Ollscoil na hÉireann, Corcaigh
National University of Ireland, Cork

ucc
Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Optical Diagnostics for Oral and Gastrointestinal Tract Pathologies

Thesis presented by

Siddra Maryam

ORCID:0000-0003-2644-2088

for the degree of

Doctor of Philosophy

Department of Physics

University College Cork

Head of Department: Professor Paul Callanan
Supervisor: Professor Stefan Andersson-Engels
Co-supervisor: Ray Burke
Clinical supervisor: Professor Richael Ni Riordain

2025

Optical Diagnostics for Oral and Gastrointestinal Tract Pathologies

Biophotonics, Tyndall National Institute.
Department of Physics.
University College Cork, UCC
Ireland.

"Science and everyday life cannot and should not be separated." — Rosalind Franklin

Contents

List of Figures

List of Tables

List of publications

This thesis is based on the following research articles, which will be referenced using Roman numerals in the text. The works of Papers I, II, III and VI have been published, whereas Paper IV and V are in the pipeline.

I **Maryam, S.**; Nogueira, M. S.; Gautam, R.; Krishnamoorthy, S.; Venkata Sekar, S. K.; Kho, K. W.; Lu, H.; Riordain, N. R.; Feeley, L.; Sheahan, P.; Burke, R.; Andersson-Engels, S.; **"Label-free optical spectroscopy for early detection of oral cancer. Diagnostics"**, 2022, 12, 2896 [1].

II **Maryam, S.**; Sekar, S. K. V.; Ghauri, M. D.; Fahy, E.; Nogueira, M. S.; Lu, H.; Beffara, F.; Humbert, G.; Riordain, R. N.; Sheahan, P.; Burke, R.; Kho; K.W.; Gautam,R.; Andersson-Engels, S.; **"Mobile multi-configuration clinical translational Raman system for oral cancer application"**, Analyst 2023, 148,1514–1523 [2].

III **Maryam, S.**; Benazza, A.; Fahy, E.; Sekar, S. K. V.; Dinish, U.; Olivo, M.; Rior- dain, R. N.; Andersson-Engels, S.; Humbert, G.; Komolibus, K.; Gautam, R **"Liquid saliva analysis using optofluidic photonic crystal fiber for detection of oral potentially malignant disorders"**, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 2025, 125788.[3]

IV **Maryam, S.**; Innocente, S.; Fahy, E.; Ghauri, M. D.; Sekar, S. K. V.; Burke, R.; Visentin, A.; Riordain, R. N.; Andersson-Engels, S.; Komolibus, K.; Gautam,R.; **"Identification of salivary biomarkers for screening of Oral Potentially Malignant Disorders through tissue correlation"**, Manuscript in preparation.

V **Maryam, S.**; Ghauri, M. D.; Fahy, E.; Innocente, S.; Burke, R.; Visentin, A.; Riordain, R. N.; Feelay, L.; Andersson-Engels, S.; Gautam,R.; Komolibus, K.;**"Diffuse Reflectance Spectroscopy for non-invasive screening of Oral Potentially Malignant Disorders"**, Manuscript in preparation.

VI Nogueira, M.S.;**Maryam, S.**; Amissah, M.; Lu, H.; Lynch, N.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S.; **"Evaluation of wavelength ranges and tissue depth probed by diffuse reflectance spectroscopy for colorectal cancer detection"**, Scientific Reports, 2021. Vol. 11, no.1, p: 1-17 [4].

Abstract

Oral potentially malignant disorders (OPMD) are a significant global health burden. They are a precursor to oral cancer and indicate a tissue with a higher risk of developing a malignancy. Early diagnosis can improve patient outcomes by minimally invasive treatment, reduced mortality rate and improved post-treatment life quality. Despite the availability of advanced technology and diagnostic tools, early detection of OPMD is limited by reliance on invasive biopsies and histopathology, which are time consuming and uncomfortable for patients, and are not suitable for widespread screening or routine monitoring in a high-risk population. Moreover, this diagnostic method is prone to inter-observer and intra-observer inconsistencies and sampling error. Although technological advancements have improved diagnostic capabilities, limitations remain in accessibility, cost-effectiveness, and sensitivity; particularly for low-resource settings. Emerging diagnostic methods, including imaging, molecular assays and optical techniques, offer promise but lack comprehensive validation and seamless integration into clinical workflow. Most available technologies fail to address critical gaps such as non-invasive early detection, real-time monitoring, and the ability to link findings to actionable clinical outcomes. Furthermore, despite the biological relevance of saliva in oral health diagnostics, its potential remains underutilized due to challenges in standardization and correlation with tissue-specific markers.

This project is an exploration of what non-invasive diagnostics can achieve, when innovative methodologies meet a persistent clinical challenge. At its core, this work reflects an attempt to democratize OPMD diagnostics, shifting narrative from invasive, resource-intensive and complex procedure to non-invasive, real-time and accessible solution. This thesis explores the development and validation of a multimodal spectroscopic approach and its application in OPMD diagnosis and screening. This multimodal approach was developed to enable *in vivo* tissue and *ex vivo* saliva analysis in the clinical setting. The developed system incorporates customized modalities of Raman and diffuse reflectance spectroscopy and proposes a multi-tiered diagnostic approach to enhance clinical decision-making. The clinical study involved the collection and analysis of saliva and *in vivo* tissue analysis from OPMD patients and healthy controls. Saliva screening was investigated as a non-invasive first-line diagnostic tool to detect biochemical alterations associated with OPMD, using Raman spectroscopy of dried and liquid samples. For individuals flagged through saliva screening, *in vivo* tissue analysis is proposed using diffuse reflectance spectroscopy (DRS) and Raman spectroscopy to offer functional insights into morphology, oxygenation and chemical changes in the tissue. Machine learning models, including Lasso regression and Linear Discriminant Analysis (LDA), were employed to identify key spectral features and classify samples with high sensitivity and specificity. This multimodal approach provides complementary information. DRS probes optical properties such as absorption and scattering that helps in depicting chromophore's (including Hb, water, lipid) concentration, oxygenation and tissue structure. Raman spectroscopy, in contrast, provides more specific biochemical information regarding conformational and compositional changes in proteins, lipids, nucleic acids and related metabolites. Although some overlap exists, their combination offers extended diagnostic potential by linking functional and molecular alterations within the same tissue region, thereby improving the diagnostic potential and risk stratification.

Building on this progression, Raman analysis of liquid saliva was advanced using photonic crystal fibers (PCFs), which amplify Raman signals to detect subtle biochemical markers in real-time. This enabled label-free, point-of-care screening that complements tissue diagnostics. Together, these techniques create a clinically relevant framework: saliva screening through dry saliva sample and PCF-based Raman spectroscopy acts as a gatekeeper for regular monitoring; and *in vivo* tissue evaluation, provides a pathway for biopsy guidance, intraoperative decision making and follow-up, addressing the challenges of sampling errors and incomplete resection.

The results of this thesis highlight the feasibility of integrating Raman and DRS into clinical workflows to complements traditional biopsy. By addressing diagnostic gaps with a focus on clinical utility, this work establishes a robust, scalable, and non-invasive approach to improve the OPMD screening, early detection, monitoring, and management of OPMD. This work provides a strong foundation for future studies aimed at

refining these techniques and expanding their clinical applicability. By bridging engineering innovation with clinical needs, this research advances the goal of making early OPMD diagnostics more accessible, accurate, and patient-friendly.

Declaration

I declare that this work that I am submitting is my own work and has not been submitted for another degree anywhere else. All external sources and references are clearly cited and acknowledged. During the preparation of this work the author used an AI-assisted tool in order to correct grammar. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of this thesis. This research has financial support from Taighde Éireann – Research Ireland under grant number SFI/15/RP/2828 and SFI/22/RP-2TF/10293.

1 Introduction

In recent years, biophotonics has emerged as a ground-breaking area of study, embracing and connecting the field of biology and photonics to offer innovative solutions for medical diagnosis, prognosis and treatment. The term 'Biophotonics' is a combination of two Greek syllables "bios" meaning "life" and "phos" meaning "light"[1]. Beyond this definition, biophotonics is a discipline of great social importance. It is a multidisciplinary field. It stands at the intersection of physics, chemistry, biology, engineering and medicine. This confluence has allowed us to solve complex medical problems leveraging the principles of light-tissue interaction[2].

Controlled light use has transformed life in many ways by opening new avenues in the early detection of diseases and offering non-invasive and real-time insights into the molecular underpinnings of various pathologies [5]. For example, fluorescence imaging and optical coherence tomography (OCT) have become crucial in cardiology, ophthalmology and oncology [6–9]. Photodynamic therapy (PDT) is also widely used to treat skin cancer and is also being explored now in other cancer applications [10]. Laser Doppler flowmetry and diffuse correlation spectroscopy are used to measure blood flow in tissue [11]. Apart from this, biophotonics has proved invaluable in the non-invasive early-stage diagnosis of cervical [12], breast [13] and skin cancer [14]. These successful clinical applications highlight the potential of biophotonics in the transformation of cancer diagnostic and treatment strategies [5].

At the forefront of this advancement is the application of biophotonics in the disease diagnosis within the oral cavity and gastrointestinal tract (GI). These areas are pivotal for human health and yet face diagnostic challenges. This thesis is a critical step towards improving patient outcomes and efficacy of treatment by exploring the potential of biophotonics diagnostics to revolutionise the screening, early detection and monitoring of disease progression in oral and GI tract. While the primary focus rests on oral cancer diagnosis with a multimodal approach consisting of Raman and diffuse reflectance spectroscopy (DRS), the thesis also extends its investigative reach into diffused optical imaging (DOI) and colorectal cancer through the lens of multiple publications in which I have contributed as a second author.

1: Photonics is a broad term that covers light-based technologies throughout the spectrum, whether they produce, control, or detect light. It also encompasses the interaction of light with matter.

2: In 2014 Eric Betzig, Stefan Hell and William Moerner were awarded the Nobel Prize in Chemistry for their development of super-resolved fluorescence microscopy

1.1 The Critical Need for Early Detection

Cancer is a common chronic disease. According to an estimate of the International Agency for Research on Cancer, World Health Organisation (WHO), in 2020, over 19 million new cancer cases and more than 9.5 million cancer deaths were reported worldwide [15]. This figure increased to 20 million new cases and 10 million deaths in 2022 [16]. Cancer can develop in various organs such as the breast, prostate, colon, head, neck, oral cavity and other vital areas. Regardless of the organs affected, early-stage diagnosis is crucial to administer timely medical intervention

and prevent complications in advanced stages and potentially save lives. This is particularly true for oral cancer, where early identification of oral potentially malignant disorders (OPMD) is essential for effective management and improved outcomes.

Oral cancer is the 16^{th} most common cancer in the world with 40% to 60% five-years survival rate [17]. OPMD represents precancerous conditions in the oral cavity and offers a critical window for early intervention. The current gold standard procedure for diagnosing OPMD and oral cancer comprises a detailed visual and physical examination through palpation followed by biopsy and histopathology. There are some non-invasive light-based detection techniques such as ViziLite and Identafi but these conventional methods sometimes lack the sensitivity and specificity necessary to diagnose OPMD [18]. On the other hand, colorectal cancer (CRC) ranks as the third most prevalent cancer in the world and is the second leading cause of cancer related deaths, resulting in 0.9 million deaths annually. The WHO reported that more than 1.9 million cases of CRC were diagnosed in 2022 [16]. Biopsy and histopathology are the gold standard procedure for colorectal cancer diagnosis. In addition, CT scan (computed tomography) [19], PET scan (positron emission tomography) [20] and MRI (magnetic resonance imaging)[21] are the imaging techniques widely used for colorectal cancer diagnosis. These diagnostic methods are expensive and often involve discomfort, lengthy procedures and significant delays in obtaining results [22, 23]. Other diagnostic techniques advocated for OPMD, oral cancer and colorectal cancer, include vital staining, cytology, chemiluminescence and optical imaging [24–27]. Some of these diagnostic techniques show potential for cancer screening and early diagnosis but none of them has yet proved better than conventional visual examination and histopathology of biopsied tissue [28, 29]. There is plenty of research devoted to developing less invasive but more effective diagnostic modalities that can facilitate selection of a more suitable lesion site for biopsy and determination of any oral pre-malignant lesion that should be completely excised.

The introduction of biophotonics, particularly through the use of Raman spectroscopy and DRS has the potential to revolutionise this landscape. For example, at the onset of the disease, the biochemical composition, chromophores and optical properties of tissue begin to change from their normal values. Raman spectroscopy can provide details about these biochemical changes in tissue and DRS can measure the changes in light scattering and absorption in tissue due to variation in haemoglobin and other chromophores. By providing information regarding these pathological changes at the molecular level, this multimodal technology can enable the identification of neoplastic changes at a stage when intervention can be more effective.

1.2 The Scope of This Thesis

This thesis aims to explore the potential of Raman spectroscopy and DRS, as a part of a multimodal approach in OPMD diagnosis, as both of these spectroscopic techniques provide complementary information. Through the analysis of saliva and *in vivo* tissue, this work investigates how

these spectroscopic techniques can serve as potential tools for the early detection of OPMD and oral cancer. Saliva analysis with its non-invasive nature could provide a patient-friendly screening process and possibly help reduce the need for biopsy, while *in vivo* tissue analysis could yield critical insights into structural and biochemical changes in tissue. It could guide the clinician to determine the most suspicious area within the extensive multifocal lesions to target biopsy and facilitate the selection of OPMD lesions that should undergo complete excision. In addition to this, such spectroscopic measurements could potentially help determine the extent of surgery in patients with confirmed oral cancers. The inclusion of a study on colorectal cancer aims to showcase the versatility and potential of biophotonic diagnostics in broader applications, underscoring its importance in disease diagnosis and surgical guidance beyond oral cancer. These multimodal spectroscopic techniques could also possibly identify the onset of a disease that may remain elusive to conventional diagnostic methods.

1.3 Hypothesis

This research proposed that the integration of Raman spectroscopy and DRS in the current clinical protocol can serve as a groundbreaking multimodal approach for OPMD screening and early-stage diagnosis. As OPMD progresses, dysplastic changes in tissue cause abnormal cell proliferation, disrupting the normal tissue structure and leading to thickening or thinning of the mucosa. Vascular changes, including an increase in blood supply and change in oxygenation levels, can contribute to tumour development. At molecular level, dysplasia leads to variation in biochemical composition, including changes in proteins, lipids and carbohydrates. I hypothesize that Raman spectroscopy can effectively capture biochemical changes in saliva and tissue, revealing key molecular markers indicative of the disease. DRS, on the other hand, can extract information about structural changes, vascularization and oxygenation within tissue with the disease onset, providing a comprehensive understanding of the state of the tissue. Together, they can provide complementary diagnostic insights that allow clinicians to make informed decisions in the targeted biopsy and early interventions to improve patient outcome. With this research, my objective was to show that such a diagnostic procedure could potentially transform regular OPMD screening and follow-up examinations.

1.4 Objectives of the thesis

The broad objectives of this thesis were as follows:

1. Conduct a comprehensive review of current diagnostic practices for oral cancer and identify the gaps in methodology and technology that this research aims to address.
2. Design and optimise a multimodal system for OPMD diagnosis. This involves the integration of various spectroscopic techniques that provide complementary information to improve diagnostic accuracy and reliability. This multimodal system could potentially

be used for diagnosis in the GI tract, however, for now only DRS was tested in colorectal cancer.
3. Development of a saliva screening protocol for OPMD diagnosis. It involves establishing a methodology for saliva analysis in dry and liquid form for disease detection by assessing the change in the levels of different biomarkers.
4. Assess the performance and reliability of Raman analysis of native saliva using photonics crystal fibre (PCF).
5. Assess the effectiveness of combining Raman spectroscopy and DRS to differentiate diseased and healthy tissues to improve the precision of diagnostic procedures.
6. Investigate DRS for the detection of CRC and assess the impact of source-to-detector distance (SDD) in probe design to effectively probe deeper tissue layers in excised human colorectal tissue.
7. Develop and validate machine learning algorithms to diagnose OPMD and for tissue recognition in the colon. This involves developing predictive models to improve diagnostic accuracy and better patient outcome.
8. Explore the potential of integrating developed technology into existing healthcare framework, assessing their ease of use, cost effectiveness, robustness, portability and ultimately improved cancer detection rate.

1.5 Results Highlights

The results obtained in this study strongly support the proposed hypothesis, demonstrating the potential of the multimodal system for OPMD screening and diagnosis. The findings reinforce the feasibility of using saliva as a screening medium. The key spectral features aligning between saliva and tissue suggest a strong correlation in disease progression. Raman spectroscopy was able to detect key biochemical alterations, with significant differences observed at 1123 cm^{-1}, 1251 cm^{-1} and 1454 cm^{-1} in liquid saliva, while DRS revealed changes in vascular and oxygenation reflected in the 540/575 nm ratio. The integration of multiple modalities significantly improves diagnostic accuracy, with Raman capturing biochemical alterations and DRS detecting vascular and structural changes. These results not only validate the hypothesis but also encourage further exploration of spectroscopic approaches for early-stage OPMD detection and screening.

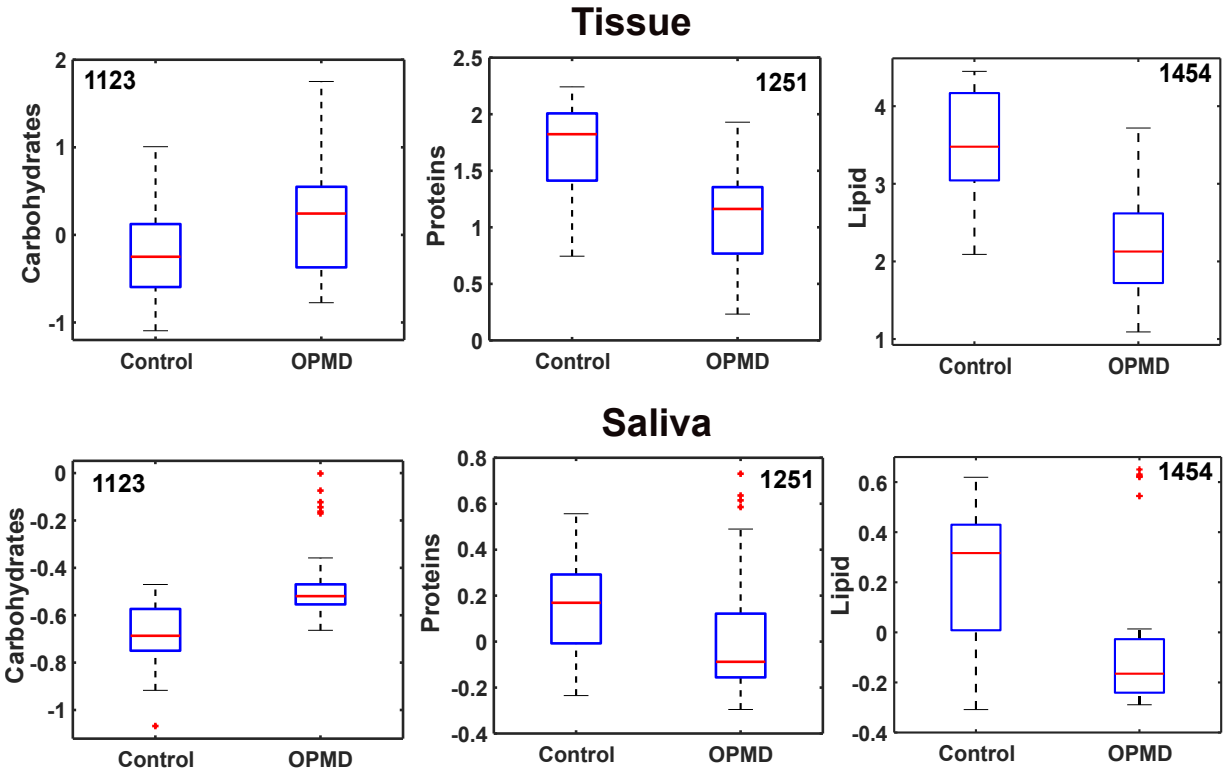


Figure 1.1: Box plot analysis of Raman peaks of liquid saliva in a PCF indicating significant changes ($p < 0.05$,) in carbohydrates, proteins and lipids in tissue (top) and saliva (bottom).

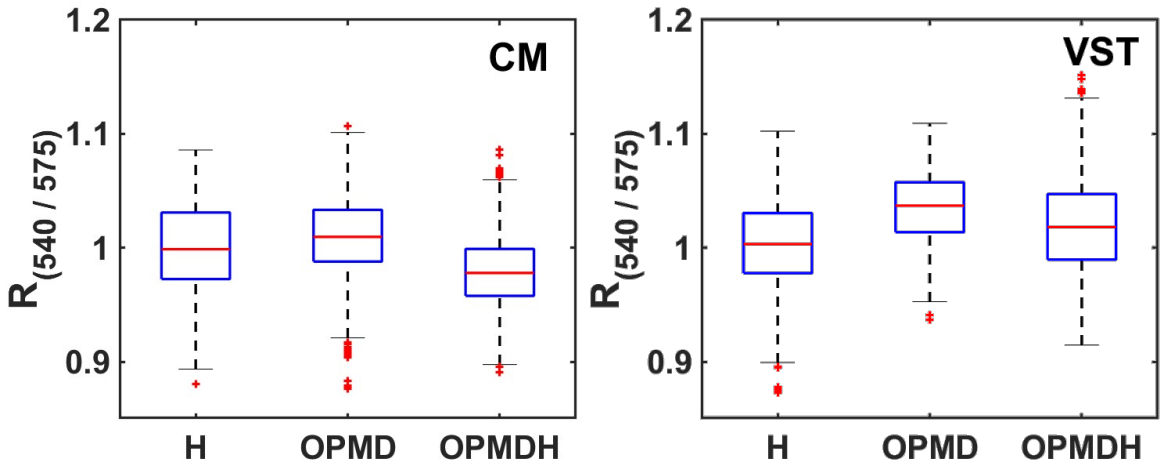


Figure 1.2: Boxplots of the reflectance ratio $R_{540/575}$ for CM (left) and VST (right) tissues. OPMD lesions show higher ratios compared to healthy and contralateral sites, indicating increased oxyhaemoglobin absorption and suggesting vascular and metabolic alterations linked to early malignant transformation.

1.6 Thesis Structure

The primary objective of this thesis is to highlight the efficacy of Raman spectroscopy and DRS in the early-stage diagnosis of oral cancer. This study presented a body of work that includes published manuscripts, detailed experimental findings and data analysis. The structure of the thesis is as follows:

Chapter 1: Introduction Outlines the significance of early diagnosis for oral cancer and the transformative potential of Biophotonics. It also sets the stage for the study.

Chapter 2: Light-tissue interaction Provides a basic understanding of Biophotonics and physics behind light-matter interaction and literature survey of the spectroscopic techniques employed in this research.

Chapter 3: Literature review and background Provides an overview of oral cancer, including its prevalence, impact, recent research and current diagnostic landscape.

Chapter 4 : Optical spectroscopy for OPMD diagnosis This chapter summarises the understanding of DRS and Raman spectroscopy in diagnosing OPMD, through recent literature.

Chapter 5: Experimental design and methodology Summarizes the multimodal spectroscopic system design and instrumentation consisting of Raman spectroscopy and DRS. Also details the experimental approach used in saliva and *in vivo* tissue analysis, including sample collection, data acquisition and analysis techniques.

Chapter 6: Results and discussion Showcases the results of saliva and tissue analysis (*in vivo*), presented through the published and submitted manuscripts. This chapter discusses the implications of these results for oral cancer detection and highlights the importance of the study in existing research and summarises the key outcome of the research.

Chapter 7: Personal perspective Summarises the personal approach and understanding regarding the research.

Chapter 7: Comments on the papers Provides an outline for published papers and manuscripts in progress.

Chapter 8: Conclusions

Chapter 8: Challenges and future perspectives Highlightes challenges in current research and methodology while providing direction and suggestions for future research avenues.

These chapters are followed by manuscript publications attached at the end of the thesis. The titles of these publications are given below.

1.7 Publications

1. **Maryam, S.**; Nogueira, M. S.; Gautam, R.; Krishnamoorthy, S.; Venkata Sekar, S. K.; Kho, K. W.; Lu, H.; Riordain, N. R.; Feeley, L.; Sheahan, P.; Burke, R.; Andersson-Engels, S.; **"Label-free optical spectroscopy for early detection of oral cancer"**, Diagnostics 2022, 12, 2896 [1].

- **About the publication** This review publication consists of a comprehensive overview of optical spectroscopy for oral cancer diagnosis, epidemiology and recent advancements in the field for diagnostic purposes. This review also highlighted the importance of salivary biomarkers for the oral cancer screening.
- **My contribution** Literature survey, manuscript writing, editing.

2. **Maryam, S.**; Sekar, S. K. V.; Ghauri, M. D.; Fahy, E.; Nogueira, M. S.; Lu, H.; Beffara, F.; Humbert, G.; Riordain, R. N.; Sheahan, P.; Burke, R.; Kho; K.W.; Gautam,R.; Andersson-Engels, S.; **"Mobile multi-configuration clinical translational Raman system for oral cancer application"**, Analyst 2023, 148,1514–1523 [2].
   - **About the publication** This manuscript is in continuation of the system development and reports the fabrication and assembly of a customised Raman system that can be modified to different configurations for *in vivo* and *ex vivo* analysis.
   - **My contribution** Methodology, system design and development, experiments, data analysis, writing and editing.

3. **Maryam, S.**; Benazza, A.; Fahy, E.; Sekar, S. K. V.; Dinish, U.; Olivo, M.; Riordain, R. N.; Andersson-Engels, S.; Humbert, G.; Komolibus, K.; Gautam, R; **"Liquid saliva analysis using optofluidic photonic crystal fiber for detection of oral potentially malignant disorders"**, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 2025, 125788 [3].
   - **About the publication** This manuscript is a groundbreaking proof-of-concept study for saliva screening. It reports a novel approach for analysing saliva samples in liquid form to detect biomarkers related to the diagnosis of OPMD. In this study a custom stage design for Raman spectroscopy of liquid saliva sample using photonics crystal fiber (PCF) is also reported.
   - **My contribution** Methodology, clinical study design, sample collection and preparations, experiments and sample analysis, data analysis, writing, review and editing.

4. Nogueira, M.S.;**Maryam, S.**; Amissah, M.; Lu, H.; Lynch, N.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S.; **"Evaluation of wavelength ranges and tissue depth probed by diffuse reflectance spectroscopy for colorectal cancer detection."** Scientific Reports, 2021. Vol. 11, no.1, p: 1-17 [4].
   - **About the publication** This study reported the evaluation of the diagnostic capabilities of DRS in colorectal cancer detection. It also describes the potential of the source-to-detecter distance in probe design to probe the deep tissue layers.
   - **My contribution** Clinical data collection, tissue probing, analysis and review.

1.8 Manuscript in preparation

1. **Maryam, S.**; Innocente, S.; Fahy, E.; Ghauri, M. D.; Sekar, S. K. V.; Burke, R.; Visentin, A.; Riordain, R. N.; Andersson-Engels, S.; Komolibus, K.; Gautam,R.; **"Identification of salivary biomarkers for screening of Oral Potentially Malignant Disorders through tissue correlation".**
   - **About the publication** This manuscript reports a novel study that relates oral cancer salivary biomarkers to those present in *in vivo* oral tissue to distinguish between OPMD and control group.
   - **My contribution** Study design, Methodology, optimisation, CREC approval, clinical data and sample collection, experiments, data analysis, writing, review and editing.
2. **Maryam, S.**; Ghauri, M. D.; Fahy, E.; Innocente, S.; Shahid, N.; Visentin, A.; Burke, R.; Riordain, R. N.; Sheahan, P.; Feelay, L.; Andersson-Engels, S.; Gautam,R.; Komolibus, K.; **"Diffuse Reflectance Spectroscopy for Non-Invasive Screening of Oral Potentially Malignant Disorders".**
   - **About the publication** This manuscript reported a unique model for selecting DRS spectroscopic features that could help in distinguishing OPMD patients from Healthy controls.
   - **My contribution** Methodology, system design and development, data collection, experiments, data analysis, writing, review and editing.

1.9 Other journal publications

The following section provides details of publications to which I have contributed, but they do not form the basis of this thesis.

1. Issatayeva, A.; **Maryam, S.**; Kho, K.W.; Andersson-Engels, S.; Cucinotta, A.; **"Suspended-core photonic crystal fibers for SERS detection of biomolecules"**; Journal of Optics, 2025, 27 [30].
2. Nogueira, M. S.; **Maryam, S.**; Amissah, M.; McGuire, A.; Spillane, C.; Killeen, S.; Andersson-Engels, S.; O'Riordain, M. **"Insights into biochemical sources and diffuse reflectance spectral features for colorectal cancer detection and localization"**. Cancers 2022, 14, 5715 [31].
3. Nogueira,M. S.; **Maryam, S.**, Amissah, M.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S.; **"Diffuse reflectance spectroscopy for colorectal cancer surgical guidance: towards real-time tissue characterization and new biomarkers"**; Analyst, 2024,Vol. 149, no.1, p:88-99 [32].
4. Nogueira, M. S.; Raju, M.; Gunther, J.; **Maryam, S.**; Amissah, M.; Lu, H.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Tissue biomolecular and microstructure profiles in optical colorectal cancer delineation"**. Journal of Physics D: Applied Physics 2021, 54, 454002 [33].

1.10 Proceeding papers

1. **Maryam, S.**; Ghauri, M. D.; Fahy, E.; Sekar, S. K. V.; Lu, H.; Russo, A.; Nogueira, M. S.; Burke, R.; Feeley, L.; Sheahen, P.; Riordain, R. N.; Andersson-Engels, S.; Kho, K.W.; Gautam,R.; **"Multimodal optical spectroscopic approach to design a complete protocol for early diagnosis of oral cancer"**. Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XXI. 2023; pp 137–140 [34].
2. **Maryam, S.**; Ghauri, M. D.; Fahy, E.; Sekar, S. K. V.; Nogueira, M. S.; Lu, H.; Russo, A.; Burke, R.; Feeley, L.; Sheahan, P.; Riordain, R. N.; Andersson-Engels, S.; Kho, K.W.; Gautam,R.; **"Biophotonics diagnostics of oral cancer using Raman spectroscopy"**. European Conference on Biomedical Optics. 2023; p 126270B [35].
3. **Maryam, S.**; Ghauri, D.; Sekar, S. K. V.; Fahy, E.; Nogueira, M. S.; Lu, H.; Beffara, F.; Humbert, G.; Burke, R.; Feeley, L.; Sheahan, P.; Riordain, R. N.; Andersson-Engels, S.; Kho, K.W.; Gautam,R.; **"Non-invasive multimodal spectroscopic diagnosis for early-stage oral cancer"**. BioOptics: Design and Application. 2023; pp DM4A–2 [36].
4. **Maryam, S.**; Nogueira, M. S.; Moorthy, S. K.; Sekar, S.; Lu, H.; Gautam, R.; Burke, R.; Andersson-Engels, S.; Riordain, R. N.; Sheahan, P. **"Multi-configuration Raman spectrometer for early-stage diagnosis of oral cancer"**. Biomedical Vibrational Spectroscopy 2022: Advances in Research and Industry. 2022; pp 20–26 [25].
5. Issatayeva, A.; Farnesi, E.; **Maryam, S.**; Cialla-May, D.; Kho, K.; Schmitt, M.; Andersson-Engels, S.; Milanese, D.; Cucinotta, A. **"SERS Measurement of miRNA with the Potential Application for Liquid Biopsy".** Sensors and Electronic Instrumentation Advances 2024, 148 [37].
6. Innocente, S.; **Maryam, S.**; Andersson-Engels, S.; Komolibus, K.; Gautam, R.;Visentin, A. **"A comprehensive pipeline to integrate preprocessing and machine learning techniques for accurate classification in Raman spectroscopy"**. Data Science for Photonics and Biophotonics. 2024; pp 51–57 [38].
7. Nogueira, M. S.; **Maryam, S.**; Amissah, M.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Intestinal anastomosis automation through diffuse reflectance spectroscopy : towards real time guidance during robotic surgery"**. High-Speed Biomedical Imaging and Spectroscopy IX. 2024; pp 66–70 [39].
8. Nogueira, M. S.; **Maryam, S.**; Amissah, M.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Colorectal cancer surgical guidance by using diffuse reflectance spectroscopy"**.Diffuse Optical Spectroscopy and Imaging IX. 2023; pp 15–20 [40].
9. Nogueira, M. S.; Raju, M.; Gunther, J.; **Maryam, S.**; Amissah, M.; Killeen, S.; Lu, H.; O'Riordain, M.; Andersson-Engels, S. **"Accurate colorectal cancer detection and delineation by probing superficial and deeper tissue biochemistry and microstructure using diffuse reflectance spectroscopy"**. Molecular-Guided Surgery: Molecules, Devices, and Applications VIII. 2022; p 1194302 [41].
10. Nogueira, M. S.; **Maryam, S.**; Amissah, M.; Lu, H.; Lynch, N.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Improving colorectal cancer detection by extending the near-infrared wavelength range**

and tissue probed depth of diffuse reflectance spectroscopy: a support vector machine approach". Optical Biopsy XX: Toward Real-Time Spectroscopic Imaging and Diagnosis. 2022; pp 12–21 [42].

11. Nogueira, M. S.; Raju, M.; Gunther, J.; **Maryam, S.**; Amissah, M.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Optical determination of superficial and deeper tissue biochemistry and microstructure for delineation and early detection of colorectal cancer"**. Translational Biophotonics: Diagnostics and Therapeutics. 2021; pp 74–76 [43].
12. Nogueira, M. S.; Li, C. L.; **Maryam, S.**; Tyndall, C.; Donovan, S.; Komolibus, K.; Gunther, J. E.; Sekar, S. K. V.; Jayet, B.; Fisher, C.; Andersson-Engels, S. **"Multidisciplinary teaching and learning biophotonics: assessing the effectiveness of fully online outreach events"**. Optical Interactions with Tissue and Cells XXXIII; and Advanced Photonics in Urology. 2022; pp 112–122 [44].
13. Saito Nogueira, M.; Amissah, M.; **Maryam, S.**; Lynch, N.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Optimization of tissue classification for colorectal cancer detection using support vector machines and diffuse reflectance spectroscopy"**. Translational Biophotonics: Diagnostics and Therapeutics 2021, 11919, 1191929 [45].
14. Saito Nogueira, M.; **Maryam, S.**; Amissah, M.; Lynch, N.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Benefit of extending near-infrared wavelength range of diffuse reflectance spectroscopy for colorectal cancer detection using machine learning"**. Translational Biophotonics: Diagnostics and Therapeutics 2021, 11919, 1191928 [46].
15. Nogueira, M. S.; Komolibus, K.; Gunther, J. E.; Sekar, S. K. V.; Jayet, B.; Tobo, A. L. P.; Li, C. L.; **Maryam, S.**; Tyndall, C.; Kennedy, D.; others; **"Online learning combining virtual lectures, at-home experiments and computer simulations: a multidisciplinary teaching and learning approach. Education and Training in Optics and Photonics"**. 2021; pp F2B–6 [47].
16. Nogueira, M. **S.; Maryam**, S.; Amissah, M.; Lu, H.; Manning, E.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. **"Broadband diffuse reflectance spectroscopy for colorectal surgical guidance"**; Optical Interactions with Tissue and Cells XXXI. 2020; p 1123802 [48].

Conclusion

This chapter has outlined the scope of this thesis, presenting the objectives, hypothesis, and structure, thereby setting the stage for the subsequent chapters. It highlights the importance of early-stage disease diagnosis in the oral and gastrointestinal tract, particularly focussing on the transformative potential of biophotonics in this domain. It also emphasised the critical need for non-invasive innovative diagnostic methods, such as Raman and DRS to address the challenges in the current diagnostic methods. With this fundamental knowledge, the following chapter will provide theoretical framework to understand spectroscopic techniques

employed in this research to improve diagnostic accuracy and patient outcomes in the context of oral diseases.

Light Tissue Interaction 2

This chapter provides an overview of the fundamental principles of light propagation in various media, focusing on biological tissues. These concepts are crucial to understand the experimental and analytical methods employed in the studies presented in this thesis. Specifically, the basics of Raman scattering, including its classical description and selection rules, form the foundation for papers II, III and IV. These papers investigate applications of Raman spectroscopy for tissue and saliva analysis. Additionally, a detailed discussion of light-tissue interaction, encompassing absorption, scattering, and radiative transport, underpins the modeling and interpretation of diffuse reflectance spectroscopy data in Papers V and VI. By addressing both theoretical and practical aspects, this chapter establishes the groundwork for the experimental and computational approaches adopted in subsequent chapters.

2.1 Light propagation in different media

Light travels in a straight line from its source but when it encounters an obstacle or when the light travels through one medium to another with a different refractive index (n), it can exhibit behaviours such as reflection, refraction, absorption and scattering.

1. **Direct transmission:** When light travels through a **homogeneous medium** with uniform optical properties, it can pass through without being absorbed, scattered, or reflected maintaining its original characteristics. Scattered light can also pass through the medium, but it will be dispersed in different directions. However, as light transitions from one material to another with a different refractive index, it may refract at the interface, with its direction determined by Snell's law [1].
2. **Specular reflection:** Specular reflection occurs when light reflects of a smooth and polished surface at the same angle as the angle of incidence. Mirror is a common example of specular reflection. This type of reflection produces clear and sharp images
3. **Total internal reflection:** When the light moves from a medium of higher refractive index to a medium of lower refractive index at an angle greater than the critical angle [2] then it is totally internally reflected.

(Figure 2.1) represents light propagation in a homogeneous medium.

4. **Scattering:** When light encounters a **heterogeneous medium**, scattering occurs due to local variations in refractive index at the microscopic scale, caused by impurities, structural irregularities, density variations. When light travels through such a medium, these refractive index variations cause the light to deviate from its original path, resulting in multiple scattering events. This results in diffusion of light throughout the medium. The degree of scattering

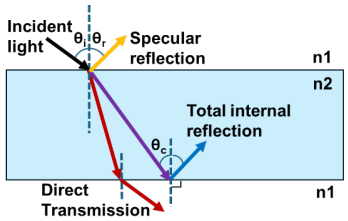


Figure 2.1: Light travelling through a medium of different refractive indices, where $n_2 > n_1$.

1: When the light enters in to a medium of higher refractive index, it bends towards the normal because of the change in the speed of light in different media. According to Snell's law, when the medium is changed, the ratio of sin of angle of incidence (θ_i) and angle of refraction (θ_r) of a wave is constant.

$$n_1 sin(\theta_i) = n_2 sin(\theta_r)$$

2: The angle of incidence for which the angle of refraction is 90° or greater is called the critical angle.

is influenced by the concentration, size and refractive index contrast of the heterogeneities relative to the wavelength of light. For example, if the scatterers are much smaller than the wavelength, Rayleigh scattering dominates, whereas Mie scattering becomes significant for larger scatterers. Figure 2.2 depicts the schematic of light propagation through a diffusive medium, where black dots represent impurities or regions with varying refractive indices.

5. **Absorption** Absorption occurs when the light energy is taken up by the medium, converting it into heat and other forms of energy, and reducing the intensity of the transmitted light.

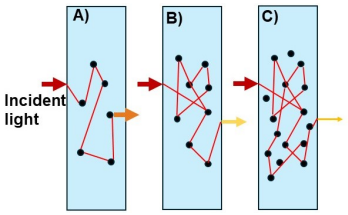


Figure 2.2: Schematic of light path within a diffusive media with different concentration of the scatterers.

2.2 Light propagation in biological tissues

Having explored how light interacts with different media, it becomes crucial to apply these principles to more complex, heterogeneous media, like biological tissues.
Tissues are highly diffusive because of their multilayered nature, rough interfaces, complex fibrous structure and critical composition. They exhibit multiple scattering, primarily because of variations in refractive index on micrometer and submicrometer scales. Subcellular components such as mitochondria and cell membranes, which are lipid-rich, contrast with the water-based intracellular and extracellular fluids, leading to significant changes in the refractive index at this level. These inhomogeneities, depending on their size relative to the wavelength of light, influence the scattering behaviour. Structures smaller than the wavelength of light, such as organelles or small protein aggregates, scatter light isotropically in all directions, following the principle of Rayleigh scattering. In contrast, larger structures, such as entire cells, mitochondria or collagen fibers, tend to scatter light predominantly in the forward direction, consistent with Mie scattering. The size and refractive index mismatch of mitochondria influence the g-factor, which describes the anisotropy[3] of scattering- how forward-scattered the light will be during each scattering event. A higher g-factor indicates more forward scattering, as is often observed in tissues with well-organized, larger subcellular structures. This multiple scattering and absorption causes the light beam to spread and attenuate as it propagates through the tissue, as depicted in Figure 2.3. Figure 2.4 represents schematics of light propagation within the tissue, highlighting scattering and absorption. Along with the complex nature of tissue and organelles, the polarization state of incident light also modulates propagation and interaction of light with the tissue.

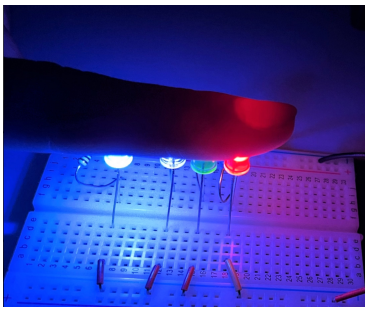


Figure 2.3: Light propagation in tissue.

3: Anisotropy describes the tendency of scattered light to be directed in a specific direction rather than being scattered uniformly in all directions. It is quantified by the anisotropic factor "g".

2.3 Energy levels

An atom or molecule can only have discrete energy values. These energy values are quantized due to the wave-particle duality of matter and are named energy levels. The quantization of energy levels is described by the Schrödinger equation. In atomic physics, principal energy levels are considered as orbits of one or more electrons surrounding a positively

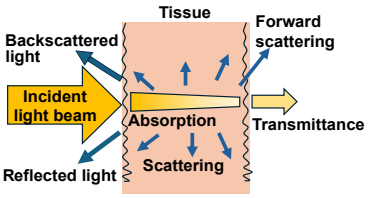


Figure 2.4: Schematic representation of the interaction of light with tissue.

charged nucleus. These orbitals are named K, L, M, N, ... shells. These shells correspond to principal quantum numbers "n" (where n = 1, 2, 3, 4, 5 ...) [49]. If there is only one electron in the atomic orbital, such as in hydrogen, energy state depends on the electrostatic interaction between nucleus and electron, and is determined by the principal quantum number n [50]. As shown in eq. 2.1

$$E_n = -hcR_H \frac{Z^2}{n^2} \tag{2.1}$$

here, E_n is the energy of n^{th} atomic orbital, h is Planck's constant, c represents the speed of light, R_H is the Rydberg constant for hydrogen[4], n is the principal quantum number and Z represents the atomic number. For atoms having more than one electron, Z is replaced by Z_{eff} due to the shielding effect caused by other electrons.
When atoms join by covalent bonds to form molecules, the atomic orbitals affect each other's energies to stabilize the system. As a result, bonding and anti-bonding molecular orbitals are generated. A molecular energy state is the sum of electronic, vibrational and rotational energies [50].

4: Rydberg constant is a physical constant. It describes the wavelength of light absorbed or emitted by an electron while transitioning between energy levels. Its value is $\approx 1.09 \times 10^7 m^{-1}$.

$$E_{total} = E_{electronic} + E_{vibrational} + E_{rotational} \tag{2.2}$$

Electronic energy levels result from a change in the distribution of electrons in molecular orbitals. Electronic transition between these levels occurs due to absorption or emission of energy in the range of ~ 1 to several eV. These energy levels of a molecule are quantized because of the wave-particle duality of matter. Their precise values are governed by quantum numbers derived from solving the Schrödinger equation [51].

$$\hat{H}\Psi = E\Psi \tag{2.3}$$

Here $\hat{H}$ is the Hamiltonian operator, Ψ is the wavefunction and E is eigenvalue [51]. The Hamiltonian for electronic energy level is:

$$\hat{H}_{electronic} = -\sum_i \frac{\Delta_i^2}{2} - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{i>j} \frac{1}{r_{ij}} + \sum_{A<B} \frac{Z_A Z_B}{R_{AB}} \tag{2.4}$$

here A, B represent two nuclei that make up the molecule; Z_A, Z_B are their atomic numbers respectively, and i, j are electrons; r_{iA} is the distance between electron i and nucleus A, while r_{ij} represents the distance between electrons i and j; R_{AB} is the distance between two nuclei A and B [52]. In eq. 2.4

- the first term represents the kinetic energy of electrons.
- the second term is due to the attraction between the electron and nuclei.
- the third term shows repulsion between electrons.
- the fourth term is repulsion between both nuclei.

The electronic energy levels can be estimated from solving this equation for a system of nuclei and electrons, leading to the quantization of energies. Each electronic state has many vibrational states and is characterized by vibrational quantum numbers "**v**". Moreover, each vibrational state has a rotational structure, as shown in Figure 2.5, and its rotational energy

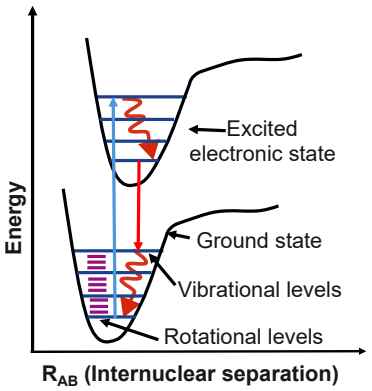


Figure 2.5: Electronic, vibrational and rotational energy levels in a diatomic molecule.

states are represented by the rotational quantum numbers "**J**".
This complex interaction within molecules can be simplified through the Born-Oppenheimer approximation by assuming that nuclear motions (vibrational, rotational and translational) are much slower than the electronic motion due to the mass difference between electrons and nuclei. This enables a stepwise analysis of energy levels, allowing the total molecular wavefunction to be separated into electronic, vibrational, and rotational components [52].

$$\Psi_{total} = \Psi_{electronic}(r, R)\Psi_{nuclear}(R) \tag{2.5}$$

where, r represents electronic coordinates, R nuclear coordinates, $\Psi_{electronic}$ is electronic wavefunction for a fixed nuclei and $\Psi_{nuclear}(R)$ is nuclear wavefunction.
Vibrational energy levels: Vibrations correspond to periodic motions of the nuclei around equilibrium bond lengths within a fixed electronic state. These motions are similar to a harmonic oscillator, and the vibrational energy levels involve energies of the range of 0.1 eV. The vibrational energy levels are given by:

$$E_v = \hbar\omega(v + \frac{1}{2}) \tag{2.6}$$

here, v = 0, 1, 2, 3 ...vibrational quantum number and ω is the vibrational angular frequency [53].

Rotational energy levels: Rotation involves the overall angular motion of the molecule. The associated energy levels are on the order of 1 meV. The rotational motion is approximated as a rigid rotor [54]. Rotational energy levels are estimated by

$$E_J = \frac{\hbar^2}{2I}J(J+1) \tag{2.7}$$

here, J = 0, 1, 2, 3 ... representing the rotational quantum number and I is the moment of inertia [5]. An energy-level diagram of a possible transition is given in Figure 2.5. The equations described in this section form the theoretical foundation for spectroscopic techniques like Raman and DRS, which probe transitions between these quantized energy levels.

5: Moment of inertia of a molecule is given by

$$I = \mu R_e^2$$

where

- R_e = equilibrium bond length
- μ = reduced mass of the system.

2.4 Absorption

Light absorption is the process by which light energy is converted to other forms of energy such as fluorescence, sound or heat. It occurs when atoms or molecules absorb energy from incident light, causing transitions from their ground states to excited states. According to quantum theory, the energy provided by the incident photon should be equal to the energy difference between the excited state (E_{10}) and the ground energy state (E_{00}) [55, 56].

Figure 2.6 represents an example of a molecule that is excited to the first vibrational state of the excited electronic state having energy E_{10} from its

ground state of energy E_{00}. This transition can be described by eq. 2.8.

$$E_{00} + E_{Photon} \rightarrow E_{10} \tag{2.8}$$

where,

$$E_{photon} = h\nu_0 \tag{2.9}$$

ν_0 is the frequency of incident light wave and h is the Planck's constant [6].

When the system is in the first electronic state (E_{10}), the decay to the ground energy state (E_{00}) results in the emission of a photon. If this photon is emitted as the molecule returns to its ground energy state after a certain lifetime without the influence of external radiation, this process is termed spontaneous emission (Figure 2.7). The emitted photon will have an energy equal to the energy difference between the two energy states as referred in eq. 2.10 but the direction of emission and polarization is random. This process is a fundamental mechanism of light emission in various types of luminescence, such as fluorescence and phosphorescence.

$$E_{10} \rightarrow E_{00} + E_{Photon} \tag{2.10}$$

In 1917 Einstein introduced the concept of stimulated emission (Figure 2.8), where emission of a photon is triggered by an external radiation of energy $h\nu_0$. An external photon interacts with an atom or molecule that is already in its excited state, prompting it to drop to a lower energy state and releasing an additional photon of the same energy in the same direction as the incident photon, giving rise to coherent output radiation. This phenomenon can be used to amplify a coherent electromagnetic wave [55–57].

2.4.1 Beer-Lambert law and coefficient of absorption

If a monochromatic light with intensity I_0 passes through an absorbing medium, some of the light is absorbed by the medium and the intensity of transmitted light is I (Figure 2.9). In spectroscopy, the term absorbance (A) is commonly used to describe the amount of light absorbed by the medium. It is the logarithmic ratio of incident light intensity (I_0) to the intensity of transmitted light (I). Sometimes, it is also termed optical density (OD). It is a dimensionless quantity and can be written mathematically as,

$$A = \log(\frac{I_0}{I}) \tag{2.11}$$

The Beer-Lambert law describes the relation between the light attenuation and properties of the medium through which the light travels. According to this law, the absorbance (A) is directly proportional to both the path length (d) of the light in centimetres (cm) and the concentration (C) of

6: Planck's constant was introduced by Max Planck in 1900. It defines the relation of frequency of photon to its energy [56, 57]. It is given by:

$$h = 6.626 \times 10^{-34} Js$$

or

$$h = 4.136 \times 10^{-15} eVs$$

where one electronVolt (eV) is the amount of energy gained or lost by an electron when moved across a potential difference of one Volt (1 eV = 1.602×10^{-19} J).

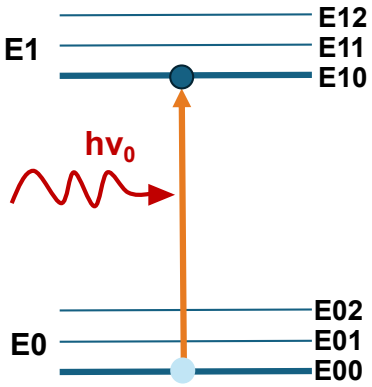


Figure 2.6: Photon absorption resulting in excitation of molecule from a ground state of energy E_{00} to a higher energy state E_{10}.

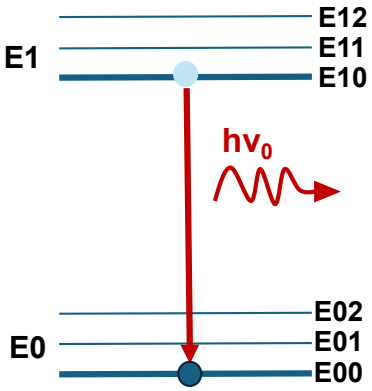


Figure 2.7: Spontaneous emission of a photon by transfer of molecule to its ground energy state E_{00}. The average lifetime for most of the energy states is of the order of 10^{-9} s.

the absorbers in the homogenous medium expressed in $(mol \cdot L^{-1})$, as shown in (eq. 2.12)

$$A \propto C \cdot d \tag{2.12}$$

$$A = \epsilon_\lambda C d \tag{2.13}$$

where ϵ_λ is a proportionality constant, termed as the extinction coefficient. It depends on wavelength λ and its units are $[cm^{-1} \cdot M^{-1}]$. It is also sometimes named as the molar absorptivity of the medium. It is an intrinsic property of the molecule and determines how strongly it can absorb light at a particular wavelength [56]. From eq. 2.13 and eq. 2.11 absorbance (A) can be written as:

$$A = \log(\frac{I_0}{I}) = \epsilon_\lambda C d \tag{2.14}$$

In biological tissues, absorption occurs because of molecular chromophores. These molecules are capable of absorbing light at specific wavelengths. Each chromophore has an absorption cross section σ_{abs} that is a quantitative measure of its ability to absorb light. Its unit is m^2. Mathematically, it can be expressed as

$$\sigma_{abs} = \frac{P_{abs}}{I_0} \tag{2.15}$$

Where I_0 is the incident light intensity in W/m^2 and P_{abs} is the power absorbed by the chromophores in W [56].

The absorption of a tissue is usually defined by its absorption coefficient μ_a $[cm^{-1}]$. It represents the likelihood of photon absorption per unit length as they travel through the tissue. It can be calculated using the absorption cross section σ_{abs} and the density ρ of the absorbing particles per unit volume:

$$\mu_a = \rho \sigma_{abs} \tag{2.16}$$

In addition to the absorption cross section of the chromophores σ_{abs}, μ_a also depends on the concentration C and the extinction coefficient ϵ_λ , which vary with the wavelength of incident light. Thus, it can provide subtle information about the molecules in the sample and their environment [56, 58]. The density (ρ) of the absorber is related to the concentration (C) of the medium in $[mol \cdot L^{-1}]$ using Avogadro's number $(N_A = 6.022 \times 10^{23})$, as described in eq. 2.17

$$\rho = C \cdot N_A \cdot 10^3 \tag{2.17}$$

Here, the factor 10^3 converts the volume unit from liters (L) to cubic centimetres (1 L = 1000 cm^3)

The transmitted light intensity I is described by the exponential decay as:

$$I = I_0 e^{-\mu_a d} \tag{2.18}$$

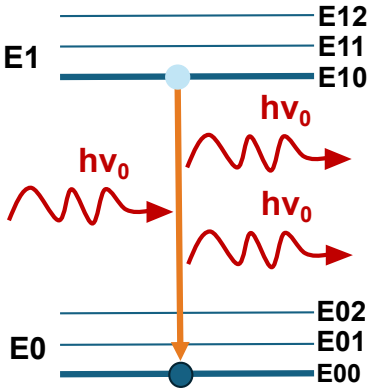


Figure 2.8: Stimulated emission of a photon by transfer of molecule to its ground energy state E_{00}.

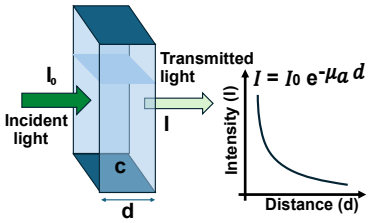


Figure 2.9: Transmission of light through a liquid sample in a cuvette.

by taking log of the eq. 2.18

$$log(\frac{I}{I_0}) = -\mu_a d \cdot log e \tag{2.19}$$

from eq. 2.19 and eq. 2.14

$$\mu_a = 2.3026 \cdot \epsilon_\lambda C \tag{2.20}$$

The reciprocal of μ_a [cm] is called absorption mean free path l_a. It represents the average distance travelled by photon before getting absorbed.

$$l_a = \frac{1}{\mu_a} \tag{2.21}$$

In tissues, common chromophores that absorb light include haemoglobin, bilirubin, myoglobin, cytochrome pigments or melanin, water and lipid [58, 59].

2.5 Transmittance

Transmittance is another term related to absorbance used in tissue optics. It is a measure of how much light is transmitted through the medium. It is the ratio of transmitted light intensity (I) after traveling a distance (d) to the incident intensity (I_0). It does not have any unit and is usually measured in terms of percentage [60]. Mathematically, it is represented as

$$T = \frac{I}{I_0} = e^{-\mu_a d} \tag{2.22}$$

Which can related to absorbance as:

$$T = \frac{I}{I_0} = 10^{-A} = 10^{-\epsilon_\lambda c d} \tag{2.23}$$

2.6 Scattering

Scattering plays a significant role in tissue optics. In addition to absorption, it is one of the key phenomena that contributes to the optical properties and appearance of the material [61]. It occurs when light passing through a medium changes its direction due to variation in the refractive index within the medium. Scattering is a process that changes the direction of the incident electromagnetic wave. As light undergoes multiple scatterings, its original path is altered significantly, transforming a uniform and collimated beam into a scattered and sometimes diffuse light propagation. This scattered light appears diffused and spreads in multiple directions. Unlike in the absorption process, light does not get attenuated during scattering but is merely redirected. The degree of change in the initial light form depends on the properties of the biological scatterers [62–64]. In tissue, scattering is primarily caused by the change in the refractive index

between different cellular and extracellular structures such as collagen fibrils or lipid-water membrane interface within and surrounding each cell [65, 66]. Thus, understanding the scattering of light in tissue can be invaluable in disease diagnosis by distinguishing between healthy and diseased tissues [62–64]. Scattering can be described by two parameters, anisotropic factor (g) and scattering coefficient (μ_s). The anisotropic factor refers to the distribution of light scattered in different directions relative to incident light and is the average value of the cosine of the scattering angles [7][67]. The scattering coefficient is the probability for a photon to scatter per unit path length in tissue. Its unit is cm^{-1} and is the product of number of scattering particles per unit volume (N) and scattering cross section $\sigma_s[m^2.]$:

$$\mu_s = N\sigma_s \tag{2.24}$$

σ_s, the cross section of a single particle is calculated as the ratio of scattered power and intensity of incident light:

$$\sigma_s = \frac{P_s}{I_0} \tag{2.25}$$

The reciprocal of scattering coefficient μ_s is the scattering mean free path l_s [cm]. It quantifies the distance covered by a photon before being scattered.

$$l_s = \frac{1}{\mu_s} \tag{2.26}$$

In turbid media such as tissue, where light behaves diffusely, both parameters scattering coefficient and anisotropic factor (g) can be combined as

$$\mu_s' = \mu_s(1 - g) \tag{2.27}$$

where

$$g = \langle cos\theta \rangle \tag{2.28}$$

μ_s' is the reduced scattering coefficient, it is a measure of effective scattering in the medium (Figure 2.10). Not all scattering events can change the direction of the photon significantly, thus this coefficient also accounts for anisotropic factor. This coefficient is particularly useful in describing light transport in biological tissues[8] where scattering is primarily in the forward direction or anisotropic [56, 68]. The transport mean free path (l^*), the average distance a photon travels before its direction becomes random, is inversely related to the reduced scattering coefficient. It is a crucial parameter to understand the light propagation in the tissue and is given by:

$$l^* = \frac{1}{\mu_s' + \mu_a} \tag{2.29}$$

Scattering can be classified as elastic and inelastic scattering [56].

7: Value of g varies between -1 and 1. Where g = 1 means that all photons would travel straight without changing direction, g = 0 represents isotropic scattering, where light is scattered uniformly in all directions, while g = -1 represents backward scattering.

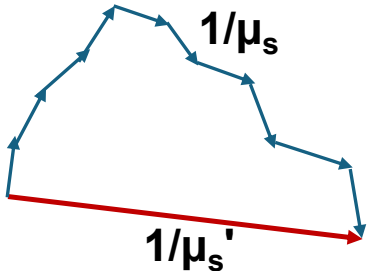


Figure 2.10: Graphical representation of μ_s and μ_s'. For biological tissues, if g = 0.9 then $\mu_s' = 0.1\mu_s$.

8: In biological tissues, the anisotropic factor (g) typically ranges between 0.8 and 0.99, reflecting the fact that scattering in tissues is predominantly in forward direction or anisotropic.

2.6.1 Elastic scattering

Elastic scattering is a fundamental phenomenon in tissue optics. It occurs when there is no change in the energy of the incident photon after

scattering. During elastic scattering, the energy and frequency of the incident photon are conserved. Elastic scattering is important for medical imaging and diagnostic applications as it can provide information about the structure and morphology of the tissue [69, 70]. Depending upon the size of the light scatterer of the medium, elastic scattering can be categorized as Rayleigh scattering and Mie scattering.

Rayleigh scattering

Rayleigh scattering[9] occurs when light is scattered by particles much smaller in size than the wavelength of light. The intensity of scattered light varies with the inverse of the 4^{th} power of the wavelength, i.e. λ^{-4}(eq. 2.30). which makes shorter wavelengths (blue and violet) scatter more than longer wavelengths (red light). In biological tissues, Rayleigh scattering is less prominent because the wavelength of light is relatively small compared to the size of the cellular component [71, 72].

9: Rayleigh scattering is named after lord Rayleigh, who described it in the 19^{th} century.

$$I_s \propto \frac{1}{\lambda^4} \tag{2.30}$$

From Figure 2.11, if light of frequency ν_0 excites a molecule, the photon emitted as a result of Rayleigh scattering will have the same energy as the incident photon.

$$E_0 = h\nu_0 \tag{2.31}$$

Mie scattering

Mie scattering plays a significant role in interaction of light with biological tissues, where the size of scatterers, such as nuclei, organelles, and extracellular matrix components, are comparable to the wavelength of incident light [66, 72]. In pathological conditions like OPMD, cancer or fibrosis, abnormal cell proliferation leads to increased cellular density, enlarged nuclei and altered extracellular architecture. These structural changes modify the scattering profile of the tissue, enhancing Mie scattering effects [73, 74]. Diffuse reflectance spectroscopy (DRS), which relies on detecting backscattered light, is sensitive to these alterations. Therefore, understanding Mie scattering is essential for interpreting DRS signals and improving its diagnostic accuracy in identifying dysplastic or malignant transformations in oral mucosa [66, 75, 76].

Sir Chandrasekara Venkata Raman won a Nobel prize in Physics in 1930, for his work on light scattering from liquids and gases and for the discovery of Raman effect.

2.6.2 Inelastic scattering

Inelastic scattering occurs when the incident photon of light loses or gains energy during scattering and the wavelength of the incident light changes after scattering due to energy transfer between a photon and the medium. This type of scattering is the foundation of Raman spectroscopy, where light is scattered from molecules with a change in wavelength that corresponds to specific molecular vibrations [77–79]. Figure 2.11 depicts the Jablonski diagram for Rayleigh and Raman scattering, where an incident photon of energy $h\nu_0$ transits the molecule to a virtual energy state[10] "VS" and then returns to the ground state by releasing a scattered

10: A virtual energy state is a non-real temporary intermediate state. It does not have a fixed energy value as in the stable molecular states. When a molecule transits through a virtual state, it quickly returns to a vibrational level of the ground state, releasing a photon of light.

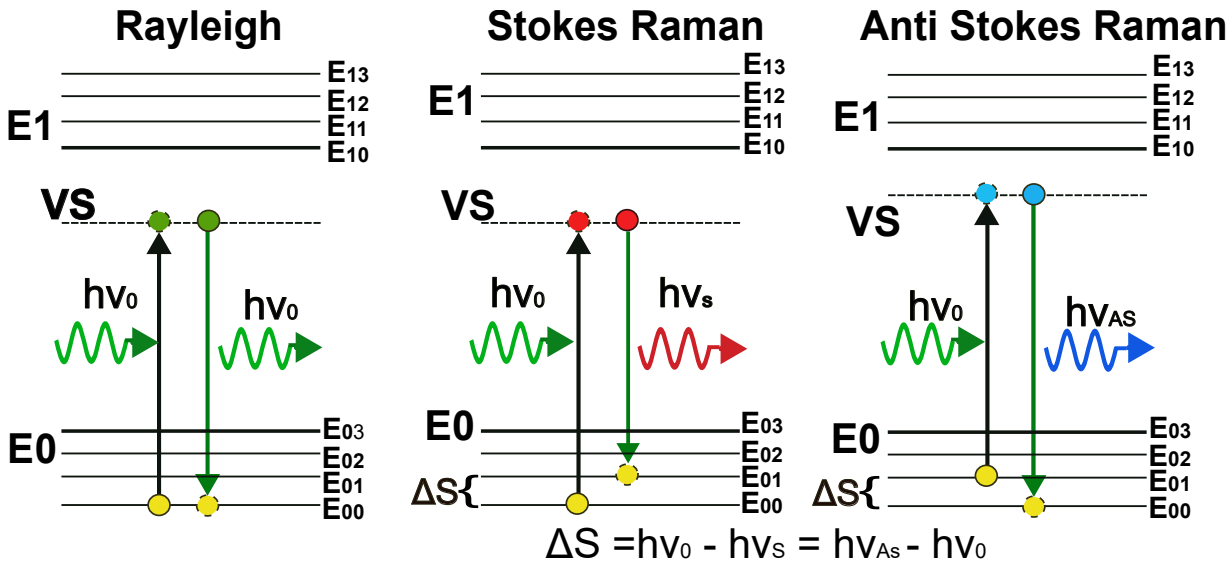


Figure 2.11: Jablonski diagram for Rayleigh and Raman scattering.

photon of the same energy in the case of Rayleigh scattering and different energy in case of Raman scattering. The difference in energy either gained (anti-Stokes) or lost (Stokes) by the photon is called the **Raman shift** and is expressed in **wavenumbers** (cm^{-1}) ΔS, Where from Figure 2.11, ΔS is

$$\Delta S = h\nu_0 - h\nu_S = h\nu_{AS} - h\nu_0 \tag{2.32}$$

Sir Chandrasekara Venkata Raman was awarded the Nobel prize in 1930. He was studying scattering of light through different gases and liquids when he noticed the presence of other lines at higher frequency, in addition to the original monochromatic incident frequency. These lines were independent of the excitation wavelength. He also noticed that the scattered light spectrum is specific to the material and is different for each molecule, and hence contains information related to the molecular structure and vibrational energy levels of the material. Only one in 10^6 of the incident photons undergoes Raman scattering [80].

Stokes scattering

Stokes scattering occurs when a photon interacts with the molecule of the medium, transiting it to a virtual energy state. The molecules then relax immediately to a vibrational level in the ground electronic state, losing energy and generating a red shifted scattered photon of lower energy $h\nu_s$. This shift towards longer wavelengths is characteristic of Stokes scattering and is fundamental to unstimulated Raman spectroscopy providing insights into molecular structures and dynamics [78]. Figure 2.12 represents different modes of molecular vibrations.

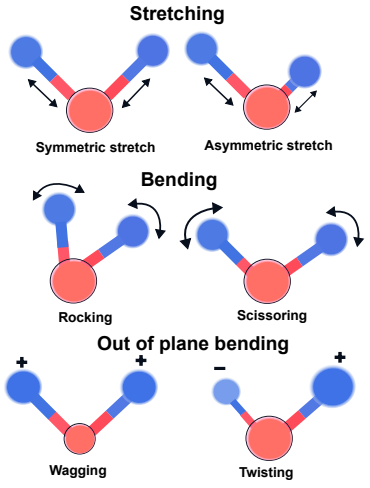


Figure 2.12: There are mainly six possible vibrational modes in a molecules, such as symmetric and asymmetric stretching, wagging, twisting, scissoring, rocking. Twisting, rocking and wagging modes do not affect the chemical bond length, but they only change their angles. Inelastic scattering is highly specific to the molecule because of distinct combination of vibrational states.

Anti-Stokes scattering

Anti-Stokes scattering occurs when the scattered photon has more energy than the incident photon. When a sample is illuminated by incident

photons, some molecules are already in the excited vibrational state. When they absorb a photon and get excited to a virtual energy state, they may relax back to a lower energy state, releasing a blue-shifted photon, which has higher energy, $h\nu_{AS}$, than the incident one.
The intensity of anti-Stokes scattering[11] depends on the population of the molecules in the excited vibrational state, which in turn depends on the temperature. Anti-Stokes lines provide information on the thermal properties of the sample and can be used to measure the temperature distribution within a sample [79]. The intensity of anti-Stokes Raman scattering increases proportionally to the number of molecules N_1 in the excited state. At room temperature, the number of molecules in the excited vibrational state decreases exponentially with the difference in the energy levels of the excited and ground state, producing a very weak anti-Stokes Raman signal [56]. The relative population in excite state(N_1) and ground state (N_0) for $h(\nu_1 - \nu_0) = h\Delta\nu$ is given by the Boltzmann distribution:

$$\frac{N_1}{N_0} = e^{(-\frac{h\Delta\nu}{kT})} \tag{2.33}$$

According to eq. 2.33, probability of Stokes transition is higher than the anti-Stokes transition at thermal equilibrium. The intensity ratio depends on the temperature according to the Boltzmann distribution.

$$\frac{I_S}{I_{AS}} \propto e^{(-\frac{h\Delta\nu}{kT})} >> 1 \tag{2.34}$$

Figure 2.13 depicts the scattering of light from a molecule and Figure 2.14 represents the relative Raman intensities of Rayleigh, Stokes and anti-Stokes vibrations. For example, the Raman spectrum of silicon chip has a strong first order Raman band at 520 cm^{-1} and a weak second order band at 960 cm^{-1} in Stokes region. However, in the anti-Stokes region only first order band is visible at - 520 cm^{-1} with a very low intensity compared to the 520 cm^{-1} peak [81]. Thus, the Stokes region is most important from a practical point of view because of higher signal-to-noise ratio.

11: Stokes and anti-Stokes lines provide the same information about a molecule, however, under normal conditions, it is customary to measure only the Stokes side of the spectrum, due to the high signal-to-noise ratio.

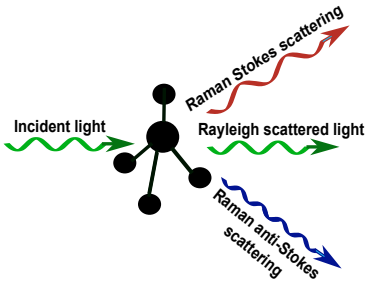


Figure 2.13: Scattering of light.

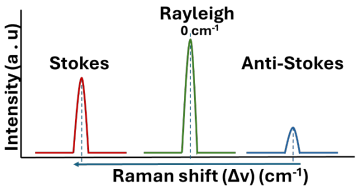


Figure 2.14: Relative Raman intensities for Rayleigh, Stokes and anti-Stokes vibrations.

2.7 Classical description of Raman scattering

In classical mechanics [56, 82], Raman scattering is elucidated by considering the interaction between a diatomic molecule and a variable electric field 'E'(Figure 2.15), represented as

$$E = E_0 cos 2\pi\nu_0 t \tag{2.35}$$

here ν_0 is the incident frequency from the laser and E_0 is the amplitude of the electric field. A changing electric field induces an electric dipole moment P in the diatomic molecule :

$$P = \alpha E \tag{2.36}$$

where α represents the "polarizability" a proportionality constant. It indicates the tendency of a molecule to acquire dipole moment under the influence of an external electric field by distorting its electronic cloud.

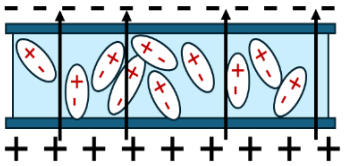


Figure 2.15: Polarization of a dielectric material under an external electric field.

This interaction between the electric field and induced dipole produces a Raman shift. Molecular polarizability is influenced by the frequency and orientation of the electric field as well as the frequency of the molecular vibrations. Substituting E from eq. 2.35 into eq. 2.36

$$P = \alpha E_0 cos 2\pi \nu_0 t \tag{2.37}$$

if the molecule starts vibrating at a frequency ν_n with an amplitude Q_0, its nuclear displacement Q can be expressed as:

$$Q = Q_0 cos 2\pi \nu_n t \tag{2.38}$$

α can be expressed in form of a Taylor series:

$$\alpha = \alpha_0 + \left(\frac{\partial \alpha}{\partial Q}\right)_0 .Q + \text{higher terms} \tag{2.39}$$

If the amplitude of vibration (Q_0) is very small, α is a linear function of Q by ignoring the higher terms of Taylor series. α_0 is the polarizability in the ground state and $\left(\frac{\partial \alpha}{\partial Q}\right)_0$ represents the rate change of the polarizability α with respect to the change in the nuclear displacement (Q) from the equilibrium position. Now, insert this α from eq. 2.39 in eq. 2.37:

$$P = [\alpha_0 + \left(\frac{\partial \alpha}{\partial Q}\right)_0 .Q].E_0 cos 2\pi \nu_0 t$$

$$P = \alpha_0 E_0 cos 2\pi \nu_0 t + \left(\frac{\partial \alpha}{\partial Q}\right)_0 .Q.E_0 cos 2\pi \nu_0 t \tag{2.40}$$

replace value of Q in eq. 2.40 from eq. 2.38: [12]

12:

$$cos(A)cos(B) = \frac{cos(A+B) + cos(A-B)}{2}$$

$$P = \alpha_0 E_0 cos 2\pi \nu_0 t + \left(\frac{\partial \alpha}{\partial Q}\right)_0 .E_0 cos 2\pi \nu_0 t.Q_0 cos 2\pi \nu_n t$$

$$P = \underbrace{\alpha_0 E_0 cos 2\pi \nu_0 t}_{\text{Rayleigh}} + \underbrace{\frac{1}{2}\left(\frac{\partial \alpha}{\partial Q}\right)_0 .Q_0.E_0[\overbrace{cos 2\pi(\nu_0 + \nu_n)t}^{\text{anti-Stokes}} + \overbrace{cos 2\pi(\nu_0 - \nu_n)t}^{\text{Stokes}}]}_{\text{Raman}} \tag{2.41}$$

According to classical theory, eq. 2.41 represents the complete dipole moment due to Raman and Rayleigh scattering. The first term in the equation is indicative of Rayleigh scattering, emitting light at frequency ν_0, while, later terms of the equation represent anti-Stokes and Stokes shifts of the Raman scattering, demonstrating the gain and loss in energy of the scattered photon [56, 82–84].
The intensity of the Stokes Raman scattering I_R can be expressed in a form reminiscent of the Beer-Lambert Law, eq 2.13, $A = \epsilon C d$. If I_0 and ν_0 represent intensity and frequency of the incident laser light, respectively. Then I_R is given by

$$I_R = K I_0 \sigma_R C V \tag{2.42}$$

Where C is the concentration of the scattering molecules, V is the volume from which the scattered light is collected and K is the proportionality

constant and depends upon the experimental conditions of Raman measurements. The coefficient σ_R[13] is the Raman scattering cross section and it can be used to calculate the ability of a molecule to scatter light[56, 85]:

$$\sigma_R \propto (\nu_0 - \Delta V)^4 \left(\frac{\partial \alpha}{\partial Q}\right)^2 \tag{2.43}$$

13: In absorption, absorptivity ϵ, and Raman cross section σ_R in Raman scattering play almost similar role and are a measure of probability of a particular transition to occur. They represent intrinsic property of a molecule of light absorption or scattering.

2.7.1 Selection rules

Raman spectroscopy can identify a molecular species from its spectral fingerprint. However, all vibrational modes can not be observed by Raman spectroscopy[14]. If the polarizability of the molecule is changed during the vibration, that vibration is Raman active. According to eq. 2.41, the vibration changing the polarizability α with the nuclear displacement Q should not be zero. i.e.

$$\frac{\partial \alpha}{\partial Q} \neq 0 \tag{2.44}$$

because from eq. 2.43

$$\sigma_R \propto \left(\frac{\partial \alpha}{\partial Q}\right)^2 \tag{2.45}$$

14: Vibrational modes that can undergo Raman transition are called Raman active, while transitions that are not permissible are Raman inactive.

If the molecular vibrations do not change polarizability α, the vibration will be Raman inactive[15]. This selection rule contrasts to infrared spectroscopy (IR spectroscopy), as both techniques can be used to probe vibrational modes. The cross section of vibrational absorption σ_{IR} in terms of dipole moment is

$$\sigma_{IR} \propto \left(\frac{\partial P}{\partial Q}\right)^2 \neq 0 \tag{2.46}$$

15: If the molecule has a change in dipole moment during vibration, then it is said to be IR active and if vibrations are associated with the change in polarizability, those are Raman active.

and the corresponding vibration will be IR active. If there is no change in dipole moment, no absorption will occur [56, 83].

There are some general observations about IR and Raman spectroscopy:

- Symmetric vibrations have relatively strong Raman signal and are not IR active.
- Asymmetric vibrations generate a weak Raman signal but strong IR signal.
- Bending vibrations generate a weaker Raman signal but strong IR signal.
- A molecule can be both Raman and IR active [83].

2.7.2 Pros and cons of Raman spectroscopy

Raman spectroscopy offers a range of advantages and disadvantages Figure 2.16, which are crucial to consider when deciding its applicability in scientific research and analysis. While it provides significant benefits through its non-invasive and detailed spectral information, it also presents several challenges that can affect its effectiveness in certain applications.

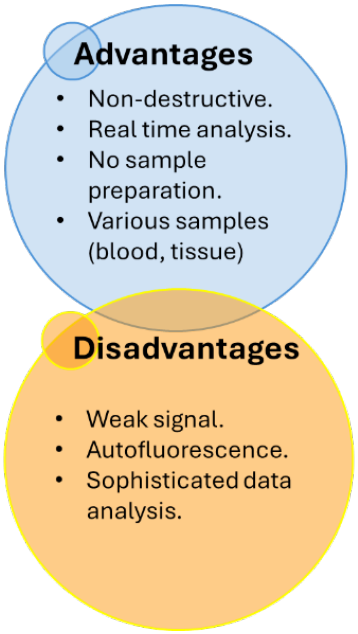


Figure 2.16: Pros and cons of Raman spectroscopy

Pros of Raman spectroscopy.

- No specific sample preparation is required.
- It is a non-destructive technique and provides control over the laser intensity.
- Can be used either with microscope or fiber optic probe.
- Water is a weak Raman scatterer enabling measurements in native state of the samples.
- Raman spectra of full spectral region from 50 cm^{-1} to 3800 cm^{-1} provide comprehensive chemical profiling [56, 83].

Cons

- High laser power is required to obtain a very weak Raman signal.
- High fluorescence background due to absorption.
- Expensive instrumentation.

In tissue, incident light undergoes multiple interactions that complicate the analysis of large specimens. As light travels or is transmitted through such turbid media, it experiences specular reflection, refraction, scattering, transmission and absorption [86]. When an oblique ray of incident light passes through a turbid medium such as from a tissue, some of the photons may pass through the sample without any significant scattering, such as in transmission. These photons maintain the direction and phase of the incident light, travel the shortest path and reach the detector earliest. These photons are called ballistic photons. In addition to these, some of the photons, called snake photons, are slightly scattered in the direction of the incident beam and are transmitted after the ballistic photons. Most photons become completely diffused, i.e. they experience large scattering angles and lose their original phase and polarization state and carry no information related to the incident beam. Diffused photons exiting the tissue on the same side as the illumination source, at a distance from the point where light entered, contribute to diffuse reflectance as shown in Figure 2.17 [56, 87]. A photon is scattered several times before it gets completely diffused. The average distance travelled by a photon before getting diffused is called the transport mean free path. An intermediate regime, termed quasi-diffusive regime also exists, where photons preserve a small information of their original phase and direction [56]. However, typically, the dimension of small animals, organs or tissue under observation is much larger than the scattering mean free path. Thus diffused photons are much greater in number than ballistic ones. To estimate the path of a photon within a biological tissue over time, different theories have been proposed.

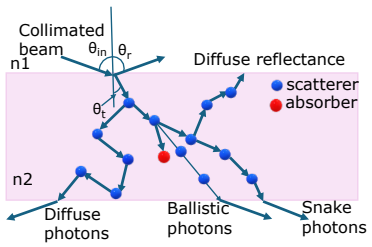


Figure 2.17: Multiple scattering of a collimated light beam in a diffusive medium.

2.8 Radiative transport equation

The Radiative Transfer Equation (RTE) is a fundamental framework for describing light propagation in biological tissues. It accounts for all the phenomena affecting light propagation, including absorption, elastic scattering, and the angular distribution of scattered light. This equation is based on the principle of energy conservation and is also termed as the Boltzmann equation [87]. The intensity of light per unit solid angle is

called radiance. RTE describes changes in radiance $L(\vec{r}, \hat{s}, t)$ at a point in the medium as a function of position and direction. The equation balances the loss of radiance due to absorption (negative terms), the gain in radiance (positive terms) from the sources within the medium, such as fluorescence. It is a differential equation and it states that when a beam of light is travelling inside a tissue or a turbid media, it loses energy through divergence and extinction, including both absorption and scattering away from the beam. It also gains some energy due to the light source within the medium and the scattering of light into the direction of interest. Polarization, non-linear effects and coherence and inelastic collisions are neglected while deriving RTE. Figure 2.18 represents the schematics of energy flowing through a unit area dA from direction $\hat{s}'$ to $\hat{s}$ within solid angle $d\Omega$. Optical properties such as scattering coefficient μ_s, coefficient of absorption μ_a, anisotropy g and refractive index n are considered time invariant but they can vary spatially[88, 89]. RTE equations describe the following events in the tissue when light enters.

1. $\hat{s}$ directed photons being absorbed or scattered. Negative sign represents loss of photons through attenuation.
2. Spatial variation of radiance along $\hat{s}$ direction at position $\vec{r}$. This term is actually the boundary conditions converted to a position term. This describes how much light enters a small volume versus exiting the same volume.
3. Photons scattered into direction $\hat{s}$ from all other directions $\hat{s}'$. The phase function is an important parameter for this.
4. Photons emitted from the source in direction $\hat{s}$.

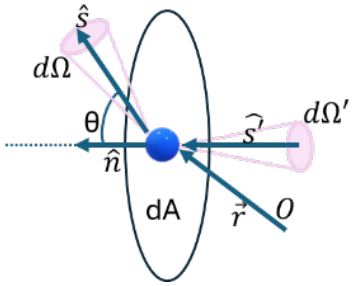


Figure 2.18: Schematic of energy flowing through a surface area dA at position $\vec{r}$ from direction $\hat{s}'$ to $\hat{s}$ within solid angle $d\Omega$. $\hat{n}$ represents the vector normal to the surface area dA.

Thus RTE can be written as:

$$\frac{\partial L(\vec{r},\hat{s},t)}{\partial t}\frac{1}{c} = -\underbrace{\mu_t L(\vec{r},\hat{s},t)}_{1} \pm \underbrace{\nabla L(\vec{r},\hat{s},t)\hat{s}}_{2} + \underbrace{\mu_s \int_{4\pi} \Phi(\hat{s}',\hat{s}) L(\vec{r},\hat{s}',t) d\Omega'}_{3} + \underbrace{S(\vec{r},\hat{s},t)}_{4} \tag{2.47}$$

Where

- $L(\vec{r}, \hat{s}, t)$ is radiance at position $\vec{r}$, in direction $\hat{s}$ and at time t.
- c is the speed of light inside the medium.
- $\mu_t = \mu_a + \mu_s$ is the tissue attenuation coefficient.
- ∇ is the gradient operator. It represents the rate of change of the radiance $L(\vec{r}, \hat{s}, t)$ in space, and the direction of the gradient points in the direction of the fastest increase in radiance. It is positive when the radiance is increasing along the direction of the light propagation and negative when the radiance is decreasing in the direction of propagation due to light being absorbed or scattered.
- $\Phi(\hat{s}', \hat{s})$ represents the phase function that describes probability of the light propagating in direction $\hat{s}'$ being scattered in direction $\hat{s}$ in the form of a solid angle $d\Omega$.
- $S(\vec{r}, \hat{s}, t)$ describe light source [56, 88–90].

