University of Strathclyde

Department of Pure and Applied Chemistry

Investigating the use of Attenuated Total Reflection Fourier-Transform Infrared (ATR-FTIR) Spectroscopy for the Rapid Diagnosis of Brain Tumours using Human Blood Serum

By

James R. Hands

A thesis presented to the Department of Pure and Applied Chemistry, University of Strathclyde, in fulfilment of the requirements for the degree of Doctor of Philosophy.

December 2015

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Date: / /

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Dr Matthew J Baker for his endless support, guidance and all of the excellent opportunities.

I would like to thank Brain Tumour North West for providing funding to make this research possible. I would like to acknowledge Professor Charles Davis, Dr Katherine Ashton, Dr Carol Walker and Professor Timothy Dawson for their help.

I give thanks to Dr Caryn Hughes and Dr Konrad Dorling for their advice and invaluable knowledge of Matlab. Special thanks to Dr Graeme Clemens for his patience, superb Matlab knowledge and for chats over coffee. I would also like to thank current and past members of the Spectral Analytics group.

Thank you to my friends at the University of Central Lancashire, particularly to Jaida her support and to Antony and Fiona for being excellent friends. Cypriot potatoes! Thanks also to Hayleigh May for being my Strathclyde/Glaswegian tour guide and for always having answers to my questions.

Finally, I would like to thank my Mum, Dad, Jenna, Dean and Jaycie for their love and endless encouragement. Thank you to Auntie Ina and Uncle George and my family in Edinburgh, Scotland, for always making me so welcome. I am so fortunate to have such an amazing family.

I dedicate this thesis to my Gran - without her unconditional love, generosity and patience I would not be where I am today. Words alone simply cannot express my appreciation of everything she has done for me. Thank you for always believing in me and encouraging me to pursue my dreams.

ABSTRACT

The ability to diagnose brain cancer rapidly from human serum would allow for short testing times and prompt results providing a responsive diagnostic environment. This study demonstrates a new method for primary and metastatic brain cancer diagnosis using 1 µl volumes of human serum and attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy. To the best of our knowledge, this is the largest study on mid-infrared spectroscopy in relation to cancer research with a 433 patient cohort consisting of 3,897 ATR-FTIR spectra. Spectral data from whole serum, 100 kiloDalton (kDa), 10 kDa and 3 kDa molecular weight cut-off filtrate samples to investigate which fraction allowed for optimum differentiation of disease state and brain tumour severity (high grade *vs*. low-grade glioma) from non-cancer. The fingerprint region (1800-1000 cm^{-1}) of the acquired data was combined with an RBF-SVM and achieved optimum sensitivities and specificities from whole serum averaging 93.75 and 96.53 % respectively when distinguishing between brain tumour severities. The 1 µl serum spots dried after 8 minutes and the acquired spectra exhibited minimal variance, especially after pre-processing. Expanding the research to, for the first time, detect from cancer *vs.* non-cancer to organ of origin of metastatic disease from the same serum sample achieved optimum sensitivities and specificities of between 80.0 and 100 % respectively. Furthermore, feature extraction fed SVM analysis of the cancer *vs*. non-cancer spectral model was performed to maximise classification accuracies to achieve improved sensitivities and specificities, in contrast to fingerprint region based SVM, of 91.5 and 83.0 % respectively.

PEER-REVIEWED PUBLICATIONS

List of Peer-Reviewed Publications

5) Rapid Discrimination of Maggots Utilising ATR-FTIR Spectroscopy
Pickering, C. L., Hands, J. R., Fullwood, L. M., Smith, J., Baker, M. J. *Forensic Science International*, 249, **2015**, 189-196

4) Vibrational Spectroscopic Methods for Cytology and Cellular Research
Clemens, G., Hands, J. R., Dorling, K. M., Baker, M. J. *Analyst*, 139, **2014**, 4411-4444

3) The Rapid Diagnosis of Gliomas via Serum Spectroscopy
Hands, J. R., Abel, P., Dawson, T., Davis, C., Lea, R. W., McIntosh, A. J., Baker, M J. *Neuropathology and Applied Neurobiology, Proceedings from the 115th Meeting of the British Neuropathological Society*, 40(1), **2014**, 25

2) Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) Spectral Discrimination of Brain Tumour Severity from Serum Samples
Hands, J .R., Dorling, K., Abel, P., M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *Journal of Biophotonics*, 7(3-4), **2014**, 189-199

1) Investigating the Rapid Diagnosis of Gliomas from Serum Samples using Infrared Spectroscopy and Cytokine and Angiogenesis Factors
Hands, J. R., Abel, P., Ashton, K., Dawson, T., Davis, C., Lea, R. W., McIntosh, A. J., Baker, M. J. *Analytical and Bioanalytical Chemistry*, 405(23), **2013**, 7347-7355

UNDER REVIEW

1) Brain Tumour Differentiation: Rapid Stratified Serum Diagnostics via Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy
Hands, J .R., Clemens, G., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Stables, R., Baker, M. J..*Neuro-oncology*

LIST OF ABBREVIATIONS AND SYMBOLS

The following table describes the various abbreviations and acronyms used throughout the thesis.

Abbreviation	Meaning
A&E	Accident and Emergency
AMI	Acute Myocardial Infraction
ATR	Attenuated Total Reflection
ATR-FTIR	Attenuated Total Reflection Fourier-Transform Infrared
BOLD	Blood oxygenation level-dependent
BMI	Body Mass Index
BTNW	Brain Tumour North West
C	Cost
CNS	Central Nervous System
CT	Computed Tomography
CV	Cross Validation
CSF	Cerebral Spinal Fluid
Di	Diamond
DNA	Deoxyribonucleic Acid
DTGS	Deuterated Triglycine Sulphate
EM	Electromagnetic
EMR	Electromagnetic Radiation
F	Female
FE	Feature Extraction/Extracted
FE-SVM	Feature Extracted-Support Vector Machine
FTIR	Fourier Transform Infrared Spectroscopy
fMRI	Functional Magnetic Resonance Imaging

GAPD45a	Growth arrest and DNA-damage-inducible protein
Ge	Germanium
GBM	Glioblastoma Multiforme
GFAP	Glial Fibrillary Acid Protein
GP	General Practitioner
GPs	General Practitioners
GUI	Graphical User Interface
H&E	Haematoxylin and Eosin
HT-FTIR	High-throughput Fourier Transform Infrared Spectroscopy
HGG	High-grade Glioma
IGFBP2	Insulin-like Growth Factor Binding Protein
IRE	Internal Reflection Element
IQGAP1	Rad GTPase-activating-like protein
kDa	kiloDalton
LGG	Low-grade Glioma
M	Male
MCT	Mercury Cadmium Telluride
Mid-IR	Mid-Infrared
MGMT	Methlyguanine methyl transferase
MMP-9	Matrix metallopeptidase-9
MRA	Magnetic Resonance Angiography
MRI	Magnetic Resonance Imaging
MVA	Multivariate Analysis
NCUH	Nottingham City University Hospital
NHS	National Health Service
NICE	National Institute of Health and Care Excellence
NIR	Near-infrared
NR	Noise Reduction

OPD	Optical Path Difference
PC	Principal Component
PCs	Principal Components
PCA	Principal Component Analysis
PCD	Programmed Cell Death
PET	Positron Emission Tomography
pkAmp	Peak Amplitude
QCL	Quantum Cascade Laser
RBF	Radial Basis Function
RBF-SVM	Radial Basis Function-Support Vector Machine
RI	Refractive Index
RPTPβ	Receptor tyrosine phosphatase
RMS	Root-mean square
RMSN	Root-mean square number
RNA	Ribonucleic Acid
Sens	Sensitivity
Spec	Specificity
STD/RSD	Standard Deviation/Relative Standard Deviation
SNR	Signal-to-Noise Ratio
SPECT	Single-photon Emission Computer Tomography
SVM	Support Vector Machine
TWW	Two-week Wait
UCLan	University of Central Lancashire
UK	United Kingdom
VN	Vector Normalisation
WHO	World Health Organisation
W	Wavenumber
YKL-40	Human Cartilage Glycoprotein-39

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF TABLES

Chapter 1

Chapter 4

Chapter 5

Chapter 6

TABLE OF CONTENTS

APPENDIX SECTION

Chapter 1

INTRODUCTION

1. General Cancer

1.1 Introduction and Epidemiology

Cancer is a collective term for a group of malignant neoplasms which is used to describe an extensive collection of diseases. Cancers are characterised by the anatomical region and cell type involved in uncontrollable cell division (1).

In the United Kingdom (UK), one in two people born after 1960 will be diagnosed with some form of cancer during their life (2). In 2012, 14.1 million new cases of cancer and 8.2 million deaths due to the disease were recorded worldwide. In the same year, some 161,823 individuals died from cancer in the UK; this devastating disease is responsible for more than 440 deaths every day in the UK, a reported one death every four minutes (3). The ever increasing global burden of cancer is predicted to soar to some 21 million diagnoses worldwide by 2030 as older aged individuals are more susceptible to the disease and an increasing aging population continues in many countries (4).

Malignant tumours are recognised as they display evidence of abnormal proliferation and lack of normal cell control mechanisms, such as the cell cycle (5). Malignant tumours are highly likely to grow fast, spread into and damage surrounding tissues and have the potential to spread to other anatomical regions via the circulatory and lymphatic systems to form secondary tumours in a process

called metastasis (5). In contrast, benign tumours are usually slow growing, do not spread to other anatomical regions and individuals can often lead a healthy life without the need for treatment (6).

1.1.1 Risk Factors

As with many diseases, there are a number of factors which contribute to the progression of cancer. The major risk factors which contribute to the development of cancer include, but are not limited to; the exposure to carcinogens, genetic factors (inherited or lifestyle influenced), tobacco/alcohol use, body weight, diet, physical activity, exposure to bacteria/viruses and age. As the population ages and life expectancy increases many types of cancer become more common due to the accumulation of mutations present in the deoxyribonucleic acid (DNA), the molecule that contains genetic information (5). Figure 1.1 shows the age of death in people for all cancers combined in the UK (2010-2012). Individuals aged 60+ are more likely to develop cancer (3).

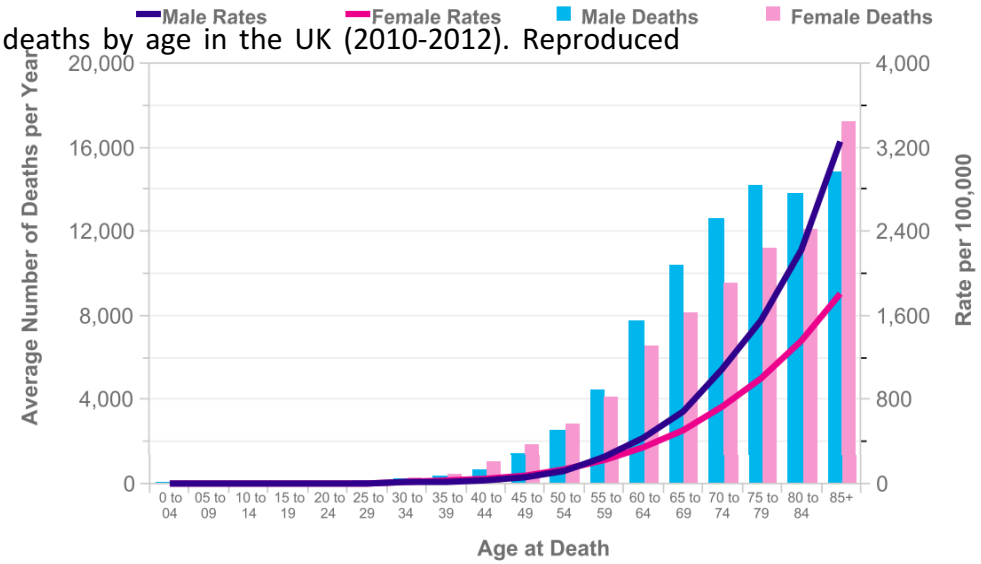


Figure 2.1 – Cancer deaths by age in the UK (2010-2012). Reproduced from (3).

1.2 Brain Cancer

1.2.1 Introduction and Statistics

Brain cancer is the biggest cancer killer of children and adults under 40 years of age (7). Gliomas are the most frequent brain tumours in adults and they account for some 70.0 % of all adult malignant primary brain tumours. Malignant gliomas are the most common and lethal of brain tumours with approximately 9,100 newly diagnosed patients each year in the UK (8). The UK annual incidence rate for all brain tumours is approximately 7 per 100,000 in the population (8). Patients diagnosed with a Glioblastoma multiforme (GBM) - the most aggressive grade of glioma with a dismal prognosis - on average have expected survival duration of approximately 8-14 months (9).

The latest incidence and mortality rates published by the Office for National Statistics (England) show that unlike most other cancers, brain tumour mortality rates have increased significantly within the last 32 years by 15.0 % and 10.0 % for men and women respectively. Incidence rates have also increased within the last 32 years by 23.0 % and 25.0 % for men and woman respectively. However, due to better treatments and early diagnosis 42.0 % of all men and women currently diagnosed survive at least one year compared to 24.0 % thirty years ago (10). Figure 1.2 shows how incidence and mortality rates for males and females have increased since 1979-2010 in England.

igure 1.2 – Brain cancer incidence and mortality rates in England (1979-2010). eproduced from (10).

Although the incidence of brain tumours is low in contrast to breast cancer or lung cancer incidence rates (Figure 1.3), on average brain and central nervous system (CNS) cancers account for the largest number of years of life lost due to cancer (Figure 1.4) (3). The brain is a major secondary site to a number of primary site cancers including the lungs, breast and bowel, thus the development of a brain tumour following metastasis is likely (11). Figure 1.3 shows the 20 most commonly diagnosed cancers in the UK (2010) and figure 1.4 the average number of years of life lost for various types of cancer. Brain and CNS cancers are responsible for the greatest number of years lost (7). As the incidence of cancer throughout the world dramatically increases in upcoming years, a significant proportion of these tumours will metastasise to the brain, thus increasing the clinical occurrences of brain

Figure 1.3 - The 20 most commonly diagnosed cancers in the UK [excluding melanoma skin cancer] (2010). Reproduced from (3).

cancer (4). The earlier most cancers are diagnosed, the greater the chance of survival; the requirement for a rapid diagnostic test is currently emerging as GPs are encouraged to suspect cancer sooner in line with new guidelines updated by the National Institute of Health and Care Excellence (NICE) (12). As of June 2015, NICE recommend that GPs refer patients to specialist centres via a symptom-based approach, a change to the traditional referral procedure of referring a patient based on which type of cancer they may have. NICE suggests that around 5,000 lives could be saved each year in England following early cancer diagnosis (12). Detecting disease early has many benefits, including the early intervention of therapeutic treatments and a reduction in mortality and morbidity rates (13). A simple, rapid and reliable diagnostic test would enable healthcare providers to diagnose cancer sooner and would relieve demand on already burdened hospital facilities.

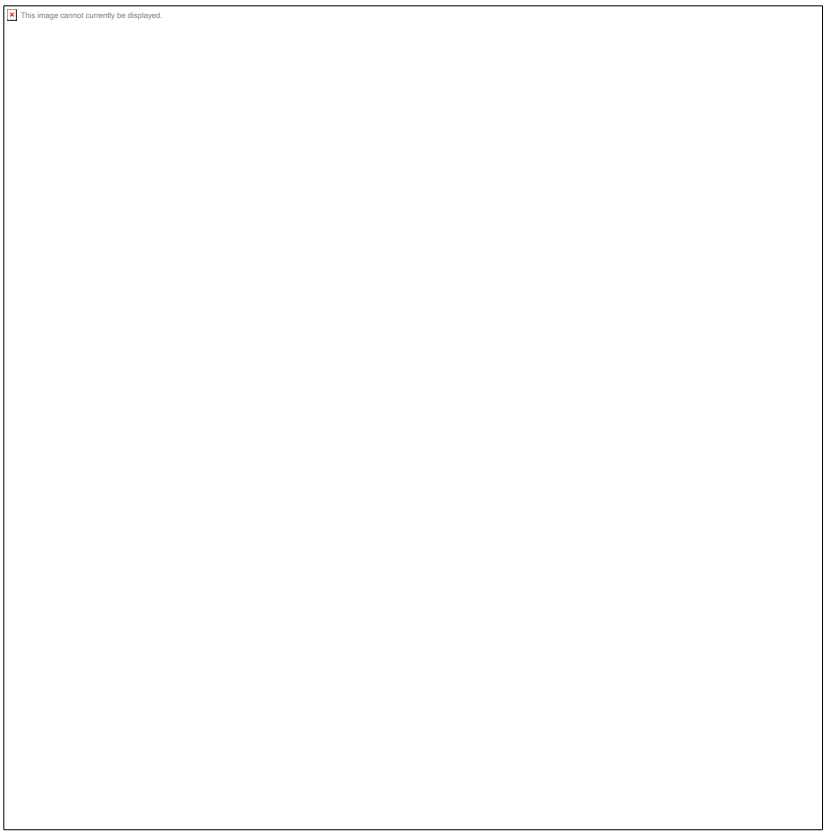
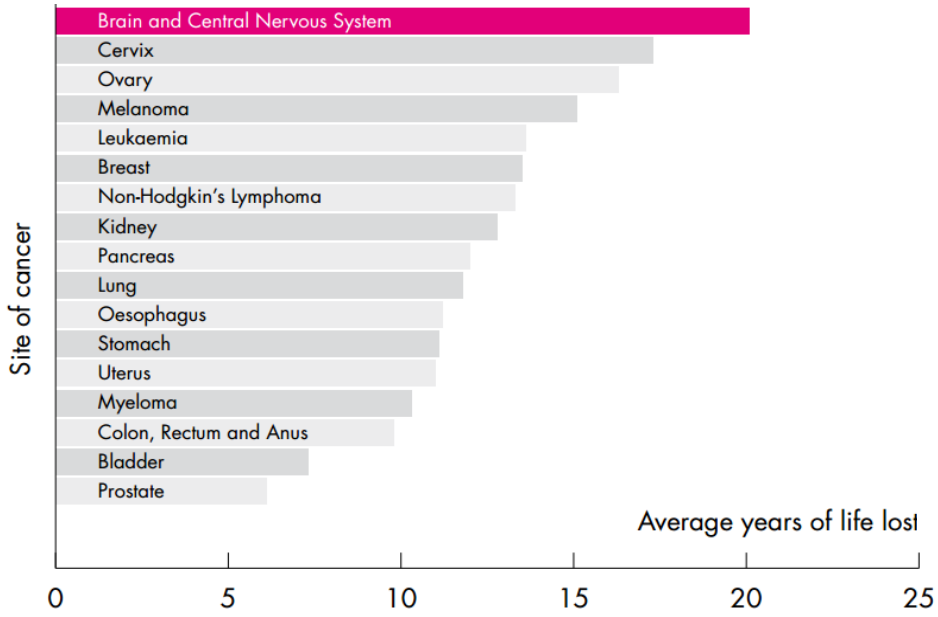


Figure 1.4 - The average years of life lost for various sites of cancer.
Reproduced from (7).

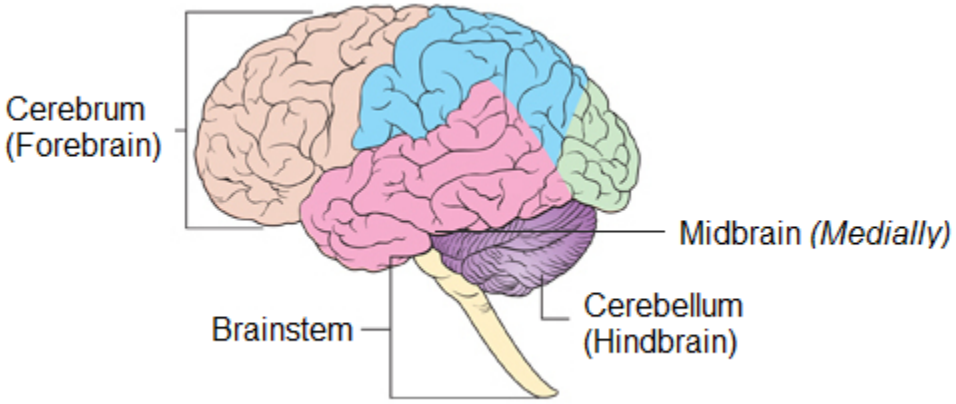


1.2.2 The Anatomy and Function of the Human Brain

The human brain is one of the most sophisticated results of biological evolution and it is responsible for all essential functions of the body. The nerves exiting the base of the brain to the body via the spinal cord allow for the control of organ function, the ability to receive/send and interpret external stimuli (pain, touch etc.) and limb movement, to name but a few (14).

The brain is the control centre of the body and this soft, jelly-like organ is protected within the cranium. For simplicity, the brain is anatomically divided into three general regions; the *forebrain* (cerebrum), *midbrain* (cerebellum) and the *hindbrain* (brainstem). Figure 1.5 shows the main areas of the human brain.

Figure 1.5 - Diagram showing the major areas of the human brain. Adapted from (19).

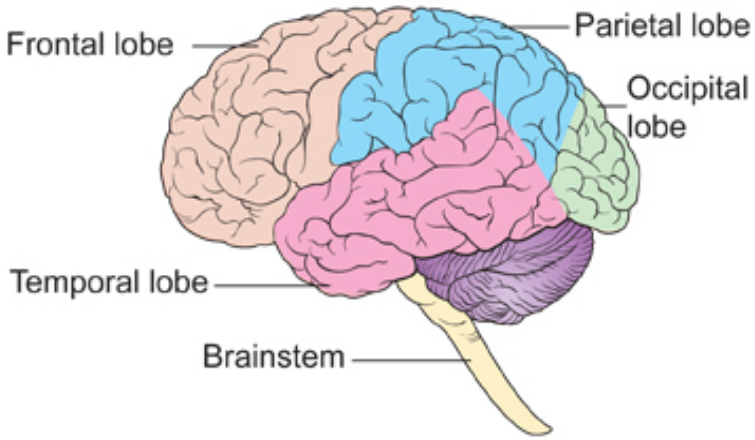


The forebrain consists of the two *cerebral hemispheres* and accounts for the major portion of the brain. Located within the *cerebral hemispheres* lies the *diencephalon*, the region of the thalamus and other glandular structures. The *midbrain* is located between the *diencephalon* and the *hindbrain*. The latter consists of the *pons*, c*erebellum* and *medulla oblongata*, commonly referred to as the *brainstem*; the region which controls critical life-sustaining functions.

Each half of the brain (medially spliced in relation to the anterior of the body's trunk) represents a cerebral hemisphere. These two halves are commonly referred to as the left and right hemispheres. The extent of the complexity of the brains regional functions remains largely unknown; however, it is known that the cerebrum controls movements, memory, speech, emotional behaviour and senses (14).

Figure 1.6 - Diagram showing the lobes of the human brain. Replicated from (19).

The left and right hemispheres are each divided into 4 general areas: the frontal lobe, temporal lobe, parietal lobe and the occipital lobe. Figure 1.6 shows the lobes of the human brain.



The frontal lobe areas are responsible for all skeletal muscle movements on the opposite side of the body to its region in the cerebral hemisphere. For example, the right cerebral hemisphere controls the movement of the left half of the body. This area is also involved with speaking and some personality traits.

The temporal lobe is the area where sounds are processed and memories stored. As is the case with the frontal lobes, the parietal lobes are the registration points of the opposite half of the body's sensory inputs (temperature, touch etc.). Finally, the occipital lobes house visual function and allow the determination of distances, colours and shapes. A tumour in any one of these four lobes can affect a person's perception of external stimuli, understanding of speech and text, and

changes to their senses; the list of symptoms due to the presence of a brain tumour is exhaustive (14).

1.2.3 Fundamentals of Brain Cancer

1.2.3.1 Primary Brain Tumours

1.2.3.1.1 Gliomas

Primary brain tumours can be broadly divided into primary glial tumours, medulloblastoma, ependymoma, germ cell tumours (germinoma and teratoma), meningioma, nerve sheath and pituitary tumours. More than 120 different types of tumour can be found in the brain, of which, gliomas account for 30-40% of all intracranial neoplasms (15). Gliomas occur in the cerebral hemispheres in > 90.0 % of cases (16), thus numerous physical and neurological symptoms often occur when a tumour affects the function of a particular lobe.

The term *glioma* refers to a primary site tumour which originates from glial cells (neuroglia) within the brain and CNS. The two main types of glial cells - astrocytes and oligodendrocytes - outnumber neuron cells and occupy approximately half of the brain (17). The three main types of malignant glioma are astrocytomas, ependymomas and oligodendrogliomas (18); the most common primary brain tumour in adults, GBM, derives from anaplastic astrocytomas through a process whereby cells undergo *anaplasia* causing them to rapidly divide and have very few morphological features of normal cells (19). A *mixed glioma* is a *tumour* with a mixture of cell types which occur in the three most frequently observed

le 1.1 - Primary brain tumour sub-types with WHO grade classification. Adapted
n (20).

tumours (19). Table 1.1 shows the sub-types of high-grade gliomas (HGG) and low-grade gliomas (LGG) described by the World Health Organisation (WHO) (20).

General Tumour Grade	WHO Grade	Grade Sub-type
Low Grade	I II II	Pilocytic astrocytoma Oligodendroglioma Astrocytoma
High Grade	III III IV	Anaplastic astrocytoma Oligodendroglioma Glioblastoma multiforme

The most aggressive and commonly diagnosed brain tumour, GBM, rarely metastasises beyond the cranium (reported 0.44 % frequency of all cases) due to the absence of lymphatic vessels in the brain and the difficulty for rogue cells to penetrate blood vessels (21).

1.2.3.1.2 Meningioma

Meningiomas are neoplasms that form from arachnoid cap cells located in the villi of the meninges, a layer of 3 connective tissues which covers the organs of the CNS. Approximately 90.0 % of all diagnosed meningiomas are benign, 5-7 % atypical (cells which are morphologically abnormal), and 1-3 % malignant. Atypical and malignant meningiomas are aggressive and rapidly progress into serious life-

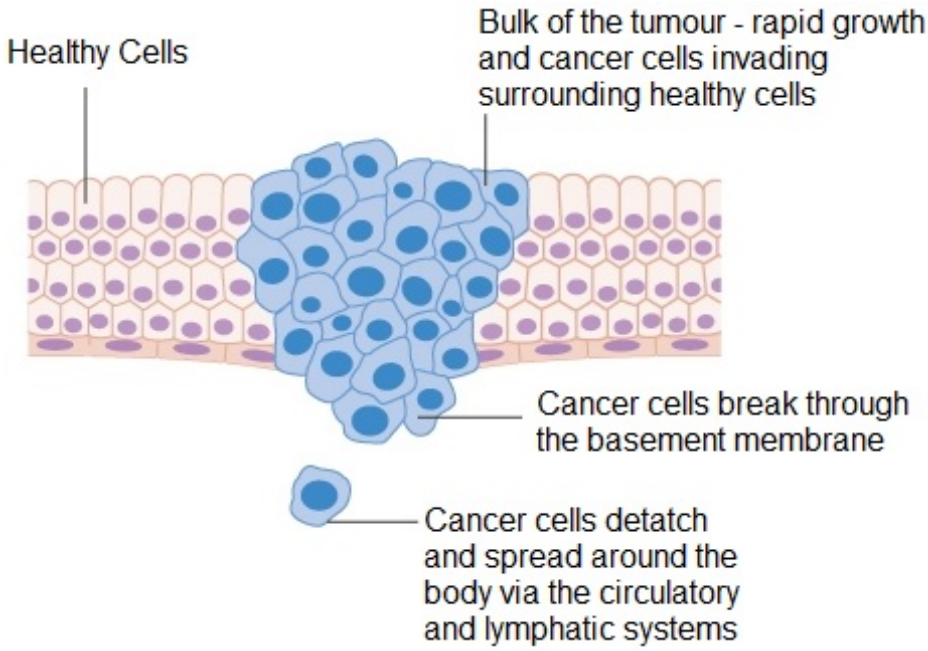
threatening tumours. A reported 40.0 % of all atypical meningiomas have a 5-year recurrence rate (22).

1.2.3.2 Metastatic Brain Tumours

Metastatic brain tumours are intracranial neoplasms that originate from cancers outside the CNS. The likelihood of a metastatic brain tumour occurring from a malignant tumour elsewhere in the body is extremely contingent on the location of the primary site tumour. Lung cancer is the most common primary site which will metastasise to the brain and is responsible for a reported 50.0 % of all brain metastases; in addition, breast cancer and malignant melanomas are responsible for some 15-25 % and 5-20 % of all brain metastases, respectively (11). The cerebral hemispheres are the location of 80-85 % of all brain metastases due to the rich blood supply and large tissue mass (23). Fewer brain metastases are located outside of the cerebral hemispheres; 10-15 % occur in the cerebellum and a meagre 1-3 % in the brain stem (24).

