used in the next iteration. The algorithm creates shadow features again using those features and determines their importance similarly to the first iteration. Finally, the Boruta algorithm stops either when all features get confirmed or rejected or it reaches a specified limit of iterations. In this study, the former stopping criterion has been used.

The permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled^{308,309}. This procedure breaks the relationship between the feature and the target, thus the drop in the model score is indicative of how much the model depends on the feature. A feature is important if shuffling its values increases the model error, because in this case the model relied on the feature for the prediction.

After determination of the reduced feature set, data were divided into training and testing sets using the repeated stratified group k-fold cross-validation (SGKCV) method, with 5 folds and 5 repeats. Using this cross-validation method, the dataset was split into five folds such that each fold contained approximately the same percentage of samples of each target class as the complete set. Therefore, the stratified k-fold technique was preferred over the k-fold which deals with classification problems with unbalanced class distributions. The folds were achieved by keeping the fraction of samples with each class constant. This five-fold approach entailed the subdivision of the training data into five equitably balanced subsets, characterized by an equitable representation of both tumorous and healthy tissue samples. The crux of this approach was to ensure that each subset encapsulated the diverse spectrum of tissue types, fostering an effective training regimen that resonates with the real-world distribution of tumour and non-tumour samples. As we deal with patient data, the stratified group k-fold cross-validation method (Figure 38) was finally chosen which attempts to create folds which preserve the percentage of samples for each class as much as possible given the constraint of non-overlapping patient groups between splits. This strategy ensured that data originating from each patient was collectively allocated to either the training or testing set, circumventing any inadvertent intermixing that might lead to data leakage or skewed assessments. The rationale underpinning this patient-wise grouping strategy is rooted in the imperative of preserving the clinical integrity of the data. By retaining data from individual patients within singular partitions, the resulting training and testing subsets mirror the

realistic distribution of patient-specific information encountered in real-world clinical scenarios.

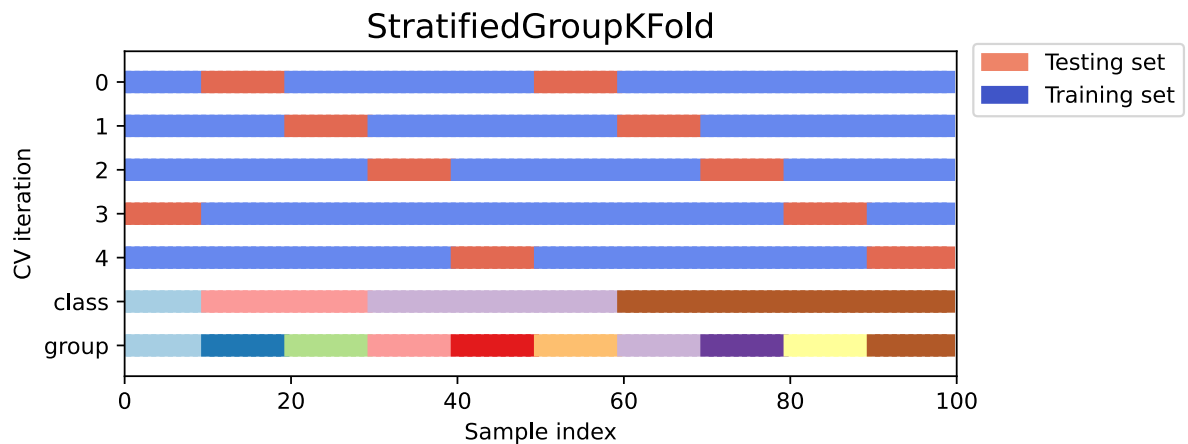


Figure 38: The process of stratified group five-fold cross-validation. When fitting a model, each row represents the split between training (blue) and validation datasets (orange). We typically estimate the final performance metrics through an average of the runs.

Our proposed framework as shown in Figure 40 was divided into two major parts: 1) stratified group k-fold cross-validation (SGKCV) and 2) machine learning classifiers. The seven main stages of this framework include data preparation, data pre-processing, feature extraction and selection, model training/building of four classifier models, applying SGKCV techniques, and performance evaluation using overall accuracy, sensitivity, specificity, and ROC curve.

Tree based classifiers such as the Random Forest (RF)³⁰⁸ Extreme Gradient Boosting (XGBoost or XGB)³¹⁰, Light Gradient Boosting Machine (LightGBM or LGBM)³¹¹ were used to differentiate tumour and non-tumour tissue. In addition, an in-house developed 1D CNN, named SpecNet, was developed and trained to distinguish healthy stomach from gastric cancer and healthy oesophagus from oesophageal cancer, as well as healthy colon tissue from colorectal cancer.

Random Forest is a probabilistic predictor that fits many decision tree classifiers and applies a mean to increase the projected accuracy and reduce overfitting³⁰⁸. The Random Forest algorithm expands upon the bagging method as it utilises both bagging and feature randomness to create an uncorrelated forest of decision trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. Feature randomness generates a random subset of features, which ensures low

correlation among decision trees. This is a key difference between decision trees and Random Forests. While decision trees consider all the possible feature splits, random forests only select a subset of those features. In the Random Forest model, a subset of data points and a subset of features is selected for constructing each decision tree. Individual decision trees are then constructed for each sample, and each of these decision trees generates an output. The final output is considered based on majority voting. An important key advantage of Random Forests is that when there is a robust number of decision trees in a Random Forest, the classifier will not overfit the model since the averaging of uncorrelated trees lowers the overall variance and prediction error.

XGBoost is a scalable decision-tree ensemble machine learning algorithm that uses a gradient boosting framework³¹⁰. XGboost and Gradient Boosting Machines (GBMs) are both ensemble tree methods that apply the principle of boosting weak learners using the gradient descent architecture. However, XGBoost improves upon the base GBM framework through systems optimisation and algorithmic enhancements. XGBoost offers regularisation, which allows to control for overfitting by introducing LASSO (L1) and Ridge (L2) penalties on the weights and biases of each tree to prevent overfitting. Another feature of XGBoost is its ability to handle sparse datasets using the weighted quantile sketch algorithm. This algorithm allows us to deal with non-zero entries in the feature matrix while retaining the same computational complexity as other algorithms like stochastic gradient descent. XGBoost also has a block structure for parallel learning that makes it easy to scale up on multicore machines or clusters. This is possible due to the interchangeable nature of loops used for building base learners; the outer loop that enumerates the leaf nodes of a tree, and the second inner loop that calculates the features. This nesting of loops limits parallelization because without completing the inner loop (more computationally demanding of the two), the outer loop cannot be started. Therefore, to improve run time, the order of loops is interchanged using initialization through a global scan of all instances and sorting using parallel threads. This switch improves algorithmic performance by offsetting any parallelization overheads in computation. It also uses cache awareness, which helps reduce memory usage when training models with large datasets. XGBoost comes with built-in cross-validation method at each iteration of the boosting process and thus, taking away the need to specify the exact optimum number of boosting iterations in a single run.

LightGBM splits the tree vertically (leaf-wise) while other boosting algorithms grow tree horizontally (level-wise)³¹¹. The two novel ideas introduced by LightGBM are Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). These techniques fulfil the limitations of the histogram-based algorithms that is primarily used in all Gradient Boosting Decision Tree (GBDT) frameworks. With GOSS, LGBM can train each tree with only a small fraction of the full dataset. Different data instances have varied roles in the computation of information gain. The instances with larger gradient will contribute more to the information gain. GOSS keeps those instances with large gradient, and only randomly drops those instances with small gradient to retain the accuracy of information gain estimation. With EFB, LGBM handles high-dimensional sparse features much more efficiently. Specifically, in a sparse feature space, many features are mutually exclusive. The exclusive features can be safely bundled into a single feature, called an Exclusive Feature Bundle. Hence, the complexity of histogram building changes from $O(data \times features)$ to $O(data \times bundle)$, while $bundle \ll features$.

The hierarchical architecture of CNNs is proving to be the most efficient and successful way to learn visual representations. Figure 39 shows the hierarchical architecture of the proposed CNN-based SpecNet for spectral classification. The network contains five layers with weights, including the input layer, convolutional layer C1, max pooling layer M2, fully connected layer F3, and output layer. Each spectral measurement in DRS can be regarded as a 2D image, and the size of the input layer is $(n1, 1)$, where $n1$ is the number of spectral bands. The first hidden convolutional layer C1 filters the input data with 20 kernels of size $k1 \times 1$. Layer C1 contains $20 \times n2 \times 1$ nodes, and there are $20 \times (k1 + 1)$ trainable parameters between layers C1 and the input layer. The second hidden layer is the max pooling layer M2 with a kernel size of $(k2, 1)$ and contains $20 \times n3 \times 1$ nodes, where $n3 = n2 / k2$. The fully connected layer F3 has $n4$ nodes, and there are $(20 \times n3 + 1) \times n4$ trainable parameters between this layer and layer M2. The output layer has $n5$ nodes, and there are $(n4 + 1) \times n5$ trainable parameters between this layer and the F3 layer. Classifying a specified spectrum requires the SpecNet with the parameters, where $n1$ and $n5$ are the spectral band size and the number of output classes of the data set, respectively.

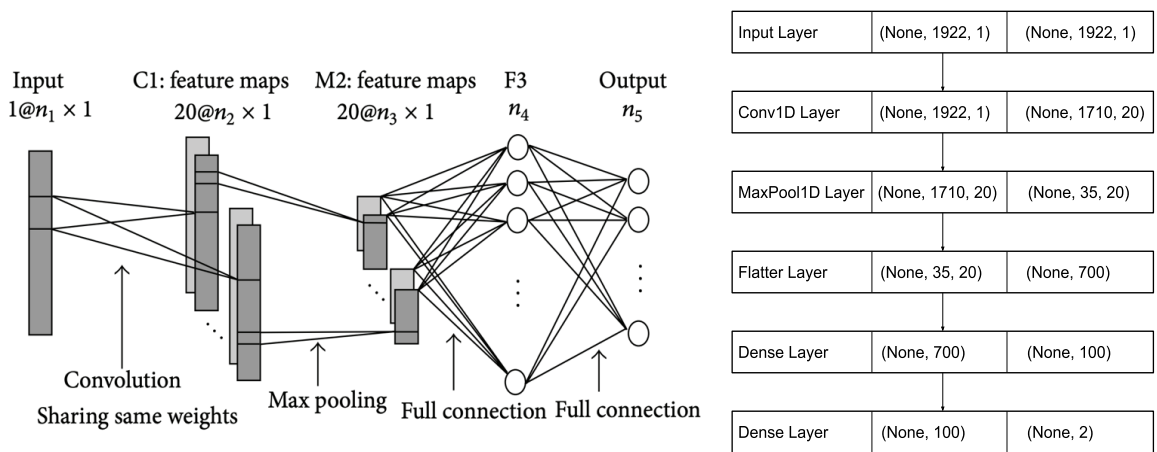


Figure 39: The architecture of the proposed one-dimensional convolutional neural network for colorectal tissue classification. The SpecNet is a five-layer convolutional neural network (CNN) for learning visual representations in diffuse reflectance spectroscopy (DRS) data. It uses 20 kernels to filter the input data, and the output layer classifies spectra based on the spectral band size and number of output classes in the dataset.

In the context of the hyperspectral data of this study, the SpecNet's input layer size is aptly described as $(n_1, 1)$, with n_1 signifying the total number of spectral bands contained within one single spectrum, that is equal to the number of wavelengths in the selected spectral range (505 for the stomach and oesophagus datasets, 1922 for the colorectal dataset). This treatment acknowledges the inherent multidimensionality of the data, condensing the spectral information across the bands into a single-channel representation akin to an image with a single height dimension.

The SpecNet's configuration includes the use of binary cross-entropy as the loss function, which is commonly used for binary classification tasks. The optimiser chosen for training is the Adam optimiser, which adapts the learning rate for each weight of the model, with an unspecified initial learning rate that needs to be defined. To mitigate the issue of overfitting and to enhance the model's generalisation ability, an early stopping mechanism is employed. This mechanism monitors the validation loss and interrupts the training process if there is no improvement in this metric for 10 consecutive training epochs, simultaneously restoring the weights from the epoch with the best validation loss. Furthermore, the model incorporates a learning rate scheduler, specifically the ReduceLROnPlateau, which reduces the learning rate by a factor of 0.1 if there is no improvement in the validation loss for 10 training epochs, with the learning rate being reduced to no lower than $1e-8$. This approach is intended to fine-tune the learning process by adjusting the learning rate in response to the validation loss performance, aiming for more efficient and effective training.

The output data of the neural network comprises binary class predictions that decisively assign each input spectrum sample to either of the two classes: healthy tissue or tumorous tissue. These binary class predictions serve as a categorical affirmation of the network's evaluation of each individual pixel's spectral attributes. Upon traversing the network's layers, which include convolutional, pooling, and fully connected components, the model collectively aggregates and processes spectral patterns representative of each tissue type. This culminates in the final layer, where the network furnishes a definitive declaration regarding the classification label for each spectrum — either "healthy" or "tumour".

The machine learning classifiers were evaluated in terms of sensitivity, specificity, overall accuracy, and the area under the curve. Overall accuracy was calculated as the proportion of correctly identified spectra over the total number of spectra. ROC curves were also plotted.

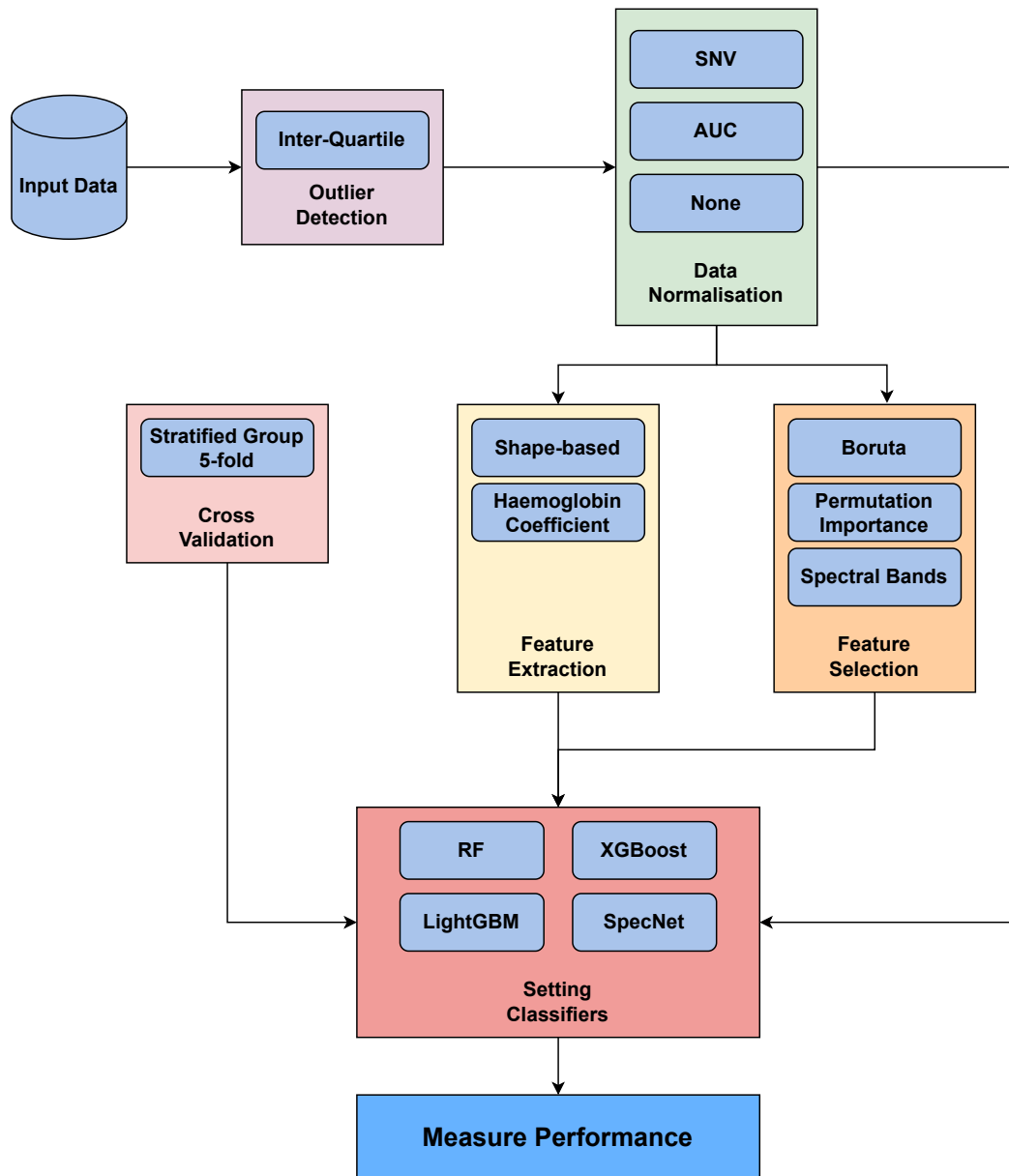


Figure 40: Data analysis framework for gastrointestinal tissue classification.

Classifiers utilising only reflectance values (**Approach 1**) were compared with classifiers utilising raw reflectance values enriched with the seven additional shape-based features (**Approach 2**), with the haemoglobin saturation coefficient tissue optical feature (**Approach 3**) or with both shape-based features and the haemoglobin saturation coefficient tissue optical feature (**Approach 4**).

4.7 Results

4.7.1 *Differentiating Healthy Upper GI Tissue from Cancer: Insights from Ex-Vivo Study*

Cohort Characteristics

A total of 64 patients were recruited for this study between July 2020 and May 2023. Of 64 recruited patients, nine were excluded due to incomplete data acquisition and three were excluded due to overall poor absorption spectra quality, which was defined based on the presence of artifacts or irregularities in intensity values as well as inconsistency when compared to known tissue standards. With these exclusions, a total of 51 patients were included in the analysis.

The mean age of these patients was 69 years (32-89) and 33 (64.7%) patients were male. A total of 30 patients (58.8%) underwent an oesophagectomy and 21 (41.2%) underwent a gastrectomy surgical procedure. The principal histology type of the cohort was adenocarcinoma (93%) and 21 patients (41.1%) had a tumour stage pT3 (TNM staging). All recruited patients had undergone neoadjuvant treatment (chemotherapy, chemoradiotherapy or radiotherapy alone) with the exception of nine patients. Additional patient characteristics can be seen in Table 7.

Table 7: Cohort characteristics of n=51 patients included in the analysis. Characteristics are presented for all patients in no. (%) as well as for gastric cancer and oesophageal cancer patients.

Characteristic	No. (%)		
	All Patients	Patients with cancer	
		Gastric	Oesophageal
Age, median (range)	69 (32-89)	69 (43-89)	69 (32-80)
Sex			
Male	33 (64.7)	14 (66.7)	19 (63.3)
Female	18 (35.3)	7 (33.3)	11 (36.7)
Procedure			
Oesophagectomy	30 (58.8)	0	30 (100)
Gastrectomy	21 (41.2)	21 (41.2)	0
Tumour Histology			
Adenocarcinoma	40 (93.0)	12 (60.0)	15 (65.2)
Squamous Cell Carcinoma	3 (4.6)	2 (10.0)	1 (4.3)
Tumour Stage (TNM staging)			
pT0	8 (15.7)	1 (4.7)	7 (23.3)
pT1	8 (15.7)	6 (28.6)	2 (6.7)
pT2	6 (11.8)	3 (14.3)	3 (10.0)
pT3	21 (41.1)	6 (28.6)	15 (50.0)
pT4	8 (15.7)	5 (23.8)	3 (10.0)
Neoadjuvant therapy			
Chemotherapy	37 (72.5)	13 (61.9)	24 (80.0)
Chemoradiotherapy	4 (7.8)	0	4 (13.3)
Radiotherapy	1 (1.9)	1 (4.7)	0
None	9 (17.6)		

Dataset Summary

Overall, 32 distinct sets of healthy stomach data, 15 sets of gastric cancer data, 26 sets of healthy oesophagus data, and 17 sets of oesophageal cancer data were included in the analysis. There was a total of 6,959 spectra for healthy stomach and 4,050 spectra for gastric cancer. In oesophagus specimens, there was a total of 5,081 spectra for healthy oesophagus and 3,597 spectra for oesophageal cancer.

The means of all spectra with standard deviation for each of the tissue classes are shown in Figure 41.

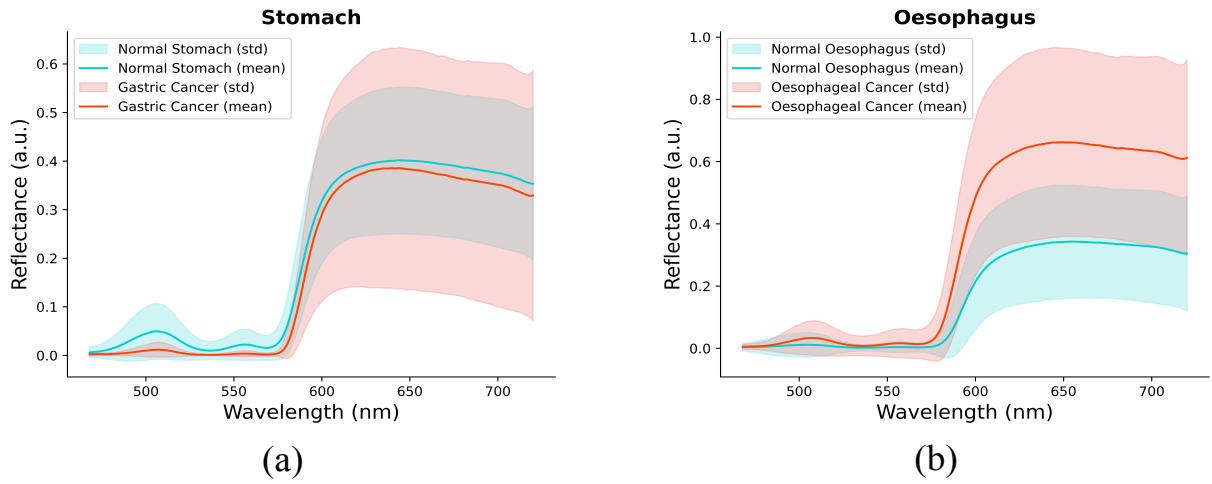


Figure 41: Mean spectra with standard deviation for (a) normal stomach and gastric cancer, and (b) normal oesophagus and oesophageal cancer.

Feature Extraction and Selection

As described in Section 4.6, the first step of the spectral analysis pipeline was to normalise the data using the SNV and AUC methods. Afterwards, the shape-based features and the haemoglobin saturation coefficient were extracted (Figure 42).

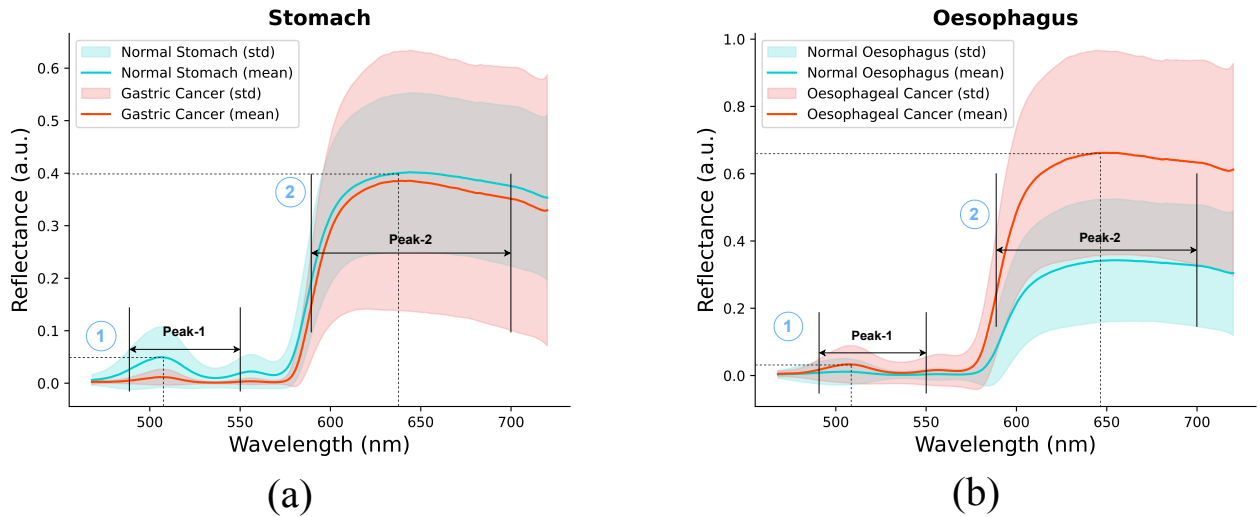


Figure 42: Illustration of shape-based features extracted for (a) stomach and (b) oesophagus datasets. (1) Depicts the secondary peak intensity value and corresponding wavelength in the 480 – 550 nm spectral range. (2) Depicts the primary peak intensity value and corresponding wavelength in the 580 – 700 nm spectral range.

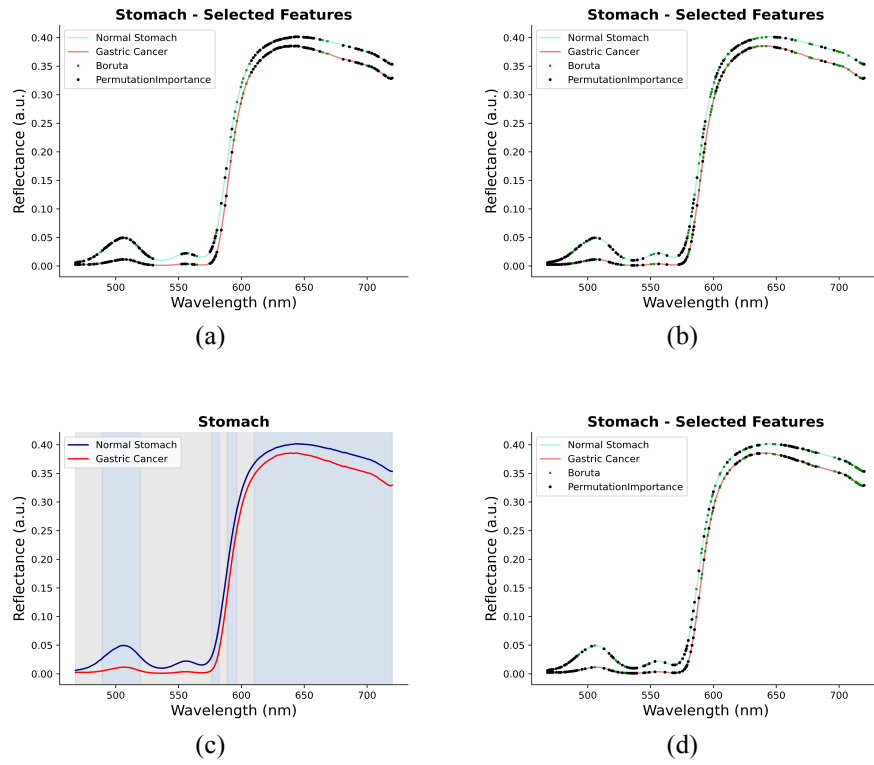


Figure 43: Selected features for the stomach dataset for each normalisation method. (a) Boruta and Permutation Feature Importance after SNV normalisation. (b) Boruta and Permutation Feature Importance after AUC normalisation. (c) Illustration of the 8 selected spectral bands. Light grey and light blue areas represent individual spectral bands. (d) Boruta and Permutation Feature Importance with no normalisation.

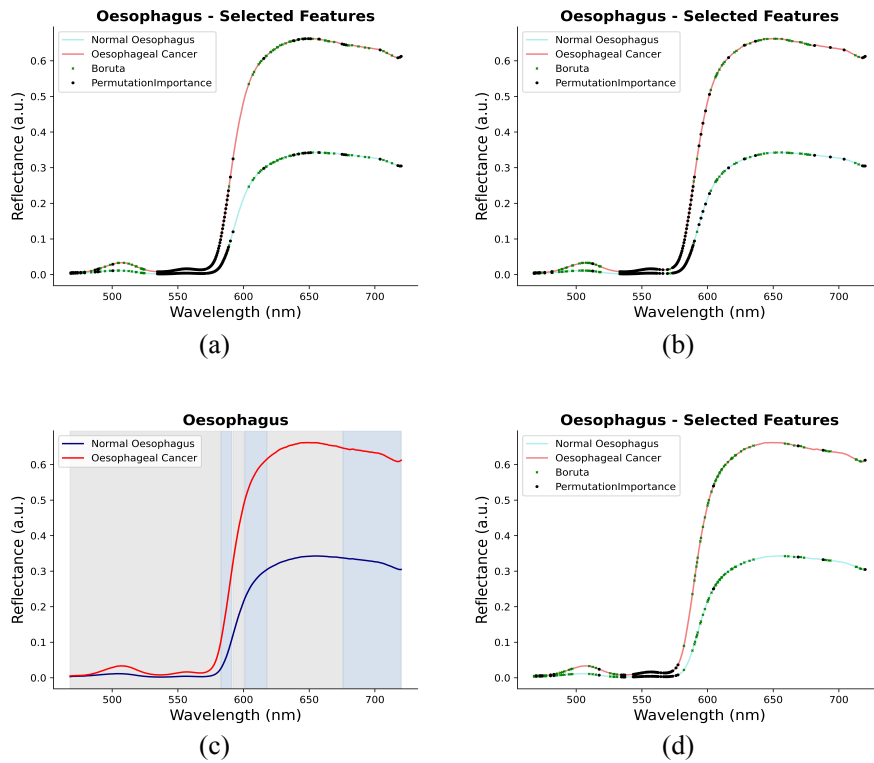


Figure 44: Selected features for the oesophagus dataset for each normalisation method. (a) Boruta and Permutation Feature Importance after SNV normalisation. (b) Boruta and Permutation Feature Importance after AUC normalisation. (c) Illustration of the 6 selected spectral bands. Light grey and light blue areas represent individual spectral bands. (d) Boruta and Permutation Feature Importance with no normalisation.

Each of the feature selection methods, namely spectral bands, Boruta and Permutation Feature Importance was applied on each dataset following SNV, AUC and no normalisation. The selected features are shown in Figure 43 and Figure 44 for the stomach and oesophagus datasets, respectively.

Exploring the Impact of Normalisation on Classification Performance

The added value of normalisation methods on the classification outcome was also examined. To this end, the increase in overall accuracy, sensitivity, specificity, and AUC for each normalisation method compared to no normalisation were analysed. This analysis was performed for the entire spectrum for the Random Forest, XGBoost, LightGBM and the SpecNet classifiers. The average increase in the aforementioned performance metrics per normalisation technique for each of the four approaches are shown in the following tables. No feature selection was considered for this investigation.

Table 8: Mean increase in overall accuracy, sensitivity, specificity, and AUC for all normalisation techniques compared to no normalisation for all classifiers for the stomach dataset.