RTE has six independent variables to define spatial and temporal radiance at any point (x, y, z from $\vec{r}$, θ and ϕ from $\hat{s}$ and t). To derive a solution for RTE, one has to reduce the number of independent variables by some reasonable assumptions, which will result in diffusion theory.

2.9 Diffusion equation

Diffusion equation (DE) is an approximation of RTE primarily used to study light propagation within tissue. It is a partial differential equation and describes the spatial and temporal propagation of radiation energy. It has two main assumptions:

- The medium considered is highly scattering. The probability of scattering must be higher than for absorption, $\mu_s' >> \mu_a$, allowing light to diffuse through multiple scattering. After multiple scattering events, the radiance is almost isotropic with a small gradient term. This phenomenon is sometimes called directional broadening.
- In a scattering medium, the change in the photon flux density is less than unity over one transport mean free path [90].

Thus, the number of independent variables is reduced in DE because light propagation is considered to be almost isotropic, hence directional dependency, apart from a gradient term, is neglected. For nearly isotropic radiance (anisotropic factor would be g=0) L is expanded to first order of spherical harmonics. The general form of diffusion equation is

$$\frac{\partial \Phi(r,t)}{\partial t} = D\nabla^2\Phi(r,t) \tag{2.48}$$

Where $\partial\Phi(r,t)$ represents quantity of interest, which in case of tissue is fluence rate at position r and time t. D represents the diffusion coefficient [16] and ∇^2 the Laplacian operator. When this DE is generalized to diffusion of light in tissue, the equation is modified to account for the unique properties of biological media, such as scattering and absorption. If $S(\overrightarrow{r},t)$ is light source in the medium, the photon diffusion equation would be:

$$\frac{1}{c}\frac{\partial \Phi(\overrightarrow{r},t)}{\partial t} + \mu_a\Phi(\overrightarrow{r},t) - D\nabla^2\Phi(\overrightarrow{r},t) = S(\overrightarrow{r},t) \tag{2.49}$$

This DE for tissue media (eq. 2.49) describes diffusion, absorption and interaction of light with scattering particles within a tissue. Effective attenuation coefficient μ_t' and terms involving distances can help to apply boundary conditions to understand spatial distribution of light. By incorporating these factors, the diffusion equation provides a robust framework for optical imaging and therapeutic applications, enabling non-invasive diagnostics and targeted treatments [56, 68, 90, 91].

16: Diffusion coefficient = $D = \frac{1}{3(\mu_a + \mu_s')}$

2.10 Light attenuation factor in tissue

Tissues possess complex structure. The propagation of light in tissue is similar to a random walk and is governed by the absorption and scattering occurring side by side with the inhomogeneity of the medium. Incident light scatters and changes direction multiple times before emerging from the tissue or getting absorbed [59]. In tissue, scattering, absorption and correlated effects are scrambled together, and the distance traveled by the incident light within the tissue sample is many times the size of the sample. This dynamic interaction affects both the penetration depth and the diffusion path of photons within the tissue matrix, profoundly

Table 2.1: Important biological chromophores that contributes to the absorption of light

Chromophores	Spectral range(nm)	Reference
Nucleic acid	210 nm, 260 nm to 270 nm	[92]
Aminoacid	Around 185 nm to 320 nm	[93]
Flavins	365 nm and around 450 nm	[94]
Bilirubin	Around 440 to 450 nm	[95, 96]
Carotenoids	450 - 500 nm	[97]
Lipids	760, 920, 1000 to 1350 nm and 1550 to 1870 nm	[98, 99]
Porphyrins	380-750 nm	[100, 101]
Oxyhaemoglobin(HbO2)	Around 415, 542, 577 nm	[102]
Deoxyhaemoglobin (Hb)	Around 430 nm and 555 nm	[102]
Proteins	260 - 300 nm, 1400 to 1700 nm, above 3000 nm	[103–106]
Melanin	400 to 700 nm	[107, 108]
Water	Around 980 nm and beyond 1200 nm, 1450 nm, 1950 nm	[109, 110]

influencing the diagnostic capabilities of optical methods [55]. Although each tissue has its unique optical signal, it is often feasible to approximate the optical properties of the tissue, based on some general components such as water, which makes 75% of the body mass and is a major chromophore. Water absorbs strongly in the NIR region ($\lambda > 1400nm$) [111]. Blood has pronounced absorption in the visible region of the spectrum due to bilirubin and haemoglobin. Therefore, absorption in tissue is usually dominated by water and blood. Other chromophores that absorb light at specific wavelengths (in lower part of the visible and UV region ($\lambda < 500nm$)) are melanin and proteins. Table 2.1 lists some common absorbers found in tissue, along with the wavelength ranges at which they exhibit high absorption [58]. Blood also has significant scattering properties. It contains disc-shaped red blood cells (RBCs) and other cellular components suspended in plasma. RBC has thickness around 2 μm and diameter 7 to 9 μm. The size and refractive index mismatches between these components and the surrounding plasma cause significant scattering of light. The volume of RBC and the extent to which they agglomerate also affect the scattering properties. This scattering contributes to the red color of blood and affects the way light propagates through tissues where blood is the primary scatterer [55, 57, 59]. This combined effect of scattering and absorption in tissue can be explained with an interaction or total attenuation coefficient μ_t, which is the decrease in incident energy per unit area due to absorption and scattering, as the distance traversed within tissue is increased.

$$\mu_t = \mu_a + \mu_s \tag{2.50}$$

Optical window

The therapeutic window is the region of the spectrum which has minimal absorption and reduced scattering. It refers to the light wavelengths that can penetrate the tissue deeper, mainly spanning between 700 and 1100 nm. Additionally, there are two comparatively less transparent and narrow windows at 1300 and 1600 nm. Although there are different opinions on this, it is crucial to consider the effective wavelength while developing a biological probe or a photoactivated drug to determine the effectiveness of particular therapeutic agents[56, 112]. Figure 2.19

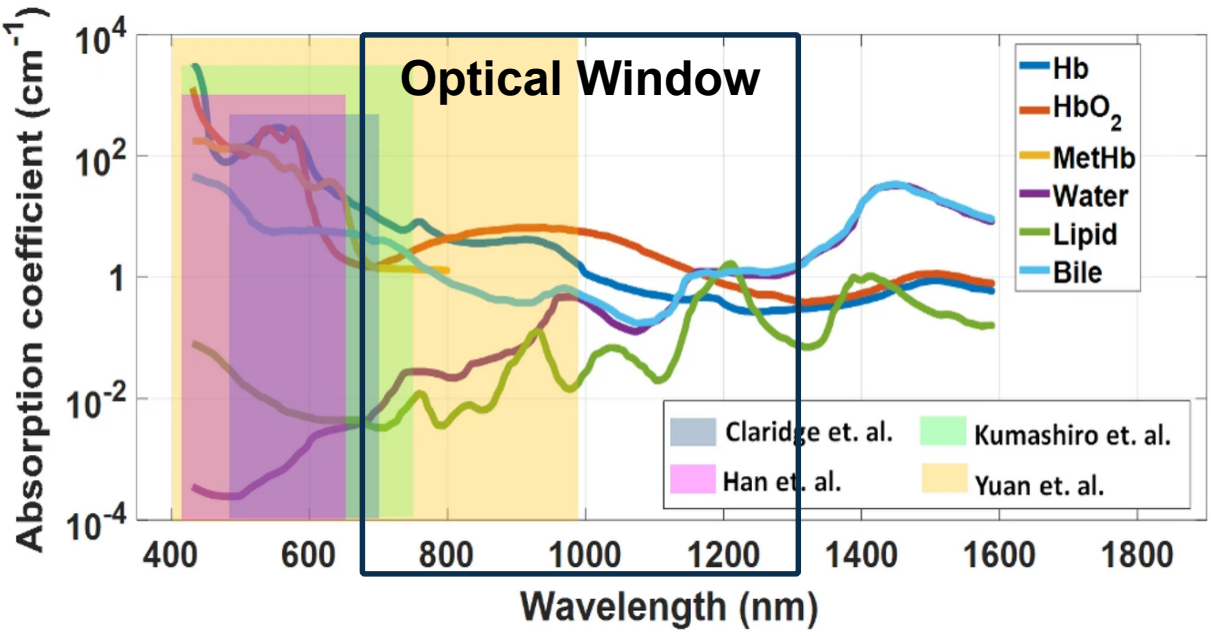


Figure 2.19: Absorption spectra of tissue chromophores and optical window [4]

presents the absorption coefficients of common components of human skin as a function of wavelength.

Conclusion

This chapter has laid the theoretical foundation for understanding the interaction of light with biological tissues. It is focused on the principles of absorption, scattering and Raman spectroscopy, as well as mathematical models such as the radiative transport and diffusion equations. These concepts are crucial for interpreting the experimental results presented in the subsequent chapters. The insights into light propagation in tissues, combined with the fundamentals of Raman scattering, directly inform the analysis and methodologies discussed in the upcoming sections of this thesis, particularly in relation to their application in early-stage cancer diagnostics using Raman and diffuse reflectance spectroscopy.

Oral Potentially Malignant Disorders and Current Diagnostic Approach

3

Chapter 3 provides a literature review on oral diseases, specifically focusing on oral cancer (OC) and oral potentially malignant disorders (OPMD). This chapter builds upon the fundamental concepts of light-tissue interaction discussed in Chapter 2, which is crucial for understanding the principles behind optical diagnostic techniques. This chapter reports the progress and aetiology of oral cancer and OPMD, along with the clinical challenges in diagnosing these conditions. The main focus is on existing diagnostic methods, including the gold standard procedures of biopsy and histopathology, and the integration of optical techniques to improve diagnostic precision. In addition, this chapter explores other diagnostic modalities, including chemiluminescence, molecular and biochemical assays, mass spectrometry, and nuclear magnetic resonance spectroscopy, which offer complementary approaches for the detection and monitoring of oral diseases. This review also sets the stage for Chapter 4, which will focus on the application of optical spectroscopy in oral diagnostics, highlighting the gap in current technology and potential of optical techniques for early detection and non-invasive screening.

3.1 Oral Cancer

Oral squamous cell carcinoma (OSCC) is the most common cancer that occurs in the oral cavity. It usually arises from surface epithelial cells of the oral cavity [113] and has a high mortality rate due to being asymptomatic until the advanced stages of the disease. It can occur in any part of the oral cavity, including the cheeks, gums, lips, tongue, floor of mouth and palate [114]. Along with other head and neck tumours it is the sixth most common cancer worldwide. According to a recent report by the World Health Organization (WHO), the number of new cases of lip and oral cavity cancer in 2022 was approximately 0.38 million, which is estimated to increase to 0.47 million by 2030 [16].

The incidence of oral cancer varies geographically as shown in Figure 3.1. More than 40% of oral cancer cases are in developing Asian countries including India, Pakistan, Bangladesh and Sri Lanka [16]. The varying incidence rate of oral cancer across different regions of the world could be attributed to distinct social practices. For example, in Eastern Europe, France and America higher rate of oral cancer is associated with excessive consumption of tobacco and alcohol [115–117]. Whereas in India, Pakistan, Bangladesh and other Asian countries, tobacco and habitual chewing of paan (a combination of betel nut, lime and tobacco) are the principal cause of oral cancer [118, 119]. Chronic paan chewing leads to repeated mechanical irritation of the oral mucosa, while chemical constituents such as nicotine induce oxidative stress, DNA damage and abnormal collagen deposition [120]. Several studies have shown a strong correlation between paan chewing and increased risk of oral submucous fibrosis [121, 122] and squamous cell carcinoma [123]. The incidence rate of oral cancer is also related to the age of the subject and increases dramatically after 40

years of age [124]. Overall low survival rate is attributed to the late stage of diagnosis [125, 126]. In Ireland, approximately 50% of patients with oral cancer were diagnosed with stage IV oral cancer between 2018 and 2020, which is an increase compared to the figures for 2015-2017 (40%) [127]. Delayed diagnosis can preclude the effectiveness of a treatment and the likelihood of positive outcomes [125, 126]. Many cases of OSCC have their origin in oral potentially malignant disorders (OPMD). Patients having OPMD have a higher chance of developing oral cancer [128, 129]. To improve the prognosis and survival rate, it is crucial to diagnose and manage OPMD properly. Since the oral cavity can be easily visualized and the lesion is easily accessible, the oral cavity is a suitable location for early-stage diagnosis. However, it is alarming that only 34% of oral cancers are diagnosed at early-stage [130, 131]. Dentists and general practitioners usually find patients with early-stage cancer or potentially malignant lesions during regular check ups. That is why there is a need to enhance awareness and understanding of these disorders among this group of clinicians and to provide them with easy-to-operate, real-time diagnostic systems, methods and possibly with a standardized screening approach that can help them decide the next course of action [132].

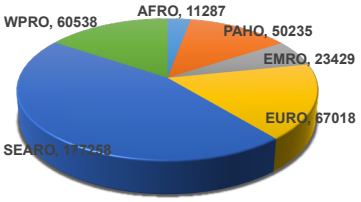


Figure 3.1: Estimated number of new cases of oral cancer in 2022 (both genders, all ages)
Africa region (AFRO), Americas region (PAHO), East Mediterranean region (EMRO), WHO Europe region (EURO), South-East Asia region (SEAR), Western Pacific region (WPRO)
All data sourced from the International Agency for research on Cancer, World Health Organization [16].

3.2 Oral potentially malignant disease (OPMD)

Warnakulasuriya *et.al.* [133] describes OPMD as a group of mucosal disorders in which morphologically altered tissue has a higher chance of developing OSCC. The concept of "precancer" was introduced in 1805, when a European panel of physicians suggested that there are some benign lesions that can develop into invasive malignant lesions if left untreated for a long time [133, 134]. In 1964, WHO began using the term "precancerous lesions" implying that OC evolves at the same location as a precancer lesion. This term was further categorized as "precancerous lesion" (including leukoplakia and erythroplakia) and "precancerous condition" (including oral submucous fibrosis, oral lichen planus (OLP), actinic keratosis etc.) [135]. However, in 2005 WHO coined a more comprehensive term of oral potentially malignant disorder (OPMD) [1], considering the fact, that only some of these lesions progress to cancer contrary to what the term 'precancerous' would suggest and that the cancer can not only appear at the affected location but anywhere in the oral cavity. This new terminology resulted in ruling out various terms used previously such as "pre-cancer", "pre-malignant", "intraepithelial lesion" and "epithelial precursor lesions" [136]. This acknowledged the fact that patients having any mucosal disorder or a specific defined lesion could have a higher potential of developing oral cancer in any part of the oral cavity in their life time. OPMD encompass a wide group of clinical conditions including leukoplakia, erythroplakia, oral lichen planus, actinic keratosis etc. [133]. Most of these OPMD lesions share similar characteristics at molecular, genomics and cytological level [136].

1: WHO classification of OPMD 2020 [136]

- Erythroplakia
- Leukoplakia
- Oral Lichen Planus (OLP)
- Proliferative Verrucous Leukoplakia (PVL)
- Oral Submucous Fibrosis (OSF)
- Actinic Cheilitis (AC)
- Discoid Lupus Erythematosus (DLE)
- Oral Graft Versus Host Disease (OGVHD).

3.3 Aetiological factors

Several lifestyle and biological factors can contribute to the development and progression of OPMD or oral cancer. The most important and well-understood aetiological factors are as follows:

- use of tobacco, including smoking cigarettes, cigars, pipes, chewing tobacco and snuff.
- heavy alcohol or use of alcohol containing mouthwash.
- excessive exposure to sunlight on the lips.
- chewing betel quid or areca nuts.
- week immune system.
- trauma and dental irritation.
- genetic factor.

Alcohol consumption and tobacco use are the primary risk factors for oral cancer; however, anatomical location of OPMD and oral cancer may also be influenced by specific aetiological factor to some extent. For instance, tobacco smoking shows the high relative risk for OSCC in the retromolar region, followed by the floor of the mouth, whereas alcohol consumption has higher relative risk for OSCC in the floor of the mouth [137, 138]. Excess sun exposure is associated with the development of malignancy on the lips [139]. Trauma, dental irritation, and chewing, betel, quid or areca nut can be highly associated with cancer or OPMD on the cheek mucosa [123]. Human papillomavirus (HPV)is also a common causative agent of OSCC in the oropharynx ; however, it cannot be associated with cancer in the oral cavity[139].

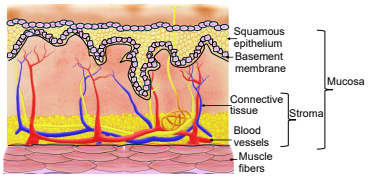


Figure 3.2: Oral mucosa consist of stratified squamous epithelium and underlying layers of connective tissue lamina propria. These two layers are separated by basement membrane from the submucosa. Most oral lesions or OPMD start from the top epithelium membrane.

3.4 Disease progression and presentation

Oral cancer usually arises in the squamous epithelium Figure 3.2. Cancer development is a multistep process [141]. Prolonged exposure to an external irritant results in increased cell proliferation, change in size of cells, nucleus enlargement and reduced cytoplasmic volume. Mutated cells escape immune surveillance and proliferate without control. This cell proliferation is called hyperplasia. If the irritant continues to affect the tissue, it causes degradation and mucosal atrophy (thinning of the epithelium) and eventually it will lead to malignant transformation or apoptosis. Cellular atypia (cellular abnormality) is typically a first sign and the further appearance of dysplasia confirms disease progression and helps gauge the risk of malignancy. [142].
In the early-stages of oral cancer many lesions may present as flat white (leukoplakic) or red (erythroplakic), or exophytic (verrucous) lesions. These lesions are usually painless but discomforting and may show variable degrees of dysplasia, which could be associated with a potential premalignancy. At a later stage, the lesion can present itself as a painful non-healing ulcer. Patients with oral potentially malignant disorder (OPMD) are more likely to develop invasive carcinomas (a precursor lesion of OSCC) compared to people with healthy mucosa [113, 128]. Leucoplakia and erythroplakia are medical terms that describe the physical appearance of the lesion, however, dysplasia is a term used in histology to describe changes in the cell morphology and the structure

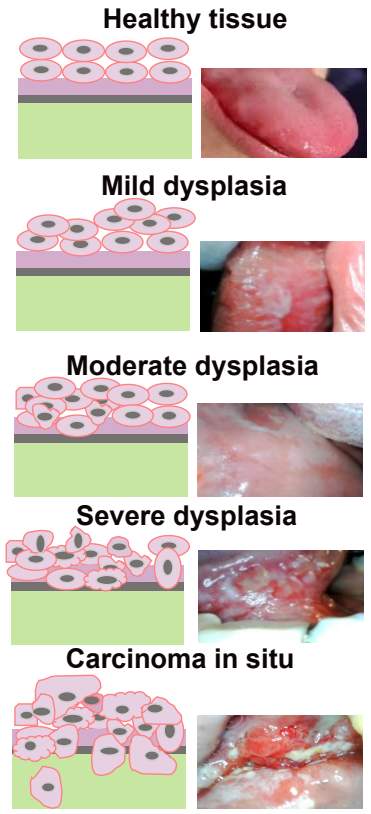


Figure 3.3: Structural alterations at different stages of dysplasia during carcinogenesis. [140]

of tissues. It is characterized by cellular irregularities and structural alteration within layers of epithelium, as shown in Figure 3.3.

Pathologists typically employ a three level grading system to categorize dysplasia: mild, moderate, severe dysplasia depending on the alteration in structural morphology and cellular structure. When dysplasia progresses to its most advanced form, it is termed carcinoma in situ [143, 144]. Table 3.1 describes the details of this classification. Almost all cases of erythroplakia and 5% to 25% of leukoplakia represent some degree of dysplasia. However, these histological classifications are subjective and semi-quantitative and are challenging to use routinely. The Smith and Pindborg grading system in Table 3.1 provides a quantitative measure of the severity of epithelial dysplasia based on specific histological features. It assigns scores to parameters such as mitotic activity, cellular atypia and architectural disturbances. Based on this score, the lesion is then classified into different grades - mild, moderate or severe, providing a standardized method for assessing the risk of malignant transformation. To reduce subjectivity in grading of dysplasia, WHO has introduced a binary system for classification of dysplasia consisting of "low risk" (negative or mild) and "high risk" (moderate or severe) dysplasia for better clinical use and reproducibility [132].

Table 3.1: Grading of various stages of dysplasia with Smith and Pindborg scoring [145].

WHO dysplasia grading	Structural alterations	Smith and Pindborg grading system scores
Squamous hyperplasia	Increased cellularity with preserved structural morphology	
No dysplasia	No Changes in structural morphology	0-10
Mild dysplasia	Extended to lower third of epithelium with mild cellular atypia	11-25
Moderate dysplasia	extended to middle third of epithelium with moderate cellular atypia	26-45
Severe dysplasia	extended to two third of epithelium with moderate or severe cellular atypia	Above 45
Carcinoma in situ	full-thickness epithelial atypia with structural irregularities; Stage 0 oral cancer	

These structural and cellular changes in OPMD indicate a higher likelihood of the lesion becoming malignant. Some healthcare professionals consider the severity of dysplasia as a gold standard, where lesions with severe dysplasia have a higher probability of developing to malignancy. However, the presence of severe dysplasia does not definitively indicate malignancy [128]. Therefore, it is necessary to manage the condition through regular monitoring, timely intervention and targeted treatment to prevent its progression to a malignant lesion.

During carcinoma progression, several histopathological changes can be observed, including the downward invasion of irregularly shaped malignant cells, keratinization, increased epithelial proliferation and an associated inflammatory response. Key pathological features that distinguish carcinoma from healthy tissue include nuclear irregularities and change in nuclear size and shape [146]. Currently, clinicians use the tumour-nodes-metastasis (TNM) staging method to estimate the survival and decide a treatment method. This TNM system is based on

- T: size of original Tumour
- N: indicates if cancer has spread to lymph nodes
- M: Presence of distant metastasis to different parts of the body from the primary location

Table 3.2 presents different stages of oral cancer with TNM classification, clinical presentation and estimated five-year survival rates [147–149].

Table 3.2: TNM staging of OSCC preposed by WHO

Stage	**Presentation**	**Size ; depth**	**TNM stage**	**5 year survival rate estimate**	**Potential optical contrast**
Stage I	Carcinoma in situ, not spread to other parts of body or lymph nodes. Very early-stage cancer.	< 2cm; < 5mm	T1, N0, M0	86%	High contrast- increased scattering and absorption in localized areas due to tissue hetroginity.
Stage II	Cancer is localized and has not metastasized or spread to lymph nodes	2 cm – 4cm; 5 mm – 10 mm	T2, N0, M0	61%	Moderate contrast-tumour growth and vascularization changes cause scattering & absorption .
Stage IIIA	Cancer has not metastasized or spread to lymph nodes	> 4 cm; > 10 mm	(T3, N0, M0)	69%	Significant optical changes - large and dense tumour structure affect scattering. Possible increased haemoglobin absorption in lymph nodes.
Stage IIIB	Cancer is localized but is present in a single lymph node close to the primary tumour	< 3cm	T1 to T3, N1, M0)		
Stage IV A	Cancer has spread to nearby tissue and to one nearby lymph node	3 cm – 6cm	T4a, N0 or N1, M0 ; or T1 to T4a, N2, M0	39%	Strong contrast from large tumour, tissue invasion & haemoglobin. Reduced contrast in distant metastasis due to tissue hetrogenity
Stage IV B	Cancer is not metastasized but present in lymph nodes or has invaded muscles and bones	> 6 cm	any T, N3, M0; or T4b,any N, M0		
Stage IV C	Cancer has spread to other parts of body				

3.5 Diagnosis

3.5.1 Gold standard diagnostic procedure-biopsy and histopathology

Table 3.1 and Table 3.2 not only present different stages of OPMD and OSCC, they also suggest that early-stage diagnosis can improve the five-year survival rate and post treatment quality of life. However, early-stage diagnosis of OPMD is challenging as these lesions often exhibit non-specific progression patterns and multifocal presentation, necessitating multiple biopsies, which are not always feasible. It is difficult to predict which potentially malignant lesion will progress to OSCC as histological grades do not always serve as a reliable indicator. For instance, severe dysplasia might not advance whereas mild dysplasia sometimes has potential to develop into OSCC [150]. At present, there is no screening program available for OPMD diagnosis or for regular follow up examinations. The gold standard diagnostic procedure for OPMD and OSCC is the same, which is a detailed physical exam and palpation, family history, followed by biopsy and histopathology. This type of diagnosis is painful, invasive, time-consuming and has its limitations. It is also subjective to sampling error [151]. Accurate diagnosis depends on factors such as biopsy site and adequate specimen submission to the laboratory, quality of specimen and correct interpretation of biopsy. Clinically innocuous oral premalignant lesions could be multifocal and can easily go undetected, or sometimes, inappropriate sampling during biopsy could result in overlooking some important features. This problem could be more intensified with the heterogeneity in the tissue [140]. Moreover, OPMD encompasses a diverse group of conditions with low to moderate risk of developing into cancer. According to a recent study, only 18% of biopsied lesions develop into cancer [152]. Therefore, time consuming, costly and invasive management strategies are not justified and should be reserved for high risk cases. Some diagnostic procedures for oral cancer and OPMD are discussed below.

3.5.2 Other diagnostic methods

3.5.3 Brush biopsy

Brush biopsy[2] is being used frequently in clinics to assess suspected lesions. It is less uncomfortable for the patient and is non-invasive. The sample is collected via a cytology brush by placing it on the tissue and rotating it by 360°. After a complete rotation sample is transferred on a glass slide along the edge. It is crucial to fix the sample within a few seconds to prevent degradation, which could compromise diagnostic accuracy [153, 154].

2: **Advantages of brush biopsy**

- Chair side test.
- Less invasive.
- Less bleeding in comparison to scalpel biopsy.
- Relatively less painful.
- can detect OPMD.

Drawbacks of brush biopsy

- Requires confirmation through scalpel biopsy.
- Time consuming.
- Topical anaesthesia may be required in case of ulcerated lesions.
- Limited to surface lesions only.
- Risk of false positive results.
- Specimen requires immediate fixation.

3.5.4 Toluidine blue (TB)

Toluidine blue is an acidophilic dye that selectively binds with nucleic acid resulting in the preferential staining of the tissue with high nucleic acid content. It is commonly used in the clinic, particularly in the case of

patients over 40 years to aid in the visualization of suspicious mucosal lesions. During the procedure, the oral cavity is rinsed with the blue dye, which is absorbed by epithelial cells. The extent of dye absorption correlates with the nucleic acid content of the cell leading to higher uptake in abnormal cells as compared to the healthy tissue [155, 156]. Although toluidine blue staining is a sensitive and minimally invasive procedure, its interpretation requires considerable experience. Moreover, the method has low specificity and is prone to false positive results [155].

3.5.5 Chemi-luminescence

This diagnostic procedure uses a chemical reaction between the target molecule (the biomarker) and a specific antibody. A catalyst linked to the antibody triggers the emission of light. The intensity and duration of the emitted light corresponds to the concentration of the target molecule. **ViziLite** has been used widely in the clinical settings, which works on the same principle and can be used with TB or acetic acid[3], because of nuclear enlargement, cancerous cells have a higher nuclear material to cytoplasmic ratio. These cells or suspected lesions, when absorb the dye and illuminated by ViziLite stick, appear white due to chemi-luminescence [160]. Although this technique is easy to use, rapid and cost-effective, it exhibits relatively low sensitivity and is prone to false-positive results. Moreover, it can not differentiate between grades of dysplasia and cancer. Figure 3.4 represents the basic principle of chemiluminescence.

Velscope is another light-based detection technique. It can detect lesions without staining. This system uses a blue light in the range 400 nm to 460 nm to excite epithelial cells. Excitation induces autofluorescence that differs between normal and pathological tissues. Increased metabolic activity in pathological tissue causes them to fluoresce less under this light and they appear dark as compared to healthy tissue [161]. VELscope operates by detecting autofluorescence from flavin adenine dinucleotides (FAD) and nicotinamide adenine dinucleotides (NADH). This technique is non-invasive, cost efficient and has potential to serve as screening tool but it also suffers form low sensitivity and cannot detect inflammatory lesions, ulcers, keratosis and lichen planus. It can be used only as an adjuvant diagnostic technique [162, 163]. Table 3.3 represents some of the non invasive or minimally invasive devices that are being used commercially for oral cancer and OPMD diagnosis.

3: Acetic acid is also sometimes used in the clinical setting for diagnosing OPMD. The area of interest is rinsed or swabbed with 5 % solution of acetic acid. Acetic acid dehydrates cells and coagulates nuclear proteins [158, 159]. Under **ViziLite** abnormal tissues often appear aceto-white, while normal tissues remain relatively unchanged. However, due to low specificity compared to TB, its utility is limited, and TB is considered far more superior than Acetic acid [158]

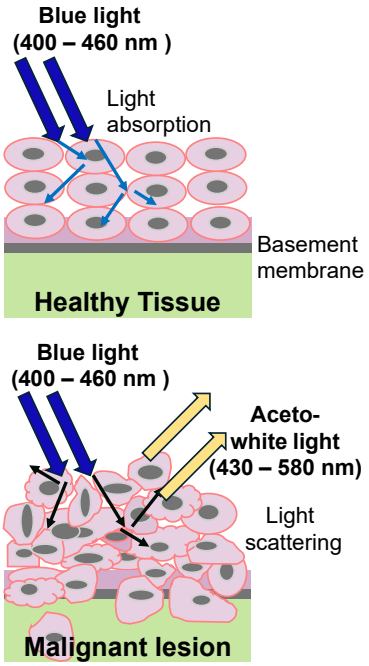


Figure 3.4: Mechanism used in chemiluminescence. Healthy tissue absorbs blue light (452–489 nm) due to the presence of natural chromophores (e.g., NADH, FAD, and haemoglobin). Whereas, in abnormal mucosa acetic acid dehydrates the cytoplasm and coagulates proteins, leading to increased light reflection and a white appearance [157].

3.5.6 Molecular and biochemical assays

Assay-based approaches, such as PCR (polymerase chain reaction), ELISA (Enzyme-linked immunosorbent assay), and immunohistochemistry, can detect specific biomarkers associated with OPMD. For instance, PCR-based assays can identify DNA alterations [167], while ELISA measures specific proteins or antigens in saliva or blood [168]. These highly sensitive techniques require complex laboratory settings, making them less suitable for real-time or point-of-care applications. There are several saliva home test kits available. Where you collect saliva and send it to a lab to be tested for cancer or related genetic changes. One of the most

Table 3.3: Commercial devices for oral cancer diagnosis

Device	Type	Advantages	Clinical use	Drawbacks	Ref.
VELscope	Fluorescence	• Easy to use • Real time	Detection of oral mucosa abnormalities	• Limited specificity • False positive results	[162, 163]
ViziLite	Fluorescence	• Easy to use • Real time	OC screening	• Limited specificity • Requires thorough follow up	[160]
OralCDx Brush Biopsy	Brush Biopsy	• Minimally invasive • Quick sample collection	OPMD and OC diagnosis	• Limited sampling depth • Require follow-up invasive biopsy	[153, 154].
Identafy	Multimodal optical imaging	• Provides comprehensive analysis	Detection of oral lesions	• High cost • Require special training	[164]
Microscan	OCT	• High resolution imaging	Evaluation of tissue layers and abnormalities	• Limited depth • High cost	[165]
CancerDetect	Saliva based assay	• Highly sensitive	Detect change in DNA or biomolecules	• Complex laboratory setting • Skilled personal	[166]

famous at-home kits is "CancerDetect Test for Oral and Throat Cancer" [166, 169]. Although these tests are available for purchase, they are not yet FDA approved and neither is included in the recommendations of the screening program by European Oral Care in Cancer Group (EOCC) or American Dental Association (ADA).

Apart from this, recent advancements in analytical methods have also provided complementary and alternative approaches for OPMD diagnosis to improve early detection, patient outcome, and reduce reliance on invasive procedures.

3.5.7 Mass Spectrometry

Mass spectrometry (MS) is a sensitive analytical tool used to identify molecules based on their mass-to-charge ratio and for their quantification. It is particularly valuable in profiling proteins, metabolites and lipids in biofluids such as urine, saliva and blood for OPMD diagnosis [170]. Proteomics studies have revealed MMP-9 and interleukins as potential diagnostic markers. Metabolomics studies using MS have identified changes in inflammatory and energy metabolites, which are indicative of malignant transformation. MS is a sensitive technique and is capable of detecting trace level biomarkers [170, 171]. However, it comes at a high cost in equipment and maintenance terms. It also requires skilled personnel for sample preparation and data analysis.

3.5.8 Nuclear magnetic resonance spectroscopy

NMR provides detailed insights regarding the biochemical composition of biological samples by analysing magnetic properties of their atomic

nuclei. It has been widely used to study metabolic alterations associated with cancer [172]. It can be used as a non-invasive diagnostic method using saliva as a sample. However, expensive instrumentation is a limitation. Also, it is less sensitive compared to MS for detecting low concentration metabolites [172].

Table 3.4 compares some non-invasive optical approaches used for OPMD and mainly for oral cancer diagnosis. It also covers their clinical use, advantages and drawbacks. This table identifies the gaps where new technologies might fit.

Table 3.4: Some non-invasive optical techniques used for OPMD and OC diagnoses

Advanced Approach	Principle	Advantages	Clinical Use in Oral Cancer	Drawbacks	Ref.
Toluidine Blue Staining	Stains dysplastic cells	• Immediate results • Easy to use	• OC screening • OPMD diagnosis	• Low specificity • May miss some lesions	[156, 173]
Fluorescence Imaging	Autofluorescence	• High sensitivity • Detailed image	• OPMD and OC diagnosis	• Affected by tissue autofluorescence • Require specific dyes	[174–177]
Optical Coherence Tomography (OCT)	Captures cross-sectional images using light waves	• High resolution imaging	• Assessment of oral tissue • Extent of lesion	• Limited penetration depth • Expensive and complex equipment	[178]
Confocal Microscopy	High-resolution imaging	• High resolution • 3D imaging	• Evaluation of cellular morphology	• Limited depth penetration • Slower imaging process • Expensive and complex	[178–180]
Spectral Imaging	Analyses spectral profiles	• Tissue differentiation	• Characterization of changes in tissue	• Expensive and complex equipment • Complex data interpretation	[181, 182]
Mass Spectrometry (MS)	Analyse mass-to-charge ratio	• High sensitivity	• Assess change in concentration of biomarkers	• Expensive instrument • Requires skilled personnel	[170, 171]
Nuclear Magnetic Resonance (NMR)	Magnetic resonance of nuclei	• Reproduceable • Non-destructive	• Metabolic profiling	• Expensive • Less sensitive	[172]
Diffuse Reflectance Spectroscopy (DRS)	Measures tissue reflectance	• Real time tissue differentiation	• Tissue differentiation	• Less spatial resolution • Affected by tissue heterogeneity	[75, 183]
Raman Spectroscopy	Analyse Molecular vibrations	• Provide chemical information • suitable for *in vivo* and *ex vivo* analysis	• Chemical fingerprint of tissue • OPMD and OC diagnosis	• Requires specialized equipment	[184–186]

Conclusion

Given these limitations and challenges associated with the current gold standard procedure and adjuvant techniques for OPMD and OSCC diagnosis, there is a pressing need for a less invasive, real-time and more accurate diagnostic procedure for early-stage diagnosis and routine check-ups. For regular screening and disease management, straightforward and low-risk diagnostic and monitoring techniques would be preferable. Such techniques would significantly improve the prognosis and post-treatment life quality for the patient. Existing diagnostic techniques such as biopsy, histopathology together with non-invasive methods, including fluorescence imaging, confocal microscopy and OCT provide valuable insights into OPMD diagnosis, but each of these techniques have limitations. Advanced techniques like MS and NMR offer molecular information related to cancer progression, but require specialized equipment and expertise, limiting their widespread use for routine screenings. In contrast, optical techniques such as diffuse reflectance spectroscopy (DRS) and Raman spectroscopy enable real-time, non-invasive diagnosis, offering faster and more accessible alternatives for OPMD screening and diagnosis. These methods provide complementary biochemical and structural information without the need for complex sample preparation or tissue excision. A multimodal approach combining Raman spectroscopy and DRS could potentially enhance early detection and improve patient outcomes. Raman spectroscopy provides biochemical information, while DRS captures the structural and optical properties of tissue. Together these techniques offer complementary insights to improve diagnostic sensitivity and specificity of OPMD screening [34, 187]. The following chapter will explore the principles and applications of DRS and Raman spectroscopy in oral diagnostics, emphasizing their role as a valuable tool in advancing the early detection and management of oral diseases.

Optical Spectroscopy for OPMD Diagnostics

4

The early and accurate diagnosis of OPMD is crucial for decision making regarding preventive measures and treatment strategies. Non-invasive, real-time screening methods, such as saliva analysis, can aid in routine checkups for early detection and disease monitoring. Building on the principles of light–tissue interactions (Chapter 2) and the overview of oral disease progression and diagnostic approaches (Chapter 3), this chapter focuses on the application of optical spectroscopy for oral diagnostics and forms the basis of manuscripts IV and V. The ability of optical techniques to provide non-invasive, real-time insights into tissue composition and biochemical changes makes them invaluable tools in detecting and monitoring oral cancer and OPMD. In particular, Raman and Diffuse Reflectance Spectroscopy (DRS) have emerged as complementary methods that offer distinct benefits in characterising structural and molecular changes in tissues and biofluids. DRS utilises the absorption and scattering of light to provide insight into tissue morphology and composition, reflecting changes in chromophores and blood supply associated with cancerous transformations. In contrast, Raman spectroscopy offers a molecular fingerprint by detecting vibrational modes of biochemical components, enabling the identification of specific molecular alterations occurring in tissue with cancer progression. The integration of these complementary techniques facilitates a comprehensive evaluation of both structural and molecular changes in oral tissues with the onset of disease, improving diagnostic sensitivity and specificity. This chapter reviews the mechanisms, technological advances, and clinical relevance of these modalities, emphasising their potential to revolutionise oral cancer diagnostics.

4.1 DRS in oral cancer diagnosis

Diffuse Reflectance Spectroscopy (DRS) is an optical diagnostic technique. By analysing the interaction of light with biological tissues, DRS provides valuable insight into tissue morphology and biochemical composition. Its non-invasive nature and ability to probe deep tissue layers make it an attractive tool for regular screening and OPMD diagnosis. Figure 4.1 represents schematics of the layers of healthy oral mucosa.

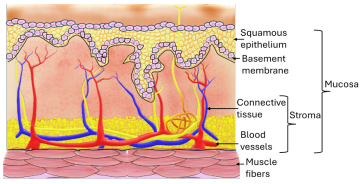


Figure 4.1: Cross section of normal mucosa.

4.1.1 Mechanisms of DRS in detecting OPMD changes

OPMD and cancerous tissue can potentially be characterised using DRS based on their unique optical properties. As tissues undergo precancerous changes, their metabolic rates often increase, requiring more oxygen and nutrients to sustain abnormal cell growth. This increased demand creates a hypoxic environment within potentially malignant tissues [188]. Hypoxia [1] triggers angiogenesis, a process in which new blood vessels are formed to deliver oxygenated blood to the affected area. However,

1: Rapid growth of malignant cells at the tumour site can outpace the formation of new blood vessels, which can lead to poor oxygen supply and cause hypoxia in the tumour. Hypoxia can also alter haemoglobin genes and oxygen-binding proteins, which further causes lower oxygenation of haemoglobin.

in cancerous tissues, angiogenesis often results in an irregular capillary network, generating a tumour-associated vascular system characterised by immature and compressed lymphatic and blood vessels [189].
As tumours grow, the disorganised vascular system cannot meet the increasing demand for nutrients and oxygen, exacerbating regional hypoxia and creating a self-perpetuating cycle of hypoxia and reoxygenation [189–191]. Changes in oxygenation, haemoglobin concentration and blood vessel density within malignant tissues are key diagnostic indicators that can be effectively captured using DRS for healthy and OPMD tissues in both *ex vivo* and *in vivo* environment [41, 192]. DRS typically operates in visible (400 - 700 nm) and near-infrared (NIR, 700-1100) regions [41]. In the visible spectrum, light absorption is dominated by deoxygenated haemoglobin (Hb) and oxygenated haemoglobin (HbO_2), both of which are abundant in blood vessels permeating the stromal tissue surrounding cells and organs. Cancerous tissues generally exhibit a higher blood content due to increased microvasculature. However, this newly formed vascular network results in inefficient oxygenation of the tumour site [189, 190]. The concentration of haemoglobin in healthy and cancerous tissues can reveal important diagnostic information. The major peaks of oxyhaemoglobin are observed at 412 nm, 542 nm, and 576 nm in both healthy and OPMD tissue. Recently, Kaniyappan et al. reported some additional peaks in DRS spectra of malignant tissue at 490 nm, 630 nm, and 690 nm. They may indicate the presence of methaemoglobin or beta-carotene, which are associated with tumour metabolism and hypoxia. Quantitatively, they have reported lower oxyhaemoglobin concentrations in malignant tissues (20.0 ± 3.0 µM) compared to healthy tissues (26.3 ± 4.3 µM), reflecting the poor oxygen supply at tumour sites caused by rapid tumour growth and insufficient vascular formation [193].

The wavelength range of 500 – 900 nm has been identified as particularly informative for diagnosing OPMD lesions, as it encompasses the absorption peaks of key chromophores, including haemoglobin and beta-carotene. Within this range, light scattering and absorption properties provide critical diagnostic information [75].

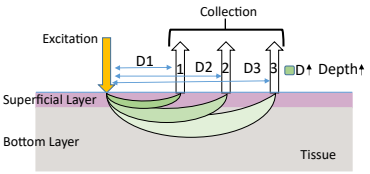


Figure 4.2: Light travelling inside the tissue. The farther the detector is from the source, the greater the distance (D) is, and light will have a greater chance to penetrate deep inside the tissue.

4.1.2 Technological advancements in DRS

Probe Configurations

One of the strengths of DRS lies in its adaptability through manipulation of the probe configuration. By adjusting the source-to-detector distance (SDD), signals can be captured from specific tissue depths as shown in Figure 4.2. It enables the distinction between superficial epithelial layers and deeper stromal layers. Optimised probe configurations can also improve diagnostic accuracy in different wavelength regions [194, 195]. For example, a study on OSCC has shown that 1 mm SDD can achieve up to 89% sensitivity, 82% specificity and 86% accuracy in detecting cancerous tissues in the 400 to 1600 nm wavelength range [196]. Current DRS systems are prone to many random and systematic errors. One of the critical factors affecting DRS measurements is the pressure exerted by the probe on the tissue. Excess pressure can alter the physiology and morphology of tissue, leading to measurement artefacts. To address this

issue, Yu et al. developed a smart optical probe capable of maintaining a consistent pressure below 1.0 psi, reducing errors and ensuring reliable measurements in oral mucosal tissues [197].

Advanced probes

Recent advancements have introduced innovative probes that enhance the diagnostic potential of DRS. For instance, Wang et al. have developed a polarisation-sensitive diffuse reflectance (rcPS-DR) probe that provides depth-selective measurements and extracts polarisation properties for comprehensive tissue analysis. This technology has demonstrated superior tissue discrimination compared to standard DRS probes. Additionally, the same group has reported designing a self-calibrating probe to measure and calibrate system responses in real-time, eliminating the effects of instrument fluctuations and improving diagnostic accuracy [198].

Integration with other techniques

Another promising innovation is the integration of DRS with other complementary diagnostic modalities. This multimodal approach has yielded significant improvements in diagnostics. For example, combining DRS with autofluorescence has achieved more than 96% accuracy in diagnoses of mucosal lesions [199]. The two modalities provide complementary information. Autofluorescence is sensitive to metabolic and biochemical changes, such as variations in NADH and FAD fluorescence, while DRS captures structural and vascular alterations through scattering and absorption properties. When the spectral data from both modalities are fused and analysed using data-driven models, such as convolutional neural networks, the combined dataset enhances tissue differentiation by jointly reflecting metabolic, morphological and haemodynamic contrasts [1, 199]. Similarly, the coupling of DRS with micro-endoscopy enhances its imaging and spectroscopic capabilities, providing a more detailed analysis of oral tissues [200]. Another exciting development is the use of multi-spectral imaging cameras to capture two-dimensional diffuse reflectance images. These systems have shown high accuracy in distinguishing healthy, OPMD, and OSCC tissues, offering a practical and efficient approach for real-time diagnostics [201].

4.1.3 Clinical relevance and practical applications

DRS holds significant clinical relevance as a non-invasive, real-time diagnostic tool for OPMD and oral cancer. Its ability to provide rapid insight into tissue physiology, haemoglobin concentration, and oxygen saturation through optical properties, makes it a valuable adjunct in the early detection of OPMD. Its compact and portable design enables bedside evaluations, making the technique particularly useful in low-resource environments where access to traditional pathology may be limited. Moreover, DRS can assist clinicians in guided biopsy, accurate surgical margin detection and monitoring treatment response by providing the contrast based on optical properties of the tissue. However, it

demonstrates limited specificity. Integrating DRS with complementary techniques such as Raman spectroscopy enhances diagnostic potential by combining structural and functional insights. DRS captures morphological and vascular changes; Raman spectroscopy provides molecular fingerprint. This multimodal approach enables more comprehensive tissue characterisation, improving tissue differentiation and disease diagnosis [1, 187].

4.1.4 Connecting experimental findings with existing literature

In this study, we observed OPMD-related changes in cellular structure, vascularization and tissue oxygenation across multiple wavelength regions. Spectral variations in 400–500 nm range likely reflect increased haemoglobin absorption, associated with enhanced vascularization and metabolic activity in the lesion [202, 203]. The 540 / 575 nm reflectance ratio, sensitive to the balance between oxy- and deoxyhemoglobin, was highest in OPMD tissue, indicating changes in haemoglobin levels and blood oxygenation due to angiogenesis and OPMD progression. In the near-infrared region (700–1100 nm), spectral variations corresponded to structural modification in deeper tissue layers, affecting light scattering and absorption. The contralateral healthy appearing tissue (OPMDH) displayed intermediate spectral behaviour, suggesting subtle field changes that are not visually apparent. This highlights potential of DRS to detect early changes in OPMD tissue. By applying machine-learning-based feature selection using Least Absolute Shrinkage and Selection Operator (LASSO), we identified 14 spectral markers within biologically relevant spectral regions previously reported in the literature [75, 193]. These features were then used to develop a Linear Discriminant Analysis (LDA) model to classify healthy and OPMD tissue across different anatomical locations. This confirms that vascular, oxygenation and scattering changes are primary indicators of OPMD. In conclusion, this study demonstrates that DRS, combined with machine learning can effectively differentiate OPMD from healthy tissue, detect early field changes and guide biopsy decisions, supporting its potential as a non-invasive, real-time screening tool for early detection and monitoring of OPMD.

4.2 Raman spectroscopy for oral cancer diagnosis

Raman spectroscopy has shown a significant potential in detecting early biochemical changes associated with the onset of oral diseases. It can provide non-invasive and real-time molecular information about the sample. Its high sensitivity to molecular vibrations can provide subtle information about the alteration of cellular components, including proteins, lipids, and nucleic acids, which are indicative of malignant transformation [204, 205]. Many studies have shown the potential of Raman spectroscopy for cancer diagnosis and differentiating OPMD from malignant or healthy tissue with high accuracy [204, 206, 207], making it an invaluable tool in the diagnosis and monitoring of OPMD and oral cancer. Figure

4.3 illustrates the typical workflow of Raman spectroscopy to analyse biological samples, highlighting its integration from sample collection and data acquisition to machine learning-based data processing and interpretation. Jeng et al. reported Raman spectroscopy of 80 samples, including 44 tumours and 36 normal cryopreserved samples from the tongue, buccal mucosa, and gingiva. Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) were used with principal component analysis (PCA) to classify the samples. They achieved 81% accuracy (sensitivity 77% and specificity 86 %) using PCA-LDA and 87% accuracy (90 % sensitivity and 83 % specificity) with PCA-QDA [204]. Cals et al. also developed a tissue classification model based on PCA-LDA for Raman spectra of OSCC on the tongue. The accuracy achieved by this model was 86% with 100 % sensitivity and 66 % specificity [208]. In 2021, Ibrahim et al. tried to stratify OSCC, benign lesions, mild, moderate and severe dysplasia. They used 72 pathological specimens and digitally removed glass and paraffin wax background signals. The accuracy achieved to distinguish OSCC from other pathologies was 70 %. However, to distinguish mild dysplasia, the accuracy achieved was 46 %, which is very close to a guess. The main distinguishing features were reported to be decreased collagen in connective tissue, decreased lipid and protein content in the epithelium, and increased nucleic acid signal with progression of disease severity. Smoking status and inflammation were also reported to affect classification efficiency [206]. The sensitivity and specificity reported in these studies for OPMD and oral cancer diagnosis using Raman spectroscopy were encouraging when compared with conventional diagnostic approaches such as visual and physical examination (sensitivity 71 - 88 % and specificity 81 - 85 %) [209], tissue autofluorescence using VELscope (sensitivity 33 % and specificity 88 %) or brush cytology (sensitivity 97 % and specificity 84 %) [210]. Therefore, Raman spectroscopy offers comparable or superior diagnostic performance, in addition to providing label-free biochemical information that reflects the underlying molecular and structural alterations associated with malignant transformation.

Christian et al. reported diagnosing OSCC with 86% sensitivity and 94% specificity using shifted-excitation Raman difference spectroscopy (SERDS). In their Raman setup, they used a variable wavelength (770 and 810 nm) diode laser for excitation. A Glan laser prism, which is transparent to s-polarised light, was used in combination with a half-wave plate for reducing incident laser intensity. During the measurement, two data sets were recorded for each measurement, one with 783 nm and the second with a slightly shifted wavelength of 785 nm. The mean spectra at 783 nm was subtracted from the mean spectra at 785 nm. The area under the Receiver Operating Characteristic (ROC) curve (AUC) for distinguishing OSCC tissue was approximately 0.95, indicating excellent discriminatory performance [211]. In a more recent study with SERDS, Matthies et al. classified OPMD and OSCC tissue against healthy oral mucosa, with the accuracy of 95% and 89%, respectively [207].

An *in vivo* approach for analysing oral mucosa with fiber-coupled Raman spectroscopy achieved 94% accuracy in diagnosing cancer. The key Raman biomarkers associated with oral diseases include increased signals from water, protein, nucleic acid and reduced Raman signal for lipids. Theses changes are obvious in the fingerprint region, as well as in the high wavenumber region of the Raman spectrum [212]. The fingerprint

region, spanning $200cm^{-1} - 1800cm^{-1}$ is rich in biochemical information corresponding to proteins, lipids, carbohydrates, nucleic acids and amino acids. Usually, healthy tissue spectra have lipid-dominated peaks, whereas malignant tissue spectra have protein-dominated peaks [204, 205]. This shift occurs because of an increase in protein content due to increased metabolic activity and protein synthesis in malignant cells. These malignant or OPMD cells have a higher amount of surface proteins in the cells, such as antigens, enzymes, antibodies and receptor proteins [205, 213]. Raman studies in oral cancer usually capture signals from the top surface layers. That is why malignant and pre-malignant tissue represents higher protein concentrations. In addition, cancerous and OPMD cells often have a disruption in the cell membrane, which alters their lipid content and causes a decrease in lipid signal [205, 214]. The signal from the bilayer membrane of healthy tissue produces a lipid enriched signal, giving rise to a weak 1655 cm^{-1} band due to C = C of lipids. However, strong and broad spectral features are observed in malignant tissue spectra at 1650 cm^{-1} attributed to amide I and at 1450 cm^{-1} due to CH_2 bending, broad peaks of amide III at 1200 to 1350 cm^{-1} and a sharp peak around 1003 cm^{-1} are all contributed by proteins [204]. Chen et al. highlighted that keratin could serve as a promising biomarker for OPMD and oral cancer diagnosis. It is a major component of fibrous structural proteins and change in its concentration can represent malignant transformation. They developed an automated diagnostic approach using Raman spectroscopy combined with multivariate curve resolution–alternating least squares (MCR-ALS) to detect keratin in oral tissue samples. This model decomposed spectra into spectrally interpretable components. By extracting keratin spectral profile from complex tissue spectra, the method achieved 92 % sensitivity and 100 % specificity in distinguishing OSCC from healthy tissue. The keratin signal in the tissue was characterized by Raman bands of prominent structural proteins, including tyrosine (830 cm^{-1} and 850 cm^{-1}), C - C stretch (890 cm^{-1} and 937 cm^{-1}), phenylalanine peak (1003 cm^{-1}), Amide III (1200 to 1350 cm^{-1}) and CH bend (1450 cm^{-1}) and Amide I (1650 cm^{-1}). These spectral features were more pronounced in OSCC tissues, reflecting increased keratin associated with malignant transformation [215].
Dai et al. used oral cancer cells and normal fibroblast cell lines to confirm the potential of adenine as a biomarker in the diagnosis of oesophageal cancer. A significant difference in the adenine signal at $735cm^{-1}$ was reported between both cell lines [216]. Another study reported a higher water content in OSCC tissue compared to normal tissue. Barosso et al. observed a broad water absorption signal from 3350 to 3550 cm^{-1} significantly higher in OSCC tissue than in the healthy surrounding tissue. To determine the water content in the tissue, they took the water to protein (2910 to 2960 cm^{-1}) ratio. They reported that this method of determining water concentration could help to determine sharp cancer margins Figure 4.4 during surgery to prevent cancer recurrence [217]. When the no cancer cells on the edge of the biopsy sample, the margin is described as negative or clean, suggesting that the cancer has been removed completely.

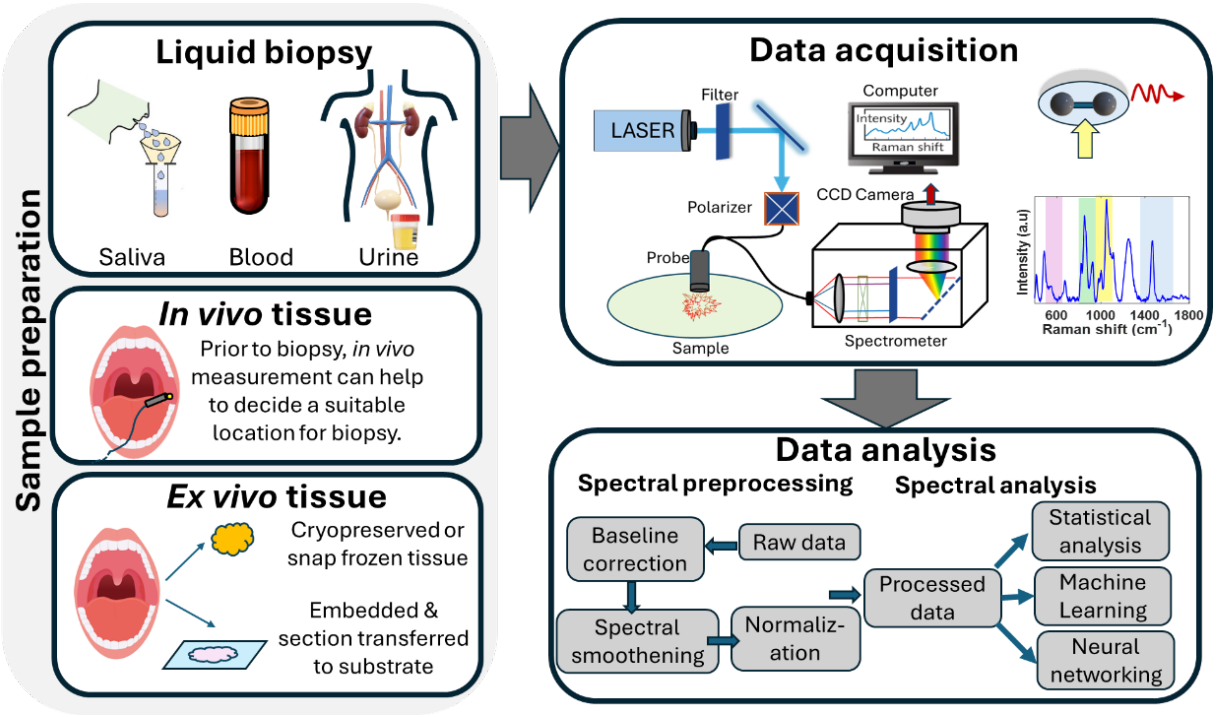


Figure 4.3: Applications of Raman spectroscopy and typical Raman analysis.

In addition to tissue analysis with Raman spectroscopy, biofluidic analysis offers an additional non-invasive tool for early-stage diagnosis. In particular, saliva screening is gaining attention for its ease of collection and rich molecular content, which can provide early-stage biomarkers for the diagnosis of oral cancer. Raman spectroscopy of biofluids such as saliva is an inevitable advancement in the non-invasive screening and monitoring of oral cancer. Given its proximity to oral lesions, saliva can capture molecular signatures of proteins, nucleic acids and other biomolecular components derived from tumour cells or their microenvironment [218]. Raman spectroscopy enables the detection of these biochemical alterations, making it a valuable tool for non-invasive screening and monitoring of oral lesions. In a proof-of-concept study, Falamas et al. discussed different protocols for preparing saliva samples. The highest signal-to-noise ratio was obtained from saliva samples that were centrifuged and lyophilized. They distinguish between OPMD and control groups based on 752, 884, 928, 989 and 1074 cm^{-1}, which were assigned to tryptophan, collagen, proline and glycogen respectively [219]. In another study from the same group, they reported a significant difference in both groups at 2064 cm^{-1}, belonging to thiocynate in Raman and FTIR spectra. They also reported additional characteristics Raman peaks for the OSCC group at 530 and 927 cm^{-1} attributed to lysozyme, 754 cm^{-1} corresponding to tryptophan and 1001 cm^{-1} associated with phenylalanine. Whereas, phosphodiester bonds stretching from DNA and RNA at 1075 cm^{-1} were prominent in the FTIR spectra [220]. Another Raman study on saliva samples reported a high signal-to-noise ratio and a better classification of lyophilised and air-dried samples[221]. Rekha et al. classified Raman data from saliva samples collected from patients suffering from oral submucous fibrosis, OSCC and the control groups. The classification accuracy achieved was up to 84% in fibrosis vs control

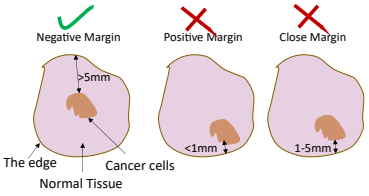


Figure 4.4: During surgery, when cancer tissue is being removed, safe margins are very important. If the pathologist cannot find any cancer cells at the edge of the tissue removed (during post-operative histology), the margins are described as clean or negative margins. However, there are no standard procedures or instructions on how wide the clear margin is. During surgery, the clinician has to rely on his expertise to decide how much border to remove to avoid any positive or close margins, which will significantly increase the chances of cancer recurrence.

groups and 94% in OSCC vs group [222]. Jaychandran et al. access the potential of Raman spectroscopy using tissue and different body fluids including saliva, blood and urine. They were able to distinguish between healthy, OSCC and OPMD groups with 97%, 78%, 90% and 93% accuracy for tissue, blood, urine and saliva, respectively. They observed that for saliva samples, the peaks at 752 cm^{-1} (oxygenated hemocyanin), 1158 and 1525 cm^{-1} (carotenoids) and 1444 cm^{-1} (mucin matrices) showed variation between three groups [223].
In recent years, advances in computational techniques and data preprocessing methods have improved the diagnostic potential of Raman spectroscopy. Usually, classical machine learning methods have been employed to analyse Raman spectra and improve classification. Building on these methods, neural networks are now also opening new eras of analysing complex Raman spectra and more accurate and automated detection of malignant changes. In a recent study, Chang et al. analysed the performance of different deep neural network models to recognise different oral tissues. They were able to distinguish healthy and OSCC tissue from different parts of the oral cavity with an accuracy of up to 93% [186].

4.2.1 Connecting experimental findings with existing literature

Our Raman spectroscopy study revealed significant spectral variations between the OPMD and control groups in both tissue and saliva samples. Notably, protein-associated peaks such as 1266 cm^{-1} and 1666 cm^{-1} showed a consistent trend in both sample types, suggesting its potential as a biomarker for OPMD. Also, 501 cm^{-1}, 541 cm^{-1} and 1464 cm^{-1}, associated with protein and carbohydrates, showed a consistent trend and strong correlation between tissue and saliva samples. These findings extend the insights from previous studies [207, 220] by demonstrating the feasibility of saliva-based screening using Raman spectroscopy, potentially simplifying the diagnostic process.

4.3 Multimodal spectroscopic approach

A multimodal spectroscopic approach combining diffuse reflectance spectroscopy (DRS) and Raman spectroscopy has emerged as a promising strategy for improving the diagnostic accuracy of oral cancer [224]. DRS provides macroscopic insights into tissue morphology and physiology, capturing changes in vascularization and oxygenation, which are indicative of disease progression [225]. Raman spectroscopy complements this by providing molecular-level information, detecting biochemical alterations such as variations in protein, lipid, and nucleic acid structure and content [1, 207]. Our recent work demonstrated that fusing DRS and Raman data using principal component analysis (PCA) significantly improved classification performance in a cohort of 69 participants, enabling more accurate differentiation between benign and malignant lesions (submitted to Photonics West 2026). This integrated approach leverages the strengths of both modalities, supporting early, non-invasive and objective diagnosis [226].

This section reviews key publications on the application of multimodal spectroscopy, highlighting the synergistic benefits of combining DRS and Raman spectroscopy in clinical practice. Greening et al. developed a multimodal instrument to monitor dysplasia progression in the oral cavity, incorporating high-resolution fluorescence imaging and sub-diffuse reflectance spectroscopy (sub-DRS). The system utilised a halogen lamp for sub-DRS and an LED (455 nm) for fluorescence imaging. The developed hybrid system was tested in clinical setting and was found to effectively provide information on the functional and structural properties of tissue [200]. In a comprehensive real-time *in vivo* study by Lin et al., a combination of DRS, autofluorescence and Raman spectroscopy along with white light imaging-guided endoscopy was utilized to detect nasopharyngeal cancer. This multimodal spectroscopic system enabled the differentiation between normal and malignant tissue from a group of 90 volunteers with a sensitivity of 98% and specificity of 95% using a PCA-LDA diagnostic algorithm. Autofluorescence imaging was observed to provide information regarding biochemical composition and metabolic state of tissues. During disease progression, morphological changes such as epithelium thickening, reduced extracellular matrix, and cellular metabolic fluorophores (flavins and nicotinamide adenine dinucleotide) and increased haemoglobin concentration were observed [224].

Our research focuses on enhancing the diagnostic potential of DRS through advanced data analysis techniques and the integration of multimodal spectral information, with the objective of improving the detection and characterization of OPMD in oral tissues.

Conclusion

This chapter has reported the advancements in DRS and Raman spectroscopy for the diagnosis of OC and OPMD, highlighting their ability to provide complementary insights into structural and molecular changes in tissues and biofluids. By reviewing the mechanisms, technological development, and clinical applications of these optical techniques, this chapter has highlighted the gaps in current technology, such as the need for integrating optical modalities in the current clinical scenario, liquid biopsy and improved probe designs for enhanced diagnostic accuracy.
In addition to the review, the findings of our experimental studies are also integrated, demonstrating how spectral variations observed in DRS and Raman spectroscopy align with existing literature while also providing novel insights. By bridging the review and our experimental results, this chapter creates a cohesive narrative that underscores the translational potential of these optical modalities.
The subsequent chapter is focused on the experimental design and methodology, describing the development of the diagnostic systems, study design and experimental setup that address the limitations identified in Chapters 3 and 4. By bridging the theory with the application, this thesis aims to improve non-invasive diagnostics through innovative system design and experimental validation.

Experimental Design and Methodology

5

Chapter 5 provides the materials and methods used in this research. It describes the design and development of multimodal spectroscopic systems, which is the foundation of all the publications this thesis is based on. This chapter is organised into several sections, starting with the system development, where the core components of the DRS and Raman systems are discussed, followed by the outline of the customised systems developed specifically for this study (Manuscript II). The section on experimental design then explains the patient recruitment, sample collection and preparation techniques for analysis. Additionally, the protocols followed for *in vivo* data collection, which were integral to the study, are also discussed. The chapter concludes with a detailed description of the spectral data analysis techniques used, including feature selection, cross-validation, and classification method applied to evaluate the performance of diagnostic systems.

5.1 System design and development

When light interacts with biological tissue, various optical phenomena such as absorption, transmission and scattering occur. Spectroscopic analysis of these optical processes can reveal information about the physical and chemical state of the sample. Spectroscopy can be categorized either by the energy of the electromagnetic waves used or by how these waves interact with the sample. Among the diverse range of spectroscopic techniques, diffuse reflectance spectroscopy (DRS) and Raman spectroscopy are particularly notable for their ability to provide biochemical and morphological information. Together, these techniques can have significant potential for point-based real-time diagnostics. This chapter discusses the development of a multimodal spectroscopic system consisting of Raman spectroscopy and DRS and methodology for data collection.

5.2 DRS system and probe design

DRS is capable of determining morphological changes and also concentrations of absorbing chromophores within the tissue. It can efficiently extract this diagnostic information. Light absorption along with refractive index of the material, shape, size, and density of cellular structures determine the spectral characteristics of backscattered light. DRS can be performed in a wide spectral range, including UV, visible, and NIR light. By adjusting the probe configuration, it is possible to obtain signals from either the superficial epithelial layer or deeper tissue layers [194, 195].

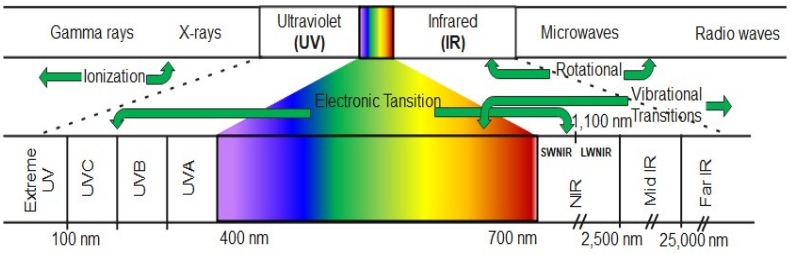


Figure 5.1: Electromagnetic spectrum and effect of different wavelength regions on the matter.

5.2.1 Principle of designing a DRS system

Designing a DRS system involves several key principles to ensure accurate and reliable measurements. Firstly, selecting an appropriate light source is crucial. It should provide stable and consistent illumination and should also cover the required spectral range (depending upon the target sample or the targeted chromophores). Typically, halogen lamps are used for visible and near-infrared (NIR) ranges and xenon lamps for ultraviolet (UV) ranges (Figure 5.1). Probe geometry is another important factor to consider. Source and detector fibers must be carefully configured to illuminate the area of interest uniformly and effectively capture the diffusely reflected light. For the *in vivo* tissue analysis, the probe should be placed perpendicular to the tissue. Additionally, a detector with high sensitivity and appropriate spectral response is required for precise measurement of the diffuse reflectance. The system should also include a mechanism for calibrating the measurements; typically a reference standard with known reflectance properties is used for the correction of any instrumental variations. Finally, employing proper data acquisition and processing techniques, including noise reduction and signal processing algorithms, is vital to extract meaningful spectral information from the collected data.

5.2.2 Customised DRS system

The DRS system used for this *in vivo* study is shown in (Figure 5.2). Tissue illumination was achieved by employing a tungsten-halogen broadband light source (HL-2000-HP, FHSA Ocean Insight, Edinburgh, United Kingdom), having emission wavelengths from 360 to 2400 nm. To guide light for tissue illumination, a low OH [1] 1-to-4 Fan-Out fiber bundles with 400 µm core multimode fiber having a linear common end (BFL44LS01- Thorlabs, Munich, Germany) was used. The same fiber probe bundle was used to collect the reflected light back to a spectrometer. Centre to centre source to detector distance (SDD) was approximately 1.2 mm (Figure 5.3). To record spectral data, it contains a highly sensitive visible wavelength spectrometer (QE Pro, Ocean Insight, Edinburgh, United Kingdom) with a range from 355 to 1100 nm and features a 1024 x 58 element CCD array. A software solution was developed in MATLAB for data acquisition. The code produced can automatically preprocess the acquired raw data. Figure 5.3 illustrates the schematic representation of the light-tissue interaction, highlighting the excitation and collection of diffuse reflectance spectroscopy signals.

1: Standard optical fibers contain hydroxyl (OH) groups that exhibit strong absorption peaks around 1383 nm and 2730 nm. Thus, low OH fibers are utilised in both the Raman and DRS systems to minimise absorption losses. This ensures more sensitive diagnostic measurements, improved signal-to-noise ratio and lower attenuation in spectral data, which is critical to detect subtle morphological changes and chromophore concentrations in tissue.

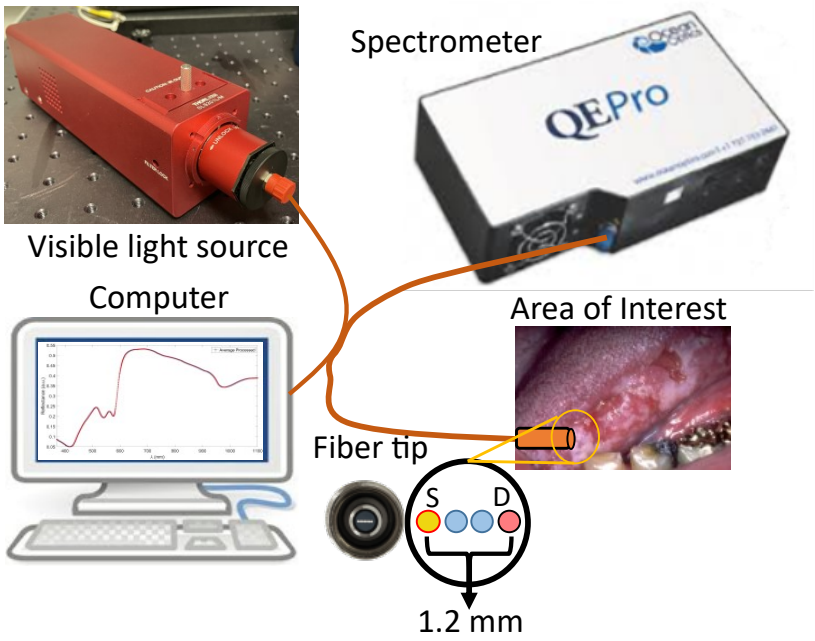


Figure 5.2: Schematic image of the DRS system and probe configuration. In probe "S" represents light source and "D" represents detection fiber.

Calibration and data acquisition

All measurements were calibrated against a Spectralon reflectance standard (WS-1-SL Diffuse Reflectance Standard, Ocean Insight, Edinburgh, United Kingdom) and corrected for the dark counts by following standard method [33]:

$$\text{Reflectance} = \frac{\text{Reflected intensity of tissue} - \text{Background intensity}}{\text{Reflected intensity of standard} - \text{Background intensity}} \tag{5.1}$$

Later, to reduce high-frequency noise Savitzky-Golay smoothing filter was applied. This filter performs a local polynomial regression within a moving window of the data, allowing for effective noise reduction while preserving the shape and height of spectral features [227]. In this study, a second-degree polynomial with a window size of 5 was used, which balances smoothing performance with retention of fine spectral details. The resulting DRS data was saved in the form of an $m \times n$ matrix. Here, m represents the number of DRS spectra (observations) and n represents the number of variables (i.e. wavelengths) from 378 to 1075 nm. All spectra were preprocessed by normalising them to zero mean value and unit variance before proper data analysis.

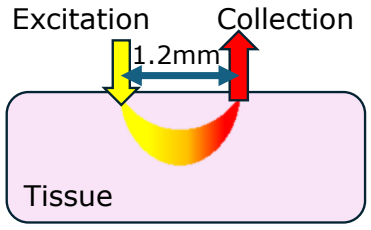


Figure 5.3: Tissue probed by DRS.

5.3 Raman system and probe design

Raman spectroscopy is adept at examining the chemical signatures of tissues. This vibrational spectroscopy provides insight into the structural

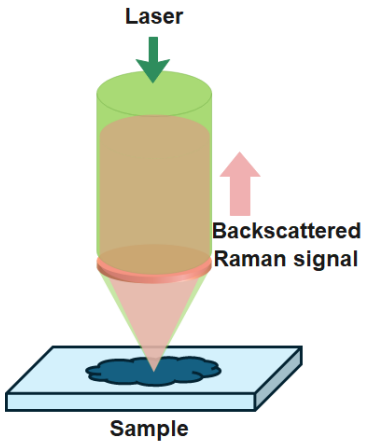


Figure 5.4: Raman signal collected in back scattering mode.

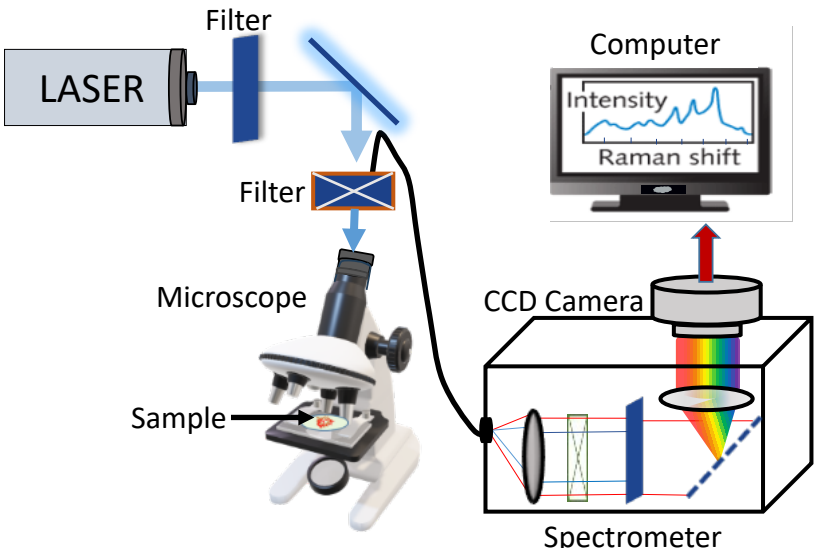


Figure 5.5: Schematic of a common commercial bench top Raman setup.

and chemical composition of the target sample. Changes in the concentration of biomolecules, mutations, or tissue deformation during cancer progression result in significant variations in Raman signals [228, 229].

5.3.1 Principle of designing a Raman system

Figure 5.4 represents principle of Raman signal collection in the backscattered mode. A modern Raman system integrates four important components: an excitation source, a spectrometer, a sample stage and a computer equipped with software to acquire and process spectral data [230], as shown in Figure 5.5. Most commercially available Raman systems are benchtop systems that are coupled with an optical microscope to focus a laser beam on the sample for observation and Raman analysis. The scattered light from the sample is usually collected in the back scattering mode [230].

The incident laser beam is focused on the sample through a microscope objective lens. The scattered light from the sample containing Raman and Rayleigh signals is collected by the same objective lens and is directed via a dichroic mirror towards the spectrometer. The collected light is filtered using a long-pass filter that rejects the Rayleigh scattered signal and allows only the Raman signal to pass through. Raman signal passes through a slit and then a collimated beam[2] is directed to a diffraction grating [3]. The grating separates the Raman signature into spectral signals, and a lens assembly focusses these signals onto a multi-element array detector [231, 232].

2: A collimated light beam (parallel rays) is directed towards diffraction grating. A collimating pack is used to collimate the light beam.

3: A diffraction grating is an optical component with regular patterns of closely spaced lines that diffracts light into several beams travelling in different directions creating a spectrum.

5.3.2 Selection of LASER wavelength

Intensity of Raman scattering is inversely related to the fourth power of the excitation wavelength [233].

$$I_R \propto \frac{1}{\lambda^4} \tag{5.2}$$

Therefore, to achieve higher Raman intensity, short wavelength lasers are preferred such as 488 nm (blue), 514 and 532 nm (green) and 633 nm (red) lasers. These short wavelength lasers are selected to achieve a better signal-to-noise ratio. 488, 514 and 532 nm laser are mainly useful for transparent liquids and polymers. However, in case of biological samples that have a lot of chromophores, these laser wavelengths also produce a huge fluorescent background [234, 235]. Figure 5.6 represents principle of fluorescence in a diatomic molecule.

In that case, using a 1064 nm excitation wavelength ensures minimum fluorescence superimposed on the Raman spectrum. However, such a long laser wavelength dramatically decreases Raman efficiency as well as detector performance [235]. That is why selecting an appropriate laser wavelength is crucial to achieve the most insightful Raman spectrum of the biological sample under observation. Commonly used wavelengths include 785 and 830 nm, especially with silicon-based back-illuminated CCD detectors [236].

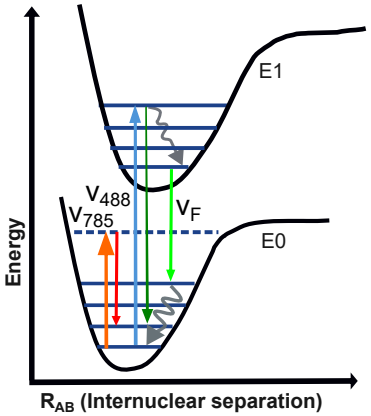


Figure 5.6: Energy levels of a diatomic molecule with bond length R. If the molecule has energy level that can absorb the incoming photon of frequency (ν_L), It will excite to a higher energy state and then relax to the ground state causing fluorescence.

5.3.3 Spectral resolution of a Raman system

The ability of a Raman system to distinguish closely spaced spectral features in a Raman spectrum is termed as the spectral resolution. The high resolution allows detecting subtle differences in vibrational modes, leading to more precise and detailed characterisation. It depends on the important component of the Raman system, i.e. a spectrometer [237].

Resolving power of a spectrometer

A typical spectrometer consists of a diffraction grating element that has a large number of equidistant fine lines or grooves to diffract the incoming light signal into a spectrum. The number of lines per unit length is called the grating constant. Increasing the grating constant leads to higher resolving power. The resolution power of a spectrometer can be described in terms of wavelength λ [238].

$$R = \frac{\lambda}{\Delta\lambda} = kN \tag{5.3}$$

where

- λ is the wavelength of light.
- $\Delta\lambda$ is the smallest difference in wavelength that can be resolved.
- k is the order of diffraction (usually 1 in the case of Raman).
- N is the total number of grooves illuminated by light.

Resolving power of a microscope (spatial resolution)

The resolving power of a microscope is determined by the numerical aperture (NA) of the objective lens. The smallest resolvable feature size R_0 is given by

$$R_0 = \frac{0.61\lambda}{NA} \tag{5.4}$$

where $NA = n \sin\theta$. n is the refractive index of the medium between the objective lens and the sample and θ is the half angle of the maximum cone of light that can enter or exit the objective lens [239].

While the grating determines the spectral resolution[4], the objective lens affects the spatial resolution [5]. Both aspects are crucial for high-quality Raman spectroscopy.

4: Spectral resolution is the ability of an instrument to distinguish between closely spaced wavelength in a spectrum. It is a measure of how nicely an instrument can separate light into its component wavelengths. It is usually represented as $\Delta\lambda$.

5: Spatial resolution is the ability of an instrument to distinguish between smallest details of an object.

5.3.4 Customised multimode Raman system

To enable early detection in primary care facilities, we envision the application of a highly adaptable spectrometry system designed to easily integrate into research settings and clinical procedures. Details of this multimodal Raman system are discussed in Paper II, **Maryam, S.; et al., "Mobile multi-configuration clinical translational Raman system for oral cancer application", Analyst 2023** [2]. This custom portable Raman system shown in Figure 5.7 is capable of both *ex vivo* and *in vivo* analysis. It is compact, portable, cost-effective and easy to assemble. It is fitted inside a console to transport it to point of care. The customised multimodal Raman system built around a 785 nm laser source (Cobolt 08 - 01 Series, Cobolt, Sweden), selected for its optimal balance between minimizing fluorescence interference and maintaining adequate Raman scattering efficiency. The system includes a diffraction grating with 1200 grooves/mm (LS - 785, Teledyne Princeton Instruments, USA) and a back-illuminated CCD detector (Blaze 400 HR, 1340 × 400 pixels, Teledyne Princeton Instruments, USA) for acquiring high-resolution spectra. A modular sample probing unit allows for easy switching between different measurement configurations, enabling flexible adaptation to various experimental setups.

- Mode (A) Raman microscope – for *ex vivo* saliva analysis in dry form, it can also be used to analyse cell pallet.
- Mode (B) Raman liquid analysis stage – for saliva supernatant analysis using photonic crystal fiber (PCF).
- Mode (C) Raman optical fiber probe – for *in vivo* tissue analysis.

These subunits can be connected to the spectrometer using optical fibers to develop a custom system for the study of oral cancer. The Raman microscope and liquid analysis stage is connected to the spectrometer through a round-to-linear fiber bundle of a 7 × 200 μ m core (Thorlabs BFL200LS02).

The three configuration modes of the customised Raman system are designed to address the diverse requirements of oral cancer diagnostics across different sample types and clinical contexts. Mode (A), the Raman microscope, is essential for high-resolution *ex vivo* analysis of dried saliva samples and cellular material. This mode provides detailed biochemical profiling at the microscopic level. Mode (B), liquid analysis stage using photonic crystal fibre (PCF), is specifically for analysing saliva supernatant, which is particularly useful for non-invasive screening. Mode (C), the optical fibre probe, is critical for in vivo tissue analysis, providing molecular insights during clinical examination. Each mode targets a unique diagnostic niche - ex vivo, biofluidic and in vivo provides complementary information. Together they provide a comprehensive platform for translational research and point-of-care screening. Their

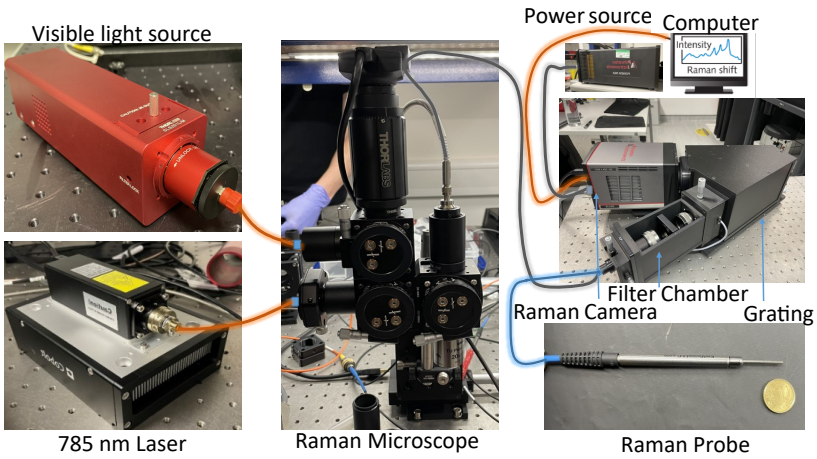


Figure 5.7: Customised Raman system with mode A (Raman microscope) and Mode C (Raman optical fiber probe).

inclusion is not only technically justified but also clinically relevant, as demonstrated in the published study [25].