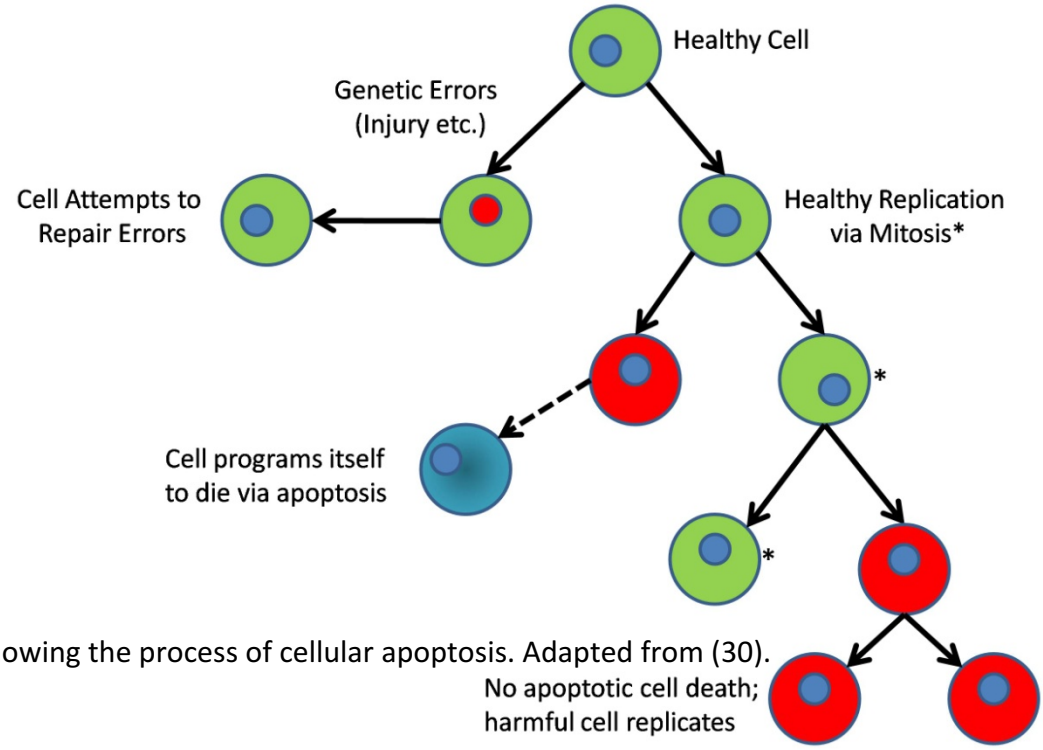
The metastatic process occurs when cancer cells from a primary tumour detach to form a deposit at a remote site via a haematogenous or lymphatic route. Figure 1.7 shows the process of metastasis from a primary site cancer (25). Approximately 50.0 % of metastatic brain cancer patients eventually have their primary sites diagnosed (26).

Figure 1.7 - Diagram showing the process of metastasis from a primary site cancer. Adapted from (25).



1.2.3.3 Programmed Cell Death

Apoptosis and regulated necrosis - termed generally as programmed cell death (PCD) - are essential for normal cell life cycle turnover in multicellular organisms (27). Programmed cell death is a vital anti-cancer process; a cellular dysfunction may potentially prevent the elimination of harmful cells, thus allowing them to continue to develop various diseases such as cancer (28). Tissue displaying *neoplasia* - the growth of abnormal tissue - often exhibits chromosomal disorganisation, ineffective cellular proliferation and unmonitored apoptosis (29). Figure 1.8 shows a schematic of the apoptotic process which can potentially lead onto harmful cells replicating uncontrollably (30).



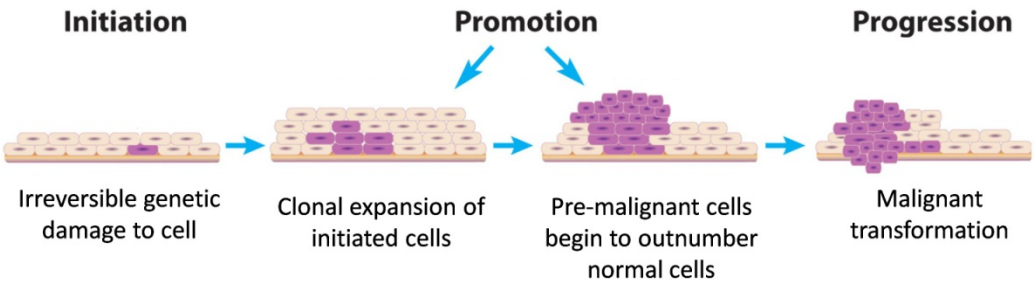
igure 1.8 - Diagram showing the process of cellular apoptosis. Adapted from (30).

In addition to the latter, another accepted model of cancer development involves stromal cell damage leading to the initiation, growth and progression of a tumour. A tumour cell's interaction with its surrounding stroma allows for vascular and tumour cell growth. The stroma is composed of a network of cellular tissues that surround and interact with the tumour; upon cellular proliferation, growth factors and angiogenesis factors that promote carcinogenesis and the process of uncontrollable cell division are released (27, 29).

Figure 1.9 - Diagram showing the process of carcinogenesis. Adapted from (33).

1.2.3.4 Carcinogenesis

The development of a tumour is a multistage process which is generally divided into three stages: initiation, promotion and progression. Initiation occurs when cellular DNA is damaged or destroyed due to the exposure of a carcinogen.



The purpose of initiation is to rapidly create irreversible genetic damage that possesses the potential to replicate under the influence of promoting agents. Promotion involves the clonal expansion of initiated cells due to the genetic alterations accumulated, thus resulting in the formation of a benign tumour. Tumour promotion agents are generally non-mutagenic but possess the ability to activate cell division. Tumour progression allows for malignant transformation to occur and for malignant cells to become more aggressive (invading surrounding tissue). Progression enables further genetic alterations to occur and the potential for metastasis (31-32). Figure 1.9 shows the basic process of carcinogenesis (33).

1.3 Diagnostic Modalities of Primary Brain Tumours

1.3.1 Symptoms of Brain Cancer

The term *brain oedema* is defined as an increase in the brain's water content (>80.0 %) which is a response to a primary brain injury or disease (34). Brain oedema leads to changes in intracranial pressure and cerebral blood flow in many diseases and is a prominent feature of brain cancer (35). An increase of intracranial pressure often contributes to neurologic dysfunction and has a negative impact on a patient's quality of life with symptoms including worsening headaches, nausea, vomiting, abnormal eye movements, seizures and strokes (36). Approximately 60.0 % of patients with brain tumours report having a headache; intracranial metastasis were found in more than 30.0 % of cancer patients whose main symptom was a headache (37). The list of symptoms linked to brain cancer is extensive, thus no one symptom is solely linked to the presence of a tumour. Currently, the diagnosis of a brain tumour is carried out via a "differential diagnosis" process, primarily to rule-out any underlying illnesses which symptoms overlap those of a more sinister diagnosis. An initial patient assessment includes medical history investigations, physical examinations, imaging studies conducted and laboratory investigations (38). Due to the many symptoms of cancer, the current diagnostic regime is extremely non-specific and subjective. In the majority of brain cancer cases, early diagnoses are not made due to the patient presenting symptoms deemed as minor and less significant, such as headaches or dizziness (37). The grade of the brain tumour is an important factor when considering how the tumour will

respond to treatment. Generally, faster growing, high-grade tumours are much more likely to metastasise and recur following treatment compared to slower-growing, low-grade tumours (39). An example of a patient's symptoms being misinterpreted could be with the eyes; the choroid, the eyes vascular structure, is a common site for ocular metastasis. The highest incidence of ocular metastatic cancer occurs from a breast primary site tumour in 40-70 % of cases (40). Patients present with symptoms including visual acuity and image distortion which may be misdiagnosed. In a number of cases, the first healthcare provider a patient visits is the optician (41) due to eyesight problems; the optician refers the patient to the hospital for further investigation (42).

1.3.2 Preoperative Diagnosis

If a patient's general practitioner (GP) suspects a brain tumour the routine procedure would be to perform one or more radiological imaging studies. Brain tumours are initially detected using a non-invasive imaging technique, such as computed tomography (CT) or magnetic resonance imaging (MRI). Imaging of the brain allows for mass, location and tumour margins to be determined (43). The following section describes the current modalities for the preoperative diagnosis of brain tumours.

1.3.2.1 Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging is often the primary imaging technique used to evaluate the presence of a brain tumour. MRI scanners utilise a combination of radiowaves, magnetic fields and specialised computer software to allow for the body's internal structures to be visualised as a 3-dimensional (3D) image. MRI is highly sensitive for the detection of abnormalities, however; it provides a large number of false positives; is costly and requires a scarce resource, liquid helium, to cool the magnet (44). False positives occur in a reported 11.0 % of metastatic and primary cancer lesions, solely based on the MRI scan (45). In addition to MRI, functional magnetic resonance imaging (fMRI) is increasingly becoming used pre-operatively to investigate patient brain function and pathology. fMRI may be advantageous over MRI in cases where presumed tumour localisation is located in or near vital neurological structures. The most commonly used fMRI technique is blood oxygenation level-dependent (BOLD) which takes advantage of the link between neuronal activity and blood flow, thus allowing the localisation of brain structures (46). Evidence of a tumour by MRI or fMRI requires confirmatory diagnosis via biopsy or tumour resection (47).

1.3.2.2 Computed Tomography (CT)

Computed tomography differs between healthy tissue and tumour tissue by examination of multiple highly detailed images of the brain. Specialised software combines the highly detailed images to generate a picture of the brain which can be

explored as required for preoperative diagnosis. A CT scan uses X-rays to create tomograms of the structures in the body, including bones, tumours and blood vessels. The circular structure around the bed on which the patient lays down houses an X-ray unit which rotates to capture scans of the internal structures. In some cases concerning the brain or abdomen, visualisation is aided by injecting an iodine-based contrast dye into the patient's circulatory system beforehand to enable clearer differentiation between tissues and blood vessels (48). As is the case with many imaging techniques, CT scans can be misinterpreted by the radiographer resulting in patients undergoing the incorrect treatments, excessive wasted costs through further testing and unnecessary surgeries. CT is not always available due to costly breakdowns nor can it provide histological evidence of tumour presence (49). The general lack of awareness with regards to knowledge of lifetime cancer risk to patients following a CT scan is relatively unknown to clinicians, despite their familiarity to the procedure; although, it has been deemed that CT procedures do not increase the incidence of cancer in patients who undergo the procedure (50). The greatest risk with CT scanning comes when it is performed on children. Children have an increased risk of low-dose radiation causing fatal cancer with an estimated 1 in 1000 paediatric fatalities occurring following a CT scan (51).

1.3.2.3 Additional Imaging Techniques

Other imaging techniques exist to provide clinicians with additional information regarding a tumours mass, precise location and exact tumour margins. Positron emission tomography (PET) is often used following a CT or MRI scan to

allow surgeons to prepare for biopsy collection by localising a specific area of interest; guided biopsy collection allows the surgeon to be more specific when taking a representative sample of the tumour, rather than collecting one which displays no tumour presence creating discrepancies between biopsy results and the medical imaging findings (52). Single-photon emission computed tomography (SPECT) can be used to differentiate between various grades of astrocytomas for example, whereby specific treatment plans affect overall patient outcomes; SPECT involves injecting the patient with a radioactive substance which interacts with healthy and cancerous cells followed by imaging with a standard CT scanner. Prior to surgery, a magnetic resonance angiography (MRA) scan would be performed to allow surgeons to visualise the blood vessels supplying a highly vascular brain tumour.

In addition to the discussed imaging techniques, it is routine for a chest X-ray to be performed to determine whether the patient has a primary lung cancer that has metastasised to the brain (38). Lung cancer is the most common primary site which will metastasise to the brain and is responsible for a reported 50.0 % of all brain metastases (11).

1.3.2.4 Biological Markers

Biological markers (biomarkers) directly reflect a spectrum of diseases from initial manifestation to more advanced stages. Biomarkers have previously been defined as “cellular, biochemical or molecular alterations that are measureable in biological media such as human tissues, cells or fluids” (53), thus biomarkers

secreted from tumours can be objectively measured to assist clinicians when identifying tumour development and aid in decision making for the most effective therapeutic pathway (54).

Biomarkers are invaluable tools for early disease diagnosis when combined with medical imaging techniques or when a multiple biomarkers assessment is performed; alone they are not universally accepted due to their low diagnostic ability and frequent false-positive results (55). Table 1.2 shows the glioma biomarkers routinely used for diagnosis and biological matter the biomarker can be found in. Biomarkers are used in hospitals for diagnostic and screening methods to identify a specific type of cancer in patients, thus biomarker analysis is only routinely given to patients whereby a specific type of cancer is suspected; however, with the complex mixture of symptoms that may arise with a brain tumour the possibility of misidentifying these symptoms is likely. Biomarkers are expected to have results with high sensitivities and specificities (54). Research has found that biomarkers for brain tumours can have sensitivities and specificities ranging from 40-88 % and 59-81 % respectively (56), although current diagnoses based on biomarkers generally provides high sensitivities and low specificities, followed by an array of advantages and disadvantages including highly labour intensive analysis (57).

le 1.2 - Glioma biomarkers used for diagnosis and biological matter the marker is identified in. Adapted from (54) .

Target Gene/Protein	Biomarker Localisation
Methlyguanine methyl transferase (MGMT)	Tumour tissue
Human cartilage glycoprotein-39 (YKL-40)	Serum or tumour tissue
Matrix metallopeptidase-9 (MMP-9)	Serum
Glial fibrillary acid protein (GFAP)	Serum or tumour tissue
Exosomes	Serum or tumour tissue
Growth arrest and DNA-damage-inducible protein (Gene: GADD45a)	Tumour tissue
Rad GTPase-activating-like protein (IQGAP1)	Tumour tissue
Insulin-like growth factor binding protein-2(IGFBP2)	Serum or tumour tissue
Receptor tyrosine phosphatase β (RPTPβ)	Tumour tissue

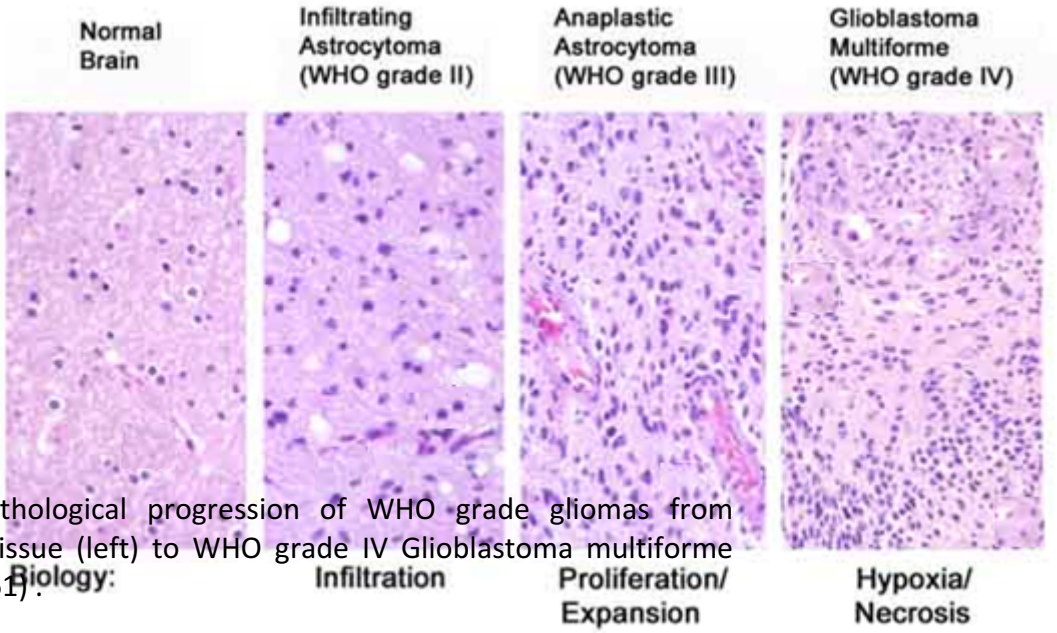
1.3.2.5 Histopathological Confirmatory Diagnosis

Following medical imaging, confirmatory diagnosis is necessary to determine whether the tumour present is malignant or benign. A biopsy is collected from the brain tumour by drilling into the skill and collecting a specimen (58), commonly via a

stereotactic technique which is known for its limitations. Stereotactic biopsies have been reported to lead to diagnostic errors resulting in patients being treated with incorrect therapies. A diagnosis based on a stereotactic biopsy is reported to be no greater than that of a diagnosis established based on patient history and other clinical findings (59). Following staining of the tissue with haematoxylin and eosin (H&E) to ease visualization of cellular components, the biopsy specimen is microscopically examined by a histopathologist to assess tissue morphology and architecture. The internationally accepted brain tumour grading system from the WHO provides histopathologists with guidelines when microscopically assessing tissue samples. The WHO describes the four grades of brain tumours as (60);

- Grade I - low proliferation; possibility of a cure following surgical resection
- Grade II - generally infiltrative (spreading into surrounding healthy tissues); relatively low-levels of proliferation; tends to possess lower and higher grades of malignancy depending on proliferation rate
- Grade III - definite evidence of histological malignancy (nuclear atypia)
- Grade IV - abundance of malignant cells; mitotically active; necrosis-prone neoplasms.

Figure 1.10 shows the histopathological progression from healthy brain tissue to WHO grade IV - glioblastoma multiforme. In grade II the astrocytoma cells present signs of multiplication and infiltration; grade III the cells begin to proliferate and, present signs of mass multiplication and become anaplastic; grade IV the cells present hyperplasia and mass necrosis (61).



igure 1.10 - Histopathological progression of WHO grade gliomas from ormal/healthy brain tissue (left) to WHO grade IV Glioblastoma multiforme right). Adapted from (61).

Histological grading does not provide clinicians with accurate prognostic and therapeutic details on an individual patient basis (62). The grading of a brain tumour is potentially underestimated or overestimated depending on the region of the tumour at which the biopsy was collected from (63). The location of the brain tumour determines which technique is performed to collect the biopsy. If the tumour is operable then a craniotomy to access and debulk the tumour may be conducted; however, if the tumour is located in an area classed as unsafe or inoperable then an X-ray guided stereotactic biopsy would be performed instead, whereby a small hole is drilled and a thin needle is guided to collect a tissue specimen (38).

The current diagnostic regime relies heavily on histopathological examination thus, diagnosis via histological examination of tissue is highly subjective (64); pathologic errors occur in a reported 1-43 % of all pathological specimens (65).

Treatment pathways are reliant on histopathological diagnosis; the misdiagnosis of a tumour could potentially severely affect a patient's overall prognostic outcome (38).

The ability to rapidly and accurately diagnose brain tumours has many benefits, including the early intervention of therapy leading to improved patient prognostic outcomes and relieving the burden on economic resources (13).

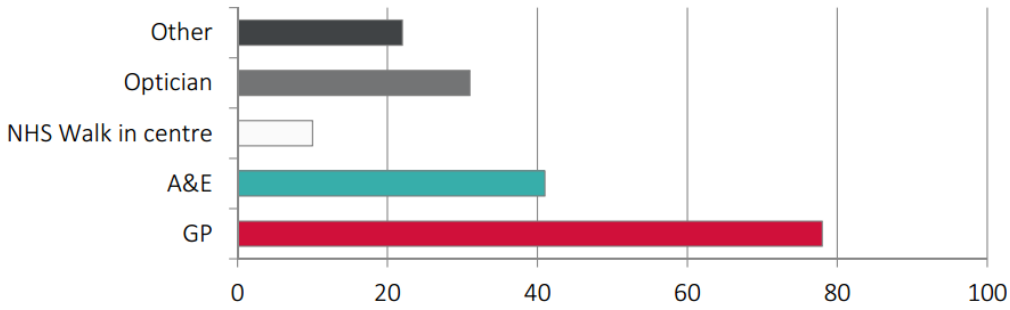
1.4 Problems with Current Brain Cancer Diagnosis

At present, England falls below many European countries in terms of overall cancer survival. The increased number of deaths due to cancer in England is due to delays in presentation, primary and secondary care delays, and delays in communication between the latter. Patients are commonly diagnosed following a visit to their GP via a two-week wait (TWW) referral system, whereby patients can expect to visit a specialist oncologist for further investigation within two weeks. A reported 38.0 % of people living with a brain tumour visited their GP on more than 5 occasions before finally being diagnosed (66), this suggests that the GP referral route is dangerously ineffective for brain tumours with opportunities to diagnose their tumour being missed (66). A GP consultation is a common route for brain tumour diagnosis, however, in many cases individuals are diagnosed via one of the following; an emergency presentation at an accident and emergency (A&E)/emergency room department; an urgent referral from a GP not made under

the TWW; detected via screening programmes (bowel, cervical, breast) or self-referral to a hospitals out-patient department. A study assessing 739,667 tumours from a range of cancers (prostate, ovary, CNS, to name a few) found that 24.0 % were diagnosed following emergency presentation; TWW and routing GP referral accounted for 26.0 and 21.0 % respectively. The one-year survival rate was significantly lower for those diagnosed at an emergency presentation compared with other diagnostic routes (67).

The point at which a brain tumour diagnosis occurs has a significant impact upon the patient's prognosis and overall outcomes. Unfortunately for the patient, symptoms of brain cancer are absent or minor during low-grade tumour development. Symptoms are more pronounced and troublesome when the tumour is advanced, high-grade and when prognosis/therapeutic intervention options are poor. Patients diagnosed at emergency presentations are commonly diagnosed with an advanced stage tumour, hence the poor one-year survival rate - treatment options may be limited, therapies ineffective or no treatments available whatsoever. The Brain Tumour Charity - the UK's largest dedicated brain tumour charitable organisation - reported that nationally some 62.0 % of people are diagnosed at an emergency presentation. Figure 1.11 shows the percentage of people in the UK who attended healthcare services available prior to diagnosis, including NHS walk in centres, opticians with eyesight problems and others referring to alternative diagnostic routes, such as outpatient appointments (66).

Figure 1.11 - The percentage of individuals with a brain tumour who attended a healthcare service on at least one occasion. Adapted from (66).



Although routinely performed, medical imaging has many benefits and a number of limitations. MRI is excellent for visualising the structures of the brain and other internal organs for tumour detection; however, false positives are not uncommon with the procedure (63). Furthermore, CT scans allow for highly detailed images of internal structures to be visualised, typically following iodine-based contrast dye being injected into the patient's blood stream to create contrast between tissues and blood vessels. Like MRI/fMRI, CT scans are subject to the experience and competency of the radiographer who is fallible.

Biological markers for brain tumour diagnosis are used to identify specific types of cancer following the presentation of symptoms suspected of cancer, although no marker is exactly specific for one type of disease. When used for diagnosis in isolation, biomarkers are not accepted universally due to their frequent false-positive results (55). Tumour markers have the potential to be elevated in patients who have a benign tumour, likewise, those who do have a malignant high-

grade tumour may not necessarily present markers in their tissue or blood to indicate the presence of a tumour. Biomarkers are best used when multiple biomarker analysis is performed alongside patient history, symptoms presented and medical imaging techniques. Despite the major limitations of tumour markers, researchers continue to focus on the discovery of new markers and the potential role they may have in screening, early cancer detection and treatment planning (68).

Patient treatment pathways are extremely reliant on histopathological analysis of a biopsy specimen; however, as previously discussed, these examinations are highly subjective with errors occurring in 1-43 % of cases (65). Overall patient prognostic outcomes rely on the correct treatment for the exact type and grade of tumour present within the brain of the suffering patient. Errors occurring with histopathological analysis can severely affect a patient's life (47).

Regardless of the problems with current brain tumour diagnoses, it is widely accepted that early diagnosis and early intervention of therapy greatly benefits patient survival prognostic outcomes and healthcare systems (13). As the majority of patients primarily attend a GP appointment (66) with symptoms not always suggestive of brain cancer, it would be extremely beneficial to both patients and healthcare providers if a rapid diagnostic tool was implemented in the GP's office for brain cancer detection. The diagnostic regime proposed throughout this thesis describes a rapid diagnostic methodology that would greatly reduce diagnosis

times, reduce mortality rates and has the potential to revolutionise the clinical environment (69).

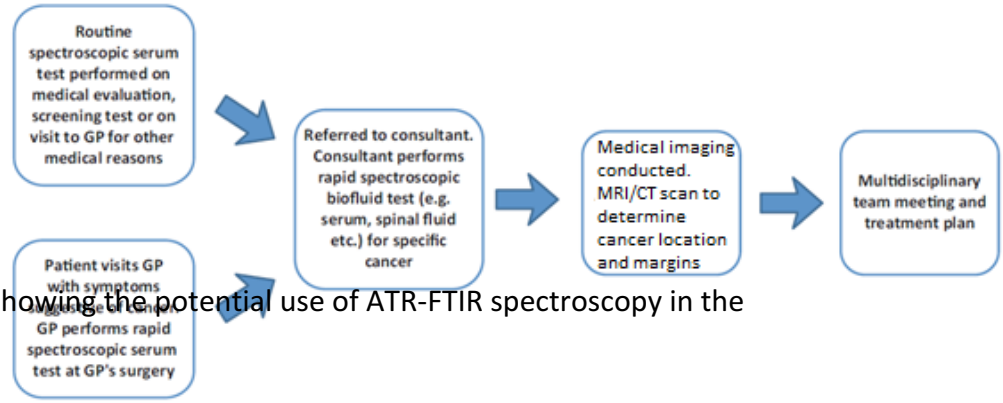
1.5 Use of Human Serum for Biological Analysis

1.5.1 Human Serum

Biofluids such as serum are easily accessible, collection is relatively non-invasive and can be performed worldwide. Human serum is a ubiquitous fluid that is routinely collected from patients for medical observations and investigations; the collection of serum is minimally invasive and its high availability makes it ideal for analysis. Serum contains more than 20,000 different proteins, the majority of the concentration from serum albumin (50g l^{-1}) and minute concentrations of 1ng l^{-1} of troponin. The low molecular weight fraction of human serum is referred to as the "peptidome" and it contains small biomolecules which are of particular interest in cancer diagnoses (70). The level of these proteins directly relates to specific disease states, thus human serum is highly suitable for cancer diagnostic purposes (71). Small molecules carried by blood serum contain a large amount of biological information, enabling this readily-accessible biofluid to be used for the early detection of a wide range of diseases. Serum has many advantages over other biological samples; repeat sampling and large sample quantities are both possible. The biochemical composition of serum is highly complex and it holds the ability to define function and phenotype, thus the potential to interpret or gain a chemical

signature unique to an individual's metabolome paves way for disease diagnosis, predictions and environmental factors contributing to disease. The biomarkers housed within serum are the focus of early detection, diagnosis, monitoring for disease recurrence, and efficiency monitoring of therapy (72).

Combining both human serum and ATR-FTIR spectroscopy offers point-of-care testing with minimal interpretation, thus allowing for serum spectroscopy to be welcomed within the clinic. ATR allows for biological fluids, such as blood serum, to be placed and dried directly on the crystal prior to spectral measurement. ATR-FTIR spectroscopy is ideal for the analysis of blood serum; it is well suited for thin film analysis which could potentially be directly translated into the clinical to make a real-world impact to patient health. In addition to its clinical suitability, ATR-FTIR spectroscopy is rapid, easy-to-use, cost effective and hand-held versions are available. The spectroscopic technique and proposed diagnostic procedure described within this thesis meets the requirements to allow it to be functional within the appropriate healthcare environments. Figure 1.12 shows a schematic of the potential use of ATR-FTIR spectroscopy in a clinical environment. When a patient visits their GP with symptoms suggestive of cancer, or for a routine medical evaluation, a spectroscopic measurement could be collected from a minute volume of blood serum for cancer detection. Following analysis, if required, the patient could be referred to the hospital for specialist oncology care where specific treatment plans and therapies can be determined (69).



ure 1.12 - Schematic showing the potential use of ATR-FTIR spectroscopy in the ıic (69).

1.5.2 Problems Associated with Serum Analysis

Although human serum is easily accessible, routinely collected and is ideal for rapid disease diagnoses, the biofluid is subject to a number of problems which the analyst should be aware of prior to spectral collection to avoid spectral distortions which are not characteristic of disease state. The issues surrounding the use of serum include the following: hemolysis (73), dilution (74), freezing, thawing/storage and transportation of the samples (75), sample preparation (76), histopathological diagnoses impacting upon disease grade of the serum sample (65) and a drying effect (77) known as the "coffee ring effect" (78).

The metabolome of human serum is varied in samples collected from patients who have different diets, exercise regimes, drug use, stress levels, age, gender, body mass index (BMI) and cardiac rhythm (73), thus these factors should be accounted for as much as possible when creating patient spectra datasets to avoid biases. Dunn *et al.* found metabolite relative concentration differences to

range from 5-200 % across 1200 patients; these differences could potentially be related to gender, age, BMI and smoking status of the blood donors (79).

Hemolysis is a large risk involved during the drawing of blood from a patient. This pre-analytical error can be avoided by careful drawing and handling of the blood sample. Hemolysis is caused through vigorous shaking of the sample tube, rough handling during transit, centrifuged at a too high speed and fluctuations in the environmental temperature at which the sample is being stored (75). Hemolysis is a process whereby the metabolites and enzymes of a sample are released from intracellular compounds, thus potentially altering the metabolomic profile of a blood sample. The increase in concentration of tryptophan and lipids in a blood sample could potentially be due to hemolysis. Yin *et al*. report that in ~18.0 % of their acquired spectral data the signals were due to the presence of extra metabolites and enzymes released during hemolysis (73); the impact of this occurring during vibrational spectroscopic analysis could significantly alter the research findings. The analyst who is conducting the spectral collection of the serum should be alert to the colour of all serum samples; haemolytic and non-haemolytic samples are easily distinguishable by their colour. The presence of free haemoglobin changes the colour of the serum from a straw-yellow to a bright red. Following blood collection from a patient, the handling and storage of the samples has the potential to alter quantitative and qualitative spectral observations (73).