Classifier	Mean (SD) (%)							
	SNV normalisation				AUC normalisation			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra								
RF	2.66 (0.03)	3.06 (0.24)	2.42 (0.02)	1.57 (0.02)	0.75 (0.24)	0.06 (0.45)	1.16 (0.12)	0.27 (0.11)
XGB	3.36 (0.07)	4.52 (0.21)	2.69 (0.008)	2.05 (0.11)	0.23 (0.07)	-0.46 (0.02)	0.64 (0.22)	0.17 (0.003)
LGBM	3.11 (0.01)	4.30 (0.12)	2.42 (0.04)	1.89 (0.08)	0.33 (0.09)	-0.39 (0.09)	0.75 (0.12)	0.21 (0.00)
SpecNet	0.18 (0.03)	1.65 (0.19)	-0.67 (0.12)	0.10 (0.03)	-6.17 (0.68)	-3.05 (0.29)	-7.99 (0.96)	-3.53 (0.73)
Approach 2 – Reflectance spectra with shape-based features								
RF	2.58 (0.07)	3.94 (0.13)	1.78 (0.13)	1.96 (0.07)	0.96 (0.11)	0.80 (0.23)	1.04 (0.26)	0.66 (0.04)
XGB	3.99 (0.25)	5.39 (0.13)	3.17 (0.32)	2.58 (0.19)	1.87 (0.004)	2.31 (0.27)	1.62 (0.10)	1.17 (0.009)
LGBM	3.82 (0.08)	5.33 (0.43)	2.94 (0.007)	2.37 (0.12)	1.45 (0.09)	1.46 (0.49)	1.44 (0.16)	0.99 (0.03)
SpecNet	-0.48 (0.06)	-1.56 (0.15)	0.14 (0.02)	-0.23 (0.05)	-6.45 (0.60)	-9.30 (0.69)	-4.79 (0.52)	-3.81 (0.65)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	2.15 (0.09)	3.17 (0.11)	1.56 (0.11)	1.65 (0.09)	0.33 (0.01)	-0.41 (0.16)	0.77 (0.03)	0.14 (0.07)
XGB	3.06 (0.26)	4.25 (0.06)	2.36 (0.29)	1.99 (0.12)	0.20 (0.06)	0.75 (0.09)	0.76 (0.22)	-0.15 (0.09)
LGBM	2.83 (0.20)	4.12 (0.24)	2.08 (0.03)	1.85 (0.09)	0.01 (0.09)	-0.77 (0.18)	0.46 (0.10)	-0.19 (0.05)
SpecNet	0.79 (0.08)	-1.32 (0.09)	2.01 (0.26)	0.07 (0.01)	-2.66 (0.24)	-4.36 (0.29)	-1.67 (0.18)	-2.35 (0.39)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	2.59 (0.12)	3.90 (0.02)	1.83 (0.17)	1.88 (0.09)	0.91 (0.15)	0.87 (0.08)	0.93 (0.23)	0.54 (0.03)
XGB	3.59 (0.20)	4.86 (0.25)	2.85 (0.09)	2.08 (0.16)	1.47 (0.15)	1.81 (0.29)	1.28 (0.10)	0.90 (0.05)
LGBM	3.59 (0.01)	4.69 (0.06)	2.95 (0.11)	2.07 (0.13)	1.35 (0.01)	1.09 (0.25)	1.50 (0.05)	0.88 (0.03)
SpecNet	2.24 (0.23)	-3.54 (0.34)	5.60 (0.61)	0.42 (0.09)	-1.24 (0.11)	-10.29 (0.81)	4.02 (0.41)	-1.70 (0.30)

In Table 8 it shows that the SNV normalisation technique improved the classification outcome significantly (mean overall increase 2.94%) for the stomach dataset using the Random Forest, XGBoost, and LightGBM classifiers compared to the AUC normalisation (mean overall increase 0.73%). However, when considering the SpecNet classifier, this improved performance trend did not persist. Contrarily, a decline in performance was observed for both normalisation techniques. Some of the reasons behind this decrease in classification performance could lie under the way SNV and AUC normalisation techniques work. SNV normalisation modifies the spectral shape, which can be crucial since the preservation of specific spectral patterns may play a pivotal role in the accurate classification of spectral data. Any alteration to these patterns could affect their repeatability across datasets, making the classifier's task more challenging. AUC normalisation, meanwhile, emphasises changing the relative intensities across wavelengths. Such transformations could inadvertently mask or dilute specific spectral features that are instrumental for accurate classification.

Table 9: Mean increase in overall accuracy, sensitivity, specificity, and AUC for all normalisation techniques compared to no normalisation for all classifiers for the oesophagus dataset.

Classifier	Mean (SD) (%)							
	SNV normalisation				AUC normalisation			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra								
RF	-0.19 (0.05)	-1.70 (0.49)	0.88 (0.14)	-0.51 (0.04)	-1.25 (0.18)	-2.36 (0.58)	-0.47 (0.14)	-1.33 (0.11)
XGB	1.44 (0.07)	0.57 (0.19)	2.05 (0.05)	0.45 (0.02)	-0.20 (0.04)	-1.31 (0.10)	0.58 (0.05)	-0.54 (0.06)
LGBM	1.04 (0.13)	-0.11 (0.37)	1.85 (0.18)	0.21 (0.09)	-0.50 (0.15)	-1.64 (0.41)	0.32 (0.09)	-0.63 (0.02)
SpecNet	-1.92 (0.25)	-2.50 (0.25)	-1.51 (0.26)	-0.44 (0.12)	-7.02 (0.77)	-15.09 (1.09)	-1.31 (0.23)	-2.52 (0.56)
Approach 2 – Reflectance spectra with shape-based features								
RF	3.29 (0.02)	2.99 (0.30)	3.50 (0.29)	1.38 (0.15)	2.29 (0.05)	1.87 (0.19)	2.60 (0.34)	0.95 (0.12)
XGB	3.56 (0.14)	3.94 (0.22)	3.29 (0.38)	1.70 (0.24)	2.70 (0.03)	2.74 (0.15)	2.67 (0.24)	1.15 (0.16)
LGBM	3.64 (0.25)	3.97 (0.17)	3.40 (0.43)	1.64 (0.19)	2.66 (0.23)	2.88 (0.29)	2.51 (0.28)	1.18 (0.19)
SpecNet	-4.99 (0.62)	-8.15 (0.70)	-2.75 (0.52)	-2.34 (0.59)	-6.76 (0.79)	-10.83 (0.86)	-3.87 (0.69)	-3.36 (0.78)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	0.76 (0.03)	-1.27 (0.48)	2.20 (0.11)	0.15 (0.03)	-0.43 (0.03)	-1.88 (0.52)	0.60 (0.25)	-1.02 (0.09)
XGB	2.55 (0.05)	1.76 (0.03)	3.11 (0.44)	1.19 (0.09)	0.83 (0.02)	-0.36 (0.16)	1.67 (0.38)	0.03 (0.002)
LGBM	2.12 (0.15)	1.14 (0.11)	2.82 (0.29)	0.89 (0.01)	0.55 (0.01)	-0.41 (0.02)	1.22 (0.21)	-0.07 (0.02)
SpecNet	-0.23 (0.03)	5.19 (0.36)	-4.07 (0.77)	-0.80 (0.20)	-4.22 (0.40)	-3.15 (0.18)	-4.98 (0.89)	-3.00 (0.61)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	3.23 (0.00)	3.22 (0.30)	3.24 (0.37)	1.35 (0.16)	2.25 (0.02)	1.93 (0.37)	2.48 (0.34)	0.92 (0.09)
XGB	3.49 (0.21)	4.11 (0.12)	3.05 (0.35)	1.61 (0.19)	2.63 (0.22)	2.93 (0.12)	2.42 (0.27)	1.10 (0.17)
LGBM	3.42 (0.27)	4.10 (0.26)	2.94 (0.52)	1.51 (0.16)	2.69 (0.20)	3.16 (0.32)	2.37 (0.23)	1.13 (0.14)
SpecNet	-5.26 (0.68)	-9.63 (0.86)	-2.16 (0.42)	-2.09 (0.56)	-7.45 (0.88)	-14.35 (1.13)	-2.56 (0.48)	-3.41 (0.80)

According to Table 9, for the oesophagus dataset, SNV normalisation technique slightly improved the classification outcome (mean overall increase 1.98%) for the stomach dataset using the Random Forest, XGBoost, and LightGBM classifiers compared to the AUC normalisation (mean overall increase 0.95%). For the SpecNet classifier, the classification performance was not improved, as for every approach and classifier only a decrease was seen among all approaches for both normalisation techniques. The reasons behind this decrease in classification performance for the SpecNet were explained above for the stomach dataset.

Because SNV normalisation seemed useful for most of the classification algorithms, and AUC normalisation did not seem useful at all, it was decided to continue with using only SNV normalisation for the comparison of different feature selection and extraction techniques in the next section.

Analysing the Influence of Feature Extraction on Classification Performance

Based on the results from the normalisation, the first analysis on feature extraction and selection was performed on non-normalised data and data normalised using the SNV normalisation technique. In Table 10 and Table 11 the results are shown for the mean increase in classification performance metrics for all feature extraction methods (approaches 2, 3, 4) compared to no feature extraction, i.e., the entire spectrum only (approach 1), for the stomach and oesophagus datasets, respectively.

Table 10: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the stomach dataset. Only non-normalised and SNV normalised data was used.

Classifier	Mean (SD) (%)							
	No normalisation				SNV normalisation			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 2 – Reflectance spectra with shape-based features								
RF	-0.15 (0.09)	-2.10 (0.45)	0.99 (0.12)	-0.43 (0.01)	-0.23 (0.01)	-1.22 (0.08)	0.34 (0.03)	-0.04 (0.04)
XGB	0.63 (0.04)	0.32 (0.06)	0.81 (0.22)	0.28 (0.02)	1.25 (0.14)	1.19 (0.02)	1.29 (0.09)	0.81 (0.10)
LGBM	0.63 (0.05)	0.50 (0.24)	0.70 (0.11)	0.32 (0.06)	1.34 (0.12)	1.53 (0.06)	1.22 (0.08)	0.81 (0.09)
SpecNet	-1.91 (0.25)	-1.15 (0.12)	-2.35 (0.37)	-0.89 (0.25)	-2.57 (0.33)	-4.36 (0.46)	-1.53 (0.23)	-1.22 (0.33)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	-0.61 (0.14)	-2.97 (0.31)	0.76 (0.14)	-0.71 (0.05)	-1.12 (0.02)	-2.86 (0.17)	-0.11 (0.00)	-0.63 (0.02)
XGB	0.40 (0.00)	-1.02 (0.34)	1.23 (0.32)	0.07 (0.01)	0.09 (0.19)	-1.29 (0.19)	0.90 (0.05)	0.01 (0.02)
LGBM	0.21 (0.15)	-1.47 (0.18)	1.19 (0.07)	0.00 (0.00)	-0.07 (0.03)	-1.65 (0.06)	0.85 (0.08)	-0.05 (0.00)
SpecNet	-4.30 (0.51)	-5.35 (0.48)	-3.69 (0.54)	-1.73 (0.43)	-3.69 (0.45)	-8.31 (0.77)	-1.01 (0.16)	-1.76 (0.45)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	0.66 (0.05)	-0.76 (0.32)	1.48 (0.03)	0.18 (0.00)	-0.05 (0.02)	-1.07 (0.07)	0.54 (0.03)	0.05 (0.03)
XGB	1.92 (0.01)	2.04 (0.14)	1.85 (0.09)	1.20 (0.04)	2.14 (0.12)	2.38 (0.18)	2.01 (0.01)	1.23 (0.08)
LGBM	1.71 (0.03)	2.09 (0.08)	1.49 (0.16)	1.03 (0.05)	2.19 (0.03)	2.48 (0.15)	2.02 (0.02)	1.21 (0.10)
SpecNet	-4.75 (0.55)	0.16 (0.02)	-7.61 (0.93)	-1.49 (0.38)	-2.69 (0.35)	-5.02 (0.52)	-1.34 (0.20)	-1.17 (0.32)

Table 11: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the oesophagus dataset. Only non-normalised and SNV normalised data was used.

Classifier	Mean (SD) (%)							
	No normalisation				SNV normalisation			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 2 – Reflectance spectra with shape-based features								
RF	-0.96 (0.14)	-0.41 (0.04)	-1.35 (0.33)	-0.64 (0.14)	2.52 (0.11)	4.29 (0.15)	1.26 (0.09)	1.24 (0.04)
XGB	-0.14 (0.05)	0.26 (0.00)	-0.43 (0.34)	-0.32 (0.14)	1.98 (0.16)	3.63 (0.40)	0.81 (0.09)	0.94 (0.08)
LGBM	-0.20 (0.24)	0.27 (0.25)	-0.54 (0.48)	-0.25 (0.10)	2.40 (0.14)	4.35 (0.29)	1.01 (0.14)	1.18 (0.18)
SpecNet	0.69 (0.10)	-0.09 (0.01)	1.25 (0.26)	0.47 (0.15)	-2.38 (0.26)	-5.74 (0.45)	0.00 (0.00)	-1.43 (0.32)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	-0.95 (0.11)	-0.41 (0.04)	-1.33 (0.38)	-0.63 (0.11)	0.00 (0.02)	0.02 (0.05)	0.02 (0.13)	-0.28 (0.05)
XGB	-0.81 (0.08)	-0.86 (0.03)	-0.77 (0.36)	-0.72 (0.15)	0.31 (0.04)	0.33 (0.13)	0.29 (0.13)	0.03 (0.08)
LGBM	-0.67 (0.13)	-0.60 (0.21)	-0.71 (0.37)	-0.60 (0.09)	0.42 (0.15)	0.65 (0.05)	0.26 (0.09)	0.08 (0.08)
SpecNet	-3.15 (0.4)	-10.0 (0.82)	1.70 (0.38)	-0.32 (0.09)	-1.46 (0.17)	-2.31 (0.20)	-0.85 (0.13)	-0.69 (0.17)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	-0.79 (0.16)	-0.30 (0.03)	-1.15 (0.40)	-0.53 (0.12)	2.62 (0.11)	4.62 (0.16)	1.21 (0.11)	1.33 (0.07)
XGB	0.19 (0.02)	0.44 (0.21)	0.00 (0.24)	-0.14 (0.09)	2.24 (0.26)	3.98 (0.53)	1.00 (0.17)	1.02 (0.08)
LGBM	0.21 (0.16)	0.42 (0.23)	0.07 (0.39)	-0.04 (0.05)	2.60 (0.25)	4.63 (0.40)	1.16 (0.31)	1.26 (0.20)
SpecNet	1.11 (0.17)	1.11 (0.13)	1.11 (0.23)	0.56 (0.18)	-2.23 (0.25)	-6.02 (0.47)	0.46 (0.07)	-1.09 (0.25)

In Table 10 and Table 11 the results are shown for the mean increase in performance metrics for all feature extraction methods compared to no feature extraction (entire spectrum only) for the stomach and oesophagus datasets, respectively. For both datasets, it can be concluded that overall using shape-based features combined with the haemoglobin saturation coefficient feature improved classification results compared to the use of the entire spectrum only. When shape-based features and haemoglobin saturation coefficient applied separately on the dataset, almost no difference compared to using the entire spectrum only for classification was seen. Moreover, when SNV normalisation is used, a higher increase in performance metrics is present using both shape-based and haemoglobin saturation coefficient features compared to no normalisation. Therefore, for the following analysis on the effect of feature selection methods on the classification performance, only the SNV normalisation technique is considered.

In Table 12 and Table 13 the results are shown for the mean increase in classification performance metrics for Boruta and Permutation Importance feature selection methods compared to no feature selection, i.e., the entire spectrum, using SNV normalisation for the stomach and oesophagus datasets, respectively.

Table 12: Mean increase in classification performance metrics for Boruta and Permutation Importance feature selection methods compared to the use of the entire spectrum for the stomach dataset. Only SNV normalised data was used.

Classifier	Mean (SD) (%)							
	Boruta feature selection				Permutation Importance feature selection			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra								
RF	0.26 (0.15)	0.41 (0.22)	0.17 (0.11)	0.06 (0.01)	0.14 (0.08)	0.25 (0.00)	0.08 (0.02)	0.07 (0.02)
XGB	0.06 (0.02)	0.01 (0.05)	0.09 (0.07)	0.08 (0.05)	0.10 (0.00)	-0.07 (0.20)	0.19 (0.10)	0.09 (0.00)
LGBM	0.05 (0.00)	-0.06 (0.08)	0.12 (0.02)	0.02 (0.04)	0.13 (0.11)	0.07 (0.18)	0.16 (0.14)	0.05 (0.02)
SpecNet	-4.51 (0.54)	-1.56 (0.18)	-6.23 (0.75)	-1.81 (0.46)	-3.51 (0.43)	-3.21 (0.35)	-3.69 (0.49)	-1.74 (0.45)
Approach 2 – Reflectance spectra with shape-based features								
RF	0.49 (0.03)	1.04 (0.12)	0.16 (0.05)	0.34 (0.03)	0.66 (0.04)	1.20 (0.05)	0.35 (0.09)	0.36 (0.02)
XGB	0.02 (0.16)	0.12 (0.06)	-0.04 (0.18)	0.03 (0.01)	0.22 (0.14)	0.62 (0.19)	0.00 (0.03)	0.06 (0.02)
LGBM	-0.08 (0.06)	-0.18 (0.07)	-0.02 (0.00)	-0.07 (0.02)	-0.10 (0.04)	-0.21 (0.15)	-0.04 (0.03)	-0.03 (0.01)
SpecNet	-1.48 (0.16)	4.12 (0.43)	-4.74 (0.52)	-0.11 (0.03)	-0.06 (0.00)	1.40 (0.13)	-0.91 (0.12)	0.09 (0.02)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	0.39 (0.05)	0.80 (0.30)	0.14 (0.05)	0.12 (0.00)	0.39 (0.05)	0.81 (0.13)	0.15 (0.07)	0.15 (0.00)
XGB	0.04 (0.15)	0.19 (0.11)	-0.05 (0.06)	0.04 (0.00)	0.02 (0.17)	0.16 (0.08)	-0.06 (0.19)	0.06 (0.01)
LGBM	0.02 (0.06)	0.11 (0.19)	-0.03 (0.05)	-0.01 (0.02)	0.11 (0.02)	0.51 (0.10)	-0.13 (0.03)	0.06 (0.02)
SpecNet	2.27 (0.26)	6.99 (0.61)	-0.48 (0.07)	1.13 (0.26)	2.09 (0.24)	7.24 (0.64)	-0.91 (0.13)	1.11 (0.25)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	0.65 (0.14)	1.44 (0.15)	0.20 (0.08)	0.38 (0.07)	0.54 (0.02)	1.08 (0.00)	0.22 (0.08)	0.34 (0.04)
XGB	0.28 (0.10)	0.30 (0.04)	0.26 (0.06)	0.12 (0.07)	0.21 (0.09)	0.37 (0.06)	0.11 (0.07)	0.09 (0.03)
LGBM	0.11 (0.21)	0.08 (0.28)	0.13 (0.07)	0.04 (0.04)	0.07 (0.03)	0.19 (0.18)	0.00 (0.07)	0.03 (0.03)
SpecNet	0.18 (0.02)	2.63 (0.25)	-1.25 (0.16)	0.09 (0.02)	-0.33 (0.04)	2.96 (0.28)	-2.25 (0.28)	-0.14 (0.03)

Table 13: Mean increase in classification performance metrics for Boruta and Permutation Importance feature selection methods compared to the use of the entire spectrum for the oesophagus dataset. Only SNV normalised data was used.

Classifier	Mean (SD) (%)							
	Boruta feature selection				Permutation Importance feature selection			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra								
RF	0.24 (0.04)	0.18 (0.26)	0.27 (0.14)	0.13 (0.00)	-0.82 (0.00)	-0.61 (0.17)	-0.96 (0.02)	-0.72 (0.04)
XGB	0.14 (0.19)	0.23 (0.17)	0.07 (0.00)	0.05 (0.00)	-0.42 (0.08)	-0.24 (0.07)	-0.56 (0.05)	-0.25 (0.04)
LGBM	0.00 (0.15)	0.27 (0.18)	-0.19 (0.07)	-0.02 (0.11)	-0.61 (0.00)	-0.36 (0.19)	-0.78 (0.09)	-0.36 (0.00)
SpecNet	-2.65 (0.29)	-2.22 (0.19)	-2.95 (0.40)	-1.22 (0.28)	-4.26 (0.44)	-2.78 (0.24)	-5.31 (0.65)	-2.42 (0.50)
Approach 2 – Reflectance spectra with shape-based features								
RF	0.41 (0.01)	1.04 (0.06)	-0.04 (0.00)	0.31 (0.04)	-0.42 (0.06)	0.30 (0.04)	-0.93 (0.07)	-0.19 (0.02)
XGB	0.19 (0.04)	0.15 (0.11)	0.22 (0.00)	0.08 (0.02)	-0.05 (0.00)	-0.12 (0.11)	0.00 (0.08)	-0.03 (0.02)
LGBM	0.03 (0.09)	0.07 (0.02)	0.01 (0.01)	0.03 (0.00)	-0.09 (0.01)	-0.11 (0.00)	-0.07 (0.08)	-0.06 (0.00)
SpecNet	6.14 (0.82)	12.22 (1.23)	1.84 (0.32)	2.78 (0.77)	-0.88 (0.09)	1.94 (0.14)	-2.89 (0.39)	-1.44 (0.25)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	0.64 (0.00)	0.83 (0.07)	0.50 (0.11)	0.42 (0.03)	-0.12 (0.00)	0.18 (0.05)	-0.33 (0.24)	-0.25 (0.07)
XGB	0.17 (0.02)	0.43 (0.07)	0.00 (0.23)	0.05 (0.04)	-0.22 (0.16)	-0.23 (0.21)	-0.21 (0.03)	-0.11 (0.02)
LGBM	0.06 (0.03)	0.36 (0.19)	-0.15 (0.05)	0.07 (0.00)	-0.32 (0.01)	-0.09 (0.18)	-0.48 (0.00)	-0.08 (0.00)
SpecNet	-2.50 (0.25)	-4.26 (0.31)	-1.25 (0.17)	-2.09 (0.39)	1.07 (0.12)	-3.06 (0.23)	4.00 (0.75)	1.11 (0.28)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	0.48 (0.03)	1.17 (0.02)	-0.02 (0.12)	0.32 (0.02)	-0.26 (0.06)	0.24 (0.09)	-0.62 (0.09)	-0.17 (0.04)
XGB	0.24 (0.21)	0.09 (0.32)	0.33 (0.06)	0.08 (0.01)	0.13 (0.09)	0.08 (0.06)	0.17 (0.09)	0.00 (0.00)
LGBM	0.13 (0.16)	0.11 (0.08)	0.14 (0.15)	0.04 (0.02)	-0.03 (0.15)	0.01 (0.16)	-0.06 (0.08)	-0.04 (0.01)
SpecNet	4.53 (0.56)	6.67 (0.53)	3.02 (0.64)	1.98 (0.52)	-1.04 (0.10)	0.56 (0.04)	-1.38 (0.21)	-1.63 (0.30)

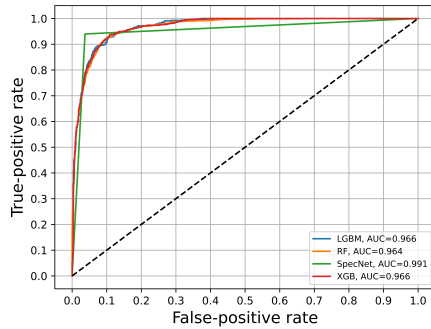
Our investigation revealed an intriguing outcome when applying Boruta and Permutation Importance to XGBoost, LightGBM and Random Forest classifiers. Contrary to initial expectations, the incorporation of these feature selection methods did not result in noticeable improvements in classification performance. The lack of discernible enhancement in classification performance could be attributed to the inherent adaptability of ensemble-based methods, which tend to effectively handle both relevant and redundant features without requiring substantial external interventions. Similar results were obtained for the spectral bands feature selection method, which are not included in the Table 12 and Table 13. The most pronounced effect was observed in the context of the SpecNet. From Table 12 and Table 13, the results are shown that feature selection improved the classification performance for the SpecNet. When augmented with Boruta and Permutation importance, the SpecNet leverages these techniques to identify and amplify the most informative spectral features that encapsulate the class-specific variations. This, in turn, empowers the network to make more precise and confident predictions.

Healthy versus Tumour Tissue Classification Results

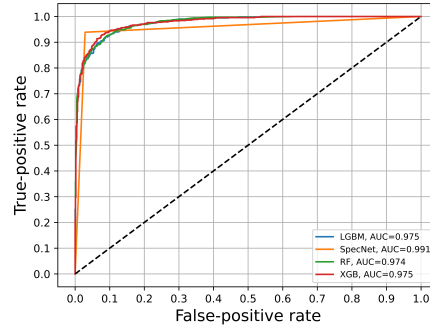
So far, normalisation and feature extraction and selection methods were examined separately. Results of the classification for stomach and oesophagus datasets are presented below. In the following tables, the classification performance metrics are shown for the best combinations of normalisation, feature extraction and selection, and classification method for each approach described in Section 4.6. Data showing the ability of multiple classifiers to detect normal versus cancerous tissue presented as mean (standard deviation). Overall accuracy is calculated as the proportion of correctly identified spectra over the total number of spectra averaged over the stratified group 5-fold cross-validation.

Table 14: Performance metrics for all classifiers using Approach 1 (entire spectrum) for stomach dataset (SNV normalisation, no feature selection) and oesophagus dataset (no normalisation, Boruta feature selection).

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	89.93 (0.56)	85.35 (1.25)	92.60 (0.68)	96.38 (0.36)	91.58 (0.64)	88.18 (1.04)	93.99 (0.65)	97.42 (0.33)
XGB	90.42 (0.58)	87.03 (1.31)	92.39 (0.64)	96.64 (0.33)	91.75 (0.62)	88.88 (1.22)	93.78 (0.69)	97.49 (0.36)
LGBM	90.31 (0.57)	86.90 (1.23)	92.30 (0.68)	96.60 (0.33)	91.65 (0.54)	89.04 (0.96)	93.49 (0.60)	97.47 (0.34)
SpecNet	95.43 (1.43)	93.99 (1.62)	96.26 (1.30)	99.07 (0.65)	95.78 (1.55)	93.89 (1.84)	97.11 (1.28)	99.14 (0.72)



(a)

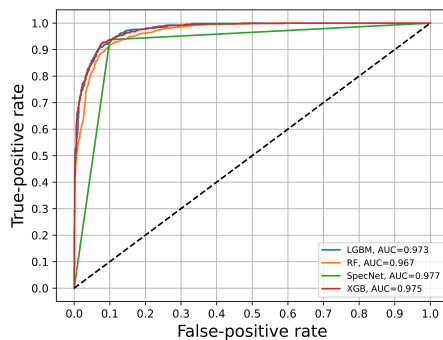


(b)

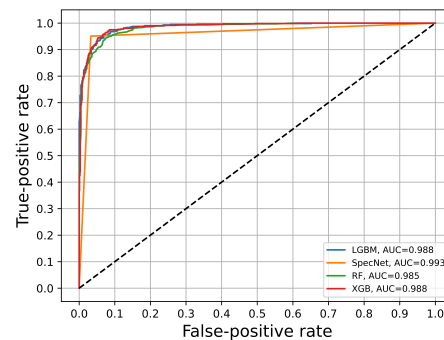
Figure 45: ROC curves with AUC values for all classifiers using Approach 1 (entire spectrum) for (a) stomach dataset (SNV normalisation, no feature selection) and (b) oesophagus dataset (no normalisation, Boruta feature selection).

Table 15: Performance metrics for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for stomach dataset (SNV normalisation, Boruta feature selection) and oesophagus dataset (SNV normalisation, Boruta feature selection).

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	90.18 (0.52)	85.17 (1.21)	93.11 (0.61)	96.68 (0.30)	94.19 (0.78)	91.63 (1.09)	95.99 (0.69)	98.47 (0.28)
XGB	91.67 (0.60)	88.34 (1.38)	93.64 (0.72)	97.50 (0.24)	94.83 (0.63)	93.12 (0.97)	96.04 (0.68)	98.77 (0.22)
LGBM	91.57 (0.52)	88.25 (1.10)	93.50 (0.60)	97.35 (0.22)	94.80 (0.68)	93.12 (1.03)	95.99 (0.67)	98.76 (0.24)
SpecNet	91.37 (1.92)	93.74 (1.65)	89.99 (2.04)	97.74 (1.02)	96.05 (1.50)	95.09 (1.66)	96.72 (1.37)	99.31 (0.64)



(a)

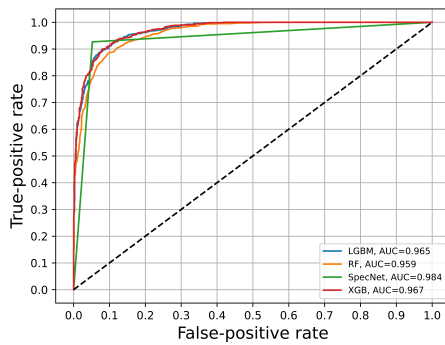


(b)

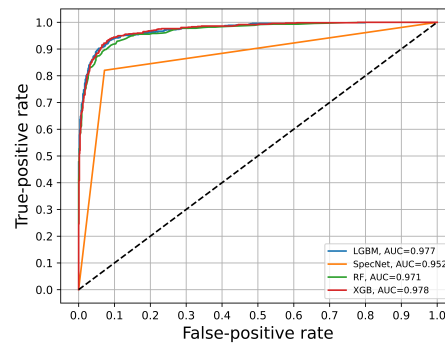
Figure 46: ROC curves with AUC values for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for (a) stomach dataset (SNV normalisation, Boruta feature selection) and (b) oesophagus dataset (SNV normalisation, Boruta feature selection).

Table 16: Performance metrics for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for stomach dataset (SNV normalisation, Boruta feature selection) and oesophagus dataset (SNV normalisation, Boruta feature selection).