Mode (A) Raman microscope for *ex vivo* analysis

The customised Raman microscope (mode A) is shown in Figure 5.8 with a laser and a visible light source connected through optical fibers. Fiber coupled illumination facilitates spot size adjustment at the sample plane. Additionally, this fiber guided method of light delivery simplifies switching to different excitation lasers. The flexible stage design is another notable feature of this system, capable of accommodating specimens of mm to tens of cm. The specimen was illuminated through an objective lens and Raman signal was collected in backscatter mode and directed to the spectrometer via 7× round-to-linear low OH fiber bundle with 200 μm core (Thorlabs BFL200LS02). Simultaneously, a portion of the light from the sample was directed to a colour detector. This configuration was used for *ex vivo* analysis of solid and liquid specimens in the reflective mode.

- **White light imaging optical path:** A broadband visible light source (Thorlabs SLS201L-broadband 360 - 2600 nm source) was used for tissue illumination for conventional optical microscopy. The incident broadband light beam entered the microscope through lens L1 (Thorlabs LA1951A, f = 25 mm) and was directed to the objective lens using a beam splitter BS (Thorlabs, BSW26R, 50:50, 350–1100 nm), dichroic mirrors DM1 (Semrock, FF757-Di01) and DM2 (Semrock LPD02-785RU). The L1 was positioned to achieve a uniform Kohler illumination of white light on the sample plane. The white light scattered from the sample was reflected by DM2 to DM1 redirecting it to a coloured CMOS camera (Thorlabs DCC1645C) through BS and a field lens L2 (Thorlabs LB1437-A, f = 150 mm) to visualise the sample under white light. The objective lens L4 (116 × 92 μm for 50× , 290 × 230 μm for 20× and 1160 × 920 μm for 5×) control field of view (FOV). The images were displayed on a computer screen using Thorcam software.

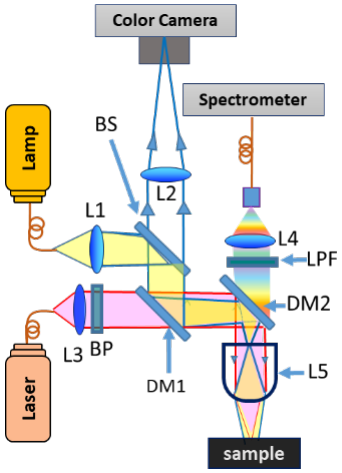
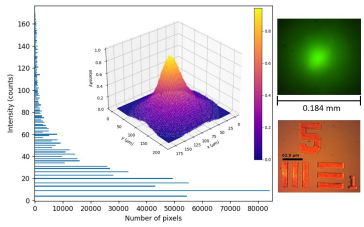


Figure 5.8: Schematic of customised Raman microscope [2].



- **Raman optical path:** Monochromatic 785 nm laser light was collimated and cleaned by lens L3 (Thorlabs A397TM-B, f = 11 mm, NA (numerical aperture) = 0.3) and laser cleaning band-pass filter BP (Semrock LD01-785/10-25) respectively. Subsequently, it was directed towards the objective lens of the microscope through DM1 and DM2 to be focused onto the sample plane. The focused excitation beam has 60 μ m spot size with 20× objective lens, as shown in Figure 5.9. From the back scattered signal, excitation light is filtered using a dichroic mirror (DM2) and a long pass filter LPF (Semrock LP02-785RU-25). Filtered Raman light focused on a low OH circular to linear fiber bundle (Thorlabs BFL200LS02) using a lens L4 (Thorlabs F810SMA-850, f = 36.2 mm). The linear side of the fiber was connected to the entrance slit of the Raman spectrometer. The Raman spectrometer subunit included a filter chamber with a long pass filter (Semrock LP02-785RU-25), a high throughput spectrometer (Teledyne Princeton Instruments Acton LS785, f/2), and a highly sensitive CCD camera (Blaze: 400HR, Teledyne Princeton Instruments, USA). The CCD pixels were vertically binned to combine the light transmitted through the entrance slit of the spectrometer [2].

The proposed Raman microscope can be used for the ex vivo analysis of any excised tissue or for liquid biopsies. Figure 5.9 represents a beam profile characterisation of our 785 nm Raman laser beam in the focus of the microscope. The beam profile has a Gaussian shape with FWHM bandwidth of 60 μm. To further reduce spot size, a single-mode input laser fiber is also incorporated for future studies. The histogram shows the number of pixels having a certain intensity and the image of a resolution target illustrates the spatial resolution of the microscope.

Mode (B) Raman liquid analysis stage

Although the dried saliva sample provides valuable information on saliva composition and biomarkers present in it, it may not accurately represent the native state of the biomolecules, as drying can change the concentration of analytes. It can also affect the molecular structure and vibrations, potentially providing misleading information. Analysing saliva in its liquid form is challenging, but it can ensure that the biochemical composition remains intact. It can also provide information regarding dynamic changes in real time, offering a more precise understanding of the biomarkers associated with oral diseases and improving the sensitivity and specificity of early cancer detection methods. Raman is a weak scattering process and the concentration of the targeted analytes is low in liquid saliva. Because of this reason, it is challenging to record a high signal-to-noise ratio Raman signal from liquid saliva sample under the microscope on a two-dimensional substrate. That is why, we have proposed another method of analysing saliva in its native form by using a Photonic Crystal Fiber (PCF) as a sample holder. PCFs are optical fibers (Figure 5.10) with a periodic array of air holes surrounding a hollow or solid core. In this study, we used silica-based suspended core PCFs with a solid core (SC-PCF) located at the intersection of three large air holes. Light is guided within this triangular core by total internal re-

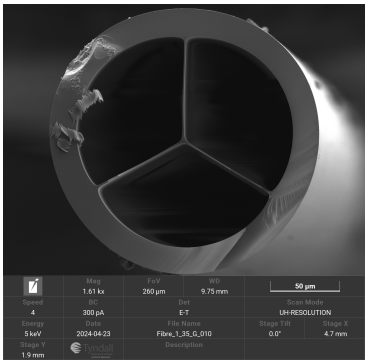


Figure 5.10: SEM image of a PCF.

flection, due to the difference between the refractive index of silica and air.

To obtain a Raman signal from liquid samples (such as saliva supernatant) in a large volume, a 3.5 μm core diameter SC-PCF with >70 μm air-channel diameter and NA = 0.5 was used. The external diameter of these SC-PCFs was 200 μm to allow easier handling. While 785 nm excitation light propagates through the solid core, Raman signals were generated by the evanescent wave that extends into the air holes. PCFs act as a sample holder to prolong the length of the light-sample interaction, thereby improving the detection sensitivity. Longer path length allows for a larger volume for sampling, for a better representation of low concentration analytes in a small sample volume (about 20–100 nl). A PCF filled with saliva can be placed under an upright Raman microscope for analysis; however, this vertical configuration presents practical challenges. To address these, a customized horizontal stage shown in Figure 5.11 was developed to facilitate more reliable and reproducible measurements. Aligning the input laser beam with the PCF core is critical for efficient Raman signal collection. The horizontal orientation simplifies the alignment by providing stable, planar positioning of the PCF. Additionally, when the PCF is positioned vertically, even slight downward movement of the saliva due to gravity can shift the sample away from the laser focus on the fiber tip, resulting in reduced Raman signal intensity and compromised measurement repeatability. The horizontal stage mitigates this issue by providing stable optical interrogation. The biofluidics stage design ensures ease of operation and consistency in the acquired Raman spectra, particularly when working with small volumes of samples such as saliva.

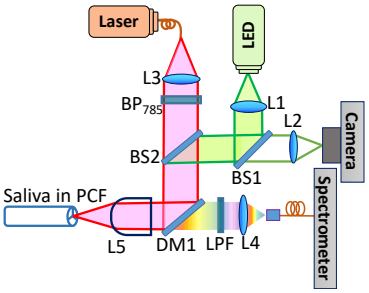


Figure 5.11: Schematic of customised PCF setup for liquid sample analysis [3].

In the PCF setup (mode B, illustrated in Figure 5.11), the fibre core was brought into focus of the objective lens by illuminating it with a green light. Green light was guided to the PCF by a lens L1 (Thorlabs AC254-030-74 B, f = 30.0 mm, Achromatic Doublet, ARC: 650 - 1050 nm), beam splitters BS1 and BS2 (CM1-BP108: Thorlabs), DM1 (Semrock LPD02-785RU) and a 50x objective lens. The image of the fiber is shown on the computer screen through the camera. The light from a single mode 785 nm laser was collimated and filtered by a lens L1 (Thorlabs AC127-025-AB-ML, f = 25 mm) and a band pass filter BP_{785} (Semrock LD01-785/10-25). It reaches the PCF core after passing through BS2, DM1 and the objective lens. The laser beam focused on the core can also be seen with the camera when the green light was turned off. The Raman signal from the sample within the PCF was collected in a back-scattered configuration. After passing through the objective lens and DM1, the Raman signal was filtered by a long pass filter (Semrock LP02-785RU-25) for the excitation wavelength and was focused on the fiber (Thorlabs BFL200LS02) going to the spectrometer, through a focusing lens (Thorlabs F810SMA-850, f = 36.2 mm) [3].

Mode (C) Raman optical fiber probe

Fibre optics-based Raman systems have a great potential for real-time *in vivo* tissue analysis during endoscopy or surgical procedure. Figure 5.12 represents mode C of this customised Raman system. It is a multi-spectroscopy fiber-optic probe (EMVision, Loxahatchee, FL, USA)

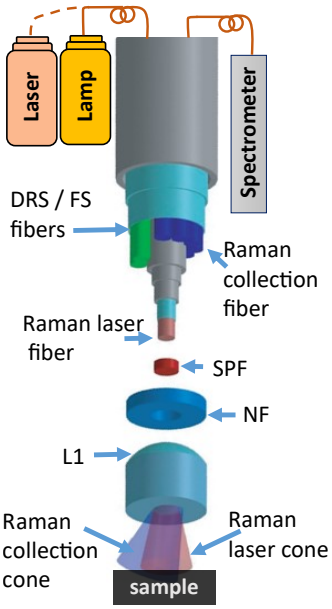


Figure 5.12: Schematic diagram of customised multi-spectroscopy fiber-optic probe [2].

utilised for *in vivo t*issue analysis. This customised probe is capable of acquiring Raman, fluorescence (FS), and DRS spectra, combining them as a multimodal spectroscopic system. At the distal end of the probe, nine 300 μm core fibers with a numerical aperture (NA) of 0.22 are arranged in a concentric ring around a similar central Raman excitation fiber. There is a short pass emission filter (SPF) in front of the central fiber to clean the input excitation laser beam. Seven of the circularly arranged fibers collect the Raman signal. The collected signal is then transmitted through a notch filter (NF) to filter out the excitation wavelength, while allowing only the light shifted to other wavelengths, like Raman signals, to pass through. The remaining two fibers are allocated for DRS and FS measurements. Furthermore, a two-component customised lens (L1) at the fiber tip ensured overlapping focus of excitation light and the area of Raman light collection from the tissue. At the frontal end of the probe, the seven Raman collection fibers are aligned linearly to transmit light to the spectrometer via its 50 μm entrance slit. This arrangement is chosen in an attempt to optimise the signal-to-noise ratio of the acquired Raman signal, which is crucial for biological samples [2].

Calibration and data acquisition

For all three modes of Raman spectroscopy, data was recorded using LightField software (Teledyne Princeton Instruments, USA). Figure 5.13 illustrates the key stages of the data analysis pipeline, including data preprocessing, classifier training, cross-validation and feature selection, which are discussed in detail in section 5.5 Calibration was critical to ensure spectral accuracy and comparability across different sample types, experimental configurations and measurement sessions. The pixel-to-wavelength (x-axis) calibration was performed using nine known emission lines from a Neon–Argon (NeAr) lamp. A third-order polynomial fit between pixel position and wavelength was applied to obtain a continuous wavelength calibration function. The calibrated wavelengths were then converted to Raman shift (cm^{-1}) using the excitation laser wavelength according to equation 5.5 and validated with acetaminophen and naphthalene standard spectrum. The y-axis (intensity) was calibrated using a standard spectrum from a NIST lamp to correct for instrument response.

$$\nu = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda}\right) \times 10^7 \tag{5.5}$$

Spikes were removed to eliminate abrupt anomalies, which can arise from cosmic rays or transient fluctuations during acquisition. For each spectrum baseline was corrected using the asymmetric least squares (ALS) method [240] (λ = 50 and P = 0.00005 for liquid saliva; λ = 100 and P = 0.00009 for dry saliva; and λ = 10 and P = 0.0001 for tissue) to remove autofluorescence. A second-degree Savitzky-Golay filter with a window width of 5 was applied to smooth the data while preserving spectral features. Finally, data was normalised with standard normal variate analysis to ensure equal contribution of all spectra to the model and eradicate inter/intra spectral variability. Then, the Raman spectral region of 400 – 1800 was selected for Raman analysis. The calibration and processing pipeline ensured consistent Raman signal quality across

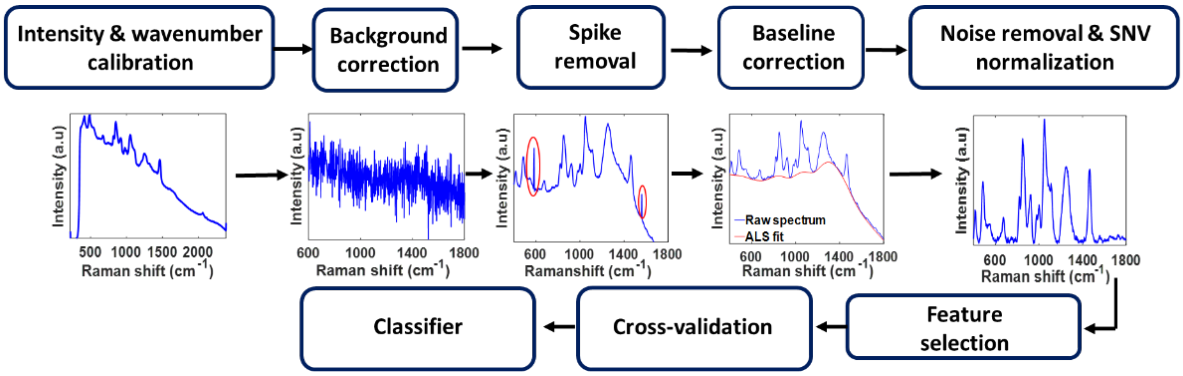


Figure 5.13: Overview of the Raman data analysis pipeline, including preprocessing steps (calibration, background correction, spike removal, baseline correction, noise reduction and SNV normalization) followed by feature selection, classifier training and cross-validation for spectral classification

all three modes, enabling reliable comparison between tissue and saliva spectra and supporting robust analysis for OPMD detection.

This system was designed to address the limitations of existing Raman setups, including their high cost and lack of adaptability for diverse clinical applications. By enabling seamless transition between different Raman modes for *in vivo* and *ex vivo* analysis, the system provides a robust platform for OPMD diagnosis. Moreover, it incorporates DRS, allowing complementary structural and morphological insights along with molecular information obtained by Raman spectroscopy. Figure 5.14 represents the customised multimodal system. This shows the two spectroscopic systems - Raman and DRS - for *in vivo* tissue analysis connected together through a single multimodal probe. The developed customised multimodal system was tested for accuracy and repeatability multiple times using different analytes, such as uric acid [36], rhodamine b [2], single cheek cells, blood smear, liquid and dry saliva [2].

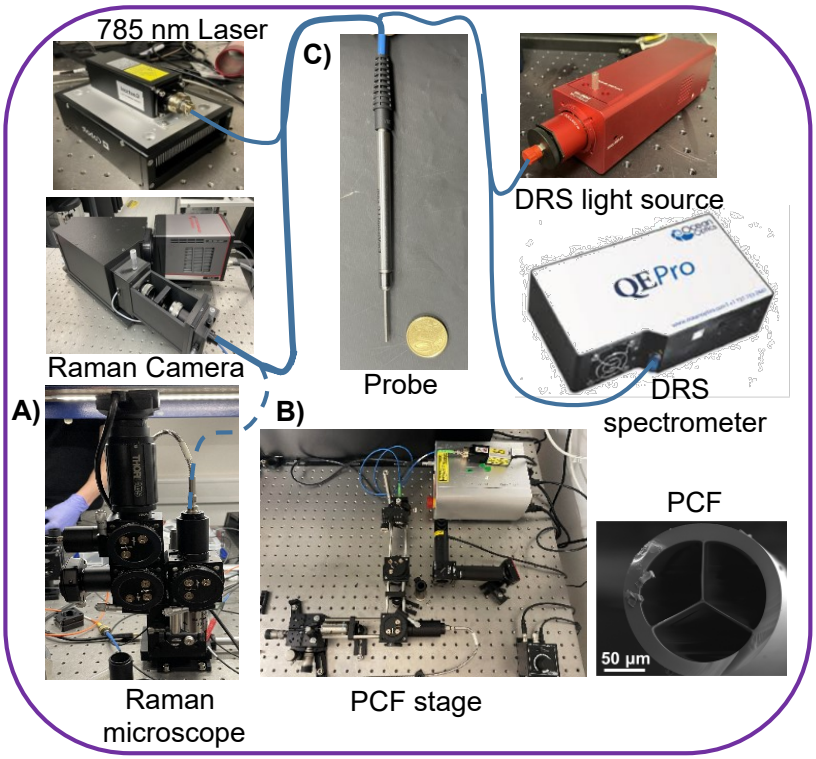


Figure 5.14: Customised multimodal spectroscopic system including three Raman modes A) microscope for dry saliva analysis B) PCF stage for liquid saliva analysis C) multimodal probe for *In vivo* tissue analysis connected with Raman and DRS spectroscopes

5.4 Experimental design

Patients presenting to Cork University Dental School and Hospital (CUDSH) and South Infirmary Victoria University Hospital (SIVUH) with oral lesions who recently underwent a biopsy of these lesions as part of their clinical care were recruited for this study. This study was approved by the Cork Research Ethics Committee (CREC) of the University College Cork (UCC) for 300 participants according to the CREC application ref.: ECM 4 (y) 13/4/2021 & ECM 3 (aaaa) 12/09/2023 including healthy volunteers and patients identified with OPMD. The study was conducted in compliance with the ethical standards established by CREC and the 1964 Helsinki Declaration and its subsequent amendments. From all participants, prior to saliva sample and data collection, informed consent was obtained.

The study involved 84 participants, 24 in the control group and 60 in OPMD group. The study is a combination of *in vivo* tissue and *ex vivo* saliva analysis, as depicted in Figure 5.15. Saliva samples were collected from all participants and *in vivo* data was collected from six anatomical locations in the oral cavity, including the cheek mucosa (CM), ventral and dorsal surface of the tongue (VST and DST), the lips (Lips), the gingiva (Gum) and the anterior floor of mouth (AFM) from both control and OPMD group. In the OPMD group, data was collected from the OPMD lesion (identified by the clinician) and the healthy contralateral site. The table below summarises the number of lesions from participants in the OPMD group on different tissues. However, as is evident, the number of patients in some tissue groups are relatively low, making them unsuitable for further statistical analysis. Therefore, to ensure unbiased and robust

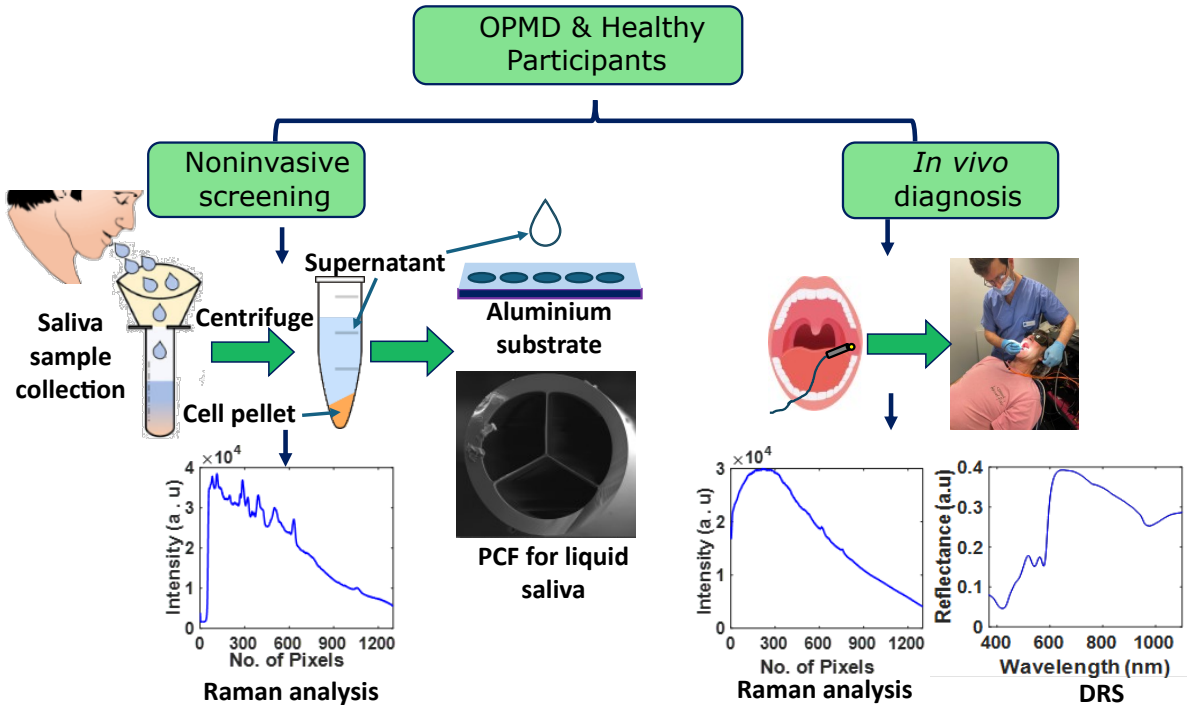


Figure 5.15: Experimental design consisting of *in vivo* tissue analysis and saliva screening.

results, only CM and VST were used for further analysis in case of DRS and only CM was considered for further analysis in case of *in vivo* Raman and saliva study. The total number of lesions on each tissue is shown in table 5.1.

Table 5.1: Number of OPMD lesions (PML) from 60 participants with respect to tissue.

Tissue	No. of potentially malignant lesions (PML)
CM	28
VST	19
DST	7
Lip	6
Gum	2
AFM	1

5.4.1 Recruitment of patients

Participants were recruited based on predefined inclusion criteria, and those who did not meet these criteria were excluded from the study.

Inclusion criteria

- Patients who signed the informed consent form.
- Patients 18 years of age or older.
- Patients who had or required a biospsy as part of their clinical care.

Exclusion criteria

- Patients who did not sign the consent form or who do not have the ability to provide informed consent.
- Pregnant women.
- Patients previously diagnosed with upper aerodigestive tract cancer or having radiotherapy.
- Any condition resulting in the patient not being able to provide a sufficient saliva sample.
- Individuals with known skin photosensitivity conditions (e.g. porphyria).
- Individuals who received oral or intravenous fluorescent medication (e.g. methylene blue or indocyanine green) in the past month.

- Individuals treated with a photosensitiser in the past 4 months.
- Patients receiving vitamin B supplement treatment.
- Patients with intraoral candida infection.

Table 5.2 represents the total number of participants and sample size for *in vivo* tissue and saliva.

	Control (H)	OPMD
Total participants	24	60
Total saliva samples	21	42
Total participants-DRS	22	60
Total participants-Raman	24	53

Table 5.2: Number of participants and sample size for Raman, DRS and saliva analysis

5.4.2 Saliva sample collection

The clinician involved in this study verified the oral hygiene of the participant and provided them with information on how to provide an unstimulated saliva sample. Detailed clinical and demographic information was recorded for each patient, including their medical history and current medications. To minimise variations, saliva samples were collected according to a standardised protocol. Patients were advised to abstain from drinks (excluding water), food and smoking one hour prior to sample collection. They were requested to avoid lipstick and other lip products. Just before sample collection, patients were advised to rinse their mouths with water for 1 to 2 min. Saliva was collected in a sterile tube by gentle drooling and immediately labelled and stored in a cold storage box at 4° C. The samples were then centrifuged for 10 min at 1000 rpm at room temperature to remove food debris and cells. The resulting cell pellet and supernatant were stored separately at -80 ° C until further analysis. Saliva samples were collected from the 24 participants in the control group and the 60 participants in the OPMD group. However, 2 participants in the OPMD group were unable to provide enough saliva samples. Four additional saliva samples were excluded from the analysis due to blood contamination and technical issues encountered during processing. Also, saliva samples collected from 3 healthy participants and 9 OPMD participants had a lot of mucus and were not suitable for this type of analysis at the moment.

So in total saliva samples from 21 healthy volunteers and 42 OPMD participants were suitable for further analysis.

5.4.3 Dry saliva sample preparation and data collection

Saliva samples were thawed at room temperature and filtered with 300 KDa filters. Five droplets of 2 μl each were applied on an aluminium substrate with subsequent drying. The prepared saliva sample was then analysed under a customised Raman microscope system (Mode A) with a 50x objective lens and 20 mW laser power. The accumulation time for each measurement was 10 sec and each spectra was averaged over 5 measurements.
In total, around 315 spectra were recorded from the control group and around 630 Raman spectra were recorded for the OPMD group. At the moment, the potential of this study is demonstrated through a subgroup analysis of data with balanced group sizes, including 19 healthy volunteers and 18 participants in OPMD who have lesions on the cheek mucosa.
After data calibration, spike removal and dark background correction, the data was baseline corrected to remove the autofluorescence background using the asymmetric least squares (ALS) method with λ = 100 and P

= 0.0009. Raman data ranging from 400-1800 cm^{-1} for 654 data points from 37 (19 + 18) participants were subjected to further analysis. To discriminate between both groups, a Lasso feature selection model was used for feature selection, and then further analysis was performed by LDA and leave one out (LOO) classifier.

5.4.4 Liquid saliva sample preparation and data collection

This was a proof-of-concept study involving a total of 11 saliva samples - 5 from the control group and 6 from patients with oral potentially malignant disorders (OPMD). To prepare the liquid saliva sample filled inside PCF, saliva samples were thawed and filtered with 300 kDA centrifuge filters. A PCF with 9 to 10 cm length was washed with acetone and dried with a syringe by pushing air inside. The filtered saliva sample was then filled inside the PCF with a syringe pump using a 17 gauge needle at $50\mu l min^{-1}$ speed. After the needle was removed, both ends of the PCF were cleaved to achieve a sharp finish. Polymer coating was removed from approximately 3 to 4 mm length of the fiber facing toward the laser to facilitate focusing laser on the core and to avoid interference from the polymer. The prepared saliva filled PCF sample was placed in the fiber holder of stage shown in Figure 5.11 and each spectrum was recorded for 4 s and five accumulations with approximately 20 mW laser power. Ten measurements were recorded from each sample.
The collected data was calibrated along the x and y axes, spikes were removed and data was corrected for dark background. The baseline was corrected using the ALS method (λ = 100 and P = 0.0009). Raman data ranging from 500 - 1800 cm^{-1} for 11 participants (5 + 6) were subjected to further analysis [3].

5.4.5 *In vivo* data collection

After collecting saliva sample, participants were subjected to *in vivo* tissue analysis with DRS and Raman spectroscopy. For both techniques, data was collected from all the six tissue types mentioned in table 5.1 including healthy tissue (H) from the control group, potentially malignant lesion (OPMD) and healthy contralateral cite (OPMDH) in the OPMD group. Before measurements were made, it was made sure that the room was absolutely dark to minimise interference in spectral data from the stray light.

Data collection with DRS

For DRS study a software solution developed in MATLAB was used for data collection and real-time data processing. Each DRS measurement was recorded over 50 ms acquisition time and averaged over 20 repetitions. Data was collected from 22 healthy volunteers and 60 OPMD volunteers. In the initial phase of the study, parameter optimisation for both Raman and DRS was the priority. As many parameters were being tested at that point, for patient's comfort and to reduce overall measurement time, approximately 9 data measurements were recorded from the first few participants and later $\approx$ 15 data measurements (DM) in both groups

Table 5.3: Number of OPMD lesion (OPMD) and spectra from OPMD and H group with respect to tissue for DRS analysis.

Tissue	No. of lesions on each tissue	OPMD (Spectra)	H (Spectra)	Total spectra (H + OPMD + OPMDH)
CM	28	393	324	1494
VST	19	270	324	1392
DST	7	102	324	1215
Lip	6	78	324	1233
Gum	2	33	324	1125
AFM	0	0	324	1074

subjected to participant's convenience and condition of the oral cavity. Detailed information on the DRS data is shown in table 5.3. In total accumulating 1944 spectra from **healthy group**.

- 54 spectra from 1 participant (9 spectra from 6 tissues = 54).
- 90 spectra from 21 participants (15 spectra from 6 tissues = 1890).

and 5589 spectra from **OPMD group**

- 54 spectra from first 5 participants (9 spectra from 6 tissues (lesion, CM, VST, DST, Lip, Gum) = 270).
- 63 spectra from next 11 participants (9 spectra from 7 tissues (lesion, CM, VST, DST, Lip, Gum, AFM)= 693).
- 105 and 90 spectra from the subsequent 43 participants (15 spectra from 7 tissues (lesion, CM, VST, DST, Lip, Gum, AFM) of 37 participants and from 6 tissues for the rest of the 2 without OPMDH contralateral site = 4065).
- 126 and 108 spectra from 4 subjects on clinician's request (18 spectra from 7 tissues (lesion, CM, VST, DST, Lip, Gum, AFM) of 2 participants and 6 tissues from the other 2 without OPMDH = 468).
- 93 spectra were recorded from 1 participant due to participant's discomfort (15 from 6 tissues (lesion, CM, VST, DST, Lip, AFM) and 3 from 1 tissue (Gum) = 93).

Data collection with Raman spectroscopy

Raman data was collected from 24 healthy participants in the control group and 53 participants in the OPMD group. Data was collected from the six anatomical locations, including the lesion and contralateral site. 3 to 5 spectral measurements were recorded from each tissue depending on the participant's comfort and condition of the oral cavity. In total, 687 spectral measurements were collected from different tissues in the control group and 1731 spectral measurements from OPMD group. The total OPMD lesions and the number of Raman spectra collected from OPMD and H are shown in table 5.4. Each spectral measurement was averaged over 9 accumulations to reduce spectral noise and the data acquisition time for each spectral recording was approximately 4 s.

Table 5.4: Total OPMD lesion (PML) and distribution of spectral data from different groups for Raman analysis.

Tissue	No. of lesions on each tissue	OPMD (Spectra)	H (Spectra)	Total spectra (H + OPMD + OPMDH)
CM	25	117	118	474
VST	15	78	104	425
DST	8	35	117	393
Lip	4	20	113	382
Gum	3	15	119	374
AFM	1	5	116	365

5.5 Spectral data analysis

5.5.1 Feature selection

Feature selection is a crucial step in building predictive models for spectroscopic data. It involves identifying the most relevant features that contribute to predicting an outcome. Feature selection can improve the performance of the developed model, while reducing complexity and computational costs. In this study, the Lasso regression model (Least Absolute Shrinkage and Selection Operator) was used for feature selection. It performs both variable selection and regularisation using a cost function and improves the prediction accuracy of the developed statistical model. It minimises the sum of squared residuals between actual and predicted values by imposing a penalty term to regression coefficient values. It results in reducing some of the coefficients to zero and thus eliminating less important features and retaining only those that have the most predictive power. [6]In mathematical terms it can be expressed as

$$\text{Cost function} = \sum_{i=1}^{n}(y_{real}^{(i)} - y_{pred}^{(i)})^2 + \alpha \sum_{j=1}^{m} |\beta_j| \tag{5.6}$$

6: Simplified version of Lasso equation 5.6 is
Lasso regression = (Sum of squared errors + α(Sum of absolute values of coefficients)

where, n is the number of DRS spectra, m is the number of features (wavelength), ($y_{real}^{(i)}$ and $y_{pred}^{(i)}$) are the true and predicted values of the ith spectrum. The term $\alpha \sum_{j=1}^{m} \beta_j$ is L1 penalty, β_j is the regression coefficient of j^{th} feature, α is a regularization parameter controlling the strength of the penalty term L1. A larger value of α leads to more pronounced regularization, driving more coefficients close to zero resulting in the development of a sparser model [241]. Conversely, smaller values of alpha reduce the amount of regularization, allowing more wavelengths to retain non-zero coefficients and thus increasing the number of features contributing to the model. The α parameter was selected using cross-validation to balance between prediction error and over- or under-fitting. LASSO regression was applied on the scaled dataset and only variables with non-zero coefficients were retained for inclusion in the predictive model. The contribution of each wavelength is quantified through LASSO weights, which guides the development of a classification model [241].

5.5.2 Cross validation and classifcation

Leave one out cross validation (LOOCV)

Following the feature selection with Lasso regression model, leave-one-out cross validation (LOOCV) was employed to evaluate the performance of the developed machine learning model. In LOOCV, applied on a set of n samples, the model is iteratively trained n times using n-1 samples for training and the remaining one sample for testing. It provides an unbiased estimate of the model's performance.

Linear discriminate analysis (LDA)

Linear discriminate analysis (LDA) was applied to classify samples on the basis of selected features. It is a statistical method to separate two or more groups by projecting the features onto a single-dimensional space. It works by maximizing the separation between classes.

Linking the analysis to our research goals

The implementation of LASSO, LOOCV and LDA was tailored to meet the primary objectives of this study, focussing on improving diagnostic accuracy for oral potentially malignant disorders (OPMD) using spectroscopic techniques. The key goals of the analysis were:

1. **Identify discriminative feature:** Primary aim of feature selection model was to identify spectral markers that could distinguish between OPMD and H tissue and saliva samples. By applying Lasso regression model, features with highest Lasso weights that had higher predictive power were selected, while minimising redundancy. It was also ensured that the selected spectral features provide robust biochemical insight.
2. **Reliable classification model:** A simple classification model was designed to categorise the data in OPMD and the control groups. The use of LDA enables a straightforward yet effective model that leverages the selected features for group selection. It ensures an accurate and interpretable clinically relevant model.
3. **Robustness through cross validation:** The LOOCV was integrated to evaluate the reliability of the developed model. It iteratively tests the model on individual samples while training on the remaining dataset. It provides an unbiased estimate of the model's performance and minimises the risk of overfitting.
4. **Align selected spectral features with diagnostic potential:** The selected features were critically examined for their relevance to OPMD associated biochemical changes, such as alterations in protein, lipids, carbohydrates and haemoglobin content. The alignment between spectral data and underlying pathological changes confirms the diagnostic potential of the multimodal spectroscopic approach.

5.5.3 Validation metrics

Validation metrics are expository for evaluating the performance and reliability of a machine learning model. This study employed various metrics to access a robust interpretation of the developed classification model. They are discussed below:

Box plots and statistical analysis of selected features

Box plot analysis was used to compare the distribution of selected features between the OPMD and the control group. The significance of the difference between both groups was confirmed by non-parametric statistical analysis.

LDA scatter plot:

LDA scatter plot represents the separation between the OPMD and H tissue in the Raman data of tissue and saliva samples in the LDA space. Each participant was represented as a data point to facilitate visual understanding of the classification performance, highlighting correctly and incorrectly classified cases.

Truth table:

The truth table summarises correctly and incorrectly classified patients. It highlights the model's ability to classify the data correctly between OPMD and control group. Each patient was represented as a data point to evaluate individual level performance.

Area under the curve (AUC):

The AUC of the Receiver Operating Characteristic (ROC) curve in the DRS study was included as an additional parameter of diagnostic performance. It reflects the trade-off between sensitivity and specificity. It also provides a summary of the model's classification capability.

Balanced accuracy

Balanced accuracy accounts for the class imbalance while calculating the accuracy. It give equal weight to both classes and ensures an unbiased measure of model performance across OPMD and control group unlike accuracy [7]. Balanced accuracy can be defined as the average of true positive rate (sensitivity) and the true negative rate (specificity), providing a more equitable measure of a model's performance, particularly for imbalanced datasets.

$$\text{Balanced Accuracy} = \frac{(\text{Sensitivity} + \text{Specificity})}{2} \tag{5.7}$$

7: Accuracy is the ratio of correctly predicted samples to total number of predicted samples:

$$\text{Accuracy} = \frac{(\text{True positive} + \text{True Negative})}{\text{Total number of samples}}$$

It may not accurately reflect the model's performance in case of unbalanced data. In such cases, higher accuracy shows that the model is biased toward the majority class

Where

$$\text{Sensitivity} = \frac{\text{True positive cases}}{(\text{True positive cases} + \text{False negative cases})} \tag{5.8}$$

$$\text{Specificity} = \frac{\text{True negative cases}}{(\text{True negative cases} + \text{False positive cases})} \tag{5.9}$$

Conclusion

This chapter detailed the design and development of the multimodal spectroscopic systems that form the backbone of this research. This system consisting of Raman spectroscopy and DRS is a step towards non-invasive, real-time diagnostic solutions for OPMD. Despite their potential, these methods still hold challenges, including sample variability, probe-tissue contact and the complexity of data interpretation. These challenges emphasise the need for continuous improvement in both hardware and data processing techniques. As we transition into Chapter 6, the subsequent discussion reports on the personal perspectives gained throughout this research, its collaborative nature and impact on bridging the gap between laboratory research and clinical application. The results of this research are summarised in the publications included at the end of the thesis, where the findings of this study are presented in detail, emphasising the clinical relevance of the proposed diagnostic approach.

Results and Discussion 6

This study explored the application of optical spectroscopic techniques for OPMD diagnosis, including DRS, Raman spectroscopy and Photonic Crystal Fiber (PCF)-enhanced Raman analysis. Using these modalities, a multimodal, non-invasive diagnostic approach is designed to improve sensitivity and specificity while addressing existing clinical gaps. Our findings provide crucial insights into the biochemical and structural alterations in saliva and OPMD tissues, highlighting the potential of biophotonics in clinical diagnostics.

We also attempted to involve Diffuse Optical Imaging (DOI) in the designed multimodal system; however, this technique required further calibration and optimisation.

6.1 Raman spectroscopy in diagnosis of OPMD

Our Raman spectroscopy analysis (presented in manuscript IV) successfully distinguished between OPMD and healthy controls in both tissue and saliva, achieving an accuracy of 78% for saliva and 89% for tissue, through Lasso feature selection and LDA classification. There were 13 selected spectral features common to both tissue and saliva that exhibited a consistent trend across both sample types, supporting the hypothesis that saliva can serve as a non-invasive alternative for OPMD screening. The spectral features at 501 cm^{-1}, 541 cm^{-1} and 1463 cm^{-1} attributed to vibrational modes of proteins and lipids (present in tissue or soluble molecules present in saliva) showed a strong correlation between saliva and tissue. These findings align with previous studies indicating that Raman spectroscopy can identify OPMD-associated biomarkers, such as changes in carbohydrates, lipids, and proteins. While prior literature has focused mainly on differentiating control and OPMD group, either in tissue or saliva, this study highlights the capability of saliva as a first-line screening tool for OPMD by identifying common spectral features in saliva and *in vivo* tissue from the same participants. These spectral changes indicate the onset of the disease and correlate the biochemical changes in saliva with those in tissue.

6.2 Photonic Crystal Fiber for Liquid Saliva Analysis

A key novelty of this work was the application of optofluidic PCF for liquid saliva analysis, demonstrating its feasibility for detecting OPMD-related biochemical signatures. Although the Raman analysis described in manuscript IV is sensitive to compositional changes in dry saliva, it struggles to capture structural alterations effectively in liquid saliva due to the low concentration of analytes [1, 242]. Drying the saliva sample could cause molecular and structural changes, affecting the

stability and integrity of biomarkers and compromising the reliability of the diagnostic assay [3, 243]. The optical confinement within the PCF core significantly enhances the interaction between incident light and liquid saliva overcoming the challenges of conventional Raman analysis. Moreover, analysis of liquid saliva can enables real-time screening and eliminates the need for time-consuming drying process. The spectral trend observed in the liquid saliva was in agreement with the spectral changes in the *in vivo* tissue of the same participants, confirming that the changes observed in the liquid saliva are due to OPMD. Significant spectral variations were observed at 1123 cm^{-1}, 1251 cm^{-1} and 1454 cm^{-1} which correspond to carbohydrates, proteins and lipids, respectively. The results suggest that PCF-based Raman spectroscopy holds potential as a real-time and non-invasive diagnostic platform for OPMD screening by enabling the monitoring of subtle molecular changes in liquid saliva.

6.3 Diffuse Reflectance Spectroscopy as a Complementary Tool

The DRS analysis further complements the potential of non-invasive spectral diagnosis. The changes observed in the DRS spectra at 400 - 600 nm and 600 - 700 nm were associated with changes in oxygenation and increased haemoglobin absorption, consistent with previous reports on vascular changes associated with dysplasia[244, 245]. The results indicate that DRS can provide structural and functional insights into tissue pathology, providing supplementary information to biochemical analysis performed by Raman spectroscopy. Significantly high 540/575 ratios in OPMD lesions compared to contralateral healthy mucosa further validate the potential of DRS for distinguishing between affected and unaffected tissues. Unlike previous studies, this research demonstrates its utility in early-stage OPMD diagnosis, establishing DRS as a complementary tool for OPMD screening, assisting in targeted biopsy and clinical guidance for margin resection during surgery. The LDA model based on LASSO features was able to classify the cheek mucosa and ventral surface of the tongue with 88 % and 86 % accuracies, respectively.

Raman spectroscopy excels in detecting biochemical changes at the molecular level from saliva and superficial layers of tissue, making it suitable for identifying OPMD-associated spectral markers. On the other hand, DRS provides critical insights into tissue oxygenation and vascular changes from deep tissue layers, offering complementary information that Raman alone cannot provide, confirming our hypothesis given in chapter 1. By integrating both Raman and DRS in a single multimodal platform, a more comprehensive diagnostic framework is structured, where Raman detects molecular changes indicative of dysplasia, and DRS strengthens the assessment by analysing vascular and oxygenation patterns. This enhances the diagnostic accuracy and improves clinical decision-making for OPMD screening.

6.4 Multimodal Spectroscopic Approach and clinical implication

Individually, DRS, Raman spectroscopy and PCF-enhanced Raman offer valuable diagnostic capabilities. When combined, they create a more comprehensive and robust diagnostic framework. The proposed multimodal approach provides a stepwise screening: (i) saliva-based Raman spectroscopy as a preliminary non-invasive test, (ii) confirmation through *in vivo* tissue analysis using Raman and DRS in patients flagged by saliva analysis, and (iii) guided biopsy based on spectral findings from *in vivo* tissue analysis. This structured approach can reduce sampling errors, unnecessary biopsies, and patient discomfort, addressing the critical limitations of current clinical practices. It could also be helpful for accurate margin detection during surgical procedures.
The correlation between salivary biomarkers and tissue pathology reinforces the concept of saliva as a non-invasive diagnostic alternative. Integrating machine learning algorithms such as Lasso and LDA enhances the diagnostic accuracy and reproducibility of the system. The developed multimodal mobile system could be seamlessly integrated into the current clinical environment, providing bedside diagnosis. Future work should focus on expanding the sample size, validating the findings in multicentre studies, and integrating automated decision support systems to make the system more reliable in clinical settings.

Conclusion

The integration of Raman spectroscopy, DRS and DOI into a single multimodal spectroscopic platform represents a significant advancement in our ability to detect OPMD. The key findings include:

- Raman spectroscopy can differentiate OPMD from the control group in both saliva (with 78 % accuracy) and tissue (with 89 % accuracy) by identifying consistent spectral markers in both sample types.
- PCF-based Raman spectroscopy enables liquid saliva screening, offering a real-time platform for biochemical analysis and reducing the number of unnecessary biopsies.
- DRS can analyse the change in oxygenation and vascularization during OPMD, complementing biochemical insights from Raman spectroscopy.
- Machine learning-based feature selection can improve diagnostic accuracy and facilitate clinical translation.
- The multimodal stepwise diagnostic approach can improve accuracy and reduce reliance on invasive diagnostic procedures.

These findings contribute to real-time, non-invasive optical diagnostics, bridging the gap between research and clinical application. By addressing existing challenges, this thesis provides the foundation for integrating spectroscopic screening into current clinical workflow to improve patient outcomes. It also opens new avenues for diagnostics in broader medical applications, including colorectal cancer. In the future, it would also be worth including fluorescence spectroscopy for metabolic and molecular insights, and Diffuse Optical Imaging (DOI) for structural imaging. This

comprehensive approach will provide multiple complementary aspects of OPMD pathology to improve diagnostic accuracy. Further research should also focus on large-scale validation, optimisation of multimodal integration, and real-time data analysis for improved clinical decision-making.

7 Personal Perspective

This project is an exploration of what non-invasive diagnostic can achieve when innovative methodologies meet a persistent clinical challenge. This work is an attempt to democratise OPMD diagnostics, shifting paradigm from invasive, resource intensive and complex procedure to non-invasive, real-time, and accessible solution. The beauty of this research lies in its simplicity, leveraging from simple optical spectroscopy like DRS and Raman and integrating them into a compact, versatile multi-configurational system with the aim to provide a patient-centred diagnostic alternative and a saliva screening procedure, addressing the clinical gaps in OPMD diagnosis. This approach goes beyond conventional biopsy by offering insight into tissue physiology and biochemical changes with high precision. The potential of this research for real-time OPMD screening and clinical applicability is critical and clear. But still there are many challenges in refining the system for robust, large scale clinical application, such as robust probes and laser systems[1]. For me, this research was not just about engineering or biology—it was about rethinking how we approach disease detection entirely.

1: Half way during the patient recruitment, the laser broke down, and further recruitment was kept on hold for 3 months.

The development of a mobile multimodal spectroscopic system was a transformative aspect of this research, addressing the limitations of existing setups, such as high cost, limited clinical adaptability and lack of portability. Moreover, this system can be used seamlessly for *in vivo* and *ex vivo* analysis. The clinical adaptability and applicability in diverse healthcare settings is a critical aspect for its application, particularly in resource-constrained environments.

What stands out to me the most is the elegance of using saliva. It is both ordinary and extraordinary, containing molecular echoes of systemic health and disease. Despite the progress achieved, it also presents inherent challenges [2]. For example, while Raman spectroscopy of saliva samples demonstrated promising potential as a non-invasive screening tool, the variability in sample preparation, the low concentration of biomarkers, and spectral noise posed significant hurdles[3]. Similarly, DRS provides significant insight into tissue morphology and blood oxygenation, however, probe pressure on the tissue and contact angle between probe and tissue remain a complex challenge. These limitations highlight the need for ongoing refinement in both hardware design and data processing algorithms.

2: Although advised not to wear lip cosmetics and to avoid food and drinks, sometimes participants forgot to adhere to these instructions, which could lead to potential contamination or variability in the collected samples, impacting the consistency and reliability of the data.

One of the most exciting aspects of this research was the alignment of disease associated chemical changes with machine learning based selected spectral feature from Raman and DRS data. Correlating the OPMD associated spectral biomarkers from saliva with the spectral biomarkers present in tissue was a cornerstone of this research. This correlation is a significant step towards reducing invasive procedures and expanding access to routine screenings. The integration of DRS has added depth to the system, enabling dual-mode assessments that capture macroscopic and molecular-level changes in tissue. It offers a combination that few diagnostic tools can claim. By combining domain knowledge and computational techniques, accuracies that reinforce the

3: Though use of saliva is an innovative approach, but it comes with a lot of uncertainties. Could the variability in saliva be due to diet, hydration, stress, medications or other factors? Can these factors hinder its reliability?

clinical relevance of these approaches were achieved. The dependency on computational models also revealed the need for robust validation strategies and potential biases due to small or unbalanced data sets. Standardising data acquisition protocols and expanding the data set is an important requirement for future studies. In addition, developing robust and standardised preprocessing techniques is essential to ensure reliable spectral analysis [4].

Looking ahead, this research lays the groundwork for a new generation of optical diagnostic tools. Future efforts could focus on integrating additional modalities, such as fluorescence spectroscopy, hyperspectral imaging, or nuclear magnetic resonance imaging (NMR) to enhance diagnostic specificity. Also, exploring additional biomarkers and refining feature selection methods could enhance the diagnostic accuracy of saliva-based assays for OPMD. Furthermore, miniaturisation of the Raman-DRS multimodal system and its integration with handheld devices or endoscopes could broaden its clinical applications, from OPMD screening to intraoperative margin evaluation. This project was a collaboration between scientists, clinicians, engineers and data analysts. It demonstrates the confluence of applied research, clinical perspective, engineering innovation and rigorous analysis to bridge the gap between laboratory research and bedside application. By addressing current limitations and building on the foundation established in this work, the potential for widespread adoption of non-invasive optical diagnostics in oral healthcare appears both promising and achievable.

4: Baseline correction and fluorescence removal is challenging in Raman spectroscopy due to the strong overlapping fluorescence signal from biological tissues, which can obscure weak Raman features.

Comments on the Papers 8

This chapter summarizes the publications resulting from this research. Each publication details distinct phases of the study, from system development to clinical application, demonstrating the potential of Raman and DRS to address gaps in OPMD diagnosis.

1. **Maryam, S.**; Nogueira, M. S.; Gautam, R.; Krishnamoorthy, S.; Venkata Sekar, S. K.; Kho, K. W.; Lu, H.; Riordain, N. R.; Feeley, L.; Sheahan, P.; Burke, R.; Andersson-Engels, S.; **"Label-free optical spectroscopy for early detection of oral cancer. Diagnostics"**, 2022, 12, 2896 [1].
   This paper provides a comprehensive overview of the potential of optical spectroscopy in clinical settings. It highlights the importance of accessible and rapid diagnostics and underscores the importance of integrating complementary spectroscopic techniques such as Raman and diffuse reflectance spectroscopy, which are central to my research, in current clinical environment. Additionally, it emphasises the use of saliva as a diagnostic medium for non-invasive screening purposes. This review summarises the important salivary biomarkers associated with oral cancer and their concentration in healthy and OSCC participants. The review advocates for the development of more advanced optical systems, incorporating sophisticated data analysis techniques and large-scale clinical validation, to make these tools suitable for routine clinical use. I wrote this manuscript. M. Nogueira, R. Gautam, S. Krishnamoorthy. S.K.V. Sekar, K.W. Kho, H. Lu, R. Ni Riordain, L. Feeley, P. Sheahan, R. Burke and S. Andersson-Engels were involved in discussions, conceptualization and editing of the manuscript.
   This publication was selected as the front cover of the journal "Diagnostics", Volume 12, Issue 12, in which it was published.

2. **Maryam, S.**; Sekar, S. K. V.; Ghauri, M. D.; Fahy, E.; Nogueira, M. S.; Lu, H.; Beffara, F.; Humbert, G.; Riordain, R. N.; Sheahan, P.; Burke, R.; Kho; K.W.; Gautam,R.; Andersson-Engels, S.; **"Mobile multi-configuration clinical translational Raman system for oral cancer application"**, Analyst 2023, 148,1514–1523 [2].
   This publication reported the development of a mobile, adaptable and cost-effective Raman system designed to address the challenges in *in vivo* diagnostics and the limitations of existing Raman setups, including their high cost and lack of adaptability for diverse clinical applications. The system incorporates three configurations, including a Raman microscope for high resolution *ex vivo* analysis, an optofluidics PCF stage to analyze saliva in liquid form and customized *in vivo* Raman multimodal probe. The development of this novel multimodal system represents a cornerstone of this research. The system was optimized and tested robustly multiple times for accuracy and reproducibility for different modes using different analytes. The developed configuration enables detailed Raman analysis from cellular-level investigations to real-time clinical

diagnostics.The study also support the integration of the developed Raman system in current clinical workflow for non-invasive screening and monitoring of OPMD. In this study, I participated in development of the multi-configuration Raman system, system refinement, experimental validation and data analysis to ensure system functionality across different configurations. Additionally, I authored the manuscript, reviewed and edited it to align the publication with high scientific and editorial standards. S.K.V. Sekar and R. Gautam were involved in system design, fabrication and validation. M.D. Ghauri and R.Gautam were also involved in clinical data collection. E. Fahy, R. Ni Riordain were involved in patient recruitment, saliva collection and *in vivo* tissue data collection. F. Beffara and G. Humbert fabricated and provided PCF. S. Andersson-Engels provided funding for this project and was involved in conceptualisation, supervision and project management. All above authors and M. Nogueira, K.W. Kho, H. Lu, L. Feeley, P. Sheahan and R. Burke were involved in discussions, conceptualisation, and editing of the manuscript.

This publication was selected for the "Researcher Development and Travel Grant from Royal Society of Chemistry in year 2023." A poster based on this study was awarded with the first prize for best research poster award at "IPIC Industry Workshop" in year 2023.

3. **Maryam, S.**; Benazza, A.; Fahy, E.; Sekar, S. K. V.; Dinish, U.S.; Olivo, M.; Riordain, R. N.; Andersson-Engels, S.; Humbert, G.; Komolibus, K.; Gautam,R.; **"Liquid saliva analysis using optofluidic photonic crystal fiber for detection of oral potentially malignant disorders"**, Published in Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy.

   This publication offers a compelling perspective on an innovative use of optofluidic photonic crystal fibers (PCFs) for Raman analysis of biofluids. This study highlights the untapped potential of PCF, to overcome traditional limitations in liquid saliva analysis due to the low concentration of bioanalytes. By maintaining saliva in the liquid state, the methodology preserves molecular integrity, ensuring the native biochemical environment of the sample. This publication reinforces the idea that saliva can serve as a non-invasive diagnostic medium, particularly when coupled with PCF. This study validates salivary biomarkers with those present in *in vivo* tissue and confirms the efficacy of PCF to enhance the Raman signal of biofluids. In this paper my contributions were paper writing, review and editing. Moreover, I, S.K.V. Sekar, R. Gautam, A. Benazza and K. Komolibus were involved in system design and fabrication and refining the system assembly. E.Fahy and R. Ni. Riordain were responsible for patient recruitment and saliva collection, and I was responsible for processing and analysing the saliva samples. I and R. Gautam also performed the Raman data pre-processing and analysis. Georges Humbert and A. Benazza designed and developed the customised PCF. R. Gautam, K. Komolibus, Dinish U.S, M. Olivo, and S. Andersson-Engels were involved in discussions and editing of the manuscript and supervision and project management.
4. **Maryam, S.**; Innocente, S.; Fahy, E.; Ghauri, M. D.; Sekar, S. K. V.; Burke, R.; Visentin, A.; Riordain, R. N.; Andersson-Engels, S.;

Komolibus, K.; Gautam,R.; **"Identification of salivary biomarkers for screening of Oral Potentially Malignant Disorders through tissue correlation"**
This manuscript is focused on exploring Raman spectroscopy of saliva samples as a non-invasive and cost-effective screening tool for OPMD and OC diagnosis. Traditional biopsy is invasive, painful and expensive. It limits its use for routine screening and early diagnosis. Saliva, on the other hand, offers a non-invasive and cost effective alternative with significant diagnostic potential. Being in direct contact with the oral lesion, it can reflect biochemical changes associated with the malignancy. By leveraging the molecular insights provided by Raman spectroscopy, saliva analysis could reduce the need for unnecessary biopsies while enhancing the accuracy of informed biopsy decisions. This approach has the potential to improve patient comfort and enable regular monitoring of high-risk individuals, contributing to better clinical outcomes and early intervention. This study presents the first comprehensive analysis correlating Raman spectral features in saliva and tissue from the same individuals diagnosed with OPMD. The results underscore the potential of Raman spectroscopy, combined with advanced feature selection techniques, in identifying critical disease markers, thereby aiding in non-invasive diagnostics. To the best of our knowledge, this is the first study that co-relate the OPMD associated salivary biomarkers with those present in the *in vivo* tissue.
The planning and design of this work was discussed collaboratively with all coauthors. I and R. Gautam contributed to study design, clinical data collection from *in vivo* tissue, conducted experimental measurements, processed and analysed saliva samples, preprocessed Raman data, developed machine learning code, performed data analysis and led the manuscript writing.Sekar, S.K. V. contributed to troubleshooting optical setup alignment and probe design. E. Fahy, D. Ghauri contributed to clinical data collection from the *in vivo* tissue. E. Fahy and R. Ni Riordain were responsible for patient recruitment saliva samples collection. L. Feelay provided insight on biopsy and the biopsy results for the recruited participants. S. Innocente, N. Shahid and A. Visentin played a key role in machine learning code development and data analysis. K. Komolibus, R.Gautam, R. Burke,R. Ni Riordain, P. Sheahan and S. Andersson-Engels were involved in conceptualization, manuscript review, editing, supervision and project management.

5. **Maryam, S.**; Ghauri, M. D.; Fahy, E.; Innocente, S.; Burke, R.; Visentin, A.; Riordain, R. N.; Feelay, L.; Andersson-Engels, S.; Gautam,R.; Komolibus, K.;**"Diffuse Reflectance Spectroscopy for non-invasive screening of Oral Potentially Malignant Disorders".**
This study reports the potential of DRS to differentiate between healthy and OPMD tissue across different anatomical location in the oral cavity. The primary objective of this publication was to highlight the key spectral features that can distinguish between two tissue groups and diagnose OPMD. This manuscript reports the increase in diffused reflection at 540 and 575 ($R_{540/575}$) nm due to increased vascularization and a change in oxygenation status in OPMD tissue. Moreover, decreased reflectance between 600 – 700

nm indicates high oxygen consumption due to increased metabolic activities. This manuscript confirms that when integrated with machine-learning based feature selection model, DRS has a potential of revolutionizing the early diagnosis of OPMD. Its non-invasive nature coupled with real time analysis, positions DRS as an effective adjunct to traditional biopsy. By enabling *in vivo* assessments, DRS can offer valuable insights into tissue oxygenation, haemoglobin content, and scattering properties, which are often indicative of malignant transformations. Furthermore, its integration into clinical workflows could facilitate regular screening and provide clinicians with guidance to identify high-risk lesions or determine biopsy sites. In this publication, I was involved in system design and customization, *in vivo* tissue data collection from human participants, data processing, analysis, machine learning code development, manuscript writing. E. Fahy, R. Ni. Riordain and P. Sheahan were responsible for patient recruitment. E. Fahy and M. D. Ghauri, R. Gautam, K. Komolibus were also involved in *in vivo* data collection. L. Feelay provided insight on biopsy and the biopsy results for the recruited participants. S. Innocente, N. Shahid and A. Visentin contributed in machine learning code development and data analysis. R. Gautam, K. Komolibus, R. Ni Riordain, P. Sheahan, R. Burke and S. Andersson-Engels were involved in conceptualisation and supervision. All authors contributed to the discussions, reviewing and editing of the manuscript.
A poster based on this study was awarded with the best research poster award at "Irish association of oral surgery conference" in year 2023.

6. Nogueira, M.S.;**Maryam, S.**; Amissah, M.; Lu, H.; Lynch, N.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S.; **"Evaluation of wavelength ranges and tissue depth probed by diffuse reflectance spectroscopy for colorectal cancer detection."** Scientific Reports, 2021. Vol. 11, no.1, p: 1-17 [4].
This study reported the evaluation of the diagnostic capabilities of DRS in colorectal cancer detection. It also describes the potential of the source-to-detecter distance in probe design to probe the deep tissue layers. Using (SVM) models, 96 % sensitivity and up to 95 % specificity was achieved by probing the superficial and deep tissue layers, This publication demonstrates the potential of DRS for CRC diagnosis. The results obtained also suggest that using a broader spectral range can enhance diagnostic accuracy by allowing deeper tissue analysis. In this publication, I was responsible for the collection of data by probing the *ex vivo* tissue and sorting the location of the data collected on the image grid. I was also involved in the discussions, data preprocessing and manuscript review. M. S. Nogueira was the lead author in this manuscript. M. Amissah was also involved in data collection process. H. Lu was responsible for ethical approval and taking patient's consent. N. Lynch and S. Killeen were the clinical experts in this study. S. Andersson-Engels and M. O'Riordain were involved in the conceptualisation and supervision.

Conclusion

This thesis presented the development and validation of a customised multimodal spectroscopic platform integrating Raman spectroscopy and diffuse reflectance spectroscopy (DRS) for the saliva screening and non-invasive diagnosis of oral potentially malignant disorders (OPMD). The combination of these complementary optical modalities enables simultaneous assessment of biochemical and structural tissue properties, offering a more comprehensive understanding of disease progression than any single technique alone.

The experimental results represent the feasibility of using the multimodal spectroscopic system to accurately diagnose pre-malignant oral lesions and to identify key spectral biomarkers associated with pathological transformation. Furthermore, the modular design of the customised Raman system, incorporating microscope, biofluidic stage, and in vivo fiber-probe configurations, extends its applicability from laboratory-based experiments to clinical assessment in real world scenario, enhancing its adaptability for diverse sample types. Beyond OPMD diagnostics, preliminary findings suggest promising translational potential in broader medical domains, including colorectal cancer. By bridging optical physics, instrumentation, and biomedical application, this work underscores the transformative role of Biophotonics in precision diagnostics, improving patient outcomes, and shaping the future of non-invasive disease detection.

Future work will focus on clinical translation through larger patient cohort, real-time data analytics and integrating artificial intelligence for automated diagnostics for decision-making. Together, these advancements hold a promise for improving early detection strategies and patient outcomes in oncology.

A

Acknowledgement

I express deepest gratitude to my supervisors, Stefan Andersson-Engels and Ray Burke, who guided me in conducting this research project. I am truly grateful for their invaluable support, guidance and constructive discussions throughout this journey. I acknowledge the financial support and facilities from Science Foundation Ireland (SFI), the Irish Photonics Centre (IPIC), the Department of Physics at UCC, and Tyndall National Institute that made this research possible. I extend my sincere appreciation to the examiners for their time and effort in reviewing this thesis. Their valuable feedback and insights have greatly contributed to the refinement and quality of this thesis.

During my time in the Biophotonics team @ Tyndall, I have had the privilege to work with exceptional individuals. A big thanks to Rekha and Kasia, my research advisors, for sharing their knowledge, guidance, valuable input and insightful discussions throughout the process. To all the members of the Oral and GI group, especially Daniyal, Simone, Kho, Marcello and Mike, with whom I have spent most of my time working, it has been a pleasure collaborating with you all. I also would like to thank Huihui for her support in obtaining ethical approval from CREC. Micael Rajo, you were the first person I met in the group, thank you for your support, encouragement and the technical talks during all these years. I am also thankful to my collaborators at CUH, Mercy University Hospital and SIVUH for their contributions, discussions and for making this research a truly interdisciplinary effort. Special thanks to Eddie and Richael without whom this research would not have been possible.

A special thanks to: Sana, for always giving a good advice and having a solution to my technical issues; Jean, for being a wonderful friend and my python teacher; Gokhan and Andrea, for the enjoyable experiences of teaching undergraduate students at UCC and for the good laugh in long working days; Kon, for having the patience to read my manuscripts and help me to communicate in a foreign language; Walter and Lorenzo, for your support in the EPE activities and for making and improving the probe designs I needed; Baptiest for the yummy cakes, you are "The Slice Maestro"; Sraha for her support in handling logistics and ensuring everything ran smoothly behind the scenes. I would also like to extend my gratitude to: Eileen, for sharing her lab skills with me and supporting me during sample processing; Alida, for providing me with the nanoparticles for my work and guiding me on the fabrication process; Aizhan and Aimeen, for visiting and teaching me from your expertise. I hope that our time together has been as much of a learning experience for you as it has been for me. Deepest thanks to Tyndall UCC student chapter and its members, it was an honour to be a part of such an active association

comprised of brilliant minds. I am also indebted to my friends, who have always believed in me and supported me throughout this journey.

On a personal note, I am deeply grateful to my family. This PhD journey would not have been possible without their love. My deepest gratitude goes to my parents, Safia and Ashfaq, whose unwavering support, prayers, and belief in me have been my strength every step of the way. I am also incredibly thankful to my parents-in-law, Naureen and Shahid, for their kindness and encouragement throughout this journey. A special thank you to my brother, who has always been my biggest cheerleader, pushing me forward whenever I doubted myself.
To my husband, Nauman, words cannot express how grateful I am for your patience, sacrifices, and unwavering belief in me, even when my PhD took so much of my time and energy. You have been my rock, holding everything together when I felt overwhelmed. And to my little Alizeh, who came into this world in the midst of it all—you have grown up alongside my PhD, and I know it hasn't been easy. Thank you for the endless love, laughter, and the reminders of what truly matters. This journey was not just mine—it was ours.

Bibliography

Here are the references in citation order.

[1] Siddra Maryam et al. 'Label-free optical spectroscopy for early detection of oral cancer'. In: *Diagnostics* 12.12 (2022), p. 2896. DOI: https://doi.org/10.3390/diagnostics12122896 (cited on pages viii, 6, 44, 45, 49, 71, 77).

[2] Siddra Maryam et al. 'Mobile multi-configuration clinical translational Raman system for oral cancer application'. In: *Analyst* 148.7 (2023), pp. 1514–1523. DOI: 10.1039/D2AN01921C (cited on pages viii, 7, 56–61, 77).

[3] Siddra Maryam et al. 'Liquid saliva analysis using optofluidic photonic crystal fiber for detection of oral potentially malignant disorders'. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* (2025), p. 125788. DOI: https://doi.org/10.1016/j.saa.2025.125788 (cited on pages viii, 7, 59, 65, 72).

[4] Marcelo Saito Nogueira et al. 'Evaluation of wavelength ranges and tissue depth probed by diffuse reflectance spectroscopy for colorectal cancer detection'. In: *Scientific Reports* 11.1 (2021), p. 798. DOI: https://www.nature.com/articles/s41598-020-79517-2 (cited on pages viii, 7, 29, 80).

[5] Laura Marcu et al. 'Biophotonics: the big picture'. In: *Journal of biomedical optics* 23.2 (2018), pp. 021103–021103. DOI: 10.1117/1.JBO.23.2.021103 (cited on page 1).

[6] Rashin Mohammadi et al. 'Fluorescence sensing and imaging with carbon-based quantum dots for early diagnosis of cancer: A review'. In: *Journal of Pharmaceutical and Biomedical Analysis* 212 (2022), p. 114628. DOI: https://doi.org/10.1016/j.jpba.2022.114628 (cited on page 1).

[7] Kiyuk Chang and Farouc Jaffer. 'Advances in fluorescence imaging of the cardiovascular system'. In: *Journal of Nuclear Cardiology* 15.3 (2008), pp. 417–428. DOI: https://doi.org/10.1016/j.nuclcard.2008.03.001 (cited on page 1).

[8] Fahmeeda Murtaza et al. 'Prognostic features of preoperative OCT in retinal detachments: a systematic review and meta-analysis'. In: *Ophthalmology Retina* 7.5 (2023), pp. 383–397. DOI: https://doi.org/10.1016/j.oret.2022.11.011 (cited on page 1).

[9] Hazel TW Chon et al. 'Fluorescence guided surgery imaging systems for breast cancer identification: A systematic review'. In: *European Journal of Surgical Oncology* 49.5 (2023), e255. DOI: https://doi.org/10.1016/j.ejso.2023.03.148 (cited on page 1).

[10] Ravleen Nagi, Arvind Muthukrishnan, and N Rakesh. 'Effectiveness of photodynamic therapy (PDT) in the management of symptomatic oral lichen planus-A systematic review'. In: *Journal of Oral Biology and Craniofacial Research* 13.2 (2023), pp. 353–359. DOI: https://doi.org/10.1016/j.jobcr.2023.03.003 (cited on page 1).

[11] Ayepa Alain Kouadio et al. 'The use of laser Doppler flowmetry to evaluate oral soft tissue blood flow in humans: A review'. In: *Archives of oral biology* 86 (2018), pp. 58–71. DOI: https://doi.org/10.1016/j.archoralbio.2017.11.009 (cited on page 1).

[12] Alisha Rahaman et al. 'Optical imaging for early detection of cervical cancer: state of the art and perspectives'. In: *Journal of Biomedical Optics* 28.8 (2023), pp. 080902–080902. DOI: 10.1117/1.JBO.28.8.080902 (cited on page 1).

[13] Afshan Shirkavand et al. 'Application of Optical Spectroscopy in Diagnosing and Monitoring Breast Cancers: A Technical Review'. In: *Clinical Spectroscopy* (2023), p. 100027. DOI: https://doi.org/10.1016/j.clispe.2023.100027 (cited on page 1).

[14] Laura Rey-Barroso et al. 'Optical technologies for the improvement of skin cancer diagnosis: a review'. In: *Sensors* 21.1 (2021), p. 252. DOI: https://doi.org/10.3390/s21010252 (cited on page 1).

[15] Adalberto Miranda-Filho and Freddie Bray. 'Global patterns and trends in cancers of the lip, tongue and mouth'. In: *Oral oncology* 102 (2020), p. 104551. DOI: https://doi.org/10.1016/j.oraloncology.2019.104551 (cited on page 1).

[16] Freddie Bray et al. 'Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries'. In: *CA: A Cancer Journal for Clinicians* n/a.n/a (). DOI: https://doi.org/10.3322/caac.21834 (cited on pages 1, 2, 30, 31).

[17] Anastasios K Markopoulos. 'Current aspects on oral squamous cell carcinoma'. In: *The open dentistry journal* 6 (2012), p. 126. DOI: 10.2174/1874210601206010126 (cited on page 2).

[18] Marta Mazur et al. 'In vivo imaging-based techniques for early diagnosis of oral potentially malignant disorders—Systematic review and meta-analysis'. In: *International Journal of Environmental Research and Public Health* 18.22 (2021), p. 11775. DOI: 10.3390/ijerph182211775 (cited on page 2).

[19] David S Weinberg et al. 'Computed tomography colonography vs colonoscopy for colorectal cancer surveillance after surgery'. In: *Gastroenterology* 154.4 (2018), pp. 927–934. DOI: 10.1053/j.gastro.2017.11.025 (cited on page 2).

[20] M Rodríguez-Fraile et al. 'FDG PET/CT in colorectal cancer'. In: *Revista Española de Medicina Nuclear e Imagen Molecular (English Edition)* 39.1 (2020), pp. 57–66. DOI: 10.1016/j.remn.2019.09.009 (cited on page 2).

[21] Søren R Rafaelsen et al. 'CT and 3 Tesla MRI in the TN staging of colon cancer: a prospective, blind study'. In: *Current Oncology* 29.2 (2022), pp. 1069–1079. DOI: 10.3390/curroncol29020091 (cited on page 2).

[22] Herbert L Fred. 'Drawbacks and limitations of computed tomography: views from a medical educator'. In: *Texas Heart Institute Journal* 31.4 (2004), p. 345. DOI: https://pmc.ncbi.nlm.nih.gov/articles/PMC548232/ (cited on page 2).

[23] MP van der Meulen et al. 'Cost-effectiveness of colonoscopy versus CT-colonography screening for colorectal cancer with observed attendance and costs.' In: *Radiology* 287.3 (2018), p. 901. DOI: 10.1148/radiol.2017162359 (cited on page 2).

[24] Tania Vanessa Pierfelice et al. 'The Diagnostic Potential of Non-Invasive Tools for Oral Cancer and Pre-cancer: A Systematic Review'. In: *Diagnostics* 14.18 (2024), p. 2033. DOI: 10.3390/diagnostics14182033 (cited on page 2).

[25] Siddra Maryam et al. 'Multi-configuration Raman spectrometer for early stage diagnosis of oral cancer'. In: *Biomedical Vibrational Spectroscopy 2022: Advances in Research and Industry*. Vol. 11957. SPIE. 2022, pp. 20–26. DOI: https://doi.org/10.1117/12.2610704 (cited on pages 2, 9, 57).

[26] Sheng Zhao et al. 'Cerenkov luminescence imaging is an effective preclinical tool for assessing colorectal cancer PD-L1 levels in vivo'. In: *EJNMMI research* 10 (2020), pp. 1–11. DOI: https://ejnmmires.springeropen.com/articles/10.1186/s13550-020-00654-w (cited on page 2).

[27] Sherley Diaz-Mercedes et al. 'Cytology Smears: An Enhanced Alternative Method for Colorectal Cancer PN Stage—A Multicentre Study'. In: *Cancers* 14.24 (2022), p. 6072. DOI: 10.3390/cancers14246072 (cited on page 2).

[28] Matthew Fleming et al. 'Colorectal carcinoma: Pathologic aspects'. In: *Journal of gastrointestinal oncology* 3.3 (2012), p. 153. DOI: 10.3978/j.issn.2078-6891.2012.030 (cited on page 2).

[29] Tanya Walsh et al. 'Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions'. In: *Cochrane Database of Systematic Reviews* 7 (2021). DOI: 10.1002/14651858.CD010276.pub2 (cited on page 2).

[30] Aizhan Issatayeva et al. 'Suspended-core photonic crystal fibers for SERS detection of biomolecules'. In: *Journal of Optics* 27 (2025). DOI: 10.1088/2040-8986/ada9fe (cited on page 8).

[31] Marcelo Saito Nogueira et al. 'Insights into biochemical sources and diffuse reflectance spectral features for colorectal cancer detection and localization'. In: *Cancers* 14.22 (2022), p. 5715. DOI: https://doi.org/10.3390/cancers14225715 (cited on page 8).

[32] Marcelo Saito Nogueira et al. 'Diffuse reflectance spectroscopy for colorectal cancer surgical guidance: towards real-time tissue characterization and new biomarkers'. In: *Analyst* 149.1 (2024), pp. 88–99. DOI: 10.1039/D3AN00261F (cited on page 8).

[33] Marcelo Saito Nogueira et al. 'Tissue biomolecular and microstructure profiles in optical colorectal cancer delineation'. In: *Journal of Physics D: Applied Physics* 54.45 (2021), p. 454002. DOI: 10.1088/1361-6463/ac1137 (cited on pages 8, 53).

[34] Siddra Maryam et al. 'Multimodal optical spectroscopic approach to design a complete protocol for early diagnosis of oral cancer'. In: *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XXI*. Vol. 12368. SPIE. 2023, pp. 137–140. DOI: https://doi.org/10.1117/12.2650536 (cited on pages 9, 41).

[35] Siddra Maryam et al. 'Biophotonics diagnostics of oral cancer using Raman spectroscopy'. In: *European Conference on Biomedical Optics*. Optica Publishing Group. 2023, 126270B. DOI: https://doi.org/10.1117/12.2670600 (cited on page 9).

[36] Siddra Maryam et al. 'Non-invasive multimodal spectroscopic diagnosis for early-stage oral cancer'. In: *Bio-Optics: Design and Application*. Optica Publishing Group. 2023, DM4A–2. DOI: https://doi.org/10.1364/BODA.2023.DM4A.2 (cited on pages 9, 61).

[37] A Issatayeva et al. 'SERS Measurement of miRNA with the Potential Application for Liquid Biopsy'. In: *Sensors and Electronic Instrumentation Advances* (2024), p. 148 (cited on page 9).

[38] Simone Innocente et al. 'A comprehensive pipeline to integrate preprocessing and machine learning techniques for accurate classification in Raman spectroscopy'. In: *Data Science for Photonics and Biophotonics*. Vol. 13011. SPIE. 2024, pp. 51–57. DOI: https://doi.org/10.1117/12.3017024 (cited on page 9).

[39] Marcelo Saito Nogueira et al. 'Intestinal anastomosis automation through diffuse reflectance spectroscopy: towards real time guidance during robotic surgery'. In: *High-Speed Biomedical Imaging and Spectroscopy IX*. Vol. 12853. SPIE. 2024, pp. 66–70. DOI: https://doi.org/10.1117/12.3005445 (cited on page 9).

[40] Marcelo Saito Nogueira et al. 'Colorectal cancer surgical guidance by using diffuse reflectance spectroscopy'. In: *Diffuse Optical Spectroscopy and Imaging IX*. Vol. 12628. SPIE. 2023, pp. 15–20. DOI: https://doi.org/10.1117/12.2670562 (cited on page 9).

[41] Marcelo Saito Nogueira et al. 'Accurate colorectal cancer detection and delineation by probing superficial and deeper tissue biochemistry and microstructure using diffuse reflectance spectroscopy'. In: *Molecular-Guided Surgery: Molecules, Devices, and Applications VIII*. Vol. 11943. SPIE. 2022, p. 1194302. DOI: https://doi.org/10.1117/12.2609547 (cited on pages 9, 43).

[42] Marcelo Saito Nogueira et al. 'Improving colorectal cancer detection by extending the near-infrared wavelength range and tissue probed depth of diffuse reflectance spectroscopy: a support vector machine approach'. In: *Optical Biopsy XX: Toward Real-Time Spectroscopic Imaging and Diagnosis*. Vol. 11954. SPIE. 2022, pp. 12–21. DOI: https://doi.org/10.1117/12.2610497 (cited on page 10).

[43] Marcelo Saito Nogueira et al. 'Optical determination of superficial and deeper tissue biochemistry and microstructure for delineation and early detection of colorectal cancer'. In: *Translational Biophotonics: Diagnostics and Therapeutics*. Vol. 11919. SPIE. 2021, pp. 74–76. DOI: https://doi.org/10.1117/12.2614592 (cited on page 10).

[44] Marcelo Saito Nogueira et al. 'Multidisciplinary teaching and learning biophotonics: assessing the effectiveness of fully online outreach events'. In: *Optical Interactions with Tissue and Cells XXXIII; and Advanced Photonics in Urology*. Vol. 11958. SPIE. 2022, pp. 112–122. DOI: https://doi.org/10.1117/12.2608138 (cited on page 10).

[45] Marcelo Saito Nogueira et al. 'Optimization of tissue classification for colorectal cancer detection using support vector machines and diffuse reflectance spectroscopy'. In: *Translational Biophotonics: Diagnostics and Therapeutics* 11919 (2021), p. 1191929. DOI: 10.1117/12.2615033 (cited on page 10).

[46] Marcelo Saito Nogueira et al. 'Benefit of extending near-infrared wavelength range of diffuse reflectance spectroscopy for colorectal cancer detection using machine learning'. In: *Translational Biophotonics: Diagnostics and Therapeutics* 11919 (2021), p. 1191928. DOI: 10.1117/12.2615031 (cited on page 10).

[47] Marcelo Saito Nogueira et al. 'Online learning combining virtual lectures, at-home experiments and computer simulations: a multidisciplinary teaching and learning approach'. In: *Education and Training in Optics and Photonics*. Optica Publishing Group. 2021, F2B–6. DOI: https://doi.org/10.1364/ETOP.2021.F2B.6 (cited on page 10).

[48] Marcelo Saito Nogueira et al. 'Broadband diffuse reflectance spectroscopy for colorectal surgical guidance (Conference Presentation)'. In: *Optical Interactions with Tissue and Cells XXXI*. Vol. 11238. SPIE. 2020, p. 1123802. DOI: https://doi.org/10.1117/12.2546803 (cited on page 10).

[49] P Tripler and G Mosca. *Physics for scientists and engineers: Electricity, Magnetism, Light, and Elementary Modem Physics*. 2004 (cited on page 14).

[50] Richard Fitzpatrick. *Quantum mechanics*. World Scientific Publishing Company, 2015, pp. 121–127 (cited on page 14).

[51] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2019 (cited on page 14).

[52] Theresa Julia Zielinski et al. *Quantum states of atoms and molecules*. 2005. DOI: https://pubs.acs.org/doi/10.1021/ed082p1880.2 (cited on pages 14, 15).

[53] Upendra B Mahatme. 'Fundamentals of Molecular Physics'. In: *Current Perspective to Physical Science Research* (), p. 64. DOI: 10.9734/bpi/cppsr/v6 (cited on page 15).

[54] John M Brown and Alan Carrington. *Rotational spectroscopy of diatomic molecules*. Cambridge university press, 2003 (cited on page 15).

[55] Chia-Wei Sun and Ching-Cheng Chuang. *Hemodynamics study based on near-infrared optical assessment*. IntechOpen, 2012 (cited on pages 15, 16, 28).

[56] C. Boudoux. *Fundamentals of Biomedical Optics: From Light Interactions with Cells to Complex Imaging Systems*. Pollux, 2017 (cited on pages 15–17, 19, 22–28).

[57] Rickson C Mesquita and Arjun G Yodh. 'Diffuse optics: fundamentals and tissue applications'. In: *Nano Optics and Atomics: Transport of Light and Mater Waves*. IOS Press, 2011, pp. 51–74 (cited on pages 16, 28).

[58] Valery V Tuchin. 'Tissue optics and photonics: light-tissue interaction'. In: *Journal of Biomedical Photonics & Engineering* 1.2 (2015), pp. 98–134 (cited on pages 17, 18, 28).

[59] Turgut Durduran et al. 'Diffuse optics for tissue monitoring and tomography'. In: *Reports on progress in physics* 73.7 (2010), p. 076701 (cited on pages 18, 27, 28).

[60] 'transmittance'. In: (2019). DOI: doi:10.1351/goldbook.T06484 (cited on page 18).

[61] Zachary A Steelman et al. 'Light-scattering methods for tissue diagnosis'. In: *Optica* 6.4 (2019), pp. 479–489 (cited on page 18).

[62] Valerii V Tuchin. 'Light scattering study of tissues'. In: *Physics-Uspekhi* 40.5 (1997), p. 495 (cited on pages 18, 19).

[63] Judith R Mourant et al. 'Predictions and measurements of scattering and absorption over broad wavelength ranges in tissue phantoms'. In: *Applied optics* 36.4 (1997), pp. 949–957 (cited on pages 18, 19).

[64] Jeremy D Rogers et al. 'Modeling light scattering in tissue as continuous random media using a versatile refractive index correlation function'. In: *IEEE Journal of Selected Topics in Quantum Electronics* 20.2 (2013), pp. 173–186 (cited on pages 18, 19).

[65] Min Xu. 'Plum pudding random medium model of biological tissue toward remote microscopy from spectroscopic light scattering'. In: *Biomedical optics express* 8.6 (2017), pp. 2879–2895 (cited on page 19).