Sample preparation including the dilution of whole, undiluted serum samples (74) is an area whereby research has been conducted (80). In their study,

Zhang *et al*. diluted each serum sample 2-fold with distilled water and allowed a volume of 5 µl microlitres to dry for 30 minutes. Sensitivities and specificities of 82.0 % and 82.5 % were achieved respectively for the diagnosis of hepatocellular carcinoma (80). Furthermore, it was found that a 3-fold dilution was most suitable for high-throughput Fourier transform spectroscopy (HT-FTIR) due to good spectral reproducibility and absorbance intensity (74). Further dilution of a serum sample past a 3-fold dilution induces the risk of a low spectral response, thus it would be unsuitable for the analysis of serum whereby small, yet significant, molecules would be at a low concentration. HT-FTIR found that undiluted whole serum samples saturate at the Amide I and Amide II bands (1700-1500 cm^{-1}) and at the H_2O band at ~3280 cm^{-1} (74). However useful HT-FTIR maybe in a pathological laboratory for the spectral analysis of multiple patients, the protocol would not be suitable for use in a doctor's office whereby time is limited and the extra steps needed for the implementation would deem the procedure as possibly troublesome, hence the drive for whole undiluted serum analytics for rapid spectral brain cancer diagnostics. In addition to dilution, the filtration of serum using molecular centrifugal devices has an impact upon spectral collection. The use of commercially available centrifugal devices allows for protein purification and molecular weight cut-off to create fractions with specific molecular weights. Bonnier *et al*. found that, when using Amicon® 0.5 devices, contaminants were introduced to the serum sample. Contaminants deriving from glycerine were observed in spectral data collected after the use of the centrifugal filters, thus the authors recommend the careful washing of the filters prior to analysis of bodily fluids (81).

Prior to the analysis of serum samples, it is assumed that the tumour type and grade of tumour within the patient from whom the blood was draw from is correct. However, as previously discussed, the error rates with histopathological diagnosis remain an issue with treatment plans and overall patient diagnosis (65), thus the incorrect diagnosis of a tumour has a direct impact upon the cancerous state of a serum sample. When a serum sample from a hospital tissue bank is received, it is classed as, for example, a cancerous WHO grade IV GBM; the spectral data from this sample would be grouped and classed to reflect this histopathological diagnosis. Furthermore, when a non-cancerous sample is collected, the process which the person donating the blood had prior to blood drawing is also of important consideration. Whether the person was in hospital for surgery unrelated to cancer; the person donated blood at a charity event to increase blood sample numbers and are unaware of any developing illnesses such as cancer, or whether the person donating blood does infact have a disease but it is unreported on hospital records and is classed as 'healthy non-cancerous' to name just a few, are issues surrounding non-cancerous samples. At present, all serum samples are received with histopathological diagnosis and often patient information (age, sex and disease grade etc.); trust is extended to the histopathologists who diagnose disease state and tumour grade (or severity).

A volume of serum dries in a non-homogenous fashion, raising concerns regarding the reproducibility and reliability of the data collected from the sample. This phenomenon is referred to as the "coffee-ring effect" (77) via capillary action as it displays an increased edge height and a significantly thinner surface towards

the centre (78). During the evaporation of water from the droplet, the induction of the Marangoni flow occurs leaving a concentrated, dense and ring-like structure along the perimeter of the sample (82). The effect of the coffee-ring formation is more pronounced during the drying of diluted serum samples, although the deposit of the drops morphology when drying is due to the hydrophobicity of the substrate, rather than the chemical composition of the serum. The cracks within the dried serum samples should be avoided when using FTIR and Raman microscopy due to the collection of abnormal Amide I and Amide II bands, and baseline issues (74). The inhomogeneous crater-shaped (coffee-ring) serum spot has a deep/broad outer ring which becomes increasingly shallower towards the centre point of the dried sample. The outer ring and inner surface has a depth of typically 8 and 5 microns respectively, determined via white light interferometry. Attenuated total reflection Fourier-transform infrared spectroscopy (ATR-FTIR) allows for acquired spectra to be representative of the sample upon the crystal, provided that the sample thickness exceeds the penetration depth of the evanescent wave. The 1 microlitre volume of serum extended over the area of the 1.8 mm ATR diamond crystal. In contrast to the ATR sampling method, transmission and transflection FTIR spectroscopy may give rise to standing wave artefacts and scattering effects from salt crystals formed during the drying process. Transmission and transflection mode FTIR also require specific sample thicknesses (76).

Ideally, the serum film on the ATR crystal would be homogenous; however, this is rarely met with manual pipetting devices. Ollesch *et al.* have developed a methodology whereby the homogeneity of a sample is extremely high due to

pipetting in nanolitre volumes with a costly, yet compact bench-top robotic system. The robotic system creates 50 nl serum films which avoid the formation of the coffee-ring effect. Following robotic serum deposition, the serum films were vacuum dried for 10 minutes. The dried serum films were ~3 µm in thickness (76), thus allowed the penetration of the 2-3 µm evanescent wave to penetrate the sample (83).

The advancement of robotic pipetting is exciting for a largely automated infrared (IR) diagnostic procedure, however, the suitability of this costly robotic device in a doctor's office remains to be seen.

1.6 Vibrational Spectroscopic Bioanalysis Applications

IR and Raman spectroscopy are both contenders to be implemented as routine point of care devices to aid diagnoses. Raman spectroscopy measured the exchange of energy with electromagnetic radiation of a certain wavelength. The exchange of energy with the sample results in a quantitative Raman shift in the same wavelength as that of the incident laser; this phenomenon is referred to as inelastic Raman scattering, a process whereby the incident particles energy is lost or increased (13). Molecular information can be obtained from a Raman spectrum, likewise, the conformation and composition of the sample is represented on the spectrum (84). A spectrum displaying the intensities and frequencies of the Raman bands are characteristic of the molecules within the analysed sample. The molecules within a sample, their concentrations within a molecule, conformations

and chemical bonding environments allowed for a Raman spectrum to be unique and characteristic to the particular sample being analysed (85).

The potential use of Raman spectroscopy as a diagnostic tool has been demonstrated by a range of studies with tissues, cells and biofluids for disease diagnosis involving: breast (86-88), brain (11, 89), asthma (90), liver (91-92) and skin (93), to name but a few. Harris *et al.* report the ability of Raman spectroscopy to discriminate between cancer and non-cancerous biological samples with sensitivities and specificities of 88.0 and 91.0 % respectively. Although successful when discriminating between disease states, Harris *et al*. note that they support the use of optical diagnosis to aid cancer diagnoses rather than replace the current pathological diagnostic route (94). Gajjar *et al.* observed significant spectral differences between brain tumour tissue and non-cancerous (healthy) brain tissue when using Raman spectroscopy. The spectral regions discriminating between cancer and non-cancerous samples include, but are not exclusive to, glycogen (~849 cm^{-1}), Amide I (1650 cm^{-1}) and phospholipids (~997 cm^{-1}) (89). Likewise, the potential of Raman spectroscopy as a diagnostic tool has been demonstrated using breast cancer patient serum samples combined with multivariate analysis (MVA). Pichardo *et al.* achieved a sensitivity of 97.0 and a specificity of 78.0 % when distinguishing between non-cancerous serum and breast cancer patient serum (95). An exciting development for Raman spectroscopy includes the use of an *in vivo* spectroscopic probe for surgical margin evaluation. An example of the use of Raman spectroscopy for tumour margin evaluation is with breast cancer surgery; the risk of local recurrence is strongly correlated with the presence of a malignant tumour to

just 1-2 millimetres (mm) of the surgical margin on the excised specimen, thus it is vital that the tumour is completely removed during surgery to prevent recurrence (96).

Like Raman, Fourier transform infrared (FTIR) spectroscopy is a rapid, reagent free, relatively inexpensive and non-destructive technique. FTIR spectroscopy is based on the molecular absorption of infrared (IR) radiation which caused the functional groups within a sample to vibrate in one of the following fashions: stretching, bending, deformation or combination, thus each molecule can produce a different spectrum which can be thought of as a fingerprint of the sample. The vibrations created due to the absorption of IR radiation will be discussed in an upcoming chapter. The absorptions and vibrations exhibited on a spectrum can be directly correlated to chemical species. Typically, the mid-infrared (mid-IR) region of the electromagnetic spectrum (4000-600 cm^{-1}) is the focus of many diagnostic studies. Spectra collected from the mid-IR region exhibit sharp, information-rich peaks due to the detection of the fundamental vibrations, in contrast to the near-infrared (NIR) region (14000-4000cm^{-1}) where overtone or harmonic vibrations would be detected (13).

Biological research performed with IR spectroscopy has focused on the diagnosis of scrapie infection (97), antemortem identification of bovine spongiform encephalopathy (98), fibromyalgia (99), arthritis (100), lung cancer (101), prostate cancer (102-105), cervical cancer (106), skin cancer (107), brain cancer (89, 108) and lung cancer diagnosis using sputum (109), to name just a few.

Backhaus *et al*. distinguished between non-cancerous (healthy) serum samples from serum collected from patients diagnosed with breast cancer; a sensitivity of 98.0 % and a specificity of 95.0 % was achieved when distinguishing between the cancerous and non-cancerous state serum samples. Backhaus *et al*. in the same study had shown the ability of IR spectroscopy to discriminate between breast cancer from other disease states such as Alzheimer's disease, hepatitis C and coronary heart disease using minute volumes of serum with sensitivities and specificities of 98.0 % and 100 % respectively (110). Disease states can be distinguished between with IR spectroscopy; Scaglia *et al.* successfully discriminated between serum spectra collected from patients with various degrees of hepatic fibrosis and normal (healthy) patients with a sensitivity of 95.2 % and a specificity of 100 % (111). Furthermore, whole serum and serum filtrate samples prepared following blood collection from patients diagnosed with one of the following: acute myocardial infarction, angina pectoris or other chest pain. The discrimination between these illnesses achieved a sensitivity and specificity of 88.5 % and 85.1 % respectively (112). The application of attenuated total reflectance (ATR), an FTIR sampling method suitable to dehydrate liquid serum due to the intense absorption of water, has been used by Gajjar *et al*. who was successfully able to differentiate between serum samples collected from patients diagnosed with either endometrial cancer, ovarian cancer or non-cancer. Results from this study were as high as 81.7 % for endometrial cancer and 96.7 % for ovarian cancer when discriminating against non-cancer control serum samples (108). One of the first studies using IR spectroscopy to diagnose disease with blood samples was able to achieve results up

to 97.6 %; whole blood samples were used in a proof-of-principle study to diagnose basal cell carcinoma. This study found that the most significant spectral band indicating the difference between cancer and non-cancer was the Amide at 1650 cm^{-1}. In another study, Petrich *et al*. observed spectral differences from serum obtained from healthy control samples and patients diagnosed with diabetes. It was determined that the significant spectral contributions when discriminating between diabetes patient serum from healthy controls is around 1026 cm^{-1}, assigned as the glucose absorption peak. A sensitivity and specificity of 80.0 % was achieved for the distinction between healthy controls and diabetes patient serum samples. (113). In addition to biofluid analysis, the potential of IR spectroscopy being used during surgery has been shown; Sobottka *et al.* shown that glioma grading is possible with IR spectroscopy and demonstrated its suitability when deciding whether or not to continue with surgical tumour resection. Tumour margins can be defined using this intraoperative tool during cerebral glioma surgery which, if implemented, would save many lives due to the detection of tumour residues which may otherwise not have been removed leading to a recurrence of the tumour. Sobottka *et al*. achieved sensitivities and specificities of 100 % and 96.9 % respectively when distinguishing between glioma residue tissues from healthy non-cancerous tissues (114). Khoshmanesh *et al*. used ATR-FTIR spectroscopy to diagnose malaria using parasite infested blood samples. The use of ATR-FTIR spectroscopy demonstrated the potential of its use as a rapid detective and quantitative tool for parasitic malarial infections. Upon deposition of the blood sample to the ATR diamond (Di), the blood was rapidly dried for 1 minute (for sample dehydration) using a blow drier (83).

The use of IR spectroscopy as a routine method of diagnoses would significantly reduce mortality and morbidity rates as well as freeing up vital economic resources. The ability to identify a disease such as brain cancer early would allow for the early intervention of therapy, thus improving patient prognosis rates. IR spectroscopy is a rapid, reagent free and non-destructive technique which has potential for routine use in a clinical practice (13) (Figure 1.12).

1.7 Aims of Research Project

The fundamental aim of the research described within this thesis is to develop a spectroscopic methodology to rapidly diagnose a cancerous disease state from a serum profile.

The initial intention of this research in the developmental stage is to primarily determine whether it is possible to differentiate between serum samples collected from either glioma or non-cancer patients with a high sensitivity and specificity using ATR-FTIR spectroscopy. Building upon the success of the ability to differentiate between cancer and non-cancerous state serum samples, the conduction of a number of studies to determine the potential use of the regime in a clinical environment is priority, thus serum drying; spectral variance and serum filtration studies are to be investigated.

Progressing the study would involve the addition of patients with various grades of brain tumour. The ability to rapidly differentiate between brain tumour

grade (high grade glioma and low grade glioma) from non-cancer would have a significant impact upon patient mortality and morbidity rates and the early intervention of therapy (13). Furthermore, the successful demonstration to identify a metastatic brain tumour which originated from outside of CNS from a simple spectroscopic method is exciting and would allow clinicians to intervene with therapy and have an understanding of the primary tumour location prior to medical imaging; currently 50.0 % of metastatic brain tumour patients have their primary site diagnosed at a later date (26).

To determine the spectroscopic regimes potential use in the clinical environment with unseen patient serum samples, the technique will be implemented with new, blind samples from a clinical collaborator whereby the serum samples have never been analysed previously. The creation of spectral datasets to diagnose disease state from a patient's blood serum is an exciting concept with many real-world applications.

Spectroscopic data collected from a wide range of patients with various types and grades of cancer or non-cancer will be used to evaluate the use of ATR-FTIR spectroscopy as a diagnostic tool utilising multivariate analysis methods and novel feature extraction processes.

References

[1] Knowles, M., Selby, P. *Introduction to the cellular and molecular biology of cancer*. Oxford University Press, 2005

[2] Ahmad, A S., Ormiston-Smith, N., Sasieni, P D. *Trends in the lifetime risks of developing cancer in Great Britain: Comparison of risk for those born in 1930 to 1960*. British Journal of Cancer, 2015. 112(5), 943-947

[3] Cancer Research UK. Cancer Statistics for the UK [Online] *Cancer Research UK:* http://www.cancerresearchuk.org/health-professional/cancer-statistics [Accessed on 6th July 2015]

[4] Popat, K., McQueen, K., Feeley T W. *The global burden of cancer*. Best Practice & Research: Clinical Anaesthesiology, 2013. 27(4), 399-408

[5] Hesketh, R. *Introduction to cancer biology*. Cambridge University Press, 2012

[6] Cancer Research UK. How Cancers Grow [Online] *Cancer Research UK:* http://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancers-grow [Accessed on 6th July 2015]

[7] Brain Tumour Research. *Report on National Research Funding [July]* [Online] *Brain Tumour Research*: http://www.braintumourresearch.org/uploads/document/BrainTumourResearchNationalFundingReportJuly_2013draft9_897.pdf [Accessed on 6th July 2015]

[8] McKinney, P A. *Brain tumours: incidence, survival, and aetiology*. Journal of Neurology, Neurosurgery, and Psychiatry, 2004. 75(2), 12-17

[9] Walid, M S*. Prognostic factors for long-term survival after glioblastoma*. The Permanente Journal, 2008. 12(4), 45-48

[10] Office for National Statistics. *Brain cancer rates have risen by a quarter over the past three decades*[Online] *Office for National Statistics (ons.gov.uk):* http://www.ons.gov.uk/ons/rel/vsob1/cancer-statistics-registrations--england--series-mb1-/no--41--2010/sty-brain-cancer-awareness.html

[11] Fullwood L M., Clemens G., Griffiths D., Ashton K., Dawson T P., Lea R W., Davis C., Bonnier F., Byrne H J., Baker M J. *Investigating the use of Raman and immersion Raman spectroscopy for spectral histopathology of metastatic brain cancer and primary sites of origin*. Analytical Methods, 2014. 6(12), 3948-3961

[12] National Institute of Health and Care Excellence. *News and features*[Online] *National Institute of Health and Care Excellence:* https://www.nice.org.uk/news/article/nice%E2%80%99s-new-

symptom%E2%80%93based-approach-will-help-to-save-thousands-of-lives-from-cancer [Accessed on 6th July 2015]

[13] Ellis, D I., Goodacre, R. *Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy*. Analyst, 2006. 131(8), 875-885

[14] Dharani, K. *The biology of thought - a neuronal mechanism on the generation of thought: a new molecular model*. Academic Press Inc., 2014

[15] Hatzikirou, H., Deutsch, A., Schaller C., Simon, M., Swanson, K. *Mathematical modelling of glioblastoma tumour development: a review*. Mathematical Models and Methods in Applied Sciences, 2005. 15(11), 1779-1794

[16] Grant, R. Overview: brain tumour diagnosis and management/royal college of physicians guidelines. Journal of Neurology, Neurosurgery, and Psychiatry, 2004. 75(2), 18-23

[17] Jessen, K R. *Glial cells*. The International Journal of Biochemistry and Cell Biology, 2004. 36, 1861-1867

[18] Bernstein, M., Berger M S. *Neuro-oncology: the essentials*. Thieme, 2008

[19] Cancer Research UK. *The Brain*. [Online] *Cancer Research UK:* http://www.cancerresearchuk.org/about-cancer/type/brain-tumour/about/the-brain [Accessed on 6th July 2015]

[20] Chen, W. *Clinical applications of PET in brain tumours*. The Journal Of Nuclear Medicine, 2007. 48(9), 1468-1481

[21] Robert, M C., Wastie, M L. *Glioblastoma multiforme: a rare manifestation of extensive liver and bone metastases*. Biomedical Imaging and Intervention Journal, 2008. 4(1), E3

[22] Wang, Y C., Chuang, C C., Wei, K C., Hsu, Y H., Hsu, P W., Lee, S T., Wu, C T., Tseng, C K., Wang, C C., Chen, Y L., Jung, S M., Chen, P Y. S*kull base atypical meningioma: Long term surgical outcome and prognostic factors*. Clinical Neurology and Neurosurgery, 2015. 128,112-116

[23] Hengel, K., Sidhu, G., Choi, J., Weedon, J., Nwokedi, E., Axiotis, C A., Song, X., Braverman, A S. *Attributes of brain metastases from breast and lung cancer*. International Journal of Clinical Oncology, 2013. 18(3), 396-401

[24] Peacock, K H., Lesser, G J. *Current therapeutic approaches in patients with brain metastases*. Current Treatment Options in Oncology, 2006. 7(6), 479-489

[25] Cancer Research UK. *How can cancers spread?*[Online] *Cancer Research UK:* http://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancer-can-spread [Accessed on 6th July 2015]

[26] Polyzoidis, K S., Miliaras, G., Pavlidis, N. *Brain metastasis of unknown primary: a diagnostic and therapeutic dilemma*. Cancer Treatment Reviews, 2005. 31(4), 247-255

[27] Sun, Y., Peng Z L. *Programmed cell death and cancer*. Postgraduate Medical Journal, 2009. 85(1001), 134-140

[28] Tower, J. *Programmed cell death in aging*. Ageing Research Reviews, 2015. 23(A), 90-100

[29] Andreeff, M., Goodrich, D W., Pardee, A B. *Chapter 3 - Cell proliferation and differentiation in Holland-Frei cancer medicine.* BC Decker Inc., 2003

[30] U.S National Library of Medicine. *The Process of Apoptosis* [Online] *Genetics Home Reference - Understanding Genetic Conditions:* http://ghr.nlm.nih.gov/handbook/illustrations/apoptosisprocess [Accessed on 6th July 2015]

[31] Dragan, Y P., Pitot, H C. *The role of the stages of initiation and promotion in phenotypic diversity during hepatocarcinogenesis in the rat*. Carcinogenesis, 1992. 13(5), 739-750

[32] Weston, A., Harris, C. *Chapter 17 - Chemical carcinogenesis in Holland-Frei cancer medicine. BC Decker Inc., 2003*

[33] Rundhaug, J E., Fischer, SM. *Molecular mechanisms of mouse skin tumor promotion*. Cancers, 2010. 2(2), 436-482

[34] Raslan, A., Bhardwaj, A. *Medical management of cerebral edema*. Neurosurgical Focus, 2007. 22(5), E12

[35] Hewitt, A., Ellory, C. *Brain oedema, intracranial pressure and cerebral blood flow*. Surgery, 2012. 30(3), 102-106

[36] Andrzejowski, J., Mateszko, O. *Seizures in critical care. A guide to diagnosis and therapeutics*. Humana Press, 2009

[37] Kaal, E C., Vecht C J. *The management of brain edema in brain tumours.* Current Opinion in Oncology, 2004. 16(6), 593-600

[38] Medifocus Inc. *Medifocus Guidebook on: Glioblastoma*. CreateSpace Independent Publishing Platform, 2014

[39] Cancer Research UK. *Statistics and outlook for brain tumours*[Online] *Cancer Research UK:* http://www.cancerresearchuk.org/about-cancer/type/brain-tumour/treatment/statistics-and-outlook-for-brain-tumours#glioma [Accessed on 6th July 2015]

[40] Chang E L., Lo S. *Diagnosis and management of central nervous system metastases from breast cancer.* Oncologist, 2003. 8(5), 398-410

[41] Rees J. *Neurological oncology*. Medicine, 2008. 32(10), 75-79

[42] Saita L., Polastri D., De Conno F. *Visual disturbances in advanced cancer patients: clinical observations*. Journal of Pain Symptom Management, 1999. 17(3), 224-226

[43] Fazel Zarandi M H., Zarinbal M., Izadi M. *Systematic image processing for diagnosting brain tumors Type-II fuzzy expert system approach*. Applied Soft Computing, 2011. 11(1), 285-294

[44] Rillo C., Gabal M., Lozano M P., Sese J., Spagna S., Diederichs J., Sager R., Chialvo C., Terry J., Rayner G., Warburton R., Reineman R. *Enhancement of the liquefaction rate in small-scale helium liquefiers working near and above the critical point*. Physics, 2015. 3(5), 051001

[45] Campos S., Davey P., Hird A., Pressnail B., Bilbao J., Aviv R I., Symons S., Pirouzmand F., Sinclair E., Culleton S.,, DeSa E., Goh P., Chow E. *Brain metastasis from an unknown primary, or primary brain tumour? A diagnostic dilemma*. Current Oncology, 2009. 16(1), 62-66

[46] Smits M. *Functional magnetic resonance imaging (fMRI) in brain tumour patients*. European Association of NeuroOncology Magazine, 2012. 2(3), 123-128

[47] Stupp R., Brada M., van den Bent M J., Tonn J C., Pentheroudakis G. *High-grade glioma: ESMO clinical practice guidelines for diagnosis, treatment and follow-up*, 2014. 21(5), 190-193

[48] National Health Service. *CT Scan - how is it performed?*[Online] *NHS Choices:* http://www.nhs.uk/Conditions/CT-scan/Pages/How-is-it-performed.aspx [Accessed 6th July 2015]

[49] Lubell D L. *Drawbacks and limitations of computed tomography*. Texas Heart Institute Journal, 2005. 32(2), 250

[50] Guttikonda R., Herts B R., Dong F, Baker M E., Fenner K B., Pohlman B. *Estimated radiation exposure and cancer risk from CT and PET/CT scans in patients with lymphoma*. European Journal of Radiology, 2014. 83(6), 1011-1015

[51] Muratore C S*. Pediatric abdominal CT scans: do it correctly. Better yet, don't do it at all*. Journal of Surgical Research, 2013. 185(2), 533-534

[52] Kobayashi K., Bhargava P., Raja S., Nasseri F., Al-Balas H A., Smith D D., George S P., Vij M S. *Image-guided biopsy: what the interventional radiologist needs to know about PET/CT*. Radiographics, 2012. 32(5), 1483-1501

[53] Mayeux R. *Biomarkers: potential uses and limitations*. NeuroRx, 2004. 1(2), 182-188

[54] Liang S., Shen G. *Biomarkers of glioma* in *Molecular Targets of CNS Tumours*. InTech Open Science, 2011

[55] Diamandis E P. *Mass spectrometry as a diagnostic and cancer biomarker discovery tool: opportunities and potential limitations*. Molecular Cell Proteomics, 2004. 3(4), 367-378

[56] Hands J R., Abel P., Ashton K., Dawson T., Davis C., Lea R W., McIntosh A J., Baker M J. *Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors*. Analytical and Bioanalytical Chemistry, 2013. 405(23), 7347-7355

[57] Barh D., Carpi A., Verma M., Gunduz M. *Cancer biomarkers: minimal and noninvasive early diagnosis and prognosis.* CRC Press, 2014

[58] Chen C C., Hsu P W., Erich Wu T W., Lee S T., Chang C N., Wei K., Chuang C C., Wu C T., Lui T N., Hsu T H., Lin T K., Lee S C., Huang Y C. *Stereotactic brain biopsy: single center retrospective analysis of complications*. Clinical Neurology and Neurosurgery, 2009. 111(10), 835-839

[59] Vaquero J., Martinez R., Manrique M. *Stereotactic biopsy for brain tumours: is it always necessary?* Surgical Neurology, 2000. 53(5), 432-437

[60] Louis D N, Ohgaki H., Wiestler O D., Cavenee W K., Burger P C., Jouvet A., Scheithauer B W., Kleihues P. *The 2007 WHO classification of tumours of the central nervous system*. Acta Neuropathologica, 2007. 114(2), 97-109

[61] BioScience. *Genetic modulation of hypoxia induced gene expression and angiogenesis: relevance to brain tumors*[Online] *Frontiers in* Bioscience: https://www.bioscience.org/2003/v8/d/942/fulltext.php?bframe=figures.htm [Accessed on 6th July 2015]

[62] Iyer, R V., Golash, A., Lea, R W., Davis, C., Roberts, P. *Diagnosis and treatment of malignant glioma* in *The molecular and cellular pathology of cancer progression and prognosis*. Research Signpost, 2006

[63] Herholz K., Langen K., Schiepers C., Mountz J M. *Brain Tumours*. Nuclear Medicine, 2012. 42(6), 356-370

[64] Krafft C., Sobottka S B., Schackert G., Salzer R. *Analysis of human brain tissue, brain tumors and tumour cells by infrared spectroscopic mapping*. Analyst, 2004. 129(10), 921-925

[65] Raab S S., Grzybicki D M., Janosky J E., Zarbo R J., Meier F A., Jensen C., Geyer S J. *Clinical impact and frequency of anatomic pathology errors in cancer diagnoses*. Cancer, 2005. 15(104), 2205-2213

[66] The Brain Tumour Charity. *Finding a better way? Improving the quality of life for people affects by brain tumours* [Online] http://www.thebraintumourcharity.org/Resources/SDBTT/news/documents/the-brain-tumour-charity-report-on-improving-quality-of-life-final-report-dec2013.pdf [Accessed on 6th July 2015]

[67] Elliss-Brookes L., McPhail S., Ives A., Greenslade M., Shelton J., Hiom S., Richards M. *Routes to diagnosis for cancer - determining the patient journal using multiple routine data sets*. British Journal of Cancer, 2012. 107(8), 1220-1226

[68] Marella S. *Prognostic and predictive markers in early detection of different types of cancers for selected organ sites*. Journal of Pharmacy and Biological Sciences, 2013. 8(4), 25-42

[69] Dorling K M., Baker M J. *Highlighting attenuated total reflection Fourier transform infrared spectroscopy for rapid serum analysis*. Trends of Biotechnology, 2013. 31(6), 327-328

[70] Bonnier F., Petitjean F., Baker M J., Byrne H J. *Improved protocols for vibrational spectroscopic analysis of body fluids*. Journal of Biophotonics, 2014. 7(3-4)

[71] De M., Rana S., Akpinar H., Miranda O R., Arvizo R R., Bunz U H F., Rotello V M. *Sensing of proteins in human serum using nanoparticle-green fluorescent protein conjugates.* Nature Chemistry, 2009. 1(6), 461-465

[72] Zhang A H., Sun H., Yan G L., Han Y., Wang X J. *Serum proteomics in biomedical research: a systematic review*. Applied Biochemistry and Biotechnology, 2013. 170(4), 774-786

[73] Yin P., Lehmann R., Xu G. *Effects of pre-analytical processes on blood samples used in metabolomics studies*. Analytical and Bioanalytical Chemistry, 2015. 407(17), 4879-4892

[74] Lovergne L., Clemens G., Lukaszewski R., Untereiner V., Sockalingum G D., Baker M J. *Investigating optimum sample preparation for infrared spectroscopic serum diagnostics*. Analytical Methods, 2015. DOI: 10.1039/C5AY00502G

[75] Mitchell B L., Yasui Y., Ci C., Fitzpatrick A L., Lampe P D. *Impact of freeze-thaw cycles and storage time on plasma samples used in mass spectrometry based biomarker discovery projects*. Cancer Information, 2005. 1, 98-104

[76] Ollesch J., Drees S L., Heise H M., Behrens T., Bruning T., Gerwert K. *FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification.* Analyst, 2013. 138(14), 4092-4102

[77] Deegan R D., Bakajin O., Dupont T F., Huber G., Nagel S R., Witten T A. *Capillary flow as the cause of ring stains from dried liquid drops.* Nature, 1997. 389, 827-829

[78] Ristenpart W D., Kim P G., Domingues C., Wan J, Stone H A. *Influence of substrate conductivity on circulation reversal in evaporating drops*. Physical Review Letters, 2007. 99(23), 234502