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	89.20 (0.53)	83.30 (1.13)	92.63 (0.64)	95.87 (0.35)	91.90 (0.68)	87.14 (1.32)	95.26 (0.81)	97.05 (0.38)
XGB	90.55 (0.54)	85.93 (1.23)	93.24 (0.74)	96.69 (0.30)	93.13 (0.69)	90.10 (1.20)	95.29 (0.88)	97.82 (0.32)
LGBM	90.27 (0.48)	85.36 (1.10)	93.12 (0.71)	96.54 (0.31)	92.85 (0.62)	89.71 (1.06)	95.08 (0.76)	97.71 (0.34)
SpecNet	94.01 (1.62)	92.67 (1.77)	94.77 (1.52)	98.44 (0.85)	88.33 (2.46)	82.03 (2.94)	92.79 (1.98)	95.18 (1.64)



(a)

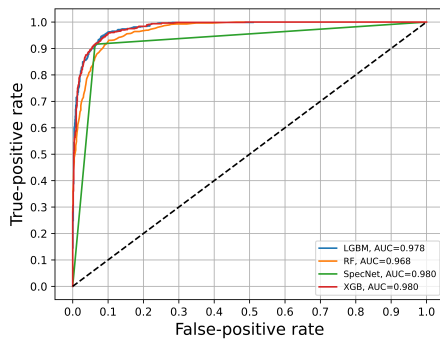


(b)

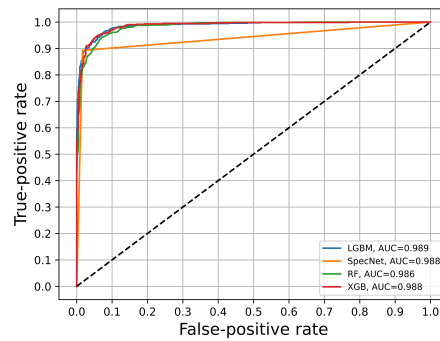
Figure 47: ROC curves with AUC values for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for (a) stomach dataset (SNV normalisation, Boruta feature selection) and (b) oesophagus dataset (SNV normalisation, Boruta feature selection).

Table 17: Performance metrics for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for stomach dataset (SNV normalisation, Boruta feature selection) and oesophagus dataset (SNV normalisation, Boruta feature selection).

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	90.54 (0.44)	85.72 (1.17)	93.33 (0.57)	96.82 (0.26)	94.36 (0.80)	92.08 (1.13)	95.98 (0.81)	98.56 (0.27)
XGB	92.84 (0.36)	89.71 (1.16)	94.67 (0.59)	97.99 (0.17)	95.13 (0.70)	93.41 (1.05)	96.35 (0.67)	98.85 (0.19)
LGBM	92.62 (0.40)	89.47 (1.10)	94.45 (0.59)	97.85 (0.20)	95.09 (0.64)	93.44 (0.98)	96.26 (0.64)	98.86 (0.20)
SpecNet	92.92 (1.75)	91.60 (1.89)	93.68 (1.66)	97.98 (0.96)	94.58 (1.74)	89.26 (2.38)	98.36 (0.98)	98.85 (0.82)



(a)



(b)

Figure 48: ROC curves with AUC values for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for (a) stomach dataset (SNV normalisation, Boruta feature selection) and (b) oesophagus dataset (SNV normalisation, Boruta feature selection).

In the pursuit of optimized performance within each approach, distinct combinations of data normalisation techniques and feature selection strategies emerged for the stomach and oesophagus datasets. For the stomach dataset, the combination of SNV normalization and no feature selection exhibited optimal performance, signifying its suitability for the Approach 1. Meanwhile, in the oesophagus dataset, the most favourable classification outcomes were attained through the fusion of no normalization and Boruta feature selection. In Approaches 2, 3, and 4, SNV normalization coupled with Boruta feature selection emerged as the superior choice for both the stomach and oesophagus datasets.

The superior classification performance achieved through the combination of SNV normalisation, and the Boruta feature selection method is rooted in their intrinsic characteristics. By centring each spectrum around its mean and scaling it by the standard deviation, SNV normalisation equalises the baseline and account for global intensity fluctuations. This process enhances the comparability of spectra across samples, minimising the impact of irrelevant variations that may confound classification models. Consequently, the application of SNV normalisation ensures that relevant spectral features, crucial for discrimination, are more pronounced, contributing to improved model accuracy. The Boruta feature selection method excels at capturing distinctive patterns in the spectral data by prioritising features that exhibit consistent discriminative power across iterations. This is crucial for identifying spectral regions that harbour essential information for classification. When these two methods combined, SNV normalisation mitigates variations in intensity, enabling Boruta to focus on relevant spectral variations linked to molecular interactions. This enables Boruta to pinpoint spectral features that are both distinct and consistent in differentiating tissue types.

A significant advancement in our study was the incorporation of real-time probe tracking during optical biopsy. As shown in Figure 49, each biopsy site was precisely tracked and then classified as normal or tumour. A graded colour map was used to visually differentiate between the two classes.

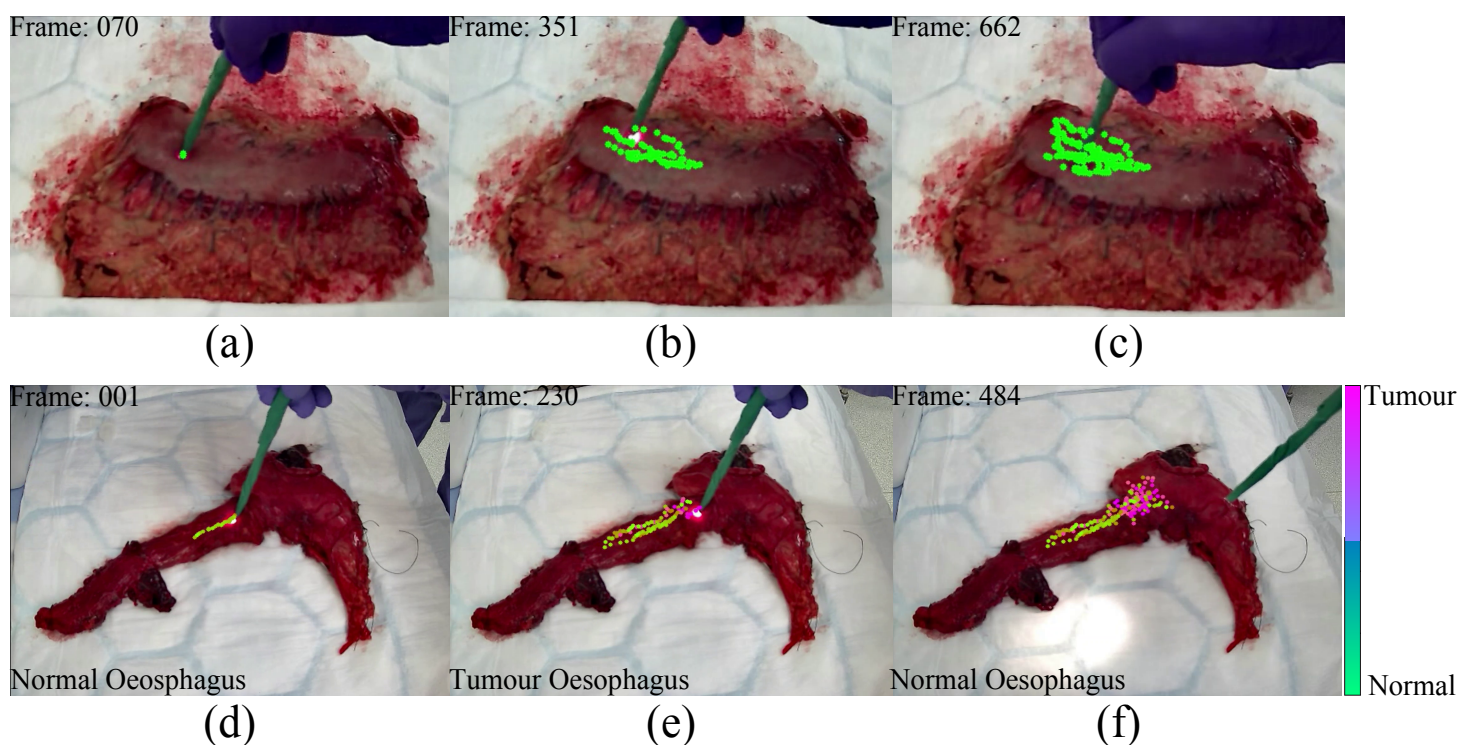


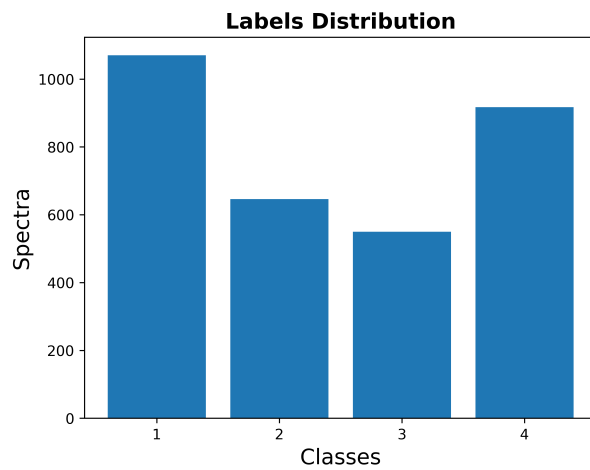
Figure 49: Illustration of real-time DRS probe tracking at each optical biopsy site coupled with the binary classification of each site on a gastric and oesophageal tissue specimen. (a) – (c) Real-time visual correlation of scanned optical biopsy sites on tissue (green dots on each frame) using the DRS probe and marker-based tracking. (d) – (f) Real-time tracking at each optical biopsy site coupled with binary classification of each site as either normal (100% green) or tumour tissue (100% pink) via a graduated colour map. Tissue type highlighted on screen in real-time (bottom left of each frame).

Multiclass Classification of Tumour Data for Pathological Staging

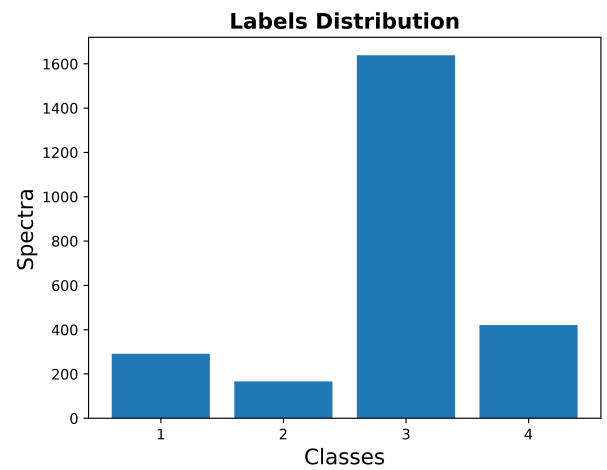
While binary classification has yielded valuable results, the intricacies of tumour heterogeneity necessitate a more nuanced approach. To this end, we delved into the realm of multiclass classification, aiming to discern subtler differentiations within tumour tissues. Specifically, we set our sights on distinguishing between pT1, pT2, pT3, and pT4 stages, representing progressive degrees of tumour invasion in the context of pathological staging.

The differentiation between these stages can provide clinicians with indispensable information for prognosis and treatment planning, thereby contributing to improved patient care and outcomes.

For gastric cancer, there was a total of 1,070 spectra for pT1, 646 spectra for pT2, 550 spectra for pT3, and 917 spectra for pT4. For oesophageal cancer, there was a total of 291 spectra for pT1, 166 spectra for pT2, 1,638 spectra for pT3, and 421 spectra for pT4. The means of all spectra for each of the tissue classes are shown in Figure 51.

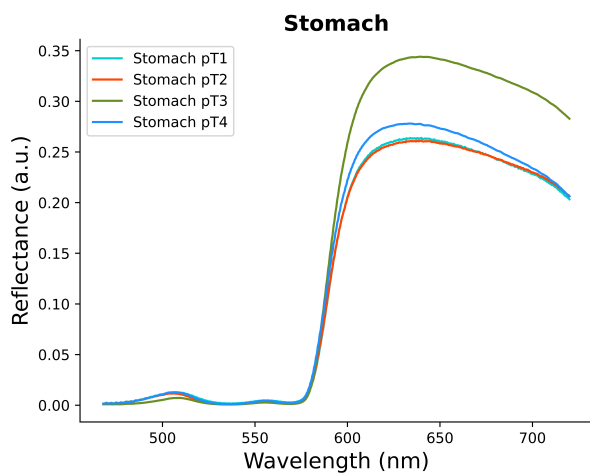


(a)

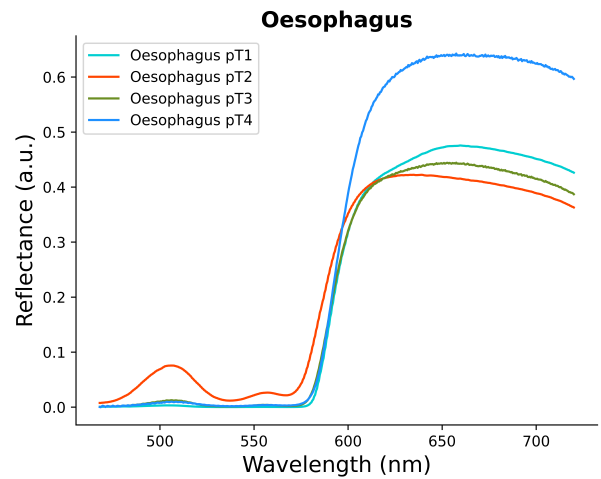


(b)

Figure 50: Labels distribution for (a) gastric and (b) oesophageal cancer pathological (TNM) staging.



(a)



(b)

Figure 51: Mean spectra of the different pathological tumour stages for (a) gastric cancer, and (b) oesophageal cancer.

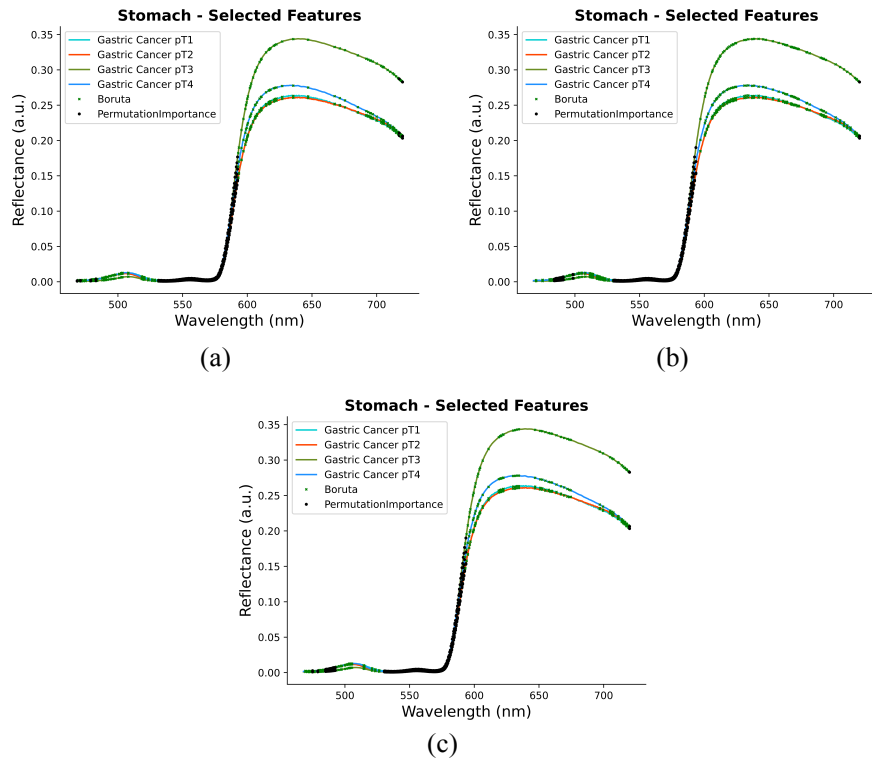


Figure 52: Selected features for the tumour pathological staging of the stomach dataset for each normalisation method. (a) Boruta and Permutation Feature Importance with no normalisation. (b) Boruta and Permutation Feature Importance after SNV normalisation. (c) Boruta and Permutation Feature Importance after AUC normalisation.

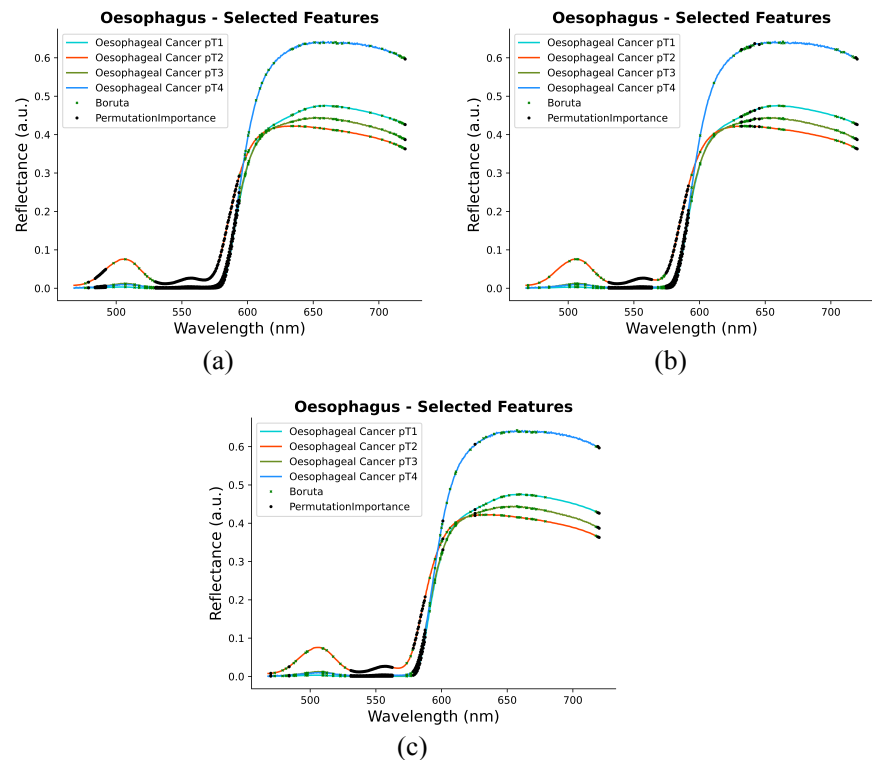


Figure 53: Selected features for the tumour pathological staging of the oesophagus dataset for each normalisation method. (a) Boruta and Permutation Feature Importance with no normalisation. (b) Boruta and Permutation Feature Importance after SNV normalisation. (c) Boruta and Permutation Feature Importance after AUC normalisation.

Results of the multiclass classification for stomach and oesophagus datasets are presented below. In the following tables, the average classification performance metrics are shown for the best combinations of normalisation, feature selection, and classification method. The algorithms are tailored to the multiclass classification through an extension of the one-vs-rest strategy. This approach transforms the multiclass problem into a set of binary classifications, allowing each algorithm to be trained on each binary classification problem and finally predictions are made using the model that is the most confident.

Table 18: Performance metrics for the multiclass spectral data classification using SNV normalisation and Boruta feature selection for both stomach and oesophagus datasets using Approach 1 (entire spectrum only).

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	89.75 (2.90)	77.99 (6.89)	92.84 (4.43)	86.34 (0.94)	95.70 (2.91)	86.66 (12.66)	94.92 (7.26)	94.27 (0.68)
XGB	90.47 (2.55)	79.66 (6.42)	93.38 (3.65)	87.29 (0.96)	96.57 (2.34)	90.15 (7.43)	96.13 (5.22)	95.43 (0.67)
LGBM	90.88 (2.41)	80.44 (6.43)	93.66 (3.54)	87.84 (1.00)	96.77 (2.20)	91.06 (6.93)	96.42 (4.68)	95.69 (0.58)
SpecNet	85.03 (4.53)	84.96 (4.46)	94.95 (2.74)	90.02 (3.80)	98.15 (1.93)	98.23 (2.15)	99.12 (1.35)	98.76 (1.58)

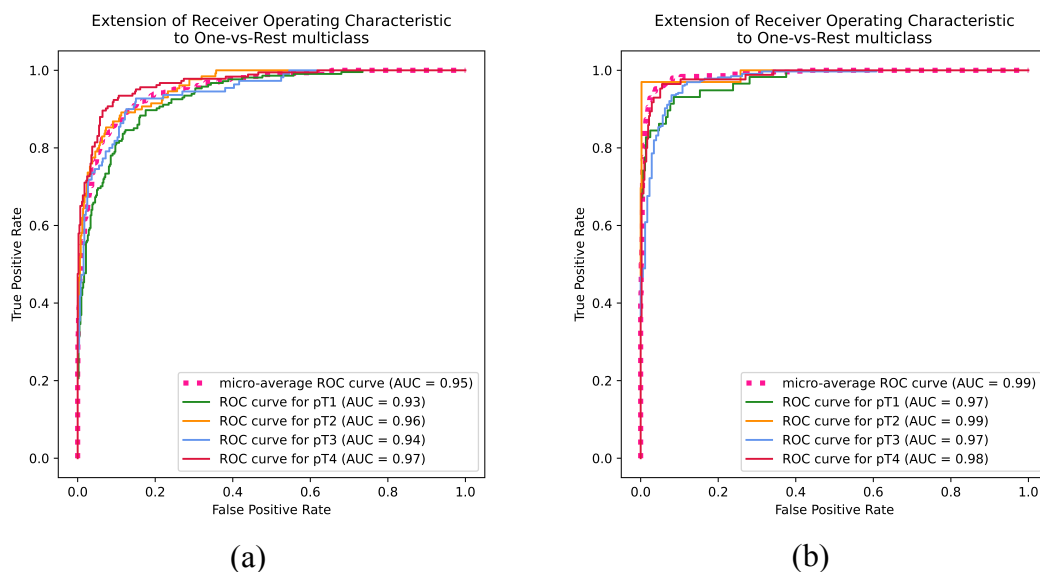


Figure 54: ROC curves for one-vs-rest multiclass classification for (a) stomach and (b) oesophagus pathological staging.

4.7.2 Distinguishing Healthy Colorectal Tissue from Cancer: Ex-Vivo Study Findings

Cohort Characteristics

A total of 23 patients were recruited for this study between April 2021 and May 2022. The majority of patients undergoing colorectal cancer surgery were male ($n=17$, 74%). The age of

patients ranged from 50 to 88 years old, with a median age of 70. Most patients (n=14, 61%) underwent an anterior resection procedure, whilst eight (35%) patients underwent a right hemicolectomy and only one (4%) patient underwent a sigmoid colectomy. There were three patients with pT0 disease on histology. These were found to be benign adenomatous polypoid lesions. Healthy tissue data was collected for these patients. Table 19 outlines the patient characteristics.

Table 19: Cohort characteristics of n=23 patients included in the analysis. Characteristics are presented for all patients in no. (%).

Characteristic	No. (%)
Age, median (range)	70 (50-88)
Sex	
Male	17 (74.0)
Female	6 (26.0)
Procedure	
Anterior Resection	14 (61.0)
Right Hemicolectomy	8 (35.0)
Sigmoid Colectomy	1 (4.0)
Tumour Histology	
Adenocarcinoma	20 (87.0)
Benign Adenomatous Polyp	3 (13.0)
Tumour Stage (TNM staging)	
pT0	3 (13.0)
pT1	5 (21.7)
pT2	2 (8.7)
pT3	11 (47.8)
pT4	2 (8.7)
Neoadjuvant therapy	
Chemotherapy	1 (4.3)
Chemoradiotherapy	0
Radiotherapy	0
None	22 (95.7)

Dataset Summary

Overall, 23 distinct sets of non-tumour and 19 sets of tumour data were included in the analysis. There was a total of 3,632 spectra for non-tumour or healthy tissue and 4,330 spectra for tumour tissue.

The means of all spectra with standard deviation for each of the tissue classes are shown in Figure 55.

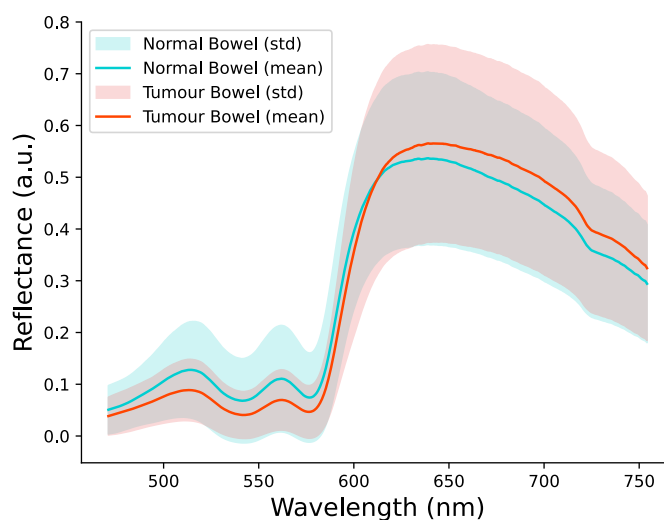


Figure 55: Mean spectra with standard deviation for normal and tumour colorectal tissue.

Feature Extraction and Selection

As described in Section 4.6, the first step of the spectral analysis pipeline was to normalise the data using the SNV and AUC methods. Afterwards, the shape-based features and the haemoglobin saturation coefficient were extracted (Figure 56).

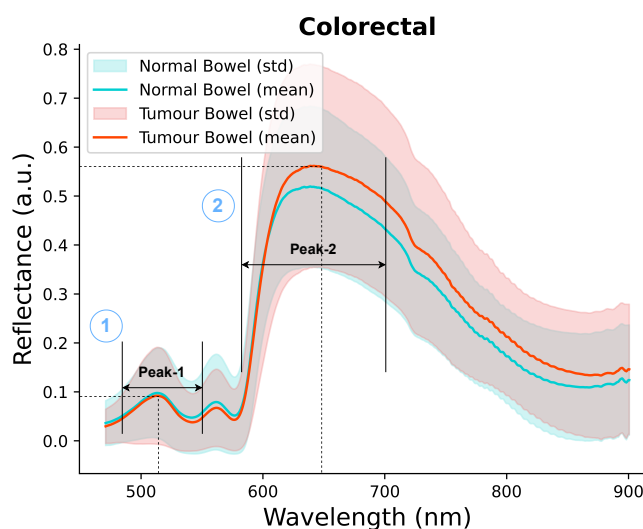


Figure 56: Illustration of shape-based features extracted for the colorectal dataset. (1) Depicts the secondary peak intensity value and corresponding wavelength in the 480 – 550 nm spectral range. (2) Depicts the primary peak intensity value and corresponding wavelength in the 580 – 700 nm spectral range.

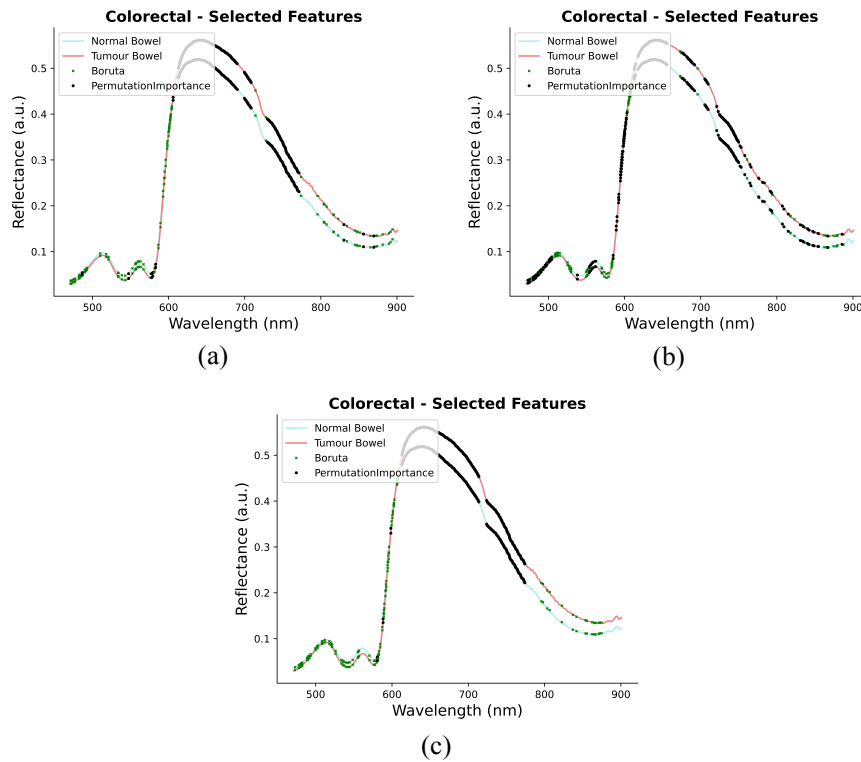


Figure 57: Selected features for the colorectal dataset for each normalisation method. (a) Boruta and Permutation Feature Importance with no normalisation. (b) Boruta and Permutation Feature Importance after SNV normalisation. (c) Boruta and Permutation Feature Importance with AUC normalisation.

Each of the feature selection methods, namely Boruta and Permutation Feature Importance, was applied on colorectal dataset following SNV and no normalisation. The selected features are shown in Figure 57 for the colorectal dataset, respectively.

Exploring the Impact of Normalization on Classification Performance

The added value of normalisation methods on the classification outcome was also examined. To this end, the increase in overall accuracy, sensitivity, specificity, and AUC for each normalisation method compared to no normalisation were analysed. This analysis was performed for the entire spectrum for the Random Forest, XGBoost, LightGBM and the SpecNet classifiers. The average increase in the aforementioned performance metrics per normalisation technique for each of the four approaches are shown in the following tables. No feature selection was considered for this investigation.

Table 20: Mean increase in overall accuracy, sensitivity, specificity, and AUC for SNV normalisation compared to no normalisation for all classifiers for the colorectal dataset.

Classifier	Mean (SD) (%)			
	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra				
RF	0.05 (0.01)	0.06 (0.03)	0.03 (0.02)	0.00 (0.00)
XGB	0.01 (0.00)	-0.05 (0.01)	0.09 (0.11)	0.00 (0.00)
LGBM	-0.15 (0.20)	-0.26 (0.12)	-0.02 (0.22)	-0.04 (0.06)
SpecNet	11.05 (0.85)	-1.85 (0.19)	26.42 (-1.57)	6.53 (0.95)
Approach 2 – Reflectance spectra with shape-based features				
RF	-0.07 (0.02)	-0.16 (0.01)	0.05 (0.02)	0.00 (0.00)
XGB	-0.02 (0.01)	-0.06 (0.03)	0.03 (0.00)	-0.02 (0.01)
LGBM	-0.08 (0.06)	-0.21 (0.12)	0.09 (0.01)	-0.07 (0.05)
SpecNet	5.15 (0.47)	-0.92 (0.10)	12.39 (0.92)	2.73 (0.45)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient				
RF	0.00 (0.02)	0.01 (0.01)	0.02 (0.09)	0.00 (0.00)
XGB	0.03 (0.01)	0.00 (0.00)	-0.05 (0.02)	0.00 (0.00)
LGBM	0.17 (0.05)	0.29 (0.18)	0.03 (0.09)	-0.02 (0.05)
SpecNet	5.57 (0.51)	9.16 (0.77)	1.28 (0.14)	4.09 (0.68)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient				
RF	0.06 (0.05)	0.05 (0.08)	0.08 (0.05)	0.00 (0.00)
XGB	-0.02 (0.01)	-0.02 (0.04)	-0.02 (0.01)	0.00 (0.00)
LGBM	0.08 (0.14)	0.11 (0.17)	0.04 (0.18)	0.06 (0.04)
SpecNet	5.19 (0.48)	9.39 (0.78)	0.18 (0.02)	3.16 (0.52)

In Table 20 it shows that SNV normalisation technique improved the classification outcome significantly for the colorectal dataset using the SpecNet classifier. For the rest of the classifiers no significant improvement was noticed.