[66] Min Xu, Tao T Wu, and Jianan Y Qu. 'Unified Mie and fractal scattering by cells and experimental study on application in optical characterization of cellular and subcellular structures'. In: *Journal of biomedical optics* 13.2 (2008), pp. 024015–024015 (cited on pages 19, 20).

[67] Daichi Fukutomi, Katsunori Ishii, and Kunio Awazu. 'Determination of the scattering coefficient of biological tissue considering the wavelength and absorption dependence of the anisotropy factor'. In: *Optical Review* 23 (2016), pp. 291–298 (cited on page 19).

[68] Hasti Yavari. 'Diffuse Reflectance Spectroscopy: Using Multivariate analysis method for determination of tissue optical properties'. In: (2016) (cited on pages 19, 27).

[69] B.J. Inkson. '2 - Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) for materials characterization'. In: *Materials Characterization Using Nondestructive Evaluation (NDE) Methods*. Ed. by Gerhard Hübschen et al. Woodhead Publishing, 2016, pp. 17–43. DOI: https://doi.org/10.1016/B978-0-08-100040-3.00002-X (cited on page 20).

[70] W. Siegel. *Fields*. 2005 (cited on page 20).

[71] Claes af Klinteberg. *On the use of light for the characterization and treatment of malignant tumours*. Lund University, 1999 (cited on page 20).

[72] R. Paschotta. *Scattering*. RP Photonics Encyclopedia. Available online at https://www.rp-photonics.com/scattering.html. DOI: 10.61835/mkl. URL: https://www.rp-photonics.com/scattering.html (visited on 05/08/2024) (cited on page 20).

[73] Jeremy D Wilson and Thomas H Foster. 'Mie theory interpretations of light scattering from intact cells'. In: *Optics letters* 30.18 (2005), pp. 2442–2444. DOI: https://doi.org/10.1364/OL.30.002442 (cited on page 20).

[74] Jonathan Shapey et al. 'Optical properties of human brain and tumour tissue: an ex vivo study spanning the visible range to beyond the second near-infrared window'. In: *Journal of biophotonics* 15.4 (2022), e202100072. DOI: https://doi.org/10.1002/jbio.202100072 (cited on page 20).

[75] Alexander V Kolpakov et al. 'Diffuse Reflectance Spectroscopy of the Oral Mucosa: In Vivo Experimental Validation of the Precancerous Lesions Early Detection Possibility'. In: *Diagnostics* 13.9 (2023), p. 1633 (cited on pages 20, 40, 43, 45).

[76] Matthew V Bills and Jeong-Yeol Yoon. 'Label-free Mie Scattering Identification of Tumor Tissue Using an Angular Photodiode Array'. In: *IEEE sensors letters* 4.7 (2020), pp. 1–4. DOI: https://doi.org/10.1109/lsens.2020.3001489 (cited on page 20).

[77] Gábor Keresztury. 'Raman spectroscopy: Theory'. In: *Handbook of vibrational spectroscopy* (2006) (cited on page 20).

[78] John M Chalmers and Peter R Griffiths. 'Handbook of vibrational spectroscopy'. In: (2002) (cited on pages 20, 21).

[79] DA Long. 'The Raman e ect: a unified treatment of the theory of Raman scattering by molecules'. In: *Chichester and New York: Wiley* (2002) (cited on pages 20, 22).

[80] Rajinder Singh. 'CV Raman and the Discovery of the Raman Effect'. In: *Physics in Perspective* 4 (2002), pp. 399–420 (cited on page 21).

[81] Alexander Rzhevskii. *Modern raman microscopy: technique and practice*. Cambridge Scholars Publishing, 2021 (cited on page 22).

[82] Ewen Smith and Geoffrey Dent. *Modern Raman spectroscopy: a practical approach, chapter 1*. John Wiley & Sons, 2019 (cited on pages 22, 23).

[83] John R Ferraro. *Introductory raman spectroscopy*. Elsevier, 2003 (cited on pages 23–25).

[84] Robin Jon Hawes Clark and Derek Albert Long. *Raman spectroscopy*. Wiley, 1988 (cited on page 23).

[85] Herman A Szymanski. *Raman spectroscopy: theory and practice*. Springer Science & Business Media, 2012 (cited on page 24).

[86] Gerd Keiser. 'Light-Tissue Interactions'. In: *Biophotonics: Concepts to Applications*. Singapore: Springer Singapore, 2016, pp. 147–196. DOI: 10.1007/978-981-10-0945-7_6 (cited on page 25).

[87] Aise Seda Artis. *Hemodynamics: New Diagnostic and Therapeutic Approaches*. BoD–Books on Demand, 2012 (cited on page 25).

[88] Elibet Ambrocio. 'A Self-Consistent Obstacle Scattering Theory for the Diffusion Approximation of the Radiative Transport Equation'. In: *Applied Mathematics. University of California, Merced. Master of Science* (2008) (cited on page 26).

[89] David A Boas, Constantinos Pitris, and Nimmi Ramanujam. *Handbook of biomedical optics*. CRC press, 2016. Chap. Forward models of light transport in biological tissue, pp. 319–327 (cited on page 26).

[90] Richard C Haskell et al. 'Boundary conditions for the diffusion equation in radiative transfer'. In: *JOSA A* 11.10 (1994), pp. 2727–2741 (cited on pages 26, 27).

[91] Lihong V Wang and Hsin-i Wu. *Biomedical optics: principles and imaging*. John Wiley & Sons, 2007 (cited on page 27).

[92] MN Kiseleva, EP Zarochentseva, and Dodonova NIa. 'Absorption spectra of nucleic acids and related compounds in the spectral region 120–280 nm'. In: *Biofizika* 20.4 (1975), pp. 561–565 (cited on page 28).

[93] Saumya Prasad et al. 'Near UV-Visible electronic absorption originating from charged amino acids in a monomeric protein'. In: *Chemical science* 8.8 (2017), pp. 5416–5433 (cited on page 28).

[94] Peter Macheroux. 'UV-visible spectroscopy as a tool to study flavoproteins'. In: *Flavoprotein protocols* (1999), pp. 1–7 (cited on page 28).

[95] BT Doumas, Tai-Wing Wu, and B Jendrzejczak. 'Delta bilirubin: absorption spectra, molar absorptivity, and reactivity in the diazo reaction.' In: *Clinical chemistry* 33.6 (1987), pp. 769–774 (cited on page 28).

[96] Giovanni Agati and Franco Fusi. 'New trends in photobiology recent advances in bilirubin photophysics'. In: *Journal of Photochemistry and Photobiology B: Biology* 7.1 (1990), pp. 1–14 (cited on page 28).

[97] Joy Udensi et al. 'Quantitative Raman analysis of carotenoid protein complexes in aqueous solution'. In: *Molecules* 27.15 (2022), p. 4724 (cited on page 28).

[98] Rami Nachabé et al. 'Validation of interventional fiber optic spectroscopy with MR spectroscopy, MAS-NMR spectroscopy, high-performance thin-layer chromatography, and histopathology for accurate hepatic fat quantification'. In: *Investigative radiology* 47.4 (2012), pp. 209–216 (cited on page 28).

[99] Robert LP Van Veen et al. 'Determination of VIS-NIR absorption coefficients of mammalian fat, with time-and spatially resolved diffuse reflectance and transmission spectroscopy'. In: *Biomedical Topical Meeting*. Optica Publishing Group. 2004, SF4 (cited on page 28).

[100] Rita Giovannetti et al. 'The use of spectrophotometry UV-Vis for the study of porphyrins'. In: *Macro to nano spectroscopy* (2012), pp. 87–108 (cited on page 28).

[101] Kazuya Ogawa, Yuria Umetsu, and Kenji Kamimura. 'Changes in the absorption spectra and colour of tetraphenylporphyrins after redox reactions'. In: *Journal of Chemical Research* 44.9-10 (2020), pp. 613–617 (cited on page 28).

[102] J.M. Schmitt. *Optical Measurement of Blood Oxygen by Implantable Telemetry*. Stanford University, 1986 (cited on page 28).

[103] Frank S Parker and Frank S Parker. 'Amides and amines'. In: *Applications of infrared spectroscopy in biochemistry, biology, and medicine* (1971), pp. 165–172 (cited on page 28).

[104] Paola Taroni et al. 'Assessment of collagen absorption and related potential diagnostic applications'. In: *European Conference on Biomedical Optics*. Optica Publishing Group. 2007, 6629_12 (cited on page 28).

[105] Amadeo B Biter et al. 'A method to probe protein structure from UV absorbance spectra'. In: *Analytical biochemistry* 587 (2019), p. 113450 (cited on page 28).

[106] A Robert Goldfarb and Leo J Saidel. 'Ultraviolet absorption spectra of proteins'. In: *Science* 114.2954 (1951), pp. 156–157 (cited on page 28).

[107] Nikiforos Kollias and Ali H Baqer. 'Absorption mechanisms of human melanin in the visible, 400–720 nm'. In: *Journal of investigative dermatology* 89.4 (1987), pp. 384–388 (cited on page 28).

[108] Hao Ou-Yang, Georgios Stamatas, and Nikiforos Kollias. 'Spectral Responses of Melanin to Ultraviolet A Irradiation'. In: *Journal of Investigative Dermatology* 122.2 (2004), pp. 492–496. DOI: https://doi.org/10.1046/j.0022-202X.2004.22247.x (cited on page 28).

[109] S Jacquemoud and SL Ustin. 'Application of radiative transfer models to moisture content estimation and burned land mapping'. In: *4th International Workshop on Remote Sensing and GIS Applications to Forest Fire Management*. Vol. 312. 2003 (cited on page 28).

[110] Gregory A Carter and Douglas C McCain. 'Relationship of leaf spectral reflectance to chloroplast water content determined using NMR microscopy'. In: *Remote Sensing of Environment* 46.3 (1993), pp. 305–310 (cited on page 28).

[111] Jan Sørensen Dam. 'Optical analysis of biological media-continuous wave diffuse spectroscopy'. In: (2000) (cited on page 28).

[112] Brondwyn S McGhie and Janice R Aldrich-Wright. 'Photoactive and luminescent transition metal complexes as anticancer agents: A guiding light in the search for new and improved cancer treatments'. In: *Biomedicines* 10.3 (2022), p. 578 (cited on page 28).

[113] Yunhan Tan et al. 'Oral squamous cell carcinomas: state of the field and emerging directions'. In: *International Journal of Oral Science* 15.1 (2023), p. 44. DOI: https://doi.org/10.1038/s41368-023-00249-w (cited on pages 30, 32).

[114] David I Conway, M Purkayastha, and IG Chestnutt. 'The changing epidemiology of oral cancer: definitions, trends, and risk factors'. In: *British dental journal* 225.9 (2018), pp. 867–873. DOI: https://www.nature.com/articles/sj.bdj.2018.922 (cited on page 30).

[115] Loredana Radoi et al. 'Population attributable risks of oral cavity cancer to behavioral and medical risk factors in France: results of a large population-based case–control study, the ICARE study'. In: *BMC cancer* 15.1 (2015), p. 827. DOI: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-015-1841-5?utm_source=chatgpt.com (cited on page 30).

[116] David Zaridze Mia Hashibe Paul Brennan. 'Risk factors for head and neck cancers in Eastern Europe, a high risk region'. In: *Cancer Research* 64 (7 supplement 2004), pp. 857–858. DOI: https://aacrjournals.org/cancerres/article/64/7_Supplement/857/515157/Risk-factors-for-head-and-neck-cancers-in-Eastern (cited on page 30).

[117] Adam Barsouk et al. 'Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma'. In: *Medical Sciences* 11.2 (2023), p. 42. DOI: https://doi.org/10.3390/medsci11020042 (cited on page 30).

[118] Apurva Garg et al. 'A review on harmful effects of pan masala'. In: *Indian journal of cancer* 52.4 (2015), pp. 663–666. DOI: 10.4103/0019-509X.178449 (cited on page 30).

[119] Anwar T Merchant and Waranuch Pitiphat. 'Total, direct, and indirect effects of paan on oral cancer'. In: *Cancer causes & control* 26.3 (2015), pp. 487–491. DOI: https://doi.org/10.1007/s10552-014-0516-x (cited on page 30).

[120] Kamal Niaz et al. 'Smokeless tobacco (paan and gutkha) consumption, prevalence, and contribution to oral cancer'. In: *Epidemiology and health* 39 (2017), e2017009. DOI: 10.4178/epih.e2017009 (cited on page 30).

[121] K Ranganathan et al. 'Oral submucous fibrosis: a case-control study in Chennai, South India'. In: *Journal of oral pathology & medicine* 33.5 (2004), pp. 274–277. DOI: https://doi.org/10.1111/j.0904-2512.2004.00116.x (cited on page 30).

[122] N Shah and PP Sharma. 'Role of chewing and smoking habits in the etiology of oral submucous fibrosis (OSF): a case-control study'. In: *Journal of oral pathology & medicine* 27.10 (1998), pp. 475–479. DOI: https://doi.org/10.1111/j.1600-0714.1998.tb01915.x (cited on page 30).

[123] Saman Warnakulasuriya and THH Chen. 'Areca nut and oral cancer: evidence from studies conducted in humans'. In: *Journal of dental research* 101.10 (2022), pp. 1139–1146. DOI: https://doi.org/10.1177/00220345221092751 (cited on pages 30, 32).

[124] Hyuna Sung et al. 'Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries'. In: *CA: a cancer journal for clinicians* 71.3 (2021), pp. 209–249. DOI: https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21660 (cited on page 31).

[125] G Tirelli et al. 'Prognosis of oral cancer: a comparison of the staging systems given in the 7th and 8th editions of the American Joint Committee on Cancer Staging Manual'. In: *British Journal of Oral and Maxillofacial Surgery* 56.1 (2018), pp. 8–13. DOI: 10.1016/j.bjoms.2017.11.009 (cited on page 31).

[126] PM Speight and PM Farthing. 'The pathology of oral cancer'. In: *British dental journal* 225.9 (2018), pp. 841–847. DOI: 10.1038/sj.bdj.2018.926 (cited on page 31).

[127] 'National Cancer Registry Ireland (2022) Cancer in Ireland 1994-2020: Annual statistical report of the National Cancer Registry. NCRI, Cork, Ireland'. In: (). DOI: https://www.ncri.ie/sites/ncri/files/pubs/NCRI_AnnualStatisticalReport_2022.pdf (cited on page 31).

[128] Nicola Cirillo. 'Precursor Lesions, Overdiagnosis, and Oral Cancer: A Critical Review'. In: *Cancers* 16.8 (2024), p. 1550. DOI: 10.3390/cancers16081550 (cited on pages 31–33).

[129] Isabel González-Ruiz et al. 'Early diagnosis of oral cancer: A complex polyhedral problem with a difficult solution'. In: *Cancers* 15.13 (2023), p. 3270. DOI: https://doi.org/10.3390/cancers15133270 (cited on page 31).

[130] Eric P Tranby et al. 'Oral cancer prevalence, mortality, and costs in Medicaid and commercial insurance claims data'. In: *Cancer Epidemiology, Biomarkers & Prevention* 31.9 (2022), pp. 1849–1857. DOI: 10.1158/1055-9965.EPI-22-0114 (cited on page 31).

[131] Pablo Varela-Centelles. *Early diagnosis and diagnostic delay in oral cancer*. 2022. DOI: 10.3390/cancers14071758 (cited on page 31).

[132] Kannan Ranganathan and Loganathan Kavitha. 'Oral epithelial dysplasia: Classifications and clinical relevance in risk assessment of oral potentially malignant disorders'. In: *Journal of Oral and Maxillofacial Pathology* 23.1 (2019), pp. 19–27. DOI: 10.4103/jomfp.JOMFP_13_19 (cited on pages 31, 33).

[133] Saman Warnakulasuriya et al. 'Oral potentially malignant disorders: A consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer'. In: *Oral diseases* 27.8 (2021), pp. 1862–1880. DOI: https://doi.org/10.1111/odi.13704 (cited on page 31).

[134] Simms Baillie and W Simms. 'Queries and responses from the Medical Committee of the Society for Investigating the Nature and Cure of Cancer'. In: *Edinburgh Med Surg J* 2 (1806), pp. 382–9 (cited on page 31).

[135] WHO Expert Committee on the Prevention of Cancer and World Health Organization. *Prevention of Cancer: Report of a WHO Expert Committee on the Prevention of Cancer*. Technical report series. World Health Organization, 1964 (cited on page 31).

[136] S1 Warnakulasuriya, Newell W Johnson, and I Van der Waal. 'Nomenclature and classification of potentially malignant disorders of the oral mucosa'. In: *Journal of oral pathology & medicine* 36.10 (2007), pp. 575–580. DOI: 10.1111/j.1600-0714.2007.00582.x (cited on page 31).

[137] Andreas Jovanovic et al. 'Tobacco and alcohol related to the anatomical site of oral squamous cell carcinoma'. In: *Journal of oral pathology & medicine* 22.10 (1993), pp. 459–462. DOI: https://doi.org/10.1111/j.1600-0714.1993.tb00125.x (cited on page 32).

[138] Luigino Dal Maso et al. 'Combined effect of tobacco smoking and alcohol drinking in the risk of head and neck cancers: A re-analysis of case–control studies using bi-dimensional spline models'. In: *European journal of epidemiology* 31 (2016), pp. 385–393. DOI: 10.1007/s10654-015-0028-3 (cited on page 32).

[139] Rasoul Ebrahimi et al. 'Lip, Oral Cavity, and Pharyngeal Cancers: Global Epidemiology, Risk Factors, and Prevention: A Narrative Review'. In: *Health Science Reports* 8.10 (2025), e71289. DOI: https://doi.org/10.1002/hsr2.71289 (cited on page 32).

[140] Kalen R Dionne et al. 'Potentially malignant disorders of the oral cavity: Current practice and future directions in the clinic and laboratory'. In: *International journal of cancer* 136.3 (2015), pp. 503–515. DOI: https://doi.org/10.1002/ijc.28754 (cited on pages 32, 36).

[141] Pratima Kumari, Priyanka Debta, and Anshuman Dixit. 'Oral potentially malignant disorders: etiology, pathogenesis, and transformation into oral cancer'. In: *Frontiers in pharmacology* 13 (2022), p. 825266. DOI: 10.3389/fphar.2022.825266 (cited on page 32).

[142] Faraz Mohammed and Arishiya T Fairozekhan. 'Oral leukoplakia'. In: (2017). DOI: https://www.ncbi.nlm.nih.gov/books/NBK442013/ (cited on page 32).

[143] Leon Barnes et al. 'World Health Organization classification of tumours: pathology and genetics of head and neck tumours'. In: (2005) (cited on page 33).

[144] Leon Barnes. *Pathology and genetics of head and neck tumours*. Vol. 9. IARC, 2005 (cited on page 33).

[145] Colin Smith and Jens J Pindborg. *Histological grading of oral epithelial atypia by the use of photographic standards*. C. Hamburgers Bogtrykkeri A/S, 1969 (cited on page 33).

[146] Melanie Boxberg et al. 'Novel prognostic histopathological grading system in oral squamous cell carcinoma based on tumour budding and cell nest size shows high interobserver and intraobserver concordance'. In: *Journal of clinical pathology* 72.4 (2019), pp. 285–294. DOI: https://jcp.bmj.com/content/72/4/285 (cited on page 33).

[147] Kathleen A Cronin et al. 'Annual report to the nation on the status of cancer, part 1: National cancer statistics'. In: *Cancer* 128.24 (2022), pp. 4251–4284. DOI: https://doi.org/10.1002/cncr.34479 (cited on page 34).

[148] Mahul B Amin et al. 'The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging'. In: *CA: a cancer journal for clinicians* 67.2 (2017), pp. 93–99. DOI: https://doi.org/10.3322/caac.21388 (cited on page 34).

[149] American Society of Clinical Oncology. *Oral and Oropharyngeal Cancer: Stages and Grades*. 2024. URL: https://www.cancer.net/cancer-types/oral-and-oropharyngeal-cancer/stages-and-grades (visited on 2024) (cited on page 34).

[150] C Scully. 'Challenges in predicting which oral mucosal potentially malignant disease will progress to neoplasia'. In: *Oral diseases* 20.1 (2014), pp. 1–5. DOI: 10.1111/odi.12208 (cited on page 36).

[151] R Macey et al. '700 Warnakulasuriya, S.; Scully, C. Diagnostic Tests for Oral Cancer and Potentially Malignant Disorders 701 in Patients Presenting with Clinically Evident Lesions'. In: *Cochrane Database Syst Rev* (2015). DOI: 10.1002/14651858.CD010276.pub2 (cited on page 36).

[152] Salman A Mian et al. 'Raman spectroscopy can discriminate between normal, dysplastic and cancerous oral mucosa: a tissue-engineering approach'. In: *Journal of Tissue Engineering and Regenerative Medicine* 11.11 (2017), pp. 3253–3262. DOI: 10.1002/term.2234 (cited on page 36).

[153] Mohammad Ehtisham and Sheeba Nissar. 'Oral Brush Biopsy: Pros & Cons'. In: *Heal talk* 8.2 (2015), pp. 29–30. DOI: https://www.researchgate.net/publication/305477696_Oral_Brush_Biopsy_Pros_Cons (cited on pages 36, 38).

[154] Satish Bhosale and Tarun Vyas. 'Application of oral CDx brush biopsy in oral cancer detection'. In: *J Dental Res Pract* 1.1 (2019), pp. 4–6 (cited on pages 36, 38).

[155] Y Linda Ma, David B Burr, and Reinhold G Erben. 'Bone histomorphometry in rodents'. In: *Principles of bone biology*. Elsevier, 2020, pp. 1899–1922. DOI: https://doi.org/10.1016/B978-0-12-814841-9.00082-8 (cited on page 37).

[156] Gokul Sridharan and Akhil A Shankar. 'Toluidine blue: A review of its chemistry and clinical utility'. In: *Journal of oral and maxillofacial pathology* 16.2 (2012), pp. 251–255. DOI: 10.4103/0973-029X.99081 (cited on pages 37, 40).

[157] Ravleen Nagi et al. 'Efficacy of light based detection systems for early detection of oral cancer and oral potentially malignant disorders: Systematic review'. In: *Medicina oral, patologia oral y cirugia bucal* 21.4 (2016), e447. DOI: 10.4317/medoral.21104 (cited on page 37).

[158] Winton San Kalluvelil and Veena Narayanan. 'Acetic acid versus toluidine blue as screening tools for oral potentially malignant disorders.' In: *Indian Journal of Cancer* 60.3 (2023). DOI: 10.4103/ijc.IJC_42_21 (cited on page 37).

[159] Jyotsna Rimal et al. 'Application of 5% acetic acid as a diagnostic adjunct in the detection of oral potentially malignant disorders'. In: *Journal of Kathmandu Medical College* 12.3 (2023), pp. 135–42. DOI: https://jkmc.com.np/ojs3/index.php/journal/article/view/1258 (cited on page 37).

[160] Nidhi Jain et al. 'Role of Chemiluminescence examination as non-invasive diagnostic tool in early detection of Leukoplakia'. In: *Journal of Oral Biology and Craniofacial Research* 8.3 (2018), pp. 177–181 (cited on pages 37, 38).

[161] Sonal Shah et al. 'The use of Velscope to assess cellular changes occuring in oral premalignancy'. In: *Journal of Oral Biology and Craniofacial Research* 10.2 (2020), pp. 99–103. DOI: 10.1016/j.jobcr.2020.03.004 (cited on page 37).

[162] Stefania Leuci et al. 'May VelScope be deemed an opportunistic oral cancer screening by general dentists? A pilot study'. In: *Journal of Clinical Medicine* 9.6 (2020), p. 1754. DOI: 10.3390/jcm9061754 (cited on pages 37, 38).

[163] Simran Sunil Singh et al. '"VELscope as an Adjuvant Chairside Diagnostic Tool"–A Clinical Case Series'. In: *Archives of Medicine and Health Sciences* (2023). DOI: 10.4103/amhs.amhs_81_23 (cited on pages 37, 38).

[164] Deepa Sharma et al. 'Evaluation of oral potentially malignant disorders with autoflorescence, reflectance spectroscopy and vital staining and their correlation with histopathology–Hospital based prospective study'. In: *Oral Oncology* 118 (2021), p. 105312. DOI: https://doi.org/10.1016/j.oraloncology.2021.105312 (cited on page 38).

[165] Barry L Ziober et al. 'Lab-on-a-chip for oral cancer screening and diagnosis'. In: *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck* 30.1 (2008), pp. 111–121. DOI: 10.1002/hed.20680 (cited on page 38).

[166] Guruduth Banavar et al. 'Detecting salivary host and microbiome RNA signature for aiding diagnosis of oral and throat cancer'. In: *Oral Oncology* 145 (2023), p. 106480. DOI: 10.1016/j.oraloncology.2023.106480 (cited on page 38).

[167] Teena Sajan et al. 'A rapid, early detection of oral squamous cell carcinoma: Real time PCR based detection of tetranectin'. In: *Molecular Biology Research Communications* 8.1 (2019), p. 33. DOI: 10.22099/mbrc.2019.31544.1365 (cited on page 37).

[168] Mariarosaria Boccellino, Alfredo De Rosa, and Marina Di Domenico. 'An ELISA Test Able to Predict the Development of Oral Cancer: The Significance of the Interplay between Steroid Receptors and the EGF Receptor for Early Diagnosis'. In: *Diagnostics* 13.12 (2023), p. 2001. DOI: 10.3390/diagnostics13122001 (cited on page 37).

[169] Guruduth Banavar et al. 'The salivary metatranscriptome as an accurate diagnostic indicator of oral cancer'. In: *NPJ genomic medicine* 6.1 (2021), p. 105. DOI: https://www.nature.com/articles/s41525-021-00257-x (cited on page 38).

[170] Shigeo Ishikawa et al. 'Identification of salivary metabolomic biomarkers for oral cancer screening'. In: *Scientific reports* 6.1 (2016), p. 31520. DOI: https://www.nature.com/articles/srep31520 (cited on pages 38, 40).

[171] Nina Ogrinc et al. 'Mass spectrometry-based differentiation of oral tongue squamous cell carcinoma and nontumor regions with the SpiderMass technology'. In: *Frontiers in Oral Health* 3 (2022), p. 827360. DOI: 10.3389/froh.2022.827360 (cited on pages 38, 40).

[172] Jopi JW Mikkonen et al. 'Potential role of nuclear magnetic resonance spectroscopy to identify salivary metabolite alterations in patients with head and neck cancer'. In: *Oncology letters* 16.5 (2018), pp. 6795–6800. DOI: 10.3892/ol.2018.9419 (cited on pages 39, 40).

[173] Stephen Gichuhi et al. 'Toluidine blue 0.05% vital staining for the diagnosis of ocular surface squamous neoplasia in Kenya'. In: *JAMA ophthalmology* 133.11 (2015), pp. 1314–1321. DOI: 10.4103/0973-029X.99081 (cited on page 40).

[174] Xiaobo Luo et al. 'Accuracy of autofluorescence in diagnosing oral squamous cell carcinoma and oral potentially malignant disorders: a comparative study with aero-digestive lesions'. In: *Scientific reports* 6.1 (2016), p. 29943. DOI: https://www.nature.com/articles/srep29943 (cited on page 40).

[175] Saygo Tomo et al. 'Fluorescence visualization as an auxiliary method to detect oral potentially malignant disorders and oral cancer'. In: *Translational Cancer Research* 11.8 (2022), p. 2477. DOI: 10.21037/tcr-22-1874 (cited on page 40).

[176] Luciana Estevam Simonato et al. 'Fluorescence visualization improves the detection of oral, potentially malignant, disorders in population screening'. In: *Photodiagnosis and photodynamic therapy* 27 (2019), pp. 74–78. DOI: 10.1016/j.pdpdt.2019.05.017 (cited on page 40).

[177] Vivek Nayyar et al. 'Use of fluorescence imaging and spectrometry in detection of oral squamous cell carcinoma and oral potentially malignant disorders'. In: *Oral Oncology Reports* 9 (2024), p. 100172. DOI: https://doi.org/10.1016/j.oor.2024.100172 (cited on page 40).

[178] Anne Lauren Clark. *Detection and diagnosis of oral neoplasia with confocal microscopy and optical coherence microscopy*. The University of Texas at Austin, 2003 (cited on page 40).

[179] Adrien Villard et al. 'Confocal laser endomicroscopy and confocal microscopy for head and neck cancer imaging: Recent updates and future perspectives'. In: *Oral Oncology* 127 (2022), p. 105826. DOI: https://doi.org/10.1016/j.oraloncology.2022.105826 (cited on page 40).

[180] Rishi S Ramani et al. 'Confocal microscopy in oral cancer and oral potentially malignant disorders: A systematic review'. In: *Oral Diseases* 29.8 (2023), pp. 3003–3015. DOI: 10.1111/odi.14291 (cited on page 40).

[181] Dalip Singh Mehta et al. 'Multimodal and multispectral diagnostic devices for oral and breast cancer screening in low resource settings'. In: *Current Opinion in Biomedical Engineering* (2023), p. 100485. DOI: https://doi.org/10.1016/j.cobme.2023.100485 (cited on page 40).

[182] Noah Bedard et al. 'Multimodal snapshot spectral imaging for oral cancer diagnostics: a pilot study'. In: *Biomedical optics express* 4.6 (2013), pp. 938–949. DOI: 10.1364/BOE.4.000938 (cited on page 40).

[183] Gnanatheepam Einstein et al. 'Diffuse reflectance spectroscopy for monitoring physiological and morphological changes in oral cancer'. In: *Optik* 127.3 (2016), pp. 1479–1485. DOI: https://doi.org/10.1016/j.ijleo.2015.11.045 (cited on page 40).

[184] Ruiying Han et al. 'Diagnostic accuracy of Raman spectroscopy in oral squamous cell carcinoma'. In: *Frontiers in Oncology* 12 (2022), p. 925032. DOI: 10.3389/fonc.2022.925032 (cited on page 40).

[185] Yuwei Zhang et al. 'Raman spectroscopy: A potential diagnostic tool for oral diseases'. In: *Frontiers in Cellular and Infection Microbiology* 12 (2022), p. 775236. DOI: https://doi.org/10.3389/fcimb.2022.775236 (cited on page 40).

[186] Xiaohan Chang et al. 'Deep learning methods for oral cancer detection using Raman spectroscopy'. In: *Vibrational Spectroscopy* 126 (2023), p. 103522 (cited on pages 40, 49).

[187] April Mordi et al. 'Design and Validation of a Multimodal Diffuse Reflectance and Spatially Offset Raman Spectroscopy System for In Vivo Applications'. In: *Journal of Biophotonics* 18.3 (2025), e202400333. DOI: https://doi.org/10.1002/jbio.202400333 (cited on pages 41, 45).

[188] Domenico Ribatti and Francesco Pezzella. 'Overview on the different patterns of tumor vascularization'. In: *Cells* 10.3 (2021), p. 639. DOI: 10.3390/cells10030639 (cited on page 42).

[189] Peter Vaupel, Ann Barry Flood, and Harold M Swartz. 'Oxygenation status of malignant tumors vs. normal tissues: Critical evaluation and updated data source based on direct measurements with pO2 microsensors'. In: *Applied Magnetic Resonance* 52.10 (2021), pp. 1451–1479. DOI: https://link.springer.com/article/10.1007/s00723-021-01383-6 (cited on page 43).

[190] Yue Li, Long Zhao, and Xiao-Feng Li. 'Hypoxia and the tumor microenvironment'. In: *Technology in cancer research & treatment* 20 (2021), p. 15330338211036304. DOI: 10.1177/15330338211036304 (cited on page 43).

[191] Xiao-Feng Li and Joseph A O'Donoghue. 'Hypoxia in microscopic tumors'. In: *Cancer letters* 264.2 (2008), pp. 172–180. DOI: https://doi.org/10.1158/0008-5472.CAN-06-4353 (cited on page 43).

[192] Scarlet Nazarian et al. 'Real-time tracking and classification of tumor and nontumor tissue in upper gastrointestinal cancers using diffuse reflectance spectroscopy for resection margin assessment'. In: *JAMA surgery* 157.11 (2022), e223899–e223899 (cited on page 43).

[193] Udayakumar Kaniyappan et al. 'Quantification of hemoglobin and its derivatives in oral cancer diagnosis by diffuse reflectance spectroscopy'. In: *Optical Biopsy XV: Toward Real-Time Spectroscopic Imaging and Diagnosis*. Vol. 10060. SPIE. 2017, pp. 170–175 (cited on pages 43, 45).

[194] Richard A Schwarz et al. 'Autofluorescence and diffuse reflectance spectroscopy of oral epithelial tissue using a depth-sensitive fiber-optic probe'. In: *Applied optics* 47.6 (2008), pp. 825–834 (cited on pages 43, 51).

[195] Adrien Wang, Vengadesan Nammalavar, and Rebekah Drezek. 'Targeting spectral signatures of progressively dysplastic stratified epithelia using angularly variable fiber geometry in reflectance Monte Carlo simulations'. In: *Journal of biomedical optics* 12.4 (2007), pp. 044012–044012 (cited on pages 43, 51).

[196] Susan G Brouwer de Koning et al. 'Toward complete oral cavity cancer resection using a handheld diffuse reflectance spectroscopy probe'. In: *Journal of biomedical optics* 23.12 (2018), pp. 121611–121611 (cited on page 43).

[197] Bing Yu et al. 'Diffuse reflectance spectroscopy of epithelial tissue with a smart fiber-optic probe'. In: *Biomedical optics express* 5.3 (2014), pp. 675–689 (cited on page 44).

[198] Bing Yu et al. 'Diffuse reflectance spectroscopy with a self-calibrating fiber optic probe'. In: *Optics letters* 33.16 (2008), pp. 1783–1785. DOI: https://doi.org/10.1364/OL.33.001783 (cited on page 44).

[199] Dongyuan Liu et al. 'A combined autofluorescence and diffuse reflectance spectroscopy for mucosa tissue diagnosis: Dual-distance system and data-driven decision'. In: *Journal of Biophotonics* 16.10 (2023), e202300086 (cited on page 44).

[200] Gage J Greening et al. 'Towards monitoring dysplastic progression in the oral cavity using a hybrid fiber-bundle imaging and spectroscopy probe'. In: *Scientific reports* 6.1 (2016), p. 26734 (cited on pages 44, 50).

[201] Manju M Stephen et al. 'Diagnostic accuracy of diffuse reflectance imaging for early detection of pre-malignant and malignant changes in the oral cavity: a feasibility study'. In: *BMC cancer* 13 (2013), pp. 1–9 (cited on page 44).

[202] N Subhash et al. 'Oral cancer detection using diffuse reflectance spectral ratio R540/ R575 of oxygenated hemoglobin bands'. In: *Journal of biomedical optics* 11.1 (2006), pp. 014018–014018 (cited on page 45).

[203] Sima Pazouki et al. 'The association between tumour progression and vascularity in the oral mucosa'. In: *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* 183.1 (1997), pp. 39–43. DOI: https://pubmed.ncbi.nlm.nih.gov/9370945/ (cited on page 45).

[204] Ming-Jer Jeng et al. 'Raman spectroscopy analysis for optical diagnosis of oral cancer detection'. In: *Journal of clinical medicine* 8.9 (2019), p. 1313 (cited on pages 45–47).

[205] Katie Hanna et al. 'Advances in Raman spectroscopy for characterising oral cancer and oral potentially malignant disorders'. In: *Expert Reviews in Molecular Medicine* 26 (2024), e25. DOI: https://doi.org/10.1017/erm.2024.26 (cited on pages 45, 47).

[206] Ola Ibrahim et al. 'The potential of Raman spectroscopy in the diagnosis of dysplastic and malignant oral lesions'. In: *Cancers* 13.4 (2021), p. 619 (cited on pages 45, 46).

[207] Levi Matthies et al. 'Optical diagnosis of oral cavity lesions by label-free Raman spectroscopy'. In: *Biomedical Optics Express* 12.2 (2021), pp. 836–851 (cited on pages 45, 46, 49).

[208] Froukje LJ Cals et al. 'Development and validation of Raman spectroscopic classification models to discriminate tongue squamous cell carcinoma from non-tumorous tissue'. In: *Oral Oncology* 60 (2016), pp. 41–47 (cited on page 46).

[209] Munira Essat et al. 'Diagnostic accuracy of conventional oral examination for detecting oral cavity cancer and potentially malignant disorders in patients with clinically evident oral lesions: Systematic review and meta-analysis'. In: *Head & neck* 44.4 (2022), pp. 998–1013. DOI: https://doi.org/10.1002/hed.26992 (cited on page 46).

[210] Shan Wang et al. 'Current advances in noninvasive methods for the diagnosis of oral squamous cell carcinoma: a review'. In: *European journal of medical research* 28.1 (2023), p. 53. DOI: https://doi.org/10.1186/s40001-022-00916-4 (cited on page 46).

[211] Medhanie Tesfay Gebrekidan et al. 'A shifted-excitation Raman difference spectroscopy (SERDS) evaluation strategy for the efficient isolation of Raman spectra from extreme fluorescence interference'. In: *Journal of Raman spectroscopy* 47.2 (2016), pp. 198–209 (cited on page 46).

[212] CI Faur et al. 'Raman spectroscopy in oral cavity and oropharyngeal cancer: A systematic review'. In: *International Journal of Oral and Maxillofacial Surgery* 51.11 (2022), pp. 1373–1381 (cited on page 46).

[213] Alecsandru Ioan Baba and Cornel Câtoi. 'Tumor cell morphology'. In: *Comparative oncology*. The Publishing House of the Romanian Academy, 2007 (cited on page 47).

[214] Xing Li et al. 'Rapid multi-task diagnosis of oral cancer leveraging fiber-optic Raman spectroscopy and deep learning algorithms'. In: *Frontiers in Oncology* 13 (2023), p. 1272305. DOI: https://doi.org/10.3389/fonc.2023.1272305 (cited on page 47).

[215] Po-Hsiung Chen et al. 'Automatic and objective oral cancer diagnosis by Raman spectroscopic detection of keratin with multivariate curve resolution analysis'. In: *Scientific reports* 6.1 (2016), p. 20097. DOI: https://doi.org/10.1038/srep20097 (cited on page 47).

[216] Wei-Yun Dai, Szetsen Lee, and Yih-Chih Hsu. 'Discrimination between oral cancer and healthy cells based on the adenine signature detected by using Raman spectroscopy'. In: *Journal of Raman Spectroscopy* 49.2 (2018), pp. 336–342 (cited on page 47).

[217] EM Barroso et al. 'Discrimination between oral cancer and healthy tissue based on water content determined by Raman spectroscopy'. In: *Analytical chemistry* 87.4 (2015), pp. 2419–2426 (cited on page 47).

[218] Rareș-Mario Borșa et al. 'Developing new diagnostic tools based on sers analysis of filtered salivary samples for oral cancer detection'. In: *International Journal of Molecular Sciences* 24.15 (2023), p. 12125. DOI: https://doi.org/10.3390/ijms241512125 (cited on page 48).

[219] A Falamas et al. 'Raman spectroscopic characterization of saliva for the discrimination of oral squamous cell carcinoma'. In: *Analytical Letters* 54.1-2 (2021), pp. 57–69. DOI: https://doi.org/10.1080/00032719.2020.1719129 (cited on page 48).

[220] A Falamas et al. 'Rapid and noninvasive diagnosis of oral and oropharyngeal cancer based on micro-Raman and FT-IR spectra of saliva'. In: *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 252 (2021), p. 119477 (cited on pages 48, 49).

[221] Arti Hole et al. 'Salivary Raman spectroscopy: standardization of sampling protocols and stratification of healthy and oral cancer subjects'. In: *Applied Spectroscopy* 75.5 (2021), pp. 581–588 (cited on page 48).

[222] Pachaiappan Rekha et al. 'Near-infrared Raman spectroscopic characterization of salivary metabolites in the discrimination of normal from oral premalignant and malignant conditions'. In: *Journal of Raman Spectroscopy* 47.7 (2016), pp. 763–772 (cited on page 49).

[223] S Jaychandran, PK Meenapriya, and S Ganesan. 'Raman spectroscopic analysis of blood, urine, saliva and tissue of oral potentially malignant disorders and malignancy-a diagnostic study'. In: *J. Oral Craniofac. Sci* 2.1 (2016) (cited on page 49).

[224] Duo Lin et al. 'Autofluorescence and white light imaging-guided endoscopic Raman and diffuse reflectance spectroscopy for in vivo nasopharyngeal cancer detection'. In: *Journal of Biophotonics* 11.4 (2018), e201700251. DOI: 10.1002/jbio.201700251 (cited on pages 49, 50).

[225] Sharmin Akter et al. 'Medical applications of reflectance spectroscopy in the diffusive and sub-diffusive regimes'. In: *Journal of Near Infrared Spectroscopy* 26.6 (2018), pp. 337–350. DOI: https://doi.org/10.1177/0967033518806637?urlappend=%3Futm_source%3Dresearchgate (cited on page 49).

[226] TSimone Innocente. 'Multimodal spectroscopy for enhanced AI-based diagnostics: leveraging DRS and Raman fusion'. In: *SPIE Photonics West*. SPIE. 2026 (cited on page 49).

[227] Neal B Gallagher. 'Savitzky-golay smoothing and differentiation filter'. In: *Eigenvector Research Incorporated* 2 (2020), p. 4 (cited on page 53).

[228] Anshuman Chandra et al. 'Unveiling the molecular secrets: a comprehensive review of Raman spectroscopy in biological research'. In: *ACS omega* 9.51 (2024), pp. 50049–50063. DOI: https://pubs.acs.org/doi/10.1021/acsomega.4c00591 (cited on page 54).

[229] Kosuke Dodo, Katsumasa Fujita, and Mikiko Sodeoka. 'Raman spectroscopy for chemical biology research'. In: *Journal of the American Chemical Society* 144.43 (2022), pp. 19651–19667. DOI: https://pubs.acs.org/doi/10.1021/jacs.2c05359 (cited on page 54).

[230] Neethu Emmanuel et al. *Fabricating a low-cost Raman spectrometer to introduce students to spectroscopy basics and applied instrument design*. 2021. DOI: https://pubs.acs.org/doi/10.1021/acs.jchemed.0c01028 (cited on page 54).

[231] CD Allemand. 'Design criteria for a Raman spectrometer'. In: *Applied Optics* 9.6 (1970), pp. 1304–1311. DOI: https://doi.org/10.1364/AO.9.001304 (cited on page 54).

[232] Patrick Martin Byrne. 'Development of a versatile Raman spectroscopy system'. PhD thesis. Dublin City University, 1998 (cited on page 54).

[233] Dustin W Shipp, Faris Sinjab, and Ioan Notingher. 'Raman spectroscopy: techniques and applications in the life sciences'. In: *Advances in Optics and Photonics* 9.2 (2017), pp. 315–428. DOI: https://doi.org/10.1364/AOP.9.000315 (cited on page 54).

[234] Edinburg Instruments, ed. *How to Choose your Lasers for Raman spectroscopy*. URL: https://www.edinst.com/resource/how-to-choose-your-lasers-for-raman-spectroscopy/#:~:text=The%20most%20commonly%20used%20laser,nm%20laser%20may%20be%20needed. (cited on page 55).

[235] Mike Hardy and Hin On Martin Chu. 'Laser wavelength selection in Raman spectroscopy'. In: *Analyst* 150.10 (2025), pp. 1986–2008. DOI: https://doi.org/10.1039/D5AN00324E (cited on page 55).

[236] Laura T Kerr, Hugh J Byrne, and Bryan M Hennelly. 'Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen'. In: *Analytical Methods* 7.12 (2015), pp. 5041–5052. DOI: https://doi.org/10.1039/C5AY00327J (cited on page 55).

[237] David Tuschel. 'Spectral resolution and dispersion in Raman spectroscopy'. In: (2020) (cited on page 55).

[238] Oscila Enabling Science, ed. *Spectral Resolution of Spectrometer*. URL: https://www.ossila.com/pages/spectrometer-resolution (cited on page 55).

[239] R Masters. 'Optical Resolution and Resolving Power: What It Is, How to Measure It, and What Limits It'. In: *Confocal microscopy and multiphoton excitation microscopy: the genesis of live cell imaging (SPIE Press, 2006)* (2006), pp. 49–54. DOI: https://doi.org/10.1117/3.660403.ch4 (cited on page 56).

[240] Rekha Gautam et al. 'Review of multidimensional data processing approaches for Raman and infrared spectroscopy'. In: *EPJ Techniques and Instrumentation* 2.1 (2015), p. 8. DOI: 10.1140/epjti/s40485-015-0018-6 (cited on page 60).

[241] Rekha Gautam et al. 'Feature selection and rapid characterization of bloodstains on different substrates'. In: *Applied spectroscopy* 74.10 (2020), pp. 1238–1251. DOI: https://doi.org/10.1177/0003702820937776 (cited on page 67).

[242] Sanchita Sil, Rekha Gautam, and Siva Umapathy. 'Applications of Raman and infrared microscopy to materials and biology'. In: *Molecular and Laser Spectroscopy: Advances and Applications*. Elsevier Inc., 2018, pp. 117–146 (cited on page 71).

[243] Tuan Phan-Xuan et al. 'Hydration-induced structural changes in the solid state of protein: A SAXS/WAXS study on lysozyme'. In: *Molecular pharmaceutics* 17.9 (2020), pp. 3246–3258. DOI: 10.1021/acs.molpharmaceut.0c00351 (cited on page 72).

[244] Noelia Campillo et al. 'Differential oxygenation in tumor microenvironment modulates macrophage and cancer cell crosstalk: novel experimental setting and proof of concept'. In: *Frontiers in oncology* 9 (2019), p. 43. DOI: https://doi.org/10.3389/fonc.2019.00043 (cited on page 72).

[245] Pavan Kumar and Asima Pradhan. 'Detection of Oral Lesions by the Diffuse Reflectance Spectroscopy on Human Oral Tissue and Saliva Samples'. In: *Journal of Tumor Medicine Prevention* 4 (2023). DOI: 10.19080/JTMP.2023.04.555626 (cited on page 72).

Paper I

diagnostics

MDPI

Review

Label-Free Optical Spectroscopy for Early Detection of Oral Cancer

Siddra Maryam [1,*], Marcelo Saito Nogueira [1], Rekha Gautam [1], Shree Krishnamoorthy [1], Sanathana Konugolu Venkata Sekar [1], Kiang Wei Kho [1], Huihui Lu [1], Richeal Ni Riordain [2,3], Linda Feeley [2,4], Patrick Sheahan [2,5], Ray Burke [1] and Stefan Andersson-Engels [1]

1 Tyndall National Institute, University College Cork, T12 R229 Cork, Ireland
2 ENTO Research Institute, University College Cork, T12 R229 Cork, Ireland
3 Cork University Dental School and Hospital, Wilton, T12 E8YV Cork, Ireland
4 Cork University Hospital, T12 DC4A Cork, Ireland
5 South Infirmary Victoria University Hospital, T12 X23H Cork, Ireland
* Correspondence: siddra.maryam@tyndll.ie

Abstract: Oral cancer is the 16th most common cancer worldwide. It commonly arises from painless white or red plaques within the oral cavity. Clinical outcome is highly related to the stage when diagnosed. However, early diagnosis is complex owing to the impracticality of biopsying every potentially premalignant intraoral lesion. Therefore, there is a need to develop a non-invasive cost-effective diagnostic technique to differentiate non-malignant and early-stage malignant lesions. Optical spectroscopy may provide an appropriate solution to facilitate early detection of these lesions. It has many advantages over traditional approaches including cost, speed, objectivity, sensitivity, painlessness, and ease-of use in clinical setting for real-time diagnosis. This review consists of a comprehensive overview of optical spectroscopy for oral cancer diagnosis, epidemiology, and recent improvements in this field for diagnostic purposes. It summarizes major developments in label-free optical spectroscopy, including Raman, fluorescence, and diffuse reflectance spectroscopy during recent years. Among the wide range of optical techniques available, we chose these three for this review because they have the ability to provide biochemical information and show great potential for real-time deep-tissue point-based *in vivo* analysis. This review also highlights the importance of saliva-based potential biomarkers for non-invasive early-stage diagnosis. It concludes with the discussion on the scope of development and future demands from a clinical point of view.

Keywords: oral cancer; Raman spectroscopy; diffuse reflectance spectroscopy; fluorescence spectroscopy; biomarkers; saliva analysis

Citation: Maryam, S.; Nogueira, M.S.; Gautam, R.; Krishnamoorthy, S.; Venkata Sekar, S.K.; Kho, K.W.; Lu, H.; Ni Riordain, R.; Feeley, L.; Sheahan, P.; et al. Label-Free Optical Spectroscopy for Early Detection of Oral Cancer. *Diagnostics* **2022**, *12*, 2896. https://doi.org/10.3390/diagnostics12122896

Academic Editors: Jyothi Tadakamadla and Sudheer Babu Balla

Received: 24 October 2022
Accepted: 18 November 2022
Published: 22 November 2022

1. Introduction

Cancer is a global health problem. According to an estimate of the International Agency for Research on Cancer, World Health Organization, more than 18 million new cancer cases and more than 9.6 million cancer deaths were reported worldwide in 2018 [1], with 19.3 million new cases and 10 million deaths in 2020 alone [2]. Oral cancer is reported to be the 16th most common cancer in the world with a five-year survival rate of only 40–60%, depending on the region of the world [2,3].

Over 90% of oral cancers are squamous cell carcinomas (SCCs), which arise from surface epithelial cells of the oral cavity [4]. Overall survival rates are strongly related to tumour stage at diagnosis [5]. The rates range from 80% for early stage cancers (I/II) to 30–50% in advanced stage cancers (III or IV) [6].

Early diagnosis of oral cancer is critical, both for optimizing survival and for quality of life after treatment. The aim of this review is to assess the current state and progress in oral cancer diagnostic techniques and discuss the importance of optical techniques in oral cancer diagnosis. This review summarizes the importance of body fluids in early detection of oral

cancer and the biomarkers that have been investigated in previous studies. In addition, major developments in label-free optical spectroscopy, including Raman spectroscopy (RS), fluorescence spectroscopy (FS), and diffuse reflectance spectroscopy (DRS), are also reviewed. This review concludes with a discussion on the scope of development and future demands from a clinical point of view.

2. Diagnosis of Oral Cancer

In the very early stages of oral cancer, the patient may present with a painless flat or discolored lesion on the oral mucosa. Patients, however, more frequently present at a later stage with a mass or non-healing ulcer intraorally, which may be painful and interfere with oral function. Enlargement of the cervical lymph nodes may indicate regional metastasis [7]. Currently, the gold standard procedure for oral cancer diagnosis is the histopathological examination of a biopsy specimen from the tumor site [8]. Biopsy can be an invasive and painful procedure and may require general anesthesia.

Early diagnosis of oral cancer is of clear benefit as the surgical excision of early stage lesions is associated with improved cancer control and survival outcomes, as well as much less morbidity and long-term sequelae for patients. However, one of the challenges with early diagnosis of oral cancer is that many lesions may present as a flat white (leukoplakic) or red (erythroplakic), or exophytic (verrucous) lesions [9]. These lesions may show variable degrees of dysplasia, which is associated with premalignant potential. Diagnosis of early invasive cancer within these lesions by pathological examination relies on the demonstration of invasion through the basement membrane; however, in early stage cancers, this invasion may be focal, thus biopsy of the lesion may miss the cancer if the biopsy is not taken from the part of the lesion corresponding to where the invasion is occurring. Furthermore, such oral premalignant lesions may be extensive, multifocal, or recurrent after initial removal [10]. Therefore, there is a pressing need for a reliable, painless, non-invasive, and cost-effective means of assessment of oral mucosal lesions. Such a technique may facilitate the identification of the most suspicious areas to target for biopsy, selection of oral premalignant lesions that should undergo complete excision, as well as define the extent of surgery for patients with confirmed oral cancers [11]. Table 1 lists a number of detection techniques that are already available and currently being used by clinicians and some that are still in the research phase.

Table 1. Techniques for oral cancer diagnosis in addition to clinical oral examination.

Diagnostic Techniques		*In Vivo*	Non-invasive	Real Time	Time Required *	Label-Free	Non-ionizing	Portability	Spatial Resolution	Detection Accuracy	Cost-Effectiveness **	References
Histopathology	Excisional biopsy	×	×	×	days	×	✓	×	µm	High	Low cost	[12,13]
	Incisional biopsy	×	×	×	days	×	✓	×	µm	High	Low cost	
	Brush biopsy	×	✓	×	days	×	✓	×	µm	High	Low cost	
	Punch biopsy	×	×	×	days	×	✓	×	µm	High	Low cost	
Vital staining	Toluidine blue	×	×	×	days	×	✓	×	µm	High	Low cost	[14]
	Iodine staining	×	×	×	days	×	✓	×	µm	High	Low cost	
	Methylene blue	×	×	×	days	×	✓	×	µm	High	Low cost	
	Lugol's iodine	×	×	×	days	×	✓	×	µm	High	Low cost	
	Acetowhite staining	×	×	×	days	×	✓	×	µm	High	Low cost	
	Double staining	×	×	×	days	×	✓	×	µm	High	Low cost	
Imaging techniques	OCT (Optical coherence tomography)	✓	✓	✓	10–20 min	✓	✓	×	µm	Medium	Expensive	[15–17]
	CT scan (computed tomography)	✓	✓	✓	10–20 min	×	×	×	Mm	Medium	Expensive	
	PET scan (positron emission tomography)	✓	✓	✓	45–60 min	×	×	×	Mm	Medium	Expensive	
	MRI (magnetic resonance imaging)	✓	✓	×	15–90 min	✓	✓	×	Mm	Low	Expensive	
Molecular analysis	Immunohistochemistry	×	×	×	days	×	✓	×	µm	Very high	Low cost	[18–20]
	In situ hybridization	×	×	×	days	×	✓	×	µm	Very high	Low cost	
	Flow cytometry	×	×	×	days	×	✓	×	µm	Very high	Low cost	
	Mass spectrometry	×	×	✓	s	✓	×	×	NA	Very high	Low cost	
	PCR (polymerase chain reaction)	×	✓	✓	min	✓	✓	✓	NA	Very high	Low cost	
Light-based detection	Tissue fluorescence (autofluorescence imaging, Velscope, Identafi)	✓	✓	✓	s	✓	✓	✓	Low	Low	Low cost	[12,15,21]
	Chemiluminiscence (ViziLite, Microlux/DL, MCE)	✓	✓	✓	s	×	✓	✓	Low	Low	Low cost	
	Optical spectroscopy	✓	✓	✓	s	✓	✓	✓	Low	Medium	Low cost	

* Time required to obtain the results after the sample collection. ** Cost/visit. NA, not available.

Optical spectroscopy, including Raman spectroscopy (RS), diffuse reflectance spectroscopy (DRS), autofluorescence spectroscopy, and optical coherence tomography (OCT), is able to provide non-invasive point-of-care analysis of structural and biochemical changes during cancer progression [22]. These morphological and biochemical changes in the epithelium serve as important biomarkers in oral cancer detection [23]. The above techniques have a potential to diagnose early stage cancer and dysplasia in real time with high sensitivity, without causing patient discomfort [22,24,25]. These techniques are also effective in analyzing bodily fluids, opening a doorway to oral cancer screening [26,27]. Post screening, non-invasive acquisition of data from multiple locations of an extensive, multifocal, and heterogeneous premalignant lesion may assist in performing an informed biopsy.

3. Optical Spectroscopy for Oral Cancer Screening *Ex Vivo*

When light interacts with a biological matter, a combination of different optical phenomena such as absorption, reflection, scattering, and fluorescence takes place. Provided that optical properties are known, spectroscopic analysis of these optical processes can provide information about the physical, chemical, and metabolic state of the tissue. Spectroscopy can be classified either by the energy of electromagnetic wave used or based on their interaction with the biological samples. Among the wide range of spectroscopic techniques, Raman spectroscopy, fluorescence spectroscopy, and diffuse reflectance spectroscopy have the ability to provide biochemical information in real-time, point-based diagnosis [28–30]. Raman, an inelastic scattering technique, probes the molecular vibration, providing information on conformational and compositional changes in the sample under investigation. In conventional Raman set-up, the monochromatic light interacts with the superficial layer of the sample. In contrast, DRS is implemented with a spatial separation between the illumination and detection fiber to enhance the collection of optical properties (absorption and scattering) of the sample in depth. For DRS, a specific optical window (650–950 nm) is preferred to avoid the excessive attenuation of light owing to dominating chromophores such as haemoglobin and water. Table 2 below represents some subjective views of the general characteristics of Raman spectroscopy, diffuse reflectance spectroscopy, and fluorescence spectroscopy for diagnostic support of oral cancers.

Table 2. General characteristics of RS, DRS, and FS in cancer diagnosis.

	Raman Spectroscopy	**Diffuse Reflectance Spectroscopy**	**Fluorescence Spectroscopy**
Basic Phenomenon	Scattering	Scattering and absorption	Photoluminescence event
Incident light	Ultraviolet, visible, NIR	Visible, NIR	Ultraviolet, visible
Detected light	Inelastically scattered Raman-shifted light	Diffusely scattered light	Emission from endogenous fluorophores
Spectral range	10–4000 cm^{-1}	200–3000 nm	360–700 nm
Typical tissue depth probed	<0.1 mm [31,32]	<1.5 mm [33]	0.1–10 mm [34]
Label-free/noninvasive/nondestructive	Yes	Yes	Yes
Real time data acquisition	Yes	Yes	Yes
Suitable for *ex vivo*/*in vivo* analysis	Yes	Yes	Yes
Photo-bleaching	No	No	Yes
Suitable for biological specimens	Yes	Yes	Yes
Suitable to incorporate with endoscopy	Yes	Yes	Yes
Suitable for water-based specimens, blood, and saliva	Yes	Only highly scattering specimens (selective wavelength regions)	Yes
Special sample preparation	No	No	No
Specificity (bandwidth)	High	Moderate	Moderate
Sensitivity	Low	High	High
Nanoparticles	Not required but can be used —SERS	Not required	Not required—can be used in SPR-enhanced fluorescence
Possible biomarkers	Lipids, protein, nucleic acid, circulating tumor cells	Water, lipid, collagen, deoxy and oxyhaemoglobin	Endogenous fluorophores—such as NADH, FAD, collagen, and porphyrins

Timely therapeutic intervention also requires technological developments that enable the detection of biochemical changes in bodily fluids and assist in biopsy. In this regard, biofluid spectroscopy is a potential method for non-invasive cancer screening and diagnostics [35]. Raman spectroscopy has the upper hand in the study of biofluids in situ as

water is a very weak Raman scatterer, as described in Table 2. There are many studies available on detecting cancer-associated biochemical changes using Raman spectroscopy in saliva [36,37], urine [38], cervical fluid [39], and blood plasma [40] with this technique [41].

The idea of conducting a simple saliva test in order to diagnose oral malignancies is extremely appealing. It offers convenient, inexpensive, and non-invasive sample collection as compared with blood and tissue biopsies. Saliva is a fluid with a very complex composition. Its major component is water—about 98%—with enzymes, electrolytes, minerals, proteins, nucleic acid, antibodies, or nasal secretions dissolved in it. It may also contain bacteria, viruses, epithelial cells, blood, or its derivatives from lesions and other biomarkers [42,43]. Its composition can represent the condition of the whole body and, therefore, it can be a very effective diagnostic medium for many diseases, particularly those in the oral cavity. In addition, molecular biomarkers present in saliva have the potential for patient stratification. In a recent study, the potential of fluorescence spectroscopy has been discussed to differentiate between OSCC, dysplasia, and control group using saliva as a specimen. This study reports significantly different fluorescence intensities associated with flavin adenine dinucleotide (FAD) and porphyrin among three groups [44]. With advancements in nanotechnology and medical science, many biomarkers associated with oral cancer have been reported, but still it is a challenge to confirm "true" biomarkers that can be identified non-invasively with high sensitivity and specificity. Previous studies show that saliva is a more effective diagnostic medium in the early stages of oral cancer than plasma owing to direct contact to the pathology site [43]. In contrast, serum could serve as a good diagnostic specimen for late stage oral cancer [45]. Another factor to consider is the reduced production of saliva in patients after treatment (surgery or radiotherapy), rendering it difficult to collect salivary specimens in follow-up patients. In a more recent study, saliva samples collected from 148 participants were subjected to Raman spectral analysis for patient stratification in healthy volunteers, tobacco habitués, and oral cancer patients. This study confirmed the identification of spectral differences among three groups [46]. In another study, Koster reported diagnostic performance accuracy of 91.7% in a cohort of head and neck cancer and benign control by unifying Raman spectra of plasma and saliva sample of each participant to assess metabolite composition [47]. Some important salivary biomarkers for oral cancer are shown in Table 3.

Table 3. Important salivary biomarkers for oral cancer.

Category	Biomarker	Healthy	Oral Cancer	Significance	Trend	Ref.
Proteins and amino acids	Phenylalanine	0.011 µmol/mL	0.105 µmol/mL	*	Increase	[48]
		4100 ng/mL	T1, T2 = 2500 ng/mL; T3, T4 = 1900 ng/mL	*	Decrease	[49]
	Tyrosine	0.112 µmol/mL	0.343 µmol/mL	*	Increase	[48]
	Tryptophan		3.81 ± 0.62 µM			[50]
	Leucin	0.015 µmol/mL	0.241 µmol/mL	*	Increase	[48]
		2300 ng/mL	T1, T2 = 600 ng/mL; T3, T4 = 500 ng/mL	*	Decrease	[49]
	Alanine	0.096 µmol/mL	0.178 µmol/mL	*	Increase	[48]
	Valine	0.038 µmol/mL	0.165 µmol/mL	NS	Increase	[48]
	Isoleucin	0.033 µmol/mL	0.236 µmol/mL	*	Increase	[48]
	Aspartic acid	0.035 µmol/mL	0.241 µmol/mL	*	Increase	[48]
	Serine	0.050 µmol/mL	0.187 µmol/mL	*	Increase	[48]
	Glycine	0.065 µmol/mL	0.288 µmol/mL	*	Increase	[48]
	Threonine	0.157 µmol/mL	0.435 µmol/mL	NS	Increase	[48]
	Arginine	0.047 µmol/mL	0.220 µmol/mL	*	Increase	[48]
	Isoleucine	0.033 µmol/mL	0.236 µmol/mL	*	Increase	[48]
	Methionine	0.012 µmol/mL	0.162 µmol/mL	*	Increase	[48]
	Albumin	0.17–0.36 g/L; mean 0.24 g/L	0.192–0.67; mean 0.36 g/L	*	Increase	[51]
		0.2 ± 0.1 mg/mL				[52]
		0.28 ± 0.19 g/dL	0.82 ± 0.41 g/dL	*	Increase	[53]
		0.8–192 mg/dL				[52]
	α-amylase	3257 ± 1682 U/mL				[52]
		65.2 mg/mL	68.07 mg/mL		Increase	[54]
		1080 ± 135.6 IU/L				[52]
	Interlukins (IL-8)	250 pg/mL	OSCC 720 pg/mL	*	Increase	[55,56]
	Interlukins (IL-6)	0 pg/mL	OC 86.5 pg/mL	*	Increase	[56]
		16 ± 3.91 # pg/mL	129 ± 66.59 # pg/mL	*	Increase	[57]

Table 3. *Cont.*

Category	Biomarker	Healthy	Oral Cancer	Significance	Trend	Ref.
	Osteopontin	35.1 ng/mL	39.23 ng/mL		Increase	[56]
	CRP (inflamation marker)	0.05–61 µg/L				[58]
	suPAR	5.22–28.1 ng/mL				[58]
	Survivin	2.44 ± 4.22 pg/mL	8.69 ± 10.15 pg/mL	*	Increase	[59]
	Kallikrien 5	~6 pg/mL	~12 pg/mL	*	Increase	[60]
	Cathepsin	9–18 * ng/mL				[61]
	Cathepsin V	~8 pg/mL	~14 pg/mL	*	Increase	[60]
	lactate dehydrogenase (LDH)	3.833 ± 1.1044 U/L	99.83 ± 49.33 U/L	*	Increase	[62]
		63.04 ± 47.4 mg/dL	1515.17 ± 765.14 md/dL		Increase	[63]
	(Endotheline-1) ET-1	0 to 9.629 fmol/m	0 to 7.554 fmol/mL	NS	Increase	[64]
		0.506–19.280 pg/mL; 4.5299 ± 3.7380 pg/mL	2.140–52.229 pg/mL; 13.51 ± 14.15 pg/mL	*	Increase	[65]
	Statherin	0.5–4.0 µg/mL; mean 0.96 µg/mL				[66]
		4.3–5.59 µM; mean 4.93 ± 0.61 µM	0–6.45 µM; mean 2.28 ± 2.86 µM	*	Decrease	[67]
	Carbohydrate antigen (CA 125)	137.12 ± 124.58 U/mL	498.10 # U/mL; 19.9–1312.32 U/mL	NS		[68]
		33.00 ± 24.37 mg/dL	888.15 ± 306.1 mg/dL	*	Increase	[63]
	Tissue polypeptide-specific antigen (TPS)	96.20 ± 71.60 U/mL	272.28 U/m #; 13.61–4706.17 U/mL	NS		[68]
	CD 44	1.09 ng/mL	7.85 ng/mL	*	Increase	[69]
Antioxidants	Glutathione	9.4 µmol/dL	8.2 µmol/dL		Decrease	[70]
	vitamin c	0.925 mg/dL	0.4787 mg/dL	*	Decrease	[71]
Carbohydrates	Fucose	0.38–17.0 mg/dL mean 2.94 mg/dL	Pre-cancer 0.112–18.46 mg/dL; mean 7.02 mg/dL	*	Increase	[72]
			OSCC 0.11–30.60 mg/dL; mean 11.66 mg/dL			
	Sialic acid	3.19 ±1.94 mg/dL	6.14 + 2.16 mg/dL	*	Increase	[73]
		3.18 mg/dL	11.66 mg/dL	*	Increase	[72]
		0.134–0.311; mean 0.189 mmol/L	0.140–0.336; mean 0.22 mmol/L	*	Increase	[73]
		21.65 ± 5.71 mg/dL	204.85 ± 60.38 mg/dL	*	Increase	[74]
		1.35 ± 1.53 mg/dL	5.30 ±1.45 mg/dL	*	Increase	[73]
Lipids	Linoleic acid	339.3 ± 267.9 ng/mL	1092.3 ± 1927.8 ng/mL		Increase	[75]
	15-HETE (Hydroxye-icosatetraenoic acids)	0.4 ± 0.8 ng/mL	5.4 ± 6.8 ng/mL	*	Increase	[75]
	Arachidonic acid	32.6 ± 26.6 ng/mL	606.9 ± 1695.7 ng/mL		Increase	[75]
	Lipo per oxidation products (MDA)	6.15–9.06 nmol/mL; mean 6.92 nmol/mL	5.56–7.78 nmol/mL; mean 6.58 nmol/mL	NS		[51]
Other biomarkers	Pyruvic acid	1.32 ± 0.10	3.49 ± 0.47	*	Increase	[76]
	Uric Acid	25.2–161.2 nmol/mL; mean 76.8	43.2–182.9; mean 93 nmol/mL	*	Increase	[51]
	Urea	4.35–8.78 mmol/L; mean 6.76 mmol/L	4.62–11.29; mean 8.66 mmol/L	*	Increase	[51]
	Total protein	0.78–1.53; mean 1.11 g/L	0.47–1.59 g/L; mean 0.84 g/L	*	Decrease	[51]
		0.47 ± 0.19 mg/mL				[52]
		0.9 ± 0.2 mg/mL				[52]
		43–710.0 mg/dL				[52]
		1.07 ± 0.59 mg/mL	1.01 ± 0.43 mg/mL	NS		[77]
		2.67 ± 0.54 mg/mL				[52]
	CYFRA-21-1	3.06 ng/mL	17.46 ± 1.46 ng/mL	*	Increase	[78]
	Basic fibroblast growth factor (bFGF)	3.17 ± 0.43 pg/mL	8.80 ± 1.26 pg/mL	*	Increase	[79]
		0.3 ± 0.3 pg/mL	OLP patients 5.9 ± 2.9 pg/mL	NS	increase	[80]

* Significant difference, NS, non-significant difference, # median.

4. Optical Spectroscopy for *In Vivo* Diagnosis

Cancer progression is a complex process in which molecular and cellular changes accumulate over time and lead to transitions of tissue from normal to malignant. The morphological and biochemical changes in the cancerous tissue such as epithelium thickening, enlargement of nucleus, and changes in the extracellular matrix (ECM) architecture act as important biomarkers in oral cancer detection [23]. Over the past few decades, optical spectroscopic techniques have shown great potential in tissue differentiation including cancer identification via probing these chemical changes. Further, the advent of new optical and data analysis techniques has provided the platform to integrate multiple spectroscopic modalities, which enables the collection of complementary information to improve the accuracy of the cancer differentiation.

Raman spectroscopy can inform the chemical signature of the tissue and provides structural and chemical composition of the targeted analytes. It relies upon inelastic scattering of light typically within the visible and NIR region. Changes in the concentration of biomolecules or any mutation in the tissue during cancer progression provide notable changes in Raman signals [81–83]. A prominent discriminating feature among healthy tissue and different stages of malignancy is a change in the concentration of proteins, lipids, and nucleic acid in the epithelium and decreased collagen with inflammation in connective tissue [84]. Recently, a software tool has been reported for the bimolecular composition analysis through Raman spectra [85,86]. Previous Raman spectra-based studies differentiated *in vivo* tissues and in vitro tissue samples using different classification models such as principal component analysis (PCA) [87], support vector machine (SVM) [88], quadratic discriminate analysis (QDA), and linear discriminant analysis (LDA) [28].

Fluorescence spectroscopy can detect metabolic changes in the tissue associated with enhanced metabolic activity and redox imbalance [89]. Such alterations cause an increase in the production of nicotinamide adenine dinucleotide phosphate (NADPH) and other metabolites [90]. This spectroscopy involves using a UV/visible light beam to excite the molecules of some particular tissue compounds—fluorophores. Following the excitation, the molecule quickly returns to a low vibrational energy level through a non-radiative vibrational relaxation. From this level, the molecule can then return to its ground state by emitting the excess energy in the form of a visible photon-fluorescence. When illuminated with an appropriate wavelength, biological tissue fluoresces because of the presence of many endogenous fluorophores including NADPH, NADH, collagen, flavins, elastin, and porphyrins [90,91]. The concentration of these fluorophores changes as the cancer progresses and can thus act as a biomarker for malignancy. Some of the fluorophores have overlapping excitation and emission spectra; however, they are more or less distinguishable by appropriate selection of excitation and emission wavelengths, which facilitates the quantification of various tissue fluorophores. The spectroscopic method is thus necessary to extract concentrations of particular fluorophores and provide a diagnostic value. If the tissue autofluorescence does not provide sufficient diagnostic accuracy, one can employ a contrast agent in the form of an exogenous fluorophore.

Diffuse reflectance spectroscopy (DRS) is based on the light that is backscattered multiple times inside the specimen and contains information about its optical properties. It can be used to extract both concentrations of absorbing tissue chromophores and morphological changes occurring in the sample. Similar to the fluorophores, chromophores can be utilized to indicate malignancy. Morphological alterations such as the uncontrolled growth of neoplastic cells or enlarged nuclei will influence the scattering properties and DRS can be employed to extract such diagnostic information efficiently. Scattering in tissue is quite complex and is linked to local changes in the refractive index. The overall change in the refractive index, shape, size, density of cellular structures, and absorption determines the spectral features of back-scattered back reflected light. DRS can operate in the UV, visible, and NIR regions. By controlling the probe configuration, one can obtain the signal from the superficial epithelial layer or from the deeper tissue layers. In a recent study, Wang et al. demonstrated the potential of a hand held DRS probe through in-depth characterization for *in vivo* diagnosis of oral diseases in a real-time clinical environment [92].

Furthermore, *in vivo* optical spectroscopy for disease diagnosis would allow clinicians to identify cancerous lesions based on spectral markers and significantly reduce the burden of biopsies. In this case, incident light is delivered to the tissue through a fiber optics probe placed in an endoscope in a localized manner. There has been significant development in this field in the last few decades. Figure 1, created with Edraw Max software, represents some typical features of the above-mentioned spectroscopic techniques.

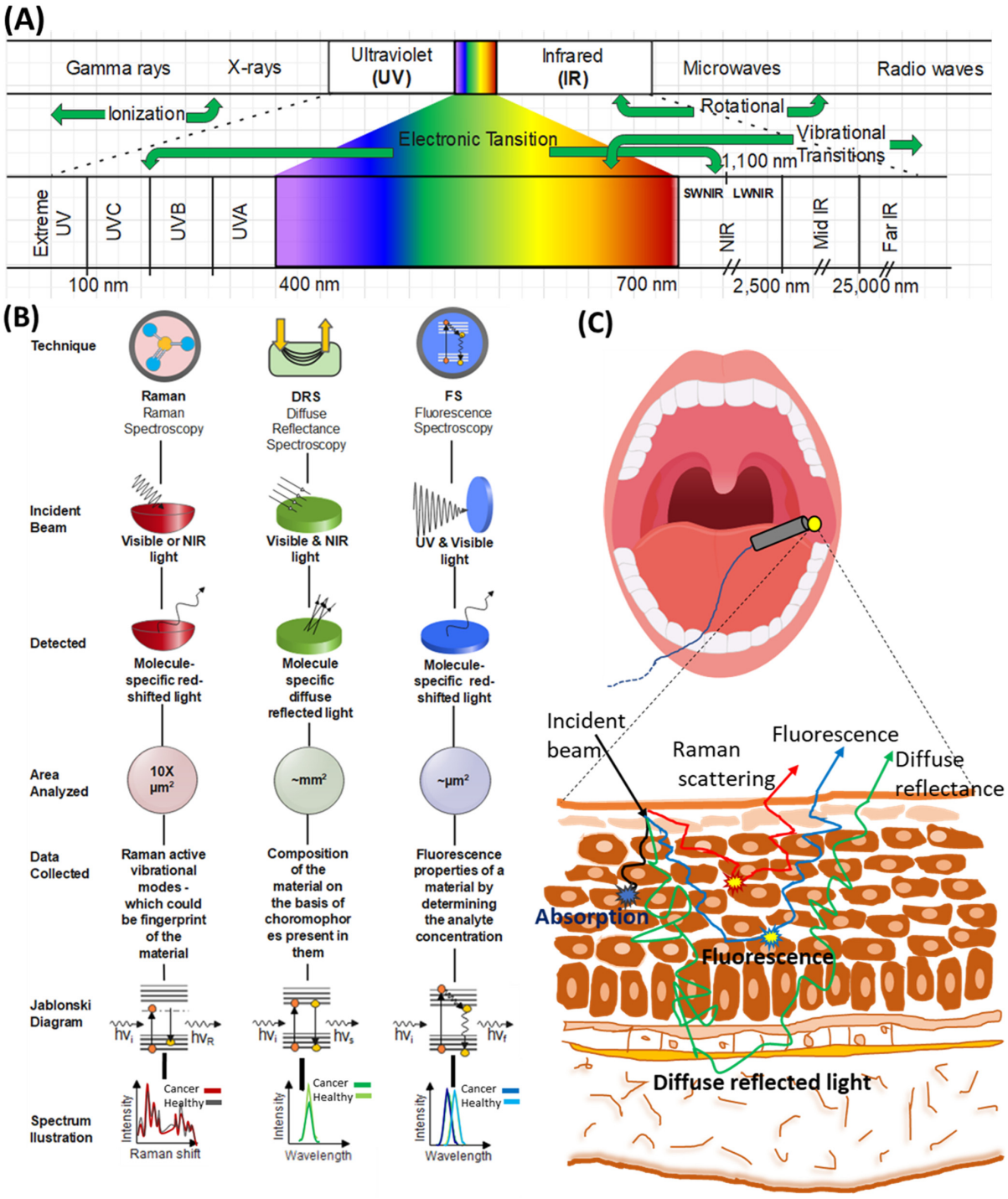


Figure 1. (**A**) Electromagnetic spectrum and light tissue interaction. (**B**) Key features of Raman spectroscopy (RS), diffuse reflectance spectroscopy (DRS), and fluorescence spectroscopy (FS). (**C**) Light tissue interaction and comparative depth of RS, DRS, and FS in oral mucosa.

5. Multimodal Spectroscopy Approach

Multimodal spectroscopy (MMS) refers to the production of signal simultaneously from multiple spectroscopy techniques. The goal of this approach is to further improve the early detection and localization of malignancies by measuring several biomarkers. In this case, the progression of various cellular events during cancer can be examined all together and in real time. Such an approach could be a promising tool to provide complementary, depth-sensitive information.

A multimodal instrument was designed by Greening et al. to monitor the progression of dysplasia in the oral cavity. This system incorporated high-resolution fluorescence imaging and sub-DRS (sub-diffuse reflectance spectroscopy). This system used two light sources—a halogen lamp for the sub-DRS modality and a 455 nm LED for fluorescence imaging. Two sets of liquid phantoms were designed to produce and validate the look-up table (LUT)-based inverse model. The sampling depth was also calculated for both probes. Subsequently, the system was validated in the clinical environment and *in vivo* sub-DRS data were collected from the inner lip of 13 volunteers. This hybrid system was reported to be capable of gathering information on functional and structural properties of tissue [93].

In a detailed real-time *in vivo* study by Lin et al., autofluorescence, DRS, and Raman spectroscopy were used in combination with white light imaging guided endoscopy for the detection of nasopharyngeal cancer. Sixty volunteers were involved in this study, 30 of whom were suffering from nasopharyngeal cancer confirmed by a biopsy immediately after the endoscopy. It was possible to distinguish between normal and cancer tissue with a very high sensitivity of approximately 98% and specificity of approximately 95% using a PCA-LDA diagnostic algorithm. It was reported that, as compared with white light imaging (WLI), autofluorescence imaging was much better at providing information on the metabolic state and biochemical composition of tissues. When normal tissue develops a malignancy, morphological changes take place in tissue such as thickening of epithelium, the extracellular matrix is decreased, cellular metabolic fluorophores such as flavins and nicotinamide adenine dinucleotide are reduced, and the concentration of hemoglobin is increased. As a consequence of the absorption of blue light by haemoglobin, a reduction in the amount of green-emitting metabolic fluorophores, and re-absorption of fluorescence by the thickened epithelium, malignant tissues tend to exhibit an overall decrease in the green autofluorescence. So, when normal tissue in the oral cavity is illuminated by a blue light, healthy tissue appears green, but as the cancer progresses, there is a gradual decrease in green fluorescence, but a proportionally less decrease in red fluorescence, hence cancerous tissue appears dark [94]. Lin et al. reported that autofluorescence imaging (AFI) helps in localizing the lesion more accurately than WLI. It also sets a ground for Raman spectroscopy and DRS to provide better guidance. They developed an integrated endoscopy system with four modalities including WLI, AFI, DRS, and Raman spectroscopy to achieve a high specificity by studying the biomolecular changes in malignant tissues. The Raman spectroscopy results showed a lower concentration of phospholipids in malignant tissue as compared with the normal surrounding tissue. However, an increased concentration of protein in α-helix, phenylalanine, and nucleic acid was observed in cancer tissue. The reflectance data were observed in the range 400–700 nm. Throughout this spectral range, the intensity of the reflected signal was higher for the normal tissue as compared with the malignant tumors, with this change becoming more prominent after a 600 nm wavelength. Detected valleys at 540 and 580 nm also represent less oxygenated hemoglobin in cancer tissues [81]. In another study by Jermyn et al., which combines DRS, Raman, and intrinsic fluorescence spectroscopy for detection of colon, lung, brain, and skin cancers, sensitivity and specificity were reported to be >99% and 93%, respectively. It was also reported that the detection capability of that system was not dependent on the cancer type, but that the system could classify the tissues on the basis of spectral features for all malignancies studied [26].

In vivo optical spectroscopy can also facilitate clinicians to differentiate among benign lesions and severe dysplasia and provide guidance for tumour classification. In one study on Raman spectroscopy performed on paraffin preserved tissue samples, Ibrahim et al. reported accuracy rates of 80–90% in the distinction of mild, moderate, and severe dysplasia in different tissue types of the same person and 60% accuracy in an interpatient study based on spectral markers. They also reported 70–80% accuracy in differentiating SCC from dysplasia and benign lesion in different tissues [95]. Furthermore, Jaychandran et al. managed to distinguish oral malignancy and premalignancy from normal control group with an accuracy of 97.4% for tissue and 93.1%, 90.5%, and 78% for saliva, urine, and

blood samples, respectively [96]. In a more recent study, Li et. al. tried to discriminate OSCC, dysplasia, and healthy mucosa by analyzing biochemical variations at different stages with near infrared Raman spectroscopy in order to evaluate the performance of different classification models such as PCA-LDA and SVM. They reported relatively higher classification accuracy for the SVM model as compared with PCA-LDA. It was also observed that oral dysplasia and OSCC have a higher content of DNA and protein than normal mucosa. However, there was no significant change in the Raman signal of various dysplastic grades [88].

In an in vitro study, Nour et al. used gold nanoparticles for detecting SCCs of the tongue through diffuse reflectance analysis and laser-induced fluorescence with 91% accuracy [97].

Tobacco usage increases the risk of damage to mucosa and the genetic changes ultimately cause abnormal cell growth. Identifying these changes may help to diagnose the lesion in the early stages. Shaiju et al. reported the importance of fluorescence spectroscopy in combination with multivariate analysis to analyze pathological changes due to tobacco abuse in the early stages. They reported that total porphyrin and hemoglobin levels in smokers are comparable to those with leukoplakia [98]. In a clinical study on thirty patients with erythroplakia and leukoplakia, it was reported that induced fluorescence in tumor could improve diagnostic accuracy. The aim of the study was to distinguish between normal and potentially malignant oral mucosa by fluorescence spectroscopy. The study reports lower fluorescence signal from the tumor center as compared with tumor borders. Similar behaviour was observed in erthroplakic lesions that have low fluorescence in comparison with normal tissue. A higher hemoglobin content could be the reason for the reduced fluorescent signal in both of these scenarios, as hemoglobin is an endogenous absorber of light. This study also points out differences in spectral features of normal patient mucosa and healthy volunteer mucosa [99,100]. Overall, the published literature highlights the potential of the spectroscopic techniques in differentiating benign premalignant lesions from invasive malignant lesions with adequate accuracy, which can be improved further by integrating multiple modalities into a system.

The importance of multimodal instruments can never be denied in medical diagnostics and there is still room for further developments. There are other published studies involving multimodal optical techniques to detect atherosclerotic plaque [101], amyloid plaques [102], breast cancer [103], and non-melanoma skin cancer in real time [104]. Now, it is time to incorporate other techniques and imaging modalities with spectroscopy for the diagnosis of oral cancer in the early stages and in real time.