[79] Dunn W B., Lin W., Broadhurst D., Begley P., Brown M., Zelena E., Vaughan A A., Halsall A., Harding N., Knowles J D., Francis-McIntyre S., Tseng A., Ellis D I., O'Hagan S., Aarons G., Benjamin B., Chew-Graham S., Moseley C., Potter P., Winder C L., Winder C L., Potts C., Thornton P., McWhirter C., Zubair M., Pan M., Burns A., Cruickshank J K., Jayson G C., Purandare N., Wu F C., Finn J D., Haselden J N., Nicholls A W., Wilson I D., Goodacre R., Kell D B. *Molecular phenotyping of a UK population: defining the human serum metabolome*. Metabolomics, 2015. 11,9-26

[80] Zhang X., Thiefin G., Gobinet C., Untereiner V., Taleb I., Bernard-Chabert B., Heurgue A., Truntzer C., Ducoroy P., Hillon P., Sockalingum G D. *Profiling serological biomarkers in cirrhotic patients via high-throughput Fourier transform infrared spectroscopy: toward a new diagnostic tool of hepatocellular carcinoma*. Translational Research, 2013. 162(5), 279-286

[81] Bonnier F., Baker M J., Byrne H J. *Vibrational spectroscopic analysis of body fluids: avoiding molecular contamination using centrifugal filtration*. Analytical Methods, 2014. 14, 5155-5160

[82] Ristenpart W D., Kim P G., Domingues C., Wan J., Stone H A. *Influence of substrate conductivity on circulation reversal in evaporating drops.* Physics Review Letters, 2007. 99(23), 234502

[83] Khoshmanesh A., Dixon M W., Kenny S., Tilley L., McNaughton D., Wood B R. *Detection and quantification of early-stage malaria parasites in laboratory infected erythrocytes by attenuated total reflectance infrared spectroscopy and multivariate analysis*. Analytical Chemistry, 2014. 86(9), 4379-4386

[84] Devpura S., Thakur J S., Sarkar F H., Sakr W A., Naik V M., Naik R. *Detection of benign epithelia, prostatic intraepithelial neoplasia, and cancer regions in radical prostatectomy tissues using Raman spectroscopy*. Vibrational Spectroscopy, 2010. 53(2), 227-232

[85] Wartewig S., Neubert R H H. *Pharmaceutical applications of Mid-IR and Raman spectroscopy*. Advanced Drug Delivery Reviews, 2005. 57(8), 1144-1170

[86] Stone N., Matousek P. *Advanced transmission Raman spectroscopy: a promising tool for breast disease diagnosis*. Cancer Research, 2008. 68, 4428

[87] Stone N., Baker R., Rogers K., Parker A W., Matousek P. *Substrate probing of calcifications with spatially offset Raman spectroscopy (SORS): future possibilities for the diagnosis of breast cancer*. Analyst, 2007. 132(9), 899-905

[88] Pichardo-Molina J L., Frausto-Reyes C., Barbosa-Garcia O., Huerta-Franco R., Gonzalex-Trujillo J L., Ramirez-Alvarado C A., Gutierrez-Juarez G., Medina-Gutierrez C. *Raman spectroscopy and multivariate analysis of serum sample from breast cancer patients*. Lasers Medical Science, 2007. 22, 229-236

[89] Gajjar K., Heppenstall L D., Pang W., Ashton K M., Trevisan J., Patel I I., Llabjani V., Stringfellow H F., Martin-Hirsch P L., Dawson T., Martin F L. *Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis*. Analytical Methods, 2013. 1(5), 89-102

[90] Sahu A., Dalal K., Naglot S., Aggarwal P., Murali Krishna C. *Serum based diagnosis of asthma using Raman spectroscopy: an early phase pilot study*. PLoS ONE, 2013. 8(11), e78921

[91] Shen A., Liao Z., Wang H., Goan I., Wu Y., Wang X., Yu Z., Hu J. *Study on the in vitro and in vivo activation of hepatic stellate cells by Raman spectroscopy*. Journal of Biomedical Optics, 2007. 12(3), 034003

[92] Wu Y M., Chen H C., Chang W T., Jhan J W., Lin H L., Liau I. *Quantative assessment of hepatic fat of intact liver tissues with coherent anti-stokes Raman scattering microscopy*. Analytical Chemistry, 2009. 81(4), 1496-1504

[93] Najssen A., Maquelin K., Santos L F., Caspers P J., Bakker Schut T C., den Hollander J C., Neumann M H., Puppels G J. *Discriminating basal cell carcinoma from*

perilesional skin using high wave-number Raman spectroscopy. Journal of Biomedical Optics, 2007. 12(3), 034004

[94] Harris A T., Rennie A., Waqar-Uddin H., Wheatley S R., Ghosh S K., Martin-Hirsch D P., Fisher S E., High A S., Kirkham J., Upile T. *Raman spectroscopy in head and neck cancer.* Head and Neck Oncology, 2010. 8(1), 126-132

[95] Pichardo-Molina J L., Frausto-Reyes C., Barbosa-Garcia O., Huerta-Franco R., Gonzalez-Trujillo J L., Ramirez-Alvarado C A., Gutierrez-Juarez G., Medina-Gutierrez C. *Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients*. Lasers Medical Science, 2007. 22(4), 229-236

[96] Keller M D., Vargis E., de Matos Granja N., Wilson R H., Mycek M A., Kelley M C., Mahadevan-Jansen A. *Development of a spatially offset Raman spectroscopy probe for breast tumour surgical margin evaluation*. Journal of Biomedical Optics, 2011. 16(7), 077006

[97] Schmitt J., Beekes M., Brauer A., Udelhoven T., Lasch P., Naumann D. *Identification of scrapie infection from blood serum by Fourier transform infrared spectroscopy*. Analytical Chemistry, 2002. 74(15), 3865-3868

[98] Lasch P., Schmitt J., Beekes M., Udelhoven T., Eiden M., Fabian H., Petrich W., Naumann D. *Antemortem identification of bovine spongiform encephalopathy from serum using infrared spectroscopy*. Analytical Chemistry, 2003. 75(23), 6673-6678

[99] Hackshaw K V., Rodriguez-Saona L., Plans M., Bell L N., Buffington C A. *A bloodspot-based diagnostic test for fibromyalgia syndrome and related disorders*. Analyst, 2013. 138(16), 4453-4463

[100] Eysel H H., Jackson M., Nikulin A., Somorjai R L., Thomson G T D., Mantsch H H. *A novel diagnostic test for arthritis: multivariate analysis of infrared spectra of synovial fluid*. Biospectroscopy, 1997. 3(2), 161-167

[101] Bird B., Miljkovic M S., Remiszewski S., Akalin A., Kon M., Diem M. *Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer*. Laboratory Investigations, 2012. 92(9), 1358-1373

[102] Gazi E., Baker M J., Dwyer J., Lockyer N P., Gardner P., Shanks J H., Reeve R S., Hart C A., Clarke N W., Brown M D. *A correlation of FTIR spectra derived from prostate cancer biopsies with Gleason grade and tumour stage*. European Urology, 2006. 50(4), 750-761

[103] Baker M J., Clarke C., Demoulin D., Nicholson J M., Lyng F M., Byrne H J., Hart C A., Brown M D., Clarke N W., Gardner P. *An investigation of the RWPE prostate derived family of cell lines using FTIR spectroscopy*. Analyst, 2010. 135, 887-894

[104] Baker M J., Gazi E., Brown M D., Shanks J H., Gardner P., Clarke N W. *FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer*. British Journal of Cancer, 2008. 99(11), 1859-1866

[105] Baker M J., Gazi E., Brown M D., Shanks J H., Clarke N W., Gardner P. *Investigating FTIR based histopathology for the diagnosis of prostate cancer*. Journal of Biophotonics, 2009. 2(1-2), 104-113

[106] Lyng F M., Faolain E O., Conroy J., Meade A D., Knief P., Duffy B., Hunter M B., Byrne J M., Kelehan P., Byrne H J. *Vibration spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool*. Experimental and Molecular Pathology, 2007. 82(2), 121-129

[107] Pallua J D., Pezzei C., Zelger B., Schaefer G., Bittner L K., Huck-Pezzei V A., Schoenbichler S A., Hahn H., Kloss-Brandstaetter A., Kloss F., Bonn G K., Huck C W. *Fourier transform infrared imaging analysis in discrimination studies of squamous cell carcinoma.* Analyst, 2012. 137(17), 3965-3974

[108] Gajjar K., Trevisan J., Owens G., Keating P J., Wood N J., Stringfellow H F., Martin-Hirsch P L., Martin F L. *Fourier-transform infrared spectroscopy coupled with a classification machine of the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer*. Analyst, 2013. 138(14), 3917-3926

[109] Lewis P D., Lewis K E., Ghosal R., Bayliss S., Lloyd A J., Wills J., Godfrey R., Kloer P., Mur L A. *Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum*. BMC Cancer, 2010. 10, 640

[110] Backhaus J., Mueller R., Formanski N., Szlama N., Meerpohl H., Eidt M., Bugert P. *Diagnosis of breast cancer with infrared spectroscopy from serum samples*. Vibrational Spectroscopy, 2010. 52(2), 173-177

[111] Scaglia E., Sockalingum G D., Schmitt J., Gobinet C., Schneider N., Manfait M., Thiefin G. *Noninvasive assessment of hepatic fibrosis in patients with chronic hepatitis C using serum Fourier transform infrared spectroscopy*. Analytical and Bioanalytical Chemistry, 2011. 401(9), 2919-2925

[112] Petrich W., Lewandrowski K B., Muhlestein J B., Hammond M E., Januzzi J L., Lewandrowski E L., Pearson R R., Dolenko B., Fruh J., Haass M., Hirschl M M., Kohler W., Mischler R., Mocks J., Ordonez-Llanons J., Quarder O., Somorjai R., Staib A., Sylven C., Werner G., Zerback R. *Potential of mid-infrared spectroscopy to aid the triage of patients with acute chest pain*. Analyst, 2009. 134(6), 1092-1098

[113] Petrich W., Staib A., Otto M., Somorjai R L. *Correlation between the state of health of blood donors and the corresponding mid-infrared spectra of the serum*. Vibrational Spectroscopy, 2002. 28(1), 117-129

[114] Krafft C., Sobottka S B., Geiger K D., Schackert G., Salzer R. *Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis*. Analytical and Bioanalytical Chemistry, 2007. 387(5), 1669-1677

[115] European Randomized Study of Screening for Prostate CancerERSPC. [Online] *European Randomized Study of Screening for Prostate* Cancer: http://www.erspc.org/prostate-cancer/erspc-background/ [Accessed on 8th July 2015]

[116] Mol B W., Lijmer J G., Evers J L., Bossuyt P M. *Characteristics of good diagnostic studies*. Seminars in Reproductive Medicine, 2003. 21(1), 17-26

Figure2.1 – The electric vector of a light wave. The horizontal arrow at zero on the *y*-axis is the direction of the undulating wave. Replicated from (1).

Chapter 2

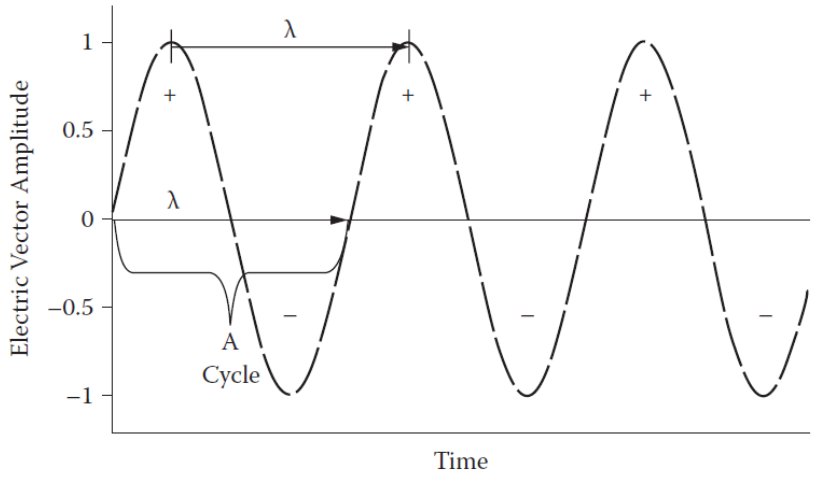
FUNDAMENTALS AND THEORY OF EXPERIMENTATION

This chapter will discuss the fundamental principles of infrared spectroscopy and of the pre-processing/data analysis methods used during this project. The fundamental principles of infrared spectroscopy will be linked to the process where interpretable spectra are acquired.

2 Infrared Spectroscopy

2.1 Fundamentals of the Electromagnetic Spectrum

A wave of electromagnetic radiation (EMR) is composed of electric and magnetic waves, commonly referred to as the electric and magnetic vectors. The electric and magnetic waves undulate, the two waves travel in planes mutually perpendicular to one other. Figure 2.1 shows the electric vector travelling in a sine wave. The amplitude of the electric vector changes over time; the sine wave has positive and negative troughs to reflect how the wave changes (1).



The motion of the sine wave is repetitive, hence the wave undergoes cycles. A cycle occurs when the sine wave begins with an amplitude of zero and crosses the zero amplitude point for a third time. Figure 2.1 illustrates the cycle of a wave.

Wavelength (λ) defines the distance travelled by a wave during a cycle; the speed of each particular wave is constant. The mid-infrared region of the electromagnetic (EM) spectrum involves radiation of approximately 10 microns. In contrast to wavelength, wavenumber measures the number of cycles which a wave encounters per unit length (measured in units of cycles per centimetres (cm))(1).

Wavelength is measured in distance per cycle and wavenumber in cycles per distance, thus the two terms are reciprocals of one another. Equation 2.1 presents how wavenumber and wavelength are reciprocals of one another where $\bar{\nu}$ = wavenumber; λ = wavelength) (1):

$$\bar{\nu} = \frac{1}{\lambda} \qquad \text{Eq. 2.1}$$

Wavenumber is proportional to the energy of a light wave. A high wavenumber peak, for example, has a higher energy resulting in the wave undergoing more cycles (1). A mid-infrared spectrum has wavenumber frequencies between 4000-400 cm^{-1} (3-30 microns), however; it is the fingerprint region at approximately 1800-1000 cm^{-1} which much research focuses on due to it being the region where complex molecular stretches and bends occur. Every compound has a different absorption pattern in this region, giving rise to the term "fingerprint region" (1).

The velocity of a single wave is measured in centimetres per second as it travels at the speed of light (c = ~3.0 X 10^{10} cm/second). A wave can be represented by its wavelength (λ) or frequency (*v*). Frequency is a measure of how many cycles a wave encounters per unit of time measured in cycles/second (sec^{-1}) or Hertz (Hz).

The different properties of light discussed are all linked to each other with the following equations:

$$c = \lambda v \qquad \text{Eq. 2.2}$$

$$v = \frac{c}{\lambda} \qquad \text{Eq. 2. 3}$$

$$\lambda = \frac{c}{v} \qquad \text{Eq. 2.4}$$

Experiments have shown that EMR behaves like waves; Young's single and double slit experiments demonstrated the wave-like nature of light. Max Planck theorised that the energy of an oscillator is discontinuous and that changes in energy must occur by means of jumping between two distinct energy states. Planck proposed that these oscillators have a specific energy (*E*) that is related to the frequency (*v*) of the oscillator, hence:

$$E = hv \qquad \text{Eq. 2.5}$$

where h = Planck's constant (6.63 x 10^{-34} Joules/second).

In brief, Einstein applied Planck's equation (*E* = h*v*) to light itself and suggested that light of a particular frequency has a given amount of energy. Einstein

concluded that light acts like as particle of energy (or “packet” of energy) and so gave rise to the term “photon” to describe a particle of light (2).

Equation 2.5 proposes that each frequency of EMR has a specific energy that is proportional to frequency and wavelength. As previously mentioned, a wave can be represented by its wavelength (λ) or frequency (*v*), but also its energy. The EM spectrum is composed of a number of regions based on increasing frequency and energy, as well as a decrease in the wavelength of the photon.

2.2 Interaction of Infrared Light and Matter

The interaction of IR radiation with a sample (or matter) is a fundamental concept within IR spectroscopy. Molecular stretches, bends and rotations that occur in a molecule following the absorption of IR radiation arise at specific frequencies characteristic of the frequency of the vibration. Following the absorption of a photon, whose energy matches the energy difference between two energy levels, the molecule can transition to a higher energy level (3). The energy of these molecular vibrations is quantised allowing for the measurement of the energy level differences following the absorption of a finite amount of energy (3).

The two atoms of a stable covalent molecule settle at a mean internuclear distance, thus the total energy of a molecule is at a minimum (4). The squeezing together and pulling apart motions of these molecules causes the repulsive force to rise rapidly and attractive force to resist separation respectively. Equation 2.6 explains the vibrational energies (E_v) of a molecule from a mean internuclear

distance (R_{eq}) following simple harmonic motions (v is the vibrational quantum number) (3):

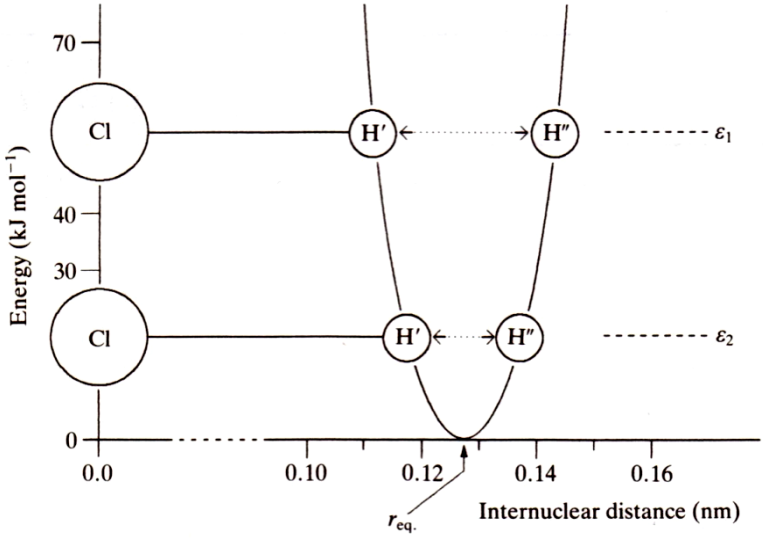
$$E_v = \left(v + \frac{1}{2}\right)\bar{v}\mathrm{cm}^{-1} \qquad \text{Eq. 2.6}$$

Consideration of this allows us to understand that a molecule can never be at a zero vibrational energy state due to the atoms never being completely at rest relative to one another. The lowest vibrational energy of a molecule at a zero vibrational energy state (ground-state)(E_0), with a vibrational quantum number of zero, still gives rise to molecular vibrations according to equation 2.7 (3).

$$E_0 = \frac{1}{2}\bar{v}\ \mathrm{cm}^{-1} \qquad \text{Eq. 2.7}$$

The distortion of a bond length requires an energy input which may be plotted against internuclear distance (Figure 2.2). Figure 2.2 shows a hydrochloric acid (HCl) molecule, whereby the chlorine atom has been permanently fixed against the *y*-axis and the hydrogen loosely placed on the *x*-axis. A larger push or pull motion of the hydrogen atom in relation to the chlorine results in an increasing energy. Increasing the energy of the HCl molecule from v_0 to v_1 allows for the displacement of the hydrogen atom, thus increasing R_{eq}. During simple harmonic oscillation, the displacement of the hydrogen molecule will remain, given that the vibrating frequency of the molecule remains constant.

Figure2.2 – Energy of the HCl molecule as the hydrogen is in a push and pull motion in relation to the chlorine atom. Replicated from (3).



The pushing and pulling motion of the bond during simple harmonic oscillation behaves accordingly to Hooke's law (Equation 2.8). According to Hooke's law, the frequency of a bonds vibration is related to the mass and force constant of the bond (or spring in Hooke's well-known "ball on a string" example). The application of Hooke's law allows us to understand the frequency of IR radiation required to successfully quantise a diatomic molecule, such as HCl, to allow it to move to a higher vibrational energy level (v_0 to v_1) (3). *k* is the force constant, *m* is the mass and *v* is the frequency of the vibration.

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \qquad \text{Eq. 2.8}$$

Hooke's law shows us that each molecule has a frequency which is characteristic of its atomic structure. The change to the masses of the atoms and

bond forces has the potential to alter a molecules wavenumber position on an IR spectrum.

During the absorption of IR radiation, a molecule is promoted to a higher vibrational energy level (Figure 2.3). In order for the molecule to absorb specific wavelengths of light, there are rules which describe whether certain transitions are allowed or not. Firstly, the primary selection rule states that the absorption of IR radiation must allow for an electric dipole moment to occur following the displacement of atoms and change in vibrational motion. The second selection rule involves quantum energy states (v); the transitions between energy states must adhere to equation 2.9. Transitions between vibrational energy states must not be greater than, or less than 1.

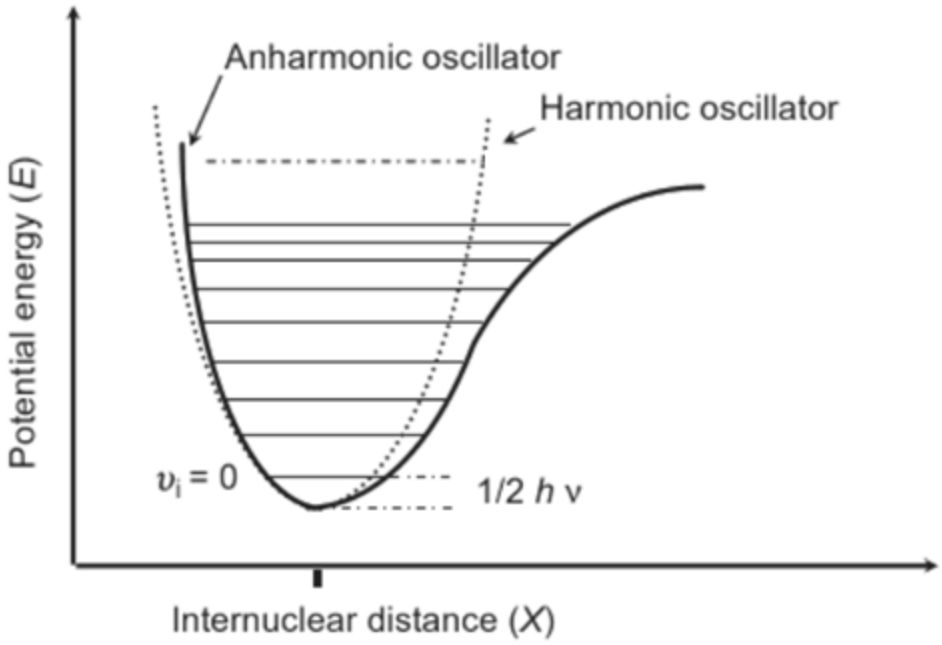
$$\Delta v \ = \pm 1 \qquad \text{Eq. 2.9}$$

Transitions between vibrational energy levels are allowed according to equation 2.9, so long as the transitions are **±**1. The Maxwell-Boltzmann distribution law states that the majority of molecules are found at the ground state (v_0) and that the dominant transition would be from v_0 to v_1 at room temperature(6). Real molecules do not obey the laws of simple harmonic motion, thus Hooke's law is not obeyed with real, non-theoretical bonds. The displacement of atoms in a molecule gets larger as the vibrational energy from IR radiation increases. If the bond between the atoms is stretched, there comes a point whereby it will break resulting in the molecule dissociating into its atoms. Hooke's law becomes redundant in place of real molecules, thus the anharmonic oscillator supersedes Hooke's concept and it

E

Figure2.3 - A potential energy diagram comparing the harmonic and anharmonic oscillators (7).

is more appropriate to explain the motion of real atoms. The anharmonic oscillator (Figure 2.3) shows a decrease in energy between the vibrational energy levels, thus decreasing the oscillation frequency as quantum state numbers increase. Figure 2.3 shows the Morse curve - the anharmonic potential-well. The energy required for transitions, *hv*, becomes smaller allowing for overtones to be lower in energy in comparison to overtones based on the harmonic oscillator theory. The intensity of overtone bands may sometimes be increased (accidental overtones) considerably due to 2 vibrational modes having frequencies very close to each other (3, 7).



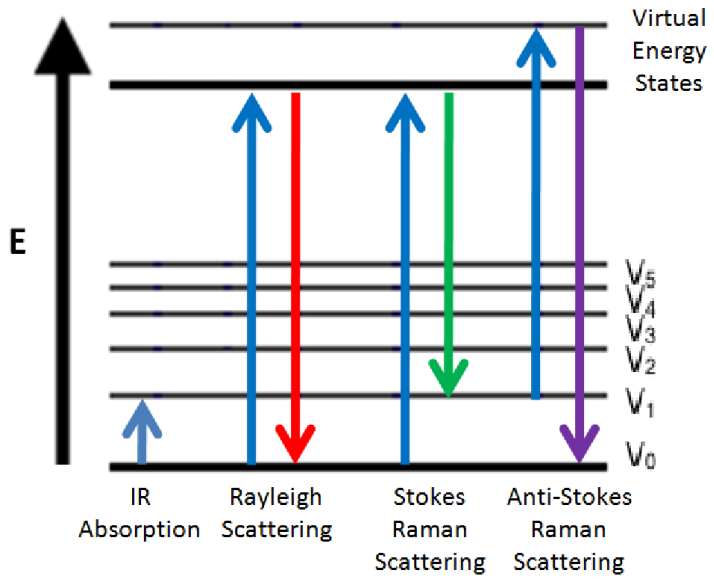
The selection rules for the anharmonic oscillator are found to be related to equation 2.9, but with the added possibility of promotion to higher vibrational energy levels. Equation 2.10 shows the anharmonic oscillator energy transitions.

$$\Delta v \ = \pm 1, \pm 2, \pm 3, \pm 4, \pm 5, \pm 6, \ldots \qquad \text{Eq. } 2.10$$

Normally the potential energy well levels$\Delta v = \pm 1, \pm 2 \text{ and } \pm 3$have the most observable intensities (3). The transition from v_0- v_1 is called the *fundamental* absorption, while those from v_0- v_2 and v_0- v_3 are called the *first* and *second overtones* respectively (3). For example, the HCl molecule has an intense fundamental absorption band at 2886 cm^{-1}, a significantly weaker band at around 5660 cm^{-1} and a very weak band at around 8350 cm^{-1}.

Figure 2.4 considers a molecule which transitions E_0 to E_1 (E = energy). E_0 to E_1 is the phase whereby IR radiation is absorbed by the sample and E_1 to E_0 is when the molecule emits the absorbed radiation. The Jablonski diagram (Figure 2.4) relates to Raman elastic (Rayleigh) and inelastic scattering; IR absorbance occurs from E_0 to E_1 (vibrational energy states) (4). Elastic (Rayleigh) and Stokes scattering are Raman scattering processes - elastic (Rayleigh) scattering occurs when the molecule returns to the same state following scattering of the excitation energy; Stokes scattering occurs when the photon has lost energy to the molecule and Anti-stokes when the photon has gained energy from the molecule (5).

Figure2.4 – The absorption and emission of infrared radiation and Raman scattering. Adapted from (4).



2.3 Fundamental Vibrations

All molecules have 3N degrees of freedom, with N representing the total number of atoms in a molecule. Each atom is placed according to 3 Cartesian co-ordinates (*x, y* and *z*). A molecule, such as water (H_20), has 9 degrees of freedom. Six of these 9 degrees of freedom include 3 rotational and 3 translational degrees; the remaining 3 belong to the fundamental vibrational modes of the molecule. For linear molecules 3N-5 (Eq. 2.11) and non-linear molecules 3N-6 (Eq. 2.12) are used to determine the vibrational modes of a molecule.

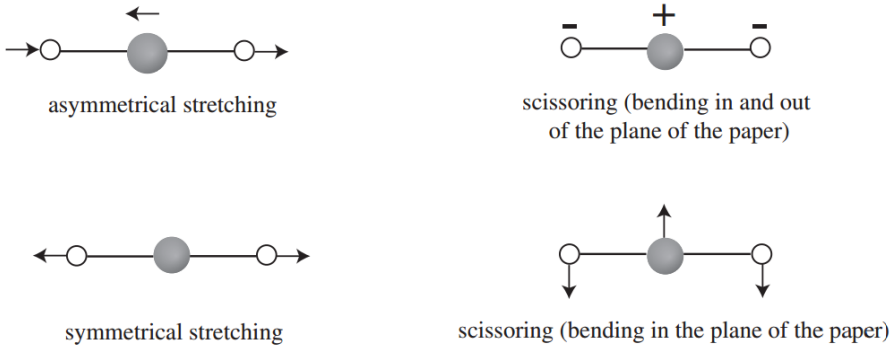
$$\text{Linear Molecule} = 3N - 5 \qquad \text{Eq. 2.11}$$

$$\text{Non} - \text{Linear Molecule} = 3N - 6 \qquad \text{Eq. 2.12}$$

There are 2 fundamental types of molecular vibration: stretching and bending. Stretching vibrations are either symmetric or asymmetric, while bending

Figure2.5 – The four vibrational modes of CO_2. Replicated from (7).

vibrations can be split up into 4 further modes (Figure 2.5). Bending vibrational modes include twisting, rocking, wagging and scissoring. Carbon dioxide (CO_2) is a linear molecule (3N-5) and has 4 fundamental vibrations. Figure 2.5 shows the 4 vibrational modes of CO_2 - 1 asymmetrical stretching, 1 symmetrical stretching and 2 modes of scissoring) (7).



The symmetric stretch vibrational mode of CO_2 is IR inactive due to the vibration causing no change in dipole moment. During symmetric stretching of the CO_2 molecule, the 2 oxygen atoms are pulled away from and pushed towards (compressed) the carbon atom concurrently resulting in no dipole moment. CO_2 has 4 fundamental vibrations but only 2 (asymmetrical stretch and a degenerate vibration from 2 bending modes) are visible on the IR spectrum. A degenerate band occurs when a molecule has equivalent modes of the same frequency which are present at the same $\bar{\nu}$ region on the IR spectrum (3, 7).