Analysing the Influence of Feature Extraction on Classification Performance

Based on the results from the normalisation, the first analysis on feature extraction and selection was performed on non-normalised data and data normalised using the SNV normalisation technique. In Table 21 the results are shown for the mean increase in classification performance metrics for all feature extraction methods (approaches 2, 3, 4) compared to no feature extraction, i.e., the entire spectrum (approach 1), for the colorectal dataset.

Table 21: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the colorectal dataset. Only non-normalised and SNV normalised data was used.

Classifier	Mean (SD) (%)							
	No normalisation				SNV normalisation			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 2 – Reflectance spectra with shape-based features								
RF	-0.16 (0.06)	0.09 (0.04)	-0.45 (0.18)	-0.05 (0.03)	-0.27 (0.09)	-0.14 (0.05)	-0.42 (0.22)	-0.05 (0.03)
XGB	0.11 (0.29)	0.23 (0.39)	-0.03 (0.12)	0.00 (0.16)	0.08 (0.30)	0.23 (0.42)	-0.09 (0.01)	-0.02 (0.17)
LGBM	-0.07 (0.08)	-0.03 (0.08)	-0.12 (0.06)	0.04 (0.02)	0.01 (0.18)	0.02 (0.16)	-0.01 (0.27)	0.00 (0.02)
SpecNet	6.20 (0.41)	0.92 (0.10)	12.48 (0.47)	3.44 (0.42)	0.29 (0.03)	1.85 (0.19)	-1.56 (0.17)	-0.35 (0.07)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	-0.13 (0.09)	0.05 (0.04)	-0.34 (0.39)	-0.03 (0.00)	-0.18 (0.12)	0.00 (0.06)	-0.39 (0.49)	0.03 (0.01)
XGB	0.50 (0.40)	0.35 (0.32)	0.67 (0.24)	0.25 (0.17)	0.46 (0.41)	0.40 (0.33)	0.53 (0.15)	0.25 (0.17)
LGBM	0.12 (0.07)	-0.19 (0.06)	0.48 (0.13)	0.23 (0.02)	0.44 (0.09)	0.36 (0.00)	0.54 (0.18)	0.25 (0.02)
SpecNet	6.24 (0.41)	-8.93 (0.74)	24.31 (1.34)	2.64 (0.31)	0.75 (0.08)	2.08 (0.21)	-0.83 (0.09)	0.21 (0.04)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	-0.14 (0.19)	-0.02 (0.21)	-0.29 (0.19)	0.00 (0.00)	-0.19 (0.20)	-0.08 (0.18)	-0.31 (0.20)	0.01 (0.00)
XGB	0.59 (0.21)	0.25 (0.20)	1.00 (0.04)	0.25 (0.11)	0.56 (0.22)	0.29 (0.25)	0.89 (0.07)	0.25 (0.11)
LGBM	0.06 (0.05)	-0.12 (0.08)	0.28 (0.20)	-0.09 (0.06)	0.49 (0.14)	0.32 (0.06)	0.69 (0.03)	0.31 (0.00)
SpecNet	6.40 (0.43)	-9.24 (0.76)	25.04 (1.41)	3.13 (0.38)	0.54 (0.06)	2.00 (0.21)	-1.19 (0.13)	-0.24 (0.05)

In Table 21 the results are shown for the mean increase in performance metrics for all feature extraction methods compared to no feature extraction (entire spectrum) for the colorectal dataset. It can be concluded that overall using shape-based features combined with the haemoglobin saturation coefficient feature improved classification results compared to the use of the entire spectrum alone. Moreover, when SNV normalisation is used, a higher increase in performance metrics is present using both shape-based and haemoglobin saturation coefficient features compared to no normalisation. Therefore, for the following analysis on the effect of feature selection methods on the classification performance, only the SNV normalisation technique is considered.

In Table 22 the results are shown for the mean increase in classification performance metrics for Boruta and Permutation Importance feature selection methods compared to no feature selection, i.e., the entire spectrum, using SNV normalisation for the colorectal dataset.

From Table 22, the results are shown that feature selection significantly improved the classification performance for all classifiers. Considering the high dimensionality of the colorectal dataset (1922 wavelengths), this increased performance induced by the feature selection methods indicates that a substantial portion of the features may indeed be redundant or noisy. By systematically identifying and excluding these redundant features,

feature selection techniques enable the models to concentrate on the essential discriminatory information within the dataset.

Table 22: Mean increase in classification performance metrics for all feature selection methods compared to the use of the entire spectrum for the colorectal dataset. Only SNV normalised data was used.

Classifier	Mean (SD) (%)							
	Boruta feature selection				Permutation Importance feature selection			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra								
RF	1.50 (0.03)	1.44 (0.22)	1.58 (0.35)	0.98 (0.08)	-1.07 (0.28)	-0.62 (0.36)	-1.61 (0.18)	-0.95 (0.17)
XGB	2.30 (0.22)	1.99 (0.38)	2.67 (0.09)	1.32 (0.20)	1.24 (0.07)	1.16 (0.20)	1.33 (0.07)	0.60 (0.09)
LGBM	1.73 (0.09)	1.76 (0.08)	1.70 (0.05)	1.06 (0.06)	0.66 (0.20)	0.74 (0.03)	0.56 (0.27)	0.27 (0.04)
SpecNet	1.18 (0.21)	0.92 (0.09)	2.94 (0.39)	1.00 (0.23)	1.55 (0.17)	2.31 (0.24)	0.64 (0.08)	0.73 (0.16)
Approach 2 – Reflectance spectra with shape-based features								
RF	1.76 (0.01)	1.42 (0.14)	2.18 (0.06)	1.02 (0.13)	-0.35 (0.20)	0.05 (0.23)	-0.81 (0.10)	-0.51 (0.11)
XGB	2.05 (0.14)	1.80 (0.07)	2.35 (0.05)	1.29 (0.14)	1.41 (0.04)	0.83 (0.05)	2.10 (0.35)	0.81 (0.02)
LGBM	1.58 (0.08)	1.21 (0.04)	1.86 (0.10)	0.91 (0.06)	1.03 (0.15)	1.03 (0.03)	1.03 (0.03)	0.44 (0.02)
SpecNet	2.43 (0.28)	2.93 (0.35)	1.83 (0.21)	1.43 (0.32)	1.38 (0.15)	-0.92 (0.10)	4.13 (0.51)	1.49 (0.33)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient								
RF	1.85 (0.04)	1.41 (0.10)	2.38 (0.51)	1.07 (0.12)	-0.58 (0.07)	-0.39 (0.18)	-0.81 (0.17)	-0.70 (0.13)
XGB	1.98 (0.16)	1.49 (0.10)	2.55 (0.13)	1.24 (0.14)	1.18 (0.04)	0.79 (0.04)	1.63 (0.19)	0.60 (0.04)
LGBM	1.58 (0.10)	1.15 (0.06)	2.10 (0.08)	1.00 (0.15)	0.65 (0.20)	0.51 (0.03)	0.81 (0.17)	0.27 (0.11)
SpecNet	1.05 (0.12)	-1.15 (0.12)	3.67 (0.47)	0.80 (0.18)	4.10 (0.53)	3.70 (0.47)	4.59 (0.62)	1.89 (0.50)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient								
RF	1.62 (0.05)	1.51 (0.24)	1.76 (0.03)	0.98 (0.10)	-0.27 (0.14)	0.18 (0.22)	-0.80 (0.01)	-0.48 (0.11)
XGB	1.94 (0.15)	1.63 (0.03)	2.31 (0.41)	1.35 (0.18)	1.30 (0.05)	1.04 (0.04)	1.61 (0.36)	0.77 (0.03)
LGBM	1.35 (0.02)	1.17 (0.09)	1.57 (0.13)	0.92 (0.11)	1.09 (0.21)	1.10 (0.10)	1.08 (0.35)	0.54 (0.69)
SpecNet	3.06 (0.37)	2.93 (0.36)	3.21 (0.39)	1.61 (0.37)	1.17 (0.13)	-0.31 (0.03)	2.94 (0.35)	1.19 (0.26)

Healthy versus Tumour Tissue Classification Results

So far, normalisation and feature extraction and selection methods were examined separately. Results of the classification for the colorectal dataset are presented below. In the following tables, the classification performance metrics are shown for the best combinations of normalisation, feature extraction and selection, and classification method for each approach described in Section 4.6. Data showing the ability of multiple classifiers to detect normal versus cancerous tissue presented as mean (standard deviation). Overall accuracy is calculated as the proportion of correctly identified spectra over the total number of spectra averaged over the stratified group 5-fold cross-validation.

Table 23: Performance metrics for all classifiers using Approach 1 (entire spectrum only) for colorectal dataset (SNV normalisation, Boruta feature selection).

Classifier	Mean (SD) %			
	Accuracy	Sensitivity	Specificity	AUC
RF	82.23 (0.95)	88.80 (1.29)	87.54 (1.56)	95.41 (0.58)
XGB	86.95 (1.08)	87.49 (1.22)	86.31 (1.58)	94.54 (0.59)
LGBM	86.75 (1.15)	87.42 (1.35)	85.94 (1.79)	94.57 (0.63)
SpecNet	91.50 (2.24)	89.22 (2.49)	94.22 (1.87)	97.62 (1.22)

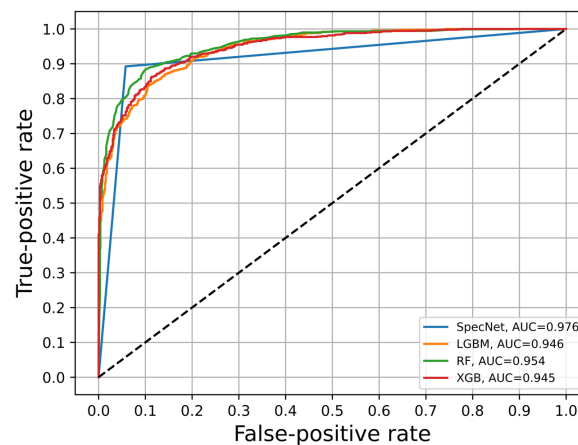


Figure 58: ROC curves with AUC values for all classifiers using Approach 1 (entire spectrum) for colorectal dataset (SNV normalisation, Boruta feature selection).

Table 24: Performance metrics for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for colorectal dataset (SNV normalisation, Boruta feature selection).

Classifier	Mean (SD) %			
	Accuracy	Sensitivity	Specificity	AUC
RF	88.28 (0.98)	88.80 (1.32)	87.67 (1.47)	95.40 (0.56)
XGB	86.79 (0.86)	87.58 (1.13)	85.86 (1.60)	94.50 (0.49)
LGBM	86.60 (1.03)	87.11 (1.42)	86.00 (1.57)	94.50 (0.59)
SpecNet	92.38 (2.13)	93.07 (2.04)	91.56 (2.23)	97.68 (1.21)

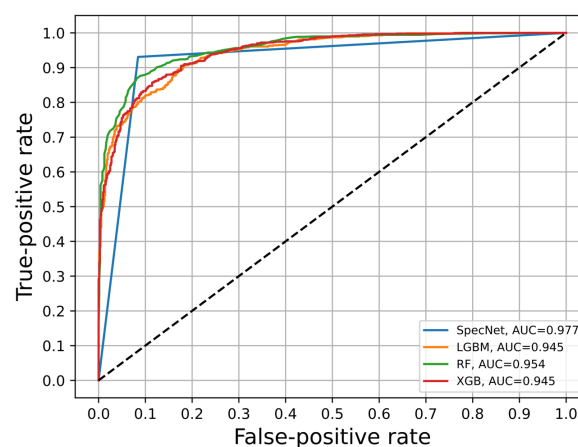


Figure 59: ROC curves with AUC values for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for colorectal dataset (SNV normalisation, Boruta feature selection).

Table 25: Performance metrics for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for colorectal dataset (SNV normalisation, Boruta feature selection).

Classifier	Mean (SD) %			
	Accuracy	Sensitivity	Specificity	AUC
RF	88.40 (1.00)	88.77 (1.36)	87.96 (1.19)	95.47 (0.56)
XGB	87.09 (0.73)	87.39 (1.17)	86.72 (1.38)	94.70 (0.48)
LGBM	87.04 (1.06)	87.16 (1.47)	86.89 (1.48)	94.76 (0.55)
SpecNet	91.46 (2.24)	89.22 (2.49)	94.13 (1.89)	97.62 (1.22)

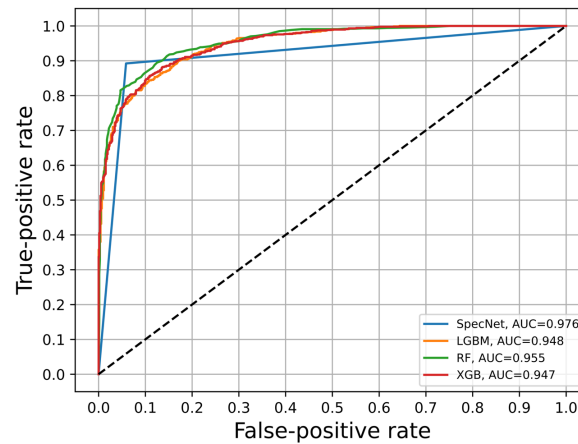


Figure 60: ROC curves with AUC values for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for colorectal dataset (SNV normalisation, Boruta feature selection).

Table 26: Performance metrics for the for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for colorectal dataset (SNV normalisation, Boruta feature selection).

Classifier	Mean (SD) %			
	Accuracy	Sensitivity	Specificity	AUC
RF	88.22 (1.02)	88.83 (1.38)	87.49 (1.39)	95.42 (0.58)
XGB	87.15 (0.93)	87.42 (1.38)	86.83 (1.33)	94.81 (0.50)
LGBM	86.86 (1.08)	87.15 (1.45)	86.51 (1.58)	94.73 (0.58)
SpecNet	93.26 (2.01)	93.23 (2.02)	93.30 (2.01)	97.97 (1.13)

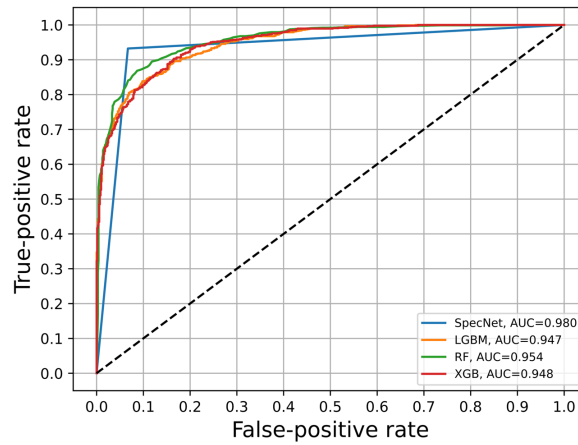


Figure 61: ROC curves with AUC values for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for colorectal dataset (SNV normalisation, Boruta feature selection).

In the pursuit of optimized performance within each approach, distinct combinations of data normalisation techniques and feature selection strategies emerged for the colorectal dataset. The combination of SNV normalization and Boruta feature selection exhibited optimal performance for reasons explained in the previous section for the upper GI datasets.

In our study, a critical aspect of the developed system was its ability to track each optical biopsy site in real-time. This is vividly illustrated in Figure 62, which demonstrates the system's ability to simultaneously perform a binary classification of each biopsy site. Each site is clearly classified as either healthy tissue, shown in solid green, or tumour tissue, shown in bright pink, using a graded colour map.

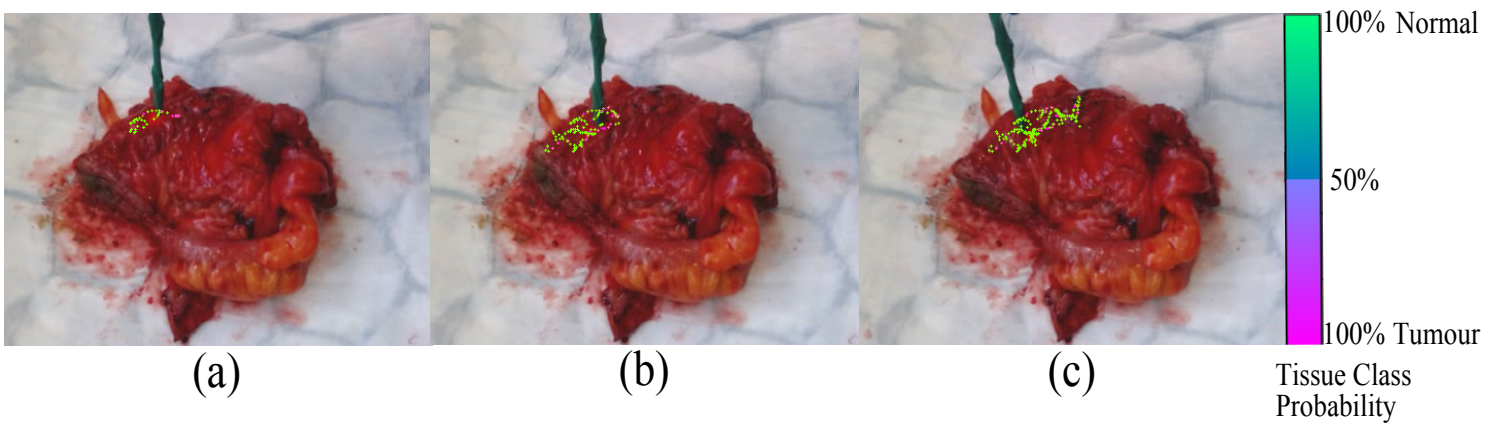


Figure 62: (a) – (c) Real-time tracking at each optical biopsy site coupled with binary classification of each site as either normal (100% green) or tumour tissue (100% pink) via a graduated colour map.

4.8 Summary

In this study, we investigated the usability of optical spectroscopy for margin assessment in an ex-vivo manner for stomach, oesophageal, and colorectal tissue.

The DRS system presented here shows suitability and potential for use intraoperatively to aid margin assessment in cancer surgery. We developed a system with high spectral resolution and high data acquisition speed. As a result, we were able to scan a large area on the tissue and differentiate tissue in real-time while creating minimal disruption to the surgical workflow. Our method for real-time probe detection and tracking has demonstrated high accuracy compared with ground truth and overcomes the point-based limitation that other DRS systems have faced^{262,265}. This enables the surgeon to follow the exact areas being sampled based on augmented data acquisition video feed. The tracking system coupled with the spectral classification could potentially be used to assess the resection margins of cancer intraoperatively. In this way, less healthy tissue could be removed, leading to a lower risk of morbidity for the patient. As such, the system has the potential to influence the intraoperative procedure and decision-making process.

The co-registration process between optical measurements and histopathology holds immense significance in the validation of innovative optical diagnostic technologies. Given the inherent tissue deformation occurring during the histopathological processing, ensuring precise alignment between optical measurements and the established gold standard necessitates the accurate correlation of H&E stained images with the optical measurements. In Chapter 4, we presented a comprehensive methodology designed to facilitate the registration process between optical measurements and the gold standard, ultimately enhancing the robustness and accuracy of the validation process for our cutting-edge optical diagnostic technology.

In the present ex-vivo studies, fibrosis (a form of connective tissue without tumour cells) was not included in the measured tissue types because of the relatively limited occurrence of fibrosis within the examined specimens, which represents a limitation of this study. Patients with advanced stage rectal and oesophagogastric cancer often receive neo-adjuvant chemo-radiotherapy that commonly leads to the development of fibrosis on the radiation-affected site. Fibrosis is challenging to distinguish from tumour tissue through the

surgeon's visual assessment alone. Therefore, it would be beneficial to include this tissue type when real-time classification is done.

A major limitation of the ex-vivo studies is the potential degradation following resection and blood supply being cut off, however, given that the entirety of the data acquisition protocol lasted approximately 10 minutes and was performed within 15 minutes following tissue resection, we believe that the impact of this was minimised, and tissue characteristics were preserved²⁸⁵.

Another limitation of these studies was the fact that they were conducted at a single institution, which inherently limits the range and diversity of the patient population studied. The implications of this are twofold: Firstly, the homogeneity of practices, resources and patient demographics within a single centre may inadvertently introduce biases or oversights. Secondly, these conditions may not accurately represent a wider, global patient and clinical landscape, potentially compromising the external validity and generalisability of the findings. A multi-centre trial would effectively address these limitations, allowing for the inclusion of subjects from different backgrounds and healthcare systems, and potentially introducing different clinical practices, thus providing a more holistic and representative dataset.

Moreover, our study methodology involved three different operators performing the sampling. While this could be seen as introducing different techniques and expertise, it is also a challenge. Each operator, with their unique experience and skills, has an individualised approach to sampling. This variability can lead to inconsistencies in the way the tissue is sampled and the precise location from which the sample is taken. This, in turn, can lead to differences in the quality of the DRS data obtained. It is worth noting that in our study, all operators underwent standardised training and a consistent protocol for data collection was carefully enforced.

While our ex-vivo studies provide important insights and advances in the field of DRS, there is an urgent need to translate these findings into real-world intra-operative scenarios. However, the in-vivo environment presents unique challenges, including physiological changes in living tissue, motion artifacts and variations due to blood perfusion. Building on our ex-vivo findings, the next chapter focuses on the in-vivo study, where we further test and refine the capabilities of our system. The ultimate goal is to ensure that the benefits observed ex-vivo, such as rapid data acquisition and real-time tissue discrimination, are effectively and reliably translated to an in-vivo surgical context.

Chapter 5: Bridging the Gap: In-Vivo DRS for Tissue Discrimination in Upper Gastrointestinal Cancer Surgery

5.1 Introduction

Chapter 4 demonstrated the proof of principle that tumour can be differentiated from non-tumour gastrointestinal tissue. The shift from ex-vivo to in-vivo diffuse reflectance spectroscopy sampling represents a pivotal transformation in the field of surgical diagnostics. This transition enables the concept of “optical biopsy”, which holds the potential to transform tissue assessment and diagnosis.

Traditionally, biopsy procedures involve the invasive extraction of tissue samples for subsequent laboratory analysis. While effective, these procedures can be discomforting, time-consuming, and carry inherent risks. The emergence of in-vivo DRS, however, introduces a paradigm shift. It enables surgeons to gain real-time, non-invasive insights into tissue properties directly at the surgical site. During surgery, in-vivo DRS provides a continuous stream of real-time data, allowing surgeons to assess tissue margins as they progress through the procedure. This precision enables the identification and removal of all malignant tissue, an imperative measure in minimising any risks associated with the potential residual cancerous cells.

However, to be smoothly integrated into the critical operating theatre environment, in-vivo DRS probes must meet strict specifications comparable to other surgical tools. In particular, they must demonstrate durability, maintain sterility, and seamlessly function within the operating theatre environment.

Much of the previously published data demonstrates the potential of DRS to discriminate healthy tissue from tumour tissue with high sensitivity (74-79%) and specificity (78-93%)^{312–319}. However, many of these studies have been conducted in a well-controlled ex-vivo setting which does not necessarily resemble intra-operative circumstances when trying to discriminate healthy tissue from tumour tissue in an individual patient. One of the

challenges that need to be addressed is the physiological differences between in-vivo and ex-vivo measurement. Due to the loss of blood flow and exposure to air, tissue physiology will be different in resected tissue. Such phenomena could impact optical parameters, especially those that are related to blood, thereby hindering one-on-one translation of ex-vivo results to an in-vivo setting^{320,321}.

In addition, the transition from the controlled ex-vivo setting to the inherently more complex in-vivo setting brings about various challenges. These challenges include variations in the pressure applied to the probe, due to physiological factors such as patient positioning and respiration. Moreover, the correlation of DRS data with the underlying pathology becomes more intricate as tissue features may change dynamically during surgery. Furthermore, controlling ambient lighting is a crucial consideration in the operating theatre, where various light sources are used, which could potentially affect the accuracy of DRS measurements.

5.2 *In-vivo Spectroscopy System*

The successful integration of in-vivo DRS probes within surgical settings depends on the fulfilment of similar rigorous standards as other surgical instruments. Here, we delve into the required specifications of the DRS probe to efficiently perform intraoperatively.

- **Durability:** The probe's construction materials should be durable enough to withstand repeated decontamination procedures, such as autoclaving or chemical disinfection, to ensure they remain in optimum working order over prolonged periods. Durability guarantees the probe's dependability during surgery reducing the likelihood of instrument failure or malfunction which could result in serious implications for patient care.
- **Sterility:** The design of the probe must be such that it allows for effective sterilisation methods including autoclaving or sterilisation with ethylene oxide (EtO) gas, to guarantee that it is free from contaminants when entering the surgical site. Suitable sterile packaging is also crucial in maintaining sterility until the probe is ready for use.
- **Ergonomics:** Equally important is the design of the probe in terms of ergonomics and interaction with the surgeon. Our DRS probe, due to its pen-shaped design, provides

a comfortable grip to ensure that surgeons can use it with precision during procedures.

In our in-vivo investigation, we built upon the DRS system detailed in Section 4.2, while making modifications to cater to the particular requirements of in-vivo sampling. The primary deviation was in the design and manufacturing of the reflection probe, an essential part of our aim to gather real-time insights within the dynamic surgical environment.

To fulfil the requirements described previously, the in-vivo DRS system consisted of a custom-made sterilisable reflection/backscattering probe (Bundle UVNSS 600, CeramOptec GmbH) with the same optical specifications as the ex-vivo one (six peripheral illumination fibres around one light collection fibre, 600 μm core diameter, 0.22 numerical aperture, 3.175 mm ferrule) (Figure 63). A tungsten halogen light source (360 – 2400 nm, typical output power 8.8 W, Ocean Optics Inc., HL-2000-HP) was connected to the six illumination fibres, while the central fibre was connected to a spectrometer (Ocean Optics Inc., USB4000) that resolves light in the spectral range 400 – 1000 nm. Both the light source and the spectrometer were controlled by a graphical user interface (GUI) developed with Python.

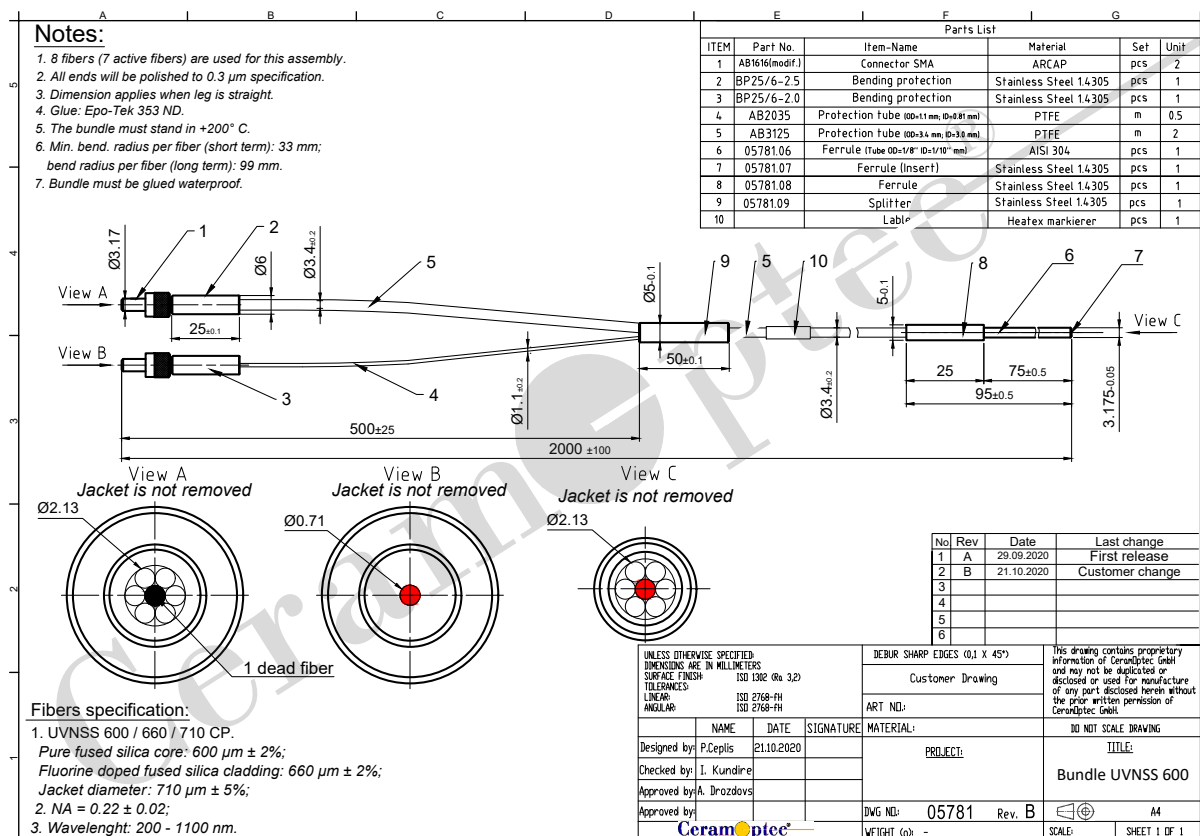


Figure 63: Schematic illustration of the sterilisable in-vivo DRS probe.

5.3 Histopathology Correlation

Histopathology correlation is a crucial element in bridging the gap between in-vivo DRS data acquisition and the definitive diagnosis of tissue pathology. This process guarantees precise alignment of the spectral information obtained during surgery with the histological characteristics of the tissue. Here, the in-vivo histopathology correlation protocol designed by Mr. Maxime Giot in collaboration with Ms. Scarlet Nazarian and the author of this thesis is outlined.

The process begins by acquiring spectra intraoperatively from the suspected tumorous area, identified by the surgeon immediately before surgical resection. To enable further correlation, the surgeon uses a sterile marker to draw a small circle around the point of spectral acquisition, which demarcates the region of interest. Subsequently, the specimen is resected, and an initial RGB photo is taken with a ruler placed next to the specimen (Figure 64). This ruler acts as a reference for pixel information in 2D images and assists in correlating measurements with the specimen.

The in-vivo acquisition is followed by ex-vivo spectral acquisition from the same region using probe tracking. The area that encompasses the previously marked circle is dyed with yellow pathology ink. Importantly, measurements are taken to quantify the distance from the proximal margin to the beginning of the yellow-inked area, from the end of the yellow-dyed area to the distal margin, and the length of the yellow-inked region in the proximal to distal plane. Additionally, a 3 cm line is marked on the specimen using a sterile marker, and then it is preserved in formalin with precautions taken to prevent deformation.



Figure 64: A gastrectomy specimen prior to fixation in formalin and processing by the pathologist. (The greater omentum is present and attached to the greater curvature on the inferior border of the stomach). The yellow-dyed area corresponds to the suspected tumour area that was sampled ex-vivo. Three measurements are annotated 1 to 3 on the figure. Line (1) is the distance from the proximal resection margin to the proximal border of the yellow-dyed area. Line (2) is the length of the yellow-dyed area in the proximal to distal plane. Line (3) is the 3 cm line drawn.