6. Discussion

Ranking as the 16th most frequently diagnosed cancer in the world [2,3], oral cancer still has a very poor prognosis [6,105] and a high rate of recurrence [106,107]. Cancer stage affects the treatment outcome. In the past few decades, many efforts have been made to develop optical spectroscopy as an assisting tool for targeted and informed biopsy as well as a guiding tool to identify surgical margins more precisely. This article provides an overview of the potential of optical spectroscopic techniques in detecting early stage oral cancer. All of the spectroscopic techniques included are green in the sense they do not employ any exogenous substances or contrast agents [108]. Raman, diffuse reflectance, and fluorescence spectroscopy are new emerging spectroscopic techniques and hold promise to facilitate timely treatment to cure a higher fraction of the oral cancers by detecting them in the earlier stages. DRS achieves better results in tissue classification roughly between 400 nm and 700 nm. Fluorescence spectroscopy allows signal detection from a very low amount of fluorophore and can achieve a sensitivity of more than 80 % in staging cancers. Because of limited chemical specificity, it is, however, sometimes hard to discriminate overlapping signals from different fluorophores, compromising the diagnostic accuracy. Moreover, not all of the biological tissue structures are fluorescent when excited using visible or NIR wavelength. On the other hand, all biomolecules have their own specific

Raman signatures, which can be probed by a single excitation wavelength. It complements mid-IR spectroscopy by detecting vibrational modes that are mid-IR inactive. It is not heavily affected by water molecules and is suitable for aqueous solutions and biological specimens. In fact, in a study on 14 patients undergoing tongue resection, Barosso et al. discriminated between healthy and cancerous tissue in the oral cavity on the basis of water content with an accuracy of 99% and specificity of 92% [109]. Raman spectroscopy along with NIRS and mid-IR spectroscopy comprise the "three sisters" of vibrational spectroscopy [108]. In NIR spectroscopy, excitation to higher vibrational states within the ground electronic state of a molecule leads to overtone and combination vibrations [110]. While Raman spectroscopy performs very well for symmetric bonds, NIRS only works for asymmetric bonds. Fluorescent spectroscopy, DRS, and Raman spectroscopy are valuable and important technologies with complementary pros and cons. Together, they can give rise to a powerful tool for effective diagnosis. Robust machine-learning-based models can be explored to fuse the informative features from the multimodal system, improving the classification and/or prediction accuracies.

7. Limitations and Future Outlook

Despite all of the advances and research, optical spectroscopies are still far from being conventionally used in clinical diagnostics for oral cancer because of instrumental limitations and lack of diagnostic validation. Each spectroscopy has its own drawbacks. Raman spectroscopy experiences strong background fluorescence and inherently weak signal. Excitation light could be in the UV, visible, and NIR regions and hence there is the need for system design, optimization, and validation for a change in Raman excitation wavelength. The UV and visible range is used for resonance enhanced Raman spectroscopy of proteins and porphyrin like molecules where enhanced signals dominate over the fluorescence background. However, for *in vivo* Raman, the NIR range is preferred to avoid excessive fluorescence background and to improve the depth of penetration. In SERS, although the signal is strong, there remain clarifications regarding how to best incorporate the metallic nanoparticles in terms of size and overall geometry to secure reproducible signals and limit toxicity. DRS signals are not linearly dependent on chromophore concentrations, making direct correlation between spectral features and diagnostic interpretation more complex, and the technique typically requires big datasets to ensure diagnostic accuracy. The biggest drawback of fluorescence spectroscopy is the risk of misdiagnosis owing to the reported somewhat lower specificity. Therefore, there is still the need for more research to incorporate different optical spectroscopies into clinical practice.

During the last few decades, optical spectroscopy has experienced substantial advancements, which can be categorized as instrumentation development and analysis approach maturation. The majority of these developments have pushed the boundary of possibilities, taking this family of techniques to a new level and redefining it. However, considering the current status of advancement, there is still the need for improvement in instrumentation, analysis approach, and standardization.

7.1. Instrumentation

Recently, instrument development is being driven in four directions; that is, miniaturization and clinical adoption, instrument sensitivity, multimodality, and further technique sophistication. There is still room for development in all of these directions, especially considering the final user will be a clinician. Miniaturized, hand-held, and portable spectrometers enable users to bring the laboratory to the sample in point-of-care diagnostics. A simplified and concise workflow to familiarize the clinician and combination of portable handheld system with sophisticated machine learning models that can provide real-time output in a language understandable to clinicians is required. Adequate training protocols need to be established learning from on-going spectroscopic clinical trials across the world. Various optical societies are working on standardizing these protocols and making them easier for the clinicians to understand. Apart from miniaturization, disposable probes, cost

efficiency, and an easy approach are also necessary. In the current scenario, each modality requires different light sources and spectrometer/detectors, which makes it expensive. However, system miniaturization and integration can still make the technology attractive from both a design/footprint and a cost perspective. Additionally, the new developments in the instrumentation area provide a hope to make it more cost effective. Increased instrument sensitivity and capability can enable new applications. The combination of different techniques may also open new dimensions in the medical science. Apart from combining multiple complementary modalities, one can also develop techniques that are more sophisticated and provide higher diagnostic accuracy. For example, standard imaging techniques lack sufficient spatial resolution and sensitivity. A possible solution to these limitations could lie in non-linear microscopy or hyperspectral imaging. Another possibility could be to employ time-domain Raman spectroscopy. Multiscale and multimodal imaging techniques that combine biochemical macroscopic imaging to morphological subcellular imaging techniques could also lead to new pathways in clinical research. Incorporating fluorescence with Raman into a single process might provide single molecule sensitivity with the specificity of Raman spectroscopy and open a new era in medical diagnostics. Tagging or incorporation of nanoparticles with these imaging modalities can also play an important role in improving the sensitivity of these techniques and the signal-to-noise ratio. In addition to this, the use of multiple wavelengths for excitation can provide more hidden information. Thus, future studies should focus on combination of optical spectroscopic techniques and multiple wavelengths to overcome the shortcomings of a single technique. Another possible future improvement could be automation of these techniques, which can make possible the analysis of multiple specimens simultaneously.

7.2. Analysis Approach/Data Analysis

Experiments have shown that spectral techniques are a valuable source of large amounts of clinical data. Years of research and technical development have provided a large number of overlapped datasets that are difficult to interpret. These data could provide useful information in medical diagnostics and potentially lead to a future paradigm shift. Various groups are making efforts to translate these spectral data into clinically relevant information. Currently, there are three general ways to translate the available spectral data. The first approach that does not require any prior data processing and knowledge of light tissue interaction is the direct analysis of raw spectral data using supervised and unsupervised machine learning methods, such as neural network, component analysis, and clustering [111]. A second method, which requires expert knowledge, involves the translation of spectral data into optical parameters such as absorption or scattering coefficient at different wavelengths and then correlating those with classically diagnosed tissue type based on statistical analyses that are more conventional [112]. A third method, which is even more complicated, includes the transformation of these physical properties into relevant biological parameters associated with molecular vibrations, chromophores, and fluorophores in tissue or other biomarkers. This approach requires detailed knowledge of biological substances within the tissue that absorb or scatter light [112,113]. Any of these three approaches can be adopted and are acceptable for clinical use if it helps the clinician to interpret the biological information hidden in the data and recognize a tissue abnormality at an early stage. The analysis can be furthered by incorporating new techniques such as explainable machine learning and artificial intelligence algorithms for better data classification efficiency and a reduced error rate. These more novel techniques can help to design an intelligent user-friendly real-time monitoring system or automatic recognition of malignancy and tumour subtypes.

7.3. Requirement of Standardization

Optical spectroscopy is still a technique under development and reliable standardization of the techniques is required along with advancement in hardware technology and analysis tools. Currently, there are different diagnostic methods, instruments (spectro-

scopes, probes, and so on) and analytical techniques, which makes it difficult to compare the results from different studies and, thereby, to fully assess the diagnostic capabilities and importance of the technology. To obtain accurate results, a standardized calibration procedure is necessary. There are few attempts to standardize the diffuse optical methods using phantoms. However, similar standardization methods need to be developed and adopted for label-free optical spectroscopy. In addition, awareness of the use of standards is required to accelerate the development of novel techniques and widespread adoption in regular clinical practice [113,114].

8. Summary and Conclusions

Biomedical optical spectroscopy provides a deep intuitive understanding about the light–tissue interaction and provides opportunities for real-time diagnostic information. The modalities including Raman, diffuse reflectance, and fluorescence spectroscopy are able to provide chemical and structural properties of analyzed biological specimens non-destructively. These optical spectroscopic techniques can facilitate the identification of lesions, which do not require biopsy owing to low or no risk of malignant progression and can help in targeted biopsy of the biologically most advanced area from within leukoplakic, erythroplakic, or verrucous lesions, especially lesions that are extensive and/or multifocal. Optical spectroscopy offers a potential mean of non-invasive diagnostics. In situ optical spectroscopic detection through endoscope is also capable of identifying malignant and premalignant lesions and has the potential to classify different tissues in real time. It can enhance a surgeon's vision by providing efficient diagnostic information and maximizing the chances of successful tumor removal with minimal harm to the surrounding healthy tissue. Body fluids, especially saliva, can serve as a very useful specimen for the early detection of oral cancer. Optical spectroscopy of blood and its components also has proven potential in the diagnosis of gastrointestinal and oral cancer. Multimodal approaches and incorporating optical modalities with other biological techniques can yield clear benefits by overcoming the limitations of individual techniques. Different mathematical models to analyze multispectral data could also serve an important role in this regard. Finally, standardization is necessary for translating these optical spectroscopic approaches and modalities into clinical diagnostic practice.

Funding: The authors are grateful for the financial support of Science Foundation Ireland through contract SFI/15/RP/2828.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge Konstantin Grygoryev and Jacqueline Gunther for externally reviewing the manuscript and providing helpful comments.

Conflicts of Interest: The authors declare no competing interest.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef]
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
3. Miranda-Filho, A.; Bray, F. Global patterns and trends in cancers of the lip, tongue and mouth. *Oral Oncol.* **2020**, *102*, 104551. [CrossRef]
4. Markopoulos, A.K. Current aspects on oral squamous cell carcinoma. *Open Dent. J.* **2012**, *6*, 126. [CrossRef]
5. Chen, X.-J.; Zhang, X.-Q.; Liu, Q.; Zhang, J.; Zhou, G. Nanotechnology: A promising method for oral cancer detection and diagnosis. *J. Nanobiotechnol.* **2018**, *16*, 52. [CrossRef]
6. Koch, F.P.; Kunkel, M.; Biesterfeld, S.; Wagner, W. Diagnostic efficiency of differentiating small cancerous and precancerous lesions using mucosal brush smears of the oral cavity—A prospective and blinded study. *Clin. Oral Investig.* **2011**, *15*, 763–769. [CrossRef]

7. Weinberg, M.A.; Estefan, D.J. Assessing oral malignancies. *Am. Fam. Physician* **2002**, *65*, 1379.
8. Oliver, R.; Sloan, P.; Pemberton, M. Verifiable CPD paper: Oral biopsies: Methods and applications. *Br. Dent. J.* **2004**, *196*, 329. [CrossRef]
9. Rethman, M.P.; Carpenter, W.; Cohen, E.E.; Epstein, J.; Evans, C.A.; Flaitz, C.M.; Graham, F.J.; Hujoel, P.P.; Kalmar, J.R.; Koch, W.M. Evidence-based clinical recommendations regarding screening for oral squamous cell carcinomas. *J. Am. Dent. Assoc.* **2010**, *141*, 509–520. [CrossRef]
10. Mian, S.A.; Yorucu, C.; Ullah, M.S.; Rehman, I.U.; Colley, H.E. Raman spectroscopy can discriminate between normal, dysplastic and cancerous oral mucosa: A tissue-engineering approach. *J. Tissue Eng. Regen. Med.* **2017**, *11*, 3253–3262. [CrossRef]
11. Bouaoud, J.; Bossi, P.; Elkabets, M.; Schmitz, S.; van Kempen, L.C.; Martinez, P.; Jagadeeshan, S.; Breuskin, I.; Puppels, G.J.; Hoffmann, C.; et al. Unmet needs and perspectives in oral cancer prevention. *Cancers* **2022**, *14*, 1815. [CrossRef]
12. Singh, S.; Ibrahim, O.; Byrne, H.J.; Mikkonen, J.W.; Koistinen, A.P.; Kullaa, A.M.; Lyng, F.M. Recent advances in optical diagnosis of oral cancers: Review and future perspectives. *Head Neck* **2016**, *38*, E2403–E2411. [CrossRef] [PubMed]
13. Samanthi. Difference Between Histopathology and Cytology. Available online: https://www.differencebetween.com/difference-between-histopathology-and-cytology/ (accessed on 21 April 2022).
14. Nitya, K.; Amberkar, V.S.; Nadar, B.G. Vital Staining-Pivotal Role in the Field of Pathology. *Ann. Cytol. Pathol.* **2020**, *5*, 058–063. [CrossRef]
15. Wilson, B.C.; Jermyn, M.; Leblond, F. Challenges and opportunities in clinical translation of biomedical optical spectroscopy and imaging. *J. Biomed. Opt.* **2018**, *23*, 030901. [CrossRef] [PubMed]
16. Lecchi, M.; Fossati, P.; Elisei, F.; Orecchia, R.; Lucignani, G. Current concepts on imaging in radiotherapy. *Eur. J. Nucl. Med. Mol. Imaging* **2008**, *35*, 821–837. [CrossRef]
17. Muzic, R.F., Jr.; DiFilippo, F.P. PET/MRI—Technical review. *Semin. Roentgenol.* **2014**, *49*, 242–254. [CrossRef]
18. Baron, J.A. Screening for cancer with molecular markers: Progress comes with potential problems. *Nat. Rev. Cancer* **2012**, *12*, 368–371. [CrossRef]
19. Furrer, D.; Sanschagrin, F.; Jacob, S.; Diorio, C. Advantages and disadvantages of technologies for HER2 testing in breast cancer specimens. *Am. J. Clin. Pathol.* **2015**, *144*, 686–703. [CrossRef] [PubMed]
20. Isaacs, C. Pros and Cons of Genetic Testing and Cancer. Available online: https://www.cancer.net/blog/2016-09/pros-and-cons-genetic-testing-and-cancer (accessed on 21 April 2022).
21. Lee, J.; Kim, B.; Park, B.; Won, Y.; Kim, S.-Y.; Lee, S. Real-time cancer diagnosis of breast cancer using fluorescence lifetime endoscopy based on the pH. *Sci. Rep.* **2021**, *11*, 16864. [CrossRef]
22. Jermyn, M.; Desroches, J.; Aubertin, K.; St-Arnaud, K.; Madore, W.-J.; De Montigny, E.; Guiot, M.-C.; Trudel, D.; Wilson, B.C.; Petrecca, K.; et al. A review of Raman spectroscopy advances with an emphasis on clinical translation challenges in oncology. *Phys. Med. Biol.* **2016**, *61*, R370. [CrossRef]
23. Xu, Z.; Huang, W.; Lin, D.; Wu, S.; Chen, M.; Ge, X.; Lin, X.; Sun, L. Discrimination of nasopharyngeal carcinoma from noncancerous ex vivo tissue using reflectance spectroscopy. *J. Spectrosc.* **2016**, *2016*, 1596845. [CrossRef]
24. Marcu, L. Fluorescence lifetime techniques in medical applications. *Ann. Biomed. Eng.* **2012**, *40*, 304–331. [CrossRef]
25. Kelly, K.; Johnson-Obaseki, S.; Lumingu, J.; Corsten, M. Oncologic, functional and surgical outcomes of primary Transoral Robotic Surgery for early squamous cell cancer of the oropharynx: A systematic review. *Oral Oncol.* **2014**, *50*, 696–703. [CrossRef] [PubMed]
26. Jermyn, M.; Mercier, J.; Aubertin, K.; Desroches, J.; Urmey, K.; Karamchandiani, J.; Marple, E.; Guiot, M.-C.; Leblond, F.; Petrecca, K. Highly accurate detection of cancer in situ with intraoperative, label-free, multimodal optical spectroscopy. *Cancer Res.* **2017**, *77*, 3942–3950. [CrossRef] [PubMed]
27. Lin, D.; Pan, J.; Huang, H.; Chen, G.; Qiu, S.; Shi, H.; Chen, W.; Yu, Y.; Feng, S.; Chen, R. Label-free blood plasma test based on surface-enhanced Raman scattering for tumor stages detection in nasopharyngeal cancer. *Sci. Rep.* **2014**, *4*, 4751. [CrossRef]
28. Jeng, M.-J.; Sharma, M.; Sharma, L.; Chao, T.-Y.; Huang, S.-F.; Chang, L.-B.; Wu, S.-L.; Chow, L. Raman spectroscopy analysis for optical diagnosis of oral cancer detection. *J. Clin. Med.* **2019**, *8*, 1313. [CrossRef]
29. Gorpas, D.; Davari, P.; Bec, J.; Fung, M.A.; Marcu, L.; Farwell, D.G.; Fazel, N. Time-resolved fluorescence spectroscopy for the diagnosis of oral lichen planus. *Clin. Exp. Dermatol.* **2018**, *43*, 546–552. [CrossRef]
30. De Koning, S.G.B.; Baltussen, E.J.; Karakullukcu, M.B.; Dashtbozorg, B.; Smit, L.A.; Dirven, R.; Hendriks, B.H.; Sterenborg, H.J.; Ruers, T.J. Toward complete oral cavity cancer resection using a handheld diffuse reflectance spectroscopy probe. *J. Biomed. Opt.* **2018**, *23*, 121611.
31. Eliasson, C.; Matousek, P. A Noninvasive Method for Deep Raman Spectroscopy of Living Tissue and Powders. Posted on August 2007. Available online: https://www.americanlaboratory.com/914-Application-Notes/1389-A-Noninvasive-Method-for-Deep-Raman-Spectroscopy-of-Living-Tissue-and-Powders/ (accessed on 21 October 2022).
32. Matousek, P.; Stone, N. Recent advances in the development of Raman spectroscopy for deep non-invasive medical diagnosis. *J. Biophotonics* **2013**, *6*, 7–19. [CrossRef]
33. Greening, G.; Mundo, A.; Rajaram, N.; Muldoon, T.J. Sampling depth of a diffuse reflectance spectroscopy probe for in-vivo physiological quantification of murine subcutaneous tumor allografts. *J. Biomed. Opt.* **2018**, *23*, 085006. [CrossRef]

34. Swartling, J.; Svensson, J.; Bengtsson, D.; Terike, K.; Andersson-Engels, S. Fluorescence spectra provide information on the depth of fluorescent lesions in tissue. *Appl. Opt.* **2005**, *44*, 1934–1941. [CrossRef]
35. Leal, L.; Nogueira, M.; Canevari, R.; Carvalho, L. Vibration spectroscopy and body biofluids: Literature review for clinical applications. *Photodiagnosis Photodyn. Ther.* **2018**, *24*, 237–244. [CrossRef]
36. Kah, J.C.Y.; Kho, K.W.; Lee, C.G.L.; James, C.; Sheppard, R.; Shen, Z.X.; Soo, K.C.; Olivo, M.C. Early diagnosis of oral cancer based on the surface plasmon resonance of gold nanoparticles. *Int. J. Nanomed.* **2007**, *2*, 785–798.
37. Connolly, J.M.; Davies, K.; Kazakeviciute, A.; Wheatley, A.M.; Dockery, P.; Keogh, I.; Olivo, M. Non-invasive and label-free detection of oral squamous cell carcinoma using saliva surface-enhanced Raman spectroscopy and multivariate analysis. *Nanomed. Nanotechnol. Biol. Med.* **2016**, *12*, 1593–1601. [CrossRef]
38. Elumalai, B.; Prakasarao, A.; Ganesan, B.; Dornadula, K.; Ganesan, S. Raman spectroscopic characterization of urine of normal and oral cancer subjects. *J. Raman Spectrosc.* **2015**, *46*, 84–93. [CrossRef]
39. Choi, S.; Park, H.K.; Min, G.E.; Kim, Y.H. Biochemical investigations of human papillomavirus-infected cervical fluids. *Microsc. Res. Tech.* **2015**, *78*, 200–206. [CrossRef]
40. Harris, A.T.; Lungari, A.; Needham, C.J.; Smith, S.L.; Lones, M.A.; Fisher, S.E.; Yang, X.B.; Cooper, N.; Kirkham, J.; Smith, D.A.; et al. Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. *Head Neck Oncol.* **2009**, *1*, 34. [CrossRef] [PubMed]
41. Zhang, Y.; Mi, X.; Tan, X.; Xiang, R. Recent Progress on Liquid Biopsy Analysis using Surface-Enhanced Raman Spectroscopy. *Theranostics* **2019**, *9*, 491. [CrossRef] [PubMed]
42. Humphrey, S.P.; Williamson, R.T. A review of saliva: Normal composition, flow, and function. *J. Prosthet. Dent.* **2001**, *85*, 162–169. [CrossRef]
43. Wang, Q.; Gao, P.; Wang, X.; Duan, Y. Investigation and identification of potential biomarkers in human saliva for the early diagnosis of oral squamous cell carcinoma. *Clin. Chim. Acta* **2014**, *427*, 79–85. [CrossRef]
44. Kumar, P. Detection of inaccessible head and neck lesions using human saliva and fluorescence spectroscopy. *Lasers Med. Sci.* **2022**, *37*, 1821–1827. [CrossRef]
45. Lee, L.; Wong, Y.; Hsiao, H.; Wang, Y.; Chan, M.; Chang, K. Evaluation of saliva and plasma cytokine biomarkers in patients with oral squamous cell carcinoma. *Int. J. Oral Maxillofac. Surg.* **2018**, *47*, 699–707. [CrossRef]
46. Hole, A.; Jadhav, P.; Pansare, K.; Noothalapati, H.; Deshmukh, A.; Gota, V.; Chaturvedi, P.; Krishna, C.M. Salivary Raman spectroscopy: Understanding alterations in saliva of tobacco habitués and oral cancer subjects. *Vib. Spectrosc.* **2022**, *122*, 103414. [CrossRef]
47. Koster, H.J.; Guillen-Perez, A.; Gomez-Diaz, J.S.; Navas-Moreno, M.; Birkeland, A.C.; Carney, R.P. Fused Raman spectroscopic analysis of blood and saliva delivers high accuracy for head and neck cancer diagnostics. *Sci. Rep.* **2022**, *12*, 18464. [CrossRef] [PubMed]
48. Reddy, I.; Sherlin, H.J.; Ramani, P.; Premkumar, P.; Natesan, A.; Chandrasekar, T. Amino acid profile of saliva from patients with oral squamous cell carcinoma using high performance liquid chromatography. *J. Oral Sci.* **2012**, *54*, 279–283. [CrossRef]
49. Wang, Q.; Gao, P.; Cheng, F.; Wang, X.; Duan, Y. Measurement of salivary metabolite biomarkers for early monitoring of oral cancer with ultra performance liquid chromatography-mass spectrometry. *Talanta* **2014**, *119*, 299–305. [CrossRef]
50. Tankiewicz, A.; Dziemiańczyk, D.; Buczko, P.; Szarmach, I.; Grabowska, S.; Pawlak, D. Tryptophan and its metabolites in patients with oral squamous cell carcinoma: Preliminary study. *Adv. Med. Sci.* **2006**, *51*, 221–224.
51. Bel'skaya, L.V.; Sarf, E.A.; Solomatin, D.V.; Kosenok, V.K. Diagnostic and Prognostic Value of Salivary Biochemical Markers in Oral Squamous Cell Carcinoma. *Diagnostics* **2020**, *10*, 818. [CrossRef] [PubMed]
52. Chiappin, S.; Antonelli, G.; Gatti, R.; Elio, F. Saliva specimen: A new laboratory tool for diagnostic and basic investigation. *Clin. Chim. Acta* **2007**, *383*, 30–40. [CrossRef]
53. Koduru, M.R.; Ramesh, A.; Adapa, S.; Shetty, J. Salivary albumin as a biomarker for oral squamous cell carcinoma and chronic periodontitis. *Ann. Med. Health Sci. Res.* **2017**, *7*, 337–340.
54. Ramya, A.S.; Uppala, D.; Majumdar, S.; Surekha, C.; Deepak, K. Are salivary amylase and pH–Prognostic indicators of cancers? *J. Oral Biol. Craniofacial Res.* **2015**, *5*, 81–85. [CrossRef] [PubMed]
55. Lin, Y.-T.; Darvishi, S.; Preet, A.; Huang, T.-Y.; Lin, S.-H.; Girault, H.H.; Wang, L.; Lin, T.-E. A review: Electrochemical biosensors for oral cancer. *Chemosensors* **2020**, *8*, 54. [CrossRef]
56. Katakura, A.; Kamiyama, I.; Takano, N.; Shibahara, T.; Muramatsu, T.; Ishihara, K.; Takagi, R.; Shouno, T. Comparison of salivary cytokine levels in oral cancer patients and healthy subjects. *Bull. Tokyo Dent. Coll.* **2007**, *48*, 199–203. [CrossRef]
57. Brailo, V.; Vucicevic-Boras, V.; Lukac, J.; Biocina-Lukenda, D.; Zilic-Alajbeg, I.; Milenovic, A.; Balija, M. Salivary and serum interleukin 1 beta, interleukin 6 and tumor necrosis factor alpha in patients with leukoplakia and oral cancer. *Med. Oral Patol. Oral Cirugía Bucal* **2012**, *17*, e10. [CrossRef]
58. Gustafsson, A.; Ajeti, V.; Ljunggren, L. Detection of suPAR in the saliva of healthy young adults: Comparison with plasma levels. *Biomark. Insights* **2011**, *6*, 119–125. [CrossRef]
59. Santarelli, A.; Mascitti, M.; Lo Russo, L.; Colella, G.; Giannatempo, G.; Bambini, F.; Emanuelli, M.; Procaccini, M.; Lo Muzio, L. Detection level of salivary survivin in patients with OSCC. *J. Carcinog. Mutagen.* **2013**, *5*, 29–85.

60. Feng, Y.; Li, Q.; Chen, J.; Yi, P.; Xu, X.; Fan, Y.; Cui, B.; Yu, Y.; Li, X.; Du, Y.; et al. Salivary protease spectrum biomarkers of oral cancer. *Int. J. Oral Sci.* **2019**, *11*, 7. [CrossRef]
61. Mirzaii-Dizgah, I.; Riahi, E. Serum and saliva levels of cathepsin L in patients with acute coronary syndrome. *J. Contemp. Dent. Pract.* **2011**, *12*, 114–119. [CrossRef]
62. Gholizadeh, N.; Ramandi, M.A.; Motiee-Langroudi, M.; Jafari, M.; Sharouny, H.; Sheykhbahaei, N. Serum and salivary levels of lactate dehydrogenase in oral squamous cell carcinoma, oral lichen planus and oral lichenoid reaction. *BMC Oral Health* **2020**, *20*, 314. [CrossRef]
63. Rakesh Yadav, S.; Deherkar, J.; Sangle, A.; Chandorkar, S. Utility of Salivary CA-125 And LDH As Tumor Markers in Oral Malignancy. *Int. J. Clin. Biomed. Res.* **2018**, *4*, 65–69. [CrossRef]
64. Hoffmann, R.R.; Yurgel, L.S.; Campos, M.M. Evaluation of salivary endothelin-1 levels in oral squamous cell carcinoma and oral leukoplakia. *Regul. Pept.* **2011**, *166*, 55–58. [CrossRef] [PubMed]
65. Cheng, Y.-S.L.; Rees, T.; Jordan, L.; Oxford, L.; O'Brien, J.; Chen, H.-S.; Wong, D. Salivary endothelin-1 potential for detecting oral cancer in patients with oral lichen planus or oral cancer in remission. *Oral Oncol.* **2011**, *47*, 1122–1126. [CrossRef]
66. Pateel, D.G.S.; Gunjal, S.; Math, S.Y.; Murugeshappa, D.G.; Nair, S.M. Correlation of salivary statherin and calcium levels with dental calculus formation: A preliminary study. *Int. J. Dent.* **2017**, *2017*, 2857629. [CrossRef]
67. Contucci, A.; Inzitari, R.; Agostino, S.; Vitali, A.; Fiorita, A.; Cabras, T.; Scarano, E.; Messana, I. Statherin levels in saliva of patients with precancerous and cancerous lesions of the oral cavity: A preliminary report. *Oral Dis.* **2005**, *11*, 95–99. [CrossRef]
68. Geng, X.-F.; Du, M.; Han, J.-X.; Zhang, M.; Tang, X.-F.; Xing, R.-D. Saliva CA125 and TPS levels in patients with oral squamous cell carcinoma. *Int. J. Biol. Markers* **2013**, *28*, 216–220. [CrossRef] [PubMed]
69. Franzmann, E.J.; Reategui, E.P.; Carraway, K.L.; Hamilton, K.L.; Weed, D.T.; Goodwin, W.J. Salivary soluble CD44: A potential molecular marker for head and neck cancer. *Cancer Epidemiol. Prev. Biomark.* **2005**, *14*, 735–739. [CrossRef]
70. Shivashankara, A. Salivary total protein, sialic acid, lipid peroxidation and glutathione in oral squamous cell carcinoma. *Biomed. Res.* **2011**, *22*, 355–359.
71. Bhat, S.; Babu, S.G.; Bhat, S.K.; Castelino, R.L.; Rao, K.; Madi, M. Status of serum and salivary ascorbic acid in oral potentially malignant disorders and oral cancer. *Indian J. Med. Paediatr. Oncol. Off. J. Indian Soc. Med. Paediatr. Oncol.* **2017**, *38*, 306. [CrossRef]
72. Sharma, M.; Sharma, E.; Prabhu, V.; Pai, V.R.; D'souza, J.M.; Harish, S.; Jose, M. Salivary L-fucose as a biomarker for oral potentially malignant disorders and oral cancer. *J. Cancer Res. Ther.* **2020**, *16*, 546. [CrossRef]
73. Arthisri, A.S.; Sathiyamoorthy, A.; Meenakshi, B.; Chandran, C.R. Ratio of salivary sialic acid to fucose as tumor markers in potentially malignant disorders and oral cancer. *Contemp. Clin. Dent.* **2020**, *11*, 131. [CrossRef] [PubMed]
74. Jacob, T.V.; Ramesh, M.; Murali, S.; Ramesh, K.; Sanjay, P.; Abraham, P. A non-invasive study to estimate and compare salivary sialic acid level as tumor marker in patients with pre-cancer and oral cancer. *J. Cancer Res. Ther.* **2016**, *12*, 634. [PubMed]
75. Metzger, K.; Angres, G.; Maier, H.; Lehmann, W.D. Lipoxygenase products in human saliva: Patients with oral cancer compared to controls. *Free Radic. Biol. Med.* **1995**, *18*, 185–194. [CrossRef]
76. Bhat, M.A.; Prasad, K.; Trivedi, D.; Rajeev, B.; Battur, H. Pyruvic acid levels in serum and saliva: A new course for oral cancer screening? *J. Oral Maxillofac. Pathol. JOMFP* **2016**, *20*, 102. [CrossRef] [PubMed]
77. Hu, S.; Arellano, M.; Boontheung, P.; Wang, J.; Zhou, H.; Jiang, J.; Elashoff, D.; Wei, R.; Loo, J.A.; Wong, D.T. Salivary proteomics for oral cancer biomarker discovery. *Clin. Cancer Res.* **2008**, *14*, 6246–6252. [CrossRef] [PubMed]
78. Rajkumar, K.; Ramya, R.; Nandhini, G.; Rajashree, P.; Ramesh Kumar, A.; Nirmala Anandan, S. Salivary and serum level of CYFRA 21-1 in oral precancer and oral squamous cell carcinoma. *Oral Dis.* **2015**, *21*, 90–96. [CrossRef] [PubMed]
79. Gupta, A.; Tripathi, A.; Patil, R.; Kumar, V.; Khanna, V.; Singh, V. Estimation of salivary and serum basic fibroblast growth factor in treated and untreated patients with oral squamous cell carcinoma. *J. Oral Biol. Craniofacial Res.* **2019**, *9*, 19–23. [CrossRef] [PubMed]
80. Magnusson, M.; Pyykkö, I.; van Setten, G.; Norlander, T.; Nastri, A.; Westermark, A. Basic fibroblast growth factor (bFGF) in saliva and oral mucosa in patients with oral lichen planus: Preliminary observations. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endodontology* **2004**, *98*, 324–332. [CrossRef]
81. Lin, D.; Qiu, S.; Huang, W.; Pan, J.; Xu, Z.; Chen, R.; Feng, S.; Chen, G.; Li, Y.; Short, M.; et al. Autofluorescence and white light imaging-guided endoscopic Raman and diffuse reflectance spectroscopy for in vivo nasopharyngeal cancer detection. *J. Biophotonics* **2018**, *11*, e201700251. [CrossRef]
82. Behl, I.; Calado, G.; Vishwakarma, A.; Traynor, D.; Flint, S.; Galvin, S.; Healy, C.M.; Pimentel, M.L.; Malkin, A.; Byrne, H.J.; et al. Classification of cytological samples from oral potentially malignant lesions through Raman spectroscopy: A pilot study. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *266*, 120437. [CrossRef]
83. Larion, M.; Dowdy, T.; Ruiz-Rodado, V.; Meyer, M.W.; Song, H.; Zhang, W.; Davis, D.; Gilbert, M.R.; Lita, A. Detection of metabolic changes induced via drug treatments in live cancer cells and tissue using Raman imaging microscopy. *Biosensors* **2018**, *9*, 5. [CrossRef]
84. Ibrahim, O.; Toner, M.; Flint, S.; Byrne, H.J.; Lyng, F.M. The Potential of Raman Spectroscopy in the Diagnosis of Dysplastic and Malignant Oral Lesions. *Cancers* **2021**, *13*, 619. [CrossRef]

85. Kuzmin, A.N.; Pliss, A.; Rzhevskii, A.; Lita, A.; Larion, M. BCAbox algorithm expands capabilities of Raman microscope for single organelles assessment. *Biosensors* **2018**, *8*, 106. [CrossRef]
86. Kuzmin, A.N.; Levchenko, S.M.; Pliss, A.; Qu, J.; Prasad, P.N. Molecular profiling of single organelles for quantitative analysis of cellular heterogeneity. *Sci. Rep.* **2017**, *7*, 6512. [CrossRef]
87. Malini, R.; Venkatakrishna, K.; Kurien, J.; Pai, K.M.; Rao, L.; Kartha, V.; Krishna, C.M. Discrimination of normal, inflammatory, premalignant, and malignant oral tissue: A Raman spectroscopy study. *Biopolym. Orig. Res. Biomol.* **2006**, *81*, 179–193. [CrossRef]
88. Li, B.; Gu, Z.-Y.; Yan, K.-X.; Wen, Z.-N.; Zhao, Z.-H.; Li, L.-J.; Li, Y. Evaluating oral epithelial dysplasia classification system by near-infrared Raman spectroscopy. *Oncotarget* **2017**, *8*, 76257. [CrossRef]
89. Ruiz-Rodado, V.; Lita, A.; Larion, M. Advances in measuring cancer cell metabolism with subcellular resolution. *Nat. Methods* **2022**, *19*, 1048–1063. [CrossRef]
90. Yuvaraj, M.; Udayakumar, K.; Jayanth, V.; Rao, A.P.; Bharanidharan, G.; Koteeswaran, D.; Munusamy, B.D.; Ganesan, S. Fluorescence spectroscopic characterization of salivary metabolites of oral cancer patients. *J. Photochem. Photobiol. B Biol.* **2014**, *130*, 153–160. [CrossRef]
91. Venugopal, C.; Nazeer, S.S.; Balan, A.; Jayasree, R. Autofluorescence spectroscopy augmented by multivariate analysis as a potential noninvasive tool for early diagnosis of oral cavity disorders. *Photomed. Laser Surg.* **2013**, *31*, 605–612. [CrossRef]
92. Wang, J. Real-time calibrating polarization-sensitive diffuse reflectance handheld probe characterizes clinically relevant anatomical locations of oral tissue in vivo. *Biomed. Opt. Express* **2022**, *13*, 105–116. [CrossRef]
93. Greening, G.J.; James, H.M.; Dierks, M.K.; Vongkittiargorn, N.; Osterholm, S.M.; Rajaram, N.; Muldoon, T.J. Towards monitoring dysplastic progression in the oral cavity using a hybrid fiber-bundle imaging and spectroscopy probe. *Sci. Rep.* **2016**, *6*, 26734. [CrossRef]
94. Zeng, H.; McWilliams, A.; Lam, S. Optical spectroscopy and imaging for early lung cancer detection: A review. *Photodiagnosis Photodyn. Ther.* **2004**, *1*, 111–122. [CrossRef]
95. Ibrahim, O. The Potential of Raman Spectroscopy in the Diagnosis of Premalignant Oral Lesions. Ph.D. Thesis, Technological University Dublin, Dublin, Ireland, December 2017.
96. Jaychandran, S.; Meenapriya, P.; Ganesan, S. Raman spectroscopic analysis of blood, urine, saliva and tissue of oral potentially malignant disorders and malignancy-A diagnostic study. *Int. J. Oral Craniofacial Sci.* **2016**, *2*, 011–014.
97. Nour, M.; Hamdy, O.; Faid, A.H.; Eltayeb, E.A.; Zaky, A.A. Utilization of gold nanoparticles for the detection of squamous cell carcinoma of the tongue based on laser-induced fluorescence and diffuse reflectance characteristics: An in vitro study. *Lasers Med. Sci.* **2022**, 1–10. [CrossRef]
98. Nazeer, S.S.; Asish, R.; Venugopal, C.; Anita, B.; Gupta, A.K.; Jayasree, R.S. Noninvasive assessment of the risk of tobacco abuse in oral mucosa using fluorescence spectroscopy: A clinical approach. *J. Biomed. Opt.* **2014**, *19*, 057013. [CrossRef]
99. Francisco, A.L.N.; Correr, W.R.; Azevedo, L.H.; Kern, V.G.; Pinto, C.A.L.; Kowalski, L.P.; Kurachi, C. Fluorescence spectroscopy for the detection of potentially malignant disorders and squamous cell carcinoma of the oral cavity. *Photodiagnosis Photodyn. Ther.* **2014**, *11*, 82–90. [CrossRef]
100. Francisco, A.L.N.; Correr, W.R.; Pinto, C.A.L.; Gonçalves Filho, J.; Chulam, T.C.; Kurachi, C.; Kowalski, L.P. Analysis of surgical margins in oral cancer using in situ fluorescence spectroscopy. *Oral Oncol.* **2014**, *50*, 593–599. [CrossRef]
101. Scepanovic, O.R.; Fitzmaurice, M.; Miller, A.; Kong, C.-R.; Volynskaya, Z.I.; Dasari, R.R.; Kramer, J.R., Jr.; Feld, M.S. Multimodal spectroscopy detects features of vulnerable atherosclerotic plaque. *J. Biomed. Opt.* **2011**, *16*, 011009. [CrossRef]
102. Summers, K.L.; Fimognari, N.; Hollings, A.; Kiernan, M.; Lam, V.; Tidy, R.J.; Paterson, D.; Tobin, M.J.; Takechi, R.; George, G.N.; et al. A multimodal spectroscopic imaging method to characterize the metal and macromolecular content of proteinaceous aggregates ("amyloid plaques"). *Biochemistry* **2017**, *56*, 4107–4116. [CrossRef]
103. Volynskaya, Z.I. Multimodal Spectroscopy: Real-Time Diagnosis of Breast Cancer during Core Needle Biopsy. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010.
104. Salomatina, E.; Muzikansky, A.; Neel, V.; Yaroslavsky, A.N. Multimodal optical imaging and spectroscopy for the intraoperative mapping of nonmelanoma skin cancer. *J. Appl. Phys.* **2009**, *105*, 102010. [CrossRef]
105. Kowalski, L.P.; Bagietto, R.; Lara, J.R.; Santos, R.L.; Silva, J.F., Jr.; Magrin, J. Prognostic significance of the distribution of neck node metastasis from oral carcinoma. *Head Neck J. Sci. Spec. Head Neck* **2000**, *22*, 207–214. [CrossRef]
106. Brennan, J.A.; Mao, L.; Hruban, R.H.; Boyle, J.O.; Eby, Y.J.; Koch, W.M.; Goodman, S.N.; Sidransky, D. Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* **1995**, *332*, 429–435. [CrossRef]
107. Wikner, J.; Gröbe, A.; Pantel, K.; Riethdorf, S. Squamous cell carcinoma of the oral cavity and circulating tumour cells. *World J. Clin. Oncol.* **2014**, *5*, 114. [CrossRef]
108. Moros, J.; Garrigues, S.; de la Guardia, M. Vibrational spectroscopy provides a green tool for multi-component analysis. *TrAC Trends Anal. Chem.* **2010**, *29*, 578–591. [CrossRef]
109. Barroso, E.; Smits, R.; Bakker Schut, T.; Ten Hove, I.; Hardillo, J.; Wolvius, E.; Baatenburg de Jong, R.; Koljenovic, S.; Puppels, G. Discrimination between oral cancer and healthy tissue based on water content determined by Raman spectroscopy. *Anal. Chem.* **2015**, *87*, 2419–2426. [CrossRef]
110. Huck, C.W. Advances of vibrational spectroscopic technologies in life sciences. *Molecules* **2017**, *22*, 278. [CrossRef]

111. Gautam, R.; Vanga, S.; Ariese, F.; Umapathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2015**, *2*, 8. [CrossRef]
112. Nogueira, M.S.; Maryam, S.; Amissah, M.; Lu, H.; Lynch, N.; Killeen, S.; O'Riordain, M.; Andersson-Engels, S. Evaluation of wavelength ranges and tissue depth probed by diffuse reflectance spectroscopy for colorectal cancer detection. *Sci. Rep.* **2021**, *11*, 798. [CrossRef]
113. Wabnitz, H.; Taubert, D.R.; Mazurenka, M.; Steinkellner, O.; Jelzow, A.; Macdonald, R.; Milej, D.; Sawosz, P.; Kacprzak, M.; Liebert, A. Performance assessment of time-domain optical brain imagers, part 1: Basic instrumental performance protocol. *J. Biomed. Opt.* **2014**, *19*, 086010. [CrossRef] [PubMed]
114. Pifferi, A.; Torricelli, A.; Bassi, A.; Taroni, P.; Cubeddu, R.; Wabnitz, H.; Grosenick, D.; Möller, M.; Macdonald, R.; Swartling, J.; et al. Performance assessment of photon migration instruments: The MEDPHOT protocol. *Appl. Opt.* **2005**, *44*, 2104–2114. [CrossRef]

Paper II

Analyst

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2023, **148**, 1514

Mobile multi-configuration clinical translational Raman system for oral cancer application

Siddra Maryam, *[a] Sanathana Konugolu Venkata Sekar,[a] M. Daniyal Ghauri,[a] Edward Fahy,[b] Marcelo Saito Nogueira,[a] Huihui Lu,[a] Flavien Beffara,[c] Georges Humbert,[c] Richeal Ni Riordain,[b,d] Patrick Sheahan,[e] Ray Burke,[a] Kiang Wei Kho,[a] Rekha Gautam [a] and Stefan Andersson-Engels [a]

Received 23rd November 2022,
Accepted 22nd February 2023

DOI: 10.1039/d2an01921c

rsc.li/analyst

Early diagnosis of oral cancer is critical to improve the survival rate of patients. Raman spectroscopy, a non-invasive spectroscopic technique, has shown potential in identifying early-stage oral cancer biomarkers in the oral cavity environment. However, inherently weak signals necessitate highly sensitive detectors, which restricts widespread usage due to high setup costs. In this research, the fabrication and assembly of a customised Raman system that can adapt three different configurations for the *in vivo* and *ex vivo* analysis is reported. This novel design will help in reducing the cost required to have multiple Raman instruments specific for a given application. First, we demonstrated the capability of a customized microscope for acquiring Raman signals from a single cell with high signal-to-noise ratio. Generally, when working with liquid samples with low concentration of analytes (such as saliva) under a microscope, excitation light interacts with a small sample volume, which may not be representative of whole sample. To address this issue, we have designed a novel long-path transmission set-up, which was found to be sensitive towards low concentration of analytes in aqueous solution. We further demonstrated that the same Raman system can be incorporated with the multimodal fibre optical probe to collect *in vivo* data from oral tissues. In summary, this flexible, portable, multi-configuration Raman system has the potential to provide a cost-effective solution for complete screening of precancer oral lesions.

1. Introduction

Oral cancer refers to pathologies affecting the tongue, floor of mouth, upper and lower alveolus, palate, and buccal mucosa.[1,2] It is described as 16th most common cancer worldwide, often detected in advanced stage, resulting in an overall survival rate of 40–60%.[3] The survival rate could be significantly improved to <80% [4–6] if detected in precancerous stage. Primary care delays in recognizing suspicious symptoms remains an ongoing problem in cancer diagnosis.[7] To address this, previously various physician-friendly point-of-care optical techniques that can provide easy, inexpensive, rapid and on-site monitoring such as fluorescence, diffuse reflectance and vibrational spectroscopy have been explored.[8] Among them, Raman spectroscopy has the potential to generate data with detailed molecular information. In addition, Raman spectroscopy is preferred for analysing bio-fluids and *in vivo* measurements as water is a weak Raman scatter due to the polar nature of the OH bond, providing minimum interference in the fingerprint region (200–1800 cm^{-1}).[9,10] Raman spectroscopy is generally a non-invasive approach used to gain chemical and structural information as well as to study molecular interactions.[11–13] It can be used for quantitative and qualitative analysis of a wide range of samples and materials. It has proven particularly useful in industrial and medical applications having resolved limitations of other spectroscopic techniques.[14] In biomedical optics and clinical applications, Raman spectroscopy in combination with multivariate analysis has proven its potential. The ability of detecting biochemical changes in tissues and cells can lead to a real time clinical diagnostic capabilities. Several studies have highlighted the importance of Raman spectroscopy for tissue differentiation in *ex vivo* and *in vivo* analysis.[15–17] Typically, a modern Raman spectrometer consist of an excitation source – usually a monochromatic laser, optical components to both transmit laser energy to specimen and collect back scattered signal, a notch filter to reject Rayleigh scattered light, a grating spectrometer and a high-end charged coupled device (CCD) camera to capture a Raman spectrum.

[a]*Tyndall National Institute, University College Cork, Cork, Ireland. E-mail: siddra.maryam@tyndall.ie*
[b]*Cork University Dental School and Hospital, Wilton, Cork, Ireland*
[c]*XLIM Research Institute, UMR 7252 CNRS/Limoges University, Limoges, France*
[d]*ENTO Research Institute, University College Cork, Cork, Ireland*
[e]*South Infirmary Victoria University Hospital, Cork, Ireland*

When the light from a laser interacts with the specimen, a small part of the incident energy is scattered inelastically (1 out of 10^6 photons) to produce a Raman spectrum.[18] In cases where the sample volume is too small or a microscopic area on a large sample needs to be analysed, a Raman spectrometry system coupled with a specially designed microscope would be desirable. Such instruments are commercially available at high cost and are bulky, non-portable, sample volume limited and are not considered suitable for point-of-care application. Additionally, Raman spectrometer coupled with the microscope are generally used as benchtop systems for *ex vivo* analysis owing to its bulkiness and limited flexibility to be used with the fibre optical probe and are therefore not suitable for clinical environment. To facilitate early detection at the primary-care level, as well as in research studies, we envision the use of a highly configurable spectrometry system suitable for seamless integration with the clinical procedure. Here, we describe an affordable, Raman-based instrument capable of both *in vivo* and *ex vivo* analysis employing single platform which can be easily wheeled to provide on-site testing and can be operated by a qualified clinician/lab technician with training and practice. Spectroscopic cell and supernatant analysis would be done to guide if biopsy sampling is necessary. After a positive test result of saliva or cell pellet analysis, *in vivo* measurements can be conducted just before tissue biopsy to guide for a suitable location of biopsy sampling and thereby provide improved sensitivity/specificity of diagnosis. Initially, the data analysis will be done by the experts, to later be automated after collecting sufficient data for machine learning models. To the best of our knowledge, this is the first report describing the design of multi-configurational Raman system addressing the limitations of the commercial clinical instruments.

The novel Raman system described here is compact, cost effective, portable and easy to assemble and align. It can easily fit into a console for transportation, thereby potentially facilitating point-of-care diagnostics. The design of our instrument is modular, consisting of the following:

Mode (A) Raman microscope – for *ex vivo* saliva cell pellet analysis.

Mode (B) Raman transmission stage – for saliva supernatant analysis using photonic crystal fibre (PCF).

Mode (C) Raman optical fibre probe – for *in vivo* tissue analysis.

This makes it suitable for both *in vivo* and *ex vivo* analysis and can be used in upright and transmission modes for liquid samples. The detailed optical design is given in the Methods section, enabling the easy adaptation by researchers in the field. In this study, we demonstrate the utility of our system in the context of oral cancer screening. The described system is only a prototype. Final system can employ centrifugation microfluidics[19] to separate cell pellet from supernatant, which will minimize human interventions. A typical work-flow of the procedure is shown in Fig. 1.

2. Experimental method

2.1 Instrumentation

Fig. 2 presents the configuration modes of the point-of-care Raman setup developed. This modular instrument consisted of: an illumination source, the sample probing subunit and

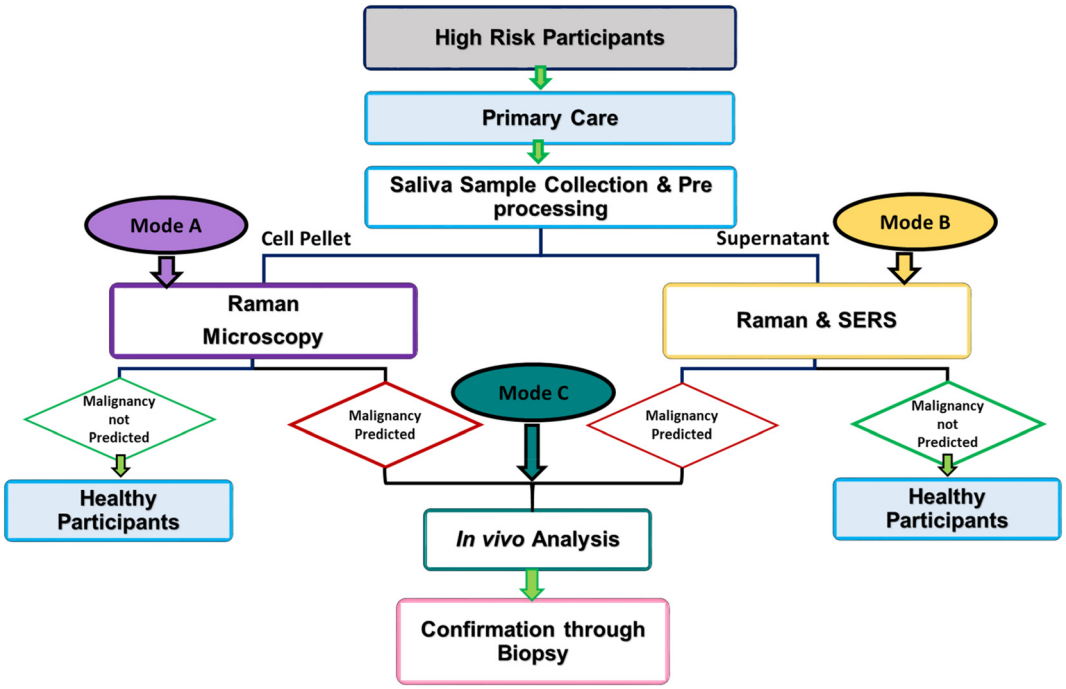


Fig. 1 Flow chart of study protocol with multi-configurational customised Raman system.

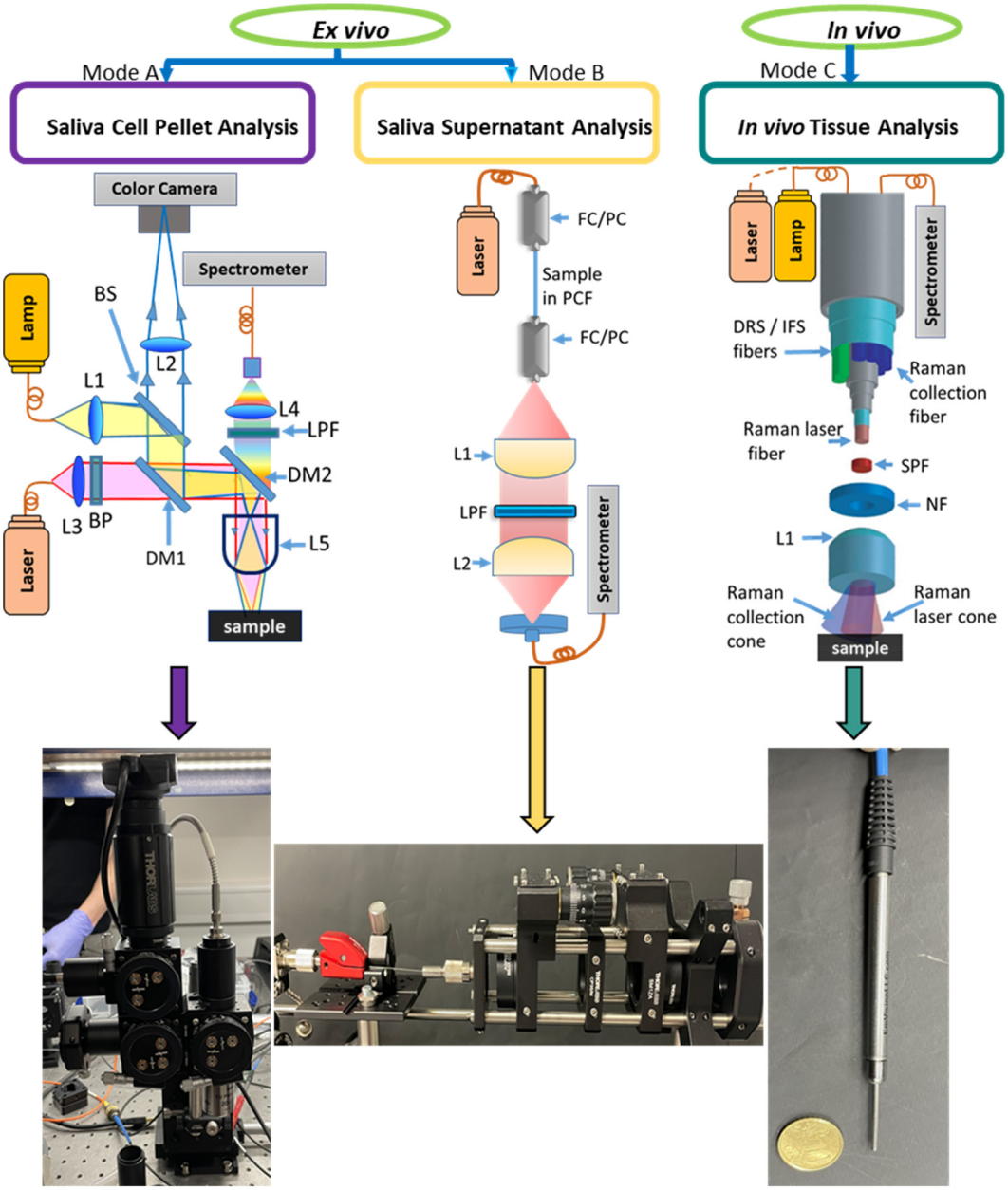


Fig. 2 Three configuration modes of the new customised Raman setup designed to include 3 plug and play modules to be coupled with common light sources and a spectrometer. These modes consist of: Mode (A) customised upright Raman microscope in combination with optical microscopy; Mode (B) transmission mode Raman stage; Mode (C) optical fibre probe for *in vivo* Raman analysis.

the spectrometer. Using the fibre optical connection between the subunits, the sample probing unit was easily switched between the three configuration modes. The illumination was achieved by a 785 nm narrow line width laser with high suppression of side bands (>60 dB suppression at 0.3 nm from the laser peak, Cobolt 08-01 Series, Cobolt, Sweden). The Raman spectrometer was equipped with a diffraction grating with 1200 groves per mm grating density (LS-785, Teledyne Princeton Instruments, USA), a 1340 × 400 pixels back illuminated CCD camera (Blaze: 400HR, Teledyne Princeton Instruments, USA). These subunits can be connected to the custom-built microscope (Fig. 2, Mode A). This is achieved fibre-optically using a round-to-linear bundle of a 7 × 200 μm core, (Thorlabs BFL200LS02). This Raman system can also utilise the customised transmission stage for the analysis of liquid specimens in transmission mode (Fig. 2, Mode B) and an optical fibre probe for the *in vivo* analysis (Fig. 2, Mode C). Currently, it takes up to five minutes to switch between configurations, mainly fibers. The light sources and calibration are the same for all three configurations. However, the final system will be more automated, which will seamlessly switch between different configurations. This prototype including

microscope and transmission mode stage is designed using off the shelf component in around 10k excluding the spectrometer and detector. The following section describes each of these configurations in detail with the novel optical design.

2.2 Mode (A) Raman microscope for *ex vivo* cell pellet analysis

Fig. 2, Mode A represents the customised Raman microscope which consists of a laser and a visible light source (for sample visualization) connected through optical fibres. Fibre coupled illumination allowed modification of the spot size on the sample plane. Further, this fibre guided light delivery makes the switching of different excitation wavelengths easy. Another important feature of this system is a flexible stage design, which can accommodate different specimen in different dimensions from mm to tens of cm. The specimen was illuminated through an objective lens and Raman scattered photons were collected *via* the same objective lens in backscattering mode and in turn, directed to the spectrometer *via* 7× round-to-linear low OH fibre bundle with 200 μm core (Thorlabs BFL200LS02). Simultaneously, a portion light from sample was patched to a CCD detector. This configuration was used for *ex vivo* analysis of solid and liquid specimens in the reflective mode.

White light imaging optical path. A broadband visible light source (Thorlabs M365LP1, 365 nm, 1350 mW mounted LED) was used to illuminate the specimen for conventional optical microscopy. The light beam entered the microscope *via* lens L1 (Thorlabs LA1951A, f = 25 mm) and redirected to the objective lens using a beam splitter BS (Thorlabs, BSW26R, 50 : 50, 350–1100 nm), dichroic mirror DM1 (Semrock, FF757-Di01) and dichroic mirror DM2 (Semrock LPD02-785RU). The L1 was arranged to provide uniform Kohler illumination of white light on the sample plane. The white light scattered from the sample was reflected by DM2 to DM1 which redirected it to a colour camera through BS and a field lens L2 (Thorlabs LB1437-A, f = 150 mm). A high-resolution coloured CMOS camera (Thorlabs DCC1645C) was used to visualize the sample under white light illumination. The field of view (FOV) was controlled using a microscope objective lens L4 (116 × 92 μm for 50×, 290 × 230 μm for 20× and 1160 × 920 μm for 5×). The images were displayed on computer screen using the Thorcam software.

Raman optical path. Monochromatic light (785 nm laser) passed through the collimation lens L3 (Thorlabs A397TM-B, f = 11 mm, NA (numerical aperture) = 0.3) and laser cleaning bandpass filter BP (Semrock LD01-785/10-25). Afterwards, the collimated excitation beam passed through the DM1 and was reflected by DM2, which was focused on the sample plane by microscope objective. The spot size of the focused excitation beam was 60 μm with 20× objective lens (Fig. 4). The Raman scattered photons from the sample passed through the objective lens and the excitation light is filtered out using a dichroic mirror (DM2) and long pass filter LPF (Semrock LP02-785RU-25). The filtered Raman light (free of excitation light) was focused on to a low OH circular to linear fibre bundle (Thorlabs BFL200LS02) using a lens L4 (Thorlabs F810SMA-850, f = 36.2 mm). The linear side of fibre was attached to the entrance linear slit of the Raman spectrometer. The Raman spectrometer subunit contained a filter chamber with a long pass filter (Semrock LP02-785RU-25), a high throughput spectrometer (Teledyne Princeton Instruments Acton LS785, f/2), and a highly sensitive deep depleted CCD camera (Blaze: 400HR, Teledyne Princeton Instruments, USA). The CCD pixels were binned vertically to combine light from linear slit. The light from the CCD is acquired using the LightField software (Teledyne Princeton Instruments, USA).

2.3 Mode (B) Raman transmission stage for saliva supernatant analysis using PCF

Fig. 2, Mode B and Fig. 3B illustrate a horizontal stage for liquid specimen analysis in transmission mode using PCF. PCF are optical fibers composed of periodic array of air-holes surrounding a solid or hollow-core. In this study, the fiber used is the simplest PCF that is composed of silica core (SC-PCF) at the intersection of three large air holes. Light is guided in the triangular core shape by the total internal reflection mechanism based on the large refractive index difference between the silica and air, and by the reduction of the thickness of the three silica struts enhancing light confinement in the core.[20,21] To integrate the inherently weak Raman signal from a liquid specimen (saliva supernatant) over a large volume, the SC-PCF with 3.5 μm core diameter, >70 μm air-channel diameter, and NA = 0.5 was filled with aqueous liquid. The external diameter (200 μm) of this SC-PCF is slightly larger than a standard fiber for allowing easier handling. In this SC-PCF, Raman signals are generated by the evanescent wave extended into the holes while excitation light (785 nm) propagates through the solid core.[20,22,23] The sample holder was used in this way to prolong light–sample interaction length, thereby improving detection sensitivity.[24] Another advantage with a longer path length than in a microscope was that the larger volume sampled can be more representative, especially for low concentration analytes. A further advantage was that still a very small volume of analyte was required (about 20–100 nl). A PCF filled with sample solution can be directly placed under the Raman microscope to collect the signal in conventional reflective mode. However, to ease the alignment of input laser beam in to the PCF, we designed a horizontal optical stage enabling the collection of signals in transmission mode. A 785 nm laser beam entered the PCF through an FC/PC connector (ferrule connector/physical contact connector). Raman signal emitted from the other end was collimated by an aspheric lens L1 (Thorlabs S-LAH64, f = 20 mm, NA = 0.54, ARC: 650–1050 nm). This lens was selected to match the NA of the PCF enhancing the collection efficiency. The collimated Raman signal, after passing through a long pass filter (LPF) (Semrock LP02-785RU-25), was collected by another lens L2 (Thorlabs N-BK7, f = 50 mm, NA = 0.23, ARC: 650–1050 nm) and focused on a low OH circular to linear fibre bundle (Thorlabs BFL200LS02). To transfer the

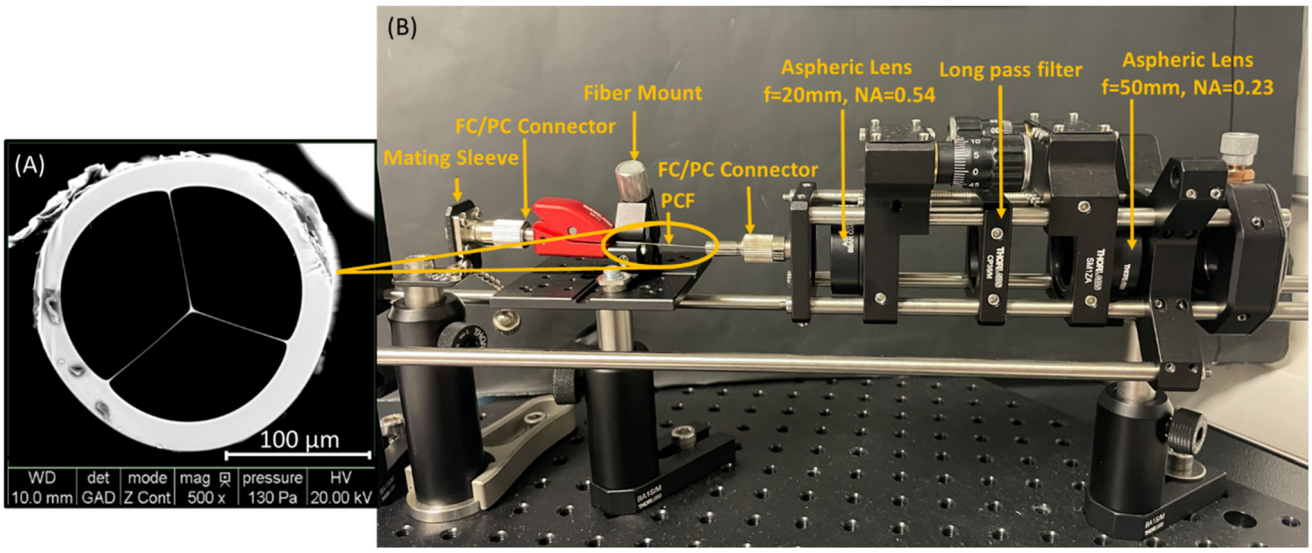


Fig. 3 (A) Cross section of a suspended core PCF designed by the authors (B) transmission mode stage for the analysis of liquid specimen injected into a PCF.

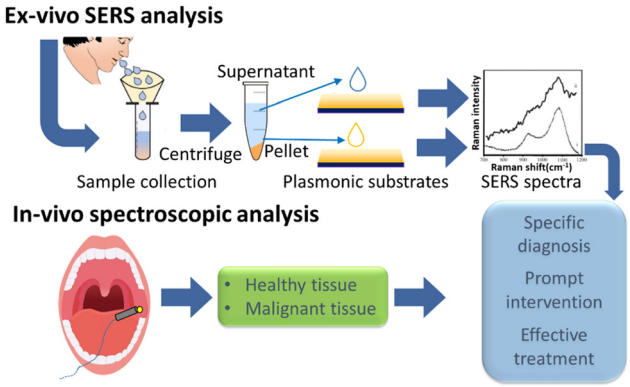


Fig. 4 Sample preparation and methodology for *ex vivo* saliva and *in vivo* tissue analysis.

signal to the spectrometer. The Fig. 3(A) shows the cross section of a typical suspended core PCF used in this study.

2.4 Mode (C) Raman optical fibre probe for *in vivo* tissue analysis

Fiber optics-based Raman spectroscopic system has a tremendous potential for real time *in vivo* tissue analysis in endoscopy or surgical guidance. However, for such applications small and flexible fiber optics probes are mandatory. This requirement has limited the clinical implementation of fiber optics based Raman technology for many years.[25] Recent advancement in fiber optics technology and miniaturization of the optical components have improved the *in vivo* implementation.[26–28] Fig. 2, Mode C shows a multi-spectroscopy fibre-optic probe (EMVision, Loxahatchee, FL, USA) used for the *in vivo* tissue analysis. This customised *in vivo* probe can be used to acquire spectra from multimodal spectroscopic techniques including Raman, fluorescence (FS) and diffuse reflectance spectroscopy (DRS). The probe consisted of 9 × 300 µm core diameter, 0.22 NA low OH silica fibres arranged in a concentric ring around a 300 µm core diameter, 0.22 NA low OH silica fibre used for Raman excitation. Seven out of nine fibres from the circular geometry collected the Raman signal passed through a donut shaped notch filter (NF), which rejected the laser line while transmitting the Raman signals from the probed tissue. The remaining two fibres can be used for DRS and FS. There was a short pass emission filter (SPF) in front of the central Raman illumination fibre to clean input laser line. A two-component customised lens L1 is placed at the probe tip to overlap the focus of excitations light and the area of collection of the Raman light from the tissue. Seven Raman collection fibres were aligned into a single line on the other end of the probe to deliver collected signals to the spectrometer *via* 50 µm slit. The accumulated signals from the seven fibres provide an optimum signal-to-noise ratio, needed for biological samples.

2.5 Sample preparation

Patients suffering from oral squamous cell carcinoma and other malignant diseases in the oral cavity going through biopsy and histopathological examination were enrolled in a study approved by the Clinical Research Ethics Committee (CREC) of University College Cork (CREC ref.: ECM 4 (y) 13/4/2021 & ECM 3 (cc) 11/05/2021). Written informed consent was obtained from human participants and study was performed in accordance with the Guidelines of Helsinki.

The performance of the multi-configuration Raman system was demonstrated for different healthy tissues of a single participant and using saliva sample collected from the same participant. A saliva specimen was collected and centrifuged at 10 000 rpm for 10 min at room temperature to separate it in supernatant and cell pellet and stored at −80 °C for further

use. For saliva sample analysis with the microscope, 3 drops of 10 μL saliva was placed and dried on a piece of aluminium foil. The cheek and red blood cells from the cell pellet were smeared on the aluminium foil to collect Raman spectra from a single cell layer.

To prepare the PCF sample for saliva analysis in transmission mode stage, a 9–10 cm long PCF with 3.5 μm core diameter and NA = 0.5 was cleaved at both ends. Its polymer jacket was removed from the ends and it was filled with saliva supernatant through capillary action.

In vivo analysis involves measurements on healthy tissues and lesion from the same participants using the fibre-optic probe. For the collection of *in vivo* Raman spectra, the probe was placed to just touch the area of interest.

2.6 Data analysis

Raman spectra were background corrected using the asymmetric least squares method,[29] noise smoothing with a Savitzky–Golay filter of polynomial order 2 and a window width of 7, and normalized to their respective mean intensity using MATLAB R2020b software (Mathworks, Inc., Natick, MA, USA).

3. Results and discussion

3.1 Mode (A) Raman microscope for *ex vivo* cell pellet analysis

Analysis of liquid biopsies has recently attracted attention as an emerging field in medical diagnosis. Biological fluids such as blood, saliva, urine and tears are rich in biomarkers that could be indicative of the health status of the sampled individual.[30–32] This provides a minimal invasive route for cancer screening and to classify cancer patients based on biomarkers identified.[33,34] The proposed Raman microscope can be used for the *ex vivo* analysis of any excised tissue or for liquid biopsies.

Fig. 5(A) illustrates Raman spectra from saliva specimen in both dried and liquid phase as well as from the mono layer of red blood cell and cheek cells as shown in Fig. 5B and C respectively. As can be seen, the Raman spectrum of saliva sample is featureless when wet but exhibits distinct peaks when dried. Several peaks characteristic of saliva are discernible.[35] For instance, peaks around 1049–1053 cm^{-1} are assignable to C–C stretching vibration of lipids. Furthermore, a well-known Raman bands characteristic of amide-I vibration of salivary proteins at around 1639–1670 cm^{-1} are also discernible. Also, unique to the saliva specimen are two additional peaks at 877 cm^{-1} and 938 cm^{-1} corresponding to a mixture of –C–O–C– and –C–C–, ring and stretching modes.[36] The Raman spectrum of red blood cell, on the other hand, was dominated by the haemoglobin signatures clearly differentiating it from cheek cell and saliva. Specifically, prominent peaks at around 1618 cm^{-1} and 1562 cm^{-1} can be assigned to the $C_a{=}C_b$ vibration of the vinyl groups, and C_β–C_β stretching modes in heme, respectively. Whereas peaks at 752 and 1395 are associated with pyrrole ring in heme. All the important differentiating Raman bands are presented in Table 1. Previously, Raman has been used to detect structural and biochemical changes in salivary analytes (*e.g.* proteins, amino acids, nucleic acids) accompanied with diseased transformation.[17] However due to generally weak signals and the strong background interference, most of the studies have been done on dried saliva samples to concentrate the analytes or utilized plasmon

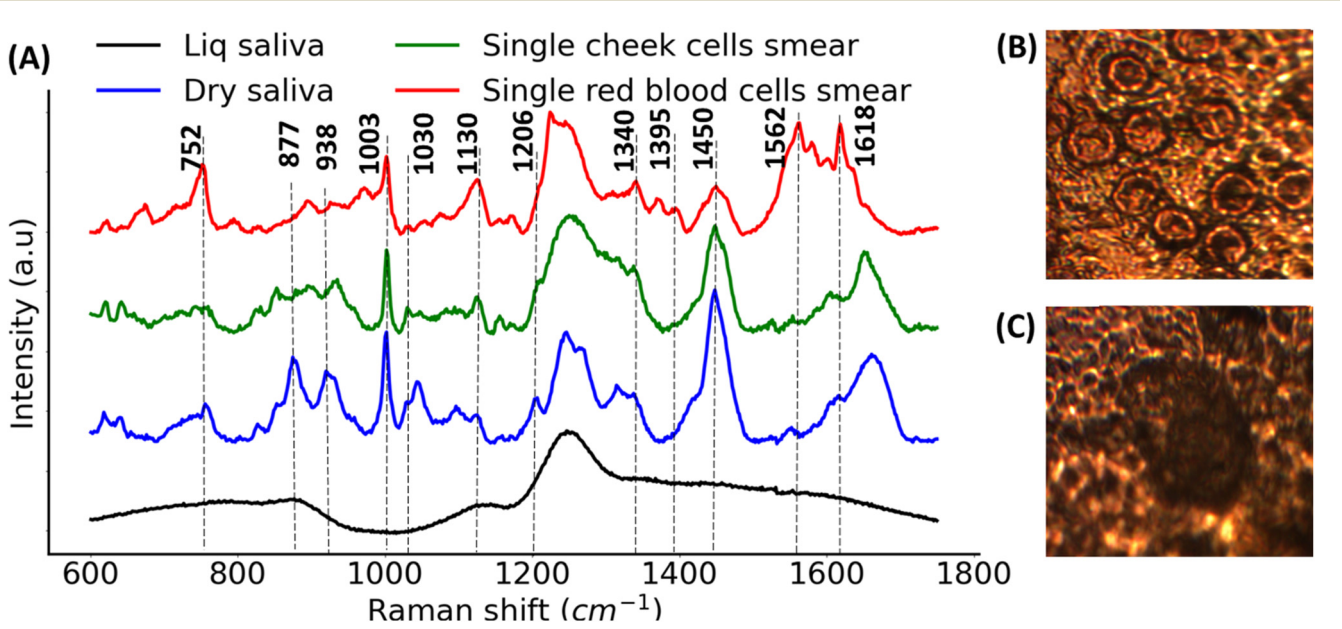


Fig. 5 (A) Pre-processed Raman spectra from saliva specimen in liquid and dry form. Right-optical microscopic view of (B) red blood cells (C) smeared cheek cell on aluminium foil.

Table 1 Spectral assignment of the dominated Raman signatures observed in experimental data[36,40]

Wavenumber (cm^{-1})	Spectral assignments
752	ν(pyrrole breathing) stretch in heme; symmetric ring breathing in tryptophan
877	C–O–C ring (polysaccharides);C–C stretching (hydroxyproline)
938	C–O–C stretching (polysaccharides); C–C stretch α-helix proteins
960	PO_4^{3-} symmetric stretch (hydroxyapatite)
1003	Symmetric ring breathing mode of phenylalanine
1030	C–H in plane bending of phenylalanine
1080–1090	C–C stretch acyl chains (*gauche*) lipids; PO_2^- stretching (DNA); $\nu(C_\beta C_1)$ asymmetric stretch (hemoglobin)
1130	C–C stretch acyl chains (*trans*) lipids; C–O stretch polysaccharides; ν(pyrrole half-ring) asymmetric stretch (hemoglobin)
1204–1209	C–C_6H_5 stretch phenylalanine and tryptophan
1220–1320	Amide III-combination of NH in-plane bending and CN stretching
1302	CH_2 deformation (lipids)
1340	Nucleic acid (adenine, guanine); CH deformation polysaccharides
1395	ν(pyrrole quarter-ring) stretch (hemoglobin)
1440	CH_2 bending mainly in lipids
1450	CH_2 bending mainly in proteins
1562	$\nu(C_\beta C_\beta)$ stretch (hemoglobin)
1618	$\nu(C_a{=}C_b)$ of vinyl groups (hemoglobin)
1656	C=C stretch mainly in lipids
1630–1690	Amide I: mainly C=O stretch in proteins
1740	C=O in lipids

enhancement to improve the weak Raman signals from low concentration analytes.[37] The drying process leads to the structural/conformational changes of the analytes (*e.g.* proteins) and SERS struggles with reproducibility. The use of low NA longer focal depth objective lens (*e.g.* 5× or 10×) have poor collection efficiency and a focal length of tens of micron implies interaction of laser light with the sample is limited and thus requires long integration time, typically tens of seconds. These are not suitable for fast detection. Recently, fiber-enhanced Raman spectroscopy (FERS) where hollow core fibers and PCF have been used to confine the excitation light along the fiber (mm) length, thereby enabling the strong and efficient process of light–matter interactions (>50 times of conventional microscopic measurements) leading to enhanced Raman signals.[38,39] The sample volume probed at the focus of an objective will always be much smaller than the total volume probed over the length of these fibers.[38] To implement FERS, we developed the novel transmission mode stage (B) in which signal gets integrated over an extended interaction length with the light.[23,38,39]

3.2 Mode (B) Raman transmission stage for saliva supernatant analysis using PCF

The interaction of light with the analytes along the PCF length is critical for sensitive detection of low concentration analytes. In such measurements one must consider light attenuation yielding a reduction of transmission signal due to scattering and absorption. It is thus important to characterize the optimum PCF length. For this, the transmission stage was used to optimize the length of PCF using vegetable oil as an analyte because its consistency and refractive index is close to that of saliva (for saliva refractive index $n \approx 1.4$ and for oil $n \approx 1.46$).[41,42] It also has a distinct Raman signal as discussed in Table 1. Different lengths of PCF ranging from 9 cm to 13 cm were tested. The results are shown in Fig. 6A, highlighting the maximum Raman signal for 9 cm and 10 cm PCF lengths. Subsequent increase in the length >10 cm of the PCF lead to gradual decrease in Raman photons due to absorption/scattering loss. The curves in the Fig. 6B represent Raman signals for different modes of our new Raman system using oil as an

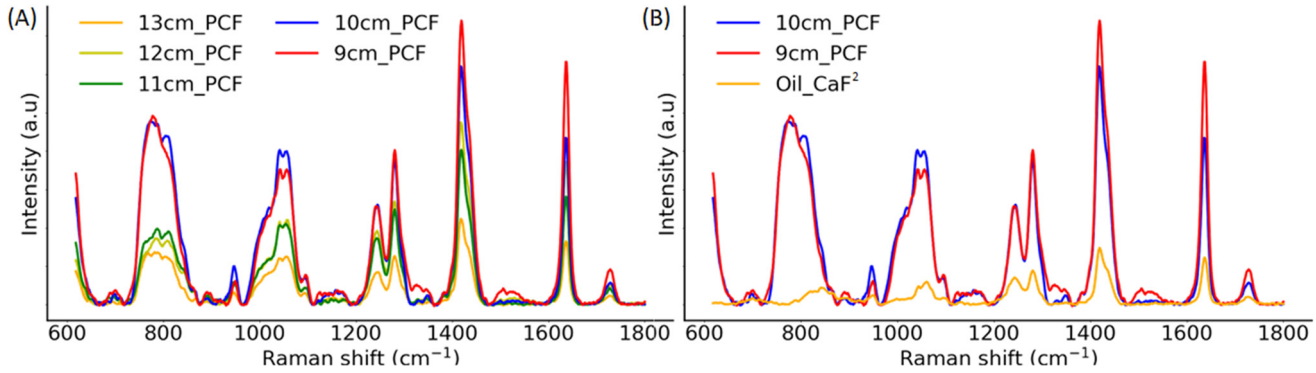


Fig. 6 Comparison of Raman spectra from (A) different length of PCF using vegetable oil as an analyte (B) different instrument configuration modes of the customized system using oil as analyte.

analyte. 9 cm and 10 cm PCF filled with oil provided the highest signals when analysed using transmission mode as compared the oil placed on a 2D CaF_2 substrate and analysed with Microscope in reflection mode. This improvement of signals in PCF compared to 2D substrates would be critical in detecting low concentration oral cancer markers in saliva.

The spectra in Fig. 7 represents Raman signals of 50 μM solution of Rhodamine B (Rhd B) in water filled in the PCF and observed in transmission mode *vs.* on a CaF_2 substrate and observed under the microscope in reflection mode. Rhd B peaks that were difficult to see on a 2D CaF_2 substrate under the microscope were significantly stronger when analysed through PCF due to increased interaction between the excitation laser and analyte over the length of 9 cm. The observed Rhd B peaks were in line with the previously published literature.[43]

3.3 Mode (C) Raman optical fibre probe for *in vivo* tissue analysis

Over the time, Raman spectra of various tissue has been observed. Most recognized Raman scatterers are blood, lipid, collagen, tyrosine elastin, myoglobin, carotenoids and nucleic acids *etc.*[17,44,45] Fig. 8 represents spectra from different tissue types in the oral cavity collected from a healthy participant with the customised fibre optic probe. The probe was gently placed on the area of interest to collect the Raman data from various tissues. 30 mW laser power was used to collect signal for a time of 4 s, to keep the energy per unit area within the permissible limit of the tissue.

The difference between different tissue types was clearly visible, especially signal from mandible bone was quite prominent at 960 cm^{-1} in the data collected from human gingiva while the lip Raman spectrum was dominated by lipid signals at 1080, 1302, 1440 and 1740 cm^{-1}.[46,47] Other subtle differences in these three different tissue types can be seen in terms of peak height and width at 1003 and 1660 cm^{-1} respectively. The molecular vibrational modes associated with these changes are tabulated in Table 1.

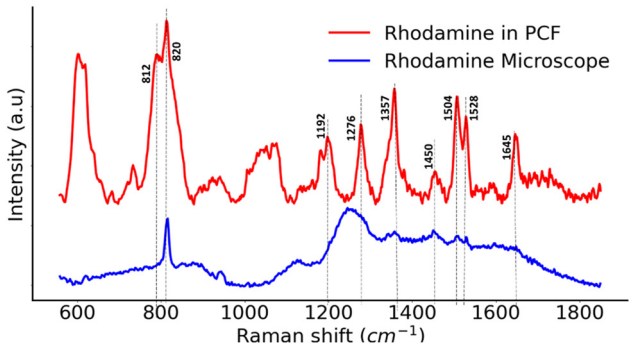


Fig. 7 Raman spectra comparison of Rhd B-filled in the PCF *vs.* on a 2 dimensional CaF_2 substrate.

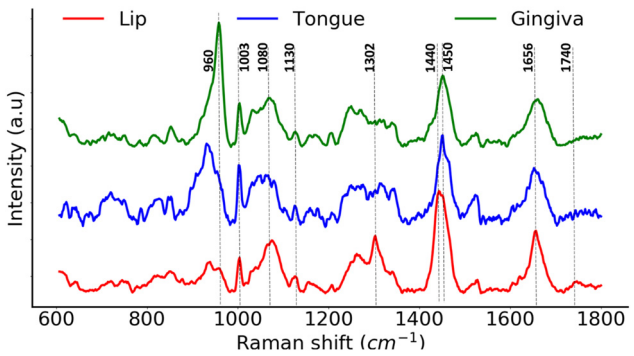


Fig. 8 Spectra from different tissue types in oral cavity collected from the new customised Raman system.