2.3.1 Complementarity of Infrared and Raman Spectroscopy

In molecules with a centre of symmetry, it is observed that vibrations are either IR or Raman active or inactive, or vice-versa - this is commonly referred to as the principle of mutual exclusion. The rule of mutual exclusion states that if a molecule has a centre of symmetry, then no transition is allowed in both IR absorption of Raman scattering, but only in one or the other. A molecule which has as centre of symmetry is carbon dioxide (CO_2). CO_2 is IR active with asymmetrical stretching and bending vibrations but it is IR inactive with symmetric stretching due to there being no change in dipole moment. In comparison, symmetric stretching of CO_2 is Raman active whilst asymmetric and bending modes are Raman inactive due to there being no change in the molecules polarisability (1).

CO_2 has 2 peaks on the IR spectrum (1 asymmetric and 2 degenerative stretching peaks) and just 1 peak (symmetric) on a Raman spectrum. The CO_2 peaks present in the IR are at 2349 cm^{-1} (asymmetric stretching) and at 667 cm^{-1} (bending) and in Raman at at 1330 cm^{-1} (symmetric). Figure 2.6 shows IR and Raman spectra displaying 2 IR peaks and 1 Raman peak (8).

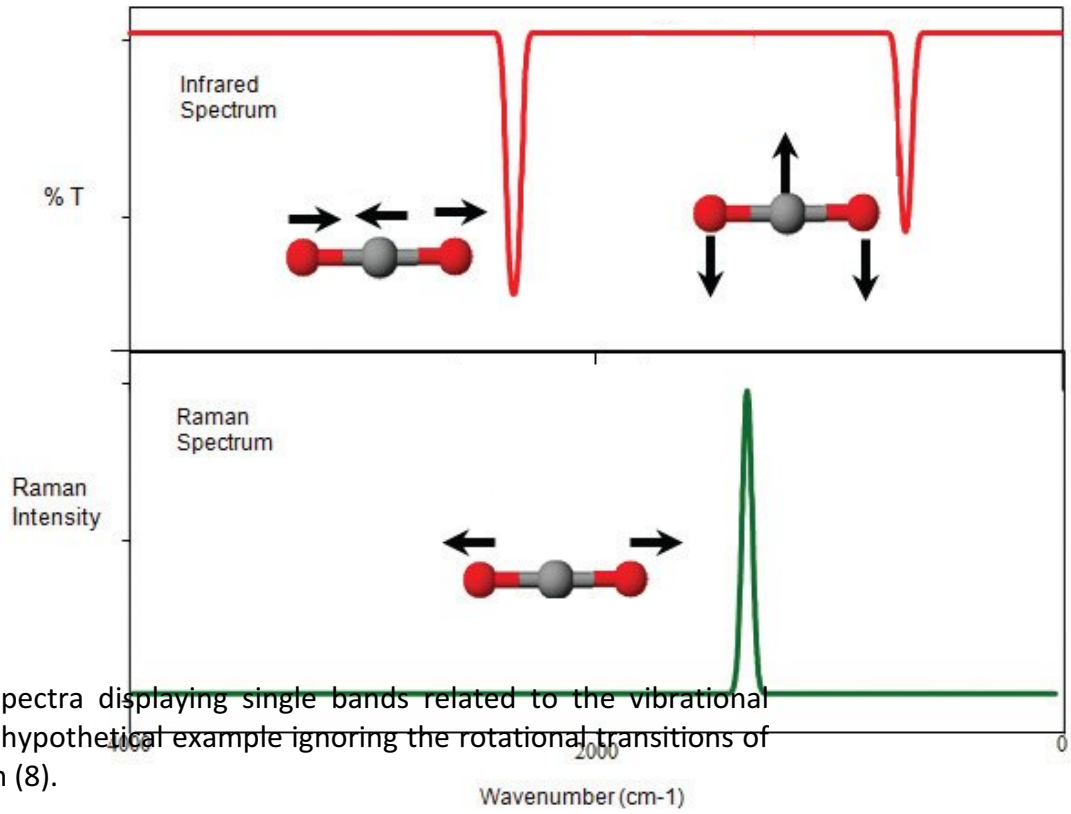


Figure2.6 – Two spectra displaying single bands related to the vibrational transitions of CO_2 (hypothetical example ignoring the rotational transitions of CO_2). Adapted from (8).

2.4 Beer-Lambert Law

Beer-Lambert law relates concentration to a samples property (thickness), absorbance and spectral property (absorptivity) for quantitative analysis using a transmission sampling method. Beer-Lambert law is the equation which relates the amount of IR radiation absorbed by a sample according to equation 2.13.

$$A = \epsilon \, l \, c \qquad \text{Eq. } 2.13$$

Figure2.7 – Transmission example of the Beer-Lambert law.
Adapted from (1).

Equation 2.13 shows the absorbance of IR radiation (A), the molar absorptivity of a sample (ϵ), the pathlength of a sample (l), and the concentration (c). $A = \epsilon l c$ relates to the concentration of molecules in a sample, whereas the absorbance of light within a sample can be calculated using equation 2.14. Equation 2.14 can be used to calculate the absorbance of a sample according to the Beer-Lambert law.

$$A = \log \frac{I_0}{I} \qquad \text{Eq. 2.14}$$

Equation 2.14 and Figure 2.6 show the absorbance of IR radiation (A), I_0 as the intensity in the background spectrum and I as the intensity in the sample spectrum.

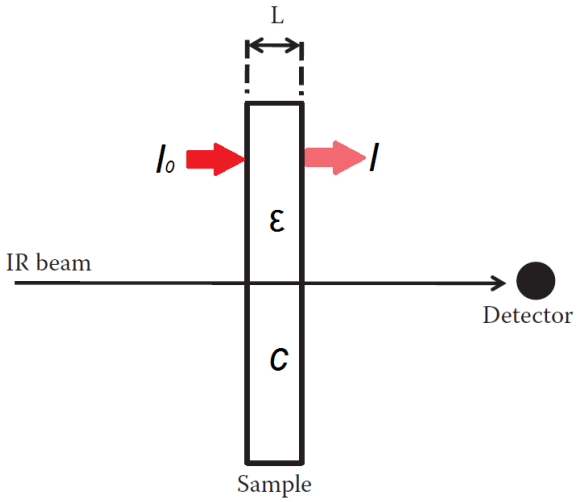


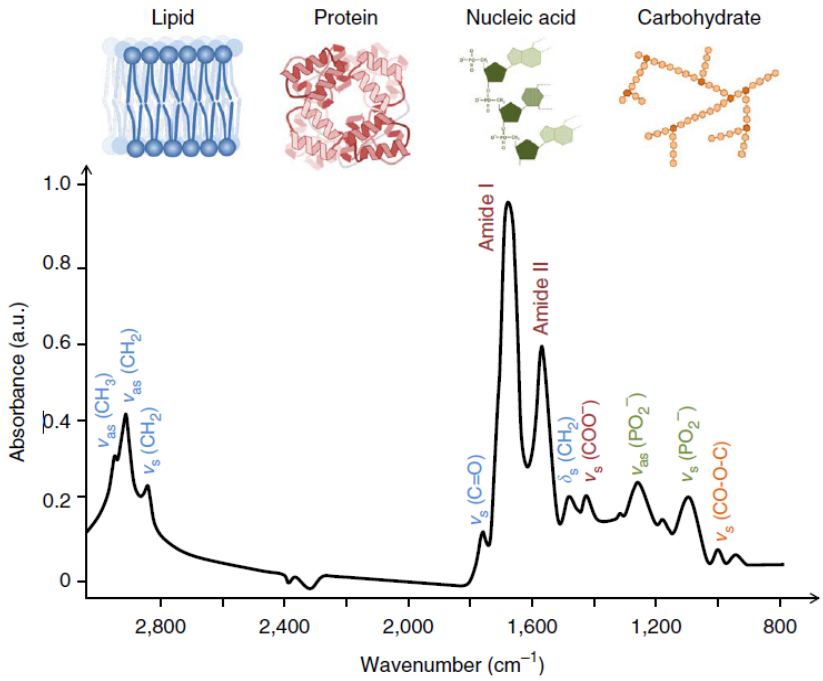
Figure 2.7 shows a beam of IR radiation being aimed at a sample with a length of l. Some of the IR beam is absorbed by the sample. Light which does pass

through the sample is focused upon a detector to obtain a sample spectrum. The Beer-Lambert law can be used to perform a calibration using the straight line equation (y= m*x* + c) when known standard concentrations are created from a known sample source, thus unknown sample can be quantitatively analysed to determine their concentration (1). Limitations causing non-linearity of the Beer-Lambert law include too high concentrations, scattering of light in the sample, fluorescence of the sample, stray light and non-monochromatic radiation (9).

2.5 Absorption Bands on a Biospectrum

The absorption of IR radiation is caused by the interaction between the electric vector (Figure 2.1) and the dipole moment of a molecule produced during molecular vibration. In addition, equation 2.8 shows that the oscillating frequency of a bond during harmonic oscillation is related to the mass and force constant of the bond, thus absorption bands on an IR spectrum are assigned to particular molecular vibrations (3). The most discriminatory spectral region to measure for the interrogation of biological samples, such as human serum, is the fingerprint region (1800-1000 cm^{-1}) and the amide I and amide II region (1700-1500 cm^{-1}), due to spectral contributions from carbohydrates, nucleic acids and proteins. The higher wavenumber region of a spectrum (3500-2500 cm^{-1}) is the region for O-H, N-H and C-H stretching vibrations; the lower wavenumber region absorption bands typically correspond to carbon skeletal molecular vibrations (10). Figure 2.8 shows a typical IR spectrum of a biological sample.

Figure2.8 – A typical IR spectrum collected from a biological sample.
Replicated from (10).

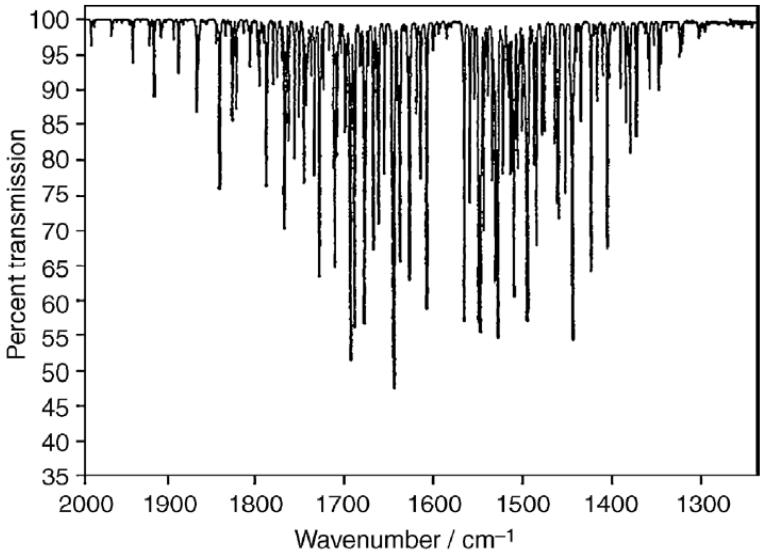


The hydrogen bonding of a molecule plays a significant role when determining its position on an IR spectrum. A molecules bond frequency increases depending on whether it has single, double or triple bonds; single bonds are the weakest with lower frequencies compared to triple bonds which have high frequencies. Amide I and amide II primary (alpha) and secondary (beta) structures can be distinguished between each other due to their hydrogen bonding. Primary alpha amide structures are placed at a higher wavenumber region on the IR spectrum due to them having a higher level of hydrogen bonding (higher frequency) in comparison to beta structures (11).

gure2.9 – Vibrational rotation spectrum of H_2O which can potentially be an terference in a spectrum. It is good practice to eliminate all traces of these olecules in the beam path of an IR spectrometer by purging the instrument (3).

2.6 Gaseous Contaminants on the IR Spectrum - Carbon Dioxide (CO_2) and Water (H_2O) Vapour

Carbon dioxide and water vapour are present in the environment in gaseous states and they absorb IR radiation allowing both rotational and vibrational rotational transitions to occur. Small molecules in a gaseous phase show considerable fine structure due to the transitions between quantised rotational energy levels. These fine structures are rarely seen in spectra of larger molecules of a gaseous phase or in molecules in a liquid state. Figure 2.9 shows a rotational-vibrational spectrum of H_2O vapour (3).



The interference of these spectral contaminants may interfere with the acquired data during statistical analysis/data processing. The collection of a background spectrum, to allow for atmospheric correction of CO_2 and water vapour, is routinely performed to account for these gaseous spectral contaminants. Subtracting the background spectrum from the sample spectrum allows for the removal of the gaseous state molecules from a spectrum.

Purging an FTIR system is an excellent way of dealing with the problems associated with gaseous CO_2 and water vapour present in the atmosphere around the instrument. A purge provides an FTIR instrument with a constant flow of CO_2 free, water vapour removed (dry via desiccation), IR inactive nitrogen gas. The purpose of a purge is to maintain a constant gaseous environment inside the FTIR instrument to eliminate inconsistencies in background levels of CO_2 and water vapour (12).

2.7 Spectral Pre-processing and Multivariate Analysis (MVA)

Prior to exploratory studies with supervised MVA, it is necessary to prepare the data. Multivariate analysis refers to the analysis of multiple variables (dependent variables to test and independent variable to determine whether they are the cause), whereby each variable is linked to one another in some form (13). It is common practice for raw spectral data to be pre-processed which may involve the creation of new variables of the raw data. This section focuses on the pre-processing and multivariate analysis methods performed during this research project.

2.7.1 Noise Reduction

Noise reduction (NR) is used following spectral collection to improve the signal-to-noise ratio (SNR) of the acquired spectra. NR was performed using an in-house written Matlab function which is based on transforming the spectral data into principal components (PCs), explained further in 3.4. The NR function allows the interpreting user to select the number of PCs they wish to use; few PCs selected leads to the greatest NR but also increases the risk of removing important spectral features (13). The principal component analysis (PCA) portion of the NR function produces scores and loadings vectors (discussed shortly in 3.4), thus allowing for the number of PCs chosen by the interpreter to be restored to the original vector size of the initial spectral dataset. A chosen number of PCs should explain the relevant chemical information of the entire dataset. Generally, the first 20-40 PCs of the data hold the majority of the spectral information (13); however, within this research the number of PCs used was decided upon by 95.0 % of the variance explained. Noise reduction can account for any instrumental factors which may have had an influence on the IR spectra collected (14).

2.7.2 Standard Normal Variate (Vector Normalisation)

The use of standard normal variate is a common pre-processing step to normalise the acquired spectral data. Vector normalisation (VN) was used during this project; VN reduces the gross variance (thickness/width) between the maximum and minimum spectra on the absorbance axis of all of the sample spectra.

The in-house written Matlab VN function used initially averages each wavenumber variable. The value of each average is then subtracted from the spectrum from which it was calculated to give a value equalling zero. Following this, each wavenumber variable is squared. The total sum of the squared wavenumber variables is calculated to give *x*. Each wavenumber variable is then divided by the square root of *x*, thus normalising the spectral dataset to a magnitude of one. Other means of normalisation do exist. Row scaling normalisation includes: ratioing to landmarks (ratioing the remainder of the spectra based upon a single intensity on a spectrum); scaling rows to a constant total (VN); block scaling; column scaling, scale based on a minimum and maximum intensity and scale 0 to 1 etc. Vector normalisation was selected to be used so that all of the original spectral data could be normalised together for every wavenumber variable, rather than landmark normalising for example, resulting in the output matrix always totalling a value of one.

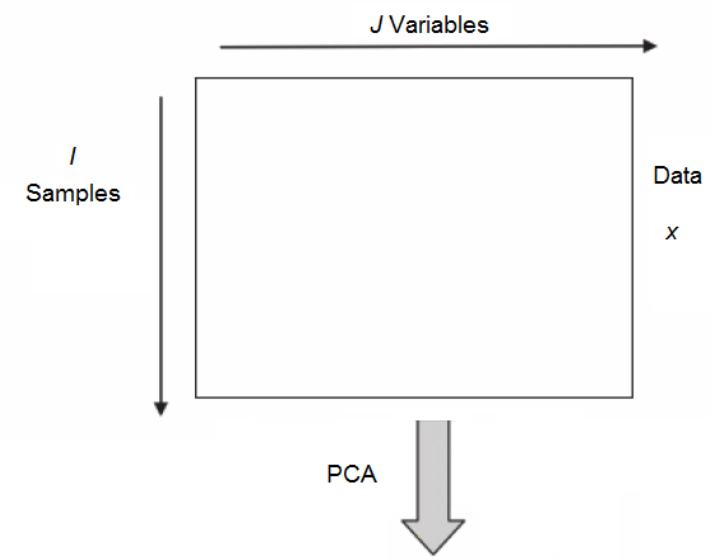
2.7.3 Principal Component Analysis (PCA)

PCA involves the transformation of the original, raw spectral data matrix according to equation 2.15 (13) and its purpose is to reduce the dimensionality of a dataset:

$$x = \mathrm{TP} + \mathrm{E} \qquad \text{Eq. 2.15}$$

where *x* is the original data matrix with dimensions *I x J*; *T* is the scores matrix and has as many rows as the original data matrix (*I x A)*; *P* is the loadings and has as

many columns as the original data matrix (*A x J*). *A* refers to the number of principal components. Figure 2.10 illustrates the PCA process. PCA is a powerful unsupervised method used to explore patterns in data and it is used to create scores and loadings matrices, *T* and *P* respectively.



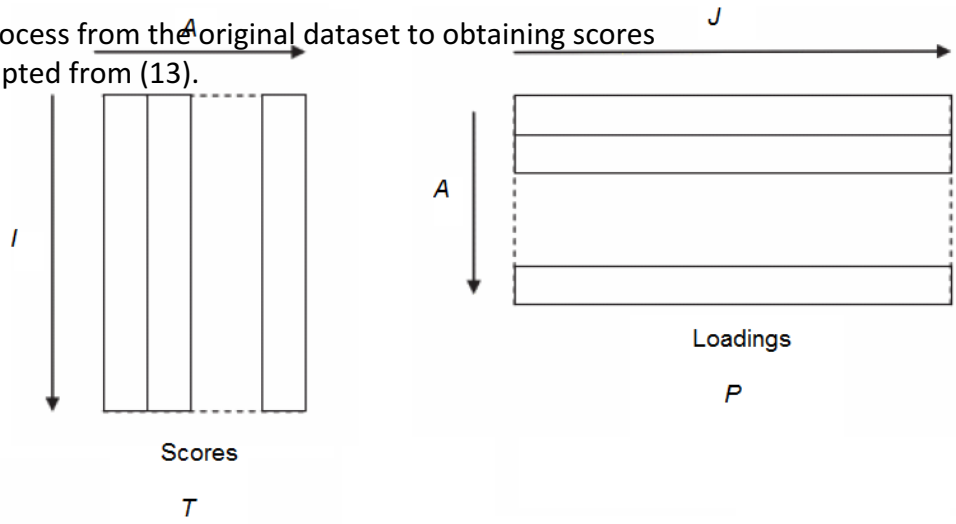


Figure2.10 – The PCA process from the original dataset to obtaining scores (*T*) and loadings (*P*). Adapted from (13).

Scores is a numerical value given to each sample spectrum for each principal component (PC), and loadings numerically describe the weight of the PC of each

input sample spectrum. A PC describes the pattern of co-variation between classes of data. A given number of PCs for analysis describes a datasets co-variation, i.e. the first PC describes the most important pattern of co-variation, whereas the second, third etc. represents the remaining patterns of co-variation (15). PCA allows for large datasets to be reduced into a number of PCs which are characteristic of the original dataset. The fundamental aim of PCA is to discover components of the data which best explain the major variations of the data matrix, for example Khanmohammadi *et al*. (2007) discusses the use of the first two PCs of their studies spectral dataset to describe 80.8 % of their datasets variance (16).

In addition to discovering PCs to explain the data, PCA involves finding mathematical functions that are related to the chemical properties of a spectrum, thus the PCs which are found to explain the datasets variance can be considered as mathematical structures (17). Following PCA, the original $\bar{v}$ variables are reduced significantly, allowing for a large dataset to be much smaller and manageable. The original dataset is mathematically modelled by the PCs, which are vectors in a multidimensional space. Due to the PC vectors being in a multidimensional space, the covariance is measured to determine the relationship between the IR spectra. An important step involved with PCA is the determination of how significant each PC is in relation to the data matrix. Eigenvalues numerically represent the contribution of each PC to the entire dataset and they describe the size of each PC, whereby the larger the size of the component the more significant it is (17). The loadings from PCA can be plotted to show the $\bar{v}$ regions which contribute to the separation along the PC axis. A loadings plot can be used to observe which spectral features are

responsible for separation of classes on a PCA plot (13). Previous research in the field of biospectroscopy has used PCA as a dimensionality reduction method prior to analysis (18-19).

2.7.4 Support Vector Machine (SVM)

A SVM is a supervised method which can produce complex curved boundaries with indefinite complexities - the more complex the boundary, the more likely it is to perfectly separate test and train datasets (13). A highly complex boundary between *x* number of classes allows for separation; however, overfitting is possible if careful consideration is not taken when selecting the parameters for the complexity of a boundary; boundaries can be set around datasets where there is no real significance, hence this occurring when distinguishing between disease state datasets could result in misleading outcomes. In this project, a radial basis function (RBF) SVM was used to separate disease classes. A RBF kernel is well-suited to classes of data where a linear hyperplane (e.g. used in linear learning machines for simple linear problems) is inadequate to separate the data. A kernel is essentially a similarity function that allows us to observe how similar classes are to one another; previous research has found that a RBF kernel coupled with SVM analysis provides excellent classification rate accuracies of 80.0 % (±18.92 %)when using complex datasets (20). In addition, previous research in the field of biospectroscopy has achieved high sensitivities and specificities when using RBF-SVMs to distinguish between cell lines (21) and for lymph node diagnostics (22).

An SVM projects the classes of data into a new higher dimensional (feature) space where a RBF is employed to find complex separating hyperplanes around the classes. The non-linear nature of an RBF is well-suited for complex datasets (13). Figure 2.11 shows two inseparable classes (left) which are projected into a hyper-dimensional feature space (middle) where a plane successfully separates the classes. Finally, the successful discrimination of the classes using a RBF is projected back into a 2-dimentional plot with the plane intact (right) (13).

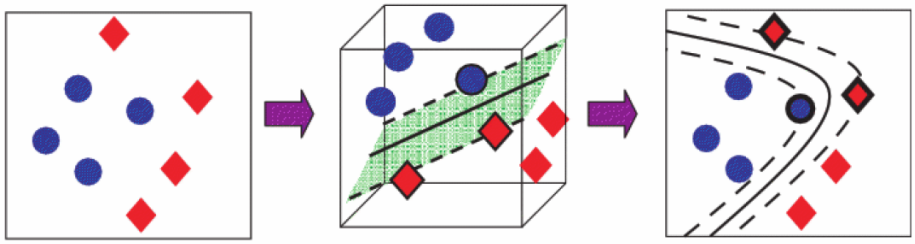
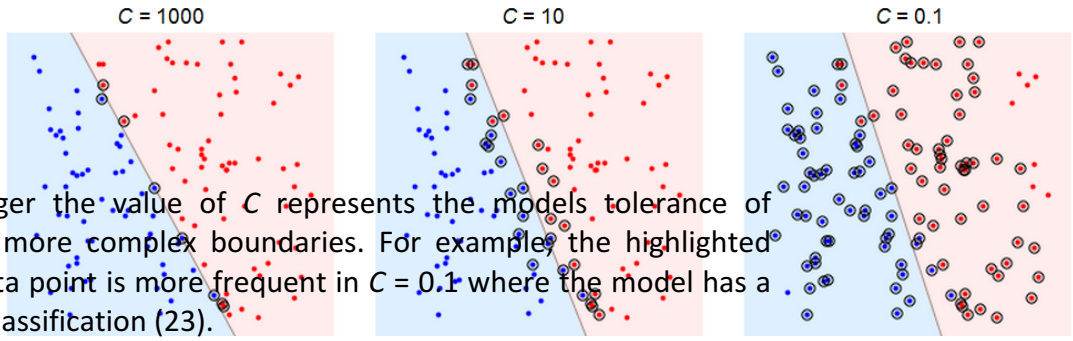


Figure2.11 – Two linearly inseparable classes (left); projection of the 2 inseparable classes into a higher dimensional feature space where a plane successfully separates the classes (middle); projecting the classes from the feature space back to 2-dimensions (right). Reproduced from (13).

The parameter, cost (*C)*, is used to determine the level of tolerance which a model has to misclassification (13). A larger *C* value represents a model having a lower tolerance of misclassification and more complex boundaries, thus a trade off exists between the complexity of the boundary and the samples misclassified or those which are near to the boundary, as illustrated in Figure 2.12 (23). High values of *C* tend to represent hard margin SVMs resulting in high misclassification error rates (13).

gure2.13 – The γ parameter influences the curvature of the SVM boundary etween the classes (24).



re2.12 – The larger the value of *C* represents the models tolerance of classification with more complex boundaries. For example, the highlighted ce around each data point is more frequent in *C* = 0.1 where the model has a tolerance to misclassification (23).

Another parameter used in RBF-SVM is the gamma (γ) value. The gamma value is related to the curvature of the SVMs plane between classes. If the value of γ is increased or decreased then the curvature of the decision boundary is altered. The result of different γ values is illustrated in Figure 2.13 (24). Figure 2.14 shows the influence of a high γ value on the curvature of a separating boundary (24).

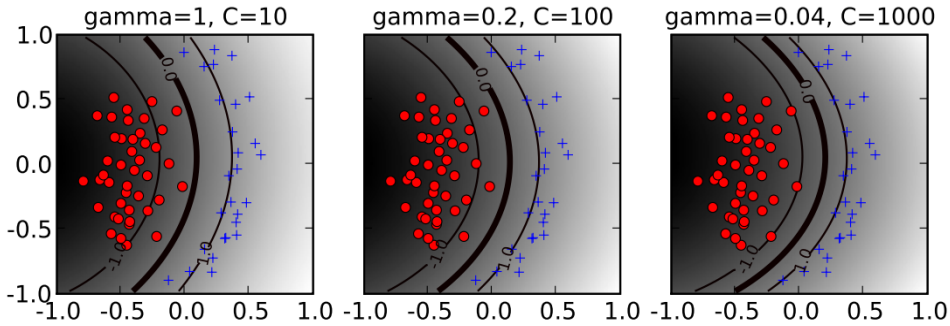
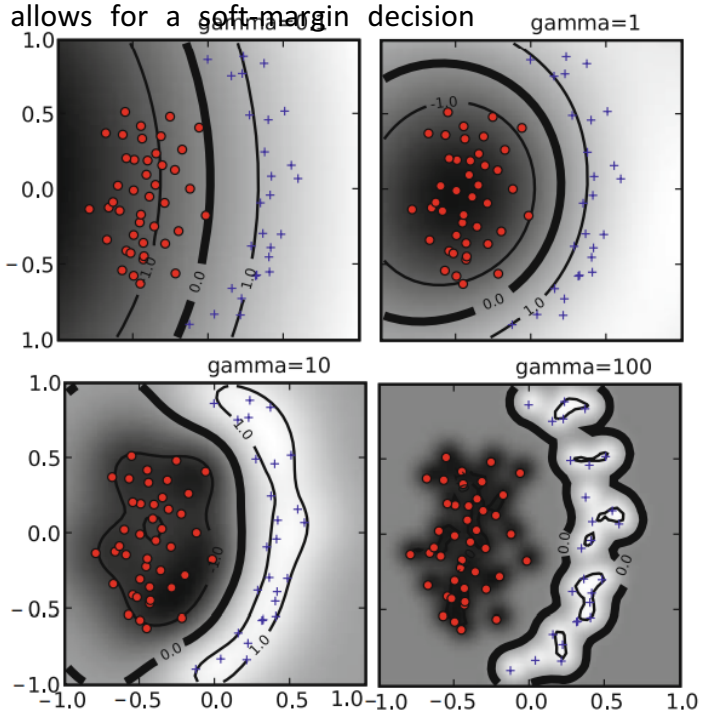


Figure2.14 – A high γ value leads to overfitting due to the increased flexibility of the separating plane (bottom left and right). A small γ value allows for a soft-margin decision boundary (24).



The consequence of decreasing the value of γ is that the curvature of the separating SVM boundary is reduced; however, if the value of *C* is increased then the separating boundary is forced to curve accordingly to accommodate the large error rate (24).

The most common SVM kernels available, apart from RBF, are polynomial and sigmoidal (24). The lowest degree of flexibility a polynomial classifier has is a linear kernel, thus the use of a linear kernel is unsuitable with a dataset with a non-linear relationship between classes (24). These kernels can be used for data which is better suited to their purpose, such as with data which is not highly complex or where a linear plane would be best to separate classes of data. The use of hard SVM

margins, as opposed to soft, to define complex class boundaries may result in overfitting. SVMs can be divided into two general categories; hard and soft margin SVMs. Hard margin SVMs require two classes of data to be completely separable with the margin's aim being to find the maximum possible distance between the two classes. To avoid overfitting, it is common practice to use a soft margin SVM which allows for a degree of misclassification whilst balancing the complexity of the model against the degree of misclassification.