Upon receipt in the histopathology laboratory, the pathologist conducted precise measurements to determine specimen dimensions. The pathologist measured the distance from the proximal margin to the yellow-inked area, or from the distal margin to the yellow-inked area. Additional measurements entail determining the length of the yellow-inked area and the length of the 3 cm line drawn pre-formalin.

The sample is sectioned to locate the area of concern for further analysis. Sequentially labelled slices (e.g., A1, A2, A3, etc.), roughly 4 mm wide, were taken from this region. The distance from the most proximal point of the sample to the first slice intended for analysis was measured. Each labelled slice was then placed in a cassette. If a slice was too large for a single cassette, it might be divided into two and noted as a composite block. The final pathology report specified slices with yellow ink as well as the specific cassette or block containing tumour tissue, if present.

After receiving the report, the post-processing steps involved calculating the specimen's shrinkage coefficient due to formalin fixation and correlating the cassettes with the corresponding slices. Measurements, both pre- and post-formalin fixation, and the shrinkage coefficient were used to identify the region on the macroscopic specimen. Video tracking data of the DRS probe were essential in accurately locating the acquired spectra on the macroscopic specimen, enabling precise correlation.

5.4 Study Design

Patients undergoing oesophageal or gastric cancer resection were prospectively recruited into the study between May 2022 and August 2023 at Hammersmith Hospital in London, United Kingdom. Prior to patient enrolment, the study received ethical approval from the HRA and Health and Care Research Wales (HCRW) under reference number 22/SC/0048. Furthermore, the principles of informed consent were rigorously upheld, with all participants providing comprehensive written consent prior to their involvement in the study.

In-vivo tissue specimens were included for patients undergoing elective surgery for either oesophageal carcinoma (adenocarcinoma or squamous cell carcinoma) or gastric adenocarcinoma. Healthy tissue data was included for analysis for cases confirmed as T0 on formal histology. It is worth noting that patients below the age of 18 and those subjected to non-elective surgical procedures were deliberately excluded from the study.

5.5 *Data Acquisition Protocol*

During surgery, the surgeon was asked to acquire DRS measurements of healthy and tumour gastrointestinal tissue. To obtain the spectra, the DRS probe was placed by the surgeon in contact with the tissue. An overview image of the tissue under study and a video recording during spectral scanning were captured by an RGB camera (Teledyne FLIR LLC, Blackfly S), which allowed for tracking of the tip of the DRS probe and mapping of optical biopsy locations in real time (Figure 65).

Spectral data were collected from oesophageal and stomach tissue intraoperatively. The sterilisable DRS probe was used on suspected healthy tissue within the sample and on macroscopically suspected cancerous tissue to obtain spectral information. DRS measurements of suspected macroscopic tumour site and healthy tissue were taken on the outer (serosal) layer. Healthy tissue measurements were taken as far from the suspected cancerous location as possible, based on visual and haptic inspection of the macroscopic tissue by a surgeon, to ensure these locations were healthy. If the surgeon was unable to localise tumour close to or at the tissue surface, no tumour measurements were performed.

The DRS probe recorded approximately 80 spectra per second. At each measurement location, twenty DRS measurements (1 spectrum per measurement) were acquired and averaged. A minimum of 200 different locations per tissue type (normal or tumour) were sampled, depending on the size of the region. When possible, DRS recordings of stomach and oesophagus were taken from the same specimen. Data were acquired in the spectral range of 420 – 720 nm with a λ separation of 0.5 nm for the oesophageal and stomach specimens.

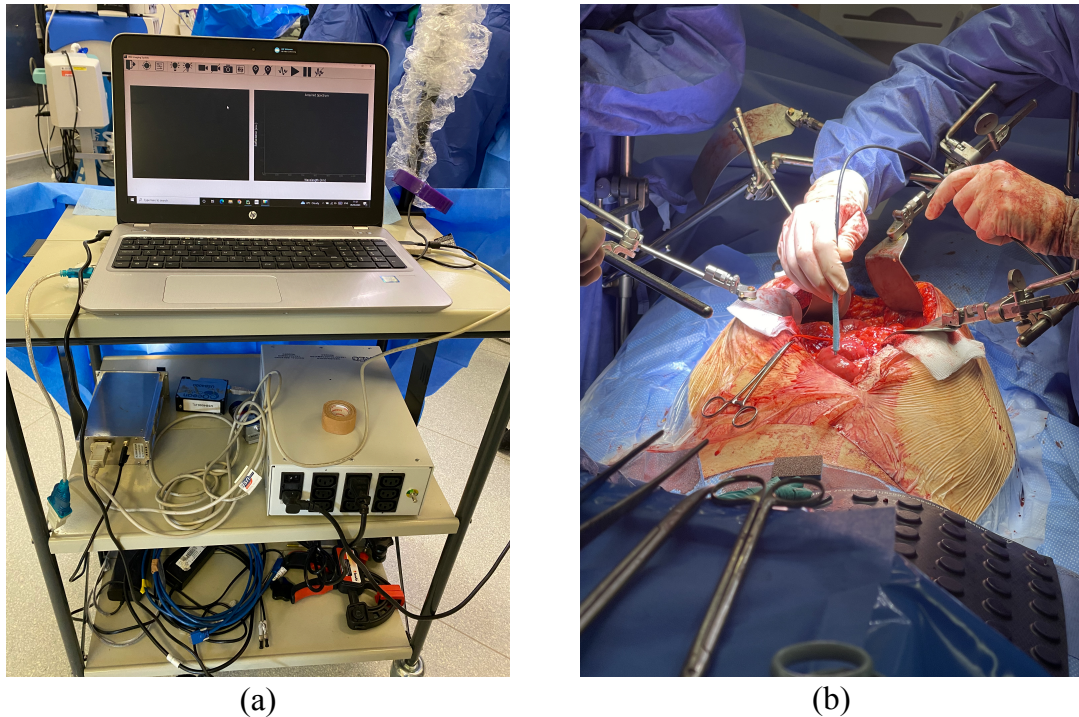


Figure 65: Photograph of (a) the DRS system setup and (b) the DRS probe during in-vivo data acquisition.

The DRS measurements were obtained in continuous mode starting in the healthy tissue, continuing through the transition zone of normal-tumour, and ending in the tumour. At each point measurement location, the acquired spectra were averaged to calculate a mean spectrum of each measurement location. To correct for the dark current, the spectrometer was calibrated prior to data acquisition²⁸⁶. A reference spectrum was acquired by direct contact measurement on a white reflectance phantom (optical-grade Spectralon® reference, Labsphere Inc.). Subsequently, to account for inter-patient comparability, background light, and signal quality variability, spectral normalisation, and noise reduction was performed for each acquired spectrum. The process has been explained in Section 4.5. DRS data were then processed using a Savitzky-Golay (or digital smoothing polynomial) filter to perform noise reduction and preserve higher order moments (spectral characteristics) of the original spectrum.

The pivotal role in patient recruitment throughout the entire study duration was assumed by Ms. Scarlet Nazarian. This responsibility was carried out in close collaboration with the author of this thesis and the proficient research team at Imperial College Healthcare NHS Trust.

5.6 Data Analysis

All data processing and analysis was performed by in-house developed software (using Python v3.9 and MATLAB v2022a). As outlined in Section 4.6, the primary data analysis process can be summarised into the following key steps.

In the acquired spectra, intensity variations might not stem from tissue-specific differences, but from factors such as the distance between the illumination source and tissue versus the Spectralon® sample used for calibration. The calibration process seeks to homogenise measurements, but when tissue thickness varies and leads to inconsistent source-to-sample distances or due to fluctuations in fibre-spectrometer coupling, these differences can produce intensity discrepancies.

To address these differences in intensity, normalisation of spectra was carried out prior to feature extraction and classification. Based on the performance recorded in Section 4.7, the SNV normalisation was utilised for in-vivo spectral data normalisation.

To deal with outliers, the automated outlier detection and removal method described in Section 4.6 was used. To tackle the issue of high dimensionality and overfitting, the use of feature extraction and selection techniques was crucial. For the in-vivo spectral dataset the shape-based and haemoglobin saturation coefficient extracted features, and the Boruta feature selection method were utilised. After determination of the reduced feature set, data were divided into training and testing sets using the repeated group stratified k-fold cross-validation (SKCV) method, with 5 folds and 5 repeats.

To differentiate between tumour and non-tumour in-vivo GI tissue, different classifiers were utilised, such as the Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine and the SpecNet. The working principles of these classifiers were outlined in Section 4.6. Finally, the performance of these classifiers was evaluated through key metrics, inclusive of sensitivity, specificity, overall accuracy, and the area under the curve (AUC).

Classifiers utilising only reflectance values (**Approach 1**) were compared with classifiers utilising raw reflectance values enriched with the seven additional shape-based features (**Approach 2**), with the haemoglobin saturation coefficient tissue optical feature (**Approach 3**) or with both shape-based features and the haemoglobin saturation coefficient tissue optical feature (**Approach 4**).

5.7 Results

Cohort Characteristics

A total of 10 patients were recruited for this study between May 2022 and August 2023. The mean age of these patients was 65 years (52-76) and 6 (60%) patients were male. A total of 7 patients (70%) underwent an oesophagectomy and 3 (30%) underwent a gastrectomy surgical procedure. In total, 5 patients (50%) had a tumour stage pT3 (TNM staging). Among the recruited patients, 8 (80%) had undergone neoadjuvant treatment (chemotherapy, chemoradiotherapy or radiotherapy alone). Additional patient characteristics can be seen in Table 27.

Table 27: Cohort characteristics of n=10 patients included in the analysis. Characteristics are presented for all patients in no. (%) as well as for gastric cancer and oesophageal cancer patients.

Characteristic	No. (%)		
	All Patients	Patients with cancer	
		Gastric	Oesophageal
Age, median (range)	65 (52-76)	54 (54-54)	71 (52-76)
Sex			
Male	4 (40.0)	1 (50.0)	4 (57.1)
Female	6 (60.0)	1 (50.0)	3 (42.9)
Procedure			
Oesophagectomy	7 (70.0)	0	7 (100.0)
Gastrectomy	2 (20.0)	2 (100.0)	0
Laparotomy	1 (10.0)	0	0
Tumour Histology			
Adenocarcinoma	9 (90.0)	2 (22.0)	7 (78.0)
Tumour Stage (TNM staging)			
pT0	0		
pT1	1 (10.0)	0	1 (14.3)
pT2	1 (10.0)	0	1 (14.3)
pT3	5 (50.0)	0	5 (71.4)
pT4	2 (20.0)	2 (100.0)	0
Neoadjuvant therapy			
Chemotherapy	7 (70.0)	2 (100.0)	5 (71.4)
Chemoradiotherapy	1 (10.0)	0	1 (14.3)
Radiotherapy	0		
None	2 (20.0)		1 (14.3)

Dataset Summary

Overall, 6 distinct sets of normal stomach data, 4 sets of gastric cancer data, 3 sets of normal oesophagus data, and 2 sets of oesophageal cancer data were included in the analysis. There were a total of 1,192 spectra for normal stomach and 541 spectra for gastric cancer. In oesophagus specimens, there was a total of 392 spectra for normal oesophagus and 229 spectra for oesophageal cancer. The means of all spectra with standard deviation for each of the tissue classes are shown in Figure 66.

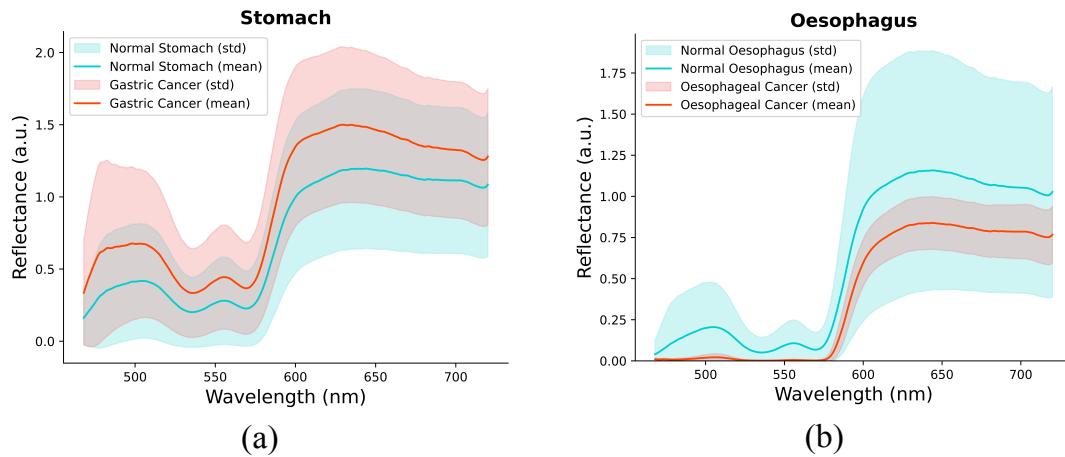


Figure 66: Mean absorption spectra with standard deviation for (a) in-vivo normal stomach and gastric cancer, and (b) in-vivo normal oesophagus and oesophageal cancer.

Feature Extraction and Selection

The first step of the spectral analysis pipeline was to normalise the data using the SNV normalisation technique. Afterwards, the shape-based features and the haemoglobin saturation coefficient were extracted (Figure 67).

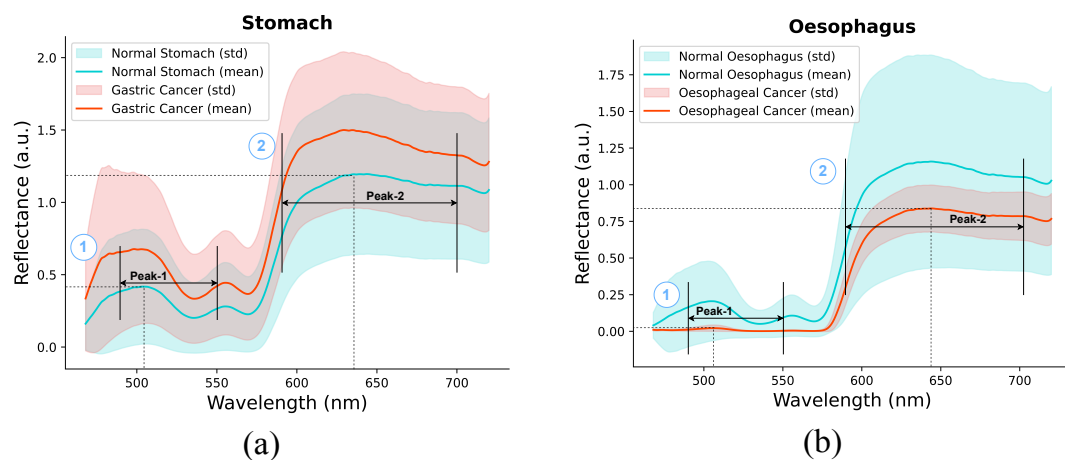


Figure 67: Illustration of shape-based features extracted for in-vivo (a) stomach and (b) oesophagus datasets. (1) Depicts the secondary peak intensity value and corresponding wavelength in the 480 – 550 nm spectral range. (2) Depicts the primary peak intensity value and corresponding wavelength in the 580 – 700 nm spectral range.

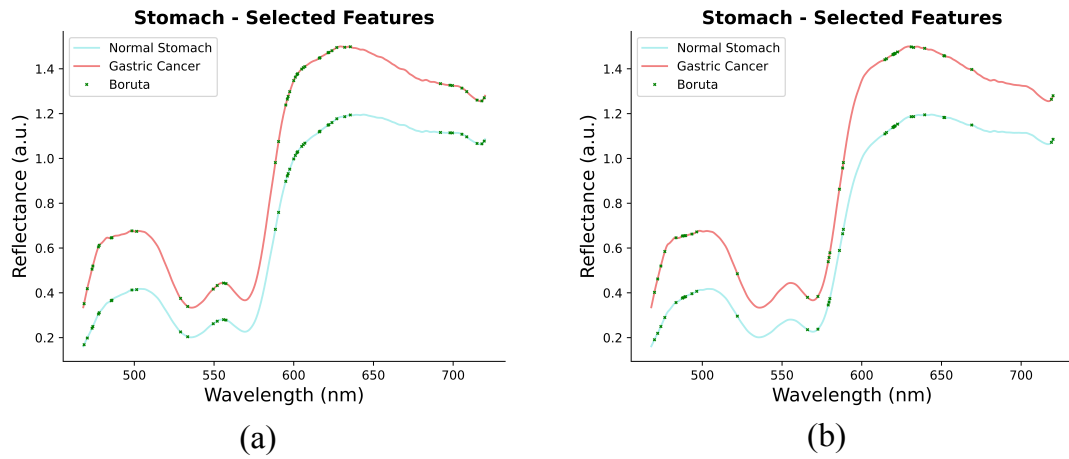


Figure 68: Selected features for the in-vivo stomach dataset for (a) SNV normalisation and (b) no normalisation using Boruta feature selection method.

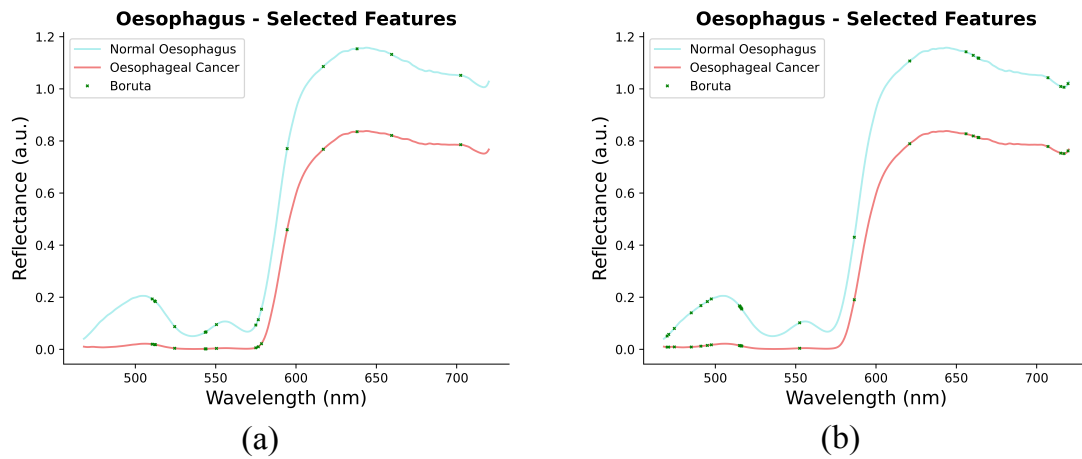


Figure 69: Selected features for the in-vivo oesophagus dataset for (a) SNV normalisation and (b) no normalisation using Boruta feature selection method.

Boruta feature selection method was then applied on each dataset following SNV normalisation. The selected features are shown in Figure 68 and Figure 69 for the in-vivo stomach and oesophagus datasets, respectively.

Exploring the Impact of Normalization on Classification Performance

The added value of normalisation methods on the classification outcome was also examined. To this end, the increase in overall accuracy, sensitivity, specificity, and AUC for the SNV normalisation method compared to no normalisation were analysed. This analysis was performed for the entire spectrum for the Random Forest, XGBoost, LightGBM and the SpecNet classifiers. The average increase in the aforementioned performance metrics for

each of the four approaches are shown in the following tables. No feature selection was considered for this investigation.

Table 28: Mean increase in overall accuracy, sensitivity, specificity, and AUC for SNV normalisation technique compared to no normalisation for all classifiers for the in-vivo stomach dataset.

Classifier	Mean (SD) (%)			
	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra				
RF	1.99 (0.50)	6.15 (0.71)	0.12 (0.17)	1.48 (0.14)
XGB	2.61 (0.08)	5.52 (0.73)	1.29 (0.00)	1.57 (0.27)
LGBM	2.30 (0.03)	5.63 (0.68)	0.79 (0.08)	1.58 (0.10)
SpecNet	-5.77 (0.98)	-14.20 (1.45)	-1.96 (0.51)	-2.43 (0.84)
Approach 2 – Reflectance spectra with shape-based features				
RF	1.99 (0.33)	5.48 (0.78)	0.42 (0.16)	1.14 (0.01)
XGB	2.75 (0.21)	5.47 (1.53)	1.51 (0.11)	1.30 (0.06)
LGBM	2.19 (0.12)	5.29 (1.21)	0.79 (0.08)	1.37 (0.11)
SpecNet	-1.73 (0.33)	-8.02 (1.15)	1.12 (0.27)	-0.58 (0.23)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient				
RF	1.77 (0.21)	5.37 (0.89)	0.13 (0.07)	1.41 (0.13)
XGB	2.78 (0.04)	5.11 (0.62)	1.73 (0.31)	1.64 (0.14)
LGBM	2.84 (0.21)	6.33 (1.27)	1.26 (0.34)	1.74 (0.14)
SpecNet	-5.96 (1.00)	-16.05 (1.71)	-1.40 (0.34)	-2.39 (0.83)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient				
RF	2.15 (0.46)	5.67 (0.61)	0.55 (0.02)	1.12 (0.00)
XGB	2.24 (0.10)	4.36 (1.43)	1.28 (0.27)	1.24 (0.03)
LGBM	2.22 (0.00)	5.14 (0.41)	0.89 (0.24)	1.37 (0.13)
SpecNet	-2.31 (0.50)	-6.79 (1.16)	-0.28 (0.08)	-0.83 (0.39)

According to Table 28 and Table 29, SNV normalisation technique improved the classification outcome significantly, for both the in-vivo stomach and oesophagus datasets, using the Random Forest, XGBoost, and LightGBM classifiers compared to no normalisation. In the case of SpecNet classifier, its classification performance showed no improvement. In fact, a noticeable decline was observed across all approaches. This decrease in performance can be attributed, in part, to the fundamental mechanisms underlying SNV normalisation technique.

Given that SNV normalisation showed performance improvement across nearly all classification algorithms, compared to no normalisation, it was decided to only use SNV normalisation for the comparison of different feature selection and extraction techniques in the next section.

Table 29: Mean increase in overall accuracy, sensitivity, specificity, and AUC for SNV normalisation technique compared to no normalisation for all classifiers for the in-vivo oesophagus dataset.

Classifier	Mean (SD) (%)			
	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra				
RF	-0.42 (0.18)	1.13 (1.73)	-1.32 (0.91)	-0.41 (0.51)
XGB	1.80 (0.37)	2.00 (1.48)	1.68 (0.36)	0.52 (0.25)
LGBM	1.58 (0.32)	1.05 (2.93)	1.89 (0.70)	0.65 (0.08)
SpecNet	-3.74 (0.82)	-7.25 (1.21)	-1.69 (0.45)	-1.49 (0.73)
Approach 2 – Reflectance spectra with shape-based features				
RF	0.10 (0.09)	1.84 (2.15)	-0.91 (0.62)	-0.01 (0.28)
XGB	3.25 (0.17)	3.66 (0.66)	3.01 (0.71)	1.23 (0.09)
LGBM	2.87 (0.66)	3.14 (2.60)	2.71 (0.87)	1.30 (0.21)
SpecNet	-1.07 (0.25)	-4.35 (0.79)	0.85 (0.24)	-0.54 (0.30)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient				
RF	0.07 (0.05)	2.10 (2.69)	-1.12 (0.46)	-0.42 (0.51)
XGB	2.45 (0.02)	2.70 (2.06)	2.30 (0.68)	0.72 (0.09)
LGBM	2.25 (0.17)	2.79 (2.39)	1.94 (0.79)	0.73 (0.06)
SpecNet	-3.21 (0.69)	-5.80 (0.92)	-1.69 (0.45)	-1.47 (0.73)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient				
RF	0.16 (0.29)	1.22 (2.33)	-0.46 (0.90)	-0.15 (0.36)
XGB	3.38 (0.22)	4.88 (1.23)	2.50 (0.31)	1.22 (0.07)
LGBM	3.45 (0.50)	4.70 (2.74)	2.71 (1.15)	1.35 (0.26)
SpecNet	-1.07 (0.25)	-1.45 (0.29)	-0.85 (0.21)	-0.88 (0.46)

Analysing the Influence of Feature Extraction on Classification Performance

Based on the results from the normalisation, the first analysis on feature extraction and selection was performed on data normalised using the SNV normalisation technique. In Table 30 and Table 31, the results are shown for the mean increase in classification performance metrics for all feature extraction methods (approaches 2, 3, 4) compared to no feature extraction, i.e., the entire spectrum (approach 1), for the in-vivo stomach and oesophagus datasets, respectively.

Table 30 and Table 31 illustrate the mean increase in performance metrics achieved by different feature extraction methods compared to not using feature extraction (i.e., using only the entire spectrum) for the in-vivo stomach and oesophagus datasets, respectively. It is worth mentioning that for both datasets, a noticeable trend emerges: the use of shape-based features either on their own or in combination with the haemoglobin saturation coefficient feature consistently improves classification outcomes. Similarly, using only the haemoglobin

saturation coefficient, did not show a noteworthy improvement in classification compared to utilising the full spectrum alone.

Table 30: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the in-vivo stomach dataset. Only SNV normalised data is used.

Classifier	Mean (SD) (%)			
	Accuracy	Sensitivity	Specificity	AUC
Approach 2 – Reflectance spectra with shape-based features				
RF	0.52 (0.16)	0.59 (0.14)	0.49 (0.10)	0.41 (0.16)
XGB	1.56 (0.31)	3.03 (1.06)	0.89 (0.07)	0.92 (0.16)
LGBM	1.92 (0.05)	3.81 (0.26)	1.06 (0.10)	0.89 (0.14)
SpecNet	3.27 (0.48)	8.02 (0.65)	1.12 (0.27)	1.56 (0.45)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient				
RF	-0.08 (0.07)	-0.29 (0.05)	0.02 (0.04)	0.06 (0.07)
XGB	0.14 (0.03)	0.22 (0.16)	0.30 (0.11)	0.06 (0.03)
LGBM	0.36 (0.18)	0.52 (0.28)	0.29 (0.08)	0.07 (0.01)
SpecNet	-0.19 (0.02)	-0.62 (0.04)	0.00 (0.00)	0.00 (0.00)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient				
RF	0.63 (0.20)	0.70 (0.00)	0.60 (0.25)	0.43 (0.13)
XGB	1.35 (0.32)	2.55 (1.03)	0.81 (0.02)	0.95 (0.13)
LGBM	1.84 (0.04)	3.44 (0.18)	1.11 (0.06)	0.88 (0.20)
SpecNet	3.85 (0.58)	10.49 (0.93)	0.84 (0.19)	1.63 (0.48)

Table 31: Mean increase in classification performance metrics for all feature extraction methods compared to the use of the entire spectrum for the in-vivo oesophagus dataset. Only SNV normalised data is used.

Classifier	Mean (SD) (%)			
	Accuracy	Sensitivity	Specificity	AUC
Approach 2 – Reflectance spectra with shape-based features				
RF	0.74 (0.09)	1.31 (0.25)	0.41 (0.05)	0.35 (0.19)
XGB	2.09 (0.40)	2.35 (0.09)	1.94 (0.82)	1.04 (0.36)
LGBM	2.13 (0.21)	2.79 (0.15)	1.73 (0.43)	1.01 (0.33)
SpecNet	2.14 (0.44)	2.90 (0.42)	1.69 (0.45)	0.95 (0.43)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient				
RF	0.48 (0.05)	0.87 (0.39)	0.25 (0.26)	0.05 (0.04)
XGB	1.06 (0.43)	1.31 (0.52)	0.92 (0.70)	0.50 (0.19)
LGBM	0.97 (0.24)	1.65 (0.41)	0.56 (0.27)	0.26 (0.11)
SpecNet	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient				
RF	0.58 (0.20)	0.87 (0.31)	0.41 (0.05)	0.32 (0.17)
XGB	2.32 (0.45)	3.22 (0.21)	1.79 (0.74)	1.20 (0.41)
LGBM	2.32 (0.00)	3.13 (0.01)	1.83 (0.44)	1.24 (0.47)
SpecNet	2.14 (0.44)	5.80 (0.92)	0.00 (0.00)	0.53 (0.23)

Table 32: Mean increase in classification performance metrics for Boruta feature selection method compared to the use of the entire spectrum for the in-vivo stomach dataset. Only SNV normalised data is used.

Classifier	Mean (SD) (%)			
	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra				
RF	0.16 (0.00)	1.03 (0.13)	0.24 (0.00)	0.05 (0.01)
XGB	0.13 (0.08)	0.89 (0.16)	-0.22 (0.00)	0.14 (0.13)
LGBM	0.39 (0.03)	0.92 (0.09)	0.15 (0.22)	0.12 (0.01)
SpecNet	2.31 (0.32)	7.41 (0.59)	0.00 (0.00)	0.51 (0.13)
Approach 2 – Reflectance spectra with shape-based features				
RF	1.02 (0.16)	3.10 (0.66)	0.07 (0.00)	0.44 (0.08)
XGB	0.14 (0.20)	0.78 (0.88)	-0.15 (0.08)	0.24 (0.07)
LGBM	0.12 (0.02)	0.33 (0.15)	0.02 (0.00)	0.11 (0.04)
SpecNet	1.92 (0.37)	7.41 (1.03)	-0.56 (0.14)	0.74 (0.33)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient				
RF	0.50 (0.09)	1.77 (0.06)	-0.08 (0.12)	0.07 (0.08)
XGB	0.17 (0.25)	1.11 (0.21)	-0.25 (0.02)	0.10 (0.14)
LGBM	0.27 (0.34)	0.59 (0.29)	0.12 (0.00)	0.09 (0.00)
SpecNet	2.50 (0.34)	6.79 (0.52)	0.56 (0.13)	0.62 (0.16)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient				
RF	1.02 (0.24)	2.95 (0.97)	0.13 (0.01)	0.47 (0.19)
XGB	0.40 (0.28)	1.26 (0.96)	0.02 (0.00)	0.21 (0.08)
LGBM	0.35 (0.02)	0.70 (0.33)	0.18 (0.20)	0.14 (0.06)
SpecNet	-2.88 (0.45)	-9.26 (0.85)	0.00 (0.00)	-0.93 (0.30)

Table 32 and Table 33 present the outcomes indicating the mean increase in classification performance metrics obtained by utilising the Boruta feature selection method in comparison to the use of the entire spectrum was taken as a reference. These results are founded on data normalised using SNV normalisation for the in-vivo stomach and oesophagus datasets, respectively.

Our study unveiled an intriguing observation about the implementation of Boruta feature selection method among all classifiers. Despite our initial assumptions, the incorporation of this method for selecting features did not result in noteworthy enhancements in classification performance. This lack of significant improvement can be attributed to the inherent versatility of ensemble-based techniques and the relatively small in-vivo datasets.

Table 33: Mean increase in classification performance metrics for Boruta feature selection method compared to the use of the entire spectrum for the in-vivo oesophagus dataset. Only SNV normalised data is used.