4. Conclusion

Design development and assembly of a versatile customised Raman system is reported in this study. This system can work in three different modes. The first mode was Raman microscope combined with conventional optical microscopy. This microscope has a flexible stage design for sample of different dimensions. We also demonstrated the benefit of the custom built liquid sample transmission stage over microscope in collecting signals from low concentration analytes. This PCF based transmission set-up enabled stronger Raman signals in a relatively small volume. Lastly, the flexible optical fibre based probe is suitable to acquire multiple (Raman, diffuse reflectance and fluorescence) spectroscopic techniques data from the same tissue location *in vivo*. In summary, this flexible Raman system is of great potential for the early diagnosis of oral cancer *via* initial screening using participants saliva followed by Raman spectral signatures guided biopsy of the lesions. The flexible and modular design can be of interest to Raman spectroscopy community and potentially in future primary care.

Author contributions

Siddra Maryam: methodology, system design and development, experiments, data analysis, writing, review and editing; Sanathana Konugolu Venkata Sekar: system design and development, review and editing; Daniyal Ghauri: data collection; Edward Fahy: data acquisition, recruiting participant, manuscript review; Marcelo Saito Nogueira: project planning, manuscript review; Huihui Lu: project planning, manuscript review; Flavien Beffara: PCF design and fabrication; Georges Humbert: PCF design and fabrication; Richeal Ni Riordain: supervision and managing clinical collaboration; Patrick Sheahan: supervision; Ray Burke: supervision and managing clinical collaboration; Kiang Wei Kho: system design and development, experiments, review and editing; Rekha Gautam: system design and development, experiment, data analysis, writing, review and editing; Stefan Andersson-Engels: funding, conceptualization, supervision, review and editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We greatly acknowledge fruitful discussions with Dr Linda Feeley. The authors will also like to thanks Jean Matias for his help in laser beam profile characterization and to Konstantin Grygoryev for his proof reading and improving the English in the manuscript. The authors are grateful for the financial support of Science Foundation Ireland through contract SFI/15/RP/2828.

References

1 P. Speight and P. Farthing, The pathology of oral cancer, *Br. Dent. J.*, 2018, **225**(9), 841–847.
2 D. Conway, M. Purkayastha and I. Chestnutt, The changing epidemiology of oral cancer: definitions, trends, and risk factors, *Br. Dent. J.*, 2018, **225**(9), 867–873.
3 C. N. E. Board, Oral and Oropharyngeal Cancer: Statistics, **https://www.cancer.net/cancer-types/oral-and-oropharyngeal-cancer/statistics#:~:text=The%20overall%205%2Dyear%20survival,HPV%20(see%20Risk%20Factors)**.
4 L. P. Kowalski, R. Bagietto, J. R. Lara, R. L. Santos, J. F. Silva Jr. and J. Magrin, Prognostic significance of the distribution of neck node metastasis from oral carcinoma, *Head Neck*, 2000, **22**(3), 207–214.
5 F. P. Koch, M. Kunkel, S. Biesterfeld and W. Wagner, Diagnostic efficiency of differentiating small cancerous and precancerous lesions using mucosal brush smears of the oral cavity—a prospective and blinded study, *Clin. Oral Investig.*, 2011, **15**(5), 763–769.
6 H. J. Byrne, I. Behl, G. Calado, O. Ibrahim, M. Toner, S. Galvin, C. M. Healy, S. Flint and F. M. Lyng, Biomedical applications of vibrational spectroscopy: Oral cancer diagnostics, *Spectrochim. Acta, Part A*, 2021, **252**, 119470.
7 M. H. Al-Azri, Delay in cancer diagnosis: causes and possible solutions, *Oman Med. J.*, 2016, **31**(5), 325.
8 L. Marcu, Fluorescence lifetime techniques in medical applications, *Ann. Biomed. Eng.*, 2012, **40**(2), 304–331.
9 G. Calado, I. Behl, A. Daniel, H. J. Byrne and F. M. Lyng, Raman spectroscopic analysis of saliva for the diagnosis of oral cancer: A systematic review, *Transl. Biophotonics*, 2019, **1**(1–2), e201900001.
10 Z. Li, M. J. Deen, S. Kumar and P. R. Selvaganapathy, Raman spectroscopy for in-line water quality monitoring—Instrumentation and potential, *Sensors*, 2014, **14**(9), 17275–17303.
11 P. Vandenabeele, *Practical Raman spectroscopy: an introduction*, Wiley Online Library, 2013.
12 A. Orlando, F. Franceschini, C. Muscas, S. Pidkova, M. Bartoli, M. Rovere and A. Tagliaferro, A comprehensive review on Raman spectroscopy applications, *Chemosensors*, 2021, **9**(9), 262.
13 S. J. Owonubi, N. M. Malima and N. Revaprasadu, Metal oxide-based nanocomposites as antimicrobial and biomedical agents, *Antibiotic materials in healthcare*, Elsevier, 2020, pp. 287–323.
14 A. Synetos and D. Tousoulis, Invasive imaging techniques, *Coronary Artery Disease: From Biology to Clinical Practice*, 2017, p. 359.
15 J. Zhang, Y. Fan, Y. Song and J. Xu, Accuracy of Raman spectroscopy for differentiating skin cancer from normal tissue, *Medicine*, 2018, **97**(34), e12022.
16 G. Donjuán-Loredo, R. Espinosa-Tanguma, M. León-Bejarano, R. Salgado-Delgado, F. J. González, E. Guevara and M. Ramírez-Elías, Raman spectroscopy for adipose tissue differentiation: a pilot study, *Optical Tomography and Spectroscopy*, Optical Society of America, 2020, p. JW3A.12.
17 S. Maryam, M. S. Nogueira, R. Gautam, S. Krishnamoorthy, S. K. Venkata Sekar, K. W. Kho, H. Lu, R. Ni Riordain, L. Feeley, P. Sheahan, R. Burke and S. Andersson-Engels, Label-Free Optical Spectroscopy for Early Detection of Oral Cancer, *Diagnostics*, 2022, **12**(12), 2896.
18 D. C. Harris and M. D. Bertolucci, *Symmetry and Spectroscopy*, Dover Publications, Inc., New York, 1978.
19 U. Y. Schaff and G. J. Sommer, Whole blood immunoassay based on centrifugal bead sedimentation, *Clin. Chem.*, 2011, **57**(5), 753–761.
20 U. S. Dinish, F. Beffara, G. Humbert, J.-L. Auguste and M. Olivo, Surface-enhanced Raman scattering-active photonic crystal fiber probe: Towards next generation liquid biopsy sensor with ultra high sensitivity, *J. Biophotonics*, 2019, **12**(11), e201900027.
21 A. V. Markin, N. E. Markina and I. Y. Goryacheva, Raman spectroscopy based analysis inside photonic-crystal fibers, *TrAC, Trends Anal. Chem.*, 2017, **88**, 185–197.
22 Y. Zhang, C. Shi, C. Gu, L. Seballos and J. Z. Zhang, Liquid core photonic crystal fiber sensor based on surface enhanced Raman scattering, *Appl. Phys. Lett.*, 2007, **90**(19), 193504.
23 F. Beffara, J. Perumal, A. Puteri Mahyuddin, M. Choolani, S. A. Khan, J. L. Auguste, S. Vedraine, G. Humbert, U. Dinish and M. Olivo, Development of highly reliable SERS-active photonic crystal fiber probe and its application in the detection of ovarian cancer biomarker in cyst fluid, *J. Biophotonics*, 2020, **13**(3), e201960120.
24 F. Beffara, G. Humbert, J.-L. Auguste, J. Perumal, U. Dinish and M. Olivo, Optimization and performance analysis of SERS-active suspended core photonic crystal fibers, *Opt. Express*, 2020, **28**(16), 23609–23619.
25 L. F. Santos, R. Wolthuis, S. Koljenović, R. M. Almeida and G. J. Puppels, Fiber-optic probes for in vivo Raman spectroscopy in the high-wavenumber region, *Anal. Chem.*, 2005, **77**(20), 6747–6752.
26 A. K. Barik, J. Lukose, R. Upadhya, M. V. Pai, V. Kartha and S. Chidangil, In vivo spectroscopy: optical fiber probes for

clinical applications, *Expert Rev. Med. Devices*, 2022, **19**(9), 657–675.

27 M. G. Shim, B. C. Wilson, E. Marple and M. Wach, Study of fiber-optic probes for in vivo medical Raman spectroscopy, *Appl. Spectrosc.*, 1999, **53**(6), 619–627.

28 M. Plesia, O. A. Stevens, G. R. Lloyd, C. A. Kendall, I. Coldicott, A. J. Kennerley, G. Miller, P. J. Shaw, R. J. Mead and J. C. Day, In vivo fiber optic Raman spectroscopy of muscle in preclinical models of amyotrophic lateral sclerosis and Duchenne muscular dystrophy, *ACS Chem. Neurosci.*, 2021, **12**(10), 1768–1776.

29 R. Gautam, D. Peoples, K. Jansen, M. O'Connor, G. Thomas, S. Vanga, I. J. Pence and A. Mahadevan-Jansen, Feature Selection and Rapid Characterization of Bloodstains on Different Substrates, *Appl. Spectrosc.*, 2020, **74**(10), 1238–1251.

30 R. Gautam, J. Y. Oh, R. P. Patel and R. A. Dluhy, Non-invasive analysis of stored red blood cells using diffuse resonance Raman spectroscopy, *Analyst*, 2018, **143**(24), 5950–5958.

31 J. M. Connolly, K. Davies, A. Kazakeviciute, A. M. Wheatley, P. Dockery, I. Keogh and M. Olivo, Non-invasive and label free detection of oral squamous cell carcinoma using saliva surface-enhanced Raman spectroscopy and multivariate analysis, *Nanomedicine*, 2016, **12**, 1593–1601.

32 S. Hu, J. A. Loo and D. T. Wong, Human body fluid proteome analysis, *Proteomics*, 2006, **6**(23), 6326–6353.

33 A. Alba-Bernal, R. Lavado-Valenzuela, M. E. Domínguez-Recio, B. Jiménez-Rodriguez, M. I. Queipo-Ortuño, E. Alba and I. Comino-Méndez, Challenges and achievements of liquid biopsy technologies employed in early breast cancer, *EBioMedicine*, 2020, **62**, 103100.

34 S. Maryam, M. S. Nogueira, S. K. Moorthy, S. Sekar, H. Lu, R. Gautam, R. Burke, S. Andersson-Engels, R. N. Riordain and P. Sheahan, Multi-configuration Raman spectrometer for early stage diagnosis of oral cancer, *Biomedical Vibrational Spectroscopy 2022: Advances in Research and Industry*, SPIE, 2022, pp. 20–26.

35 H. J. Koster, A. Guillen-Perez, J. S. Gomez-Diaz, M. Navas-Moreno, A. C. Birkeland and R. P. Carney, Fused Raman spectroscopic analysis of blood and saliva delivers high accuracy for head and neck cancer diagnostics, *Sci. Rep.*, 2022, **12**(1), 18464.

36 Z. Movasaghi, S. Rehman and I. U. Rehman, Raman Spectroscopy of Biological Tissues, *Appl. Spectrosc. Rev.*, 2007, **42**(5), 493–541.

37 S. Feng, S. Huang, D. Lin, G. Chen, Y. Xu, Y. Li, Z. Huang, J. Pan, R. Chen and H. Zeng, Surface-enhanced Raman spectroscopy of saliva proteins for the noninvasive differentiation of benign and malignant breast tumors, *Int. J. Nanomed.*, 2015, **10**, 537.

38 H. Ding, D. J. J. Hu, X. Yu, X. Liu, Y. Zhu and G. Wang, Review on All-Fiber Online Raman Sensor with Hollow Core Microstructured Optical Fiber, *Photonics*, 2022, 134.

39 B. G. Eleftheriades, E. E. Storey and A. S. Helmy, Label-Free Spontaneous Raman Sensing in Photonic Crystal Fibers with Nanomolar Sensitivity, *Adv. Opt. Mater.*, 2022, **10**(6), 2101117.

40 B. R. Wood, P. Caspers, G. J. Puppels, S. Pandiancherri and D. McNaughton, Resonance Raman spectroscopy of red blood cells using near-infrared laser excitation, *Anal. Bioanal. Chem.*, 2007, **387**(5), 1691–1703.

41 M. Cárdenas, T. Arnebrant, A. Rennie, G. Fragneto, R. K. Thomas and L. Lindh, Human saliva forms a complex film structure on alumina surfaces, *Biomacromolecules*, 2007, **8**(1), 65–69.

42 L. Coelho, D. Viegas, J. L. C. O. Santos and J. M. M. M. de Almeida, Monitoring of high refractive index edible oils using coated long period fiber grating sensors, *Optical Sensors 2015*, SPIE, 2015, pp. 248–254.

43 M. Liu, R. Xiang, Y. Lee, K. Otsuka, Y.-L. Ho, T. Inoue, S. Chiashi, J.-J. Delaunay and S. Maruyama, Fabrication, characterization, and high temperature surface enhanced Raman spectroscopic performance of SiO_2 coated silver particles, *Nanoscale*, 2018, **10**(12), 5449–5456.

44 J. Desroches, M. Jermyn, M. Pinto, F. Picot, M.-A. Tremblay, S. Obaid, E. Marple, K. Urmey, D. Trudel and G. Soulez, A new method using Raman spectroscopy for in vivo targeted brain cancer tissue biopsy, *Sci. Rep.*, 2018, **8**(1), 1792.

45 H. Zeng, J. Zhao, M. Short, D. I. Mclean, S. Lam, A. Mcwilliams and H. Lui, Raman spectroscopy for in vivo tissue analysis and diagnosis, from instrument development to clinical applications, *J. Innovative Opt. Health Sci.*, 2008, **1**(01), 95–106.

46 P. Meksiarun, B. B. Andriana, H. Matsuyoshi and H. Sato, Non-invasive quantitative analysis of specific fat accumulation in subcutaneous adipose tissues using Raman spectroscopy, *Sci. Rep.*, 2016, **6**(1), 1–8.

47 M. V. Schulmerich, J. H. Cole, K. A. Dooley, M. D. Morris, J. M. Kreider, S. A. Goldstein, S. Srinivasan and B. W. Pogue, Noninvasive Raman tomographic imaging of canine bone tissue, *J. Biomed. Opt.*, 2008, **13**(2), 020506.

Paper III

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 332 (2025) 125788

Contents lists available at ScienceDirect

Spectrochimica Acta Part A:
Molecular and Biomolecular Spectroscopy

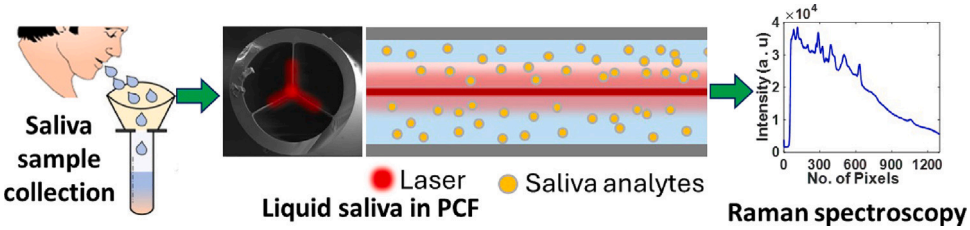
journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

Liquid saliva analysis using optofluidic photonic crystal fiber for detection of oral potentially malignant disorders

Siddra Maryam [a,b,*], Amine Benazza [d], Edward Fahy [c], Sanathana Konugolu Venkta Sekar [a,b], Dinish U.S. [e], Malini Olivo [e], Richeal Ni Riordain [c], Stefan Andersson-Engels [a,b], Georges Humbert [d], Katarzyna Komolibus [a], Rekha Gautam [a,**]

[a] Biophotonics@Tyndall National Institute, UCC, Dyke Parade, Cork, T12R5CP, Ireland
[b] School of Physics, University College Cork, College road, Cork, T12CY82, Ireland
[c] Cork University Dental School and Hospital, Wilton, Cork, T12E8YV, Ireland
[d] XLIM Research Institute, CNRS Limoges University, Albert Thomas, Limoges, 87000, France
[e] Translational Biophotonics Laboratory, A*STAR Skin Research Labs (A*SRL), Agency for Science, Technology and Research (A*STAR), 31 Biopolis Way, #07-01, Nanos, Singapore 138669, Republic of Singapore

GRAPHICAL ABSTRACT



HIGHLIGHTS

- PCF-based Raman spectroscopy for non-invasive liquid saliva analysis.
- Prolonged light-saliva interactions along the fiber to enhanced Raman signal.
- Paired Saliva and *in vivo* tissue analysis, showing consistent molecular trends.

ARTICLE INFO

Keywords:
Optofluidic photonic crystal fiber
PCF
Saliva
OPMD
Oral cancer
Raman spectroscopy

ABSTRACT

Oral potentially malignant disorders (OPMD) serve as early indicators of oral cancer. These conditions require reliable, non-invasive and real-time diagnostic methods for effective detection and screening. Saliva, as an easily accessible and non-invasive biofluid, has emerged as a promising tool to detect OPMD-associated biomarkers. This proof-of-concept study investigated the application of suspended-core optofluidic photonic crystal fibers (PCF) for Raman spectroscopy to distinguish between saliva from healthy controls and OPMD patients. This novel approach provides enhanced Raman signals through prolonged interactions between the excitation light and the saliva sample along the length of the PCF. Raman spectra of liquid saliva samples were collected from eleven participants, including six OPMD patients and five healthy controls. Notable spectral differences were identified at 1123 cm^{-1}, 1251 cm^{-1}, and 1454 cm^{-1}, which correspond to carbohydrates, proteins, and lipids, respectively. *In vivo* tissue measurements were recorded as a reference for comparative analysis from the same patients. Our findings suggest that PCF-based Raman spectroscopy holds promise as

* Corresponding author at: Biophotonics@Tyndall National Institute, UCC, Dyke Parade, Cork, T12R5CP, Ireland.
** Corresponding author.
E-mail addresses: siddra.maryam@tyndall.ie (S. Maryam), rekha.gautam@tyndall.ie (R. Gautam).

https://doi.org/10.1016/j.saa.2025.125788
Received 9 October 2024; Received in revised form 6 December 2024; Accepted 21 January 2025
Available online 28 January 2025

a non-invasive diagnostic platform for oral cancer screening by enabling the monitoring of subtle molecular changes in liquid saliva.

1. Introduction

A group of conditions characterized by abnormal changes in the oral tissue with a potential to develop into cancer are termed as oral potentially malignant disorders (OPMD). These changes can manifest as thickening of the oral mucosa lining, lumps in the oral cavity or in the neck, unusual red or white patches on the tongue, lips, gums, oral mucosa or on floor of mouth. Among many types of OPMD, oral submucous fibrosis, leukoplakia and actinic cheilitis are the most common [1,2] . Some OPMD can increase the chance of developing an invasive oral cancer by 35% [3,4]. However, effective lifestyle changes can slow down or even reverse the progression to cancer, especially in early-stage of dysplasia [5,6]. This is why early-stage diagnosis and treatment of OPMD are crucial to avoid their malignant transformation. Current gold standard procedure for diagnosing and staging of oral cancer and OPMD involves physical examination followed by biopsy and histopathology [7]. However, at present there is no reliable non-invasive screening procedure for disease monitoring or for the detection of potential recurrence after the initial diagnosis or treatment. Saliva, as an easily available and non-invasive biofluid, has emerged as a promising tool to detect OPMD associated biomarkers [8]. One significant advantage of using saliva for oral cancer screening is its direct interaction with lesions, which provides specific biomarkers indicative of the presence and progression of OPMD, thereby mimicking the changes that occur in those lesions [9,10].

Optical non-destructive diagnostic tools, such as Raman spectroscopy, hold significant promise for analyzing saliva in the context of oral cancer screening [11–13]. Although the current Raman measurement method is sensitive to the composition, structure, and conformation of molecules, it struggles to capture molecular information effectively due to the low concentration of analytes present in liquid saliva [14,15]. While the process of drying saliva samples can enhance the concentration of analytes for improved detection of Raman signals [11,13], this process unfortunately leads to molecular structural changes, potentially affecting stability and integrity of biomarkers. Moreover, changes in the structure and concentration of the analytes due to the drying process can lead to inconsistent results and compromise the reliability of the diagnostic assays [16,17]. For this reason, analyzing saliva in its liquid form will facilitate real-time saliva screening, eliminating the time-consuming and potentially destructive drying process [18]. Additionally, identifying native state molecular markers may enhance our understanding of the primary molecular mechanisms responsible for disease development and progression. Therefore, balancing the need for concentrated samples with the preservation of conformational integrity could lead to more accurate diagnostic outcomes.

To address the challenge of detecting low concentration analytes in liquid saliva, we propose a novel method of utilizing optofluidic photonic crystal fibers (PCF) as a sample holder. PCF have recently attracted significant attention as a potential platform for chemical and biological sensing due to their unique optical properties, which enhance light-sample interaction while at the same time reducing the required sample volume [19]. In this study, we used custom optofluidic suspended-core PCF with a small triangular solid core located at the junction of three hollow air channels [20,21], which can be filled with the saliva sample. In such configuration, the excitation light is tightly guided within the core, while also enabling evanescent field to extend into the surrounding regions where the liquid sample is present [22]. This design enhances light-saliva interactions and allows for improved detection sensitivity through prolonged light-saliva interaction length. To the best of our knowledge, this is the first report on the analysis of liquid saliva conducted without the addition of interfering substrates, such as nanoparticles, for signal enhancement. Standard surface-enhanced Raman spectroscopy (SERS) can introduce additional signals from nanoparticles or other impurities, which may obscure weaker analyte signals [23,24]. This effect is also common in nano roughened metallic planar substrates, especially if the nanostructures are formed by random process. Some commercial SERS substrates also exhibit background signals due to certain impurities formed during fabrication. Additionally, label-free SERS results can be biased towards molecules (e.g. uric acid, ergothioneine) with a higher affinity for the substrates [25], while variations in Raman bands make spectral interpretation difficult for SERS saliva spectra [23]. In contrast, the PCF-based approach presented here enables the acquisition of Raman signals from native saliva while preserving molecular structure and achieving a Raman bandwidth comparable to that of conventional Raman spectroscopy, without the complexities and biases associated with SERS [24,26].

2. System design and development

To facilitate precise laser focusing and PCF alignment, we have designed a horizontal Raman stage setup shown in Fig. 1 [27]. This design also helps to avoid the effect of gravity on the saliva filled inside the PCF. The system consists of two channels: a guiding channel and a Raman channel. Guiding channel consists of a green LED (Thorlabs: M530L4) and an optical camera (ThorLabs: CS165MU) for locating the PCF core. The green light is first passed through a lens L1 (Thorlabs: AC254-030-B), from where a parallel beam is directed by beam-splitters BS1 and BS2 (ThorLabs: CM1-BP108) to a dichroic mirror DM1 (Semrock LPD02-785RU). DM1 reflects the light towards L5 objective lens (Nikon 50x /0.60), focusing it onto the PCF core. The fiber is mounted on an XYZ translation stage for precise alignment, while the PCF core can be observed on the optical camera (ThorLabs: CS165MU) through L2 lens (ThorLabs: AC254-100-AB). The second channel guides the Raman excitation laser to the PCF core and collects the backscattered Raman signal from saliva sample. For Raman illumination a single mode 785 nm laser (IPS: 1804B000 FATBOY) is used. The laser beam is collimated by L3 lens (Thorlabs: AC127-025-AB-ML) and filtered by a band pass filter BP_{785} (Semrock LD01-785/10-25). The incident laser light then passes through the BS2 and is directed towards the objective by DM1. The system is aligned in such a way that the green light and the laser follow the same path from BS2 onward. Once the PCF core is in focus for the green light, it is nearly aligned for the Raman laser as well, however, fine-tuning of the z-axis is required due to the wavelength-dependent (530 nm vs 785 nm) differences in the focal point. The Raman laser was aligned to the suspended-core along the z-axis using dim green light for visualization. Once the laser was aligned with the PCF core, the green light was turned off to collect Raman data. The laser beam focused on the PCF core can also be seen on the optical camera, when the green light is turned off. Backscattered Raman signals from the saliva-filled PCF, are collected by the objective and after passing through DM1, they are filtered for any remaining excitation light by a long pass filter (Semrock LP02-785RU-25). The filtered Raman signal is then refocused by a L4 lens (Thorlabs: F810SMA-850) on a low-OH circular to linear optical fiber bundle (Thorlabs BFL200LS02). The linear side of the fiber is connected to the 50 μm entrance slit of

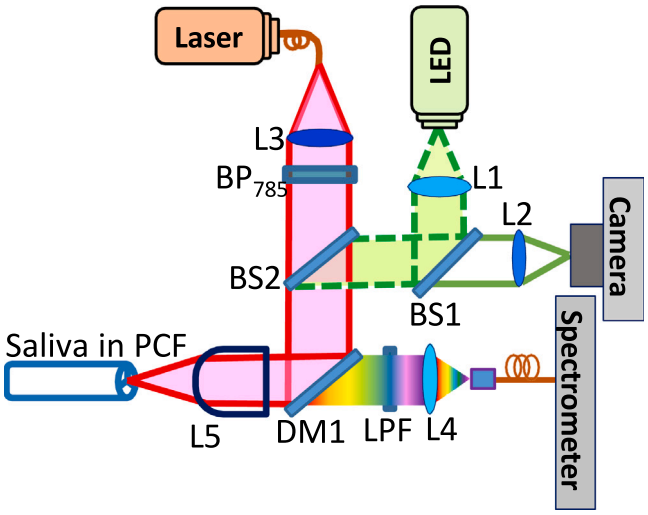


Fig. 1. Schematic of customized PCF stage for liquid sample analysis, consisting of two input arms (i) green LED guiding channel; (ii) Raman excitation channel with a backscattered Raman detection path.

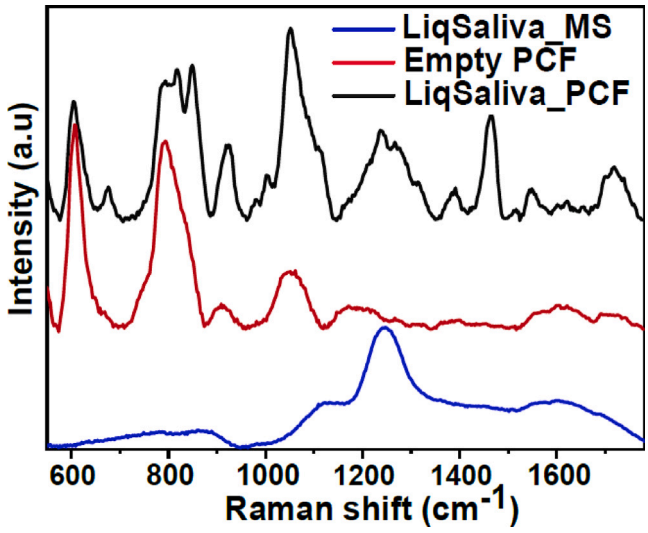


Fig. 2. Comparison of average Raman spectra of liquid saliva sample on a 2D aluminum substrate under a Raman microscope (MS) (blue), empty PCF (red) and liquid saliva in PCF (black). Important Raman features of saliva are visible in PCF data.

the Raman spectrometer. The Raman spectrometer consists of a filter chamber, having a long pass filter (LPF) (Semrock LP02-785RU-25), a spectrometer (Teledyne: LS785, f/2) and a Charge-Coupled Device (CCD) camera (Teledyne: Blaze 400HR). To collect the light from a linear slit and to improve signal-to-noise ratio, the individual pixels on the CCD are binned together vertically to generate a Raman spectrum.

3. Experimental design

This study was approved by Cork Research Ethics Committee (CREC) of University College Cork (UCC) according to CREC application reference: ECM 4 (y) 13/4/2021 & ECM 3 (aaaa) 12/09/2023 and include healthy volunteers as well as patients identified with OPMD. This study was conducted in compliance with the ethical standards established by CREC and the 1964 Helsinki declaration and its subsequent amendments. In this proof-of-concept study, six OPMD patients that have been recently biopsied and five healthy volunteers, visiting Cork University Dental School and Hospital, were enrolled. Informed consent was obtained from all the volunteers prior to the measurements. The participants were advised to fast for at least 1 h and rinse thier mouth with water for 1 minute before providing the saliva samples. Unstimulated saliva was collected by gentle drooling process and then *in vivo* tissue data was recorded. Collected saliva samples were centrifuged to remove any debris or solid particles and stored at −80 °C immediately.

3.1. Sample preparation

For further analysis, saliva samples were thawed and filtered using a 300 kDa membrane filter (Millipore MPE300025) to eliminate contamination from food particles while preserving the majority of proteins and other biomolecules [28]. In this experiment, PCFs fabricated at XLIM Research Institute - Limoges, France were used [21]. The external diameter of the used PCF was 200 μm for easy handling. Three different core sizes – 2.5 μm, 3.0 μm and 3.5 μm – were tested to determine the optimum core size for further experimentation. For each PCF core size, three fibers were prepared using saliva sample from one of the healthy participant resulting in a total of 9 PCFs being examined. To optimize the PCF length, a 3.5 μm core was used with lengths ranging from 4 cm to 16 cm with the saliva from the same healthy participant. To prepare PCF for saliva analysis, PCF fibers were cleaved from both ends and glued in to 25G needles. These fibers were then washed for a minute with acetone. The left-over acetone was removed and fibers were dried by pushing air inside the PCF with a new syringe. Later, filtered saliva sample was filled inside a syringe and injected in the PCF at a constant speed of 50 μL/min with a syringe pump. The polymer coating was then carefully stripped from one end and fiber was cleaved again to avoid any interference effect from the polymer and from the overflowing saliva droplet [20]. The prepared PCF fiber was then placed on the XYZ stage and brought to focus with the help of green light for data acquisition.

3.2. Data acquisition

The data was recorded from each prepared fiber for 4 s with 22 mW laser power using the LightField software (Teledyne Vision Solution). Each spectrum was averaged over 5 accumulations for better signal-to-noise ratio. *In vivo* tissue data was also recorded for 3.87 s with 30 mW power with a customized Raman systems and a multimodal probe described elsewhere [29]. Collected data was calibrated along x-axis and y-axis using Neon-Argon and white light lamps from IntelliCal spectral calibration system (Teledyne) and by standard spectra of acetaminophen and naphthalene. Data was corrected for the dark background and baseline using asymmetric least square method [29] ($\lambda = 50 and$ p = 0.00005). Finally, data was normalized with standard normal variate analysis and Savitzky–Golay filter (order = 2, window = 7) was used for noise filtering. Since the patients' saliva data was not normally distributed, the Kruskal–Wallis test used as a non-parametric method for statistical evaluation. A p-value of <0.05 was considered statistically significant.

4. Results and discussion

Fig. 2 shows a comparison of mean Raman spectra of liquid saliva sample on a two-dimensional (2D) aluminum substrate under a typical Raman microscope, an empty PCF and filtered liquid saliva in a PCF analyzed with the customized Raman system. This figure clearly illustrates the spectral differences between the two measurement methods. Raman signal of liquid saliva on a 2D substrate is predominantly featureless with the only prominent peak observed around 1246 cm^{-1}. This feature is generally attributed to the background Raman signal originating from the optics of the system, which becomes noticeable when the signals from the samples under investigation are weak [29].

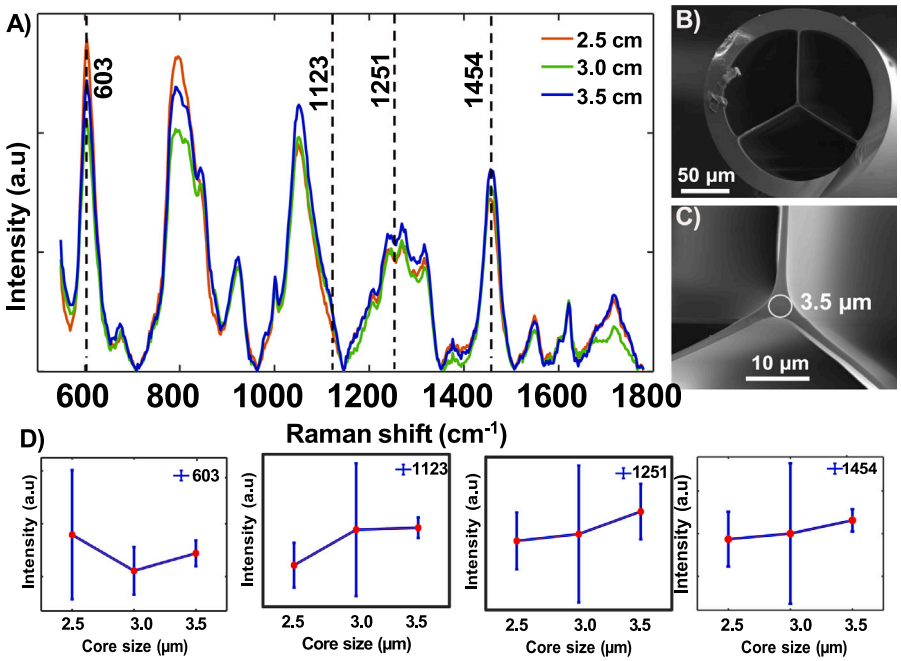


Fig. 3. (A) Comparison of Raman signal of PCFs with 2.5, 3.0 and 3.5 μm core size. Each core size was tested with three saliva samples derived from the same participant's saliva. (B) SEM image of 3.5 μm PCF. (C) Magnified SEM image of 3.5 μm core PCF. (D) Comparison of Raman intensity of 2.5, 3.0 and 3.5 μm PCF at 603, 1123, 1251 and 1454 cm^{-1}. Error bars represent standard deviation from mean intensity for each core size from multiple measurements across three samples.

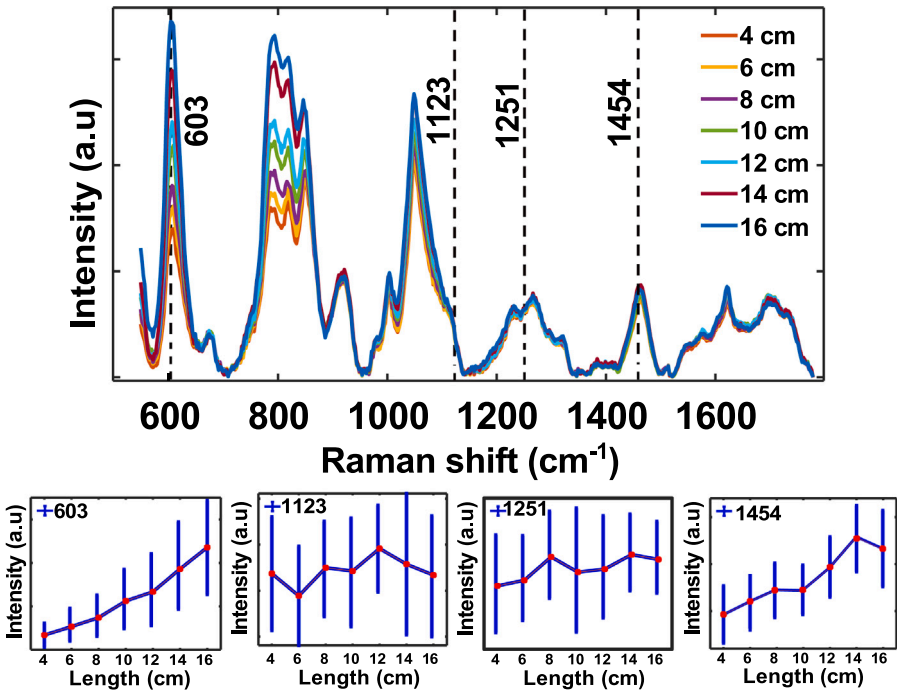


Fig. 4. Comparison of average Raman intensity for different fiber lengths from 4 cm to 16 cm (top). Each fiber length was tested with three saliva samples derived from the same participant's saliva. Graphs at the bottom of the figure represent average intensity at 603, 1123, 1251 and 1454 cm^{-1} plotted against the fiber length. Error bars show standard deviation of mean intensity for each fiber length, from multiple measurements across three samples.

On the other hand, Raman spectrum of saliva analyzed using PCF revealed distinct spectral features likely due to enhanced signals from the prolonged interaction of the excitation light with saliva along the entire length of the PCF. Some of these pronounced Raman spectral features and their assignments are shown in Table 1. The PCF data shows important Raman spectral features of human filtered saliva, as observed by Carlomagno et al. [28]. Two additional Raman bands at 603 cm^{-1} and 796 cm^{-1} are associated with the silica background of the PCF [30], as these signals are also clearly evident in the spectrum of an empty PCF.

These results confirm the potential of PCF for analyzing liquid saliva samples and enhancing signal quality. To further optimize the signal strength from liquid saliva, we conducted experiments using PCF fibers of various core sizes and lengths. The average Raman signal for each fiber core is presented in Fig. 3A, while Fig. 3B shows a scanning electron microscope (SEM) image of the 3.5 μm PCF with its core

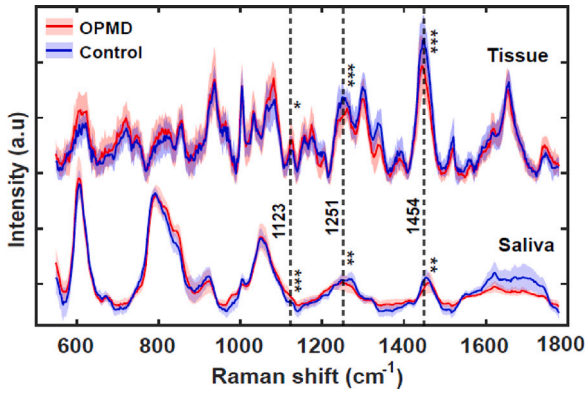


Fig. 5. Average Raman signal of control (four participants) and OPMD group (six participants) with shaded area representing standard deviation of tissue (top) and saliva in PCF (bottom). The labeled peaks represent the features that are significantly different in both groups (* represents $p < 0.05$, ** represents $p < 0.01$ and *** represents $p < 0.001$).

Table 1
Peak assignment of prominent Raman peaks in liquid saliva filled inside PCF [30–36].

Raman shift (cm^{-1})	Peak assignment
603	Si–O–Si bending silica
796	Si–O bending silica
855	C–O–C carbohydrates, Tyrosine
1003	Symmetric breathing mode of phenylalanine
1050	C–C stretching lipids, Si–O silica
1123	C–O carbohydrates
1251	Amide III protein
1454	C–H bending/deformation in lipids or protein
1548	C_{β}-C_{β} hemoglobin, Tryptophan
1660	Amide I protein

highlighted in Fig. 3C. Fig. 3D displays the Raman intensities of three major components of the saliva (carbohydrates, proteins and lipids) at 1123, 1251 and 1454 cm^{-1} along with silica band at 603 cm^{-1} for all the three PCFs (2.5, 3.0 and 3.5 μm). The error bars represent standard deviation of multiple measurements from three fibers for each core size, prepared in a same way, with the saliva sample from the same healthy participant. This comparison highlights the increase in Raman signal strength with PCF core size, indicating that a core size of 3.5 μm is the best of the tested diameters for saliva sample analysis with this geometry. The 3.5 μm core size also exhibits a comparatively lower standard deviation for the analyzed Raman features, which may be attributed to the increased consistency of focusing locations and better coupling efficiency as described by F. Beffara et al. [21]. Based on this analysis, we decided to proceed with the 3.5 μm core size PCF for further analysis.

Subsequently, to determine an optimized length of 3.5 μm core size PCF for saliva analysis, different fiber lengths ranging from 4 cm to 16 cm were evaluated. Fig. 4 shows average Raman intensity for each prepared fiber length. Error bars represents standard deviation of mean intensity for each fiber length, from multiple measurements across three samples. The results indicate that the silica background at 603 cm^{-1} increases with the fiber length, and the saliva signals at 1123, 1251 and 1454 cm^{-1} increase with fiber length up to 12 cm, 8 cm and 14 cm, respectively, and then level off (Fig. 4). Taking this into account, we decided to utilize 9–10 cm length for further experimentation to strike a balance between minimizing interfering silica signals and ensuring robust saliva analysis with satisfactory signal intensity across the entire spectrum. This intermediate fiber length reduces silica signals while providing sufficient interaction volume for enhanced Raman signal detection. Additionally, this length is also suitable for handling and mechanical stability ensuring consistent and reproducible results. Therefore, selection of 9–10 cm PCF length for further analysis is a practical and strategic choice that balances experimental objectives with operational considerations.

Our primary objective is to harness the capabilities of PCF for liquid saliva analysis, providing insights into both molecular composition and structural changes, which are critical for the early detection of OPMD. To achieve this, it is essential to identify spectral markers specifically related to OPMD. Since saliva mimics the composition of oral tissue, it presents an advantage as a potential screening tool for OPMD. However, saliva composition can also be influenced by other health issues [28]. To specifically target the spectral markers related to OPMD, we incorporated *in vivo* tissue measurements from the same participants for comparison. This approach provides a robust comparative analysis that validates the biochemical changes observed in saliva due to OPMD. Fig. 5 presents the preliminary Raman data from the remaining four control and six OPMD participants, with tissue data shown at the top and saliva data at the bottom for better comparison.

Noticeably, three spectral markers related to carbohydrates, proteins and lipid were identified, showing significant differences ($p < 0.05$) between two groups. The corresponding statistical analysis plots for these spectral markers are shown in Fig. 6. We observed that the features representing carbohydrates at 855 and 1123 cm^{-1} are more pronounced in OPMD spectra [36]. In OPMD patients, an increase in carbohydrate levels, which are critical components of cell membranes and play significant roles in cellular processes such as adhesion and signal transduction, has been observed previously [37,38]. The enhanced expression of these saccharide molecules on cancer cell membranes facilitates disease progression and their entry into the bloodstream and saliva [37,38].

Amide III band at 1251 cm^{-1} shows decreasing trend in OPMD saliva compared to the control group, indicating decrease in ordered protein. Additionally, a decrease in the peak at 1454 cm^{-1}, which is generally associated with C-H bending/deformation in proteins and lipids, is observed. Similar changes including decrease in 1454 cm^{-1} and protein bands have been observed previously in Raman spectroscopy analysis of oral lesions [39–42]. Increased protein breakdown and protein denaturation has been seen in OPMD patients, causing significant reduction in proteins levels in biofluids [43–45]. Further, oxidation of primary biomolecules, such as proteins and lipids, plays a significant role in the progression of cancer and inflammation, as well as in the proliferation of cells that utilize existing stores to maintain the structural and functional integrity of cell membranes [41,43,46]. However, levels of proteins and lipids in saliva are also influenced by the age, tobacco use and alcohol consumption [43,44]. Yet, further research is required to identify underlying reasons contributing to reduction in salivary lipid and protein in OPMD patients [44].

In this study, we presented preliminary results indicating significant differences between OPMD and control group, establishing a proof-of-concept for utilizing liquid saliva in non-invasive OPMD screening. While many studies have analyzed saliva using Raman spectroscopy, most have either utilized dried or lyophilized saliva or incorporated SERS substrates [23,47,48]. In cases where liquid saliva is analyzed, these studies often struggle to achieve a favorable signal-to-noise ratio [48]. Additionally, it is important to note that weak Raman signals are affected by etaloning when excited with NIR laser light in combination with thin back-illuminated CCD detectors [49]. This effect is more prominent when measuring weak Raman signals in the presence of a large fluorescence background. Such distortions can mimic or obscure Raman peaks, leading to misinterpretation of the data [30]. The present study seeks to address these challenges and improve the reliability of Raman signals from native saliva. However, it suffers from several limitations, including interference from silica signals, variability in excitation laser coupling, and a small sample size. The silica background complicates the analysis of saliva signals below 1000 cm^{-1}. To reduce the silica background, we aim to increase the fraction of evanescent light and reduce fiber length or develop a novel fiber design with enhanced light-analyte interactions to achieve stronger Raman

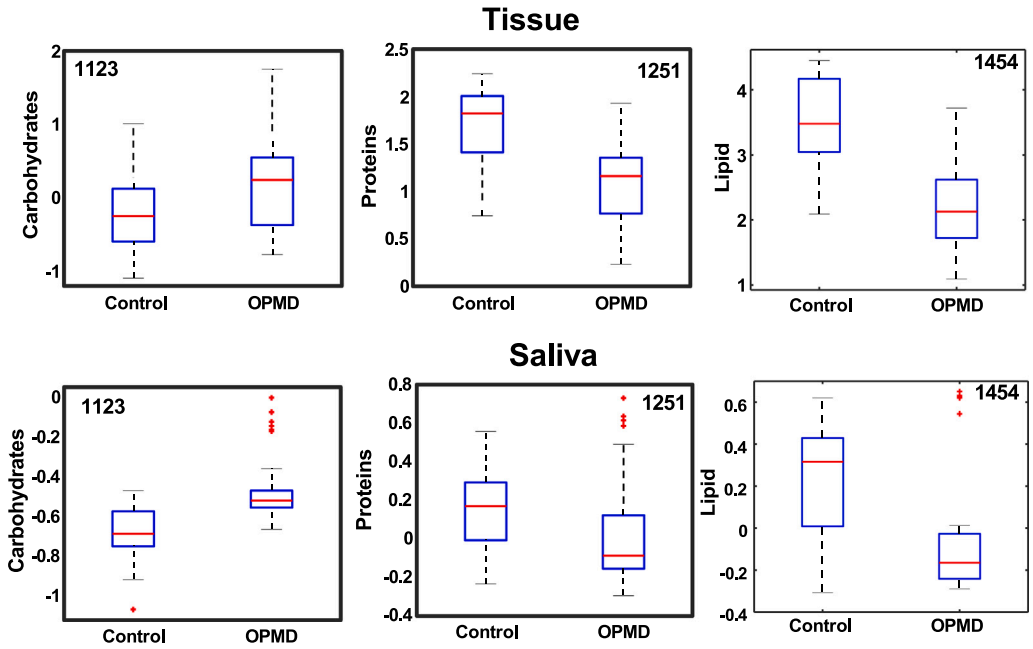


Fig. 6. Box plot analysis of Raman peaks indicating significant changes ($p < 0.05$,) in carbohydrates, proteins and lipid in tissue (top) and saliva (bottom).

signals. Further data decomposition, feature selection methods and implementing other data analysis techniques could help distinguish saliva signals from the less variable silica signals. We also plan to incorporate a second camera to visualize light transmission through the PCF, which may provide a quantitative method to analyze light coupling efficiency. As this manuscript presents a proof-of-concept study with preliminary data from only 10 participants, this sample size is insufficient for robust machine learning analysis. Future work will also involve a larger patient cohort, and with an expanded dataset, incorporating more variables and sophisticated algorithms could enhance diagnostic accuracy.

5. Conclusion

We have demonstrated the potential of PCF based-optofluidic Raman spectroscopy probe as a non-invasive tool for early diagnostics of OPMD by saliva analysis. This novel method relies on saliva in its liquid state, mitigating the disadvantages of dry saliva and SERS-based measurements, which can potentially enable real-time diagnosis. Using the best of the tested PCF with core size of 3.5 μm and length of 10 cm, we identified carbohydrates, proteins and lipids as three biomarkers which can differentiate between OPMD and healthy groups. These results were in agreement with data obtained from *in vivo* tissue Raman spectroscopy from the same participants confirming the changes observed in saliva are due to OPMD. This research underscores the importance of advanced diagnostic approaches that offer increased patient comfort, real-time analysis and improved methods for oral cancer monitoring. This study serves as a proof-of-concept demonstration for non-invasive saliva screening of OPMD. However, the proposed technique could be a promising tool for several clinical applications where biofluids, such as plasma or cerebrospinal fluid, hold important information about the disease state.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant number SFI/15/RP/2828 and SFI/22/RP-2TF/10293. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The authors also acknowledge Institutional funding support from the French National Research Agency (programs ANR-18-EURE-0017 TACTIC, ANR-21-CE42-0027-01), the Singaporean National Research Foundation (program NRF2021-NRF-ANR002), the International Research Project “FiberMed” (CNRS, A*STAR, Univ. Limoges) and the Nouvelle Aquitaine (program FibOsome).

Data availability

Data will be made available on request.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used AI assisted tool in order to correct grammar. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Fernanda Weber Mello, Andressa Fernanda Paza Miguel, Kamile Leonardi Dutra, André Luís Porporatti, Saman Warnakulasuriya, Eliete Neves Silva Guerra, Elena Riet Correa Rivero, Prevalence of oral potentially malignant disorders: a systematic review and meta-analysis, J. Oral Pathol. Med. 47 (7) (2018) 633–640, http://dx.doi.org/10.1111/jop.12726.

[2] Yunhan Tan, Zhihan Wang, Mengtong Xu, Bowen Li, Zhao Huang, Siyuan Qin, Edouard C Nice, Jing Tang, Canhua Huang, Oral squamous cell carcinomas: state of the field and emerging directions, Int. J. Oral Sci. 15 (1) (2023) 44, http://dx.doi.org/10.1038/s41368-023-00249-w.

[3] Hisham M Mehanna, Tim Rattay, Joel Smith, Christopher C McConkey, Treatment and follow-up of oral dysplasia—a systematic review and meta-analysis, Head Neck: J. Sci. Spec. Head Neck 31 (12) (2009) 1600–1609, http://dx.doi.org/10.1002/hed.21131.

[4] S Warnakulasuriya, T Kovacevic, P Madden, VH Coupland, M Sperandio, E Odell, H Møller, Factors predicting malignant transformation in oral potentially malignant disorders among patients accrued over a 10-year period in south east England, J. Oral Pathol. Med. 40 (9) (2011) 677–683, http://dx.doi.org/10.1111/j.1600-0714.2011.01054.x.

[5] Pasquale Marino, Mariangela Mininni, Giovanni Deiana, Graziella Marino, Rosa Divella, Ilaria Bochicchio, Alda Giuliano, Stefania Lapadula, Alessandro Rocco Lettini, Francesca Sanseverino, Healthy lifestyle and cancer risk: modifiable risk factors to prevent cancer, Nutrients 16 (6) (2024) 800, http://dx.doi.org/10.3390/nu16060800.

[6] R. James Barnard, Prevention of cancer through lifestyle changes, Evidence-Based Complement. Altern. Med. 1 (3) (2004) 233.

[7] Kalen R Dionne, Saman Warnakulasuriya, Rosnah Binti Zain, Sok Ching Cheong, Potentially malignant disorders of the oral cavity: Current practice and future directions in the clinic and laboratory, Int. J. Cancer 136 (3) (2015) 503–515, http://dx.doi.org/10.1002/ijc.28754.

[8] Paola Dongiovanni, Marica Meroni, Sara Casati, Riccardo Goldoni, Douglas Vieira Thomaz, Nermin Seda Kehr, Daniela Galimberti, Massimo Del Fabbro, Gianluca M Tartaglia, Salivary biomarkers: novel noninvasive tools to diagnose chronic inflammation, Int. J. Oral Sci. 15 (1) (2023) 27, http://dx.doi.org/10.1038/s41368-023-00231-6.

[9] Livia Barenghi, Francesco Spadari, Aldo Bruno Giannì, Matteo Vidali, Ferruccio Ceriotti, Saliva: Challenges, possibilities, and limits of the diagnostic use Part 2–pre-analytical and analytical aspects, Biochim. Clin. 47 (1) (2023) https://biochimicaclinica.it/en/saliva-challenges-possibilities-and-limits-of-the-diagnostic-use-part-2-pre-analytical-and-analytical-aspects/.

[10] Caizhi Liao, Xiaofeng Chen, Ying Fu, Salivary analysis: An emerging paradigm for non-invasive healthcare diagnosis and monitoring, Interdiscip. Med. 1 (3) (2023) e20230009, http://dx.doi.org/10.1002/INMD.20230009.

[11] Genecy Calado, Isha Behl, Hugh J. Byrne, Fiona M. Lyng, Raman spectroscopic characterisation of non stimulated and stimulated human whole saliva, Clin. Spectrosc. 3 (2021) 100010, http://dx.doi.org/10.1016/j.clispe.2021.100010.

[12] Kiang Wei Kho, Olivo Malini, Ze Xiang Shen, Khee Chee Soo, Surface enhanced Raman spectroscopic study (SERS)of saliva in the early detection of oral cancer, in: Optical Diagnostics and Sensing V, vol. 5702, SPIE, 2005, pp. 84–91, http://dx.doi.org/10.1117/12.590142.

[13] Bhavik Vyas, Ana Khatiashvili, Lisa Galati, Khoa Ngo, Neil Gildener-Leapman, Melinda Larsen, Igor K Lednev, Raman hyperspectroscopy of saliva and machine learning for Sjögren's disease diagnostics, Sci. Rep. 14 (1) (2024) 11135, http://dx.doi.org/10.1038/s41598-024-59850-6.

[14] Siddra Maryam, Marcelo Saito Nogueira, Rekha Gautam, Shree Krishnamoorthy, Sanathana Konugolu Venkata Sekar, Kiang Wei Kho, Huihui Lu, Richeal Ni Riordain, Linda Feeley, Patrick Sheahan, et al., Label-free optical spectroscopy for early detection of oral cancer, Diagnostics 12 (12) (2022) 2896, http://dx.doi.org/10.3390/diagnostics12122896.

[15] Sanchita Sil, Rekha Gautam, Siva Umapathy, Applications of Raman and infrared microscopy to materials and biology, in: Molecular and Laser Spectroscopy: Advances and Applications, Elsevier Inc., 2018, pp. 117–146.

[16] Tarun Tejasvi Mutukuri, Nathan E Wilson, Lynne S Taylor, Elizabeth M Topp, Qi Tony Zhou, Effects of drying method and excipient on the structure and physical stability of protein solids: Freeze drying vs. spray freeze drying, Int. J. Pharm. 594 (2021) 120169, http://dx.doi.org/10.1016/j.ijpharm.2020.120169.

[17] Tuan Phan-Xuan, Ekaterina Bogdanova, Anna Millqvist Fureby, Jonas Fransson, Ann E Terry, Vitaly Kocherbitov, Hydration-induced structural changes in the solid state of protein: A SAXS/WAXS study on lysozyme, Mol. Pharm. 17 (9) (2020) 3246–3258, http://dx.doi.org/10.1021/acs.molpharmaceut.0c00351.

[18] Yu Han, Xi-Ling Li, Minghui Zhang, Jing Wang, Su Zeng, Jun Zhe Min, Potential use of a dried saliva spot (DSS) in therapeutic drug monitoring and disease diagnosis, J. Pharm. Anal. 12 (6) (2022) 815–823.

[19] Ana M Cubillas, Sarah Unterkofler, Tijmen G Euser, Bastian JM Etzold, Anita C Jones, Peter J Sadler, Peter Wasserscheid, Philip St J Russell, Photonic crystal fibres for chemical sensing and photochemistry, Chem. Soc. Rev. 42 (22) (2013) 8629–8648.

[20] US Dinish, Flavien Beffara, Georges Humbert, Jean-Louis Auguste, Malini Olivo, Surface-enhanced Raman scattering-active photonic crystal fiber probe: towards next generation liquid biopsy sensor with ultra high sensitivity, J. Biophotonics 12 (11) (2019) e201900027, http://dx.doi.org/10.1002/jbio.201900027.

[21] Flavien Beffara, Jayakumar Perumal, Aniza Puteri Mahyuddin, Mahesh Choolani, Saif A Khan, Jean-Louis Auguste, Sylvain Vedraine, Georges Humbert, US Dinish, Malini Olivo, Development of highly reliable SERS-active photonic crystal fiber probe and its application in the detection of ovarian cancer biomarker in cyst fluid, J. Biophotonics 13 (3) (2020) e201960120, http://dx.doi.org/10.1002/jbio.201960120.

[22] Alexey V. Markin, Natalia E. Markina, Irina Yu Goryacheva, Raman spectroscopy based analysis inside photonic-crystal fibers, TRAC Trends Anal. Chem. 88 (2017) 185–197, http://dx.doi.org/10.1016/j.trac.2017.01.003.

[23] Alois Bonifacio, Silvia Cervo, Valter Sergo, Label-free surface-enhanced Raman spectroscopy of biofluids: fundamental aspects and diagnostic applications, Anal. Bioanal. Chem. 407 (2015) 8265–8277, https://link.springer.com/article/10.1007/s00216-015-8697-z.

[24] Rekha Gautam, Deepika Chaturvedi, Sanchita Sil, Nikki Kuhar, Saumya Singh, Siva Umapathy, Characterization of aggregating agents towards sensitive optical detection of tryptophan using lab-on-a-chip, in: Photonics, vol. 9, MDPI, 2022, p. 648.

[25] Stefano Fornasaro, Valter Sergo, Alois Bonifacio, The key role of ergothioneine in label-free surface-enhanced Raman scattering spectra of biofluids: a retrospective re-assessment of the literature, FEBS Lett. 596 (10) (2022) 1348–1355, http://dx.doi.org/10.1002/1873-3468.14312.

[26] Judith Langer, Dorleta Jimenez de Aberasturi, Javier Aizpurua, Ramon A Alvarez-Puebla, Baptiste Auguié, Jeremy J Baumberg, Guillermo C Bazan, Steven EJ Bell, Anja Boisen, Alexandre G Brolo, et al., Present and future of surface-enhanced Raman scattering, ACS Nano 14 (1) (2019) 28–117.

[27] Amine Benazza, Flavien Beffara, Jean-Louis Auguste, Malini Olivo, US Dinish, Georges Humbert, Reliable and easy-to-use SERS spectroscopy probe using a tapered opto-fluidic photonic crystal fiber, Opt. Express 32 (3) (2024) 3440–3450, http://dx.doi.org/10.1364/OE.501911.

[28] C Carlomagno, PI Banfi, A Gualerzi, S Picciolini, Eleonora Volpato, M Meloni, A Lax, E Colombo, N Ticozzi, F Verde, et al., Human salivary Raman fingerprint as biomarker for the diagnosis of amyotrophic lateral sclerosis, Sci. Rep. 10 (1) (2020) 10175, http://dx.doi.org/10.1038/s41598-020-67138-8.

[29] Siddra Maryam, Sanathana Konugolu Venkata Sekar, M Daniyal Ghauri, Edward Fahy, Marcelo Saito Nogueira, Huihui Lu, Flavien Beffara, Georges Humbert, Richeal Ni Riordain, Patrick Sheahan, et al., Mobile multi-configuration clinical translational Raman system for oral cancer application, Analyst 148 (7) (2023) 1514–1523.

[30] Ravi teja Vulchi, Volodymyr Morgunov, Rajendhar Junjuri, Thomas Bocklitz, Artifacts and anomalies in Raman spectroscopy: A review on origins and correction procedures, Molecules 29 (19) (2024) 4748, http://dx.doi.org/10.3390/molecules29194748.

[31] Ewelina Wiercigroch, Ewelina Szafraniec, Krzysztof Czamara, Marta Z Pacia, Katarzyna Majzner, Kamila Kochan, Agnieszka Kaczor, Malgorzata Baranska, Kamilla Malek, Raman and infrared spectroscopy of carbohydrates: A review, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 185 (2017) 317–335, http://dx.doi.org/10.1016/j.saa.2017.05.045.

[32] Shangyuan Feng, Shaohua Huang, Duo Lin, Guannan Chen, Yuanji Xu, Yongzeng Li, Zufang Huang, Jianji Pan, Rong Chen, Haishan Zeng, Surface-enhanced Raman spectroscopy of saliva proteins for the noninvasive differentiation of benign and malignant breast tumors, Int. J. Nanomedicine (2015) 537–547, http://dx.doi.org/10.2147/IJN.S71811.

[33] Zanyar Movasaghi, Shazza Rehman, Ihtesham U. Rehman, Raman spectroscopy of biological tissues, Appl. Spectrosc. Rev. 42 (5) (2007) 493–541, http://dx.doi.org/10.1080/05704920701551530.

[34] Krzysztof Czamara, Katarzyna Majzner, Marta Z Pacia, K Kochan, Agnieszka Kaczor, M Baranska, Raman spectroscopy of lipids: a review, J. Raman Spectrosc. 46 (1) (2015) 4–20, http://dx.doi.org/10.1002/jrs.4607.

[35] Wilmar van Grondelle, Sophie Lecomte, Carmen Lopez-Iglesias, Jose-Maria Manero, Roland Cherif-Cheikh, Maité Paternostre, Céline Valéry, Lamination and spherulite-like compaction of a hormone's native amyloid-like nanofibrils: spectroscopic insights into key interactions, Faraday Discuss. 166 (2013) 163–180, http://dx.doi.org/10.1039/c3fd00054k.

[36] Giuseppe Pezzotti, Raman spectroscopy in cell biology and microbiology, J. Raman Spectrosc. 52 (12) (2021) 2348–2443, http://dx.doi.org/10.1002/jrs.6204.

[37] Tharun Varghese Jacob, Maya Ramesh, S Murali, K Ramesh, PR Sanjay, Philip Abraham, A non-invasive study to estimate and compare salivary sialic acid level as tumor marker in patients with pre-cancer and oral cancer, J. Cancer Res. Ther. 12 (2) (2016) 634–639, http://dx.doi.org/10.4103/0973-1482.148697.

[38] Mudita Sharma, Eklavya Sharma, Vishnudas Prabhu, Vinitha Ramanath Pai, Jyothi MP D'souza, Sindhu Harish, Maji Jose, Salivary L-fucose as a biomarker for oral potentially malignant disorders and oral cancer, J. Cancer Res. Ther. 16 (3) (2020) 546–550, http://dx.doi.org/10.4103/jcrt.JCRT_552_17.

[39] Levi Matthies, Medhanie T Gebrekidan, Jasper F Tegtmeyer, Nicolai Oetter, Maximilian Rohde, Tobias Vollkommer, Ralf Smeets, Waldemar Wilczak, Florian Stelzle, Martin Gosau, et al., Optical diagnosis of oral cavity lesions by label-free Raman spectroscopy, Biomed. Opt. Express 12 (2) (2021) 836–851.

[40] Hugh J Byrne, Isha Behl, Genecy Calado, Ola Ibrahim, Mary Toner, Sheila Galvin, Claire M Healy, Stephen Flint, Fiona M Lyng, Biomedical applications of vibrational spectroscopy: Oral cancer diagnostics, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 252 (2021) 119470, http://dx.doi.org/10.1016/j.saa.2021.119470.

[41] Surapaneni K Sai, Suresh Babburi, G Deepthi, SRKK Nandan, Shyam P Reddy, Praveena Adusumilli, Lipid profile in oral potentially malignant disorders and oral squamous cell carcinoma–A prognostic view, J. Oral Maxillofac. Pathol. 26 (4) (2022) 464–469, http://dx.doi.org/10.4103/jomfp.jomfp_116_22.

[42] Isha Behl, Genecy Calado, Alison Malkin, Stephen Flint, Sheila Galvin, Claire M Healy, Marina Leite Pimentel, Hugh J Byrne, Fiona M Lyng, A pilot study for early detection of oral premalignant diseases using oral cytology and Raman micro-spectroscopy: assessment of confounding factors, J. Biophotonics 13 (11) (2020) e202000079, http://dx.doi.org/10.1002/jbio.202000079.

[43] Chandramani B. More, Palak H. Shah, Rashmi Venkatesh, Estimation of serum protein in oral potentially malignant disorders and oral malignancy–A cross-sectional study, J. Clin. Diagn. Res.: JCDR 11 (2) (2017) ZC17.

[44] H Çevik-Aras, Shehed Musa, Richard Olofsson, Annica Almståhl, Ulrica Almhöjd, Patients with oral lichen planus display lower levels of salivary acidic glycoproteins than individuals without oral mucosal disease, Clin. Oral Investig. 28 (1) (2023) 2, http://dx.doi.org/10.1007/s00784-023-05411-6.

[45] Juliana Araújo da Silva Simoura, Alessandra Laís Pinho Valente Pires, Lísia Daltro Borges Alves, Franco Arsati, Ynara Bosco de Oliveira Lima-Arsati, Jean Nunes dos Santos, Valéria Souza Freitas, Psychological profile and α-amylase levels in oral lichen planus patients: A case–control preliminary study, Oral Dis. 29 (3) (2023) 1242–1249, http://dx.doi.org/10.1111/odi.14081.

[46] Ling Qian, Bo Qian, Juanyong Xu, Jingjing Yang, Guoying Wu, Yuping Zhao, Qinglan Liu, Zhiran Yuan, Yuan Fan, Huaiqi Li, Clinical relevance of serum lipids in the carcinogenesis of oral squamous cell carcinoma, BMC Oral. Health 23 (1) (2023) 200, https://bmcoralhealth.biomedcentral.com/articles/10.1186/s12903-023-02859-6.

[47] A Falamas, CI Faur, M Baciut, H Rotaru, M Chirila, S Cinta Pinzaru, M Hedesiu, Raman spectroscopic characterization of saliva for the discrimination of oral squamous cell carcinoma, Anal. Lett. 54 (1–2) (2021) 57–69.

[48] Hanna J Koster, Antonio Guillen-Perez, Juan Sebastian Gomez-Diaz, Maria Navas-Moreno, Andrew C Birkeland, Randy P Carney, Fused Raman spectroscopic analysis of blood and saliva delivers high accuracy for head and neck cancer diagnostics, Sci. Rep. 12 (1) (2022) 18464, https://www.nature.com/articles/s41598-022-22197-x.

[49] Christine Massie, Keren Chen, Andrew J. Berger, Calibration technique for suppressing residual etalon artifacts in slit-averaged Raman spectroscopy, Appl. Spectrosc. 76 (2) (2022) 255–261, https://opg.optica.org/as/abstract.cfm?uri=as-76-2-255.

Paper IV

Identification of Salivary Biomarkers for Screening of Oral Potentially Malignant Disorders Through Tissue Correlation

Siddra Maryam*,[†] Simone Innocente,[†] Edward Fahy,[‡] M. Daniyal Ghauri,[†] Sanathana Konugolu Venkata Sekar,[†,§] Ray Burke,[†] Andrea Visentin,[¶] Richeal Ni Riordain,[‡] Stefan Andersson-Engels[a],[†,§] Katarzyna Komolibus,[†] and Rekha Gautam[*,†]

†*Biophotonics@Tyndall, Tyndall National Institute, UCC, Dyke Parade, Cork, Ireland, T12R5CP.*

‡*Cork University Dental School and Hospital, Wilton, Cork, Ireland, T12E8YV.*

¶*Department of Computer Science, UCC, College road, Cork, Ireland, T12 CY82.*

§ *School of Physics, UCC, College Road, Ireland, T12 YN60.*

E-mail: *siddra.maryam@tyndall.ie;*rekha.gautam@tyndall.ie

Abstract

The development of a reliable screening procedure for oral potentially malignant disorders (OPMD) is crucial for early diagnosis, timely intervention and improved patient outcomes. While saliva is a promising diagnostic medium for non-invasive screening, it can also be influenced by other health conditions. This study explores the potential of Raman spectroscopy for OPMD screening by analyzing biochemical changes in both saliva and tissue. By correlating saliva and tissue spectra from the same participants, we identified key OPMD-associated salivary biomarkers using least absolute shrinkage

and selection operator (Lasso) regression for feature selection. Notably, Raman peaks at 501 cm^{-1}, 541 cm^{-1} and 1463 cm^{-1} showed strong correlation between saliva and tissue. Linear Discriminant Analysis (LDA) was used for classification between healthy and OPMD group. The model achieved 78% accuracy for saliva and 89% for tissue, with key biochemical markers showing a consistent trend across both mediums. The notable correlation between saliva and tissue spectra highlights potential of saliva screening.

Introduction

Oral potentially malignant disorders (OPMD) are lesions in the oral cavity with increased risk of malignant transformation. These disorders include conditions such as leukoplakia, submucosal fibrosis, actinic cheilitis or dysplasia. Early detection is crucial for timely intervention, reducing the burden of invasive treatments and improving patient outcomes. The gold standard diagnostic methods for OPMD is visual examination, palpation and histopathology of tissue biopsies. While most reliable, this set of examinations is subjective, invasive, often requires local anesthesia and causes patient discomfort. Additionally, it is also time-consuming, making it unsuitable for routine screening or monitoring disease progression. Given that OPMD is a rapidly progressing disease, even the slightest delay in diagnosis or treatment can have significant consequences.[1]

It is also a considerable financial burden, with the average cost of biopsy ranging between $ 750 and $ 1,000,[2,3] despite the fact that 92 % of OPMD cases do not progress to malignancy,[4] raising concerns about the cost-effectiveness of biopsy as a sole diagnostic method, especially in resource-limited settings. Relying only on biopsy and histopathology also imposes a strain on the healthcare system in regions where access to specialists and laboratories is limited.

These challenges emphasize the need for non-invasive, real-time and cost-effective alternatives such as saliva screening, that can streamline the diagnosis and monitoring of OPMD while reducing the financial and logistic burden from patients and the healthcare system.[5]

Saliva being an easily accessible body fluid has recently gained attention as a potential diag-

nostic medium for various diseases. It could be particularly useful for diagnosing OPMD due to its direct interaction with lesions in the oral cavity, making it a valuable tool in detecting specific biomarkers that reflect the pathological state of oral tissue. Several studies have explored saliva for oral cancer and OPMD related biomarkers.[6–9]

However, saliva-based diagnostic methods also face challenges, such as standardization of sampling protocols, validating biomarkers, and results interpretation because of the multifactorial origin of salivary biomarkers. Further research is required to develop a reliable and accurate saliva screening method.

In recent years, optical methods such as Raman spectroscopy have shown potential in diagnosing oral cancer and OPMD by detecting subtle molecular changes, such as changes in levels of proteins, carbohydrates, lipids and nucleic acid, which together are referred as molecular fingerprints, in various biological samples.[10] It works well in both *ex vivo* and *in vivo* environments due to its non-invasive nature and ability to obtain signals from the oral cavity and body fluids,[11] avoiding the risk of bleeding, allergic reaction and other complications that can arise after a conventional tissue biopsy.[12]

This study explores the potential of saliva as a screening tool for OPMD by identifying salivary biomarkers in Raman spectra and correlating them with *in vivo* tissue Raman data from the same participants. Lasso (least absolute shrinkage and selection operator) feature selection model identified key spectral differences between OPMD and healthy controls. Comparing saliva and tissue spectra revealed consistent trends in protein, lipid and carbohydrate signatures, strengthening the hypothesis that saliva can be a potential surrogate for non-invasive screening.

Materials and methods

This study was conducted in a collaboration between Tyndall National Institute and Cork University Dental School and Hospital (CUH) with ethical approval from the Cork Research Ethics Committee (CREC) of University College Cork (UCC) (CREC reference: ECM 4 (y) 13/4/2021 & ECM 3 (aaaa) 12/09/2023). This study adhered to the ethical standards outlined by CREC, the 1964 Helsinki Declaration and its subsequent amendments. This study included 19 healthy controls and 18 OPMD volunteers, all of whom provided explicit informed consent. Only patients scheduled for biopsy or those who had undergone biopsy were recruited in the study. The clinician involved in this study confirmed the oral hygiene of the participant and collected detailed demographic and clinical information, including medical history and diagnosis. Participants were advised to refrain from food, drinks (except water), smoking or using oral hygiene products and any lip cosmetics for at least one hour prior to sample collection. After rinsing the mouth for 1 to 2 mins, unstimulated saliva was collected in sterile tubes and stored at 4°C. The saliva samples were then centrifuged at 10,000 rpm for 10 minutes to remove cells and any solid particles. The resulting cell pellet and supernatant were stored separately at -80°C. *In vivo* Raman spectra was recorded using a handheld Raman probe[13] from six oral tissues: cheek mucosa (CM), lip, gingiva, dorsal surface of tongue (DST), ventral surface of tongue (VST) and anterior floor of mouth (AFM) in both control and OPMD groups. For OPMD patients, additional data were collected from the lesion and a contra-lateral site. However, for the purpose of this study, only CM data was analyzed due to the limited sample availability from other tissue sites in the OPMD group. Figure 1 presents the experimental work flow.

Raman analysis

***In vivo* tissue:** After saliva sample collection, the participants were comfortably seated in a dental chair. The area of interest in the oral cavity was gently dried with sterilized

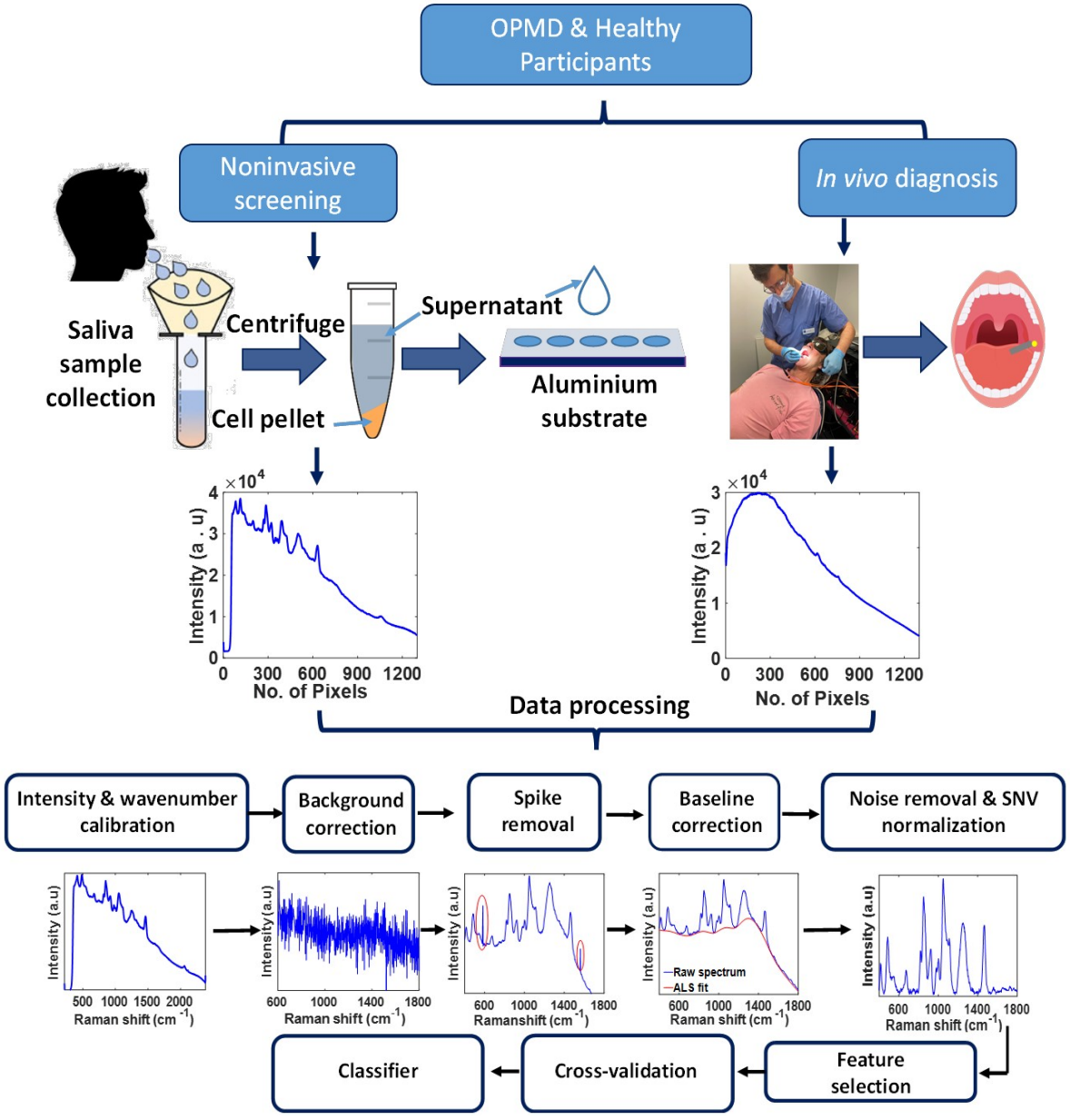


Figure 1: Experimental workflow and data analysis

medical grade gauze before and between measurements to minimize saliva interference. The clinician placed the Raman probe[13] on the tissue to record the data. Instrument parameters including laser power, acquisition time and number were optimized based on data from a few initial participants. Five spectra were recorded from each tissue type including OPMD lesion, contralateral site and healthy tissue with 30 mW power for 0.43 seconds averaged over 9 acquisitions.

Saliva samples: Frozen saliva samples were thawed at room temperature and filtered with 300 KDa centrifugal filters to ensure homogeneity and remove large particles.[14] Each sample was prepared layering five 2 μl droplets of filtered saliva subsequently on an aluminum substrate to let dry. Dried saliva samples were then analyzed at room temperature with the customized Raman microscope using a 785 nm laser wavelength and a 50x objective lens.[13] Each spectrum was recorded for 10 sec and then averaged over 5 sequentially recoded spectra from each position. The excitation power was 20 mW. Three spectra were recorded per droplet, totaling 15 spectra from five saliva droplets per sample.

Data analysis

The system was calibrated for wavelength dependent sensitivity using a source with a National Institute of Standards and Technology (NIST) traceble white light spectrum; and wavelength-scale using a Neon Argon (NeAr) lamp from Intellical spectral calibration system, as well as well-defined substances exhibiting known spectra with distinct Raman peaks, such as acetaminophen and naphthalene. Dark background (BG) was subtracted from the collected data and spectral spikes (cosmic rays) were removed. Baseline correction was performed by asymmetric least square method with λ = 100, P = 0.00009 for saliva and λ = 10 and P = 0.0001 for tissue to remove autofluorescence background. A second order Savitzky-Golay filter with window width 5 was applied to reduce noise, followed by standard normal variate (SNV) normalization to minimize spectral variability. The Raman spectral range from 400 – 1800 cm^{-1} was selected for further analysis. The saliva data consists of

555 Raman spectra from 37 participants, while tissue dataset include 175 spectra from the same participants.

Analysis methodology

To identify key spectral features distinguishing between OPMD and control groups, Lasso regression model was employed. Lasso eliminates less relevant features by adding a penalty term proportional to the sum of regression coefficients and preserve only those features that are crucial for classification. Features selected at the dips of Raman spectra were removed because they are believed to not be very informative and likely unrelated to any biochemical changes in tissue or saliva. To improve the analysis, additional features were incorporated using clinical domain knowledge ensuring the inclusion of chemically relevant biomarkers linked to disease progression. These features included Raman bands associated with proteins (such as 501 and 826 cm^{-1}), carbohydrates (541 cm^{-1}) and lipids (1081 cm^{-1}) Linear Discriminant Analysis (LDA) was then applied to evaluate the efficacy of the selected features in classifying the samples using leave-one-out (LOO) cross validation model on patient level. Where one patient data was excluded in each iteration to access the model's predictive performance.

Results and discussion

Figure 2 presents the normalized mean Raman spectra of saliva and tissue from control (blue) and OPMD (red) cohorts, showing noticeable intensity differences in several regions, around 500, 825, 1050, 1120, 1270, 1450, 1580, and 1660 cm^{-1}. To further investigate, a Lasso regression model was applied to extract key features across the entire spectral range, identifying markers with diagnostic potential. In Figure 2 A and B, the Lasso-selected features are marked as dotted lines on the spectra, emphasizing the spectral regions contributing most to the differentiation between control and OPMD cohorts. These Lasso-selected features were

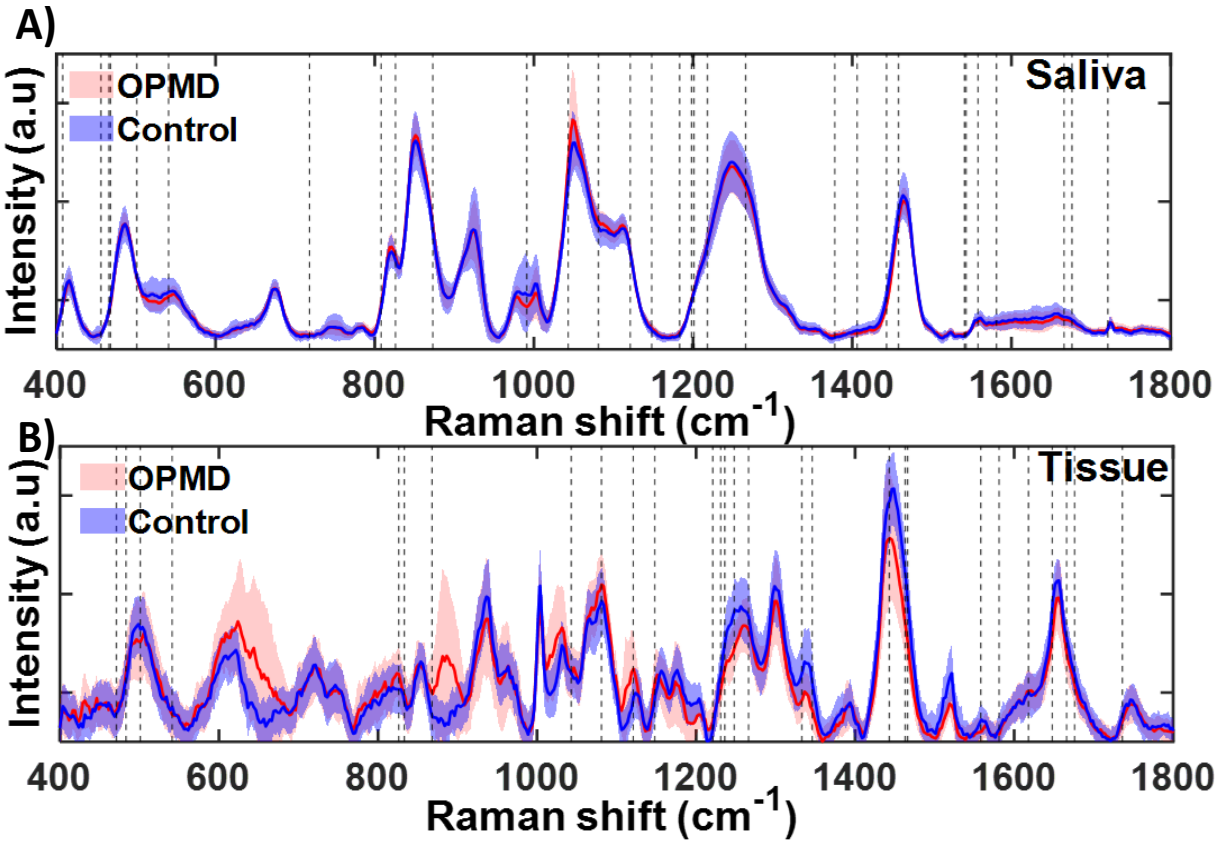


Figure 2: Average Raman spectra of A) saliva and B) tissue from OPMD (red) and control (blue) groups. Key Lasso features, excluding those at dips and additional chemically relevant features are marked on the spectra. The shaded regions represent standard deviations, indicating spectral variability across samples within each group.

further integrated with visually distinct spectral regions identified in the averaged spectra of tissue samples. The exact locations of these visually apparent regions were determined based on their statistical significance in differentiation of OPMD from control tissue.