2.8 SpecToolbox for Feature Extraction

The SpecToolbox, written in collaboration with Dr Ryan Stables (Birmingham City University), is an in-house written piece of software developed to work in Matlab for the analysis of spectroscopic data. Feature extraction (FE) enables the user to extract the relevant discriminatory spectral information into smaller vectors of information, while removing redundant data. Although highly correlated, PCA vectors do not necessarily explain the key regions of the spectrum. Data reduction is performed prior to using the SpecToolbox by using PCA, however, determining the important spectral changes between groups of data can be difficult when evaluating loading plots, thus variable ranking is performed. FE may also improve classification accuracies. The main function of the SpecToolbox is to provide FE to elucidate and rank the relevant spectral information from recorded data. Figure 2.15 shows the graphical user interface (GUI) of the SpecToolbox.

spectra in the SpecToolbox.

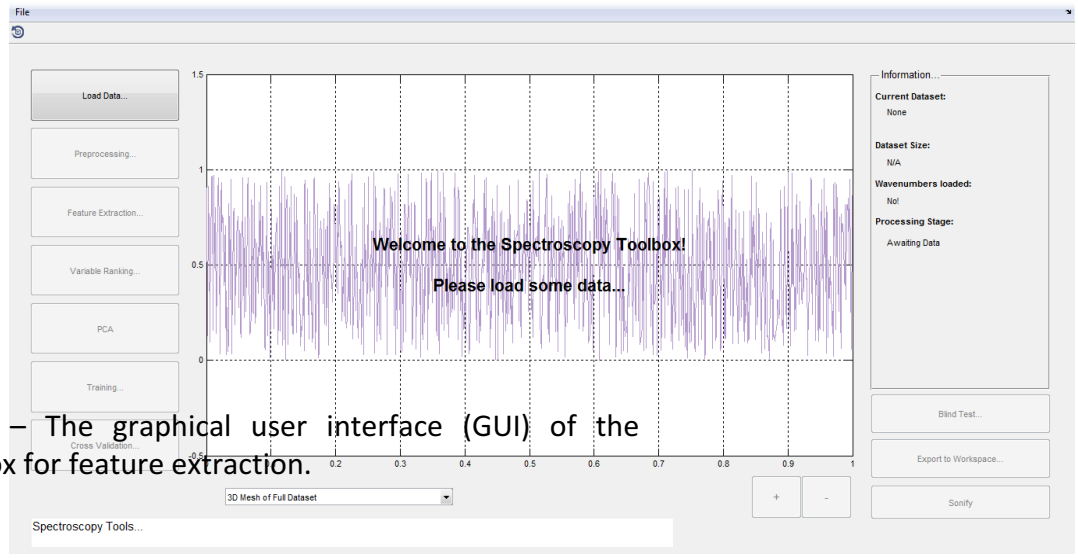
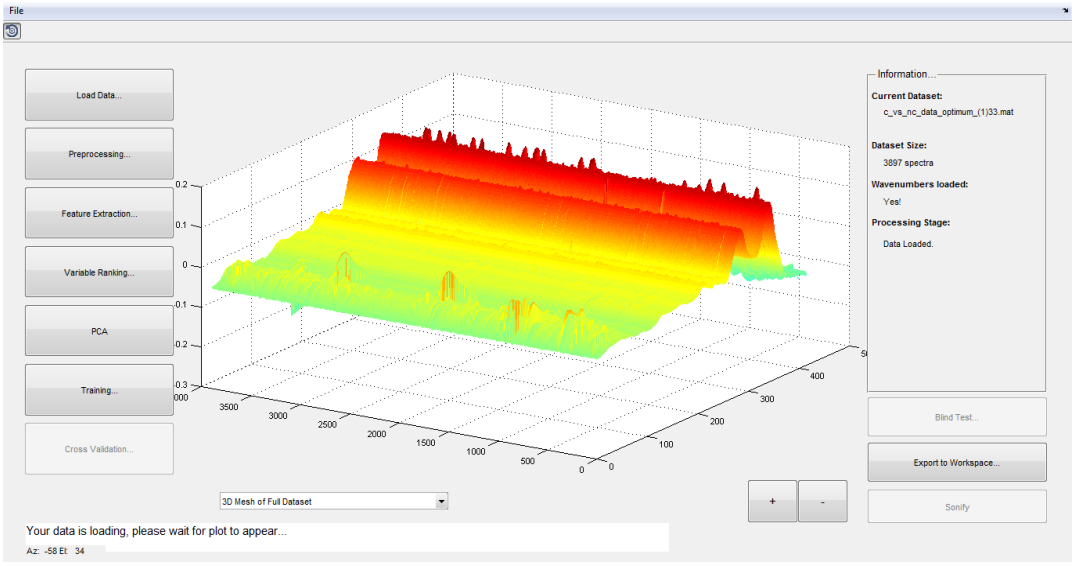


Figure2.15 – The graphical user interface (GUI) of the SpecToolbox for feature extraction.

Following the input of spectral data in the SpecToolbox, the user can extract features based upon the $\bar{\nu}$ regions of interest, thus allowing for redundant information to be removed from analysis. Figure 2.16 shows IR data loaded into the SpecToolbox.



Variable ranking via information gain highlights the wavenumber regions to the user which are most discriminatory between the spectral classes by means of Pearson's coefficient correlation and entropy. Variable ranking lists the wavenumber features in descending order based upon their role in discriminating between various classes - the top variable ranking feature is the most discriminatory between the classes. Pearson's correlation coefficient is frequently used in clinical chemistry when comparing between the strength associations between spectral bands (5). The similarities between spectra can be determined with spectral numerical values. All points of the user selected/highlighted spectral bands are taken into consideration during the coefficients calculation of similarity, thus the approach is more efficient than visual comparison based on band intensity. A Pearson coefficient value is a numerical indication of the strength and direction of the linear relationship between variables. Coefficient values are varied between -1 and 1, with -1 representing a perfect negative relationship, +1 a perfect positive relationship and 0 representing no relationship (25). Furthermore, entropy measures the level of impurity in a group. The higher the entropy of a wavenumber region, the more information rich the region is. Performing information gain on a spectral dataset allows for the understanding of how important a given wavenumber region (feature) is (26). Following this step, the regions of the spectrum upon which the user wishes to perform FE can be selected. Figure 2.17 shows the selection of spectral bands on a 2-dimensional plot of the mean spectra.

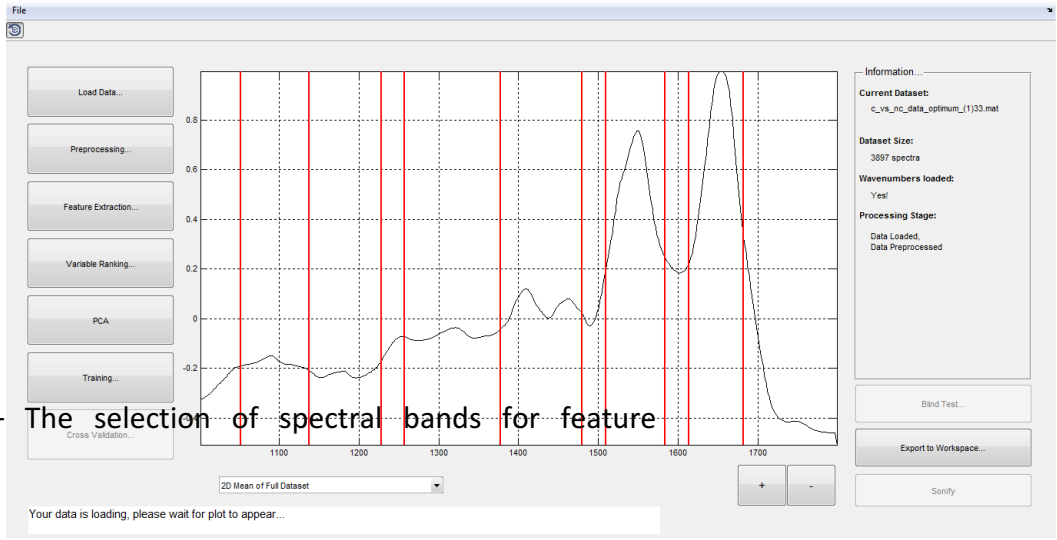


Figure2.17 – The selection of spectral bands for feature extraction.

Following the selection of the discriminatory bands, FE is performed. During FE spectral descriptors such as peak kurtosis, peak skew, peak centroid, peak frequency, peak amplitude and RMS energy from each selected band are captured. The feature information is ranked and scored in descending order to best describe how each feature explains the variance between the classes. The feature data matrix can be exported to the Matlab workspace, saved and used for FE-fed SVM analysis. The peak shapes/descriptors and FE-fed SVM will be discussed in a later chapter of this thesis. Alternative methods of FE include random forests which have been found to increase classification accuracies and reduce redundant data from analysis (27). The SpecToolbox was designed with a user-friendly GUI for simplicity; its use as a tool for FE in disease detection is presented within this thesis.

References

[1] Smith, B C. *Fundamentals of Fourier-transform infrared spectroscopy.* 2nd Edition. CRC Press, 2007

[2] Ball, D W. *The basics of spectroscopy*. SPIE Press, 2001

[3] Banwell, C., McCash E. *Fundamentals of molecular spectroscopy*. 4th Edition. McGraw-Hill Higher Education, 1994

[4] Stuart, B H. *Infrared spectroscopy: fundamentals and applications*. Wiley, 2004

[5] Griffiths, P R., Shao, L. *Self-weighted correlation coefficients and their application to measure spectral similarity*. Applied Spectroscopy, 2009. 63(8), 916-919

[6] Sathyanarayana, D N. *Vibrational spectroscopy: theory and application*. New Age International Pvt Ltd Publishers, 2007

[7] Larkin, P. J. *IR and Raman spectroscopy - principles and spectral interpretation*. Elsevier, 2011

[8] Donahue M., Botonjic-Sehic E., Wells D., Brown C W. *Understanding infrared and Raman spectra of pharmaceutical polymorphs* [Online] http://www.americanpharmaceuticalreview.com/Featured-Articles/37183-Understanding-Infrared-and-Raman-Spectra-of-Pharmaceutical-Polymorphs/ [Accessed on 30th August 2015]

[9] Swinehart, D. F. *The Beer-Lambert Law*. Journal of Chemical Education, 1962, 39(7)

[10] Baker, M J., Trevisan, J., Bassan P., Bhargava, R., Butler, H J., Fielden, P R., Fogarty, S W., Fullwood, N J., Heys, K A., Hughes, C., Lasch, P., Martin-Hirsch, P L., Obinaju, B., Sockalingum, G D., Sule-Suso, J., Strong, R J., Walsh, M J., Wood, B R., Gardner, P., Martin, F L. *Using Fourier-transform IR spectroscopy to analyse biological materials*. Nature Protocols, 2014. 9(8), 1771-1791

[11] Barth, A. Zscherp, C. *What vibrations tell us about proteins*. Quarterly Reviews of Biophysics, 2002. 35(4), 369-430

[12] Parker Balston. *Effect of purging a sealed and desiccated FTIR spectrometer sample compartment* [Online] http://www.parker.com/literature/Balston%20Filter/AGS/AGS%20Technical%20Articles/PDFs/Effect_of_Purging_FTIR_Spectrometer.pdf [Accessed on 10th July 2015]

[13] Brereton, R. *Chemometrics for pattern recognition*. Wiley-Blackwell, 2009

[14] Goodacre, R. VecNorm Function (2000) written by Goodacre, R. University of Manchester. Matlab Software

[15] Martens, H., Martens, M. *Multivariate Analysis of Quality - An Introduction*. John Wiley & Sons., 2001

[16] Khanmogammadi, M., Nasiri, R., Ghasemi, K., Samani, S., Bagheri Garmarudi, A. *Diagnosis of basal cell carcinoma by infrared spectroscopy of whole blood samples applying soft independent modeling class anaology*. Journal of Cancer Research in Clinical Oncology, 2007. 133(12), 1001-1010

[17] Brereton, R. *Chemometrics: data analysis for the laboratory and chemical plant*. Wiley, 2003

[18] Gajjar, K., Heppenstall, L. D., Pang, W., Ashton, K. M., Trevisan, J., Patel, I. I., Llabjani, B., Stringfellow, H. F., Martin-Hirsch, P. L., Dawson, T., Martin, F. L. *Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis*. Analytical Methods, 2012. 6(5), 89-102

[19] Lovergne L., Clemens G., Lukaszewski R., Untereiner V., Sockalingum G D., Baker M J. *Investigating optimum sample preparation for infrared spectroscopic serum diagnostics*. Analytical Methods, 2015. DOI: 10.1039/C5AY00502G

[20] Gajjar K., Trevisan J., Owens G., Keating P J., Wood N J., Stringfellow H F., Martin-Hirsch P L., Martin F L. *Fourier-transform infrared spectroscopy coupled with a classification machine of the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer*. Analyst, 2013. 138(14), 3917-3926

[21] Baker, M. J., Clarke, C., Demoulin, D., Nicholson, J. M., Lyng, F. M., Byrne, H. J., Hart, C. A., Brown, M. D., Clarke, N. W., Gardner, P. *An investigation of the RWPE prostate derived family of cell lines using FTIR spectroscopy*. Analyst, 2010. 135(5), 887-894

[22] Sattlecker, M., Bessant, C., Smith, J., Stone, N. *Investigation of support vector machines and Raman spectroscopy for lymph node diagnostics*. Analyst, 2010. 135(5), 895-901

[23] StackOverflow. *SVM - Hard or Soft Margins* [Online] http://stackoverflow.com/questions/4629505/svm-hard-or-soft-margins [Accessed on 6th July 2015]

[24] Ben-Hur, A., Weston, J. *A user's guide to support vector machines*. Humana Press, 2010

[25] University of Strathclyde Humanities and Social Sciences. *Pearson r-correlation coefficient* [Online] http://www.strath.ac.uk/aer/materials/4dataanalysisineducationalresearch/unit4/pearsonr-correlationcoefficient/ [Accessed on 23rd July 2015]

[26] Computer Science & Engineering at University of Washington. *Information Gain*, 2015 [Online] http://homes.cs.washington.edu/~shapiro/EE596/notes/InfoGain.pdf [Accessed on 10th August 2015]

[27] Joelsson, S. R. *Feature Selection for Morphological Feature Extraction using Random forests*. Signal Processing Symposium, 2006, NORSIG 2006, Proceedings of the 7th Nordic. 2006. 138-141

Chapter 3

INSTRUMENTATION

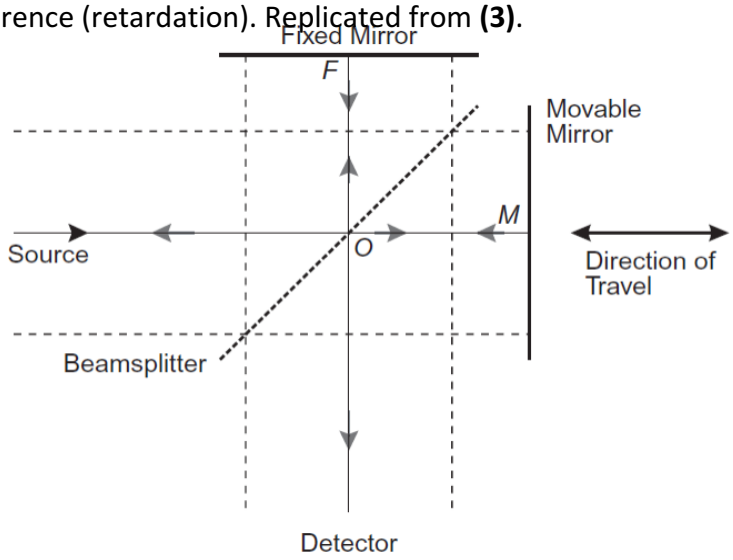
This chapter will discuss the Michelson interferometer and describe the factors which influence optimal spectral acquisition, commonly known as the FTIR trading rules. The ATR sampling method will be described and the FTIR instruments used to collect spectral data from patient serum samples will be presented. Additionally, this chapter will discuss background spectra and the JASCO and Agilent FTIR instruments used during this project will be presented.

3 Instrumental Theory

3.1 Michelson Interferometer

FTIR spectroscopic instrumentation is fundamentally based on the use of an IR globar source, a detector, a reference laser and a Michelson interferometer; however, recently broadly tunable mid-infrared quantum cascade lasers (QCLs) have been integrated into spectroscopic imaging systems allowing for simpler instrumentation as no interferometer is required (1). The Michelson interferometer is a beam-splitting device which divides the beam of IR radiation into two paths, followed by recombining them both after a path difference has been introduced (2). The range of IR intensities exiting the interferometer is measured by a detector based upon path difference (3). Figure 3.1 shows a schematic diagram of a Michelson interferometer and its internal components.

Figure3.1- Schematic diagram showing the Michelson interferometer. *O-M* and *O-F* show the optical path difference (retardation). Replicated from **(3)**.



The Michelson interferometer is composed of a fixed mirror and a movable mirror. The IR beam is collimated by a lens as it exits the source, to allow the light rays to be parallel to each other, prior to it being split in to two separate beams by the beamsplitter - partially onto the movable mirror and partially onto the fixed mirror. The beams of IR radiation return to the beamsplitter where they interfere, thus producing an interferogram, and again are further partially reflected and transmitted. Ideally, a beamsplitter would consist of non-absorbent films and transmit/reflect at 50 %). The moving mirror can either be set to move at a constant velocity or paused for equal amounts of time and stepped rapidly; a moving mirror moving at a velocity of >0.1 $cm.s^{-1}$ is common place to allow for rapid continuous scanning (4).

In an ideal scenario, IR radiation would be perfectly collimated and the beamsplitter described as above; in such a case, the path difference between *O-M*

and *O-F* on Figure 3.1 is described as the optical path difference (OPD) (retardation). Since the optical path difference is uniform for every parallel input beams, it compensates for IR radiation which is not perfectly collimated. When the moveable and fixed mirrors are at an equal distance from the beamsplitter, the beams recombining at the beamsplitter interfere with one another perfectly. In such an instance, every intensity from the source reaches the detector and there is a zero path difference (zero retardation). It is at the point of zero retardation whereby the centreburst on an interferogram is produced (3). The purpose of the beamsplitter is to split an infrared beam in to two and recombine it again in to one single beam. The vast majority of beamsplitters consist of a film of Ge in-between two infrared transparent windows which transmit and reflect equal measures of IR radiation. Many of these infrared inactive windows are composed of potassium bromide (KBr) due to its transparency across a wide spectral range (400 cm^{-1} in to the NIR); below 400cm^{-1} the KBr windows strongly absorb IR radiation, therefore, this region of the spectrum is not collected. Another reason for a water vapour free purge is that KBr windows are hygroscopic (atmospheric water absorbing), which would eventually lead to water collecting on the KBr windows resulting in the limiting of IR light reaching the sample and detector. Zinc selenide (ZnSe) is a suitable alternative to KBr as it is not hygroscopic, however, it is significantly more expensive per window (£120-£1350 depending on thickness) compared to KBr (£50-£160) (5).

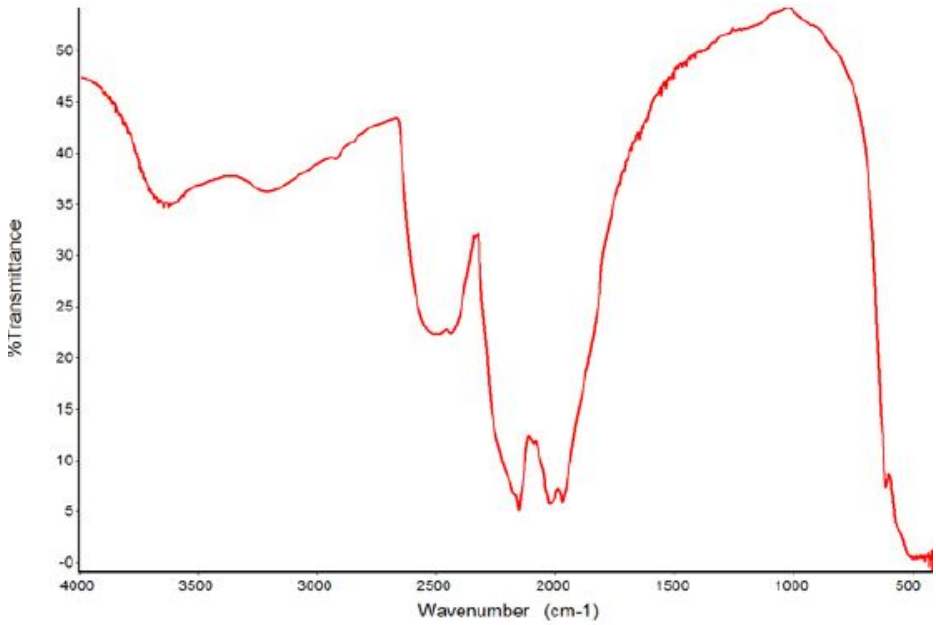
An interferogram is a plot of IR intensity against OPD; a samples characteristic absorptions are contained within the wavelengths of an

Figure3.2 - An example of a background spectrum

interferogram. When a sample is placed in the path of the IR beams interference, the loss of a beams intensity due to the absorption of the frequencies of light to the sample are measured by the detector. In summary, Fourier-transform is a mathematical function which is used to produce an IR spectrum from an interferogram.

3.2 Background Spectrum

If no sample is placed in the path of the interfering beam then the resulting spectrum is called a background spectrum. Figure 3.2 gives an example of a background spectrum (2).



A background spectrum represents the environment from inside and around the instrument and it is collected to allow for atmospheric correction. As described in chapter 2, an IR spectrum contains environmental spectral contributions from atmospheric gases such as CO_2 and H_2O vapour (2). Figure 3.3 shows the peaks of the described gaseous molecules. The two interferograms acquired from both the background and the sample are divided together, followed by the Fourier-transform calculation, thus resulting in a spectrum where environmental influences on the spectrum the have been accounted for (2).

3.3 Attenuated Total Reflectance (ATR) - FTIR Spectroscopy

Attenuated total reflectance (ATR) is as sampling mode used in IR spectroscopy based upon the internal reflection of IR radiation through an internal reflection element (IRE) such as Di or Germanium (Ge). The major benefit of ATR use in the analysis of human serum is that it is easy to use, cost effective and has rapid spectral acquisition times (6). The ATR crystal is ideal for drying biological samples onto to allow for sample dehydration to occur, thus reducing the spectral inhibition of highly absorbent water. During the drying process, intimate contact occurs between the sample and the IRE, enabling the evanescent wave to successfully penetrate the sample (7). ATR-FTIR was used over transmission and transflection sampling modes as they have major disadvantages. Transmission and transflection sampling modes suffer from Mie scattering and an effect known as the electric field standing wave during measurement due to sample thicknesses (height above

surface) due to particle size, respectively (8). ATR was used throughout this project however; a limitation of using the ATR sampling mode is that high-throughput sampling is not currently an option due to no accessory yet being available. Unless a multi-ATR crystal sampling device was created, however, single crystal ATR sampling is unlikely to be used for high-throughput in the clinical due it its impracticality (9).

A beam of IR light travels through a crystal of high refractive index (RI) (n_c), which then interacts with the sample placed upon the surface (n_s). θ_i represents the angle of incidence and θ_R the angle of refraction; as the angle of incidence increases, so does the angle of refraction. In order for total internal refraction to occur, the angle of refraction, θ_R, must exceed the critical angle (θ_C). The critical angle, θ_c, is calculated based upon the RI of the sample and the ATR crystal (equation 3.1) (2).

$$\theta_C = \sin^{-1}\frac{n2}{n1} \qquad \text{Equation 3.1}$$

where n1 = RI of ATR; n2 = RI of sample. High refractive materials are selected for the ATR crystal to allow for a minimal critical angle. Figure 3.3 shows the process of ATR with the aim being total internal refraction.

Figure3.3 - An example of an ATR crystal displaying the critical angle. Adapted from **(2)**.

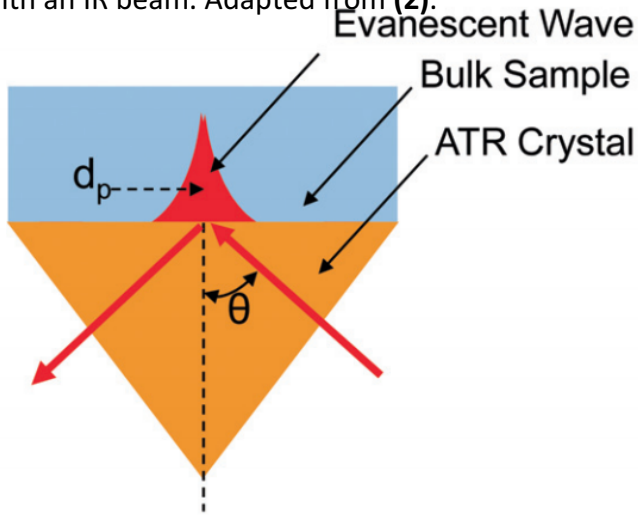
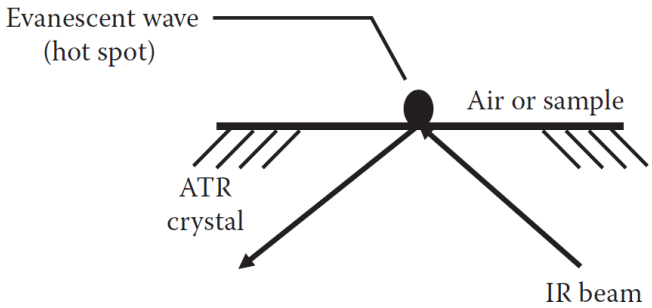


Figure3.4 - An example of an ATR crystal with an IR beam. Adapted from **(2)**.

If the RI of the ATR crystal is higher than the RI of the sample placed on top of the crystal, then internal reflectance occurs. Internal reflectance creates an evanescent wave which extends upwards above the ATR crystal into the sample. When the sample placed upon the crystal absorbs IR energy, the evanescent wave is attenuated. Following attenuation, the IR beam continues to the detector showing specific IR wavelengths where the sample has absorbed energy (2). Figure 3.4 presents an example of an ATR crystal with a path of IR radiation.



The distance which the evanescent wave can penetrate a sample is around 2-3 microns (10). The depth of penetration is an important consideration with ATR-FTIR as the evanescent wave extends from the IRE and penetrates into the sample (2). Equation 3.2 shows the equation for depth of beam penetration.

$$\mathrm{DP} = \frac{1}{[2\pi[\mathrm{W}\mathrm{n}_{\mathrm{c}}(\sin^2\theta - \mathrm{n}^2{}_{\mathrm{sc}})1/2]}$$ Equation 3.2

where DP is depth of penetration; W = wavenumber; n_c = RI of ATR crystal, θ = angle of incidence and $n_{sc} = \theta_C$ (2).

There are several factors which effect the spectra obtained via ATR-FTIR, they include: the ATR crystal material (RI index); RI of the sample; wavelength of the IR radiation; angle of incidence (θ_i) and the efficiency of contact with the sample (11).

Due to the exponential decrease of the evanescent wave as the distance from the surface of the crystal increases, an intimate contact between the sample and the crystal allows for the efficient penetration of the evanescent wave into the sample. Intimate contact between human serum and the ATR crystal is easily achieved due to the liquid biofluid drying on the surface. A drying study of human serum was conducted during this research and will be discussed in a subsequent chapter.

The sampling depth of ATR is wavelength dependent, and increases with decreasing wavenumber. The depth of penetration depends on the angle of

incidence. Varying the angle of incidence of the light and the RI of the IRE can control the sampling depth - increasing either decreases the sampling depth (12). The wavelength dependent behaviour of the evanescent wave leads to ATR-FTIR spectral band intensities decreasing with increasing wavenumbers; this is not the case with transmission experiments of the exact sample, thus the RI of the IRE effects acquired spectra (12). During the project described within this thesis, to correct for this, an in-built software ATR-correction method automatically corrected band position and shape (11). The RI of an ATR crystal has 2 major influences on an ATR spectrum. A high RI crystal allows for the critical angle (θ_C) to be kept to a minimum; it is important for the angle of incidence to exceed θ_C to avoid spectral distortions created from the combination of ATR and external reflectance.

A high RI ATR crystal will decrease the depth of penetration by the evanescent wave, thus decreasing the effective pathlength. Effective pathlength is used during ATR-correction by approximating a samples expected spectrum with transmission whereby the thickness of a sample is directly related to a spectrums absorbance intensities. The purpose of ATR-correct is to increase the peak intensities at higher wavenumbers on the IR spectrum relative to a spectrum acquired via transmission mode collection (2). Increasing the depth of penetration increases the absorbance intensities on a spectrum, thus reducing the use of approximation of ATR-correct on an acquired ATR spectrum (11).

The angle of incidence also plays a role in the collection of an ATR spectrum. An ATR crystal should be chosen with a RI to allow the angle of incidence to exceed

the critical angle (equation 3.7). Using equation 3.7, an ATR crystal which can achieve a greater angle of incidence should be selected to ensure that the spectra collected represents the sample. The angle of incidence effects the number of reflections within the ATR crystal; the greater the angle of incidence, the less intense the absorbance's are on the ATR spectrum, furthermore, a high angle of incidence decreases the evanescent wave penetration depth, thus reducing the absorbance intensities (11).