Classifier	Mean (SD) (%)			
	Accuracy	Sensitivity	Specificity	AUC
Approach 1 – Reflectance spectra				
RF	0.42 (0.27)	0.95 (0.35)	0.10 (0.60)	0.14 (0.20)
XGB	0.32 (0.60)	0.18 (0.49)	0.41 (0.21)	0.30 (0.21)
LGBM	-0.39 (0.47)	0.00 (1.63)	-0.61 (0.36)	0.20 (0.27)
SpecNet	2.67 (0.56)	5.80 (0.92)	0.85 (0.21)	1.12 (0.53)
Approach 2 – Reflectance spectra with shape-based features				
RF	1.39 (0.14)	1.30 (1.17)	1.43 (0.70)	0.74 (0.33)
XGB	0.32 (0.02)	0.53 (0.38)	0.21 (0.44)	0.30 (0.22)
LGBM	0.55 (0.39)	0.88 (0.71)	0.36 (0.05)	0.30 (0.16)
SpecNet	-2.67 (0.54)	-1.45 (0.22)	-3.39 (0.84)	-1.41 (0.62)
Approach 3 – Reflectance spectra with haemoglobin saturation coefficient				
RF	0.77 (0.48)	1.05 (0.61)	0.61 (0.60)	0.27 (0.13)
XGB	0.32 (0.39)	0.97 (1.00)	-0.05 (0.39)	-0.07 (0.07)
LGBM	-0.13 (0.31)	0.18 (0.69)	-0.31 (0.42)	0.24 (0.06)
SpecNet	-3.21 (0.55)	-10.14 (1.04)	0.85 (0.21)	-1.90 (0.68)
Approach 4 – Reflectance spectra with shape-based features and haemoglobin saturation coefficient				
RF	1.71 (0.08)	1.75 (0.37)	1.68 (0.74)	0.88 (0.37)
XGB	0.77 (0.15)	0.97 (0.04)	0.66 (0.07)	0.25 (0.22)
LGBM	0.48 (0.11)	0.96 (0.57)	0.20 (0.02)	0.18 (0.13)
SpecNet	1.60 (0.38)	-1.45 (0.26)	3.39 (1.01)	-0.27 (0.12)

Healthy versus Tumour Tissue Classification Results

Up to this point, we have assessed normalisation, feature extraction and feature selection methods individually. In the current section, we present the classification results for the in-vivo stomach and oesophagus dataset. The following tables present the performance metrics for the most effective combinations of normalisation, feature extraction, feature selection and classification method, as described in Section 4.6. The results demonstrate the ability of the selected classifiers to differentiate between normal and tumour tissue.

Table 34: Performance metrics for the spectral data classification using Approach 1 (reflectance values) using SNV normalisation and Boruta feature selection.

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	92.86 (1.61)	85.03 (3.23)	96.41 (1.52)	97.52 (0.86)	91.60 (2.27)	87.85 (4.14)	93.78 (2.62)	97.54 (1.08)
XGB	93.70 (1.13)	87.18 (3.22)	96.66 (1.34)	97.74 (0.66)	91.92 (1.74)	87.42 (3.92)	94.54 (2.86)	97.47 (1.13)
LGBM	94.02 (1.16)	87.88 (2.95)	96.81 (1.04)	98.05 (0.65)	91.85 (2.17)	88.05 (5.42)	94.09 (2.17)	97.84 (0.81)
SpecNet	93.65 (2.10)	87.65 (2.83)	96.37 (1.61)	97.59 (1.32)	92.51 (3.77)	89.86 (4.33)	94.07 (3.39)	98.37 (1.81)

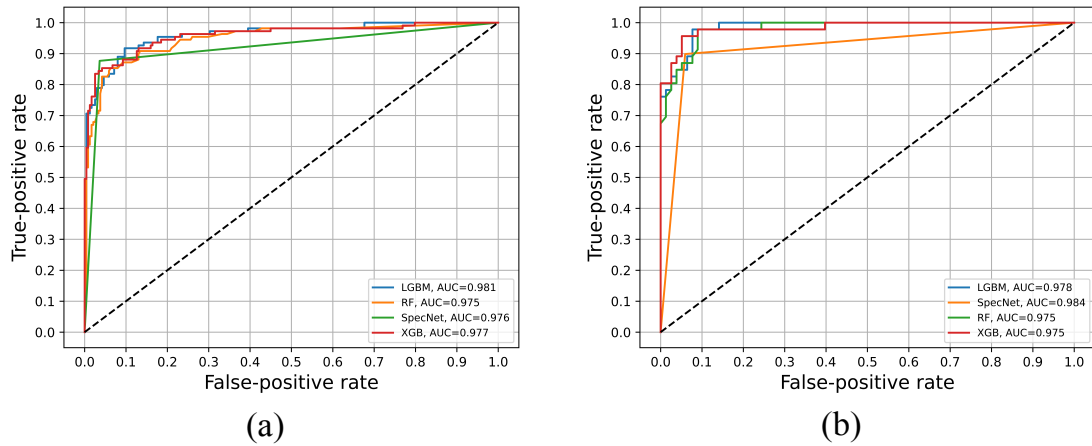


Figure 70: ROC curves with AUC values for all classifiers using Approach 1 (entire spectrum) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.

Table 35: Performance metrics for the spectral data classification using Approach 2 (reflectance values and shape-based extracted features) using SNV normalisation and Boruta feature selection.

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	94.23 (1.29)	87.69 (2.84)	97.20 (1.43)	98.31 (0.60)	93.30 (2.05)	89.51 (4.71)	95.51 (2.47)	98.50 (0.76)
XGB	95.27 (1.11)	90.09 (2.88)	97.62 (1.19)	98.77 (0.56)	94.01 (1.96)	90.13 (4.88)	96.28 (2.69)	98.52 (0.76)
LGBM	95.66 (1.10)	91.09 (2.63)	97.73 (1.16)	98.93 (0.56)	94.91 (1.87)	91.70 (4.36)	96.79 (2.04)	98.96 (0.59)
SpecNet	96.54 (1.57)	95.68 (1.74)	96.93 (1.48)	99.38 (0.67)	89.30 (4.43)	85.51 (5.05)	91.53 (3.99)	96.78 (2.53)

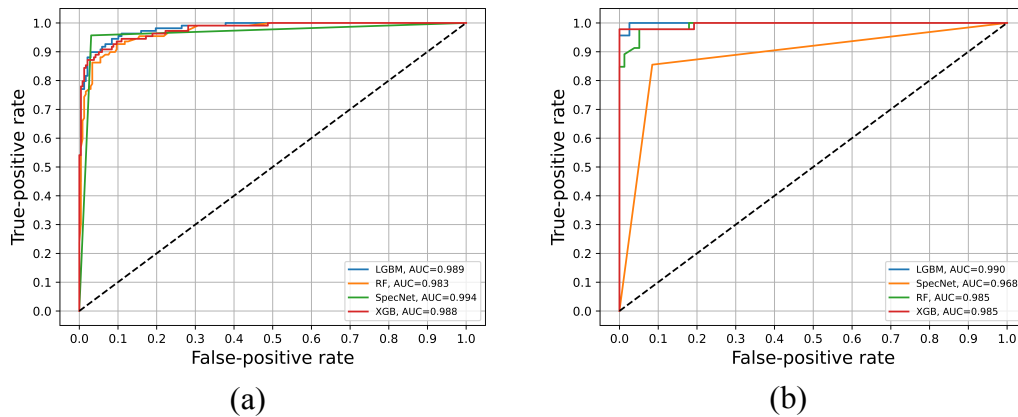


Figure 71: ROC curves with AUC values for all classifiers using Approach 2 (reflectance values and shape-based extracted features) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.

Table 36: Performance metrics for the spectral data classification using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) using SNV normalisation and Boruta feature selection.

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	93.11 (1.46)	85.48 (3.34)	96.58 (1.36)	97.60 (0.85)	92.43 (1.57)	88.82 (2.80)	94.54 (2.37)	97.73 (1.11)
XGB	93.88 (0.94)	87.18 (2.69)	96.93 (1.21)	97.76 (0.62)	92.98 (2.30)	89.52 (4.89)	95.00 (2.77)	97.60 (1.21)
LGBM	94.25 (1.29)	88.06 (3.04)	97.06 (1.33)	98.09 (0.65)	93.08 (2.25)	89.87 (4.89)	94.95 (2.68)	98.14 (0.92)
SpecNet	93.65 (2.10)	86.42 (2.94)	96.93 (1.48)	97.69 (1.29)	86.63 (4.88)	73.91 (6.29)	94.07 (3.39)	95.35 (3.02)

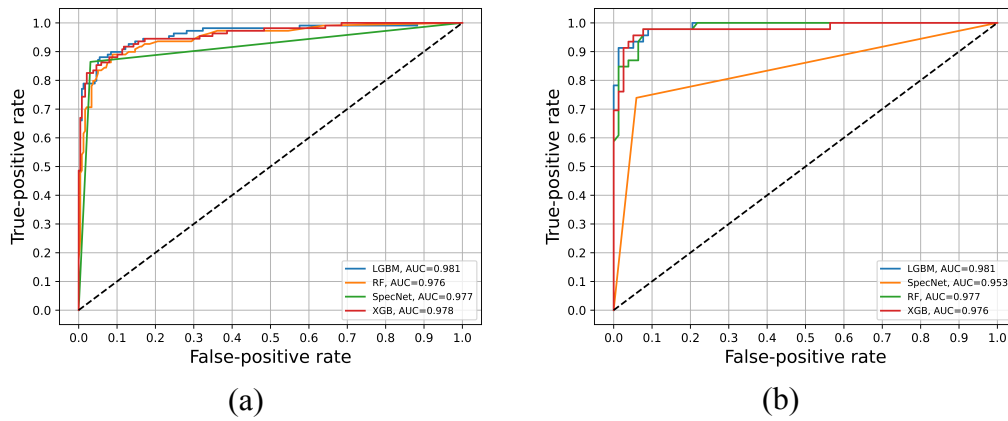


Figure 72: ROC curves with AUC values for all classifiers using Approach 3 (reflectance values and haemoglobin saturation coefficient extracted tissue optical feature) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.

Table 37: Performance metrics for the spectral data classification using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) using SNV normalisation and Boruta feature selection.

Classifier	Mean (SD) (%)							
	Stomach				Oesophagus			
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC
RF	94.34 (1.17)	87.66 (2.39)	97.38 (1.27)	98.35 (0.52)	93.46 (2.12)	89.52 (3.86)	95.77 (2.43)	98.60 (0.74)
XGB	95.33 (1.17)	90.10 (2.99)	97.70 (1.31)	98.76 (0.57)	94.69 (2.04)	91.44 (4.58)	96.58 (2.40)	98.63 (0.71)
LGBM	95.81 (1.15)	91.09 (2.89)	97.95 (1.13)	98.95 (0.52)	95.04 (1.82)	92.14 (4.38)	96.74 (2.10)	99.06 (0.49)
SpecNet	92.31 (2.29)	81.48 (3.34)	97.21 (1.42)	97.78 (1.27)	93.58 (3.51)	88.41 (4.59)	96.61 (2.59)	97.51 (2.23)

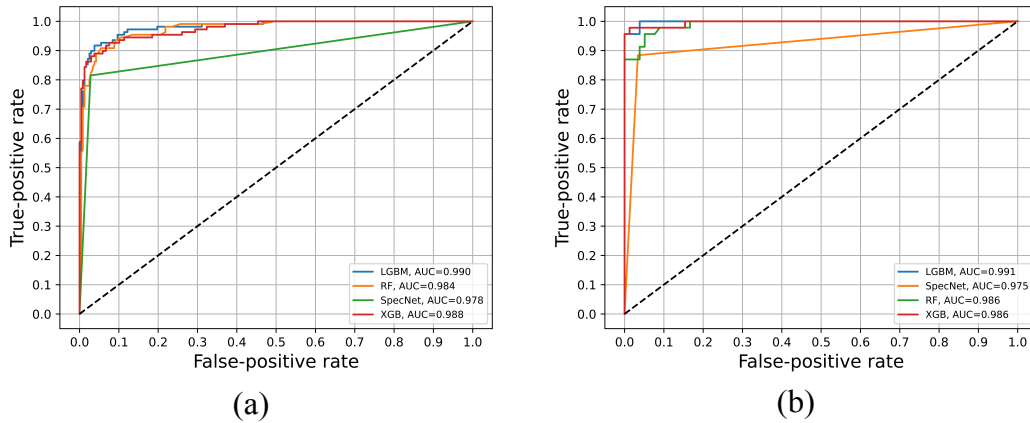


Figure 73: ROC curves with AUC values for all classifiers using Approach 4 (reflectance values, shape-based extracted features, and haemoglobin saturation coefficient extracted tissue optical feature) for (a) in-vivo stomach dataset and (b) in-vivo oesophagus dataset using SNV normalisation and Boruta feature selection.

Remarkably, for all approaches, the combination of SNV normalisation and Boruta feature selection consistently emerged as the superior choice for both the in-vivo stomach and oesophagus datasets.

The high classification accuracy achieved by integrating SNV normalisation and the Boruta feature selection method is fundamentally based on their inherent features. SNV normalisation, which centres each spectrum around its mean and scales it by the standard deviation, ensures the baseline is uniform and compensates for overall intensity changes. This important procedure improves the consistency of spectra across samples, thus reducing the effect of extraneous differences that could confuse classification models. Therefore, implementing SNV normalization guarantees that the significant spectral characteristics, crucial for differentiation, become more prominent, thereby enhancing model precision.

On the contrary, the Boruta technique for feature selection stands out in its ability to detect unique patterns within spectral data by giving priority to features that exhibit consistent discriminative power across iterations. This attribute is crucial for identifying spectral regions containing critical information for classification purposes.

When both of these methods are employed simultaneously, SNV normalization efficiently reduces variations in intensity, allowing Boruta to focus on relevant spectral variations related to molecular interactions.

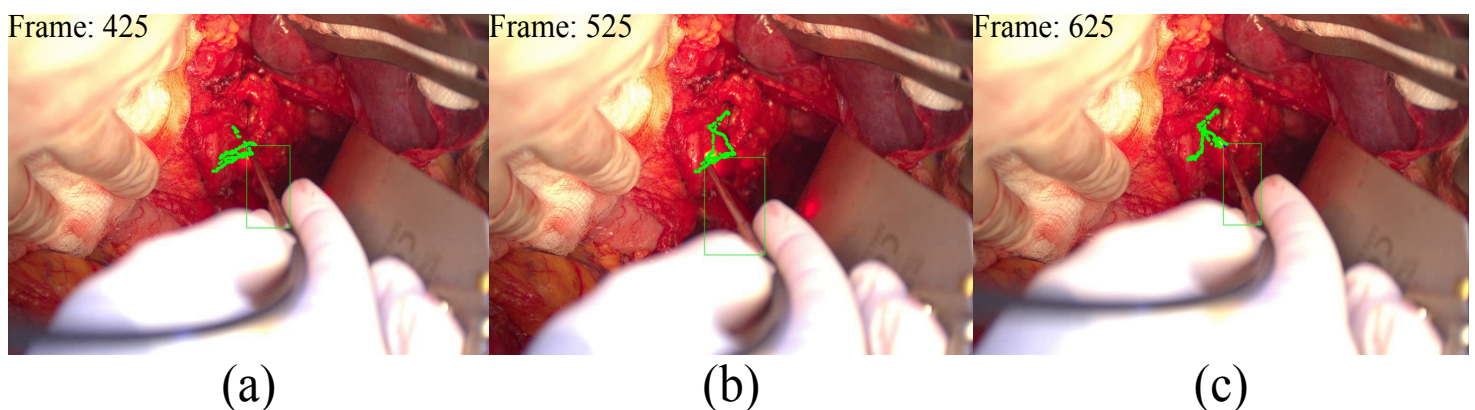


Figure 74: Illustration of real-time DRS probe marker-less tracking at each optical biopsy site during an oesophagectomy. (a) – (c) Real-time visual correlation of scanned optical biopsy sites on tissue (green dots on each frame) using the DRS probe and marker-less tracking.

5.8 Summary

To the best of our knowledge this is the first in-vivo study using DRS to distinguish tumour tissue from healthy tissue in gastric and oesophageal cancer surgery. The results of our in-vivo study using DRS to differentiate tumour tissue from healthy tissue in gastric and oesophageal cancer surgery are indeed promising. We achieved a sensitivity and specificity of 88.11% and 97.09% for the stomach and 88.35% and 95.05% for the oesophagus. There

results suggest that DRS has great potential for real-time tissue discrimination during gastrointestinal cancer surgery.

When comparing the current in-vivo study to the results obtained previously in the ex-vivo study (Section 4.7.1), the accuracies seem similar, ranging from 89% to 94%. However, it should be acknowledged that the in-vivo study had a relatively small sample size of only 10 patients. This limited sample size may have contributed to the high accuracy rates obtained. Typically, when working with larger datasets, performance metrics tend to stabilise and may decrease slightly due to increased variability in patient characteristics, tumour types and other factors. As we expand the study to include more patients, it is reasonable to expect that the performance metrics, while still promising, may show some variability and possibly a gradual decline. This is a common phenomenon in medical research and highlights the importance of validating these results in larger and more diverse patient populations to establish the robustness and generalisability of the DRS technology for gastrointestinal tumour identification and therefore margin assessment.

A primary limitation that could influence the classification performance of the in-vivo study stems from the inherently less controlled intra-operative environment. This setting, as opposed to more controlled ones (ex-vivo), can lead to challenges in achieving precise histopathology correlation. Consequently, this might result in reduced accuracy. This is very difficult to circumvent. To derive a more reliable and accurate differentiation between healthy and tumour tissue, it is imperative to expand the study by including a larger patient cohort. Increasing the number of patients not only increases the likelihood of obtaining clear histopathological findings but may also improve the ratio of patients with clear histopathology to those with ambiguous results. The larger the study sample, the more robust and generalisable the results will be, reducing the impact of inherent variability in individual patient responses.

The translation of DRS into clinical practice has taken some time. This in-vivo study represents a significant leap forward in the integration of DRS into clinical application. It demonstrates the feasibility of using DRS during gastrointestinal cancer surgery to discriminate between healthy and tumour tissue in real-time. This study opens the door to further development of the technology and paves the way for potential incorporation into intelligent smart tools. The integration of DRS into surgical instruments promises to improve surgical precision and decision-making. It also extends the applicability of DRS to minimally

invasive procedures, such as laparoscopic and robotic surgery, where real-time tissue discrimination could greatly benefit patient outcomes.

Chapter 6: Conclusions and Future Work

6.1 *Conclusions*

In this doctoral thesis, we present a pioneering exploration of the advantages and challenges associated with the use of diffuse reflectance spectroscopy as a margin assessment technique in the context of gastrointestinal cancer surgery. This system has been carefully designed to provide high spectral resolution and rapid data acquisition. These technical capabilities allowed us to efficiently scan large areas of tissue, enabling real-time tissue differentiation with minimal disruption to the surgical procedure. Our research efforts represent a significant milestone as, to the best of our knowledge, they represent the first systematic investigation of DRS for this specific intraoperative clinical application.

In particular, fibre-optic DRS demonstrated the capability of distinguishing tumour from healthy tissue even in the challenging in-vivo environment. As the cancer treatment landscape evolves with advances in neoadjuvant therapies, the need for intraoperative tissue classification may appear to be diminishing. In addition, the introduction of population screening is expected to detect gastrointestinal tumour at earlier stages. A recent study on the use of Raman spectroscopy showed high accuracy in screening GI tumours, suggesting that early detection could significantly enhance the precision of margin mapping by identifying serum components unique to cancer patients³²². While these developments might seem to reduce the need for intraoperative tissue classification, the presence of advanced tumours that necessitate surgical intervention underscores its ongoing importance. This technology's impact on margin screening is significant, offering the potential to more accurately define tumour margins during surgery, thus improving surgical outcomes and patient prognosis in cases where complete tumour removal is critical.

Our primary objective was to contribute to the overarching and long-term goal of providing direct diagnostic feedback on resection margins in gastrointestinal specimens through the application of DRS. By providing real-time insight into the presence of residual malignant tissue, our ultimate goal was to enable surgical teams to make informed decisions during the initial operation, thereby improving the precision and effectiveness of cancer surgery.

A central focus of this thesis has been to establish a robust and high correlation between optical measurements and histopathological data. This effort was aimed at constructing a reliable and meticulously curated spectral dataset that could serve as an excellent basis for rigorously testing the extensive capabilities of DRS as a margin assessment technique.

A successful intraoperative margin assessment technique should be able to provide margin assessment without the need for an experienced operator or interpreter of the results. This criterion embodies the broader objective of streamlining and democratising the use of such techniques in the real-world clinical setting. The ability to operate autonomously, with minimal reliance on specialist expertise, is of paramount importance in the context of intraoperative margin assessment. It ensures that the technique can be seamlessly integrated into the surgical workflow, where time is often of the essence. Ideally, surgeons and OR staff should be able to easily access and interpret the assessment results, enabling them to make informed decision in real-time.

In this thesis, spectral data was analysed using a machine learning methodology that uses experimental data to develop classification algorithms. Initially, our data analysis was focused on excised tissue specimens. This was a deliberate choice, guided by the favourable correlation observed between these spectral measurements and histopathological assessment. In addition, this approach was supported by the availability of an ample number of both tumour and healthy tissue measurements, which facilitated rigorous analysis. In Chapter 5, we made a significant step towards the realisation of our overarching goal – the ability to provide real-time feedback during gastrointestinal cancer surgery. During this phase, we expanded our data collection efforts to include in-vivo gastrointestinal specimens.

The development of a robust machine learning methodology for tissue discrimination requires careful consideration of several critical factors. These factors are key to ensuring the accuracy, reliability, and real-world applicability of the methodology. Firstly, one of the main concerns is the risk of overfitting, where the classification algorithms are overly tailored to the training data. To mitigate this challenge and ensure the models' generalisation capability, we employed several strategies. One primary method was the use of cross-validation, specifically stratified group 5-fold cross-validation with 5 repeats, which helps in evaluating the model's performance on unseen data by dividing the data into 5 subsets and training the model 5 times, each time using a different subset as the test set and the remaining as the

training set. Additionally, we focused on hyperparameter optimisation to fine-tune the models' parameters for optimal performance. This process involved systematically searching through a range of hyperparameters to find the combination that results in the best model performance on the validation set. Grid search technique was used to explore the hyperparameter space efficiently. Moreover, we closely monitored the standard deviation (SD) across the cross-validation folds. A very small SD in model performance metrics (i.e., sensitivity, specificity) across the folds indicated that our model was stable and not overly fitted to a particular subset of the training data. This stability across different subsets of data demonstrates the model's ability to generalise well to new, unseen data. Through these methods, we ensured that our machine and deep learning models were not only tailored to perform well on the training data but also maintained their performance and reliability on unseen test data, thereby mitigating the risk of overfitting and enhancing their real-world applicability. Second, the representativeness of the training data plays a crucial role in the development of a reliable machine learning algorithm. Although it can be difficult to set an exact threshold for representativeness, it is generally accepted that larger datasets are more likely to reflect the complexity of real-world scenarios. Consequently, our efforts have focused on accumulating a substantial amount of data to improve the algorithm's ability to handle the diversity of tissue types encountered in practice. Thirdly, the use of ground truth labels in supervised learning is paramount. It is essential that these labels accurately represent the characteristics of the measured data. The reliability and representativeness of the ground truth data served as a cornerstone of our approach, ensuring that the machine learning algorithm learned from high quality references. Finally, the performance of any machine learning methodology is intrinsically linked to the quality of the data used for training. With this in mind, careful pre-processing of the raw spectral data was undertaken. This pre-processing involved data normalisation, outliers detection and removal, feature extraction and selection. These efforts were aimed at ensuring that the input data provided a solid foundation on which the machine learning models could be built.

In this research, a dynamic approach to data processing was adopted to match the incremental collection of patient data. Rather than processing the data as a complete set once all the data had been collected, the data were analysed in batches, each corresponding to the datasets of 10-20 patients. This periodic approach not only allowed for timely evaluation and recalibration, but also provided insight into the adaptability of the classification model to

newly acquired data. As more batches of patient data were included in the analysis, there was a trend towards stabilisation of the classification performance. Significant variability was seen in the early batches, reflecting the models' learning curve. However, subsequent batches showed less variation in results, indicating a trend towards convergence. Our methodology for assessing the variation in learning with incremental patient data collection incorporated a dynamic analysis approach. By evaluating performance metrics across batches, we could effectively monitor the models' learning progression and adaptation. This rigorous evaluation framework ensured our models' performance stabilised over time, demonstrating its potential reliability and clinical relevance. This stabilisation suggests that the models have achieved a degree of generalisation, becoming less influenced by the peculiarities of individual batches and more able to identify broader, consistent patterns. After processing the initial batches, the models were found to consistently process new patient data with reliable accuracy. This consistent performance underlines the potential reliability and clinical relevance of the approach.

Our real-time probe detection and tracking methodology has demonstrated exceptional accuracy when compared to ground truth data both ex-vivo and in-vivo. This achievement represents a significant advance over the point-based limitations of other DRS systems^{262,265}. Overcoming the challenges of an in-vivo environment is not trivial. Several factors such as the natural movement of tissues and organs, the presence of blood and other fluids can optically interfere with the readings. However, where other DRS systems have found limitations, our methodology has found opportunities. The traditional point-based limitations of many optical spectroscopy systems often result in a narrow focus, potentially missing broader tissue variations. Our approach provides surgeons with a holistic view. The real-time, enriched data acquisition video feeds allow precise localisation of optical biopsy sites, ensuring that each optical biopsy provides a comprehensive understanding of the tissue in question. This detection and tracking system, coupled with the classification of each sampled site on the tissue, has the potential to revolutionise intraoperative cancer margin assessment.

In contrast to the current standard of frozen sectioning, which can incur significant costs, in excess of £3,000 per patient³²³, the system we have developed offers a cost-effective alternative. Our portable system ensures ease of use and eliminates the need for additional staff or labour. The initial investment for our DRS system is an one-off cost of £8,000, making

it a significantly more economical option. Importantly, the system can be used for multiple sampling events, significantly reducing the cost per patient. In addition, our DRS system is completely non-invasive, preserving tissue integrity and structure without disruption. This is particularly important when analysing small tumours, ensuring that the tissue sent to histopathology remains unaltered for full assessment.

In conclusion, this work highlights the significant potential of the DRS modality as a powerful tool for probing tissue. The findings presented here lay the foundation for its potential integration into standard clinical procedures for the screening and diagnosis of tissue pathologies. As research in this field advances, we hope that these findings will contribute to the continued evolution of clinical practice, ultimately benefiting patients through improved and more accurate interventions.

6.2 *Future Work*

The large tissue deformation that occurs between the optical measurements and histopathological margin assessment is a well-known challenge in the validation of optical techniques. Specifically, we have developed a protocol for correlating the H&E stained tissue sections with the white light images that were contemporaneously captured alongside the optical measurements. In the future, we intend to improve and optimise this registration process. This is particularly important as achieving precise alignment between histopathological data and optical measurements is fundamental to the full validation of our DRS technique. Such methodology will include both affine and deformable registration techniques specifically tailored to account for tissue deformation. This advanced approach promises to significantly improve the correlation between optical measurements and histopathological data, surpassing the performance of the current histopathology correlation protocol. This transformative step forward will result in the generation of a comprehensive spectral database characterised by a remarkably high degree of correlation with histopathological findings. Overcoming this challenge is pivotal for the validation and clinical viability of our DRS technique intraoperatively.

It is worth noting that this histopathology correlation approach has broader implications for other optical probes as well. Several optical techniques currently face the challenge of aligning in-vivo measurements with subsequent histopathological evaluations.

By successfully implementing and refining this correlation protocol, we could establish a gold standard methodology that is not limited to our DRS probe, but it also adaptable to other diagnostic probes. The broad application of this correlation approach underlines its potential to revolutionise the field of optical diagnostics beyond our immediate research focus.

Lack of external validation is a major barrier to the safe implementation and routine use of artificial intelligence classification models in clinical practice³²⁴. As such, external validation of our classification models on a separate cohort of patients is a crucial step towards the clinical translation of this technology. It is also vital that decision analysis studies and key stakeholder interviews are conducted to identify the optimal use-cases for the technology, the barriers, and facilitators to its adoption in the future^{325,326}.

Diffuse reflectance spectroscopy proved to be a suitable candidate for the real-time detection of malignancies in gastrointestinal tissue. Its diagnostic performance when applied to gastrointestinal tissue specimens demonstrated remarkable accuracy, and it excelled in terms of efficiency, requiring minimal time for both data acquisition and analysis. However, one of the main challenges remains the discrimination between cancerous and fibrosis tissue.

Fibrosis, characterised by the formation of excessive fibrous connective tissue, can exhibit spectral signatures that overlap with those of both healthy and tumour tissues, complicating its recognition. Despite its clinical relevance, it is important to clarify that our analysis did not include measurements of fibrotic spectral data due to the difficulties associated with accurately identifying fibrosis in the spectral data. The lack of fibrosis data undoubtedly left a significant gap in our investigation, as without fibrosis data we were unable to make direct distinctions between tumour and fibrosis.

The difficulty in distinguishing between cancer and fibrosis is a critical consideration as it directly affects the accuracy of margin assessment during surgery. Surgeons need precise information to accurately determine the boundaries of malignant tissue. If fibrosis is closely associated with tumour tissue or if its presence could complicate the accurate assessment of tumour margins, a surgeon might decide to remove it to ensure clear margins. The presence of fibrosis in relation to the tumour can complicate margin assessment, as it might be difficult to determine whether fibrosis is part of the tumour's microenvironment or a separate pathological process. These challenges posed by fibrosis underscore the importance of continued research and refinement of DRS techniques, with a focus on improving the classification algorithms' ability to differentiate between these two tissue types.

However, if research is successful in demonstrating the ability to distinguish between tumour tissue and fibrosis, fibre-optic DRS holds great promise for surgical applications. DRS, for example, could be seamlessly integrated into surgical instruments, providing surgeons real-time insights into tissue composition just millimetres before they operate. This capability could significantly improve surgical precision and decision-making, potentially leading to improved patient outcomes.

Moreover, the incorporation of DRS into endoscopic and laparoscopic procedures marks a significant advance in the field of gastrointestinal cancer surgery. By seamlessly integrating DRS into these modalities, surgeons can access real-time tissue assessment capabilities during these minimally invasive procedures, providing an invaluable tool for precision and accuracy in cancer detection and margin assessment. This has the potential to minimise the extent of resection while ensuring complete removal of cancerous tissue, sparing healthy tissue, and reducing the risk of post-operative complications. In addition, the introduction of DRS into laparoscopic procedures will have an impact on the early detection and management of GI cancers. DRS-enabled laparoscopes can facilitate optical biopsy during routine staging laparoscopy, offering the potential for early cancer detection, staging and treatment planning without the need for separate biopsy procedures. Another important aspect to highlight is our recent efforts to explore the feasibility of using our DRS probe in a laparoscopic training environment. Our research group is currently in the process of conducting a feasibility study in this area, which is intended to lay the groundwork for future real-world applications. In parallel, an exploratory study has been initiated to evaluate the DRS probe in real patient scenarios during staging laparoscopic procedures.