A comparison of selected features of tissue and saliva revealed thirteen common, statistically significant features, with slight variations in their exact positions due to differences in the spectral profiles of the two sample types. Figure 3 displays the statistical analysis plots for these common features, highlighting their significance in distinguishing OPMD from control. The Kruskal-Wallis non-parametric test was used for significance analysis, and $p<0.05$ was considered significant. These shared features suggest that saliva could serve as a non-invasive screening tool for identifying individuals at risk, facilitating early detection while reducing the number of biopsies. The biochemicals associated with these spectral markers (features) are discussed below.

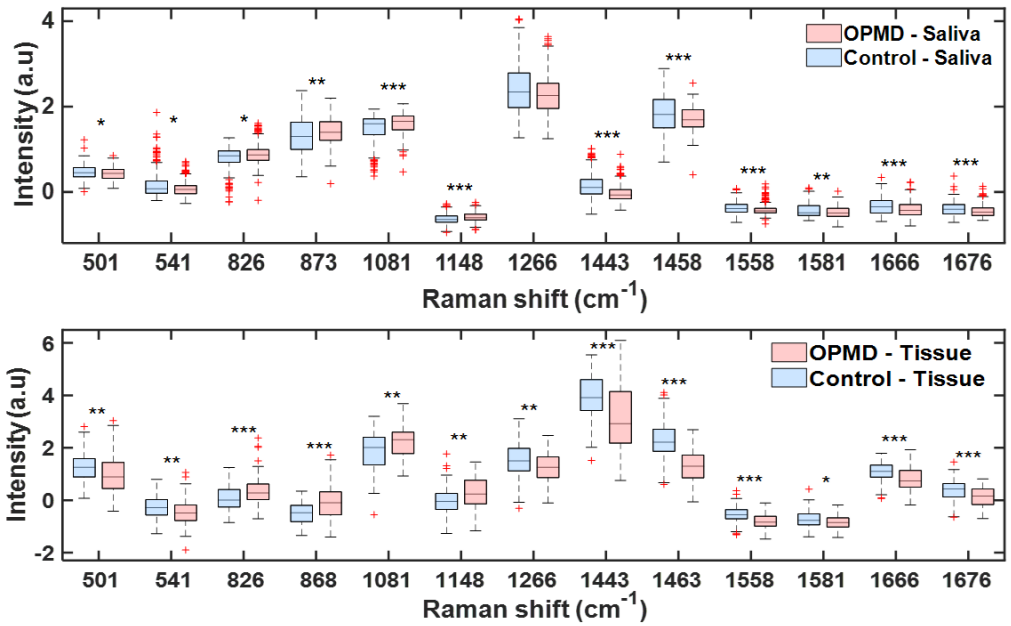


Figure 3: Box plot showing the distribution of selected Raman features for saliva (top) and tissue (bottom) and comparing OPMD (red) and control (blue) groups. The plots highlight the statistical difference and trend between both groups across the selected features. Each feature was selected with Lasso feature selection model and domain knowledge. Visual comparison emphasizes that the trend of the selected features is common in saliva and tissue.

Proteins and amino acids: Analysis of spectral changes in saliva and tissue of patients with oral potentially malignant disorders (OPMD) compared to healthy individuals provides significant insights into the biochemical alterations associated with malignancy. Notably, the observed decrease in the amide III (1266 cm^{-1}) and amide I (1666 cm^{-1}) bands indicate protein denaturation and breakdown, suggesting an increase in proteolytic activity.[15–18] Furthermore, a decline in the gauche-gauche-gauche conformers of disulfide protein linkages at 501 cm^{-1} supports the concept of increased protein disorder[19,20]

This degradation of proteins is likely associated with a significant increase in amino acid-related peaks, particularly those corresponding to tyrosine (826 cm^{-1}) and hydroxyproline (873 cm^{-1}).[21,22] As the demand for glucose and amino acids rises to support cell proliferation and lesion growth, this leads to the breakdown of proteins into amino acids and a notable decrease in overall protein levels within the saliva of OPMD patients. Elevated levels of amino acids, including tyrosine, have also been observed earlier with oral cancer progression.[23,24] As saliva is in direct contact with oral tissues, these spectral signatures may mirror changes occurring in the tissue itself, making saliva a potential non-invasive screening tool for early-stage OPMD detection.

Lipids: Lipids are principal cellular molecules involved in enzyme activity, energy storage, cell division, stabilising DNA chains and maintaining structural and functional integrity of the cell membrane.[25] The lipid profile in the OPMD spectra of tissue and saliva represents distinct metabolic changes that could be significant for the disease diagnosis.

Raman peak at 1081 cm^{-1} is more prominent in OPMD group. This gauche band is associated with C-C stretch in lipids and is sensitive to lipid chain conformations. This could be associated with oxidation of lipids, which significantly contributes to the progression of cancer and inflammation. Furthermore, 1450 cm^{-1} region in tissue and saliva is associated with the CH_2 bending mode in fatty acid chains.[26–29] These peaks show a decrease in the OPMD group, which could be the result of lipid utilization in membrane biogenesis in rapidly dividing cells.[30,31] These spectral changes suggest that OPMD is not only associated with

the change in lipid content but also with lipid structures.[25,32]

Carbohydrates: Raman peaks at 541 cm^{-1} and 1463 cm^{-1} are associated with glycogen and saccharides. In the OPMD group, reduction in glycogen levels suggests impaired energy storage and utilization, potentially indicating cellular stress and metabolic disturbances associated with disease progression.[33] Rapid cell proliferation, increase glycogen breakdown under hypoxic condition through anerobic glycolysis, resulting in less ATP (adenosine triphosphate) production and driving increased glucose demand.[34]

Haemoglobin: In OPMD, Raman peaks at 1558 and 1581 cm^{-1}, associated with heme, are decreasing significantly, suggesting changes in blood composition within affected tissue.[35] The decrease in the OPMD group implies reduced hemoglobin concentration or change in blood flow in OPMD tissue due to ineffective vascularization.

Carotenoids: Raman peak at 1148 cm^{-1} depicts a significant increase in OPMD group for both tissue and saliva. This peak is attributed to the C-C stretching mode of carotenoids. The elevated intensity in OPMD may indicate metabolic changes, oxidative stress, or an adaptive cellular response to disease progression.[36]

These selected features were employed to differentiate between the OPMD and control groups using LDA. The leave-one-participant-out cross-validation method was applied to test the LDA model rigorously. Figure 4 (A and B) represents scatter plots showcasing the LDA classification, where each point corresponds to an individual participant. Correctly classified cases are marked in blue and red, while misclassified cases are indicated by green and yellow. The truth tables in Figure 4 (C and D) summarize classification outcomes, revealing that the model achieved 78 % accuracy for saliva and 89 % accuracy for tissue samples. The higher accuracy in *in vivo* tissue reflects more pronounced biochemical changes in tissue as a primary cite compared to saliva, reflecting the localized effect of OPMD. Nevertheless, saliva remains a valuable diagnostic tool for non-invasive screening with reasonable accuracy. It shares distinct biochemical markers with tissue, enabling effective classification. The overlap of significant features between saliva and tissue underscores that saliva reflects comparable

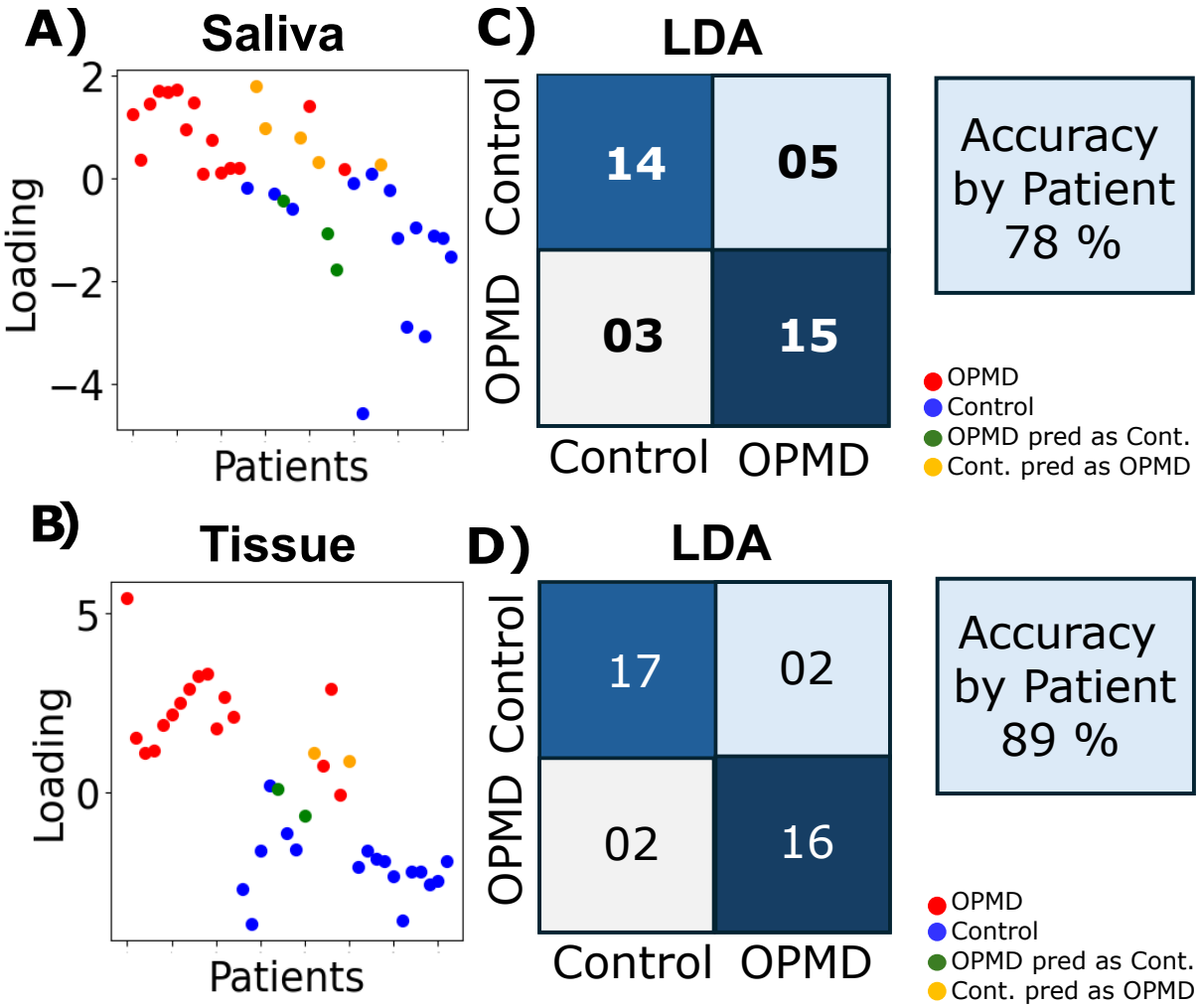


Figure 4: LDA scatter plot showing patient classification on LDA loadings for A) saliva B) tissue. Each point represents a patient with correctly labeled OPMD (red), Control (blue) and misclassified points (yellow and green). Right side of the figure shows confusion matrix of C) saliva D) tissue. The confusion matrices summarize the classification performance, with their respective overall accuracy on patient level in the box adjacent to the confusion matrices.

structural and metabolic changes associated with OPMD found in tissue samples. To identify highly correlating features between tissue and saliva, a correlation analysis was performed, as shown in Figure 5 . This correlation not only supports the potential of using saliva as a non-invasive screening tool but also depicts the consistency of Raman spectral features across different biological media.

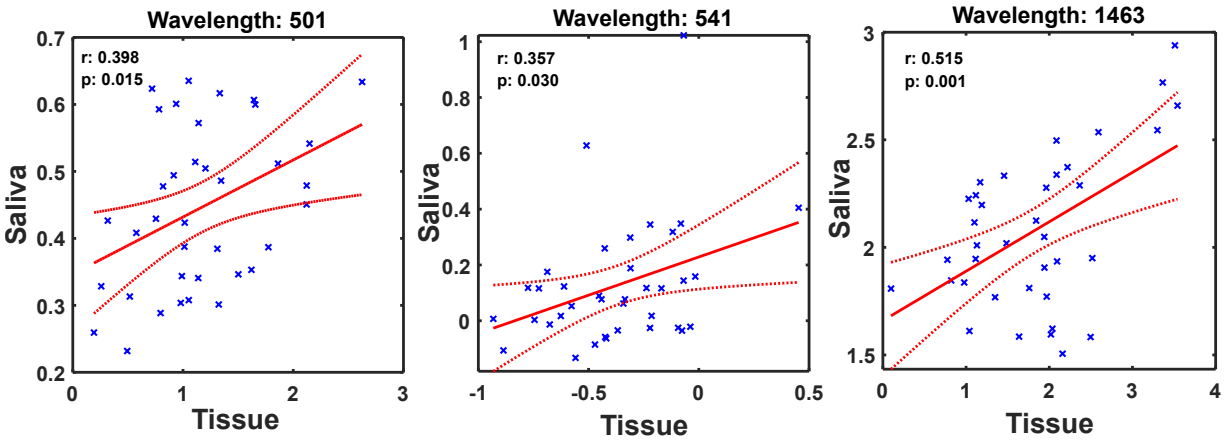


Figure 5: Association between selected Raman features from tissue along x-axis and saliva along y-axis. Raman intensities from peaks correspond to proteins and carbohydrates. Correlation coefficients (Pearson) and corresponding P values are indicated on each panel. Lines represent fitting by linear regression with 95 % confidence band. A strong correlation suggests that a common biomarker between saliva and tissue can assist in OPMD detection.

Conclusion

This study demonstrates the potential of Raman spectroscopy combined with Lasso regression in identifying key biomarkers associated with OPMD, aiding non-invasive diagnostics. LDA provides a visual representation of the classification performance for saliva and tissue samples. The model achieved accuracy of 78% for saliva and 89 % for tissue. The higher accuracy in oral tissue analysis reflects more pronounced biochemical changes in tissue compared to saliva. However, saliva still offers a non-invasive alternative with a reasonable accuracy, which could be improved by exploring additional biomarkers or by improving the feature selection technique.

Future perspective

Future studies could focus on improving classification by integrating additional biomarkers and refining feature selection. Larger sample sizes and comprehensive statistical analysis are currently underway to enhance the reliability of this dual-medium approach for OPMD screening.

Acknowledgement

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant number SFI/15/RP/2828 and SFI/22/RP-2TF/10293.

References

(1) Stefanuto, P.; Doucet, J.-C.; Robertson, C. Delays in treatment of oral cancer: a review of the current literature. *Oral surgery, oral medicine, oral pathology and oral radiology* **2014**, *117*, 424–429.

(2) `https://www.corkoralsurgery.ie/treatments-and-fees/`, Treatment and fee.

(3) `https://frazerdental.ie/fees-services/`, Biopsy service for Mouth Cancer - Dr. Hani Jamah.

(4) Iocca, O.; Sollecito, T. P.; Alawi, F.; Weinstein, G. S.; Newman, J. G.; De Virgilio, A.; Di Maio, P.; Spriano, G.; Pardiñas López, S.; Shanti, R. M. Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. *Head & neck* **2020**, *42*, 539–555.

(5) Ralbovsky, N. M.; Lednev, I. K. Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning. *Chemical Society Reviews* **2020**, *49*, 7428–7453.

(6) Hu, S.; Arellano, M.; Boontheung, P.; Wang, J.; Zhou, H.; Jiang, J.; Elashoff, D.; Wei, R.; Loo, J. A.; Wong, D. T. Salivary proteomics for oral cancer biomarker discovery. *Clinical Cancer Research* **2008**, *14*, 6246–6252.

(7) Riccardi, G.; Bellizzi, M. G.; Fatuzzo, I.; Zoccali, F.; Cavalcanti, L.; Greco, A.; Vincentiis, M. d.; Ralli, M.; Fiore, M.; Petrella, C.; others Salivary biomarkers in oral squamous cell carcinoma: a proteomic overview. *Proteomes* **2022**, *10*, 37.

(8) Khurshid, Z.; Zafar, M. S.; Khan, R. S.; Najeeb, S.; Slowey, P. D.; Rehman, I. U. Role of salivary biomarkers in oral cancer detection. *Advances in clinical chemistry* **2018**, *86*, 23–70.

(9) Piyarathne, N. S.; Rasnayake, R.; Angammana, R.; Chandrasekera, P.; Ramachandra, S.; Weerasekera, M.; Yasawardene, S.; Abu-Eid, R.; Jayasinghe, J. P.; Gupta, E. Diagnostic salivary biomarkers in oral cancer and oral potentially malignant disorders and their relationships to risk factors–A systematic review. *Expert Review of Molecular Diagnostics* **2021**, *21*, 789–807.

(10) Das, R. S.; Agrawal, Y. Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational spectroscopy* **2011**, *57*, 163–176.

(11) Krafft, C.; Sergo, V. Biomedical applications of Raman and infrared spectroscopy to diagnose tissues. *Spectroscopy* **2006**, *20*, 195–218.

(12) McGregor, H. C.; Short, M. A.; Lam, S.; Shaipanich, T.; Beaudoin, E.-L.; Zeng, H. Development and in vivo test of a miniature Raman probe for early cancer detection in the peripheral lung. *Journal of biophotonics* **2018**, *11*, e201800055.

(13) Maryam, S.; Sekar, S. K. V.; Ghauri, M. D.; Fahy, E.; Nogueira, M. S.; Lu, H.; Beffara, F.; Humbert, G.; Riordain, R. N.; Sheahan, P.; others Mobile multi-configuration clinical translational Raman system for oral cancer application. *Analyst* **2023**, *148*, 1514–1523.

(14) Carlomagno, C.; Banfi, P.; Gualerzi, A.; Picciolini, S.; Volpato, E.; Meloni, M.; Lax, A.; Colombo, E.; Ticozzi, N.; Verde, F.; others Human salivary Raman fingerprint as biomarker for the diagnosis of Amyotrophic Lateral Sclerosis. *Scientific Reports* **2020**, *10*, 10175.

(15) Çevik-Aras, H.; Musa, S.; Olofsson, R.; Almståhl, A.; Almhöjd, U. Patients with Oral Lichen Planus display lower levels of salivary acidic glycoproteins than individuals without oral mucosal disease. *Clinical Oral Investigations* **2023**, *28*, 2.

(16) Talungchit, S.; Buajeeb, W.; Lerdtripop, C.; Surarit, R.; Chairatvit, K.; Roytrakul, S.; Kobayashi, H.; Izumi, Y.; Khovidhunkit, S.-o. P. Putative salivary protein biomarkers

for the diagnosis of Oral Lichen Planus: a case-control study. *BMC Oral Health* **2018**, *18*, 1–14.

(17) Simoura, J. A. d. S.; Pires, A. L. P. V.; Alves, L. D. B.; Arsati, F.; Lima-Arsati, Y. B. d. O.; Santos, J. N. d.; Freitas, V. S. Psychological profile and α-amylase levels in Oral Lichen Planus patients: A case–control preliminary study. *Oral Diseases* **2023**, *29*, 1242–1249.

(18) Lieu, E. L.; Nguyen, T.; Rhyne, S.; Kim, J. Amino acids in cancer. *Experimental & molecular medicine* **2020**, *52*, 15–30.

(19) Schlücker, S.; Liang, C.; Strehle, K.; DiGiovanna, J.; Kraemer, K.; Levin, I. Conformational differences in protein disulfide linkages between normal hair and hair from subjects with trichothiodystrophy: a quantitative analysis by Raman microspectroscopy. *Biopolymers: Original Research on Biomolecules* **2006**, *82*, 615–622.

(20) Koster, H. J.; Guillen-Perez, A.; Gomez-Diaz, J. S.; Navas-Moreno, M.; Birkeland, A. C.; Carney, R. P. Fused Raman spectroscopic analysis of blood and saliva delivers high accuracy for head and neck cancer diagnostics. *Scientific reports* **2022**, *12*, 18464.

(21) Rekha, P.; Aruna, P.; Brindha, E.; Koteeswaran, D.; Baludavid, M.; Ganesan, S. Near-infrared Raman spectroscopic characterization of salivary metabolites in the discrimination of normal from oral premalignant and malignant conditions. *Journal of Raman Spectroscopy* **2016**, *47*, 763–772.

(22) Maryam, S.; Benazza, A.; Fahy, E.; Sekar, S. K. V.; Dinish, U.; Olivo, M.; Riordain, R. N.; Andersson-Engels, S.; Humbert, G.; Komolibus, K.; others Liquid saliva analysis using optofluidic photonic crystal fiber for detection of oral potentially malignant disorders. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2025**, 125788.

(23) Yatsuoka, W.; Ueno, T.; Miyano, K.; Enomoto, A.; Ota, S.; Sugimoto, M.; Uezono, Y. Time-course of salivary metabolomic profiles during radiation therapy for head and neck cancer. *Journal of Clinical Medicine* **2021**, *10*, 2631.

(24) Bel'skaya, L. V.; Sarf, E. A.; Loginova, A. I. Diagnostic value of salivary amino acid levels in cancer. *Metabolites* **2023**, *13*, 950.

(25) Granero Fernandez, M.; Lopez-Jornet, P. Association between smoking, glycaemia, blood lipoproteins and risk of oral leukoplakia. *Australian Dental Journal* **2017**, *62*, 47–51.

(26) Czamara, K.; Majzner, K.; Pacia, M. Z.; Kochan, K.; Kaczor, A.; Baranska, M. Raman spectroscopy of lipids: a review. *Journal of Raman spectroscopy* **2015**, *46*, 4–20.

(27) Zúñiga, W. C.; Jones, V.; Anderson, S. M.; Echevarria, A.; Miller, N. L.; Stashko, C.; Schmolze, D.; Cha, P. D.; Kothari, R.; Fong, Y.; others Raman spectroscopy for rapid evaluation of surgical margins during breast cancer lumpectomy. *Scientific reports* **2019**, *9*, 14639.

(28) Fan, L.; Wang, H.; Wu, X.; Wang, S.; Han, Y.; Wang, J. Raman spectroscopic study of benign and malignant ovarian tissues. *Laser Physics* **2022**, *32*, 035601.

(29) Ibrahim, O.; Toner, M.; Flint, S.; Byrne, H. J.; Lyng, F. M. The potential of Raman spectroscopy in the diagnosis of dysplastic and malignant oral lesions. *Cancers* **2021**, *13*, 619.

(30) Abdolsamadi, H.; Rafieian, N.; Goodarzi, M. T.; Feradmal, J.; Davoodi, P.; Jazayeri, M.; Taghavi, Z.; Hoseyni, S.-M.; Ahmadi-Motamayel, F. Levels of salivary antioxidant vitamins and lipid peroxidation in patients with oral lichen planus and healthy individuals. *Chonnam medical journal* **2014**, *50*, 58–62.

(31) Sai, S. K.; Babburi, S.; Deepthi, G.; Nandan, S.; Reddy, S. P.; Adusumilli, P. Lipid profile in oral potentially malignant disorders and oral squamous cell carcinoma–A prognostic view. *Journal of Oral and Maxillofacial Pathology* **2022**, *26*, 464–469.

(32) Mesgarzadeh, A. H.; Akbarzadeh, A.; Rasipour, A.; Rasipour, T.; Mehdizadeh, A.; Shaaker, M. Secretory phospholipase-A2 and fatty acid composition in oral reactive lesions: a cross-sectional study. *Cancer Cell International* **2017**, *17*, 1–7.

(33) Yoshimura, N.; Yamada, S.-i.; Aizawa, H.; Xiao, T.; Nishimaki, F.; Kurita, H. Glycogen metabolism in an oral dysplastic/cancerous (iodine-negative) epithelium: Glycogen was consumed in the pentose phosphate pathway, not in glycolysis. *Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology* **2019**, *31*, 288–294.

(34) Nie, H.; Hu, X.; Xiong, H.; Zeng, L.; Chen, W.; Su, T. Change and pathological significance of glycogen content in oral squamous cell carcinoma and oral submucous fibrosis. *Tissue and Cell* **2024**, *87*, 102337.

(35) Gautam, R.; Oh, J.-Y.; Marques, M. B.; Dluhy, R. A.; Patel, R. P. Characterization of storage-induced red blood cell hemolysis using Raman spectroscopy. *Laboratory medicine* **2018**, *49*, 298–310.

(36) Albanidou, F. E.; Markopoulos, A.; Kayavis, I. Antioxidant micronutrient levels in the serum of patients with oral lichen planus. *Balkan Journal of Stomatology* **2003**, *7*, 191–193.

Paper V

Diagnostic potential of diffuse reflectance spectroscopy and feature selection in identifying oral potentially malignant disorders

Siddra Maryam*[a,b], M. Daniyal Ghauri[a,b], Edward Fahy[c], Simone Innocente[a,d], Ray Burke[a], Andrea Visentin[a,d], Richeal Ni Riordain[c], Linda Feeley[e], Stefan Andersson-Engels[a,b], Rekha Gautam[a], Katarzyna Komolibus[a]

[a]Tyndall National Institute, Dyke Parade, Cork, Ireland, T12R5CP.
[b]School of Physics, UCC, College Road, Ireland, T12 YN60.
[c]Cork University Dental School and Hospital, Wilton, Cork, Ireland, T12E8YV.
[d]Department of Computer Science, UCC, College road, Cork, Ireland, T12 CY82.
[e]Cork University Hospital, Wilton, Cork, Ireland, T12E8YV.

Abstract.

0.1 Significance

Early diagnosis of oral potentially malignant disorders (OPMD) is essential to prevent progression to oral cancer. This requires objective, non-invasive tools capable of identifying subtle, subclinical physiological alterations in oral tissue.

0.2 Aim

This study investigated the potential of diffuse reflectance spectroscopy (DRS) combined with machine learning-based feature selection to analyse spectral features of relevance for detection of OPMD-associated changes in different anatomical sites.

0.3 Approach

DRS spectra were acquired from the cheek mucosa (CM) and ventral surface of the tongue (VST) in 47 patients with OPMD and 22 healthy controls. Least Absolute Shrinkage and Selection Operator (LASSO) was used to select key spectral features, which were subsequently used to classify OPMD tissue and healthy controls by Linear Discriminant Analysis (LDA).

0.4 Results

Distinct spectral differences were observed between healthy and OPMD tissues, consistent with alterations in oxygenation, hemoglobin absorption, and vascularization. The developed model achieved classification accuracies of 87% for CM and 89% for VST. The contralateral sites exhibited heterogeneous spectral responses, suggesting the presence of early field effects.

0.5 Conclusion

Although most diagnostic signatures were identified within the visible spectral range, DRS revealed subtle biochemical and structural alterations beyond visual detection, enabling more objective tissue characterization. Incorporating LASSO-selected features enhanced the model's robustness and classification performance.

Keywords: Diffuse reflectance spectroscopy, oral potentially malignant disorders, oral cancer, feature selection, non-invasive diagnostics.

*Siddra Maryam, siddra.maryam@tyndall.ie **Katarzyna Komolibus, katarzyna.komolibus@tyndall.ie

1 Introduction

Oral cancer is the 16th most common cancer worldwide, with a survival rate strongly associated with the stage of the disease at the time of diagnosis.[1,2] More than 45% of the cancer cases have metastasised to the lymph nodes when first diagnosed.[2,3] Early-stage diagnosis remains challenging due to the lack of obvious clinical signs and the limited sensitivity of visual examination or commercially available oral screening methods. More than 90% of oral cancers constitute squamous cell carcinomas (OSCC), which originate from genetic mutations in epithelial cells[4] that trigger uncontrolled cell proliferation. In some cases, the body's defence mechanisms are unable to repair DNA damage or eliminate these abnormal cells through apoptosis, leading to neoplastic transformation.[5] A critical stage in oral carcinogenesis is the development of oral potentially malignant disorders (OPMD), a group of precancerous conditions that increase the risk of progression to OSCC. Identifying these early-stage abnormalities can help facilitate timely intervention and improve patient outcomes.

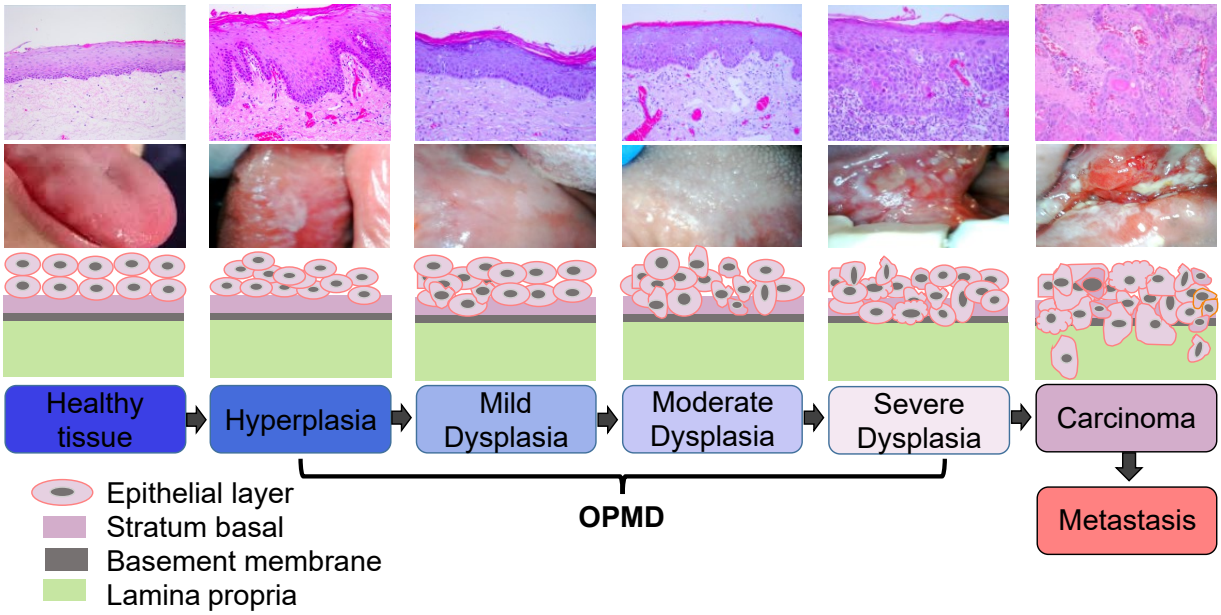


Fig 1 Disease progression in oral tissue.

OPMD often presents as persistent white or red patches, or as easily bleeding sores on the lips or within the oral cavity, such as leukoplakia and lichen planus. As the disease progresses, these lesions can develop into non-healing ulcers, frequently accompanied by numbness, tenderness, pain and difficulty in swallowing. Figure 1 illustrates the typical progression of these lesions within oral tissue. The diagnostic procedure for OPMD involves reviewing the patient's medical history and conducting a visual examination. However, a definitive diagnosis relies on histopathological evaluation of the biopsy sample. Despite its high accuracy, biopsy is invasive, time-consuming, impractical for multifocal lesions and sometimes prone to variability in sampling and interpretation.[6] Traditional imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) can help to assess the extent and progression of the lesions, but are expensive and unsuitable for screening and routine monitoring. Given these limitations, non-invasive diagnostic approaches are being widely explored.

Several studies have explored exhaled breath analysis to detect volatile organic compounds that exhibit distinct concentration profiles associated with disease progression[7,8] . Other research focuses on blood- and saliva-based biomarkers, such as circulating tumor cells, DNA and RNA fragments;[9–11] specific proteins[12] including CD44,[13] Cyfra 21,[14] CA-125;[15] inflammatory markers such as C-reactive proteins.[16] These biomarkers are typically analysed using complicated laboratory techniques like enzyme-linked immunosorbent assays (ELISA),[16] next generation sequencing[17] or polymerase chain reaction (PCR).[18] Salivary IL-1β and IL-8 have also been reported as promising biomarkers for diagnosing oral cancer.[19] Moreover, saliva-based home test kits, such as "CancerDetect Test for Oral and Throat Cancer"[20,21] have been introduced. However, most of these methods lack FDA approval or widespread clinical acceptance and often struggle to detect biomarkers at ultra-low concentrations during the OPMD stage.[22,23] Therefore, there remains an

urgent need for a cost-effective, non-invasive, real-time screening tool that can be easily integrated into clinical workflows.

Among various emerging technologies, optical spectroscopy offers a promising, label-free approach to tissue characterisation, providing non-invasive insights into biochemical and morphological changes associated with disease progression. Techniques such as fluorescence, Raman spectroscopy and diffuse reflectance spectroscopy (DRS) have demonstrated potential in differentiating between normal and abnormal oral tissues.[12] Fluorescence-based adjunctive optical screening devices, such as Vizilite and VELscope, have been developed to enhance the visualisation of oral mucosa abnormalities. While offering improved lesion contrast, their diagnostic accuracy remains limited. On the other hand, DRS encodes information related to tissue's biochemical and structural properties including haemoglobin concentration, oxygenation levels and scattering changes due to alterations in epithelial thickness and cellular morphology, typically indicative of OPMD progression.[24] Some works have explored the combination of fluorescence with DRS.[25–27] For example, a commercially available device, Identafi, combines autofluorescence from epithelial and stromal layers and reflectance in 500-600 nm region to visualise vasculature with only moderate sensitivity and specificity in differentiating oral lesions.[28] Previous studies focusing solely on DRS have reported effective discrimination between normal, OPMD and malignant tissues.[29,30] Many of the studies have concentrated on buccal mucosa and tongue lesions - the most frequent OPMD sites - reporting in-vivo accuracy values in a wide range from 75%[31] up to 98%.[30,32] Other studies explored the potential of DRS as a tool for the evaluation of resection margins,[33] biopsy guidance[34] with encouraging results. While majority of the works analyse the full DRS spectrum (typically 400 - 900 nm), others focus on ratio metric changes related to haemoglobin absorption peaks.[35,36] However, DRS data are inherently complex and high-dimensional, often containing redundant and

correlated spectral features. These characteristics pose challenges for direct classification, often leading to model over-fitting or reduced diagnostic accuracy. Therefore, identifying the most informative spectral features that contribute to tissue differentiation is essential for developing robust and interpretable diagnostic models.

Unlike prior DRS-based studies that primarily analyse entire spectra, this work systematically identities and validates the most diagnostically informative spectral regions using feature selection. In this study, we investigate DRS as a potential tool for the regular screening of OPMD and as an adjunct tool for biopsy guidance. Using broadband (380 - 1100 nm) DRS measurements acquired *in vivo* from 82 participants, combined with feature selection and machine learning (ML) algorithms, we identify the most discriminative spectral regions for tissue characterisation. This analysis reveals spectral biomarkers linked to early pathological changes, facilitating objective and interpretable approach to OPMD diagnosis. The primary objectives of this study are: (1) to acquire and analyze broadband DRS spectra from normal and OPMD oral tissues; (2) to identify the most relevant spectral features contributing to tissue differentiation; and (3) to evaluate classification performance using both the entire broadband spectra and the optimised feature set. Overall, this work aims to establish a non-invasive, real-time, and cost-effective diagnostic framework to improve early detection and clinical management of OPMD.

2 Materials and methods

2.1 Clinical study design

This *in vivo* study involved 82 participants - 22 healthy controls and 60 OPMD patients with lesions present in different anatomical locations in the oral cavity. Two of the OPMD participants had multifocal lesions. Ethical approval was granted by the Cork Research Ethics Committee

(CREC) of University College Cork (UCC) according to ref.: ECM 4 (y) 13/4/2021 & ECM 3 (kkkkk) 09/05/2023. The research was conducted according to the CREC ethical guidelines and 1964 Helsinki Declaration and its subsequent amendments. Written informed consent, clinical and demographic information, including diagnosis, medical history, smoking and alcohol status were obtained from all participants. Participants were advised to refrain from food or drinks (except water) at least an hour before data collection. They also all rinsed their mouths with water for one minute before the examination. A dental surgeon confirmed the general condition of the oral cavity, including the identification of lesions and healthy tissue. During this standard procedure, biopsy and histopathology were considered for confirmaing the disease status. DRS measurements were conducted before any biopsies were collected.

2.2 Measurements workflow

For data collection, the clinician gently dried the tissue with sterilized gauze and carefully positioned the probe on the tissue. Data was collected from six tissue types, including the cheek mucosa (CM), lips, gingiva (Gum), dorsal and ventral surface of the tongue (DST and VST, respectively) and anterior floor of mouth (AFM) from the control (H) and OPMD groups. The tissues in OPMD group were further subdivided into lesion (OPMD) and contralateral site (OPMDH - confirmed by the clinician). Initially, to optimize the procedure and keep the overall measurement time to a minimum, 9 spectra were recorded from each tissue for the first 17 participants. For the remaining patients, we aimed to collect 15 spectra from each tissue, with the actual number depending on the condition of the oral cavity, the availability of contralateral tissue, and the comfort of the participant. In total, 7409 spectra were collected during the study from six locations/tissue types. The total number of DRS measurements per tissue type is summarized in Table 1.

Table 1 Total number of DRS spectra from each tissue type.

Tissue	CM	VST	DST	Lip	Gum	AFM
Total number of spectra	1386	1376	1215	1233	1125	1074
Total number of lesions	28	19	7	6	2	0
Number of OPMD spectra	375	258	102	78	33	0
Number of OPMDH spectra	777	798	789	831	768	750
Number of H spectra	324	320	324	324	324	324

For classification purposes, data collected from all tissues were labelled into the following three categories: H – all measurements from the control group, OPMD – all lesions from the OPMD group, OPMDH – all healthy tissues from the OPMD group, including the contralateral site. For this preliminary study, we focused only on CM and VST tissues, as the number of lesions in the other anatomical locations was insufficient for comprehensive analysis, but will be included in future studies. A total of 1056 spectra from CM (324 H, 375 OPMDH and 357 OPMDH) and 836 spectra from VST (320 H, 258 OPMD and 258 OPMDH) were analysed. Although OPMDH spectra were collected from all healthy-appearing tissues in the OPMD group, only the contralateral site corresponding to the lesion location (e.g., the opposite side of CM or VST) was considered here to ensure anatomical comparability.

2.3 Diffuse reflectance spectroscopic system and probe design

DRS system (Figure 2) used for this *in vivo* study is based on a highly sensitive spectrometer (QE Pro, Ocean Insight) ranging from 355 nm to 1100 nm with a 1024 x 58 element CCD array. The tissue was illuminated by a tungsten-halogen broadband light source (HL-2000-HP, FHSA Ocean Insight), with an emission wavelength range from 360 nm to 2400 nm. A low OH 1-to-4 Fan-Out fibre bundle with 400 µm core multimode fibre with linear common end (BFL44LS01- Thorlabs) delivered and collected light from the tissue. Although the probe contained three detector fibres, only the fibre with a 1.2 mm source–detector distance (SDD) was employed for this study. This

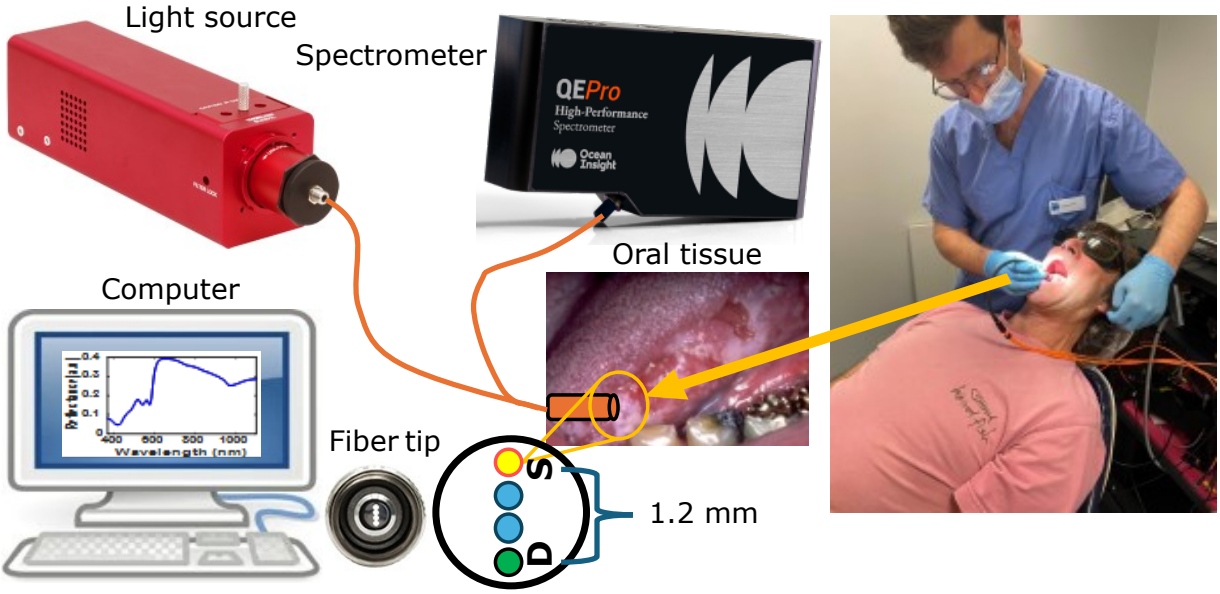


Fig 2 Schematic image of the DRS system and probe configuration.

distance was selected as it provides information regarding the epithelial layer (typically between 400 and 250μm for CM and VST[27,37,38]) and superficial stroma, which are the most relevant for distinguishing healthy tissue from OPMD.

2.4 DRS data collection and preprocessing

Each DRS measurement was recorded for 50 ms and averaged over 20 repetitions. Data was calibrated using a Spectralon reflectance standard (WS-1-SL Diffuse Reflectance Standard, Ocean Insight) and corrected for dark counts (which was measured by recording spectra of reflectance standard with the light source switched off) by using standard reflectance formula given in Eq.1 :

$$R = \frac{I_{tissue} - I_{background}}{I_{standard} - I_{background}} \tag{1}$$

where R is the calibrated reflectance, I_{tissue} is the intensity of the reflectance collected from tissue, $I_{background}$ is the intensity of the background with the illumination source off, $I_{standard}$ is the reflectance intensity collected using the standard. The data was truncated to 377 - 1075 nm to avoid

the influence of a poor signal-to-noise ratio related to low input light intensity and detector sensitivity. Subsequently, a second-degree polynomial Savitzky-Golay filter was applied with window size 5 and spectra were binned over 5 pixels to filter out high-frequency noise and approximate the resolution of the instrument; and finally the data was normalized to zero mean and unit variance. The resultant DRS data was saved in the form of an n $\times$ m matrix, where n represents the number of DRS spectra (observations) and m represents the number of variables (wavelengths). To check the statistical difference between groups Kruskal-Wallis test was used in MATLAB.

2.5 DRS feature selection

After data pre-processing, each DRS spectrum consisted of m = 150 wavelengths (features). To reduce dimensionality and identify the most relevant features for classification purposes, the Least Absolute Shrinkage and Selection Operator (LASSO) feature selection model was applied. LASSO minimizes the sum of squared residuals along with an L1 penalty on the coefficients. The L1 penalty shrinks some of the coefficient estimates towards zero, thereby allowing feature selection.[39] The β_j for each predictor variable is estimated by minimizing the following cost function:

$$\text{Cost function} = \sum_{i=1}^{n}(y_{real}^{(i)} - y_{pred}^{(i)})^2 + \alpha \sum_{j=1}^{m} |\beta_j| \tag{2}$$

where, n is the number of DRS spectra, m is the number of features (wavelength), ($y_{real}^{(i)}$ and $y_{pred}^{(i)}$) are the true and predicted values of the ith spectrum. The term $\alpha \sum_{j=1}^{m} \beta_j$ is L1 penalty, β_j is the regression coefficient of j^{th} feature, α is a regularization parameter controlling the strength of the penalty term L1. A larger value of α leads to more pronounced regularization, driving more coefficients close to zero, resulting in the development of a sparser model. Conversely, smaller values of alpha reduce the amount of regularization, allowing more wavelengths to retain

non-zero coefficients and thus increasing the number of features contributing to the model. The α parameter was selected using cross-validation to balance between prediction error and over- or under-fitting. LASSO regression was applied on the scaled dataset and only variables with non-zero coefficients were retained for inclusion in the predictive model. The contribution of each wavelength is quantified through LASSO weights, which guides the development of a classification model.[40]

2.6 Classification models

Following feature selection, both tissues H and OPMD were classified by Linear Discriminate Analysis (LDA), which maximizes the separation between the two groups based on the selected features. In addition, LDA is particularly effective when the two groups have distinct but overlapping properties, as observed in spectral data from healthy and OPMD tissues. The model was validated using the leave-one-out cross-validation (LOOCV) method to ensure robustness and minimize bias. The performance of the LASSO-LDA model was confirmed in the form of a confusion matrix and receiver operator curves (ROC). Data processing, feature selection and classification tasks were performed using Python (version 3.12) with standard scientific libraries including NumPy, Pandas, sciPy, scikit-learn, Seaborn and Matplotlib for data handling, analysis, modelling and visualization.

3 Results and discussion

DRS spectra are sensitive to changes in tissue composition and microstructure in OPMD lesions caused by alterations in light scattering and absorption properties. Figure 3 illustrates the average DRS spectra for CM and VST tissues in three groups: healthy (H), participants with OPMD

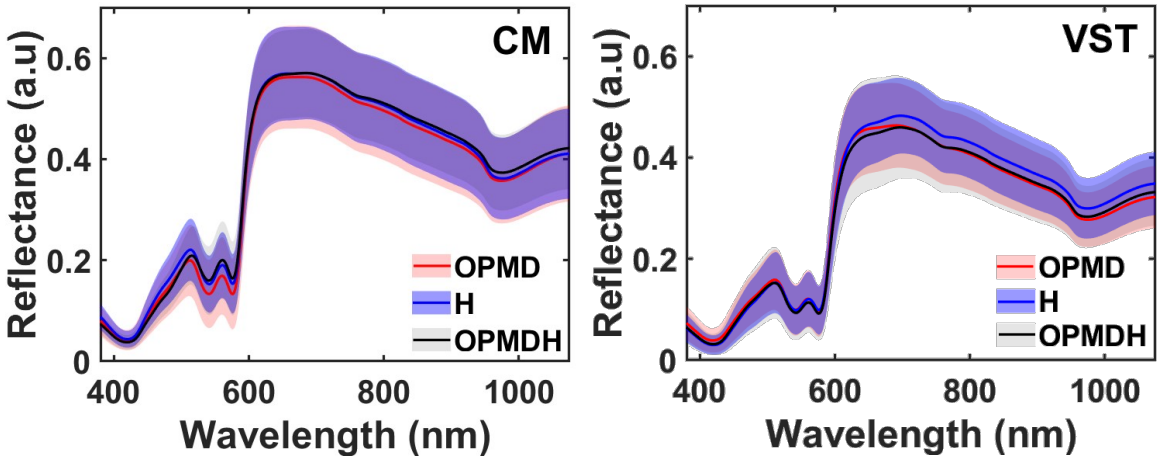


Fig 3 Mean DRS spectra of CM and VST. Blue, red, and black curves correspond to healthy (H), OPMD, and contralateral site (OPMDH), respectively. Shaded areas represent standard deviation across all measurements.

and contralateral site of the lesion (OPMDH), respectively. Although both the CM and VST are considered as nonkeratinized sites, they were analyzed separately due to anatomical differences in epithelial thickness and underlying vascularity.[41] The presented spectra were normalised using the reflectance standard, preserving physically meaningful intensity differences. The standard deviation (shown as shaded areas in the graph) of each group demonstrates high intra- and inter-patient variation related to tissue heterogeneity, typically observed in DRS. All three groups exhibit very similar spectral shapes with small variations in intensities across different wavelength ranges. In general, CM tissue exhibits a higher reflectance than VST, indicating a higher attenuation of light in the latter, which can be a result of thinner epithelial layer and stronger absorption in the underlying stroma. Intensity-wise, the mean DRS spectrum of the contralateral side (OPMDH) resembles more healthy tissue in the case of CM, and that of OPMD tissue in the case of VST.

The DRS spectrum below 600 nm is dominated by haemoglobin absorption, which is an important biomarker for OPMD progression due to frequent neoangiogenesis and changes in the oxygen consumption and delivery occurring within the lesion. For CM, healthy tissue in this spectral re-

gion consistently exhibits a higher reflectance, while the OPMD group typically has the lowest reflectance, indicating increased blood content and vascularization in the lesion. In contrast, for VST, OPMD exhibited the highest reflectance in the 400–500 nm region compared to H and OPMDH, while in the region of 500-600 nm all three groups overlap. This relative change in OPMD reflectance likely exhibits site-specific differences in scattering and absorption, potentially linked to altered epithelial morphology in tongue lesions. To quantify these differences, the ratio of reflectance values at 540 and 575 nm ($R_{540/575}$), representative of relative haemoglobin absorption, was calculated for both CM and VST as shown in figure 4. Although both peaks are related to oxyhaemoglobin, their relative intensities change depending on the balance between oxy- and deoxyhaemoglobin, making them a sensitive indicator of tissue oxygenation status. Changes in this ratio can therefore highlight vascular and metabolic alterations associated with the progression of OPMD. In CM, OPMD lesions showed the highest $R_{540/575}$ compared to H and OPMDH, indicating relative change in oxygenation and increased blood volume within the lesion, which is in agreement with previous studies.[30,35,42] A similar but more pronounced trend was also observed in VST, with OPMD exhibiting the highest ratio. OPMDH exhibits intermediate behaviour suggesting subtle changes in haemoglobin absorption and oxygenation in the clinically normal-appearing contralateral site. This could be a sign of early vascular changes in OPMDH and aligns with the concept of field cancerization, where molecular and histopathological changes in the normal-appearing mucosa adjacent to oral lesions may predispose these areas to the development of future lesions.[43–45] For instance, Siar et al. reported that patients with OPMD are at greater risk of developing a second lesion, most probably in the contralateral anatomical site.[43] Another study reported that half of the patients with OPMD were more vulnerable to field change and had an increased risk of developing second or multiple lesions.[46] Collectively, these results indicate that subtle vascular

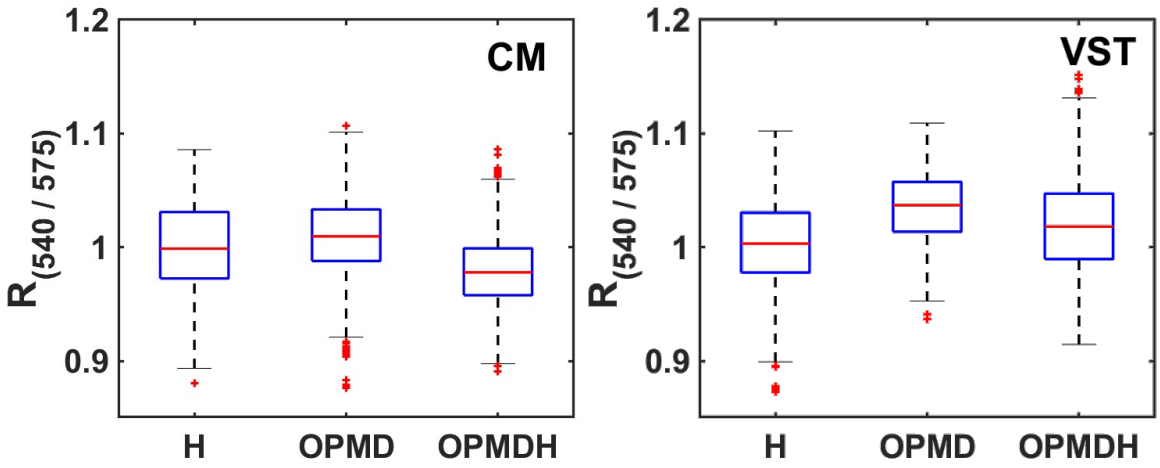


Fig 4 Boxplots of the reflectance ratio $R_{540/575}$ for CM (left) and VST (right) tissues. OPMD lesions show higher ratios compared to healthy and contralateral sites, indicating increased oxyhaemoglobin absorption and suggesting vascular and metabolic alterations linked to early malignant transformation.

and oxygenation changes in the contralateral site may reflect early or preclinical changes associated with the formation of lesion. These findings highlight the potential of DRS to detect early, spatially-resolved tissue alterations that are not visually apparent.[25,30]

Beyond the oxyhemoglobin absorption peaks, distinct site- and tissue-specific changes were observed in the near-infrared (NIR) region from 600 to 1075 nm. In this region, absorption is on average lower and dominated by water,[47,48] making reflectance more sensitive to scattering changes and increasing the penetration depth.[49] Although OPMDH is more similar to H in the case of CM tissue, it consistently exhibited higher reflectance compared to OPMD and H, which confirms the underlying structural and compositional changes in the contralateral tissue even in the absence of visible lesions. OPMD was lower than H throughout this region. Higher keratinization and thicker epithelial layer in CM[50] can mask the subtle scattering-related changes yielding same reflectance profiles for OPMD and healthy tissue. However, the reduction could be due to scattering from changes in stromal composition, microvascular density consistent with previous studies showing altered scattering in premalignant oral tissues.[51] In VST, group-specific optical changes were more

prominent in the NIR region. Healthy tissue exhibited the highest reflectance throughout the entire spectral range above 620 nm. OPMDH have a trend similar to OPMD, indicating subtle alterations in the contralateral site relative to both lesions and healthy controls confirming the field cancerization phenomena once again reported in oral premalignant lesions.[43,46] The changes in OPMD tissue likely arise from localized variations in haemoglobin concentration, microvascularization and structural changes in the epithelium and stroma.[24] The subtle dip around 760 nm in both tissues is associated with deoxyhaemoglobin absorption and can be observed consistently in both tissues. A subtle flattening of the 760 nm dip was observed in CM OPMD tissue compared to H tissue, indicating altered deoxyhaemoglobin absorption or changes in tissue scattering associated with the lesion.".[52] The dip at 760 nm points to localized hypoxia or inefficient vascular function due to immature tumour-associated vasculature and suggests an imbalance in oxygen supply and demand.[53]

These spectral differences and the ratio validate key DRS spectral regions to differentiate OPMD from healthy tissue. However, to move beyond visual assessment and ensure objectivity and scalability, a systematic approach is required to determine diagnostically relevant spectral regions. To this end, we employed LASSO for spectral feature selection, which allowed us to identify wavelengths that contribute most significantly to distinguishing spectra between OPMD and healthy tissue. Figure 5 represents the LASSO-selected features overlaid on averaged DRS spectra for H vs OPMD tissues with their respective LASSO weights shown as a bar graph on the right. The algorithm selected 14 most important features for the classification of CM and VST. These features were selected using regularization parameter alpha ($\alpha = 0.005$) for both CM and VST. The LASSO weights highlight the importance of each selected feature for healthy vs OPMD classification. In both tissues, LASSO-selected features correspond to spectral regions associated

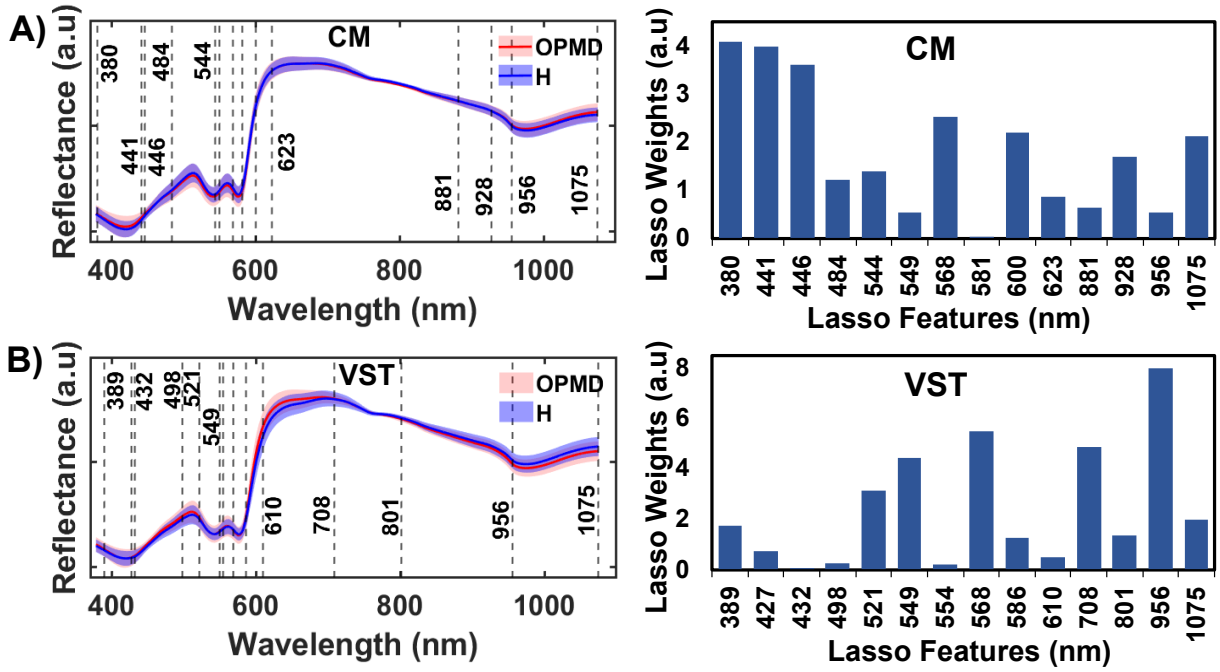


Fig 5 LASSO selected features projected on scaled average DRS spectra of H (blue) and OPMD (red) of A) CM and B) VST with their respective LASSO weights shown as a bar plot on the right.

with haemoglobin absorption, oxygenation status and scattering properties. In particular, in CM the highest weights were achieved for wavelengths in the short wavelength region (377 - 500 nm). On the other hand, for VST bands related to oxyhemoglobin in the 500-600 nm range and water absorption peak appear dominant. Several of the features with varying weights were selected for both tissue types, such as 549 nm, 568 nm, 956 nm and 1075 nm. The selection of features around these wavelengths reinforces the hypothesis that vascular changes and oxygen metabolism are key to distinguishing OPMD from H tissue, regardless of the anatomical location. Additionally, the features selected between 700 – 1100 nm suggest changes in the scattering properties of the deeper tissue layers, potentially indicating malignant progression.[30,42,47,49] The ability of LASSO to identify and prioritize wavelengths provides biological interpretability and guidance for designing compact optical devices for clinical use. For example, a portable multispectral imaging system with discrete bands instead of the full broadband spectrum can be designed for rapid screening

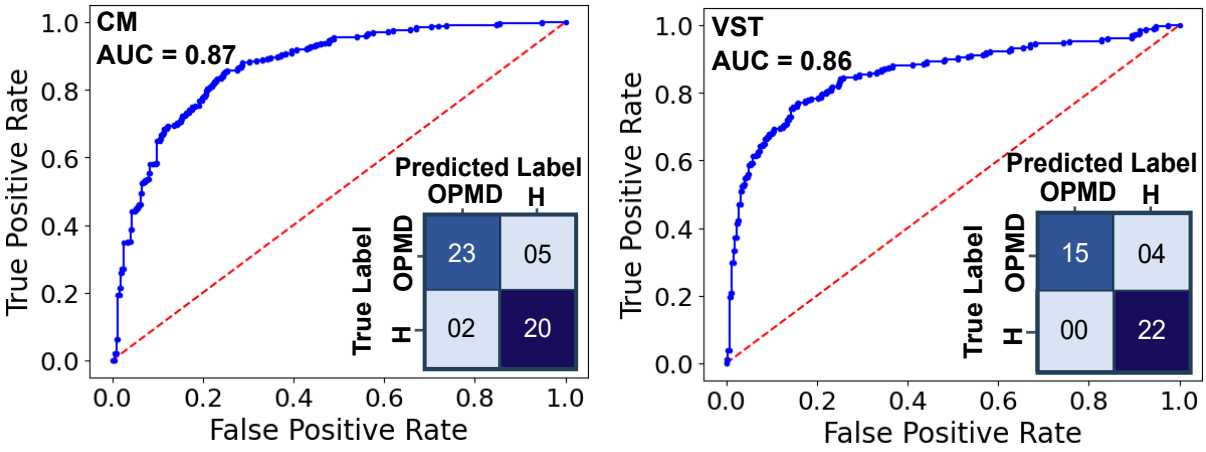


Fig 6 Confusion matrix and ROC curve for CM (left) and VST (right) for LASSO selected features.

procedures.[54] Some of the selected wavelengths are closely spaced (e.g. 441 and 446 nm for CM), making them even more suitable as a choice for narrow-band approach.

Following the LASSO feature selection, LDA was applied to classify healthy (H) and OPMD tissue using a leave-one-patient-out approach achieving an accuracy of 87% and 89% for CM and VST, respectively. The corresponding confusion matrices and receiver operator characteristic (ROC) curves shown in Figure 6 validate the performance of the model. The ROC curve indicates an excellent classification ability with area under the curve (AUC) of 0.87 and 0.86 for CM and VST, respectively, underscoring the potential of DRS for OPMD screening and diagnosis. Table 2 details the classification performance using sensitivity, specificity and accuracy at both patient and spectral levels for CM and VST. For CM, LASSO selected features outperformed full-spectrum analysis, achieving an accuracy of 87%, compared to 81% for all wavelengths. In VST, the accuracy for LASSO features reached 89% compared to 77% for the entire wavelength range. These results indicate that full spectrum analysis leads to over-fitting of the model and poor generalization capability, while classification on LASSO-selected features of importance yields more robust results. This underscores the potential of DRS and feature selection for OPMD screening proce-

Table 2 Comparison of the classification results for CM and VST tissues.

Tissue	Classification input	Sensitivity	Specificity	Accuracy by spectra	Balanced accuracy by patient
CM	LASSO features	82.2 %	90.9 %	79.0 %	87.0 %
	All wavelengths	89.3 %	72.7 %	80.8 %	81.0 %
VST	LASSO features	79.0 %	100 %	81.2 %	89.0 %
	All wavelengths	68.4 %	86.4 %	76.9 %	77.0 %

dures.

To further understand the physical aspect of the LASSO-selected features and their behaviour in the three studied groups, boxplots were generated for H, OPMD, and OPMDH (figure 7) for all 14 features. These plots illustrate the measurement distribution and differences between the lesion, healthy, and contralateral sites for CM and VST. While the model was trained on H and OPMD data, examining OPMDH spectra at the selected wavelengths allows for assessment of field effects in normal-appearing contralateral site and provides insight into early optical changes associated with OPMD. Many of the selected features demonstrate significant differences between classes. In the case of CM tissue, all peaks selected between 377 - 600 nm show significant differences between H and OPMD, with some also showing significant differences for H and OPMDH. In the case of VST, only wavelengths up to 430 nm and above 800 nm were statistically significant. However, some features with high LASSO weights did not show large absolute differences between H and OPMD, and in contrast, some of the features with low LASSO scores turned out to be significant. This indicates that LASSO captures subtle multivariate contributions rather than relying solely on univariate significance. The behaviour of OPMDH group doesn't follow a clear pattern; it trends closer to H in some spectral ranges, while in others aligns more with OPMD. This variability underscores the heterogeneous nature of contralateral tissue,[46] which may carry some optical signatures of field changes without a visible lesion.[55,56]

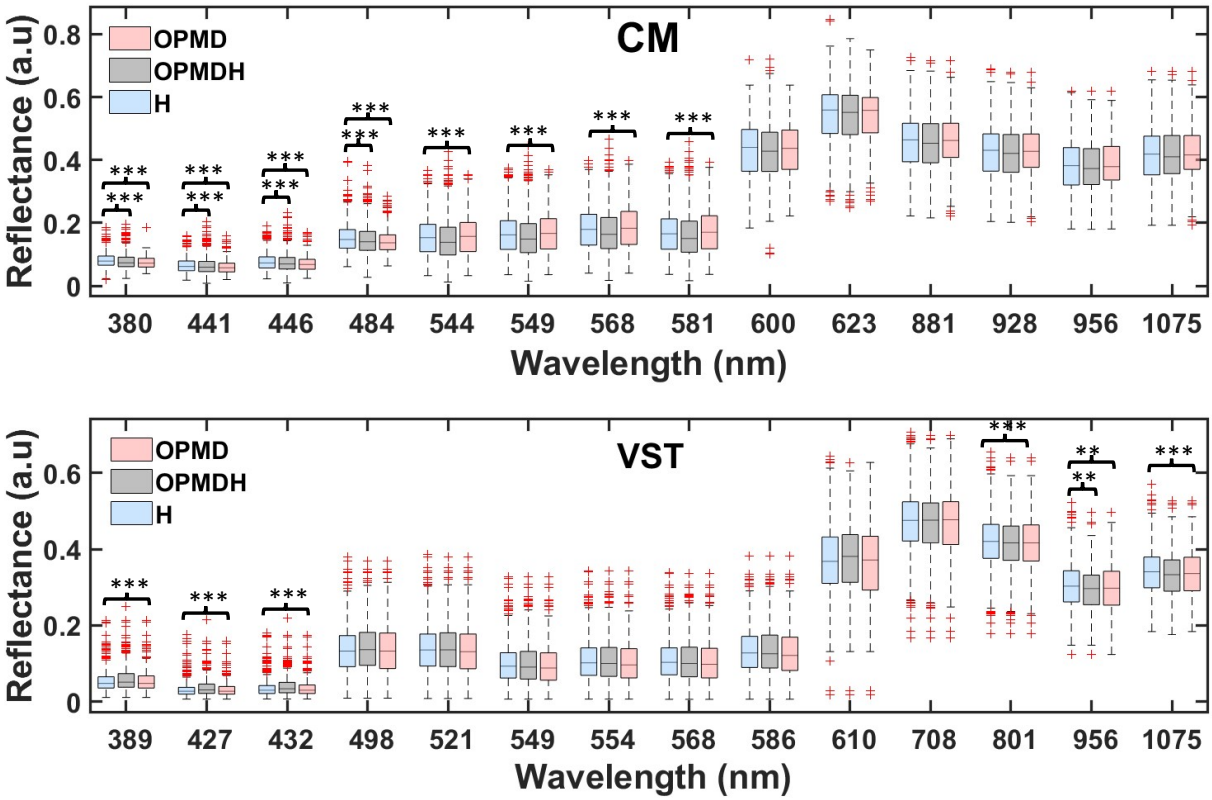


Fig 7 Boxplots of LASSO-selected spectral features for CM (top) and VST (bottom). Each box represents the interquartile range, the horizontal line indicates the median, whiskers show the minimum and maximum values, while '+' sign represents outliers.

Although most diagnostic signatures were within the visible spectrum, DRS captures subtle optical differences in tissue that are not visible to the human eye. These variations arising from early changes in oxygenation, vascularization and tissue scattering, can precede the morphological alterations, allowing DRS to detect early dysplastic transformations even in clinically normal-appearing tissue. In addition, unlike visual examination, DRS provides standardized and reproducible measurements that support consistent interpretation across users and clinical settings. This objectivity is particularly valuable in improving diagnostic yield by guiding biopsy toward optically abnormal sub-regions within heterogeneous lesions.

4 Conclusion

This study demonstrates the potential of DRS as a valuable tool for OPMD screening. By applying Lasso feature selection and Linear Discriminant Analysis (LDA), key spectral features that could distinguish OPMD tissue from healthy tissue were identified. The results show that DRS, particularly with selected features, offers higher sensitivity and accuracy in classifying OPMD in cheek mucosa (CM) and ventral surface of the tongue (VST), showcasing its clinical importance. This approach also identifies optical changes in normal-appearing contralateral tissue (OPMDH) and its heterogeneous nature, suggesting subtle field changes that may precede overt lesions. Spectral differences were linked to vascular and structural changes such as altered oxygenation and increased haemoglobin absorption, all associated with OPMD progression. With machine-learning-driven feature selection, DRS offers a real-time, non-invasive approach complementing traditional biopsy and enabling regular in vivo screening and clinical decision-making.

Future work will focus on incorporating additional tissue sites and a larger patient cohort, including different OPMD stages to validate the robustness and generalizability of the OPMD screening model. Refining of the machine learning algorithms by using more advanced feature selection methods and cross-validation techniques. Building on the parallel Raman study conducted on the same patient cohort and integrating DRS with Raman spectroscopy could enhance diagnostic precision through a multimodal approach. DRS can also be used in margin detection during clinical procedures where preserving healthy tissue is critical.

Acknowledgements

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant number SFI/15/RP/2828 and SFI/22/RP-2TF/10293 and

EUH2020 MSCA-ITN project 'Photonics for healthcare: multiscale cancer diagnosis and therapy' (PHAST-ETN) (Grant No. 860185). For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

1 D. I. Conway, M. Purkayastha, and I. Chestnutt, "The changing epidemiology of oral cancer: definitions, trends, and risk factors," *British dental journal* **225**(9), 867–873 (2018).

2 E. P. Tranby, L. J. Heaton, S. L. Tomar, *et al.*, "Oral cancer prevalence, mortality, and costs in medicaid and commercial insurance claims data," *Cancer Epidemiology, Biomarkers & Prevention* **31**(9), 1849–1857 (2022).

3 S. Maryam, M. D. Ghauri, E. Fahy, *et al.*, "Biophotonics diagnostics of oral cancer using Raman spectroscopy," in *European Conference on Biomedical Optics*, 126270B, Optica Publishing Group (2023).

4 A. Bugshan and I. Farooq, "Oral squamous cell carcinoma: metastasis, potentially associated malignant disorders, etiology and recent advancements in diagnosis," *F1000Research* **9** (2020).

5 H. Ram, J. Sarkar, H. Kumar, *et al.*, "Oral cancer: risk factors and molecular pathogenesis," *Journal of maxillofacial and oral surgery* **10**, 132–137 (2011).

6 S. Wang, M. Yang, R. Li, *et al.*, "Current advances in noninvasive methods for the diagnosis of oral squamous cell carcinoma: a review," *European journal of medical research* **28**(1), 53 (2023).

7 S. Mentel, K. Gallo, O. Wagendorf, *et al.*, "Prediction of oral squamous cell carcinoma based on machine learning of breath samples: a prospective controlled study," *BMC Oral Health* **21**, 1–12 (2021).

8 N. Dharmawardana, C. Woods, D. I. Watson, *et al.*, "A review of breath analysis techniques in head and neck cancer," *Oral oncology* **104**, 104654 (2020).

9 S. Y. Oh, S.-M. Kang, S. H. Kang, *et al.*, "Potential salivary mrna biomarkers for early detection of oral cancer," *Journal of clinical medicine* **9**(1), 243 (2020).

10 R. Lawrence, M. Watters, C. R. Davies, *et al.*, "Circulating tumour cells for early detection of clinically relevant cancer," *Nature Reviews Clinical Oncology* **20**(7), 487–500 (2023).

11 P. Rajendran, R. Sekar, H. A. Zahra, *et al.*, "Salivaomics to decode non-coding rnas in oral cancer. a narrative review," *Non-coding RNA research* **8**(3), 376–384 (2023).

12 S. Maryam, M. S. Nogueira, R. Gautam, *et al.*, "Label-free optical spectroscopy for early detection of oral cancer," *Diagnostics* **12**(12), 2896 (2022).

13 C. Chen, S. Zhao, A. Karnad, *et al.*, "The biology and role of cd44 in cancer progression: therapeutic implications," *Journal of hematology & oncology* **11**, 1–23 (2018).

14 P. Adusumilli, S. Babburi, A. Venigalla, *et al.*, "Estimation of salivary and serum cyfra 21-1 levels in patients with oral squamous cell carcinoma," *Journal of Oral and Maxillofacial Pathology* **27**(1), 98–102 (2023).

15 A. Ahmad, U. B. Ghaffar, S. D. Hiware, *et al.*, "A prospective study for assessment of saliva ca-125 level in oral cancer patients," *Age* **20**(30), 7 (2021).

16 S. R. Gosavi and A. A. Torkadi, "Serum c-reactive protein in oral submucous fibrosis and

oral squamous cell carcinoma: A cross-sectional study," *Journal of Oral and Maxillofacial Pathology* **24**(1), 46–51 (2020).

17 D. Qin, "Next-generation sequencing and its clinical application," *Cancer biology & medicine* **16**(1), 4 (2019).

18 T. Sajan, S. Murthy, R. Krishnankutty, *et al.*, "A rapid, early detection of oral squamous cell carcinoma: Real time PCR based detection of tetranectin," *Molecular Biology Research Communications* **8**(1), 33 (2019).

19 P. Singh, J. K. Verma, and J. K. Singh, "Validation of salivary markers, il-1β, IL-8 and Lgals3bp for detection of oral squamous cell carcinoma in an indian population," *Scientific reports* **10**(1), 7365 (2020).

20 G. Banavar, O. Ogundijo, C. Julian, *et al.*, "Detecting salivary host and microbiome rna signature for aiding diagnosis of oral and throat cancer," *Oral Oncology* **145**, 106480 (2023).

21 "Cancerdetect test for oral & throat cancer."

22 J. M. Yoshizawa, C. A. Schafer, J. J. Schafer, *et al.*, "Salivary biomarkers: toward future clinical and diagnostic utilities," *Clinical microbiology reviews* **26**(4), 781–791 (2013).

23 D. Bastías, A. Maturana, C. Marin, *et al.*, "Salivary biomarkers for oral cancer detection: an exploratory systematic review," *International Journal of Molecular Sciences* **25**(5), 2634 (2024).

24 A. Amelink, O. Kaspers, H. Sterenborg, *et al.*, "Non-invasive measurement of the morphology and physiology of oral mucosa by use of optical spectroscopy," *Oral oncology* **44**(1), 65–71 (2008).

421 25 D. V. Messadi, F. S. Younai, H.-H. Liu, *et al.*, "The clinical effectiveness of reflectance optical
422 spectroscopy for the in vivo diagnosis of oral lesions," *International journal of oral science*
423 **6**(3), 162–167 (2014).
424 26 J. L. Jayanthi, N. Subhash, M. Stephen, *et al.*, "Comparative evaluation of the diagnostic
425 performance of autofluorescence and diffuse reflectance in oral cancer detection: a clinical
426 study," *Journal of biophotonics* **4**(10), 696–706 (2011).
427 27 D. Liu, J. Zheng, Q. Zhang, *et al.*, "A combined autofluorescence and diffuse reflectance
428 spectroscopy for mucosa tissue diagnosis: Dual-distance system and data-driven decision,"
429 *Journal of Biophotonics* **16**(10), e202300086 (2023).
430 28 D. Sharma, J. Rimal, I. K. Maharjan, *et al.*, "Evaluation of oral potentially malignant dis-
431 orders with autoflorescence, reflectance spectroscopy and vital staining and their correlation
432 with histopathology–hospital based prospective study," *Oral Oncology* **118**, 105312 (2021).
433 29 E. Fahy, *Diffuse reflectance spectroscopy in the identification of oral potentially malignant*
434 *disorders*. PhD thesis, University College Cork (2024).
435 30 J. Jayanthi, G. Nisha, S. Manju, *et al.*, "Diffuse reflectance spectroscopy: diagnostic accuracy
436 of a non-invasive screening technique for early detection of malignant changes in the oral
437 cavity," *BMJ open* **1**(1), e000071 (2011).
438 31 A. V. Kolpakov, A. A. Moshkova, E. V. Melikhova, *et al.*, "Diffuse reflectance spectroscopy
439 of the oral mucosa: In vivo experimental validation of the precancerous lesions early detec-
440 tion possibility," *Diagnostics* **13**(9), 1633 (2023).
441 32 L. Zhang, Q. Chang, Q. Zhang, *et al.*, "Sub-diffuse reflectance spectroscopy combined with

machine learning method for oral mucosal disease identification," *Lasers in Surgery and Medicine* **57**(4), 339–351 (2025).

33 S. G. Brouwer de Koning, E. J. Baltussen, M. B. Karakullukcu, *et al.*, "Toward complete oral cavity cancer resection using a handheld diffuse reflectance spectroscopy probe," *Journal of biomedical optics* **23**(12), 121611–121611 (2018).

34 N. Subhash, S. Anand, R. Prasanna, *et al.*, "Bimodal multispectral imaging system with cloud-based machine learning algorithm for real-time screening and detection of oral potentially malignant lesions and biopsy guidance," *Journal of Biomedical Optics* **26**(8), 086003–086003 (2021).

35 R. Mallia, S. S. Thomas, A. Mathews, *et al.*, "Oxygenated hemoglobin diffuse reflectance ratio for in vivo detection of oral pre-cancer," *Journal of biomedical optics* **13**(4), 041306–041306 (2008).

36 N. Subhash, J. Mallia, S. S. Thomas, *et al.*, "Oral cancer detection using diffuse reflectance spectral ratio r540/ r575 of oxygenated hemoglobin bands," *Journal of biomedical optics* **11**(1), 014018–014018 (2006).

37 D. Di Stasio, D. Lauritano, H. Iquebal, *et al.*, "Measurement of oral epithelial thickness by optical coherence tomography," *Diagnostics* **9**(3), 90 (2019).

38 J. Wang, Y. Xu, and S. A. Boppart, "Review of optical coherence tomography in oncology," *Journal of biomedical optics* **22**(12), 121711–121711 (2017).

39 R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society Series B: Statistical Methodology* **58**(1), 267–288 (1996).

40 R. Gautam, D. Peoples, K. Jansen, *et al.*, "Feature selection and rapid characterization of bloodstains on different substrates," *Applied spectroscopy* **74**(10), 1238–1251 (2020).

41 S. McGee, V. Mardirossian, A. Elackattu, *et al.*, "Anatomy-based algorithms for detecting oral cancer using reflectance and fluorescence spectroscopy," *Annals of Otology, Rhinology & Laryngology* **118**(11), 817–826 (2009).

42 P. Chandra Sekhar, J. Betsy, J. Presanthila, *et al.*, "Discrimination of periodontal diseases using diffuse reflectance spectral intensity ratios," *Journal of Biomedical Optics* **17**(2), 027001–027001 (2012).

43 C. H. Siar, M. C. Mah, and P. P. Gill, "Risk of the contralateral mucosa in patients with oral potentially malignant disorders," *Asian Pac J Cancer Prev* **12**(3), 631–5 (2011).

44 M. Mohan and N. Jagannathan, "Oral field cancerization: an update on current concepts," *Oncology reviews* **8**(1), 244 (2014).

45 M. Cankovic, M. P. Ilic, N. Vuckovic, *et al.*, "The histological characteristics of clinically normal mucosa adjacent to oral cancer," *Journal of cancer research and therapeutics* **9**(2), 240–244 (2013).

46 R. V. Sune, A. D. Indurkar, R. R. Bhowate, *et al.*, "An evaluation of field cancerization in patients with oral cancer by "mirror image" biopsy," *Journal of Datta Meghe Institute of Medical Sciences University* **12**(2), 148–153 (2017).

47 S. L. Jacques, "Optical properties of biological tissues: a review," *Physics in Medicine & Biology* **58**(11), R37 (2013).

48 S. Jacques, "Optical penetration depth of light into skin tissue," (2008).

49 U. Kaniyappan, E. Gnanatheepam, P. Aruna, *et al.*, "Quantification of hemoglobin and its derivatives in oral cancer diagnosis by diffuse reflectance spectroscopy," in *Optical Biopsy XV: Toward Real-Time Spectroscopic Imaging and Diagnosis*, **10060**, 170–175, SPIE (2017).

50 H. Clausen, D. Moe, K. Buschard, *et al.*, "Keratin proteins in human oral mucosa," *Journal of Oral Pathology & Medicine* **15**(1), 36–42 (1986).

51 J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Springer (2006).

52 S. Prahl *et al.*, "Optical absorption of hemoglobin," (1999).

53 D. M. Mancini, L. Bolinger, H. Li, *et al.*, "Validation of near-infrared spectroscopy in humans," *Journal of Applied Physiology* **77**(6), 2740–2747 (1994).

54 G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics* **19**(1), 010901–010901 (2014).

55 M. Peralta-Mamani, A. Terrero-Perez, R. M. A. Tucunduva, *et al.*, "Occurrence of field cancerization in clinically normal oral mucosa: A systematic review and meta-analysis," *Archives of Oral Biology* **143**, 105544 (2022).

56 B. J. Braakhuis, M. P. Tabor, J. A. Kummer, *et al.*, "A genetic explanation of slaughter's concept of field cancerization: evidence and clinical implications," *Cancer research* **63**(8), 1727–1730 (2003).

List of Figures

List of Tables

Paper VI

scientific reports

Check for updates

OPEN

Evaluation of wavelength ranges and tissue depth probed by diffuse reflectance spectroscopy for colorectal cancer detection

Marcelo Saito Nogueira[1,2]✉, Siddra Maryam[1,2], Michael Amissah[1,2], Huihui Lu[1], Noel Lynch[3], Shane Killeen[3], Micheal O'Riordain[3] & Stefan Andersson-Engels[1,2]

Colorectal cancer (CRC) is the third most common type of cancer worldwide and the second most deadly. Recent research efforts have focused on developing non-invasive techniques for CRC detection. In this study, we evaluated the diagnostic capabilities of diffuse reflectance spectroscopy (DRS) for CRC detection by building 6 classification models based on support vector machines (SVMs). Our dataset consists of 2889 diffuse reflectance spectra collected from freshly excised ex vivo tissues of 47 patients over wavelengths ranging from 350 and 1919 nm with source-detector distances of 630-µm and 2500-µm to probe different depths. Quadratic SVMs were used and performance was evaluated using twofold cross-validation on 10 iterations of randomized training and test sets. We achieved (93.5 ± 2.4)% sensitivity, (94.0 ± 1.7)% specificity AUC by probing the superficial colorectal tissue and (96.1 ± 1.8)% sensitivity, (95.7 ± 0.6)% specificity AUC by sampling deeper tissue layers. To the best of our knowledge, this is the first DRS study to investigate the potential of probing deeper tissue layers using larger SDD probes for CRC detection in the luminal wall. The data analysis showed that using a broader spectrum and longer near-infrared wavelengths can improve the diagnostic accuracy of CRC as well as probing deeper tissue layers.

Colorectal cancer (CRC) is third most common type of cancer worldwide and the second most deadly. CRC comprised 10.2% (1.85 million) of the diagnosed cancer cases and 9.2% (0.88 million) of the cancer related deaths in 2018. In the same year, CRC had the second highest 5-year prevalence (10.9% or 4.79 million of the cases)[1,2]. The number of incident cases is estimated to increase by 63.4% until 2040, whereas the number of deaths is estimated to increase by 71.5% in the same period[3]. The projected increased incidence of CRC and the improvement on patient prognosis generated by early CRC detection[4] has increased the interest in the development of novel techniques to detect precancerous lesions and lesions early in the polyp-carcinoma sequence[5].