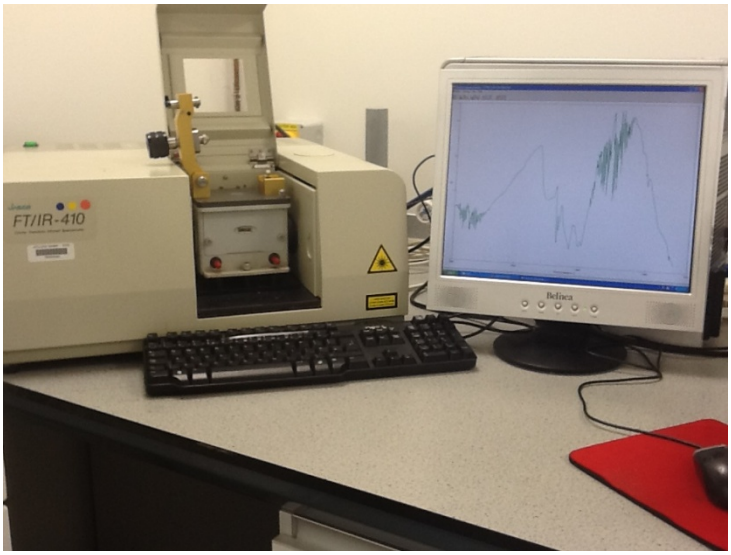
3.4 FTIR Instrumentation

During this PhD project, two FTIR spectrometers have been used to collect spectral data from pooled and patient serum samples. In subsequent research chapters, the specific FTIR spectrometer used to collect the spectral data will be stated.

3.4.1 JASCO FTIR Instrument

During the early stages of the PhD project, a JASCO FTIR-410 Specac ATR-single reflection diamond Golden Gate™ spectrometer was used to collect all absorption data. Figure 3.5 shows an image of the JASCO FTIR instrument with a computer monitor showing an infrared background spectrum.

igure3.5 - JASCO FTIR-410 Specac ATR spectrometer with computer monitor.



All data was collected in the range of 4000-400 cm^{-1}, at a resolution of 4 cm^{-1} and over 32 co-added scans. Prior to each spectral collection, a background absorption spectrum was collected for atmospheric correction.

3.4.2 Agilent Cary 600 Series FTIR Instrument

An Agilent Cary 600 Series FTIR instrument with a PIKE MIRacle™ single reflection ATR sampling accessory was also used to collect spectral data from patient serum samples. It is important to note that spectral data collected from both the JASCO and Agilent FTIR systems was not combined together for data analysis in this project; the wavenumber bins on both instruments do not match, thus all required sample spectra was re-collected upon installation of the Agilent FTIR system. Figure 3.6 shows an image of the Agilent Cary 600 Series FTIR spectrometer (13) and Figure 3.7 an image of the PIKE MIRacle™ single reflection ATR sampling accessory (14).

Figure3.6 - Agilent Cary 600 Series FTIR spectrometer (inside PIKE MIRacle™ ATR accessory). Reproduced from **(13)**.

Figure3.7 - PIKE MIRacle™ ATR accessory. Reproduced from **(14)**.

The Agilent FTIR system is a research grade instrument which can deliver excellent SNR and excellent spectral resolution with fast internal optic speeds to allow for optimal spectral collection (13). All data was collected in the range of 4000-400 cm^{-1}, at a resolution of 4 cm^{-1} and over 32 co-added scans. Prior to each spectral collection, a background absorption spectrum was collected for atmospheric correction.

3.4.3 Infrared Sources and Detectors

Both JASCO and Agilent systems used the same type of infrared source. A mid-infrared emitting light source, commercially known as Globar, is a resistively

heated silicon carbide rod which was used throughout this project. Globar sources are heated to approximately 1300 Kelvin and emit mid-IR radiation in the range of 50-6000 cm^{-1}(3).

Briefly, the purpose of a detector is to convert an intensity of light into an electrical signal (voltage). The JASCO-410 FTIR spectrometer and Agilent Cary 600 Series FTIR spectrometer both used a deuterated triglycine sulphate (DTGS) detector. A DTGS is optimal for use in the mid-infrared region when using KBr beamsplitter windows. DTGS detectors are suitable for ATR analysis, but they are known for their level of noise contribution to IR spectra when performing IR microscopy. Ideally, a mercury cadmium telluride (MCT) detector should be used during microscopy experiments due it being a photoconductor. An MCT detector produces a current when IR radiation is absorbed. Comparing an MCT detector to a DTGS detector, where the current generated varies with temperature, MCT detectors are reported to be 4 times faster and 10 times less noisy - however, MCT detectors are several thousand pounds more expensive and their detection range cuts off at around 700 cm^{-1}, thus concealing a part of the mid-IR region. DTGS detectors cut off at around 400 cm^{-1} making it more suitable for use with fingerprint region collection via ATR mode spectroscopy (2).

References

[1] Clemens, G., Bird, B., Weida, M., Rowlette, J., Baker, M. J. *Quantum cascade laser-based mid-infrared spectrochemical imaging of tissues and biofluids*. Spectroscopy Europe/Asia, 2014. 26(4), 14-20

[2] Smith, B C. *Fundamentals of Fourier-transform infrared spectroscopy.* 2nd Edition. CRC Press, 2007

[3] Griffiths, P. R., De Haseth, J. A. *Fourier transform infrared spectroscopy*.2nd Edition. Wiley, 2007

[4] Banwell, C., McCash E. *Fundamentals of molecular spectroscopy*. 4th Edition. McGraw-Hill Higher Education, 1994

[5] Edmund Optics. *UV and IR windows*. [Online] http://www.edmundoptics.co.uk/optics/windows-diffusers/ultraviolet-uv-infrared-ir-windows/potassium-bromide-kbr-windows/3328/ [Accessed on 3rd August 2015]

[6] Dorling, K. M., Baker, M. J. *Highlighting attenuated total reflection Fourier transform infrared spectroscopy for rapid serum analysis*. Trends in Biotechnology, 2013. 31(6), 327-328

[7] Shaw, R. A., Mantsch, H. H. *Infrared spectroscopy in clinical and diagnostic analysis* in Encyclopaedia of Analytical Chemistry, 2006

[8] Pilling, M. J., Bassan, P., Gardner, P. *Comparison of transmission and transflectance mode FTIR imaging of biological tissue*. Analyst, 2015. 7(140),2383-2392

[9] Hughes, C., Brown, M., Clemens, G., Henderson, A., Monjardez, G., Clarke, N. W., Gardner, P. *Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum*. Journal of Biophotonics, 2014. 7(3-4), 180-188

[10] Khoshmanesh, A., Dixon, M. W., Kenny, S., Tilley, L., McNaughton, D., Wood, B. R. *Detection and quantification of early-stage malaria parasites in laboratory infected erythrocytes by attenuated total reflectance infrared spectroscopy and multivariate analysis*. Analytical Chemistry, 2014. 86(9), 4379-4386

[11] SpectraTech. *Introduction to attenuated total internal reflectance (ATR)* [Online] http://www.nicoletcz.cz/userfiles/file/Aplikace/Introduction%20ATR%20SpectraTech.pdf [Accessed 4th August 2015]

[12] Larkin, P. *Infrared ad Raman spectroscopy: principles and spectral interpretation*. Elsevier, 2011

[13] Agilent Technologies. *Agilent Cary 600 Series FTIR* [Online] http://www.chem.agilent.com/Library/brochures/Agilent_FTIR_660-670-680_Brochure.pdf [Accessed 4th August 2015]

[14] PIKE Technologies. *MIRacle™ single reflection ATR* [Online] http://www.piketech.com/atr-miracle-tm-single-reflection.html [Accessed 4th August 2015]

Chapter 4

AN ORTHOGONAL APPROACH: INVESTIGATING THE POTENTIAL USE OF ATR-FTIR FOR RAPID GLIOMA DIAGNOSIS

4.1 Introduction and Aims

This chapter investigates the potential use of ATR-FTIR spectroscopy to identify cancer from non-cancer from whole (unfiltered) and 10 kDa (kiloDalton) filtered serum samples.

Prior to our spectral analysis of the human serum samples our collaborator Peter Abel at the University of Central Lancashire used a Bioplex immunoassay to provide cytokine and angiogenesis factor levels that differ between serum samples from glioma and non-cancer patients. Specifically angiopoietin, follistatin, leptin and interleukin-8 factors were found to be most discriminatory. The cytokines and angiogenesis factors responsible for discrimination have varying molecular weights; hence we perform a filtration study to assess which fraction of serum is most suitable for our analyses. To study this, an RBF-SVM was used to investigate the spectral data was collected from patient serum samples. Sampling protocols were determined and optimised (e.g. drying time for serum dehydration) and whole (unfiltered) serum and filtrate aliquots were analysed to determine the most discriminatory fraction of serum.

The work presented within this chapter has been published in the Journal of Analytical and Bioanalytical Chemistry, 2013, Volume 405(23); 7347-7355.

Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors.

4.2 Experimental Details

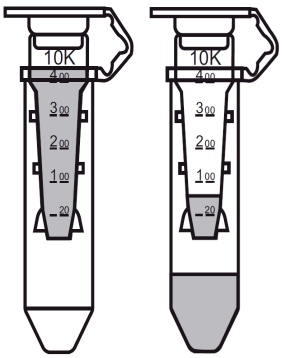
4.2.1 Biological Samples

Human blood was collected from 49 patients diagnosed with a WHO grade GBM brain tumour and from 25 patients who do not have cancer (non-cancerous state). All serum samples were collected from the Walton Centre Research Tissue Bank (Liverpool, England) and Brain Tumour North West (BTNW) Tissue Bank (Preston, England) with full ethical approval (BTNW and Walton Centre ethical application #1108). Only those patients with a newly diagnosed GBM were included in the study - no patients had a recurring tumour. All patient blood samples were collected pre-operatively and left to clot at room temperature for 30-120 minutes from blood draw. Centrifugation of blood samples at 1200x*g* for 10 minutes was performed to separate the clot, followed by the serum being aliquoted into cryovials. All samples were snap frozen using liquid nitrogen and were stored at a temperature of -80°C. Samples were collected from the hospital on dry-ice and stored in a -80°C. When required, samples were thawed at room temperature for 30 minutes, gently shaken and a volume pipetted for analysis. Serum samples did not undergo the freeze-thaw cycle on more than 3 occasions prior to spectral collection. The latter was consistent throughout the project. Appendix 1 shows patient demographic data. The patient dataset has an entire age mean of 60.18 years. The non-cancer patients have a mean age of 60.16 years; likewise, the GBM patients

have a mean age of 60.18. Appendix 1 displays each patient's gender, age and diagnosis - at the time of sample collection, further patient information (BMI, weight, treatment and medication use) was unavailable from our collaborating tissue bank.

4.2.2 Serum Preparation for Filtration Study

All whole serum samples were thawed for 30 minutes prior to spectral collection 10 kDa filtration 1 ml aliquots were prepared to investigate the most discriminatory fraction of serum using Amicon Ultra-0.5 mL centrifugal filters (purchased from Millipore Limited, UK) (Figure 4.1). Centrifugal filters filter out components of the serum above the cut-off point of the filters membrane (i.e. 10 kDa), allowing components below the filter membrane cut-off point to pass through. Each whole serum sample had a 10 kDa filtration aliquot prepared by pipetting 0.5 mL of the whole serum in to the filtration device and centrifuging at 14,000 rpm for 15 minutes. Figure 4.1 shows a centrifugal filter where a volume of serum (0.5ml) is pipetted into the top (left) and centrifuged so that the filter retains all serum constituents greater than the kDa range, thus only allowing through the serum filtrate which contains constituents below the maximum range.



;ure 4.1 - Centrifugal filters used to filter serum constituents greater than the oDalton range's cut off point. Reproduced with permission from Millipore Ltd.

4.2.3 Drying Study

Effective spectral collection via ATR-FTIR spectroscopy requires intimate contact between the ATR Di crystal and the sample due to the penetration of the ~2.0 micron evanescent wave (1)*;* as such a drying of human serum was investigated. Whole normal (non-cancerous) human mixed pooled serum (0.2 µL sterile filtered, CS100-100, purchased from TCS Biosciences, UK) was used in volumes of one microlitre to determine the optimal drying time necessary for quality spectral collection, in addition, 10 kDa serum filtrate aliquots were also prepared to determine drying time of a filtrate sample. Following the collection of spectral data from dried serum films on the ATR crystal, Virkon® (purchased from Antec Int., Suffolk, UK) and absolute ethanol (purchased from Fisher Scientific, Loughborough, UK) were used to clean the crystal. In total, 666 spectra (74 patients with 9 spectra each) were collected from all whole serum samples, and another 666 spectra from the serum filtrate aliquots.

4.2.4 Non-Cancer (Control) Sample Analysis

A control study was conducted to observe whether ATR-FTIR spectroscopy could distinguish sex (male *vs*. female) and young *vs*. old 1 microlitre volumes of patient serum. All spectral data was acquired from the described method in 4.2.5. The analysis of the 25 non-cancerous state serum samples within this study will provide us with an insight regarding any other potential sources of classification. The spectral data acquired was split up into two datasets - one dataset for sex determination and one to observe patients in a young *vs*. old manner. To allow for a

sufficient number of patients in both young and old classes, it was decided that patients younger than 50 years old were 'young' and patients older than 50 years were 'old', thus 7 patients were considered young and 18 patients 'old'.

4.2.5 ATR-FTIR Spectral Collection

A JASCO FTIR-410 Specac ATR Di Golden Gate™ spectrometer was used to collect absorption data from all 74 serum samples in this initial study. All data was collected in the range of 4000-400 cm^{-1}, at a resolution of 4 cm^{-1} and using 32 co-added scans. Prior to each spectral collection, a background absorption spectrum was collected for atmospheric correction.

All serum samples were analysed in a random order within the whole and 10 kDa sample sets. For each sample, the described procedure was repeated 3 times: a 1 microlitre spot of patient serum was pipetted onto the ATR crystal and allowed to dry for 8 minutes, at which point 3 spectra were collected. A background absorption was collected prior to analysis and in between each 3 sets of spectra acquired to allow for atmospheric correction. Following the collection of spectral data from dried serum films on the ATR crystal, Virkon® (purchased from Antec Int., Suffolk, UK) and absolute ethanol (purchased from Fisher Scientific, Loughborough, UK) were used to clean the crystal. In total, 666 spectra (74 patients with 9 spectra each) were collected from all whole serum samples, and another 666 spectra from the serum filtrate aliquots.

4.3 Drying Study Results

In total, 666 spectra (74 patients with 9 spectra each) were collected from all whole serum samples, and another 666 spectra from the serum filtrate aliquots. Effective spectral collection via ATR-FTIR spectroscopy requires intimate contact between the ATR Di crystal and the sample due to the penetration of the ~2.0 micron evanescent wave (1). Figure 4.2 shows offset ATR-FTIR spectra of whole and 10 kDa serum.

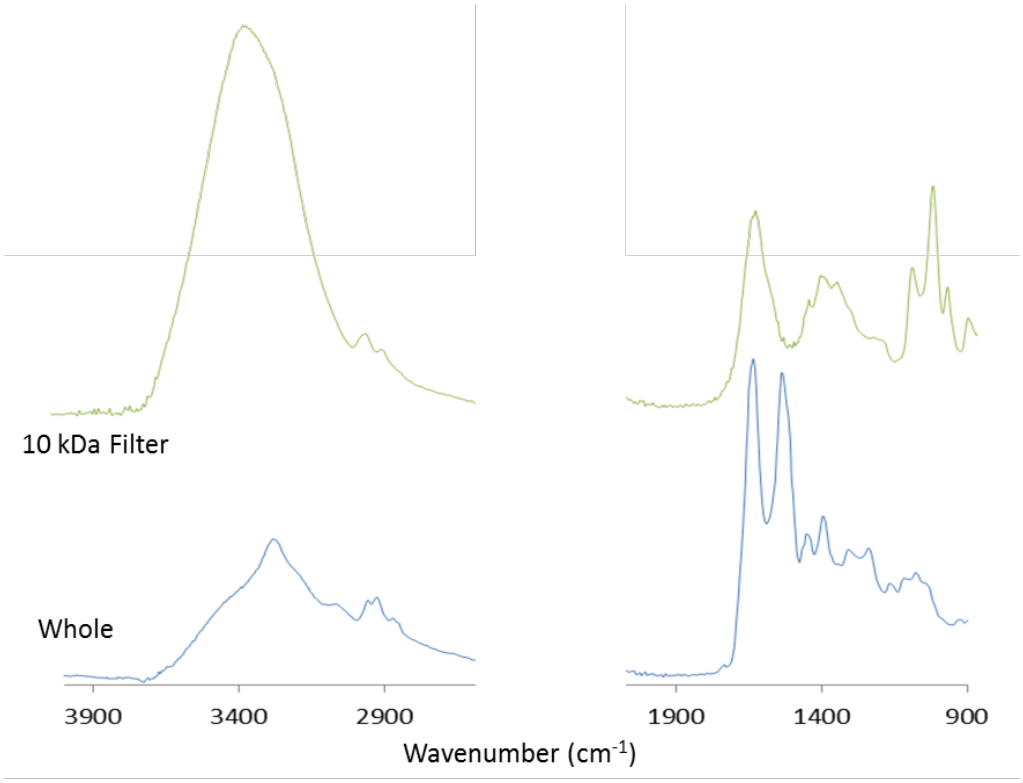


Table 4.1 shows proposed biomolecular assignments for the dried whole serum spectrum and Table 4.2 for dried 10 kDa serum. The major differences between the whole and 10 kDa serum samples include the presence of the Amide II

ble 4.1 - The proposed biomolecular assignments of the whole dried serum film
ectrum (Figure 4.2) (8-18).

band (1637 cm^{-1}) in the whole serum spectrum but not in the 10 kDa serum filtrate spectrum. The lipid CH_2 band (1307 cm^{-1}) is also present in whole serum but not in the serum filtrate samples. The 10 kDa serum filtrate spectrum holds an intense band at 1032 cm^{-1}, tentatively assigned as C-O vibration of ribonucleic acid (RNA) and ribose/glucose (8-18).

Wavenumber (cm^{-1})	**Proposed Biomolecular Assignment**
3441	O-H stretch of hydroxyl groups
3286	CH_3 stretch (antisymmetric), fatty acids, lipids, proteins
3071	CH_3 stretch (antisymmetric), fatty acids, lipids, proteins
2958	CH_2 stretch (antisymmetric), fatty acids, lipids, proteins
2933	CH_2 (antisymmetric), methylene group of membrane phospholipids
2866	CH_2 stretch (antisymmetric), fatty acids, lipids, proteins
1741	C=O stretch (symmetric), fatty acids, lipids, proteins
1637	Amide I
1536	Amide II
1449	CH_2 deformation of methylene group, lipids
1399	CH_3 (asymmetric), lipids
1307	Lipids CH_2 twist, protein amide II band, cytosine, adenine
1204	C-C, C-H, PO_2^- phosphodiester
1170	C-O (antisymmetric), lipids
1018	C-O (glucose)
924	C-C-N backbone, C-C, glucose

Table 4.2 - The proposed biomolecular assignments of the 10 kDa dried serum film spectrum (Figure 4.1) (8-18).

Wavenumber (cm^{-1})	**Proposed Biomolecular Assignment**
3308	O-H stretch of hydroxyl groups
2941	CH_3 (antisymmetric), fatty acids, lipids, proteins
2883	CH_2 (antisymmetric), methylene group of membrane phospholipids
1641	Amide I
1451	CH_2 deformation of methylene group, lipids
1403	CH_3 deformation, lipids
1345	Adenine, CH deformation
1215	C-C, C-H bend
1107	PO_2 symmetric stretch
1032	C-O RNA, ribose/glucose vibration
991	C-C, C-N, PO_3 stretch
916	C-C, glucose

One microlitre volumes of pooled serum were pipetted onto the ATR crystal and a spectrum was collected at 0, 2, 4, 8, 16 and 32 minute intervals to observe spectral changes during the drying process. One biological repeat and two technical repeats were collected per 1 µL of dried serum. The drying experiment was repeated 70 times to gain representative spectra at specific times during drying of the sample. It was determined that after 8 minutes of drying at ambient room temperature (~18°C) a one microlitre serum film was stable with no further spectral

changes being observed upon visual spectral inspection. The dried serum film at 8 minutes (Figure 4.3) matches the representative film described by Lie *et al*. (2).

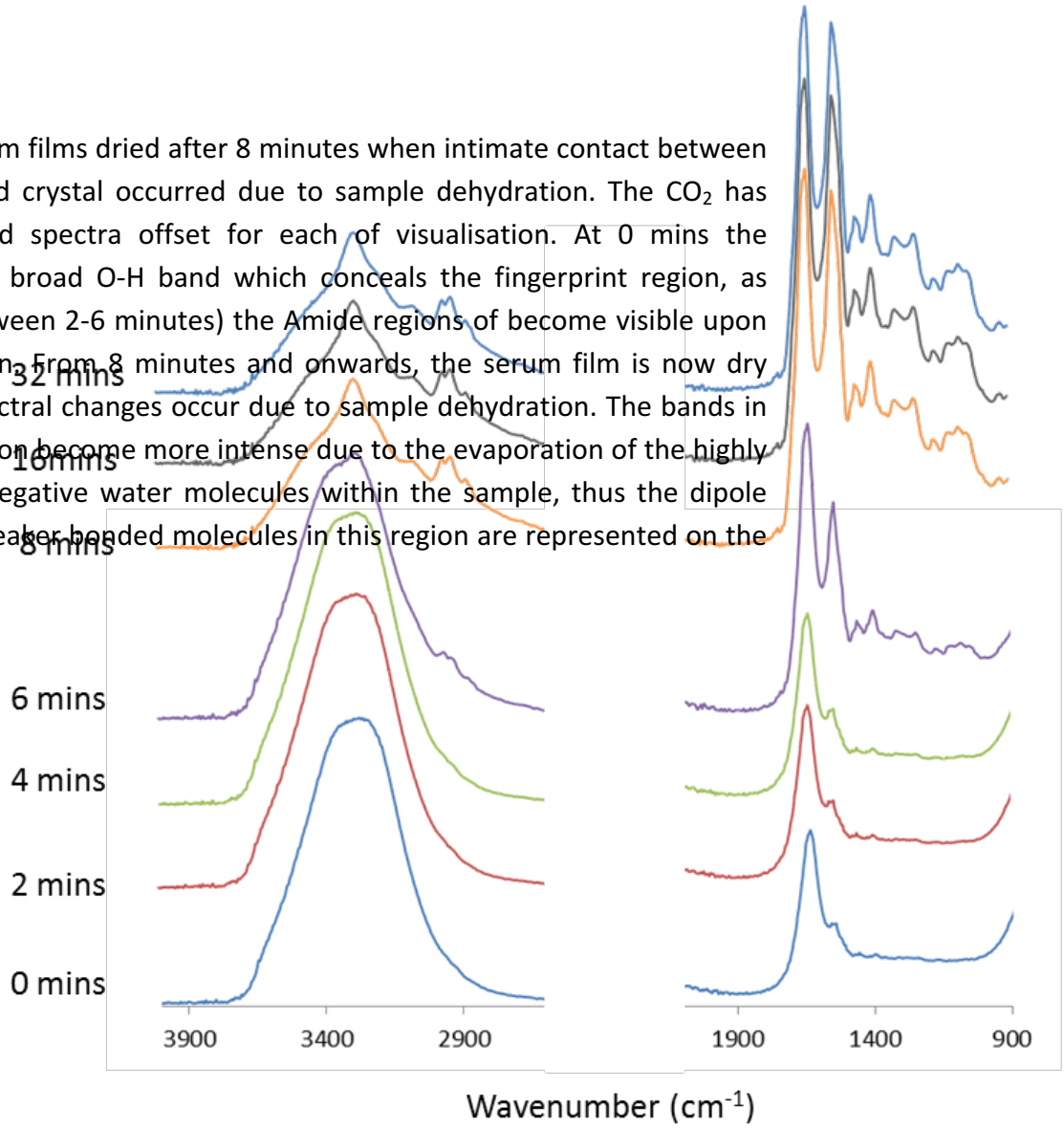
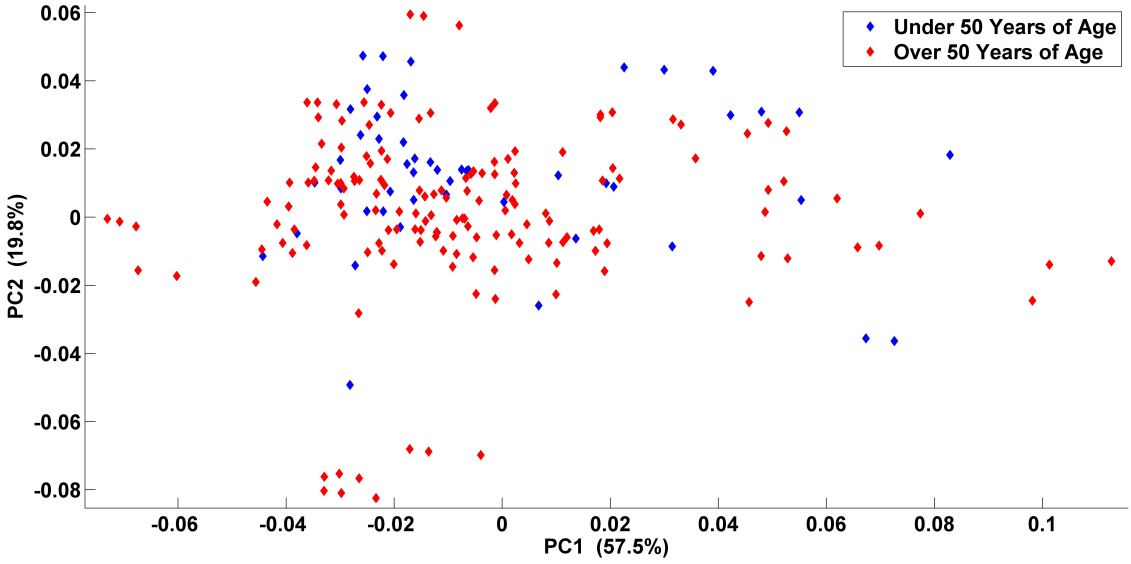


Figure 4.3- All serum films dried after 8 minutes when intimate contact between the serum film and crystal occurred due to sample dehydration. The CO_2 has been removed and spectra offset for each of visualisation. At 0 mins the spectrum shows a broad O-H band which conceals the fingerprint region, as drying occurs (between 2-6 minutes) the Amide regions of become visible upon sample dehydration. From 8 minutes and onwards, the serum film is now dry and no further spectral changes occur due to sample dehydration. The bands in the fingerprint region become more intense due to the evaporation of the highly polar and electronegative water molecules within the sample, thus the dipole moments of the weaker bonded molecules in this region are represented on the IR spectrum

(control) patients who are under 50 years of age - 'young' (blue) and patients

4.4 Non-Cancer (Control) Sample Analysis Results

As described in 4.2.4, the analysis of the non-cancerous control samples was conducted to determine any other potential causes of classification, such as patient age or sex. Figure 4.4 shows a PCA scores plot of the non-cancerous patients.



Figu

cate

than 50 years of age are largely based between -0.02 and 0, furthermore, the *y*-axis, PC2, does not describe separation between the two classes. It would be expected to observe a separation between younger and older aged individuals; however, the sample data here has not demonstrated this, thus it can be concluded that the age of the patients within this non-cancer classification group are not influencing the classification of test set data (discussed further in 4.5).

(control) male (blue) and female (red) serum samples (Appendix 1 for individual patient ages).

Figure 4.5 shows a PCA scores plot of the non-cancerous control serum sample data to show separation between male and female patients.

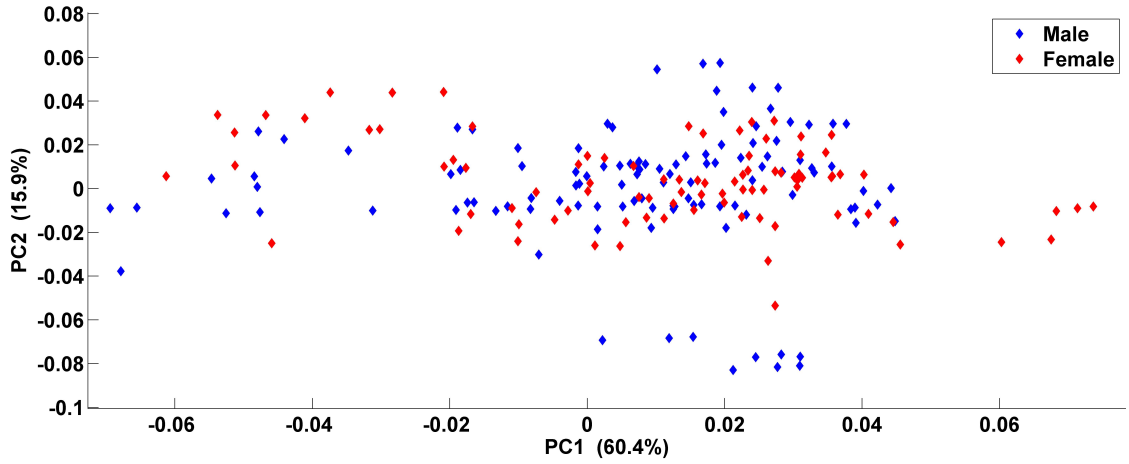


Figure 4.5 shows no separation between male and females non-cancerous patients. The *x*-axis, PC1, shows that the vast majority of the patients are between -0.02 and 0.045 with no clear separation. Figure 4.5 does not present a clear separation between male and female patient serum spectra; therefore, it is reasonable to conclude that the sex of the patients within the 25 patient non-cancer serum dataset does not influence the classification of the patients within the test set data.

4.5 ATR-FTIR Data Pre-processing and Multivariate Analysis

Pre-processing and multivariate analysis were performed on the raw spectral data using Matlab (version 7.11 [R2010b]) (The MathWorks, Inc., USA) using in-house written software and LibSVM 3.12.