In the current phase of our research, we have introduced augmented reality technology to enhance the surgeon's visualisation and decision-making process during procedures involving optical biopsy sites. This was achieved by superimposing classified optical biopsy sites, represented by a graded colour map, onto the live video feedback from the white light RGB camera. This approach provides valuable spatial information to the surgeon, aiding them to identify regions of interest in real-time.

As we look to the future and consider the potential transition to virtual reality (VR) technology for this application, several exciting possibilities emerge. VR offers a more immersive and three-dimensional experience compared to AR, which overlays digital

information onto the real-world view. Here's how the integration of VR could enhance the application of DRS in surgical procedures:

- **Enhanced depth perception:** VR headsets can provide surgeons with an immersive 3D view of the surgical field. This enhanced depth perception can be particularly beneficial when interpreting DRS data, allowing for a more intuitive understanding of tissue structures and optical biopsy site locations in relation to the surrounding anatomy.
- **Multimodal integration:** VR can facilitate the integration of multiple imaging modalities such as DRS, medical imaging (e.g., MRI or CT scans) and real-time video feeds. Surgeons can access a comprehensive dataset within a single VR interface, providing a more holistic understanding of the surgical context.

It is important to note that the transition to VR for DRS applications in surgery would require careful consideration of factors such as user interface design, data integration, system reliability and user training. In addition, the development of VR applications for medical use should adhere to strict safety and regulatory standards to ensure patient welfare. Nevertheless, the potential benefits of integrating VR with DRS in surgical procedures are promising and could lead to more accurate and informed decision making in the operating room.

Another exciting frontier in flexible endoscopy is the advent and integration of greppable probes. Designed for enhanced dexterity and precision, these probes can be manoeuvred with greater ease and control. Surgeons can visualise and navigate challenging intraoperative environments with increased accuracy. Combined with real-time tracking capabilities and DRS data, these greppable probes promise to increase the precision and efficiency of endoscopic procedures, making them indispensable tools in modern surgical scenarios.

Bibliography

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394–424 (2018).
2. Ferlay J, Ervik M & Lam F. Global cancer Observatory: cancer today. *International Agency for Research on Cancer* (2018).
3. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
4. Raftopoulos, S. C., Segarajasingam, D. S., Burke, V., Ee, H. C. & Yusoff, I. F. A Cohort Study of Missed and New Cancers After Esophagogastroduodenoscopy. *American Journal of Gastroenterology* **105**, 1292–1297 (2010).
5. Arnold, M., Soerjomataram, I., Ferlay, J. & Forman, D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut* **64**, 381–387 (2015).
6. Colquhoun, A. *et al.* Global patterns of cardia and non-cardia gastric cancer incidence in 2012. *Gut* **64**, 1881–1888 (2015).
7. Salamat, M. S. Robbins and Cotran: Pathologic Basis of Disease, 8th Edition. *J Neuropathol Exp Neurol* **69**, 214–214 (2010).
8. SARBIA, M. Pathology of upper gastrointestinal malignancies. *Semin Oncol* **31**, 465–475 (2004).
9. Khushalani, N. I. Cancer of the Esophagus and Stomach. *Mayo Clin Proc* **83**, 712–722 (2008).
10. Gallo, A. Updates on esophageal and gastric cancers. *World J Gastroenterol* **12**, 3237 (2006).
11. Nagini, S. Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention. *World J Gastrointest Oncol* **4**, 156 (2012).
12. Amin, A. *et al.* Gastric adenocarcinoma missed at endoscopy. *J R Coll Surg Edinb* **47**, 681–4 (2002).
13. Peters, Y. *et al.* Barrett oesophagus. *Nat Rev Dis Primers* **5**, 35 (2019).
14. Wheeler, J. B. & Reed, C. E. Epidemiology of Esophageal Cancer. *Surgical Clinics of North America* **92**, 1077–1087 (2012).
15. Herrera, V. & Parsonnet, J. Helicobacter pylori and gastric adenocarcinoma. *Clinical Microbiology and Infection* **15**, 971–976 (2009).
16. Desai, R. K., Tagliabue, J. R., Wegryn, S. A. & Einstein, D. M. CT evaluation of wall thickening in the alimentary tract. *Radiographics* **11**, 771–83; discussion 784 (1991).
17. Noh, H. M., Fishman, E. K., Forastiere, A. A., Bliss, D. F. & Calhoun, P. S. CT of the esophagus: spectrum of disease with emphasis on esophageal carcinoma. *RadioGraphics* **15**, 1113–1134 (1995).
18. Wakelin, S. J. *et al.* A comparison of computerised tomography, laparoscopic ultrasound and endoscopic ultrasound in the preoperative staging of oesophago-gastric carcinoma. *Eur J Radiol* **41**, 161–167 (2002).
19. Picus, D., Balfe, D. M., Koehler, R. E., Roper, C. L. & Owen, J. W. Computed tomography in the staging of esophageal carcinoma. *Radiology* **146**, 433–8 (1983).

20. Daffner, R., Halber, M., Postlethwait, R., Korobkin, M. & Thompson, W. CT of the esophagus. II. Carcinoma. *American Journal of Roentgenology* **133**, 1051–1055 (1979).
21. Reed, C. E. & Eloubeidi, M. A. New techniques for staging esophageal cancer. *Surgical Clinics of North America* **82**, 697–710 (2002).
22. Hordijk, M. L., Zander, H., van Blankenstein, M. & Tilanus, H. W. Influence of Tumor Stenosis on the Accuracy of Endosonography in Preoperative T Staging of Esophageal Cancer. *Endoscopy* **25**, 171–175 (1993).
23. Napier, K. J. Esophageal cancer: A Review of epidemiology, pathogenesis, staging workup and treatment modalities. *World J Gastrointest Oncol* **6**, 112 (2014).
24. Tio, T. L. *et al.* Endosonography and computed tomography of esophageal carcinoma. Preoperative classification compared to the new (1987) TNM system. *Gastroenterology* **96**, 1478–86 (1989).
25. Rösch, T. Endosonographic staging of esophageal cancer: a review of literature results. *Gastrointest Endosc Clin N Am* **5**, 537–47 (1995).
26. Saunders, H. S., Wolfman, N. T. & Ott, D. J. Esophageal cancer. Radiologic staging. *Radiol Clin North Am* **35**, 281–94 (1997).
27. Misra, S., Choi, M., Livingstone, A. S. & Franceschi, D. The role of endoscopic ultrasound in assessing tumor response and staging after neoadjuvant chemotherapy for esophageal cancer. *Surg Endosc* **26**, 518–522 (2012).
28. Kim, T. J., Kim, H. Y., Lee, K. W. & Kim, M. S. Multimodality Assessment of Esophageal Cancer: Preoperative Staging and Monitoring of Response to Therapy. *RadioGraphics* **29**, 403–421 (2009).
29. Kato, H. *et al.* Comparison between positron emission tomography and computed tomography in the use of the assessment of esophageal carcinoma. *Cancer* **94**, 921–8 (2002).
30. Souquet, J. C. *et al.* Endoscopic ultrasonography in the preoperative staging of esophageal cancer. *Endoscopy* **26**, 764–6 (1994).
31. Lerut, T. *et al.* Histopathologic Validation of Lymph Node Staging With FDG-PET Scan in Cancer of the Esophagus and Gastroesophageal Junction. *Ann Surg* **232**, 743–752 (2000).
32. Quint, L. E., Hepburn, L. M., Francis, I. R., Whyte, R. I. & Orringer, M. B. Incidence and distribution of distant metastases from newly diagnosed esophageal carcinoma. *Cancer* **76**, 1120–1125 (1995).
33. Korst, R. J. *et al.* Proposed Revision Of The Staging Classification For Esophageal Cancer. *J Thorac Cardiovasc Surg* **115**, 660–670 (1998).
34. Kuszyk, B. S. *et al.* Portal-phase contrast-enhanced helical CT for the detection of malignant hepatic tumors: sensitivity based on comparison with intraoperative and pathologic findings. *American Journal of Roentgenology* **166**, 91–95 (1996).
35. Flanagan, F. L. *et al.* Staging of esophageal cancer with 18F-fluorodeoxyglucose positron emission tomography. *American Journal of Roentgenology* **168**, 417–424 (1997).
36. Downey, R. J. *et al.* Whole Body ¹⁸ FDG-PET and the Response of Esophageal Cancer to Induction Therapy: Results of a Prospective Trial. *Journal of Clinical Oncology* **21**, 428–432 (2003).
37. Kelsen, D. P., Bains, M. & Burt, M. Neoadjuvant chemotherapy and surgery of cancer of the esophagus. *Semin Surg Oncol* **6**, 268–273 (1990).

38. Bollschweiler, E. Treatment of early gastric cancer in the Western World. *World J Gastroenterol* **20**, 5672 (2014).
39. Sasako, M. *et al.* Five-Year Outcomes of a Randomized Phase III Trial Comparing Adjuvant Chemotherapy With S-1 Versus Surgery Alone in Stage II or III Gastric Cancer. *Journal of Clinical Oncology* **29**, 4387–4393 (2011).
40. Ajani, J. A. *et al.* Esophageal and Esophagogastric Junction Cancers. *Journal of the National Comprehensive Cancer Network* **9**, 830–887 (2011).
41. D'Amico, T. A. Surgery for esophageal cancer. *Gastrointest Cancer Res* **2**, S6-9 (2008).
42. NCCN Guidelines, https://www.nccn.org/guidelines/category_1.
43. Gastric Cancer. National Comprehensive Cancer Network (NCCN) Guidelines. https://www.nccn.org/professionals/physician_gls/pdf/gastric.pdf. (2020).
44. Jayanathan, M. *et al.* MAGIC versus MacDonald treatment regimens for gastric cancer: Trends and predictors of multimodal therapy for gastric cancer using the National Cancer Database. *The American Journal of Surgery* **219**, 129–135 (2020).
45. Merkow, R. P. *et al.* Use of Multimodality Neoadjuvant Therapy for Esophageal Cancer in the United States: Assessment of 987 Hospitals. *Ann Surg Oncol* **19**, 357–364 (2012).
46. van Hagen, P. *et al.* Preoperative Chemoradiotherapy for Esophageal or Junctional Cancer. *New England Journal of Medicine* **366**, 2074–2084 (2012).
47. Hsu, F.-M. *et al.* Retrospective Analysis of Outcome Differences in Preoperative Concurrent Chemoradiation With or Without Elective Nodal Irradiation for Esophageal Squamous Cell Carcinoma. *International Journal of Radiation Oncology*Biophysics* **81**, e593–e599 (2011).
48. Shapiro, J. *et al.* Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol* **16**, 1090–1098 (2015).
49. Ajani, J. A. *et al.* Esophageal and esophagogastric junction cancers, version 2.2019, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network* **17**, 855–883 (2019).
50. Al-Batran, S.-E. *et al.* Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. *The Lancet* **393**, 1948–1957 (2019).
51. Thomassen, I. *et al.* Peritoneal carcinomatosis of gastric origin: A population-based study on incidence, survival and risk factors. *Int J Cancer* **134**, 622–628 (2014).
52. Yoo, C. H., Noh, S. H., Shin, D. W., Choi, S. H. & Min, J. S. Recurrence following curative resection for gastric carcinoma. *British Journal of Surgery* **87**, 236–242 (2002).
53. Hsu, H.-Y. *et al.* Postoperative Adjuvant Therapy Improves Survival in Pathologic Nonresponders After Neoadjuvant Chemoradiation for Esophageal Squamous Cell Carcinoma: A Propensity-Matched Analysis. *Ann Thorac Surg* **102**, 1687–1693 (2016).
54. Lu, M. *et al.* Nimotuzumab plus paclitaxel and cisplatin as the first line treatment for advanced esophageal squamous cell cancer: A single centre prospective phase <sc>II</sc> trial. *Cancer Sci* **107**, 486–490 (2016).
55. McNamara, M. J. *et al.* The relationship between pathologic nodal disease and residual tumor viability after induction chemotherapy in patients with locally

- advanced esophageal adenocarcinoma receiving a tri-modality regimen. *J Gastrointest Oncol* **7**, 196–205 (2016).
56. Alnaji, R. M. *et al.* Pathologic Complete Response Is an Independent Predictor of Improved Survival Following Neoadjuvant Chemoradiation for Esophageal Adenocarcinoma. *Journal of Gastrointestinal Surgery* **20**, 1541–1546 (2016).
 57. Rawla, P. & Barsouk, A. Epidemiology of gastric cancer: global trends, risk factors and prevention. *Gastroenterology Review* **14**, 26–38 (2019).
 58. Islami, F. *et al.* Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA Cancer J Clin* **68**, 31–54 (2018).
 59. Liu, B. *et al.* Updated Methodology for Projecting U.S.- and State-Level Cancer Counts for the Current Calendar Year: Part I: Spatio-temporal Modeling for Cancer Incidence. *Cancer Epidemiology, Biomarkers & Prevention* **30**, 1620–1626 (2021).
 60. Miller, K. D. *et al.* Updated Methodology for Projecting U.S.- and State-Level Cancer Counts for the Current Calendar Year: Part II: Evaluation of Incidence and Mortality Projection Methods. *Cancer Epidemiology, Biomarkers & Prevention* **30**, 1993–2000 (2021).
 61. Wolf, A. M. D. *et al.* Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin* **68**, 250–281 (2018).
 62. Edwards, B. K. *et al.* Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* **116**, 544–573 (2010).
 63. Dorundi, S. & Bannerjea, A. Colorectal cancer: early diagnosis and predisposing causes. *Surgery (Oxford)* **24**, 131–136 (2006).
 64. J. Ferlay, M. Ervik & F. Lam. “Global Cancer Observatory: Cancer Today.” https://gco.iarc.fr/today/data/factsheets/cancers/10_8_9-Colorectum-fact-sheet.pdf. *Cancer Today* (2018).
 65. Monson, J. R. T. *et al.* Practice Parameters for the Management of Rectal Cancer (Revised). *Dis Colon Rectum* **56**, 535–550 (2013).
 66. Siegel, R. L. *et al.* Colorectal cancer statistics, 2017. *CA Cancer J Clin* **67**, 177–193 (2017).
 67. Douaiher, J. *et al.* Colorectal cancer-global burden, trends, and geographical variations. *J Surg Oncol* **115**, 619–630 (2017).
 68. Hossain, Md. S. *et al.* Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. *Cancers (Basel)* **14**, 1732 (2022).
 69. Testa, U., Pelosi, E. & Castelli, G. Colorectal Cancer: Genetic Abnormalities, Tumor Progression, Tumor Heterogeneity, Clonal Evolution and Tumor-Initiating Cells. *Medical Sciences* **6**, 31 (2018).
 70. Malki, A. *et al.* Molecular Mechanisms of Colon Cancer Progression and Metastasis: Recent Insights and Advancements. *Int J Mol Sci* **22**, 130 (2020).
 71. Nguyen, H. & Duong, H. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (Review). *Oncol Lett* (2018) doi:10.3892/ol.2018.8679.
 72. Vogelstein, B. *et al.* Genetic Alterations during Colorectal-Tumor Development. *New England Journal of Medicine* **319**, 525–532 (1988).

73. Keum, N. & Giovannucci, E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* **16**, 713–732 (2019).
74. Lanza, G., Messerini, L., Gafà, R. & Risio, M. Colorectal tumors: The histology report. *Digestive and Liver Disease* **43**, S344–S355 (2011).
75. Manne, U., Shanmugam, C., Katkoori, V. R., Bumpers, H. L. & Grizzle, W. E. Development and progression of colorectal neoplasia. *Cancer Biomarkers* **9**, 235–265 (2011).
76. NIH. Advances in Colorectal Cancer Research: Stages of Colorectal Cancer. <https://www.nih.gov/researchtraining/advances-colorectal-cancer-research> (2021).
77. Slattery, M. Stage of colon cancer at diagnosis: implications for risk factor associations? *Int J Epidemiol* **27**, 382–387 (1998).
78. Ammerata, G. *et al.* Hyperthermic intraperitoneal chemotherapy and colorectal cancer: From physiology to surgery. *World J Clin Cases* **10**, 10852–10861 (2022).
79. Potter MA. Principles of the surgical management of cancer. in *Principles and Practice of Surgery* 86–97 (2018).
80. Lee, J., Lee, S. & Bae, Y. Multiple Margin Positivity of Frozen Section Is an Independent Risk Factor for Local Recurrence in Breast-Conserving Surgery. *J Breast Cancer* **15**, 420 (2012).
81. Gage, I. *et al.* Pathologic margin involvement and the risk of recurrence in patients treated with breast-conserving therapy. *Cancer* **78**, 1921–8 (1996).
82. Neuzillet, Y. *et al.* Positive surgical margins and their locations in specimens are adverse prognosis features after radical cystectomy in non-metastatic carcinoma invading bladder muscle: results from a nationwide case-control study. *BJU Int* **111**, 1253–1260 (2013).
83. Partridge, M. *et al.* Detection of minimal residual cancer to investigate why oral tumors recur despite seemingly adequate treatment. *Clin Cancer Res* **6**, 2718–25 (2000).
84. Tringale, K. R., Pang, J. & Nguyen, Q. T. Image-guided surgery in cancer: A strategy to reduce incidence of positive surgical margins. *WIREs Systems Biology and Medicine* **10**, (2018).
85. Orosco, R. K. *et al.* Positive Surgical Margins in the 10 Most Common Solid Cancers. *Sci Rep* **8**, 5686 (2018).
86. Grant, Y. *et al.* Patient-level costs in margin re-excision for breast-conserving surgery. *Br J Surg* **106**, 384–394 (2019).
87. Yu, J. *et al.* Cost Analysis of a Surgical Consensus Guideline in Breast-Conserving Surgery. *J Am Coll Surg* **225**, 294–301 (2017).
88. Weledji, E. P. How multimodal treatment improves surgery for oesophageal cancer—a narrative review. *Dig Med Res* **4**, 10–10 (2021).
89. Gravante, G. *et al.* Rectal cancers with microscopic circumferential resection margin involvement (R1 resections): Survivals, patterns of recurrence, and prognostic factors. *J Surg Oncol* **114**, 642–648 (2016).
90. Homan, J. *et al.* Variation in circumferential resection margin: Reporting and involvement in the South-Netherlands. *Eur J Surg Oncol* **41**, 1485–92 (2015).
91. Adam, I. J. *et al.* Role of circumferential margin involvement in the local recurrence of rectal cancer. *Lancet* **344**, 707–11 (1994).
92. Bernstein, T. E., Endreseth, B. H., Romundstad, P., Wibe, A. & Norwegian Colorectal Cancer Registry. What is a safe distal resection margin in rectal cancer patients

- treated by low anterior resection without preoperative radiotherapy? *Colorectal Dis* **14**, e48-55 (2012).
93. Nagtegaal, I. D. *et al.* Circumferential margin involvement is still an important predictor of local recurrence in rectal carcinoma: not one millimeter but two millimeters is the limit. *Am J Surg Pathol* **26**, 350–7 (2002).
 94. Nash, G. M. *et al.* Close distal margin and rectal cancer recurrence after sphincter-preserving rectal resection. *Dis Colon Rectum* **53**, 1365–73 (2010).
 95. Vernava, A. M., Moran, M., Rothenberger, D. A. & Wong, W. D. A prospective evaluation of distal margins in carcinoma of the rectum. *Surg Gynecol Obstet* **175**, 333–6 (1992).
 96. Marijnen, C. A. M. *et al.* Radiotherapy does not compensate for positive resection margins in rectal cancer patients: report of a multicenter randomized trial. *Int J Radiat Oncol Biol Phys* **55**, 1311–20 (2003).
 97. Mukkai Krishnamurty, D. & Wise, P. E. Importance of surgical margins in rectal cancer. *J Surg Oncol* **113**, 323–332 (2016).
 98. Tramontano, A. C. *et al.* Esophageal cancer treatment costs by phase of care and treatment modality, 2000-2013. *Cancer Med* **8**, 5158–5172 (2019).
 99. Carvalho, C. & Glynne-Jones, R. Challenges behind proving efficacy of adjuvant chemotherapy after preoperative chemoradiation for rectal cancer. *Lancet Oncol* **18**, e354–e363 (2017).
 100. Hanly, P., Céilleachair, A. Ó., Skally, M., O'Neill, C. & Sharp, L. Direct costs of radiotherapy for rectal cancer: a microcosting study. *BMC Health Serv Res* **15**, 184 (2015).
 101. Gehrman, J., Angenete, E., Björholt, I., Lesén, E. & Haglind, E. Cost-effectiveness analysis of laparoscopic and open surgery in routine Swedish care for colorectal cancer. *Surg Endosc* **34**, 4403–4412 (2020).
 102. Raab, S. S. *et al.* Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer* **104**, 2205–13 (2005).
 103. Frierson, H. F. *et al.* Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *Am J Clin Pathol* **103**, 195–8 (1995).
 104. Robbins, P. *et al.* Histological grading of breast carcinomas: a study of interobserver agreement. *Hum Pathol* **26**, 873–9 (1995).
 105. Elmore, J. G. *et al.* Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA* **313**, 1122–32 (2015).
 106. Elmore, J. G. *et al.* Variability in Pathologists' Interpretations of Individual Breast Biopsy Slides: A Population Perspective. *Ann Intern Med* **164**, 649 (2016).
 107. Pleijhuis, R. G. *et al.* Obtaining adequate surgical margins in breast-conserving therapy for patients with early-stage breast cancer: current modalities and future directions. *Ann Surg Oncol* **16**, 2717–30 (2009).
 108. Maloney, B. W. *et al.* Review of methods for intraoperative margin detection for breast conserving surgery. *J Biomed Opt* **23**, 1–19 (2018).
 109. St John, E. R. *et al.* Diagnostic Accuracy of Intraoperative Techniques for Margin Assessment in Breast Cancer Surgery: A Meta-analysis. *Ann Surg* **265**, 300–310 (2017).
 110. Butler-Henderson, K., Lee, A. H., Price, R. I. & Waring, K. Intraoperative assessment of margins in breast conserving therapy: a systematic review. *Breast* **23**, 112–9 (2014).

111. Shipp, D. W. *et al.* Intra-operative spectroscopic assessment of surgical margins during breast conserving surgery. *Breast Cancer Research* **20**, 69 (2018).
112. Esbona, K., Li, Z. & Wilke, L. G. Intraoperative Imprint Cytology and Frozen Section Pathology for Margin Assessment in Breast Conservation Surgery: A Systematic Review. *Ann Surg Oncol* **19**, 3236–3245 (2012).
113. Versteegden, D. P. A. *et al.* Performance characteristics of specimen radiography for margin assessment for ductal carcinoma in situ: a systematic review. *Breast Cancer Res Treat* **166**, 669–679 (2017).
114. Janssen, N. N. Y. *et al.* Feasibility of Micro-Computed Tomography Imaging for Direct Assessment of Surgical Resection Margins During Breast-Conserving Surgery. *Journal of Surgical Research* **241**, 160–169 (2019).
115. Marcus, D. *et al.* Point-of-Care Diagnosis of Endometrial Cancer Using the Surgical Intelligent Knife (iKnife)—A Prospective Pilot Study of Diagnostic Accuracy. *Cancers (Basel)* **14**, 5892 (2022).
116. Tzafetas, M. *et al.* The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease. *Proceedings of the National Academy of Sciences* **117**, 7338–7346 (2020).
117. St John, E. R. *et al.* Diagnostic Accuracy of Intraoperative Techniques for Margin Assessment in Breast Cancer Surgery. *Ann Surg* **265**, 300–310 (2017).
118. Pleijhuis, R. G. *et al.* Obtaining Adequate Surgical Margins in Breast-Conserving Therapy for Patients with Early-Stage Breast Cancer: Current Modalities and Future Directions. *Ann Surg Oncol* **16**, 2717–2730 (2009).
119. Chan, B. K., Wiseberg-Firtell, J. A., Jois, R. H., Jensen, K. & Audisio, R. A. Localization techniques for guided surgical excision of non-palpable breast lesions. *Cochrane Database of Systematic Reviews* **2016**, (2015).
120. St John, E. R. *et al.* Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. *Breast Cancer Research* **19**, 59 (2017).
121. Phelps, D. L. *et al.* The surgical intelligent knife distinguishes normal, borderline and malignant gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS). *Br J Cancer* **118**, 1349–1358 (2018).
122. Vaysse, P.-M. *et al.* Evaluation of the Sensitivity of Metabolic Profiling by Rapid Evaporative Ionization Mass Spectrometry: Toward More Radical Oral Cavity Cancer Resections. *Anal Chem* **94**, 6939–6947 (2022).
123. Paizs, P. *et al.* Abstract 4536: Rapid evaporative ionization mass spectrometry (REIMS) analysis of the mucosal lipidome has a high diagnostic accuracy for adenomas and early colorectal cancer. *Cancer Res* **78**, 4536–4536 (2018).
124. Wang, H. *et al.* Lipidomic study and diagnosis of hepatocellular carcinoma tumor with rapid evaporative ionization mass spectrometry. *Electrophoresis* **44**, 1057–1067 (2023).
125. Mesurolle, B. *et al.* Sonography of Postexcision Specimens of Nonpalpable Breast Lesions: Value, Limitations, and Description of a Method. *American Journal of Roentgenology* **186**, 1014–1024 (2006).
126. Londero, V. *et al.* Surgical specimen ultrasound: Is it able to predict the status of resection margins after breast-conserving surgery? *The Breast* **19**, 532–537 (2010).
127. Treanor, D. & Quirke, P. Pathology of Colorectal Cancer. *Clin Oncol* **19**, 769–776 (2007).

128. Peravali, R. & Hall, N. Colorectal cancer: features and investigation. *Medicine* **43**, 299–302 (2015).
129. Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* **66**, 271–289 (2016).
130. Voutilainen, M. E. & Juhola, M. T. Evaluation of the diagnostic accuracy of gastroscopy to detect gastric tumours: clinicopathological features and prognosis of patients with gastric cancer missed on endoscopy. *Eur J Gastroenterol Hepatol* **17**, 1345–1349 (2005).
131. Raftopoulos, S. C., Segarajasingam, D. S., Burke, V., Ee, H. C. & Yusoff, I. F. A Cohort Study of Missed and New Cancers After Esophagogastroduodenoscopy. *American Journal of Gastroenterology* **105**, 1292–1297 (2010).
132. Todd, J. A., Richards, C. J., Dixon, A. & Robinson, R. J. Gastric ulcer and malignancy - is there a need for follow-up endoscopy? *Aliment Pharmacol Ther* **19**, 989–991 (2004).
133. Bernards, N. *et al.* No improvement in median survival for patients with metastatic gastric cancer despite increased use of chemotherapy. *Annals of Oncology* **24**, 3056–3060 (2013).
134. Thomassen, I. *et al.* Peritoneal carcinomatosis of gastric origin: A population-based study on incidence, survival and risk factors. *Int J Cancer* **134**, 622–628 (2014).
135. Flock, S. T., Wilson, B. C. & Patterson, M. S. Total attenuation coefficients and scattering phase functions of tissues and phantom materials at 633 nm. *Med Phys* **14**, 835–841 (1987).
136. Bargo, P. R. *et al.* In vivo determination of optical properties of normal and tumor tissue with white light reflectance and an empirical light transport model during endoscopy. *J Biomed Opt* **10**, 034018 (2005).
137. Bashkatov, A. N. *et al.* Optical properties of human stomach mucosa in the spectral range from 400 to 2000 nm. in (ed. Tuchin, V. V.) 653513-653513–11 (2007). doi:10.1117/12.740966.
138. Wei, H.-J. Determination of optical properties of normal and adenomatous human colon tissues *in vitro* using integrating sphere techniques. *World J Gastroenterol* **11**, 2413 (2005).
139. Drezek, R. *et al.* Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, DNA content, and chromatin texture. *J Biomed Opt* **8**, 7 (2003).
140. Drezek, R., Dunn, A. & Richards-Kortum, R. Light scattering from cells: finite-difference time-domain simulations and goniometric measurements. *Appl Opt* **38**, 3651 (1999).
141. Salomatina, E., Jiang, B., Novak, J. & Yaroslavsky, A. N. Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range. *J Biomed Opt* **11**, 064026 (2006).
142. Thueler, P. *et al.* In vivo endoscopic tissue diagnostics based on spectroscopic absorption, scattering, and phase function properties. *J Biomed Opt* **8**, 495 (2003).
143. Charvet, I. *et al.* Diagnosis and grading of gastritis by non-invasive optical analysis. *Eur J Gastroenterol Hepatol* **16**, 1189–1198 (2004).
144. Pitris, C. *et al.* Feasibility of optical coherence tomography for high-resolution imaging of human gastrointestinal tract malignancies. *J Gastroenterol* **35**, 87–92 (2000).

145. DACOSTA, R. S., WILSON, B. C. & MARCON, N. E. New optical technologies for earlier endoscopic diagnosis of premalignant gastrointestinal lesions. *J Gastroenterol Hepatol* **17**, S85–S104 (2002).
146. Abe, S. E. *et al.* Margin re-excision and local recurrence in invasive breast cancer: A cost analysis using a decision tree model. *J Surg Oncol* **112**, 443–448 (2015).
147. Georgakoudi, I. & Van Dam, J. Characterization of dysplastic tissue morphology and biochemistry in Barrett's esophagus using diffuse reflectance and light scattering spectroscopy. *Gastrointest Endosc Clin N Am* **13**, 297–308 (2003).
148. Pocock, G. & Richards, C. D. *The Human Body : An Introduction for the Biomedical and Health Sciences*. (Oxford University Press, 2009).
149. Cope, M. *et al.* Methods of Quantitating Cerebral Near Infrared Spectroscopy Data. in 183–189 (1988). doi:10.1007/978-1-4615-9510-6_21.
150. Regensburger, A. P., Brown, E., Krönke, G., Waldner, M. J. & Knieling, F. Optoacoustic Imaging in Inflammation. *Biomedicines* **9**, 483 (2021).
151. Hale, G. M. & Querry, M. R. Optical Constants of Water in the 200-nm to 200-μm Wavelength Region. *Appl Opt* **12**, 555 (1973).
152. Brown, J. Q. *et al.* Quantitative Optical Spectroscopy: A Robust Tool for Direct Measurement of Breast Cancer Vascular Oxygenation and Total Hemoglobin Content *In vivo*. *Cancer Res* **69**, 2919–2926 (2009).
153. Volynskaya, Z. *et al.* Diagnosing breast cancer using diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy. *J Biomed Opt* **13**, 024012 (2008).
154. Chang, V. T.-C. *et al.* Quantitative Physiology of the Precancerous Cervix In Vivo through Optical Spectroscopy. *Neoplasia* **11**, 325–332 (2009).
155. Reif, R. *et al.* Analysis of changes in reflectance measurements on biological tissues subjected to different probe pressures. *J Biomed Opt* **13**, 010502 (2008).
156. Schwarz, R. A. *et al.* Noninvasive evaluation of oral lesions using depth-sensitive optical spectroscopy. *Cancer* **115**, 1669–1679 (2009).
157. Zonios, G. *et al.* Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. *Appl Opt* **38**, 6628 (1999).
158. Zhu, C., Palmer, G. M., Breslin, T. M., Harter, J. & Ramanujam, N. Diagnosis of breast cancer using diffuse reflectance spectroscopy: Comparison of a Monte Carlo versus partial least squares analysis based feature extraction technique. *Lasers Surg Med* **38**, 714–724 (2006).
159. Zhu, C., Palmer, G. M., Breslin, T. M., Xu, F. & Ramanujam, N. Use of a multiseparation fiber optic probe for the optical diagnosis of breast cancer. *J Biomed Opt* **10**, 024032 (2005).
160. Bigio, I. J. *et al.* Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results. *J Biomed Opt* **5**, 221 (2000).
161. Tromberg, B. J. *et al.* Non-Invasive In Vivo Characterization of Breast Tumors Using Photon Migration Spectroscopy. *Neoplasia* **2**, 26–40 (2000).
162. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (Springer US, Boston, MA, 1999). doi:10.1007/978-1-4757-3061-6.
163. Wang, L. V., Wu, H.-I. & Masters, B. R. Biomedical Optics, Principles and Imaging. *J Biomed Opt* **13**, 049902 (2008).
164. Ramanujam, N. Fluorescence Spectroscopy of Neoplastic and Non-Neoplastic Tissues. *Neoplasia* **2**, 89–117 (2000).