Although current screening methods reduce the risk of CRC-associated mortality[6], their effectiveness varies due to issues related to suboptimal screening compliance, limitations of test performance and lack of accessibility[4]. With this in mind, initiatives have been taken to increase the range of diagnostic methods including blood and stool-based tests. The results of these tests show only the potential presence of the cancer and further investigation is required to locate potential tumors. Tumors can be visualized and identified by using minimally-invasive methods such as computed tomographic colonography, double contrast barium enema, and capsule endoscopy[4]. However, colonoscopy with biopsy remains the gold standard diagnostic test for evaluation of colonic disease.

Colonoscopy is predominantly a very safe procedure with a very low risk profile of serious complications but does pose significant inconveniences for patients from logistical perspective of bowel preparation, as the patients need to be accompanied home after sedation. Recent research efforts have focused on developing non-invasive techniques to identify cancerous lesions by comparing the techniques capable of generating tissue classification models closely matching the biopsy results. In particular, optical spectroscopy techniques can provide real-time and cost-effective tissue identification, which can provide more accurate and faster than conventional detection methods such as visualization, palpation, and histology[7–20]. One of these techniques is diffuse reflectance spectroscopy (DRS).

[1]Tyndall National Institute, Lee Maltings, Dyke Parade, Cork, Ireland. [2]Department of Physics, University College Cork, College Road, Cork, Ireland. [3]Department of Surgery, Mercy University Hospital, Cork, Ireland. ✉email: marcelosaitonogueira@gmail.com

Number of patients	47
Gender	
Male	32
Female	15
Age (years)	
Median	69
Minimum	40
Maximum	89
Interquartile range	13.5
Cancer types	
Adenocarcinoma	46
Carcinoma	1
Tumor stage	
pT1	5
pT2	7
pT3	26
pT4	9

Table 1. Patient demographics and tumor characteristics.

DRS is an optical method capable of tissue identification based on the biochemical composition, oxygenation and microstructure of the tissue. DRS technology is based on delivering light to the tissue and capturing the reflected light that travels inside it. The reflected light contains information about the light scattering (related to the tissue microstructure) and absorption (associated with its biomolecular content). The combination of scattering and absorption (optical properties) as well as the geometry of the light source and detector defines the probed depth into the tissue[21]. Since the optical properties vary with the wavelength of the light source, the wavelength range determines the number of biomolecules that can be investigated using DRS.

Previous studies used DRS to distinguish tumor tissue from healthy surrounding tissues in the oral cavity (head and neck cancer)[7], breast[8,13], lung[14], and liver[15,16]. Recent studies have also investigated the detection of colorectal cancer during surgery[9–12,17–20]. In terms of primary colorectal cancer, DRS and related optical spectroscopic techniques have mainly been used to guide colonoscopy by distinguishing the normal mucosa and malignant tissue inside the colon (luminal side). Several of these studies have demonstrated sensitivity and specificity as high as 100% in a limited number of samples and collected spectra[9–12,17–20]. Other optical spectroscopic techniques applied to CRC detection during colonoscopy include fluorescence spectroscopy (FS), time-resolved fluorescence spectroscopy (TR-FS), Raman spectroscopy (RS), Fourier-transform infrared spectroscopy (FTIR)[22–38]. Numerous fluorescence spectroscopy studies achieved both sensitivity and specificity close to 100% for diagnosis[22]. On the other hand, most of the studies performed require contrast agents and/or ultraviolet illumination. This requirement restricts the optical guidance time to the safe limits of ultraviolet exposure and the time contrast agents remain accumulated in the cancerous lesion. Although the diagnostic abilities of label-free FS, TR-FS, RS and FTIR spectroscopy have been investigated for a limited number of samples, the tissue classification performance has also demonstrated sensitivity and specificity up to 100%[23–28,31–38]. Challenges on the repeatability and reproducibility of the classification performance arise from increasing the data collection and analyzing the factors contributing to higher sensitivities and specificities. Factors for DRS include the measured wavelength range as well as the tissue depth investigated by using specific source-to-detector distances (SDD).

Most of the prior studies using DRS and related optical spectroscopic techniques probed only the superficial colorectal tissue by using fiber optic probes with small SDD or collecting hyperspectral images from wide-illuminated tissue areas[39–50]. In addition, previous optical spectroscopy studies have investigated relatively narrow visible and near-infrared wavelength ranges compared to the wavelength range exploited in this study (from 350 to 1919 nm).

In this study, we developed a multivariate analysis model for colorectal cancer detection in ex vivo specimens for future application in colonoscopy settings. We compared its classification performance with previous DRS, elastic scattering spectroscopy, near-infrared spectroscopy and hyperspectral imaging (HSI) studies of other research groups. This performance was evaluated for the broad wavelength range from 350 to 1919 nm as well as ranges used in other studies[39–50]. Furthermore, we investigate the usefulness of the probed tissue depth on cancer identification by using fiber optic probes with SDDs of 630 μm (small SDD for superficial tissue measurements) and 2500 μm (large SDD). Our DRS spectra dataset contained 2889 spectra of freshly excised ex vivo tissues of 47 patients. To the best of our knowledge, this is the first DRS study to investigate the potential of probing deeper tissue layers using larger SDD probes for CRC detection during colonoscopy.

Methodology

Clinical study protocol and research ethics. The study included 47 patients undergoing bowel resection for cancer of the colon or rectum at Mercy University Hospital (Cork, Ireland). Patient demographics are shown in Table 1. The study was performed under the approval of the Clinical Research Ethics Committee of

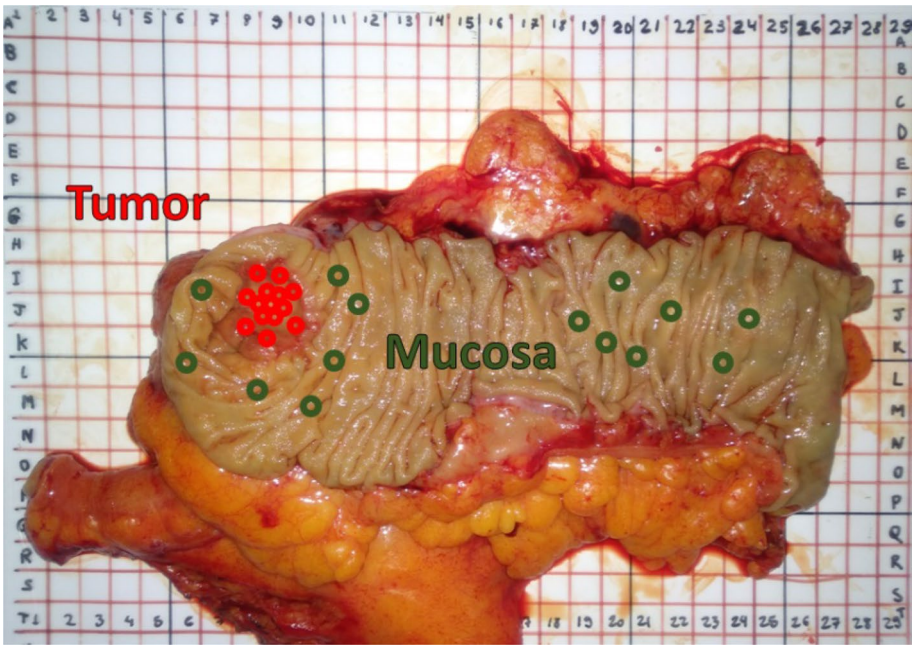


Figure 1. Mucosa and tumor measurement locations on the specimen. Measurements were distributed over a large area in order to obtain a robust dataset containing tissue heterogeneity information.

University College Cork. All methods were performed in accordance with the relevant guidelines/regulations. Informed consent was obtained from all participants of the study. The specimen was removed 15–25 min after the blood supply to the specimen had been cut off. Once the specimen was removed from the patients, its lumen was exposed and cleaned with slow running water in order to remove any remaining feces and blood in the surface of the specimen. Then, healthy and cancerous tissues were identified by experienced surgeons prior to the data collection. The time between the specimen removal and the start of the data collection was 40 min on average. The pH was not stabilized by saline solution. The data was collected from ex vivo mucosal/submucosal tissues and tumors of the specimen within an average time of 60 min after surgical resection. Test conditions were kept as uniform as possible throughout the data collection by keeping the tissue moisture with a wet wipe. Approximately 15 sites were measured for healthy tissues (mucosa/submucosa) and 15 for cancerous tissues over a typical area of 100 cm^2 (Fig. 1).

The location of every measurement was recorded on a photograph taken with a red–green–blue (RGB) camera. After the acquisition of all optical DRS data, the specimen was returned to the Pathology Department for processing and analysis according to standard protocols. The ground truth of every measurement was obtained by histopathology analysis.

Diffuse reflectance spectroscopy system and probes. Our DRS system comprised of a tungsten-halogen broadband light source (HL-2000-HP, Ocean Optics, Edinburgh, United Kingdom) with an emission spectrum ranging from 350 to 2400 nm, which delivers and captures light via optical fibers coupled to both light source and detectors (spectrometers). First, the excitation light is sent through the fiber optic probe. Next, the diffuse-reflected light is collected by optical fibers in the same probe. These fibers deliver the diffuse-reflected light to a visible-wavelength spectrometer (QE-Pro, Ocean Optics, Edinburgh, United Kingdom) and a near-infrared spectrometer (NIR-Quest, Ocean Optics, Edinburgh, United Kingdom). Finally, the light intensity detected by the spectrometers is pre-processed ("Data preprocessing and analysis" section) in order to obtain the tissue reflectance spectra. A schematic drawing of the DRS system is shown in Fig. 2.

In order to obtain the reflectance spectrum preferentially from the superficial tissue (tens of microns deep), we used a quadrifurcated 600-µm-core Low-OH-Silica fiber optic probe (BF46LS01 1-to-4 Fan-Out Bundle, Thorlabs, Munich, Germany) with 630 µm source-to-detector distance (SDD; fiber center-to-center distance represented in Fig. 2). One of the optical fibers was used for illumination, two other fibers were used for collection, and the remaining one was not used during the reflectance measurements. We also used a 2500 µm SDD probe containing one source fiber in the center and 10 collection fibers surrounding it (five for the visible-wavelength detection alternating with five for the near-infrared detection), as seen in Fig. 2. The configuration allows probing deeper tissue layers while achieving higher efficiency on the light collection.

The depth interrogated by each probe was estimated by using a spectral fitting algorithm to extract the optical properties from our DRS measurements. This algorithm was based on forward Monte Carlo simulations of steady-state light transport in multilayered tissues (MCML)[51]. In particular, we used optimized simulations accelerated by graphics processing unit (GPU) reported by Alerstam et al.[52] by considering tissue as a semi-infinite homogeneous medium with the optical properties (absorption coefficient μ_a and scattering coefficient

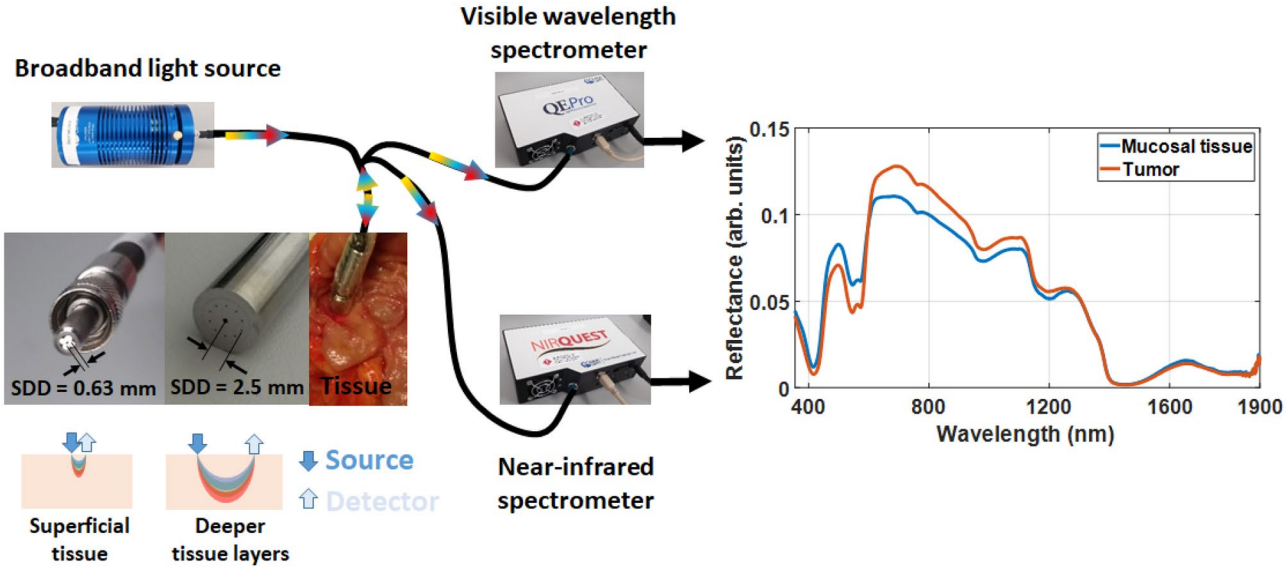


Figure 2. Schematic image of the DRS system. The reflected light can be collected in a wide range of wavelengths and thus, allow the investigation of a larger variety of chromophores in biological tissues.

μ_s) extracted by using a spectral fitting algorithm described in previous studies[12,14,15]. Fluence maps of 5 × 5 mm were generated with simulations using 10 million photon packets, refractive index of the outer medium (η_{out}) of 1, tissue refractive index (η_{rel}) of 1.4, anisotropy factor (g) of 0.9, and 50 µm of radial and depth resolution. Based on these fluence maps, the photon hitting density maps of 3.75 mm × 5 mm were computed by converting cylindric to cartesian coordinates and multiplying the fluence maps at the positions of the source and collection fibers. Then, the position of the maximum photon hitting density value at the mean position between the source and detector was taken as the average probed depth for each wavelength. Then, the minimum and maximum probed depth were reported in this study in order to show the independence of the datasets acquired with each probe and the importance of evaluating the classification performance at each tissue probed volume.

Optical data collection protocol. Prior to and after each run of clinical data collection, we collected the background and reference measurements. The reference measurements were taken on a specialized holder able to keep a fixed distance of 2 mm and 6 mm between the small SDD or large SDD probe and the reflectance standard (FWS-99-01c, Avian Technologies LLC, New London, USA), respectively. The fixed distance allows the comparison of the reflectance data across patients. Prior to each set of tissue measurements, the probes were covered with polyvinyl chloride (PVC) film to avoid possible contamination. The same probes were used for every clinical measurement throughout this study. During the tissue data collection, measurements were performed by positioning the probe 90 degrees from the tissue surface. A total of 1363 spectra were collected for the small SDD probe (630 µm SDD) and 1526 for the large SDD probe (2500 µm SDD).

Data preprocessing and analysis. The data preprocessing comprised of the background subtraction and subsequent division of the captured signal (tissue reflected intensity) by the reflected intensity of the reference (reflectance standard):

$$Reflectance(\lambda) = \frac{Tissue\ reflected\ intensity - Background\ intensity}{Reference\ reflected\ intensity - Background\ intensity}.$$

Once the reflectance spectra were obtained for both visible and near-infrared spectra, these spectra were merged based on the overlapping spectral region between the two spectrometers (from 1095 to 1130 nm). The merging was performed by first doing an interpolation of the overlapping region of the two spectrometers. The interpolated data were used for a weighted sum following:

$$Reflectance(\lambda) = \sum_{i=0}^{100} \frac{\left((100 - i) \times reflectance\ of\ \ VIS\ spectrometer + i \times reflectance\ of\ \ NIR\ spectrometer\right)}{100}$$

The result is a smooth reflectance curve where the reflectance measured by each spectrometer has more contribution in the wavelength regions where they are most sensitive. In order to prepare the reflectance data for classification, the data was centered and scaled/normalized between − 1 and + 1. This scaling ensures the contribution of data at each wavelength is similar for building the classification models used in this study.

Our data analysis consisted of building tissue classification models based on support vector machines (SVMs) and evaluating their accuracy. The training set for the classification was $\{\boldsymbol{x}_i^m, y_i\}_{i=1}^n$ with n samples (e.g. number of reflectance spectra) and with m features (e.g. number of wavelengths). The SVM classifier finds the best

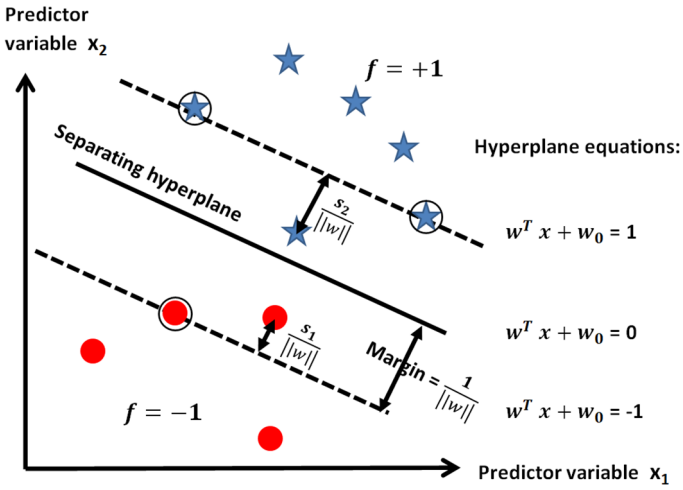


Figure 3. Graphical explanation of SVM hyperplane and accepted margins as defined by parameters involved on the hyperplane equations and slack variables s_i. The stiff condition of hard margins represented in this schematic is relaxed by slack variables, which act as local quantities that add flexibility to the margin.

hyperplane that separates two types of samples described by the class labels $y_i \in \{-1, +1\}$ (e.g. normal mucosa and tumor) based on a predictor variable $\boldsymbol{x}_i^m$ (e.g. reflectance values), where i = 1, 2, …, n. The hyperplane was calculated by:

$$\min_{w,w_0,s_i} \left(\frac{1}{2}||w^2|| + C_i \sum_i s_i \right) \text{subject to } y_i \left(\boldsymbol{w}^T \boldsymbol{x} + w_0 \right) \geq 1 - s_i \text{ and } s_i \geq 0,$$

where $\boldsymbol{w} \in \boldsymbol{R^p}$ and w_0 are parameters that define the hyperplane, C_i is the regularization parameter (or penalty strength) and s_i are the slack variables. s_i was used to measure the level of error accepted on the hyperplane margin, while C defines the penalty for misclassifications. The two parameters together determine the width of the accepted margin on the hyperplane (Fig. 3).

The classification is given by a sign function which determines the side of the hyperplane a given sample falls into:

$$f(\boldsymbol{x}) = sign\left(\sum_{i=1}^{n} y_i C_i K(\boldsymbol{x}, x_i) + w_0 \right)$$

where the $K(x_i, x_j)$ represents the Kernel function:

$$K(x_i, x_j) = y_i K(\boldsymbol{x}, x_i) = \emptyset_i \emptyset_j,$$

where $\emptyset$ is the transformed space for non-linear mapping.

The Kernel function for polynomial kernels is given by:

$$K(x_i, x_j) = (x_i x_j + 1)^p$$

where p is the order of the polynomial. The Kernel function for Gaussian kernels is defined by:

$$K(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2\sigma^2} \right).$$

Classification models based on SVMs were built by using the algorithms available in the classification learner app of MATLAB R2016a. All SVM classifiers used the penalty strength C (or Box constraint level in MATLAB) as 1. Linear, quadratic and cubic SVMs used MATLAB automatic heuristic procedure based on subsampling to select an appropriate Kernel scale factor. The software divides all elements of the predictor matrix (scaled reflectance values) by the Kernel scale factor and applies the kernel norm to compute the Gram matrix of the Kernel function. Fine, medium and coarse Gaussian SVMs used Kernel scale factors of 0.5, 2 and 8, respectively.

The models for the broadest wavelength range of this study (from 350 to 1919 nm) were built by several types of SVM algorithms using linear, quadratic, cubic, fine Gaussian, medium Gaussian and coarse Gaussian Kernel functions. In addition, quadratic SVMs were used to develop models on wavelength ranges covered by previous studies reported in the literature. The performance assessment of all generated classification models took into account their sensitivity, specificity, accuracy and area under the receiver operating characteristic curve (AUC). The performance parameters were obtained by using twofold cross-validation for 10 iterations with random

Type of SVM	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Linear	95.2 ± 1.9	91.8 ± 1.1	93.6 ± 1.5	0.970 ± 0.007
Quadratic	93.5 ± 2.4	94.0 ± 1.7	93.8 ± 2.0	0.971 ± 0.014
Cubic	93.8 ± 0.5	93.3 ± 1.9	93.6 ± 1.2	0.968 ± 0.011
Fine Gaussian	86.4 ± 2.3	90.1 ± 1.3	88.2 ± 2.1	0.960 ± 0.007
Medium Gaussian	90.7 ± 0.3	88.0 ± 1.7	89.4 ± 1.0	0.965 ± 0.005
Coarse Gaussian	80.8 ± 3.7	79.9 ± 2.5	80.4 ± 3.1	0.878 ± 0.023

Table 2. Comparison of the classification performance of SVM classifiers for the dataset of the small SDD probe (n = 1363) containing reflectance values from 350 to 1919 nm.

Type of SVM	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Linear	93.1 ± 0.2	91.9 ± 0.6	92.5 ± 0.4	0.963 ± 0.004
Quadratic	96.1 ± 1.8	95.7 ± 0.6	95.9 ± 1.2	0.987 ± 0.005
Cubic	94.9 ± 0.7	95.5 ± 0.4	95.2 ± 0.5	0.983 ± 0.004
Fine Gaussian	82.4 ± 1.9	95.1 ± 1.3	88.4 ± 1.6	0.970 ± 0.005
Medium Gaussian	94.7 ± 0.8	89.6 ± 0.6	92.3 ± 0.7	0.975 ± 0.005
Coarse Gaussian	90.5 ± 1.5	85.1 ± 0.8	87.9 ± 1.2	0.935 ± 0.005

Table 3. Comparison of the classification performance of SVM classifiers for the dataset of the large SDD probe (n = 1526) containing reflectance values from 350 to 1919 nm.

sampling. In order to do this, the dataset was first separated into training and tests sets of equal size. Next, the model was generated using the training set. Then, the model was applied to classify and validate the test set. Afterwards the test and training sets were switched and the same process repeated. At the end of this process, the output was the mean of each classification performance parameter. Then, the process is repeated 10 times. Then, the mean and standard deviation of the of the output of the 10 iterations were determined and reported in this study. The reproducibility of these parameters was assessed by the obtained standard deviations. Although fivefold cross-validation is the usual type of validation, this study used twofold cross-validation in order to show the patterns used for tissue differentiation in our dataset can be found by using 50% of the data as training set instead of 80%. Twofold cross-validation also allows the model to be tested in a larger dataset compared to fivefold cross-validation, which means that successful classification will occur only if the model is robust enough to describe half of the dataset by using half for training. In this case, successful classification means stronger potential of generalization of the model upon increase in sample size, especially compared to validation results using more than 50% of the dataset for training and less than 50% for testing.

Comparison with previous studies. The classification performance for colorectal cancer detection achieved in this study was compared to similar studies investigating DRS, elastic scattering, near-infrared spectroscopy (reflectance modality) and hyperspectral imaging[39–50]. This comparison involved listing the type of tissue evaluation (in vivo or ex vivo), number of patients, wavelength range, source-to-detector distance, types of tissue analyzed, number of analyzed spectra, tissue types used for classification, and diagnostic performance metrics (sensitivity, specificity, accuracy, and AUC). In order to consider comparable probed depths and biomolecules, the performance metrics of this study took into account the classification using similar SDD and wavelength ranges. The classifier used for all the comparisons was the quadratic SVM.

Results

Tissue classification. In order to choose which classifier would be used for comparison of the performance achieved with different wavelength ranges, we first evaluated classification models using SVMs and the entire wavelength range of the diffuse reflectance spectra. The performance of SVMs with linear, quadratic, cubic, fine Gaussian, medium Gaussian and coarse Gaussian Kernel functions on the dataset were assessed for the small SDD probe ($n_{normal\ tissue} = 728$ and $n_{tumor} = 635$) and the large SDD probe ($n_{normal\ tissue} = 804$ and $n_{tumor} = 722$). The probed depth for the small probe ranges from 300 to 1000 μm, whereas the one for the large probe ranges from 800 to 2000 μm (data not shown). Then, the tissue volume interrogated by each probe is different and can affect the classification performance parameters. Sensitivity, specificity, accuracy, and AUC are shown in Tables 2 and 3.

Table 2 shows that the classification performance (small SDD probe) for the polynomial (linear, quadratic and cubic) SVMs is higher than the Gaussian SVMs. The use of the linear SVM leads to higher sensitivity values, whereas quadratic and cubic SVMs can be used to obtain higher performance metrics in exchange for lower sensitivity. The increase in all performance metrics when comparing the coarse Gaussian SVMs with the fine

Analyzed wavelengths	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
From 350 o 760 nm	84.7 ± 2.4	87.0 ± 2.3	85.8 ± 2.4	0.922 ± 0.014
From 350 to 800 nm	87.4 ± 2.1	87.3 ± 2.7	87.4 ± 2.4	0.933 ± 0.009
From 400 to 440 nm and from 540 to 580 nm	84.2 ± 2.9	86.3 ± 2.2	85.2 ± 2.6	0.917 ± 0.012
From 400 to 1000 nm	90.5 ± 1.2	92.0 ± 0.5	91.2 ± 0.9	0.963 ± 0.004
From 405 to 665 nm	85.6 ± 1.4	90.0 ± 0.7	87.7 ± 1.1	0.930 ± 0.010
From 405 to 750 nm	86.8 ± 0.5	89.0 ± 0.3	87.8 ± 0.4	0.927 ± 0.005
From 470 to 700 nm	82.3 ± 2.2	86.0 ± 0.8	84.1 ± 1.6	0.900 ± 0.008
From 415 to 425 nm, from 495 to 505 nm, from 525 to 545 nm, and wavelengths 465 nm and 625 nm	74.5 ± 0.4	83.0 ± 0.8	78.5 ± 0.6	0.863 ± 0.005
405.235 nm, 406.42 nm, 408.794 nm, 414.736 nm, 422.48 nm, 468.76 nm, 559.489 nm, 577.506 nm, 594.342 nm, 957.948 nm, and 1000.31 nm	78.2 ± 0.7	86.0 ± 0.5	81.9 ± 0.6	0.903 ± 0.005
From 900 to 1300 nm	86.9 ± 3.5	88.0 ± 3.7	87.4 ± 3.6	0.921 ± 0.020
From 1000 to 1919 nm	92.1 ± 1.2	92.3 ± 1.4	92.2 ± 1.3	0.950 ± 0.011
From 350 to 1919 nm	93.5 ± 2.4	94.0 ± 1.7	93.8 ± 2.0	0.971 ± 0.014

Table 4. Performance metrics of each wavelength range analyzed using the quadratic SVM classifier on the dataset of the small SDD probe.

and medium Gaussian SVMs suggests the optimum decision boundary is reasonably narrow. The narrower the boundary is, the higher the specificity and the lower the sensitivity achieved.

Table 3 indicates the classification performance for the large SDD probe has similar trends as that of the small SDD probe in terms of the comparison between polynomial and Gaussian SVMs. On the other hand, the decrease in the performance metrics when decreasing the Kernel scale factors (i.e. using finer Gaussian SVMs) is not as pronounced. In addition, the quadratic SVM model achieved the highest performance metrics. Since the quadratic SVM led to the highest performance for both probes, we used this SVM for comparing the performance metrics with other studies.

The comparison between the best performance metrics of the small and large SDD probes indicated an average gain of 2.6% sensitivity, 1.7% specificity, 2.1% accuracy and 0.16 AUC for tissue classification between 350 and 1919 nm. This suggests probing deeper tissue layers with this wavelength range may improve colorectal cancer detection during colonoscopy.

The performance metrics ((96.1 ± 1.8)% sensitivity, (95.7 ± 0.6)% specificity, (95.9 ± 1.2)% accuracy and 0.987 ± 0.005 AUC) for the large SDD probe are sufficiently high to show potential comparison with other techniques. Furthermore, these metrics are robust as they are based on twofold cross validation of a 1526 spectra dataset. The same robustness is shown for the dataset of the small SDD probe. The classification performance can be further improved by combining DRS with other optical techniques such as fluorescence spectroscopy (FS), time-resolved fluorescence spectroscopy, Raman spectroscopy, Fourier-transform infrared spectroscopy.

Usefulness of the extended wavelength range. Based on the classification performance on the wavelength ranges of previous studies[39–50], we analyzed the potential benefit of using the extended wavelength for colorectal cancer (CRC) detection. In order to be consistent with the findings of previous studies, this analysis consisted of the comparison of the sensitivity, specificity, accuracy and AUC using the small SDD probe (Table 4).

Table 4 shows the performance metrics for the wavelength ranges covering near-infrared wavelengths are higher than those achieved by analyzing only the visible wavelength range. Moreover, the broader and longer the near-infrared range evaluated, the higher the performance is achieved. The assessment of wavelengths from 1000 to 1919 nm showed similar tissue classification performance compared to the entire wavelength range analyzed in this study (between 350 and 1919 nm). This fact could not be observed by solely analyzing the diagnostic ability of previous studies, as comparison across studies was impossible due to insufficient overlap among investigated wavelength ranges.

Discussion

Relevant biochemical and structural differences for CRC detection. From a clinical perspective, biomolecular changes probed by optical techniques[53–79] can potentially obviate the need for multiple biopsies or polypectomies of normal mucosa as well as identify sessile serrated polyps, which may be difficult to recognize during colonoscopy at times[7–20]. The potential of optical spectroscopy for colorectal cancer (CRC) detection in ex vivo specimens or in vivo during colonoscopy has been evaluated for superficial tissues (small SDD probes) in several wavelength ranges[39–50]. To the best of our knowledge, this is the first study that uses DRS information from deeper tissue layers (2500 μm SDD) to detect cancerous tissues in the luminal wall of ex vivo specimens for colonoscopy applications. Even though light penetrates deeper for certain wavelengths, if small SDD probes are used, the collected light is predominantly reflected from the tissue surface (Fig. 4). Figure 4 shows that the probed depth in the near-infrared region where the maximum probed depth is achieved for both small and large probes. The probed depth of the small probe typically varies between 0.5 and 1 mm (> 0.8 mm for most wavelengths), while that of the large probe varies mostly between 0.5 and 1.9 mm (> 1.4 mm for most wavelengths; data not shown).

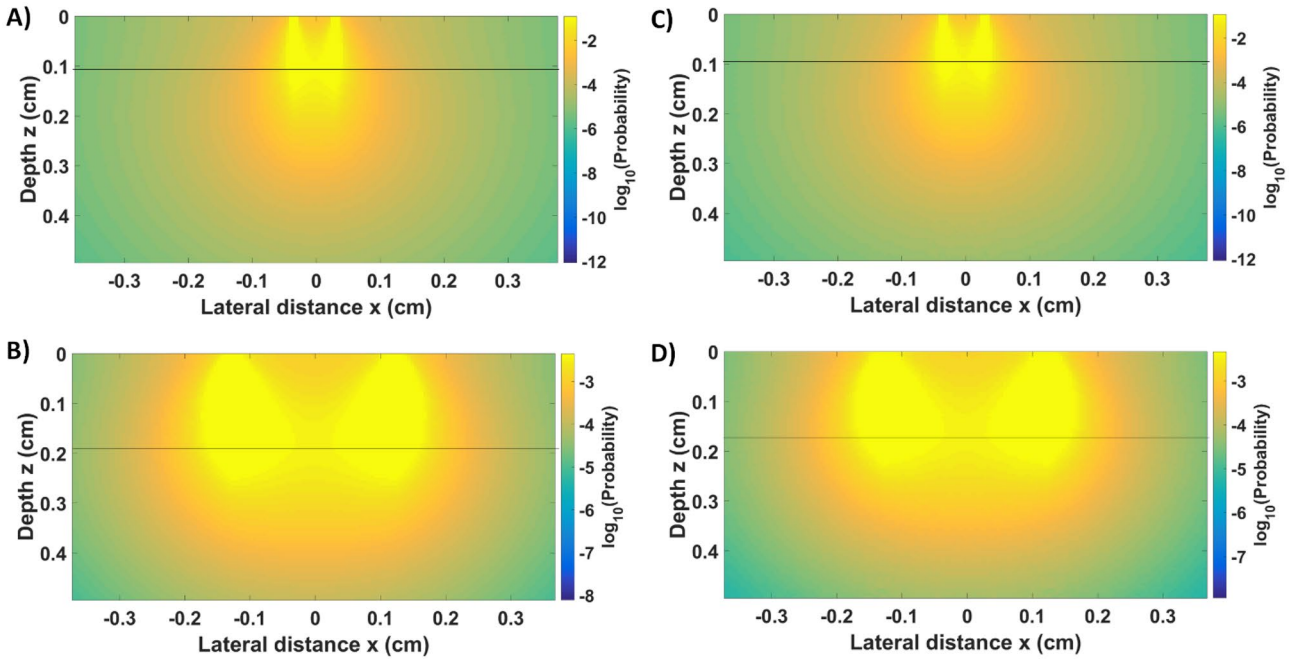


Figure 4. Photon hitting density (PHD) map generated from fluence maps of forward Monte Carlo simulations of light propagation in mucosa (**A**,**B**) tumor (**C**,**D**) tissues at 1150 nm. The black line shows the position of maximum in the middle of the PHD map, which indicates the probed depth at SDD for the (**A**) small probe (630-µm SDD) and (**B**) large probe (2500-µm SDD).

Previous studies used small SDD probes and visible/near-infrared wavelength ranges to investigate signals related to the biochemical composition (related to light absorption) and microstructure (associated to light scattering) of the tissue surface. This biochemical composition includes concentrations of oxyhemoglobin, deoxyhemoglobin, met-hemoglobin, bile, bilirubin, β-carotene, water, lipid, collagen, elastin, and other biomolecules. The tissue microstructure can be investigated due to refractive index mismatches caused by multicellular structures (such as vessels, fibers, etc.), cells size, number of organelles (e.g. mitochondria), composition of tissue layers, extracellular matrix, cell and organelle membranes, collagen fibers and fibrils, and other factors. Once factors such as tissue microstructure, biochemical composition and probed depth are better understood, novel designs of DRS and hyperspectral imaging systems can be proposed to improve the diagnostic accuracy for CRC.

In terms of colonoscopy applications, primary CRC needs to be differentiated from the surrounding normal mucosa (the inner lining of the colon). The intraluminal optical guidance may be affected by other tissue layers apart from the mucosa, especially when using large SDD probes. The colorectal wall is composed of layers of mucosa, submucosa, muscularis propria, and serosa ordered from the intraluminal to the extraluminal side[49]. Each tissue layer differs in their microstructure and biomolecular content which changes the scattering and absorption properties, respectively.

In normal mucosal tissues, the light that enters the colon from the luminal side is first absorbed by mucosal biomolecules (e.g. blood in mucosal capillaries) and scattered by collagen fibers and fibrils, epithelial cells, organelles, and cytoskeleton[80]. The diffuse-reflected light is composed of the scattering within the mucosal layer (Fig. 5) and the backscattered light at the boundary between mucosa and submucosa (refractive index mismatch). The diffuse-transmitted light enters the submucosal layer to be mostly absorbed by water and large submucosal vessels while scattered by the submucosal cellular and subcellular structures. Backscattering happens once more on the interface between the submucosa and muscle layers, which increases the percentage of diffuse-reflected light. In the muscle layer, large collagen bundles can scatter light forward. A small fraction of the light that was not absorbed by the muscle and serosa and backscattered within the muscle layer or in the interface with serosa is diffuse-reflected and may reach the light detector. The detected fraction of diffuse-reflected light of all tissue layers gives the DRS or elastic scattering spectroscopy signal. In tumor tissues, this fraction can be changed upon variation of optical properties resulting from biomolecular and structural associated to carcinogenesis[49].

Most of the precancerous changes on the healthy colorectal mucosa originate in the epithelial tissue[80]. As the cancer progresses, changes due to angiogenesis as well as size and density of collagen fibers can be observed. Angiogenesis generates alterations in blood oxygen saturation and increases the density, volume and disorganization of the blood microvasculature. Cancerous tissues tend to have a high metabolism and an increase in blood volume in the tumor periphery, which can prevent the oxygen and nutrient supply to the center of larger tumors. This lack of oxygen and nutrient supply leads to necrosis and subsequent loss of central microvasculature in late-stage cancers[49]. Most of the tumors probed in this study are of advanced stages which may exhibit more pronounced structural and biomolecular changes. However, these changes are expected to progress over tumor stages and thus, the identification of parameters for optimized early-stage tumor detection requires feasibility studies including tumors of diverse stages such as the present study. Parameters to be optimized include the

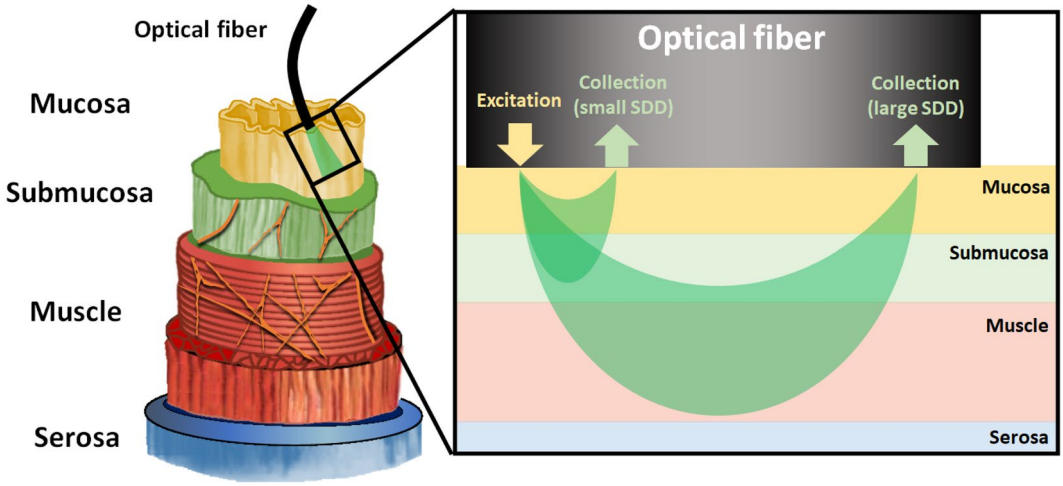


Figure 5. Tissue layers of the colon and rectum (left) and the light-tissue interaction volume interrogated by using the small SDD probe and the large SDD probe. The probed tissue depth increases with the SDD. Variations in the diffuse reflectance signal caused by carcinogenesis were evaluated based on the structural and biomolecular changes of the tissue layers investigated by each probe.

extended wavelength range and larger SDD explored in this study and discussed in more detail in "Wavelength selection and probed depth" section. In addition, DRS and optical spectroscopy are in the process of being introduced as a feasible tool for cancer detection in general, as already discussed in "Introduction" section. The use of DRS and hyperspectral imaging for CRC detection is discussed in the "Wavelength selection and probed depth", "Comparison with DRS-related optical spectroscopy studies" and "Comparison with hyperspectral imaging studies" sections. Limitations due to the dataset based on tumors visible by naked eye are discussed in the "Limitations and validity of this study" section.

In terms of changes to be probed with optical techniques for early-stage cancer detection, Backman and Roy[80] have shown that changes associated with carcinogenesis involve microvessel density in the mucosa (pericryptal network) and the superficial submucosa as well as structural changes on the mucosal crypts and stroma around 200–300 μm from the tissue surface (information from pictures of histology slides shown by the authors). Other suggested optically detectable alterations cited by the authors included nanostructural and microstructural changes in the chromatin structure, collagen fiber crosslinking and cytoskeleton.

Wavelength selection and probed depth. Previous DRS studies targeting cancer detection for colonoscopy guidance reported possible differentiation between luminal normal and premalignant or tumor tissue by using hemoglobin concentration and blood saturation information alone[39,40,81–83]. Similar considerations are made by studies evaluating hyperspectral imaging (HSI) for the same application[41,49,50]. Blood-related parameters are detected on the visible wavelength range, where the blood absorption is stronger (Fig. 6). Prior DRS and HSI studies focused on analysis using visible light, except by Chen et al.[45–47], Ehlen et al.[48] and Yuan et al.[42]. The typically targeted near-infrared wavelength region contains information about the water and lipid absorption. Longer wavelengths or shorter wavenumber also contain information about protein and carbohydrate content. The wavelength ranges used in prior studies are shown in Fig. 6.

In addition to the biochemical information of the wavelength ranges, the DRS signal contains tissue information on depths depending on the optical properties (absorption and scattering coefficients) and source-to-detector distance (SDD). Since the scattering coefficient is large on the visible wavelength range[53], most of the light in this range is diffuse-reflected close to the source of light. Then, using small SDD probes allows efficient collection of visible emitted light and large SDD probes will collect visible light from deeper layers. However, the tissue scattering coefficient is relatively low in the near-infrared wavelength range, which allows the small SDD probes to collect the diffuse-reflected light from the deeper tissue layers, whereas large SDD probes collect light from even deeper layers. Therefore, the DRS probed depth is mainly determined by the SDD of the fiber optic probe.

Although DRS and HSI collect similar tissue information, there are major differences with respect to the probed depth and reflectance standardization. DRS is taken from point measurements in contact with the tissue. In this case, adjusting the probe SDD will control the depth to be probed. Yet, hyperspectral images are taken from a certain tissue area and, thus, each tissue point has a different distance from the detector[85]. This means the tissue reflectance is not standardized, as the standardization process requires every tissue point to be at a fixed distance from both source and detector. Since HSI works by illuminating a tissue area right below the light source and detecting the reflected intensity from the same area, the detected light comes primarily from the tissue surface[85]. Similarly, to the smaller SDD probes for DRS, the wavelength range (and optical properties) will have a strong effect on the probed tissue depth.

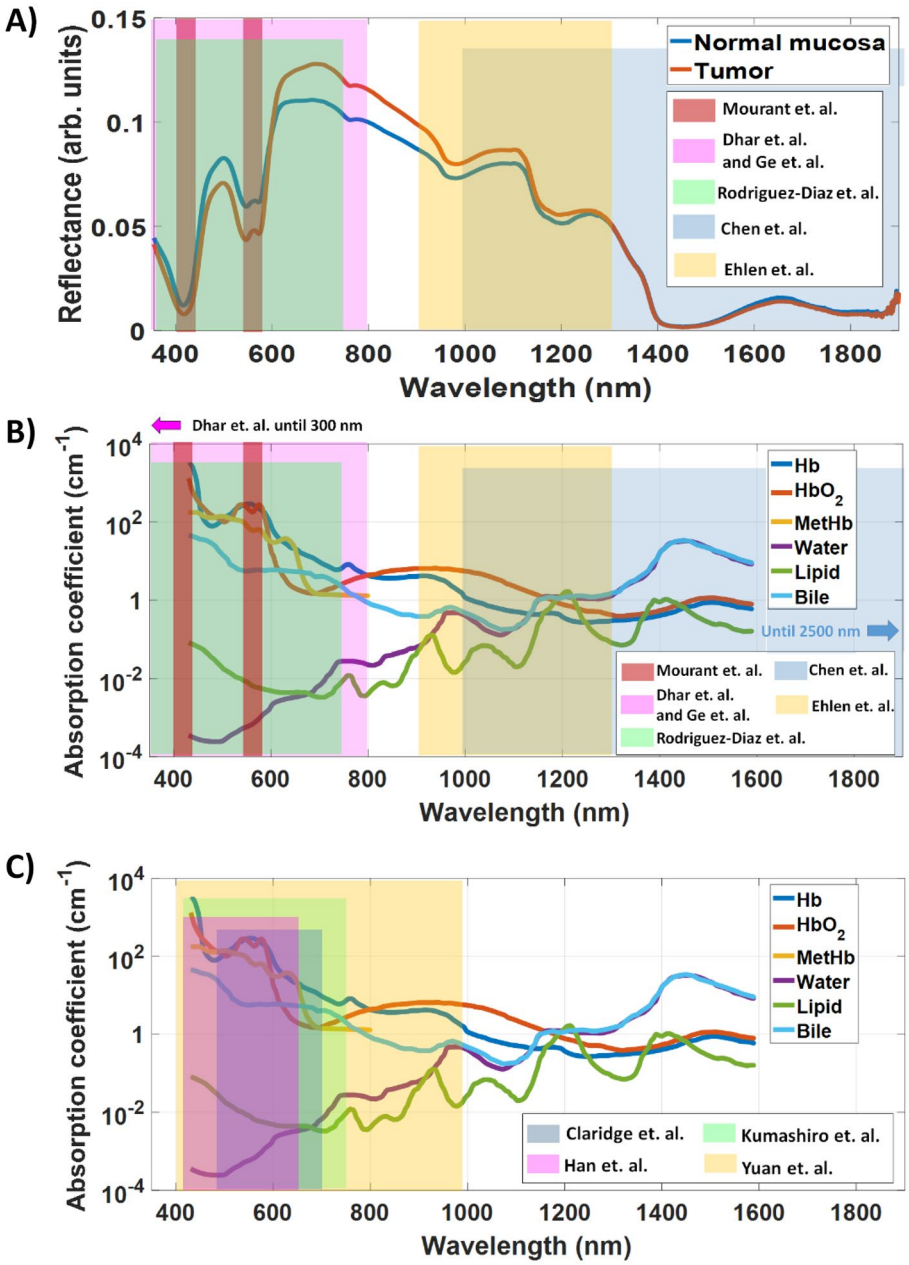


Figure 6. Wavelength ranges analyzed in previous DRS studies shown on the (**A**) mean spectra of normal mucosa and tumor (small SDD probe) collected in this study and (**B**) absorption spectra of endogenous tissue chromophores (biomolecules)[84]. (**C**) Absorption spectra of endogenous tissue chromophores showing the wavelength ranges investigated on previous hyperspectral imaging studies.

Comparison with DRS-related optical spectroscopy studies. By using quadratic SVMs, we analyzed the performance of classification models using the wavelength range of previous studies. Since prior studies used only small SDD probes, we used the dataset of the 630-μm-SDD probe to build the classification models. As already shown in Fig. 6, 4 of these studies focused their analysis on features of the visible wavelength range and near-infrared (NIR) range until 800 nm, whereas 4 studies analyzed the wavelengths higher than 900 nm. In general, studies probing wavelengths lower than 800 nm exhibit lower classification performance for tissue differentiation compared to studies in the NIR range above 900 nm (Table 5). Yet, it is important to note that most of the studies including visible wavelengths were conducted in vivo and with a larger sample size compared to the NIR studies above 900 nm.

The classification performance metrics achieved in this study is comparable to most of the previous studies (Table 5), especially those probing NIR wavelengths above 900 nm and those including visible wavelengths with similar sample size. Non-comparable classification performance metrics are observed in studies by Mourant et al.[39], who have investigated a substantially lower amount of spectra and patients, and Rodriguez-Diaz et al.[44],

Authors	Tissue evaluation	Patients	Wavelength range (analysis)	Source-to-detector distance (SDD)	Tissues considered for analysis	Number of spectra analyzed	Classification	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Mourant et al.[39]	In vivo	15	Analyzed range: 400–440 nm and 540–580 nm collected range: 300–750 nm)	300–400 µm	21 normal mucosa, 31 colitis, inflammatory bowel disease, Crohn's disease, 1 hyperplastic tissue, 6 adenomas and 1 adenocarcinoma	60 (1 spectrum per tissue site)	Non-cancer and cancer and tissues	100	98	–	–
Nogueira et al.[56]	Ex vivo	47	400–440 nm and 540–580 nm	630 µm	47 tumor and 47 normal tissues	1363 (728 normal mucosa and 635 cancer spectra)	Normal mucosa and cancer tissue	84.2 ± 2.9	86.3 ± 2.2	85.2 ± 2.6	0.917 ± 0.012
Dhar et al.[40]	In vivo	45	300–800 nm		138 colonic sites	483 (290 normal, 19 hyperplastic, 69 adenomatous polyps, 74 chronic colitis, and 31 colorectal cancer spectra)	Cancerous and normal tissue	80	86	–	–
							All pathological tissues and normal tissue	92	82	–	–
Ge et al.[43]	In vivo	153	350–800 nm	16 randomly placed fibers for tissue excitation and collection	107 non-neoplastic tissues (84 normal mucosa, 23 hyperplastic polyps) and 53 neoplastic tissues (44 adenomatous polyps, 9 adenocarcinomas)	244 (9 adenocarcinomas, 90 adenomatous polyps, 33 hyperplastic polyps, and 84 normal colon tissue spectra)	Neoplastic tissue from non-neoplastic tissue	79–91	74–91	–	–
							Adenomatous polyps from hyperplastic polyps	85–91	72–75	–	–
Nogueira et al.[56]	Ex vivo	47	350–800 nm	630 µm	47 tumor and 47 normal tissues	1363 (728 normal mucosa and 635 cancer spectra)	Normal mucosa and cancer tissue	87.4 ± 2.1	87.3 ± 2.7	87.4 ± 2.4	0.933 ± 0.009
Rodriguez-Diaz et al.[44]	In vivo	83	330–760 nm	250 µm	85 neoplastic tissue, 80 hyperplastic polyps and 53 polypoid colonic mucosa	218 spectra averaged from 1090 measurements (5 spectra per tissue)	Neoplastic and non-neoplastic tissues	91.5	92.2	91.9	-
Nogueira et al.[56]	Ex vivo	47	350–760 nm	630 µm	47 tumor and 47 normal tissues	1363 (728 normal mucosa and 635 cancer spectra)	Normal mucosa and cancer tissue	84.7 ± 2.4	87.0 ± 2.3	85.8 ± 2.4	0.922 ± 0.014
Chen et al.[45]	Ex vivo	20	1000–2500 nm (wavenumber interval from 4000 cm^{-1} to 10,000 cm^{-1})	Multiple (7.0 mm^2 of sampling area)	78 cancer and 108 normal tissues	78 tumor spectra and 108 normal tissue spectra	Normal and cancer tissues	92.3	100		–
Chen et al.[46]	Ex vivo	–	1000–2500 nm	Multiple (7.0 mm^2 of sampling area)	55 cancer and 55 normal tissue slices (stabilized in 4% formaldehyde)	110 (55 cancer and 55 normal tissue spectra)	Normal and cancer tissues	100	78.6	89.1	–
Chen et al.[47]	Ex vivo	–	1000–2500 nm	Multiple (7.0 mm^2 of sampling area)	58 paraffin tissue sections	58	Normal and cancer tissues	92.8	86.7	–	–
Nogueira et al.[56]	Ex vivo	47	1000–1919 nm	630 µm	47 tumor and 47 normal tissues	1363 (728 normal mucosa and 635 cancer spectra)	Normal mucosa and cancer tissue	92.1 ± 1.2	92.3 ± 1.4	92.2 ± 1.3	0.95 ± 0.011

Continued

Authors	Tissue evaluation	Patients	Wavelength range (analysis)	Source-to-detector distance (SDD)	Tissues considered for analysis	Number of spectra analyzed	Classification	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Ehlen et al.[48]	Ex vivo	10	Analyzed range: 900–1300 nm (collected range: 900–1720 nm)	400 µm + fiber cladding	10 tumor and 10 normal tissues	95 (50 tumor and 45 normal tissue spectra)	Normal mucosa and cancer tissue (DRS and FS)	98	93	96	–
							Normal mucosa and cancer tissue (only DRS)	93	68	82	
Nogueira et al.[56]	Ex vivo	47	900–1300 nm	630 µm	47 tumor and 47 normal tissues	1363 (728 normal mucosa and 635 cancer spectra)	Normal mucosa and cancer tissue	86.9 ± 3.5	88.0 ± 3.7	87.4 ± 3.6	0.921 ± 0.020

Table 5. Comparison of clinical data, instrument specifications and classification performance metrics across optical spectroscopy studies. The dataset of this study (represented by the Nogueira et al.[56] in the table) was evaluated in the wavelength ranges of previous studies and for comparison of classification results.

who have used 250 µm of fiber SDD and probed different types of tissue compared to the present study. In terms of purely NIR studies, the group of Chen et al.[45–47] showed the tissue discrimination achieved by using random forest, Adaboost and SVM classifiers. As expected, the closest classification performance compared to the present study was the SVM classifier. Still, Ehlen et al.[48] found 6.1% higher sensitivity, 20% lower specificity and 5.4% lower accuracy compared to the present study, which may be originated by using 400-µm fiber SDD and substantially smaller sample size (number of patients and spectra).

The accuracy achieved in our study tends to be higher (i.e. due to increased sensitivity and/or specificity) compared to the in vivo DRS studies (Table 5). It is unclear whether higher accuracy is obtained due to differences in the tissue evaluation procedure (in vivo or ex vivo), number of patients, region where the study was conducted, types of interrogated tissue or machine learning model used in each study. At all cases, our classification model contains the highest number of investigated spectra and therefore, the largest representation over heterogeneity caused by probe contact pressure effects, probe positioning and other factors contributing to DRS signal variation during contact measurements.

Comparison with hyperspectral imaging studies. In order to compare the performance metrics achieved by wavelength ranges of hyperspectral imaging (HSI) studies, we used quadratic SVM models build with the dataset of the 630-µm-SDD probe. The equivalent comparison comes from the information hyperspectral imaging gets from the superficial tissue, which is comparable to what is obtained by using the short SDD probe of our study. Most of the HSI studies have exploited the reflectance on the visible wavelength region and thus, based their tissue identification on blood absorption and relatively high superficial tissue scattering (Fig. 6, Table 6). Kumashiro et al.[50] extended the wavelength range up to 750 nm and Yuan et al.[42] extended the range further up to 1000 nm. Although the study of Yuan et al. investigated the water absorption signal at longer wavelengths, the study was limited by ex vivo samples of 4 patients.

Studies including data on wavelengths between 400–470 nm showed higher classification performance compared to the study of Claridge et al.[49], who used only wavelengths longer than 470 nm. This agrees with the results of our study and suggests that the blood content and oxygenation in superficial tissue layers highly contributes to the discrimination between mucosa and tumor tissues. In addition, even though Han et al.[41] investigated a substantially narrower wavelength range than Yuan et al.[42], they achieved higher classification performance in their in vivo study compared to an ex vivo study. This performance suggests that blood oxygenation changes in vivo, especially those probed between 405 and 665 nm, contribute significantly to the tissue differentiation. Still, further investigation is required to understand the origin of the contrast between normal and cancerous tissues by using HSI.

Previous HSI studies have investigated a significantly smaller patient population compared to DRS studies. However, the classification performance metrics obtained in this study as well as their changes due to wavelength ranges were comparable when considering similar specifications with respect to the probe depth investigated by either imaging from a distance or using lower SDD fiber probes. Compared to the performance metrics that could be achieved in this study by using the wavelength range between 405 and 665 nm, Han et al.[41] obtained significantly higher metrics on their in vivo study by using support vector machines. Conversely, the ex vivo tissue classification performed on a wider wavelength range (from 405 to 750 nm) by Kumashiro et al.[50] led to much lower performance metrics compared to our study.

Current hyperspectral imaging studies are limited by the number of patients, and lack of standardization on imaged angles, distance from the tissue and SDDs for each pixel. Therefore, comparison with performance metrics achieved by point spectroscopy (which standardizes measurement parameters and has proven to be robust among a large number of patients) still requires further investigation to address those limitations. On the other hand, HSI has the potential to cause a significant decrease in measurement times by acquiring data of multiple points at once. In order to keep the differentiation achieved by previous spectroscopic studies, HSI has to be used at optimum wavelengths for application-specific tissue identification.

Authors	Tissue evaluation	Patients	Wavelength range (analysis)	Normal and cancerous tissues considered for analysis	Number of images and spectra analyzed	Classification	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Claridge et al.[49]	Ex vivo	8	470–700 nm	7 adenocarcinomas, 2 adenomatous polyps, 1 adenoma, 1 tubulo-villous adenoma, 1 neoplastic polyp and 1 hyperplastic polyp and surrounding normal mucosa	–	Colorectal lesions and normal mucosa	85	78	–	–
Nogueira et al.[56]	Ex vivo	47	470–700 nm	47 tumor and 47 normal tissues	1363	Normal mucosa and cancer tissue	82.3 ± 2.2	86.0 ± 0.8	84.1 ± 1.6	0.900 ± 0.008
Kumashiro et al.[50]	Ex vivo	21	405–750 nm	40 tumor and 95 normal mucosa sites	135 (40 tumor and 95 normal mucosa images)	Normal mucosa and cancer tissue	72.5	82.1	–	–
						All pathological and normal tissues	75.0	55.8	–	–
Nogueira et al.[56]	Ex vivo	47	405–750 nm	47 tumor and 47 normal tissues	1363	Normal mucosa and cancer tissue	86.8 ± 0.5	89.0 ± 0.3	87.8 ± 0.4	0.927 ± 0.005
Han et al.[41]	In vivo	12	405–665 nm	21 colorectal tumors or adenomatous polyps and their surrounding normal mucosa	–	Normal mucosa and cancer tissue (all hyperspectral images)	96.9 ± 2.3	91.5 ± 4.1	94.5 ± 4.9	–
	In vivo	12	415–425 nm, 495–505 nm, 525–545 nm and the wavelengths 465 nm and 625 nm	21 colorectal tumors or adenomatous polyps and their surrounding normal mucosa	–	Normal mucosa and cancer tissue (selected wavelength bands)	94.7 ± 4.2	89.1 ± 3.2	92.9 ± 5.4	–
Nogueira et al.[56]	Ex vivo	47	405–665 nm	47 tumor and 47 normal tissues	1363	Normal mucosa and cancer tissue	85.6 ± 1.4	90.0 ± 0.7	87.7 ± 1.1	0.930 ± 0.010
Nogueira et al.[56]	Ex vivo	47	415–425 nm, 495–505 nm, 525–545 nm and the wavelengths 465 nm and 625 nm	47 tumor and 47 normal tissues	1363	Normal mucosa and cancer tissue	74.5 ± 0.4	83.0 ± 0.8	78.5 ± 0.6	0.863 ± 0.005
Yuan et al.[42]	Ex vivo	4	Analyzed: 405.235 nm, 406.42 nm, 408.794 nm, 414.736 nm, 422.48 nm, 468.76 nm, 559.489 nm, 577.506 nm, 594.342 nm, 957.948 nm, and 1,000.31 nm (collected: 400 nm and 1000 nm)	–	–	Cancerous and non-cancerous tissues	93.8	87.5	90.6	–
Nogueira et al.[56]	Ex vivo	47	400–1000 nm	47 tumor and 47 normal tissues	1363	Normal mucosa and cancer tissue	90.5 ± 1.2	92.0 ± 0.5	91.2 ± 0.9	0.963 ± 0.004
Nogueira et al.[56]	Ex vivo	47	405.235 nm, 406.42 nm, 408.794 nm, 414.736 nm, 422.48 nm, 468.76 nm, 559.489 nm, 577.506 nm, 594.342 nm, 957.948 nm, and 1,000.31 nm	47 tumor and 47 normal tissues	1363	Normal mucosa and cancer tissue	78.2 ± 0.7	86.0 ± 0.5	81.9 ± 0.6	0.903 ± 0.005

Table 6. Comparison of clinical data, instrument specifications and classification performance metrics across hyperspectral imaging studies and this study. The dataset of this study (represented by the Nogueira et al.[56] in the table) was evaluated in the wavelength ranges of previous studies and for comparison of classification results.

Limitations and validity of this study. Our paper has a number of limitations. First, this study does not include non-cancer pathology, which will be investigated in future research to incorporate data for improved tissue identification and possibly accurate sampling of areas to be biopsied. The measurements of this study have been performed in tumors of diverse stages with predominance of advanced stage tumors (Table 1). The tumor sites to be measured were identified by a combination of palpation and naked eye determination by experienced surgeons who demarcated the tumor region. However, this determination does not decrease the validity of our dataset and robust classification model, which is based on a larger dataset compared to previous studies (Tables 5, 6) and comprises variations on DRS spectra due to parameters discussed in this section. With this in mind, tissue alterations occurring in early-stage cancer (e.g. changes on the tissue microstructure due to cell proliferation) are expected to be the same as advanced cancer but with lower magnitude. Our feasibility study shows that these alterations are detectable and improved detection is achieved by larger SDD probes as well as extended wavelength ranges into the near-infrared.

Since the tissue blood oxygenation may change in ex vivo tissues during our measurements, we evaluated its changes over time in a pilot observation of the DRS signal in 3 patients and found no significant variations during the first 15 min of our data collection (measurements taken every 5 min and data not shown). Even though tissue oxygenation may be different between our research and prior in vivo studies due to collection of DRS measurements of ex vivo tissues, the trend observed in our data matches with that found in previous studies in general. Variations due to tissue blood oxygenation are expected to be similar to what was obtained by Baltussen et al.[9], who reported that blood content and StO_2 increased in the measured ex vivo within 1 h after resection in relation to in vivo tissues. Other factors changing ex vivo spectra include tissue dehydration, which was reduced as much as possible in this study by maintaining the tissue moisture with a wet wipe. However, it is important to remember that Baltussen et al.[9] used a 1290-µm SDD probe and investigated fat, tumor, and healthy colorectal wall tissues.

In terms of variations observed in every technique requiring contact measurements, probe contact pressure and sufficient optical contact may be challenging to accomplish during in vivo endoscopy. If we consider that the large SDD probe collects data from a larger tissue volume and pressure is applied more uniformly during measurements, a reduced sensitivity to optical contact and probe positioning is expected. In this case, higher repeatability of measurements with the large SDD may be associated with the higher classification performance with the large SDD probe. Yet, the larger probed volume means the probe contact pressure may lead to variations due to the thickness of tissue layers. The variations commented above were previously reported[62] and need to be further investigated in order to apply automated corrections to DRS spectra. Still, the measurements performed in this study incorporate variations due to probe contact pressure and temperature decrease in ex vivo tissues and a dataset substantially larger (1363 spectra for the small SDD probe and 1526 for the large SDD probe) than prior studies was generated in order to build a robust classification model.

Finally, other types of classification methods (e.g. deep neural networks) can be used to build classification models that can easily be run in the real-time. In fact, we tested other types of classification methods which resulted in sensitivities and specificities higher than 85% (data not shown). High performance metrics independent on the classification method used means our data clearly shows the discrimination between mucosa and cancer tissues. Therefore, we believe that future in vivo studies advancing research presented here should consider using larger SDD probes and extended wavelength ranges in the near-infrared for differentiation between mucosa and cancer tissues. As a factor for clinical translation, the incorporation of 2.5-mm-SDD probes is feasible, as the size of colonoscope channels range from 2.8 to 4.2 mm[86,87].

Conclusions

In this study, we evaluated the usefulness of the extended wavelength range for improving CRC detection using DRS and compared results achieved in this study with previous research. By using probing the superficial tissue, we obtained (93.5 ± 2.4)% sensitivity, (94.0 ± 1.7)% specificity and 0.971 ± 0.014 AUROC, whereas (96.1 ± 1.8)% sensitivity, (95.7 ± 0.6)% specificity and 0.987 ± 0.005 AUROC was achieved by sampling deeper tissue layers. To the best of our knowledge, this is the first DRS study to investigate the potential of probing deeper tissue layers using larger SDD probes for CRC detection in the luminal wall. Our study was conducted in ex vivo tissues, while it is straight forward to extend this methodology to an in vivo examination during endoscopy. Future studies employing diffuse reflectance spectroscopy, elastic scattering spectroscopy, near-infrared spectroscopy, hyperspectral imaging and spatial frequency domain imaging can exploit enhanced tumor detection due to the use of large SDD probes and the broadband wavelength range illustrated in this study. In a practical perspective, this study could potentially be used to develop a probe for CRC detection during colonoscopy. Real-time tissue classification is enabled by automated SVM model coupled with a DRS instrument capable of displaying the result of a single reading in about 2–3 s. By integrating this capability into a flexible fiberoptic probe which could be passed down a scope working channel, optical spectroscopy can obviate the need for multiple biopsies or polypectomies of normal mucosa as well as identify more subtle mucosal abnormalities such as sessile serrated polyps for example, which may be difficult to recognize during colonoscopy.

Received: 10 August 2020; Accepted: 4 December 2020
Published online: 12 January 2021

References

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**(6), 394–424 (2018).
2. Ferlay, J. *et al. Global Cancer Observatory: Cancer Today* (International Agency for Research on Cancer, Lyon, 2018).
3. WHO. *Estimated Number of Incident Cases from 2018 to 2040, All Cancers, Both Sexes, All Ages* (WHO, Geneva, 2020).
4. Issa, I. A. & Noureddine, M. Colorectal cancer screening: An updated review of the available options. *World J. Gastroenterol.* **23**(28), 5086 (2017).
5. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319**(9), 525–532 (1988).
6. Bond, J. H. Fecal occult blood test screening for colorectal cancer. *Gastrointest. Endosc. Clin.* **12**(1), 11–21 (2002).
7. De Koning, S. G. B. *et al.* Toward complete oral cavity cancer resection using a handheld diffuse reflectance spectroscopy probe. *J. Biomed. Opt.* **23**(12), 121611 (2018).
8. De Boer, L. L. *et al.* Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries. *Breast Cancer Res. Treat.* **152**(3), 509–518 (2015).
9. Baltussen, E. J. M. *et al.* Comparing in vivo and ex vivo fiberoptic diffuse reflectance spectroscopy in colorectal cancer. *Transl. Biophoton.* **1**(1–2), e201900008 (2019).
10. Baltussen, E. J. M., Sterenborg, H. J. C. M., Ruers, T. J. M. & Dashtbozorg, B. Optimizing algorithm development for tissue classification in colorectal cancer based on diffuse reflectance spectra. *Biomed. Opt. Express* **10**(12), 6096–6113 (2019).
11. 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**(7), 559–565 (2015).
12. Langhout, G. C. *et al.* Colorectal cancer identified using optical spectroscopy. *Ann. Oncol.* **25**, 206 (2014).
13. Soares, J. S. *et al.* Diagnostic power of diffuse reflectance spectroscopy for targeted detection of breast lesions with microcalcifications. *Proc. Natl. Acad. Sci.* **110**(2), 471–476 (2013).
14. Spliethoff, J. W. *et al.* Improved identification of peripheral lung tumors by using diffuse reflectance and fluorescence spectroscopy. *Lung Cancer* **80**(2), 165–171 (2013).
15. Evers, D. J. *et al.* Optical sensing for tumor detection in the liver. *Eur. J. Surg. Oncol.* **39**(1), 68–75 (2013).
16. Tanis, E. *et al.* In vivo tumor identification of colorectal liver metastases with diffuse reflectance and fluorescence spectroscopy. *Lasers Surg. Med.* **48**(9), 820–827 (2016).
17. 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**(10), 106014 (2017).
18. Baltussen, E. J. M. *et al.* Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery. *J. Biomed. Opt.* **24**(1), 16002 (2019).
19. Baltussen, E. J. M. *et al.* Using diffuse reflectance spectroscopy to distinguish tumor tissue from fibrosis in rectal cancer patients as a guide to surgery. *Lasers Surg. Med.* **52**, 604 (2019).
20. Baltussen, E. J. M. *et al.* Tissue diagnosis during colorectal cancer surgery using optical sensing: An in vivo study. *J. Transl. Med.* **17**(1), 1–10 (2019).
21. Li, W., Liu, Y. & Qian, Z. Determination of detection depth of optical probe in pedicle screw measurement device. *Biomed. Eng. Online* **13**(1), 1–11 (2014).
22. Prosst, R. & Gahlen, J. Fluorescence diagnosis of colorectal neoplasms: a review of clinical applications. *Int. J. Colorectal Dis.* **17**(1), 1–10 (2002).
23. Cothren, R. M. *et al.* Detection of dysplasia at colonoscopy using laser-induced fluorescence: A blinded study. *Gastrointest. Endosc.* **44**(2), 168–176 (1996).
24. Saini, M. K., Juneja, M. & Mittal, A. *Colorectal Cancer Detection from Colonoscopy Videos: A Comprehensive Survey of Computer-Aided Diagnosis Techniques* (2018).
25. Molckovsky, A., Wong Kee Song, L. M., Shim, M. G., Marcon, N. E. & Wilson, B. C. Diagnostic potential of near-infrared Raman spectroscopy in the colon: Differentiating adenomatous from hyperplastic polyps. *Gastrointest. Endosc.* **57**(3), 396–402 (2003).
26. Widjaja, E., Zheng, W. & Huang, Z. Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines. *Int. J. Oncol.* **32**(3), 653–662 (2008).
27. Li, S. *et al.* Identification and characterization of colorectal cancer using Raman spectroscopy and feature selection techniques. *Opt. Express* **22**(21), 25895–25908 (2014).
28. Salman, A., Shufan, E., Sahu, R. K., Mordechai, S. & Sebbag, G. Insights on colorectal cancer relapse by infrared microscopy from anastomosis tissues: Further analysis. *Vib. Spectrosc.* **83**, 17–25 (2016).
29. Eker, C. *et al.* Clinical spectral characterisation of colonic mucosal lesions using autofluorescence and δ aminolevulinic acid sensitisation. *Gut* **44**(4), 511–518 (1999).
30. Andersson, P. S. *et al.* Fluorescence endoscopy instrumentation for improved tissue characterization. *Med. Phys.* **14**(4), 633–636 (1987).
31. Mayinger, B. *et al.* Endoscopic light-induced autofluorescence spectroscopy for the diagnosis of colorectal cancer and adenoma. *J. Photochem. Photobiol. B Biol.* **70**(1), 13–20 (2003).
32. Shao, X., Zheng, W. & Huang, Z. In vivo diagnosis of colonic precancer and cancer using near-infrared autofluorescence spectroscopy and biochemical modeling. *J. Biomed. Opt.* **16**(6), 67005 (2011).
33. Wang, C.-Y., Lin, J.-K. & Chiang, H. K. Autofluorescence spectroscopy to identify normal and cancerous colorectal tissues. In *Proc. 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)* Vol. **2**, 967–969 (1998).
34. Kwek, L. C. *et al.* High-sensitivity and specificity of laser-induced autofluorescence spectra for detection of colorectal cancer with an artificial neural network. *Appl. Opt.* **44**(19), 4004–4008 (2005).
35. Coda, S. *et al.* Fluorescence lifetime spectroscopy of tissue autofluorescence in normal and diseased colon measured ex vivo using a fiber-optic probe. *Biomed. Opt. Express* **5**(2), 515–538 (2014).
36. Tang, G., Lu, H., Zhang, Y., Yan, S. & Chen, Z. Detection of human colonic adenoma by laser-induced autofluorescence integrated with diffuse reflectance spectroscopy. *Biomed. Photonics Optoelectron. Imaging* **4224**, 174–178 (2000).
37. Wallace, M. B. & Kiesslich, R. Advances in endoscopic imaging of colorectal neoplasia. *Gastroenterology* **138**(6), 2140–2150 (2010).
38. Aihara, H., Tajiri, H. & Suzuki, T. Application of autofluorescence endoscopy for colorectal cancer screening: Rationale and an update. *Gastroenterol. Res. Pract.* **2012**, 1 (2012).
39. Mourant, J. R. *et al.* Elastic scattering spectroscopy as a diagnostic tool for differentiating pathologies in the gastrointestinal tract: Preliminary testing. *J. Biomed. Opt.* **1**(2), 192–200 (1996).
40. Dhar, A. *et al.* Elastic scattering spectroscopy for the diagnosis of colonic lesions: Initial results of a novel optical biopsy technique. *Gastrointest. Endosc.* **63**(2), 257–261 (2006).
41. Han, Z. *et al.* In vivo use of hyperspectral imaging to develop a noncontact endoscopic diagnosis support system for malignant colorectal tumors. *J. Biomed. Opt.* **21**(1), 16001 (2016).

42. Yuan, X. *et al.* Hyperspectral imaging and SPA–LDA quantitative analysis for detection of colon cancer tissue. *J. Appl. Spectrosc.* **85**(2), 307–312 (2018).
43. Ge, Z., Schomacker, K. T. & Nishioka, N. S. Identification of colonic dysplasia and neoplasia by diffuse reflectance spectroscopy and pattern recognition techniques. *Appl. Spectrosc.* **52**(6), 833–839 (1998).
44. Rodriguez-Diaz, E. *et al.* Endoscopic histological assessment of colonic polyps by using elastic scattering spectroscopy. *Gastrointest. Endosc.* **81**(3), 539–547 (2015).
45. 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).
46. Chen, H., Lin, Z., Mo, L. & Tan, C. Identification of colorectal cancer using near-infrared spectroscopy and adaboost with decision stump. *Anal. Lett.* **50**(16), 2608–2618 (2017).
47. Chen, H., Tan, C., Wu, H., Lin, Z. & Wu, T. Feasibility of rapid diagnosis of colorectal cancer by near-infrared spectroscopy and support vector machine. *Anal. Lett.* **47**(15), 2580–2593 (2014).
48. Ehlen, L. *et al.* Synergy of fluorescence and near-infrared spectroscopy in detection of colorectal cancer. *J. Surg. Res.* **242**, 349–356 (2019).
49. Claridge, E. & Hidović-Rowe, D. Model based inversion for deriving maps of histological parameters characteristic of cancer from ex-vivo multispectral images of the colon. *IEEE Trans. Med. Imaging* **33**(4), 822–835 (2013).
50. Kumashiro, R. *et al.* Integrated endoscopic system based on optical imaging and hyperspectral data analysis for colorectal cancer detection. *Anticancer Res.* **36**(8), 3925–3932 (2016).
51. Wang, L., Jacques, S. L. & Zheng, L. MCML—Monte Carlo modeling of light transport in multi-layered tissues. *Comput. Methods Progr. Biomed.* **47**(2), 131–146 (1995).
52. Alerstam, E. *et al.* Next-generation acceleration and code optimization for light transport in turbid media using GPUs. *Biomed. Opt. Express* **1**(2), 658–675 (2010).
53. Jacques, S. L. Optical properties of biological tissues: A review. *Phys. Med. Biol.* **58**(11), R37 (2013).
54. Richards-Kortum, R. & Sevick-Muraca, E. Quantitative optical spectroscopy for tissue diagnosis. *Annu. Rev. Phys. Chem.* **47**(1), 555–606 (1996).
55. de Paula Campos, C. *et al.* Fluorescence spectroscopy in the visible range for the assessment of UVB radiation effects in hairless mice skin. *Photodiagn. Photodyn. Ther.* **20**, 21–27 (2017).
56. Saito Nogueira, M., Cosci, A. & Kurachi, C. Assessment of oxidative stress and metabolic rates in liver grafts using time-resolved fluorescence spectroscopy. *Biophotonics Photonic Solut. Better Health Care* **6**, 144 (2018).
57. Kurachi, C., Pires, L., Nogueira, M. S. & Pratavieira, S. Lifetime fluorescence for the detection of skin lesions. *Biomed. Opt.* **2014**, BS4B.3 (2014).
58. da Silva, A. P., Saito Nogueira, M., Jo, J. A., Salvador Bagnato, V. & Mayumi Inada, N. Optical based diagnosis and treatment of onychomycosis. *Biomed. Opt.* **2016**, 37 (2016).
59. de Andrade, C. T. *et al*. Optical spectroscopy of radiotherapy and photodynamic therapy responses in normal rat skin shows vascular breakdown products. In *Proc. SPIE*, 969410 (2016).
60. Ono, B. A., Nogueira, M., Pires, L., Pratavieira, S. & Kurachi, C. Subcellular localization and photodynamic activity of Photodithazine (glucosamine salt of chlorin e6) in murine melanoma B16–F10: An in vitro and in vivo study. *Opt. Methods Tumor Treat. Detect. Mech. Tech. Photodyn. Ther.* **1047616**, 44 (2018).
61. Nicolodelli, G. *et al.* Fluorescence lifetime evaluation of whole soils from the Amazon rainforest. *Appl. Opt.* **56**(24), 6936–6941 (2017).
62. Saito Nogueira, M. *et al.* Diffuse reflectance spectroscopy for determination of optical properties and chromophore concentrations of mice internal organs in the range of 350 nm to 1860 nm. *Biophotonics Photonic Solut. Better Health Care* **6**, 125 (2018).
63. Carvalho, L. F. C. S., Nogueira, M. S., Neto, L. P. M., Bhattacharjee, T. T. & Martin, A. A. Raman spectral post-processing for oral tissue discrimination—A step for an automatized diagnostic system: Erratum. *Biomed. Opt. Express* **8**(11), 5218 (2017).
64. Carvalho, L. F. C. S., Nogueira, M. S., Neto, L. P. M., Bhattacharjee, T. T. & Martin, A. A. Raman spectral post-processing for oral tissue discrimination—A step for an automatized diagnostic system. *Biomed. Opt. Express* **8**(11), 5218–5227 (2017).
65. Saito Nogueira, M. *et al.* Portable fluorescence lifetime spectroscopy system for in-situ interrogation of biological tissues. *J. Biomed. Opt.* **22**(12), 1 (2017).
66. Nogueira, M. S. *Fluorescence Lifetime Spectroscopy for Diagnosis of Clinically Similar Skin Lesions* (Instituto de Física de São Carlos, Física Aplicada, 2016).
67. Felipe, L. & Nogueira, M. S. New insights of Raman spectroscopy for oral clinical applications. *Analyst* **143**, 6037–6048 (2018).
68. Leal, L. B. Vibration spectroscopy and body biofluids: Literature review for clinical applications. *Photodiagn. Photodyn. Ther.* **24**, 237–244 (2018).
69. Carvalho, L. F. C. S. *et al.* In vivo Raman spectroscopic characteristics of different sites of the oral mucosa in healthy volunteers. *Clin. Oral Investig.* **23**, 1–11 (2018).
70. Nogueira, M. S., Pratavieira, S. & Kurachi, C. Fluorescence lifetime spectroscopy for evaluation of photoaging process at sun exposed and non-exposed skin. Livro de Resumos. In *SIFSC - Integrated Graduation and Graduation Week of the São Carlos Institute of Physics* (2015).
71. Nogueira, M. S. *et al.* Portable fluorescence microendoscope system for smartphones and imaging processing software. Tech. Summ. In *SPIE Biophotonics South America* (2015).
72. Nogueira, M. S., Pratavieira, S. & Kurachi, C. Evaluation of photoaging at sun exposed and non-exposed skin using a fluorescence lifetime spectroscopy system. Livro de Resumos. In *SIFSC - Integrated Graduation and Graduation Week of the São Carlos Institute of Physics* (2015).
73. Saito Nogueira, M., Texeira Rosa, R. G., Pratavieira, S. S., de D'Almeida, C. P. C. D. P. & Kurachi, C. Assembly and characterization of a fluorescence lifetime spectroscopy system for skin lesions diagnostic. *Biophotonics S. Am.* **9531**, 95313 (2015).
74. Cosci, A. *et al.* Time-resolved fluorescence spectroscopy for clinical diagnosis of actinic cheilitis: erratum. *Biomed. Opt. Express* **7**(10), 4210–4219 (2016).
75. Cosci, A. *et al.* Time-resolved fluorescence spectroscopy for clinical diagnosis of actinic cheilitis. *Biomed. Opt. Express* **7**(10), 4210–4219 (2016).
76. Saito Nogueira, M. *et al.* Evaluation of actinic cheilitis using fluorescence lifetime spectroscopy. In *Proc. SPIE*, 97031U (2016).
77. Pires, L., Nogueira, M. S., Pratavieira, S., Moriyama, L. T. & Kurachi, C. Time-resolved fluorescence lifetime for cutaneous melanoma detection. *Biomed. Opt. Express* **5**(9), 3080 (2014).
78. Saito Nogueira, M. & Kurachi, C. Assessing the photoaging process at sun exposed and non-exposed skin using fluorescence lifetime spectroscopy. In *Proc. SPIE*, 97031W (2016).
79. Almeida, C. D. P. D., Campos, C., Nogueira, M. S. & Kurachi, C. Time-resolved and steady-state fluorescence spectroscopy for the assessment of skin photoaging process. *Biophotonics S. Am.* **9531**, 1–7 (2015).
80. Backman, V. & Roy, H. K. Advances in biophotonics detection of field carcinogenesis for colon cancer risk stratification. *J. Cancer* **4**(3), 251 (2013).
81. Roy, H. K. *et al.* Spectroscopic microvascular blood detection from the endoscopically normal colonic mucosa: biomarker for neoplasia risk. *Gastroenterology* **135**(4), 1069–1078 (2008).

82. 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**(4), 2805–2817 (2009).
83. Zonios, G. *et al.* Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. *Appl. Opt.* **38**(31), 6628–6637 (1999).
84. Nachabé, R. *et al.* Effect of bile absorption coefficients on the estimation of liver tissue optical properties and related implications in discriminating healthy and tumorous samples. *Biomed. Opt. Express* **2**(3), 600–614 (2011).
85. Baltussen, E. J. M. *The Clinical Implementation of Optical Spectroscopy in Colorectal Cancer Surgery* (University of Twente, Enschede, 2020).
86. Varadarajulu, S. *et al.* GI endoscopes. *Gastrointest. Endosc.* **74**(1), 1–6 (2011).
87. Kay, M. & Wyllie, R. Colonoscopy, polypectomy, and related techniques. In *Pediatric Gastrointestinal and Liver Disease* (eds Wyllie, R. *et al.*) 650–667 (Elsevier, Amsterdam, 2011).

Acknowledgements

We acknowledge Eduardo Moriyama for the advice and logistics of the project associated with this manuscript, Shree Krishnamoorthy for the discussions on the same project, Haiyang Li for the design and assembly of the holder of the DRS system, Andrew McGuire for specimen retrieval of the second set of patients of this study and all the logistics related to it, Vivienne Curran, Aoife Foyle, Una McAuliffe, Evelyn Flanagan, Shauni Fitzgerald, and Edmund Manning for all the logistics of clinical procedures, research ethics documents, and search for patient information, Jacqueline Gunther for externally reviewing the manuscript and providing helpful comments.

Author contributions

M.S.N. Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing—Original Draft, Writing—Review & Editing, Visualization, Supervision, Project administration, S.M. Investigation, Data Curation, Writing—Review & Editing, M.A. Investigation, Data Curation, Software, Writing—Review & Editing, H.L. Methodology, Investigation, Resources, Data Curation, Writing—Review & Editing, Supervision, N.L. Methodology, Investigation, Resources, Data Curation, Writing—Review & Editing, S.K. Investigation, Resources, Data Curation, M.O.R. Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing—Review & Editing, Supervision, Project administration, Funding Acquisition. S.A.E. Conceptualization, Methodology, Resources, Writing—Review & Editing, Supervision, Project administration, Funding Acquisition.

Funding

This study received financial support from Science Foundation Ireland (SFI): Grant ID SFI/15/RP/2828.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.S.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.