165. Robert W. Redmond. *Introduction to Fluorescence and Photophysics*. (CRC Press, 2003).
166. Brown, S. Biological Spectroscopy. *Biochem Educ* **13**, 93 (1985).
167. Sharma, P. K., Nass, R. & Schmidt, H. Effect of solvent, host precursor, dopant concentration and crystallite size on the fluorescence properties of Eu(III) doped yttria. *Opt Mater (Amst)* **10**, 161–169 (1998).
168. Mary-Ann Mycek & Brian W. Pogue. *Handbook of Biomedical Fluorescence*. (CRC Press, 2003). doi:10.1201/9780203912096.
169. Liu, K., Zhao, Q., Li, B. & Zhao, X. Raman Spectroscopy: A Novel Technology for Gastric Cancer Diagnosis. *Front Bioeng Biotechnol* **10**, (2022).
170. Shipp, D. W., Sinjab, F. & Notingher, I. Raman spectroscopy: techniques and applications in the life sciences. *Adv Opt Photonics* **9**, 315 (2017).
171. Kouri, M. A. *et al.* Raman Spectroscopy: A Personalized Decision-Making Tool on Clinicians' Hands for In Situ Cancer Diagnosis and Surgery Guidance. *Cancers (Basel)* **14**, 1144 (2022).
172. Utzinger, U. & Richards-Kortum, R. R. Fiber optic probes for biomedical optical spectroscopy. *J Biomed Opt* **8**, 121 (2003).
173. Roy, H. K. & Backman, V. Spectroscopic Applications in Gastrointestinal Endoscopy. *Clinical Gastroenterology and Hepatology* **10**, 1335–1341 (2012).
174. Bargo, P. R., Prahl, S. A. & Jacques, S. L. Optical properties effects upon the collection efficiency of optical fibers in different probe configurations. *IEEE Journal of Selected Topics in Quantum Electronics* **9**, 314–321 (2003).
175. Zhu, C., Liu, Q. & Ramanujam, N. Effect of fiber optic probe geometry on depth-resolved fluorescence measurements from epithelial tissues: a Monte Carlo simulation. *J Biomed Opt* **8**, 237 (2003).
176. Chen, Y. *et al.* Raman Spectroscopy Analysis of the Biochemical Characteristics of Molecules Associated with the Malignant Transformation of Gastric Mucosa. *PLoS One* **9**, e93906 (2014).
177. Dukor, R. K. Vibrational Spectroscopy in the Detection of Cancer. in *Handbook of Vibrational Spectroscopy* (ed. Griffiths, P. R.) (Wiley, 2001). doi:10.1002/0470027320.s8107.
178. Evers, D., Hendriks, B., Lucassen, G. & Ruers, T. Optical spectroscopy: current advances and future applications in cancer diagnostics and therapy. *Future Oncology* **8**, 307–320 (2012).
179. Hendriks, B. H. W. *et al.* Nerve detection with optical spectroscopy for regional anesthesia procedures. *J Transl Med* **13**, 380 (2015).
180. Peter Lasch & Janina Kneipp. *Biomedical Vibrational Spectroscopy*. (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008). doi:10.1002/9780470283172.
181. Green, B., Cobb, A. R. M., Brennan, P. A. & Hopper, C. Optical diagnostic techniques for use in lesions of the head and neck: review of the latest developments. *British Journal of Oral and Maxillofacial Surgery* **52**, 675–680 (2014).
182. Shim, M. G., Wong Kee Song, L.-M., Marcon, N. E. & Wilson, B. C. In vivo Near-infrared Raman Spectroscopy: Demonstration of Feasibility During Clinical Gastrointestinal Endoscopy. *Photochem Photobiol* **72**, 146 (2000).
183. Rodriguez-Diaz, E., Bigio, I. J. & Singh, S. K. Integrated optical tools for minimally invasive diagnosis and treatment at gastrointestinal endoscopy. *Robot Comput Integr Manuf* **27**, 249–256 (2011).

184. Lewis, P. D. *et al.* Evaluation of FTIR Spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC Cancer* **10**, 640 (2010).
185. Pennathur, A., Gibson, M. K., Jobe, B. A. & Luketich, J. D. Oesophageal carcinoma. *The Lancet* **381**, 400–412 (2013).
186. Ishigaki, M. *et al.* Diagnosis of early-stage esophageal cancer by Raman spectroscopy and chemometric techniques. *Analyst* **141**, 1027–1033 (2016).
187. Shaheen, N. J. Upper Endoscopy for Gastroesophageal Reflux Disease: Best Practice Advice From the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med* **157**, 808 (2012).
188. Kendall, C. *et al.* Evaluation of Raman probe for oesophageal cancer diagnostics. *Analyst* **135**, 3038 (2010).
189. Wani, S. *et al.* Patients With Nondysplastic Barrett’s Esophagus Have Low Risks for Developing Dysplasia or Esophageal Adenocarcinoma. *Clinical Gastroenterology and Hepatology* **9**, 220-227.e1 (2011).
190. Hvid-Jensen, F., Pedersen, L., Drewes, A. M., Sørensen, H. T. & Funch-Jensen, P. Incidence of Adenocarcinoma among Patients with Barrett’s Esophagus. *New England Journal of Medicine* **365**, 1375–1383 (2011).
191. Almond, L. M. & Barr, H. Advanced endoscopic imaging in Barrett’s oesophagus. *International Journal of Surgery* **10**, 236–241 (2012).
192. Almond, L. M. *et al.* Endoscopic Raman spectroscopy enables objective diagnosis of dysplasia in Barrett’s esophagus. *Gastrointest Endosc* **79**, 37–45 (2014).
193. Georgakoudi, I. *et al.* Fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in patients with Barrett’s esophagus. *Gastroenterology* **120**, 1620–1629 (2001).
194. Lovat, L. B. Elastic scattering spectroscopy accurately detects high grade dysplasia and cancer in Barrett’s oesophagus. *Gut* **55**, 1078–1083 (2005).
195. Teh, S. K. *et al.* Diagnostic potential of near-infrared Raman spectroscopy in the stomach: differentiating dysplasia from normal tissue. *Br J Cancer* **98**, 457–465 (2008).
196. Teh, S. K. *et al.* Near-infrared Raman spectroscopy for early diagnosis and typing of adenocarcinoma in the stomach. *British Journal of Surgery* **97**, 550–557 (2010).
197. Dhar, A. *et al.* Elastic scattering spectroscopy for the diagnosis of colonic lesions: initial results of a novel optical biopsy technique. *Gastrointest Endosc* **63**, 257–261 (2006).
198. Wang, H.-W. *et al.* Diffuse reflectance spectroscopy detects increased hemoglobin concentration and decreased oxygenation during colon carcinogenesis from normal to malignant tumors. *Opt Express* **17**, 2805 (2009).
199. Chowdary, M. V. P. *et al.* Discrimination of Normal and Malignant Mucosal Tissues of the Colon by Raman Spectroscopy. *Photomed Laser Surg* **25**, 269–274 (2007).
200. Li, L. Q. *et al.* Oxygenated hemoglobin diffuse reflectance ratio for in vitro detection of human gastric pre-cancer. *Laser Phys* **20**, 1667–1672 (2010).
201. Bergholt, M. S. *et al.* Raman endoscopy for in vivo differentiation between benign and malignant ulcers in the stomach. *Analyst* **135**, 3162 (2010).
202. Rodriguez-Diaz, E. *et al.* Endoscopic histological assessment of colonic polyps by using elastic scattering spectroscopy. *Gastrointest Endosc* **81**, 539–547 (2015).
203. Kuiper, T. *et al.* Automatic optical diagnosis of small colorectal lesions by laser-induced autofluorescence. *Endoscopy* **47**, 56–62 (2014).

204. Rath, T. *et al.* In vivo real-time assessment of colorectal polyp histology using an optical biopsy forceps system based on laser-induced fluorescence spectroscopy. *Endoscopy* **48**, 557–562 (2016).
205. Lim, L., Bajbouj, M., von Delius, S. & Meining, A. Fluorescein-enhanced autofluorescence imaging for accurate differentiation of neoplastic from non-neoplastic colorectal polyps: a feasibility study. *Endoscopy* **43**, 419–424 (2011).
206. Alam, I. S. *et al.* Emerging Intraoperative Imaging Modalities to Improve Surgical Precision. *Mol Imaging Biol* **20**, 705–715 (2018).
207. Nema, S. & Vachhani, L. Surgical instrument detection and tracking technologies: Automating dataset labeling for surgical skill assessment. *Front Robot AI* **9**, (2022).
208. Qiu, L., Li, C. & Ren, H. Real-time surgical instrument tracking in robot-assisted surgery using multi-domain convolutional neural network. *Healthc Technol Lett* **6**, 159–164 (2019).
209. Chmarra, M. K., Grimbergen, C. A. & Dankelman, J. Systems for tracking minimally invasive surgical instruments. *Minimally Invasive Therapy & Allied Technologies* **16**, 328–340 (2007).
210. Shaeffer, D. K. MEMS inertial sensors: A tutorial overview. *IEEE Communications Magazine* **51**, 100–109 (2013).
211. Ren, H. & Kazanzides, P. Investigation of Attitude Tracking Using an Integrated Inertial and Magnetic Navigation System for Hand-Held Surgical Instruments. *IEEE/ASME Transactions on Mechatronics* **17**, 210–217 (2012).
212. Geist, E. & Shimada, K. Position error reduction in a mechanical tracking linkage for arthroscopic hip surgery. *Int J Comput Assist Radiol Surg* **6**, 693–698 (2011).
213. Gomes, M. P. S. F., Kumar, P. V. S., Timoney, A. G. & Davies, B. L. Tool tracking for endoscopic surgery. *Transactions of the Institute of Measurement and Control* **25**, 281–292 (2003).
214. David Forsyth & Jean Ponce. *Computer Vision: A Modern Approach*. (Prentice Hall Professional Technical Reference, 2002).
215. Yilmaz, A., Javed, O. & Shah, M. Object tracking. *ACM Comput Surv* **38**, 13 (2006).
216. Sznitman, R. *et al.* Data-Driven Visual Tracking in Retinal Microsurgery. in 568–575 (2012). doi:10.1007/978-3-642-33418-4_70.
217. Rieke, N. *et al.* Real-time localization of articulated surgical instruments in retinal microsurgery. *Med Image Anal* **34**, 82–100 (2016).
218. Rutala, W. A. & Weber, D. J. Disinfection and Sterilization in Health Care Facilities: An Overview and Current Issues. *Infect Dis Clin North Am* **30**, 609–37 (2016).
219. Bouget, D., Allan, M., Stoyanov, D. & Jannin, P. Vision-based and marker-less surgical tool detection and tracking: a review of the literature. *Med Image Anal* **35**, 633–654 (2017).
220. Mountney, P., Giannarou, S., Elson, D. & Yang, G.-Z. Optical Biopsy Mapping for Minimally Invasive Cancer Screening. in 483–490 (2009). doi:10.1007/978-3-642-04268-3_60.
221. Huang, B. *et al.* Tracking and visualization of the sensing area for a tethered laparoscopic gamma probe. *Int J Comput Assist Radiol Surg* **15**, 1389–1397 (2020).
222. Bouarfa, L., Akman, O., Schneider, A., Jonker, P. P. & Dankelman, J. In-vivo real-time tracking of surgical instruments in endoscopic video. *Minimally Invasive Therapy & Allied Technologies* **21**, 129–134 (2012).

223. Horgan, C. C. *et al.* Image-guided Raman spectroscopy probe-tracking for tumor margin delineation. *J Biomed Opt* **26**, (2021).
224. Zhang, L., Ye, M., Chan, P.-L. & Yang, G.-Z. Real-time surgical tool tracking and pose estimation using a hybrid cylindrical marker. *Int J Comput Assist Radiol Surg* **12**, 921–930 (2017).
225. Jayarathne, U. L., Chen, E. C. S., Moore, J. & Peters, T. M. Robust, Intrinsic Tracking of a Laparoscopic Ultrasound Probe for Ultrasound-Augmented Laparoscopy. *IEEE Trans Med Imaging* **38**, 460–469 (2019).
226. Hulke, U. & Gupta, A. Single camera based motion tracking for minimally invasive surgery. in *22nd Mediterranean Conference on Control and Automation* 356–361 (IEEE, 2014). doi:10.1109/MED.2014.6961397.
227. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* **22**, 1330–1334 (2000).
228. Zhang, X. & Payandeh, S. Application of visual tracking for robot-assisted laparoscopic surgery. *J Robot Syst* **19**, 315–328 (2002).
229. Cartucho, J. *et al.* An enhanced marker pattern that achieves improved accuracy in surgical tool tracking. *Comput Methods Biomech Biomed Eng Imaging Vis* **10**, 400–408 (2022).
230. Tonet, O., Thoranaghatte, R. U., Megali, G. & Dario, P. Tracking endoscopic instruments without a localizer: A shape-analysis-based approach. *Computer Aided Surgery* **12**, 35–42 (2007).
231. Guo-Qing Wei, Arbter, K. & Hirzinger, G. Real-time visual servoing for laparoscopic surgery. Controlling robot motion with color image segmentation. *IEEE Engineering in Medicine and Biology Magazine* **16**, 40–45 (1997).
232. Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. in *Proceedings of the Seventh IEEE International Conference on Computer Vision* 666–673 vol.1 (IEEE, 1999). doi:10.1109/ICCV.1999.791289.
233. Mamone, V. *et al.* Robust Laparoscopic Instruments Tracking Using Colored Strips. in 129–143 (2017). doi:10.1007/978-3-319-60928-7_11.
234. Principal Component Analysis.
235. Welch, G. & Bishop, G. Welch & Bishop , An Introduction to the Kalman Filter 2 1 The Discrete Kalman Filter In 1960. in (1994).
236. Jo, K., Choi, Y., Choi, J. & Chung, J. W. Robust Real-Time Detection of Laparoscopic Instruments in Robot Surgery Using Convolutional Neural Networks with Motion Vector Prediction. *Applied Sciences* **9**, 2865 (2019).
237. Zhao, Z., Cai, T., Chang, F. & Cheng, X. Real-time surgical instrument detection in robot-assisted surgery using a convolutional neural network cascade. *Healthc Technol Lett* **6**, 275–279 (2019).
238. Bouget, D., Allan, M., Stoyanov, D. & Jannin, P. Vision-based and marker-less surgical tool detection and tracking: a review of the literature. *Med Image Anal* **35**, 633–654 (2017).
239. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. (2013).
240. Girshick, R. Fast R-CNN. (2015).
241. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell* **39**, 1137–1149 (2017).

242. He, K., Zhang, X., Ren, S. & Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans Pattern Anal Mach Intell* **37**, 1904–1916 (2015).
243. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. (2015).
244. Liu, W. *et al.* SSD: Single Shot MultiBox Detector. in 21–37 (2016). doi:10.1007/978-3-319-46448-0_2.
245. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. Focal Loss for Dense Object Detection. (2017).
246. Tian, Z., Shen, C., Chen, H. & He, T. FCOS: Fully Convolutional One-Stage Object Detection. (2019).
247. Dong, X., Yan, S. & Duan, C. A lightweight vehicles detection network model based on YOLOv5. *Eng Appl Artif Intell* **113**, 104914 (2022).
248. Nikkath Bushra, S., Shobana, G., Uma Maheswari, K. & Subramanian, N. Smart Video Surveillance Based Weapon Identification Using Yolov5. in *2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC)* 351–357 (IEEE, 2022). doi:10.1109/ICESIC53714.2022.9783499.
249. Zhaoxin, G., Han, L., Zhijiang, Z. & Libo, P. Design a Robot System for Tomato Picking Based on YOLO v5. *IFAC-PapersOnLine* **55**, 166–171 (2022).
250. Mahaur, B. & Mishra, K. K. Small-object detection based on YOLOv5 in autonomous driving systems. *Pattern Recognit Lett* **168**, 115–122 (2023).
251. Qi, D., Tan, W., Yao, Q. & Liu, J. YOLO5Face: Why Reinventing a Face Detector. (2021).
252. Yican Zhong. Image-based Optical Probe Detection and Tracking Using Deep Learning. (Imperial College London, 2022).
253. Feng, Z., Kittler, J., Awais, M., Huber, P. & Wu, X. Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2235–2245 (2017).
254. Wang, R. J., Li, X. & Ling, C. X. Pelee: A Real-Time Object Detection System on Mobile Devices. (2018).
255. Elfving, S., Uchibe, E. & Doya, K. Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning. (2017).
256. Gkouzionis, I. *et al.* A YOLOv5-based network for the detection of a diffuse reflectance spectroscopy probe to aid surgical guidance in gastrointestinal cancer surgery. *Int J Comput Assist Radiol Surg* (2023) doi:10.1007/s11548-023-02944-9.
257. Yun, S. *et al.* CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. (2019).
258. Zhang, H., Cisse, M., Dauphin, Y. N. & Lopez-Paz, D. mixup: Beyond Empirical Risk Minimization. (2017).
259. Spliethoff, J. W. *et al.* Monitoring of tumor radio frequency ablation using derivative spectroscopy. *J Biomed Opt* **19**, 097004 (2014).
260. Baltussen, E. J. M. *et al.* Diffuse reflectance spectroscopy as a tool for real-time tissue assessment during colorectal cancer surgery. *J Biomed Opt* **22**, 1 (2017).
261. de Boer, L. L. *et al.* Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries. *Breast Cancer Res Treat* **152**, 509–518 (2015).
262. Spliethoff, J. W. *et al.* Improved identification of peripheral lung tumors by using diffuse reflectance and fluorescence spectroscopy. *Lung Cancer* **80**, 165–171 (2013).

263. Georgakoudi, I. & Feld, M. S. The combined use of fluorescence, reflectance, and light-scattering spectroscopy for evaluating dysplasia in Barrett's esophagus. *Gastrointest Endosc Clin N Am* **14**, 519–537 (2004).
264. Panjehpour, M. *et al.* Spectroscopic diagnosis of esophageal cancer: New classification model, improved measurement system. *Gastrointest Endosc* **41**, 577–581 (1995).
265. Baltussen, E. J. M. *et al.* Diffuse reflectance spectroscopy as a tool for real-time tissue assessment during colorectal cancer surgery. *J Biomed Opt* **22**, 1 (2017).
266. Baltussen, E. J. M. *et al.* Tissue diagnosis during colorectal cancer surgery using optical sensing: an in vivo study. *J Transl Med* **17**, 333 (2019).
267. Langhout, G. C. *et al.* Differentiation of healthy and malignant tissue in colon cancer patients using optical spectroscopy: A tool for image-guided surgery. *Lasers Surg Med* **47**, 559–565 (2015).
268. Schols, R. M., Dunias, P., Wieringa, F. P. & Stassen, L. P. S. Multispectral characterization of tissues encountered during laparoscopic colorectal surgery. *Med Eng Phys* **35**, 1044–1050 (2013).
269. Zonios, G. *et al.* Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. *Appl Opt* **38**, 6628 (1999).
270. Chen, H. *et al.* Diagnosis of colorectal cancer by near-infrared optical fiber spectroscopy and random forest. *Spectrochim Acta A Mol Biomol Spectrosc* **135**, 185–191 (2015).
271. Hidović-Rowe, D. & Claridge, E. Modelling and validation of spectral reflectance for the colon. *Phys Med Biol* **50**, 1071–1093 (2005).
272. Wells, W. A. *et al.* Validation of novel optical imaging technologies: the pathologists' view. *J Biomed Opt* **12**, 051801 (2007).
273. Jerjes, W. K. *et al.* The future of medical diagnostics: review paper. *Head Neck Oncol* **3**, 38 (2011).
274. Gurcan, M. N. *et al.* Histopathological Image Analysis: A Review. *IEEE Rev Biomed Eng* **2**, 147–171 (2009).
275. Krishnaswamy, V., Laughney, A. M., Wells, W. A., Paulsen, K. D. & Pogue, B. W. Scanning in situ Spectroscopy platform for imaging surgical breast tissue specimens. *Opt Express* **21**, 2185 (2013).
276. Pilling, M. J. *et al.* Infrared spectral histopathology using haematoxylin and eosin (H&E) stained glass slides: a major step forward towards clinical translation. *Analyst* **142**, 1258–1268 (2017).
277. Wilson, B. C., Jermyn, M. & Leblond, F. Challenges and opportunities in clinical translation of biomedical optical spectroscopy and imaging. *J Biomed Opt* **23**, 1 (2018).
278. Lu, G. *et al.* Framework for hyperspectral image processing and quantification for cancer detection during animal tumor surgery. *J Biomed Opt* **20**, 126012 (2015).
279. Naranjo, V. *et al.* Stained and infrared image registration as first step for cancer detection. in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* 420–423 (IEEE, 2014). doi:10.1109/BHI.2014.6864392.
280. Halicek, M. *et al.* Deformable registration of histological cancer margins to gross hyperspectral images using demons. in *Medical Imaging 2018: Digital Pathology* (eds. Gurcan, M. N. & Tomaszewski, J. E.) 22 (SPIE, 2018). doi:10.1117/12.2293165.

281. Lu, G., Halig, L., Wang, D., Chen, Z. G. & Fei, B. Hyperspectral imaging for cancer surgical margin delineation: registration of hyperspectral and histological images. in (eds. Yaniv, Z. R. & Holmes, D. R.) 90360S (2014). doi:10.1117/12.2043805.
282. Phipps, J. E. *et al.* Automated detection of breast cancer in resected specimens with fluorescence lifetime imaging. *Phys Med Biol* **63**, 015003 (2017).
283. Laughney, A. M. *et al.* Automated classification of breast pathology using local measures of broadband reflectance. *J Biomed Opt* **15**, 066019 (2010).
284. Laughney, A. M. *et al.* Scatter Spectroscopic Imaging Distinguishes between Breast Pathologies in Tissues Relevant to Surgical Margin Assessment. *Clinical Cancer Research* **18**, 6315–6325 (2012).
285. Hsiung, P.-L., Nambiar, P. R. & Fujimoto, J. G. Effect of tissue preservation on imaging using ultrahigh resolution optical coherence tomography. *J Biomed Opt* **10**, 064033 (2005).
286. Lu, G. & Fei, B. Medical hyperspectral imaging: a review. *J Biomed Opt* **19**, 010901 (2014).
287. Savitzky, Abraham. & Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* **36**, 1627–1639 (1964).
288. Nachabé, R. *et al.* Real-Time In Vivo Characterization of Primary Liver Tumors With Diffuse Optical Spectroscopy During Percutaneous Needle Interventions. *Invest Radiol* **50**, 443–448 (2015).
289. de Boer, L. L. *et al.* Using DRS during breast conserving surgery: identifying robust optical parameters and influence of inter-patient variation. *Biomed Opt Express* **7**, 5188 (2016).
290. Tanis, E. *et al.* Real-time in vivo assessment of radiofrequency ablation of human colorectal liver metastases using diffuse reflectance spectroscopy. *European Journal of Surgical Oncology (EJSO)* **42**, 251–259 (2016).
291. Spliethoff, J. W. *et al.* In vivo characterization of colorectal metastases in human liver using diffuse reflectance spectroscopy: toward guidance in oncological procedures. *J Biomed Opt* **21**, 097004 (2016).
292. Spliethoff, J. W. *et al.* Real-time In Vivo Tissue Characterization with Diffuse Reflectance Spectroscopy during Transthoracic Lung Biopsy: A Clinical Feasibility Study. *Clinical Cancer Research* **22**, 357–365 (2016).
293. Evers, D. J. *et al.* Optical sensing for tumor detection in the liver. *European Journal of Surgical Oncology (EJSO)* **39**, 68–75 (2013).
294. Nachabé, R. *et al.* Diagnosis of breast cancer using diffuse optical spectroscopy from 500 to 1600 nm: comparison of classification methods. *J Biomed Opt* **16**, 087010 (2011).
295. Lu, G. *et al.* Detection of Head and Neck Cancer in Surgical Specimens Using Quantitative Hyperspectral Imaging. *Clinical Cancer Research* **23**, 5426–5436 (2017).
296. Bard, M. P. L. *et al.* Improving the specificity of fluorescence bronchoscopy for the analysis of neoplastic lesions of the bronchial tree by combination with optical spectroscopy: preliminary communication. *Lung Cancer* **47**, 41–47 (2005).
297. Evers, D. J. *et al.* Diffuse Reflectance Spectroscopy: A New Guidance Tool for Improvement of Biopsy Procedures in Lung Malignancies. *Clin Lung Cancer* **13**, 424–431 (2012).
298. Brouwer de Koning, S. G. *et al.* Toward complete oral cavity cancer resection using a handheld diffuse reflectance spectroscopy probe. *J Biomed Opt* **23**, 1 (2018).

299. Baltussen, E. J. M. *et al.* Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery. *J Biomed Opt* **24**, 1 (2019).
300. Liu, Z., Wang, H. & Li, Q. Tongue Tumor Detection in Medical Hyperspectral Images. *Sensors* **12**, 162–174 (2011).
301. de Boer, L. L. *et al.* Towards the use of diffuse reflectance spectroscopy for real-time in vivo detection of breast cancer during surgery. *J Transl Med* **16**, 367 (2018).
302. Chen, H., Lin, Z., Mo, L., Wu, T. & Tan, C. Near-Infrared Spectroscopy as a Diagnostic Tool for Distinguishing between Normal and Malignant Colorectal Tissues. *Biomed Res Int* **2015**, 1–7 (2015).
303. Barnes, R. J., Dhanoa, M. S. & Lister, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl Spectrosc* **43**, 772–777 (1989).
304. Rinnan, Å., Berg, F. van den & Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* **28**, 1201–1222 (2009).
305. Kho, E. *et al.* Hyperspectral Imaging for Resection Margin Assessment during Cancer Surgery. *Clinical Cancer Research* **25**, 3572–3580 (2019).
306. Michal Kawka. Real-time detection of oesophageal and gastric cancer using diffuse reflectance spectroscopy: ex-vivo validation study. (Imperial College London, 2021).
307. Kursu, M. B. & Rudnicki, W. R. Feature Selection with the Boruta Package. *J Stat Softw* **36**, (2010).
308. Breiman, L. Random Forests. *Mach Learn* **45**, 5–32 (2001).
309. Fisher, A., Rudin, C. & Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. (2018).
310. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. (2016) doi:10.1145/2939672.2939785.
311. Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. in *Advances in Neural Information Processing Systems 30 (NIP 2017)* (2017).
312. Haka, A. S. *et al.* In vivo Margin Assessment during Partial Mastectomy Breast Surgery Using Raman Spectroscopy. *Cancer Res* **66**, 3317–3322 (2006).
313. Kong, K. *et al.* Towards intra-operative diagnosis of tumours during breast conserving surgery by selective-sampling Raman micro-spectroscopy. *Phys Med Biol* **59**, 6141–6152 (2014).
314. Keller, M. D. *et al.* Autofluorescence and diffuse reflectance spectroscopy and spectral imaging for breast surgical margin analysis. *Lasers Surg Med* **42**, 15–23 (2010).
315. Zysk, A. M. *et al.* Intraoperative Assessment of Final Margins with a Handheld Optical Imaging Probe During Breast-Conserving Surgery May Reduce the Reoperation Rate: Results of a Multicenter Study. *Ann Surg Oncol* **22**, 3356–3362 (2015).
316. Brown, J. Q. *et al.* Quantitative Optical Spectroscopy: A Robust Tool for Direct Measurement of Breast Cancer Vascular Oxygenation and Total Hemoglobin Content *In vivo*. *Cancer Res* **69**, 2919–2926 (2009).
317. Lue, N. *et al.* Portable Optical Fiber Probe-Based Spectroscopic Scanner for Rapid Cancer Diagnosis: A New Tool for Intraoperative Margin Assessment. *PLoS One* **7**, e30887 (2012).

318. Laughney, A. M. *et al.* Spectral discrimination of breast pathologies in situ using spatial frequency domain imaging. *Breast Cancer Research* **15**, R61 (2013).
319. Nichols, B. S. *et al.* A Quantitative Diffuse Reflectance Imaging (QDRI) System for Comprehensive Surveillance of the Morphological Landscape in Breast Tumor Margins. *PLoS One* **10**, e0127525 (2015).
320. Salomatina, E. & Yaroslavsky, A. N. Evaluation of the in vivo and ex vivo optical properties in a mouse ear model. *Phys Med Biol* **53**, 2797–2807 (2008).
321. Palmer, G. M., Marshek, C. L., Vrotsos, K. M. & Ramanujam, N. Optimal methods for fluorescence and diffuse reflectance measurements of tissue biopsy samples. *Lasers Surg Med* **30**, 191–200 (2002).
322. Liu, K. *et al.* Evaluation of Raman spectroscopy combined with the gated recurrent unit serum detection method in early screening of gastrointestinal cancer. *Analyst* **148**, 6061–6069 (2023).
323. DiNardo, L. J., Lin, J., Karageorge, L. S. & Powers, C. N. Accuracy, Utility, and Cost of Frozen Section Margins in Head and Neck Cancer Surgery. *Laryngoscope* **110**, 1773–1776 (2000).
324. Collins, G. S. *et al.* External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* **14**, 40 (2014).
325. Singh, R. P., Hom, G. L., Abramoff, M. D., Campbell, J. P. & Chiang, M. F. Current Challenges and Barriers to Real-World Artificial Intelligence Adoption for the Healthcare System, Provider, and the Patient. *Transl Vis Sci Technol* **9**, 45 (2020).
326. Lin, C., Lin, I.-C. & Roan, J. Barriers to Physicians' Adoption of Healthcare Information Technology: An Empirical Study on Multiple Hospitals. *J Med Syst* **36**, 1965–1977 (2012).