The fingerprint region (1800-1000 cm^{-1}) was extracted and selected for analysis to allow. The fingerprint region is known to produce different spectral patterns and troughs in response to different components of compounds (1). A visual quality test was performed to check for gross spectral error. To do this, the maximum absorbance point (~3290 cm^{-1}) of the O-H region was observed to have absorbance units between 0.17 - 0.23; Amide I and II to have absorbance units between 0.35 - 0.45 and the spectrum to have a smooth baseline with little or no noise contribution. Spectra which have gross errors could potentially skew a diagnosis based on non-biochemical information. Noise reduction (50 PCs) and vector normalisation was performed on the data to mathematically correct for inconsistencies due to changing environment and instrumental factors.

Spectra were organised into two sets on a patient level to allow the SVM model to be trained with the majority of the data prior to being challenged with the test set, as follows:

- Training set - 450 spectra (2/3 of the entire dataset)
- Test set - 216 spectra (1/3 of the entire dataset)

An RBF-SVM was used for its ability to easily optimise the cost and gamma model parameters. Leave-one-out cross validations were performed using the train sets of the whole serum and 10 kDa serum filtrate samples to determine the optimum cost and gamma values (Table 4.3). Table 4.3 shows the SVM parameters used for the SVM kernel within this chapter. Cross-validation accuracy is conducted to test the model to understand how it will perform with an independent dataset. The SVM model was trained using the optimum cost and gamma values prior to being tested with the blind dataset.

e 4.3 - Optimal cost and gamma values for whole serum and 10 kDa serum filtrate
m, presented with cross validation accuracies

	Whole Serum Dataset	**10 kDa Serum Dataset**
Optimum Cost	1024	16384
Optimum Gamma	8	16
Cross-Validation Accuracy	99.80 %	99.30 %

4.6 ATR-FTIR Diagnostic Model

In order to assess and measure the quality of the spectral discrimination between the GBM and non-cancer serum states, sensitivities and specificities were used.

Sensitivity measures the ability of a model to correctly classify (equation 4.1), and specificity measures its ability not to misdiagnose (equation 4.2).

$$\text{Sensitivity} = \frac{\text{True Postives}}{\text{True Positives} + \text{False Negatives}} \quad \text{Equation 4.1}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad \text{Equation 4.2}$$

where:

- ***True Positives*** is the number of spectra where samples were correctly identified as cancerous;
- ***True Negatives*** is the number of spectra where the samples have been correctly identified as non-cancerous;
- ***False Positives*** is the number of spectra where the model has incorrectly identified the samples as cancerous when it is non-cancerous;
- ***False Negatives*** is the number of spectra where the model has incorrectly identified samples as non-cancerous when it is cancerous.

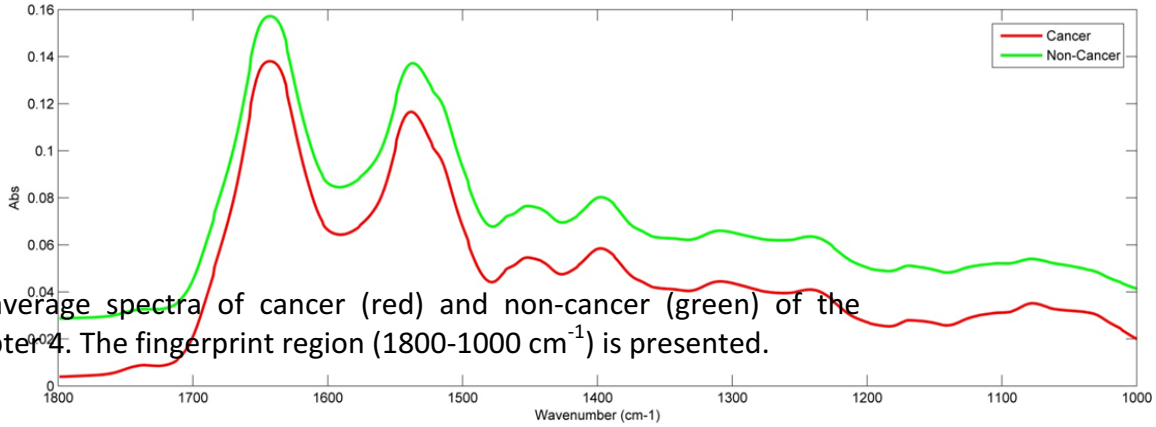
4.6.1 Whole Serum Diagnostic Ability

The whole serum RBF-SVM diagnostic model misclassified 21 of the 216 test set data when it was challenged with the blind test dataset, thus resulting in spectral-based sensitivities and specificities of 88.2 % and 94.4 %. Table 4.4 shows the patients whereby spectra were misclassified from the whole serum blind test.

le 4.4 - Patient level results - patients' numbers showing the number of classified spectra per patients from the blind test set for the whole serum gnostic model

	Patient Number	Misclassified Spectra
GBM	12012	1/9
Patients	12019	2/9
	12055	**5/9**
	11156	**9/9**
Non-Cancer	600	3/9
Patients	607	1/9

With regards to Table 4.5, if more than 4 out of 9 patient spectra are classed as GBM or non-cancer then this is was considered as the diagnosis. Table 4.5 shows that only 2 GBM patients (12055 and 11156) were misclassified as non-cancer; no non-cancer patients were diagnosed as GBM. Sensitivities and specificities achieved on a patient level were 87.5 % and 100.0 % respectively. Figure 4.6 shows two average spectra from the cancer and non-cancer classes of spectra within this chapter.



igure 4.6 - Two average spectra of cancer (red) and non-cancer (green) of the atients within chapter 4. The fingerprint region (1800-1000 cm^{-1}) is presented.

The spectral which occur between the two-classes are assigned to 4 discriminatory wavenumber regions - the Amide II band/CH_2 deformation at 1490-1535 cm^{-1}, Amide II band at 1535-1570 cm^{-1}, Amide I band at 1600-1630 cm^{-1} and C-C, C-O, DNA/RNA contributions between 1040 and 1230 cm^{-1}. The Bioplex immunoassay data collection from these samples (discussed further in 4.7) shows that there is a significant concentration increase in certain cytokines and angiogenesis factors within the serum collected from cancer patients compared to non-cancer patients. The slight spectral changes observed on Figure 4.4 between cancer *vs*. non-cancer patients could possibly be due to the increase of PECAM-1, PDGF, leptin and interleukin-8 being secreted into a patient's bloodstream due to the development of cancer. As discussed earlier in 4.2.1, patient BMI, weight and medication use is unknown from these samples, thus analysis of this data without this patient information does now allow the definitive conclusion that the spectral changes are solely a biochemical response to cancer, rather than patient medication use and BMI allowing for this separation.

ıle 4.5 - Patient level results - patients' numbers showing the number of classified spectra per patients from the blind test set for the 10 kDa filtrate serum ɡnostic model

4.6.2 Filtrate Serum Diagnostic Ability

The 10 kDa serum filtrate RBF-SVM diagnostic model misclassified more spectra in comparison to the whole diagnostic model. The diagnostic model misclassified 38 out of 216 spectra when challenged with the blind test dataset, thus a sensitivity of 81.4 % and a specificity of 81.9 % was achieved on a spectral level. Table 4.6 shows the patients whereby spectra were misclassified from the 10 kDa filtrate serum blind test.

Table 4.5 shows that just 4 patients diagnosed with GBM were misclassified as non-cancer and only 1 non-cancer patient diagnosed as GBM, thus a patient-level based sensitivity and specificity of 78.9 % and 88.9 % was achieved respectively.

	Patient Number	**Misclassified Spectra**
GBM	**11121**	**6/9**
Patients	**11154**	**6/9**
	12019	**6/9**
	12052	**8/9**
Non-Cancer	572	2/9
Patients	**593**	**6/9**
	594	**4/9**

In comparison to the 10 kDa serum filtrate aliquot samples, the whole serum RBF-SVM diagnostic model achieved better spectral and patient-level sensitivities and specificities; however, due to the relatively high sensitivities and specificities achieved with the 10 kDa serum RBF-SVM diagnostic model, there is reason to indicate that sub-fraction biomolecular constituents of whole serum may provide interesting targets for diagnosis.

4.7 Orthogonal Approach

Our collaborator, Peter Abel, at the University of Central Lancashire (England) conducted a study using a Bioplex immunoassay, with exactly the same whole serum samples as those used in the spectroscopic study to investigate the significant cytokine and angiogenesis factors enabling for the differentiation between GBM and non-cancer patients. For brevity, it was found that the most significant component of the serum samples allowing for discrimination was follistatin. Follistatin achieved the highest sensitivity and specificity with 88.0 and 81.0 % respectively. In addition to follistatin, other cytokines and angiogenesis factors also allowed for discrimination - angiopoietin, interleukin-8 and leptin are among those biochemical factors which are able to distinguish between GBM and non-cancerous disease state. The cytokines and angiogenesis factors found to be responsible for discrimination between GBM and non-cancer have various molecular weights. Follistatin has a molecular weight of 35-70 kDa (3); interleukin-8 a molecular weight of 8.4 kDa (4); angiopoietin 55 kDa (5); leptin 16 kDa (6) and

PECAM-1 130 kDa (7), thus it was decided to initially investigate a 10 kDa cut-off fraction of the whole serum via ATR-FTIR spectroscopy where all biochemical components of the filtrate has a molecular weight less than 10 kDa. The BioPlex immunoassay study found that cancer patient serum samples have a significantly higher concentration of PECAM-1, PDGF, leptin, interleukin-8, follistatin and angiopoietin compared to non-cancerous serum state samples. The signalling pathway of angiogenesis in high-grade GBM patients are mainly controlled by molecules which include angiopoietin, PDGF, interleukin-8, follistatin, hepatocyte growth factor and tumour necrosis factors. The discussed biomolecules are attributed with brain tumour development and tumour angiogenesis.

Table 4.6 shows those patients which were misclassified (patient level) based upon serum follistatin levels and from serum ATR-FTIR spectroscopic diagnosis. The total number of patients misclassified for both techniques is few, thus a two-stage diagnostic method utilising ATR-FTIR spectroscopy and follistatin levels has great potential for rapid brain cancer diagnoses.

le 4.6 - Patient level results - patient numbers who were misclassified based on statin levels and whole serum ATR-FTIR diagnostic model outcomes

	False Negatives	False Positives
Follistatin	11205	585, 594, 607
Level	11202	609, 589
(95% CI)		
Whole	12055, 11159	-
Serum		
ATR-FTIR		

4.7 Discussion

The work described within this chapter has shown that spectroscopic signatures acquired from ATR-FTIR and patient serum samples can be used to diagnose GBM to sensitivities and specificities as high as 87.5 % and 100.0 % respectively. ATR-FTIR spectroscopy can provide a diagnosis in just 10 minutes of serum deposition onto the ATR crystal and within 4 hours with a Bioplex immunoassay study. The orthogonal diagnostic procedure described here is based upon very different phenomena - the serum concentration of cytokines and angiogenesis factors and the molecular response to vibrational spectroscopy of whole and filtrate serum. A limitation presented within this chapter and all patient blood serum samples used throughout this thesis is the lack of individual patient information, e.g. BMI, drug use, diet, lifestyle, etc. Patients who have a blood sample collected pre-operatively are highly likely to be on an array of medication for

cancer or have been treated with chemotherapy/radiotherapy, thus impacting upon the blood serum sample acquired. In contrast, a patient who donated a non-cancerous sample will not have the physiological changes induced by the intake of cancer medications and therapies. To summarise, cancer patients are likely to be on pain killers, steroids and have different regional physiological functions in comparison to those patients which have not been diagnosed with cancer. The age and sex of the patients also has the potential to affect the acquired data. It would be reasonable to assume that older aged patients take medications on a regular basis in comparison to younger adults. As discussed in Chapter 1, the human metabolome impacts upon the concentration of biomolecules within blood serum based on the donor's diet, fitness regime and BMI etc. This chapter focuses on a proof-of-principle concept of using ATR-FTIR spectroscopy for rapid brain tumour identification from minute volumes of blood serum.

There are a number of factors which can greatly impact the analysis of human serum samples. These factors include hemolysis, sample filtration, freezing/thawing, storage/transportation and histopathological diagnoses. The problem of hemolysis was considered throughout this project and a visual examination of all blood serum samples prior to spectral analysis was conducted - all samples were straw-yellow coloured prior to analysis. The filtration step in this chapter provided sensitivities and specificities which were not as high as those achieved with whole serum; in addition, the filtration step is time consuming and requires costly molecular filters. Subsequent chapters will further analyse serum

filtration to determine which fraction of whole serum is most discriminatory for disease identification.

The whole serum ATR-FTIR spectral diagnostic models misclassified 2 GBM patients - 12055 and 11159. A potential reason for the misclassification of these 2 GBM patients could relate to the age and sex of the patients. Both patients are male; 12055 is 25 years old and 11159 is 57 years old. Additionally, these 2 patients were both successfully classified as GBM patients based upon the blood serum follistatin concentration. It should not be dismissed that both patients could potentially have medical complications alongside their GBM tumour, thus they could be taking different medications compared to those patients with just a primary GBM, although this information was not available from the tissue bank. Both patients could be extremely active or be completely inactive - their BMI values, lifestyle and medication-use has the potential to play a significant role in the reasoning behind their misclassification, likewise, these factors could potentially be key to the classification of a patient spectrum to a certain disease class. A definitive classification of solely cancer or non-cancer for a patient classification is not possible without interrogating the discussed additional patient information, which is unavailable from the tissue bank. Future work on this project would aim to acquire detailed patient information (unavailable at time of analysis) from the tissue bank to further investigate patient classifications.

For a single patient serum sample, if the same diagnosis is found from both tests then the overall diagnosis has a greater diagnostic strength and the patient can

be referred for further diagnostic techniques such as CT, MRI etc. As both the Bioplex immunoassay and spectral study achieved such high sensitivities and specificities, should a conflicting overall diagnosis occur between both there is still reason to further investigate the potential of the patient having cancer.

References

[1] Smith, B. C. *Fundamentals of Fourier transform infrared spectroscopy*. 2nd Edition. CRC Press, 2011.

[2] Liu, K. Z., Shaw, R. A. Man, A. Dembinski, T. C., Mantsch, H. H. *Reagent-free, simultaneous determination of serum cholesterol in HDL and LDL by infrared spectroscopy*. Clinical Chemistry, 2002. 48(3), 499-506

[3] Santa Cruz Biotechnology, Inc. *Follistatin (H-144): sc030194*, 2015 [Online] http://datasheets.scbt.com/sc-30194.pdf [Accessed on 13th August 2015]

[4] Sigma-Aldrich. *Interleukin-8 human*, 2015 [Online] http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/2/i1645dat.pdf [Accessed on 13th August 2015]

[5] abcam®. *Angiopoietin-1 (ab8451)*, 2015 [Online] http://www.abcam.com/angiopoietin-1-antibody-ab8451.html [Accessed 13th August 2015]

[6] Liu, C., Liu, X. J., Barry, G., Ling, N., Maki, R. A., De Souza, E. B. *Expression and characterization of a putative high affinity human soluble leptin receptor*, 1997. 138(8), 3548-3554

[7] Santa Cruz Biotechnology, Inc. *PECAM-1 (M-20): sc-1506*, 2015 [Online] http://datasheets.scbt.com/sc-1506.pdf [Accessed on 13th August 2015]

[8] Baker M J., Gazi E., Brown M D., Shanks J H., Gardner P., Clarke N W. *FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer*. British Journal of Cancer, 2008. 99(11), 1859-1866

[9] Diessel, E., Willmann, S., Kamphaus, P., Kurte, R., Damm, U., Heise, H. M. *Glucose quantification in dried-down nanoliter samples using mid-infrared attenuated total reflection spectroscopy*. Applied Spectroscopy, 2004. 58(4), 442-450

[10] Shaw, R. A., Mantsch, H. H. *Infrared spectroscopy in clinical and diagnostic analysis* in Encyclopaedia of Analytical Chemistry, 2006

[11] Goormaghtigh, E., Raussens, V., Ruysschaert, J. M. *Attenuated total reflection infrared spectroscopy of proteins and lipids in biological membranes*. Biochim Biophys Acta, 1999. 1422(2), 105-185

[12] Rohleder, D., Kocherscheidt, G., Gerber, K. Kiefer, W. Kohler, W., Mocks, J., Petrich, W. *Comparison of mid-infrared and Raman spectroscopy in the quantitative analysis of serum*. Journal of Biomedical Optics, 2005. 10(3), 031108

[13] Naumann, D. *FT-infrared and FT-Raman spectroscopy in biomedical research*. Applied Spectroscopy Reviews, 2001. 36(2-3), 239-198

[14] Meade, A. D., Lyng, F. M., Kneif P., Byrne, H. J. *Growth substrate induced functional changes elucidated by FTIR and Raman spectroscopy in in-vitro cultured human keratinocytes*. Analytical and Bioanalytical Chemistry, 2007. 387(5), 1717-1718

[15] Bellisola, G., Sorio, C. *Infrared spectroscopy and microscopy in cancer research and diagnosis.* American Journal of Cancer Research, 2012. 2(1), 1-21

[16] Tamm, L. K., Tatulian, S. A. *Infrared spectroscopy of proteins and peptides in lipid bilayers*. Quarterly Reviews of Biophysics, 1997. 30(4), 365-429

[17] Petrich W., Lewandrowski K B., Muhlestein J B., Hammond M E., Januzzi J L., Lewandrowski E L., Pearson R R., Dolenko B., Fruh J., Haass M., Hirschl M M., Kohler W., Mischler R., Mocks J., Ordonez-Llanons J., Quarder O., Somorjai R., Staib A., Sylven C., Werner G., Zerback R. *Potential of mid-infrared spectroscopy to aid the triage of patients with acute chest pain*. Analyst, 2009. 134(6), 1092-1098

[18] Gajjar K., Trevisan J., Owens G., Keating P J., Wood N J., Stringfellow H F., Martin-Hirsch P L., Martin F L. *Fourier-transform infrared spectroscopy coupled with a classification machine of the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer*. Analyst, 2013. 138(14), 3917-3926

Chapter 5

INVESTIGATING ATR-FTIR SPECTROSCOPY TO DISCRIMINATE BRAIN TUMOUR SEVERITIES FROM WHOLE AND FILTRATE SERUM SAMPLES

5.1 Introduction and Aims

In the previous chapter, ATR-FTIR was used to differentiate between cancer and non-cancer state using whole and 10 kDa filtrate serum samples. This aims of this chapter include the investigation of ATR-FTIR spectroscopy to discriminate between cancer and non-cancer, and to also discriminate between brain tumour severities from whole (unfiltered) serum and 100 kDa, 10 kDa and 3 kDa serum filtrate aliquots. Additionally, this chapter describes a variance study which was performed to assess the reproducibility of the spectra acquired from human serum.

The purpose of conducting a tiered serum filtrate study was to discover the most discriminatory fraction of proteins present in serum for disease discrimination based upon the identified discriminatory protein regions found in the previous chapter. The cytokines and angiogenesis factors found to be responsible for discrimination between GBM and non-cancer in chapter 4 have various molecular weights, thus this chapter investigates whether removing biomolecular constituents of human serum via filtration will enhance the spectral response of smaller biomolecules to allow for improved disease discrimination.

The work presented within this chapter has been published in the Journal of Biophotonics, 2014, Volume 7(3-4); 189-199. *Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples.*

5.2 Experimental Details

5.2.1 Variance Study

To assess the reproducibility and variance of spectra acquired from human serum a variance study was conducted. The aim of the variance study was to use standard deviation (STD) values to represent the consistency of spectra collected from human serum.

5.2.1.1 Materials and Methods of Variance Study

Normal (non-diseased) human mixed pooled serum (0.2 µl sterile filtered, CS100-100, purchased from TSCBiosciences UK) was used in this study. A 1 µl volume of serum was pipetted onto the ATR-FTIR single reflection diamond and dried for 8 minutes prior to spectral collection. All 1 µl volumes of serum were dried for 8 minutes in accordance to the drying study described in 4.3 of chapter 4. All spectra were collected using a JASCO FTIR-410 spectrometer fitted with a Specac ATR diamond Golden-Gate™ in the range of 4000-400 cm^{-1}, at a resolution of 4 cm^{-1} and over 32 co-added scans. Per 1 µl volume of serum, 3 spectra were collected. In total, 50 different 1 µl volumes of missed pooled serum were analysed with a background collected in between each and the dried serum films removed from the

diamond with absolute ethanol. In total, 150 ATR-FTIR spectra were collected from 50 x 1 µl spots of the mixed pooled serum. Mixed pooled serum is purchased from a biofluid supplier company where the blood serum has been extracted from blood collected from blood donors.

5.2.2 Addition of Low-Grade Glioma Patients

The addition of these low-grade glioma patients to the patient dataset is a development as it allows for the investigation into grading brain cancer severities and the determination of the most discriminatory fraction of serum for doing so via ATR-FTIR spectroscopy.

5.2.2.1 Materials and Methods - Low-Grade Serum

In addition to the cancer (GBM) and non-cancer human serum samples used in chapter 4 (Appendix 1), this chapter includes 23 low-grade patients to allow for the discrimination of brain tumour severities to be investigated. All serum samples were collected from the Walton Centre Research Tissue Bank (Liverpool, England) and Brain Tumour North West (BTNW) Tissue Bank (Preston, England) with full ethical approval (BTNW and Walton Centre ethical application #1108). As stated in chapter 4, all 23 low-grade patient blood samples were collected pre-operatively and left to clot at room temperature for 30-120 minutes from blood draw. Centrifugation of blood samples at 1200x*g* for 10 minutes was performed to separate the clot, followed by the serum being aliquoted into cryovials. All samples

ble 5.1 - Patient demographic data - each patient's age, gender (F: female, M: ıle) and tumour diagnosis is presented

were snap frozen using liquid nitrogen and were stored at a temperature of -80°C.

Table 5.2 shows patient demographic data of the 23 low-grade patients.

Patient Number	Gender	Age at Sample Collection	Diagnosis
11006	F	34	Astrocytoma
11017	M	41	Astrocytoma
11020	M	50	Astrocytoma
11084	F	31	Astrocytoma
12118	M	20	Astrocytoma
12001	F	60	Astrocytoma
12156	F	44	Astrocytoma
12164	F	20	Astrocytoma
12318	F	23	Astrocytoma
12100	M	37	Oligoastrocytoma
12060	F	53	Oligoastrocytoma
12329	F	28	Oligoastrocytoma
11109	M	46	Oligodendroglioma
12103	F	31	Oligodendroglioma
12121	M	59	Oligodendroglioma
12136	M	41	Oligodendroglioma
12324	M	43	Oligodendroglioma
1043	M	21	Ganglioglioma
1044	M	29	Infiltrating edge of glioma
1046	F	19	LGG diagnosis only
1047	F	54	Ependymoma
1048	F	35	LGG diagnosis only
1049	M	29	LGG diagnosis only

The average patient age of the entire sample set (GBM, low-grade and non-cancer samples) is 54.62 years. Table 5.3 provides demographic details by cancer group.

Table 5.2 - Number, age and gender data of patient samples from each tumour grade.

Tumour Grade	Number of Subjects	Age range/mean age	Gender
Normal (Non-cancer)	25	26-87/59.1 years	15 male, 10 female
Low-Grade	23	19-60.3/36.9 years	11 male, 12 female
High-Grade	49	24.7-78.8/60.1 years	29 male, 20 female

5.2.3 Serum Filtrate Preparation

All whole serum samples were thawed prior to spectral collection and 100 kDa, 10 kDa and 3 kDa filtration aliquots were prepared using Amicon Ultra-0.5 mL centrifugal filters (purchased from Millipore Limited, UK) - as described in 1.2 of chapter 4. Centrifugal filters filter out components of the serum above the cut-off point of the filters membrane (i.e. 100 kDa), allowing components below the filter membrane cut off point to pass through. Each whole serum sample (high-grade, low-grade and control) had a filtration aliquot prepared by pipetting 0.5 mL of the whole serum in to the filtration device and centrifuging at 14,000 rpm for; 10 minutes, 15 minutes, and 30 minutes for 100 kDa, 10 kDa and 3 kDa filter devices

respectively. Spectra were collected in a random order within the serum sample sets. For each sample, a 1 μl serum spot was dried for 8 minutes on the ATR-FTIR crystal, at which time 3 spectra were collected. This procedure was repeated three times per sample. As a result, for each sample 9 spectra were collected.

Prior to spectral collection, a background absorption spectrum was collected (for atmospheric correction) before the 1 μl was pipetted onto the ATR-FTIR crystal, thus a background was collected per serum replicate. The dried serum film was washed off the crystal in between each procedure using Virkon® disinfectant (purchased from Antec Int., Suffolk, UK) and absolute ethanol.

Spectra were acquired in the range of 4000–400 cm^{-1}, at a resolution of 4 cm^{-1} and averaged over 32 co-added scans. In total, 3375 ATR-FTIR spectra were collected from all whole and filtration serum samples. Table 5.4 shows the total number of spectra and patients in each serum grade and filtration category. The number of patients reduces as serum filtrate aliquots are prepared due to serum availability.

Table 5.3 - The number of spectra collected and number of patients (in brackets) for each filtrate composition for the range of cancer serum severities being analysed

	Whole Serum	**100 kDa Serum**	**10 kDa Serum**	**3 kDa Serum**
High-Grade Serum	441 (49)	423 (47)	423 (47)	405 (45)
Low-Grade Serum	207 (23)	207 (23)	198 (22)	198 (22)
Normal (Non-cancer) Serum	225 (25)	225 (25)	225 (25)	198 (22)
Total	873 (97)	855 (95)	846 (94)	801 (89)

The 100 kDa serum filtrate preparation will contain all of the biomolecules which have a molecular weight below 100 kDa, allowing for the analysis of serum which has had the abundant albumin molecule removed. Likewise, the 10 kDa and 3 kDa serum filtrate preparations contain only those biomolecules which have a molecular weight less than the cut-off range of the filter. As previously discussed in 4.7 of Chapter 4, the cytokines and angiogenesis factors deemed to be significant when distinguishing between non-cancer and GBM blood serum were interleukin-8, angiopoietin, leptin, PECAM-1, PDGF, HGF and follistatin. For example, follistatin with a molecular weight of 35-70 kDa, angiopoietin (55 kDa) and leptin (16 kDa) will be present in the 100 kDa preparation (*along with smaller molecular weighted biomolecules); the 10 kDa filtrate preparation will contain interleukin-8 (8.4 kDa)*, and the 3 kDa preparation heparin fragments associated with carcinogenesis*. The presence of these previously identified significant cytokines and angiogenesis factors via Bioplex immunoassay gives promise to the prepared serum filtrate samples holding biomolecules which may allow for successful discrimination between disease state and tumour severity.

5.3 Results of Variance Study

Figure 5.1 (A-B) shows ATR-FTIR raw and pre-processed spectra in the fingerprint region of 1800–1000 cm^{-1}. The 2 spectra display an average spectrum (coloured error margin (coloured black).

vector normalised with error margin (STD) in the region of 1800–900cm^{-1}.

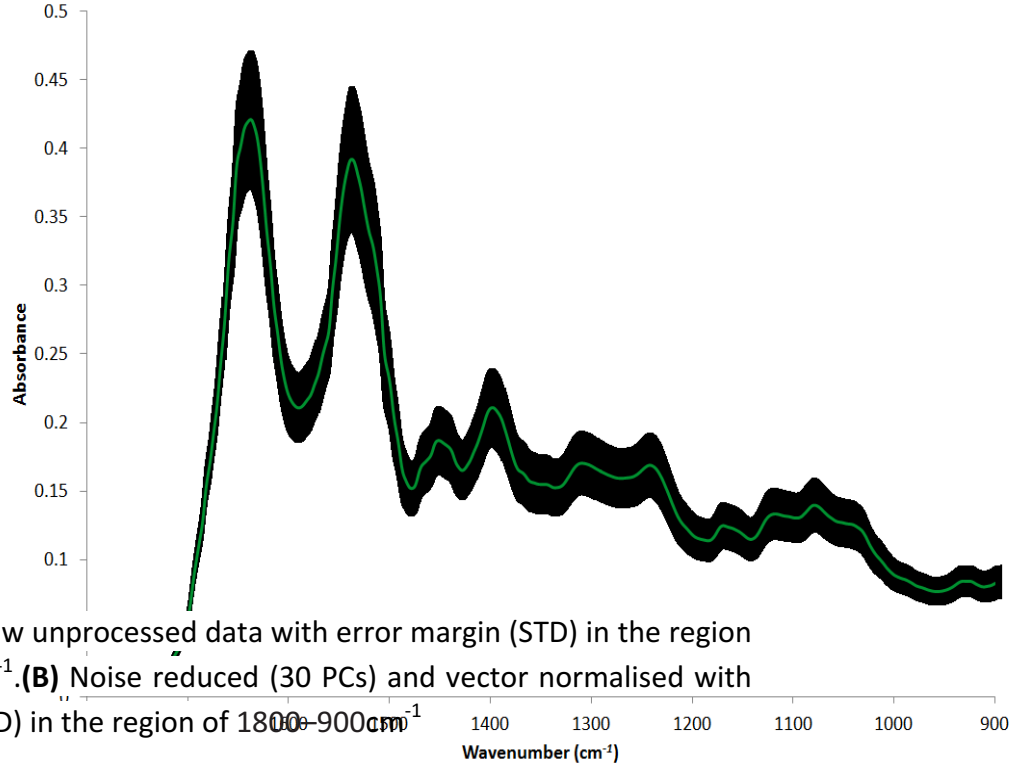
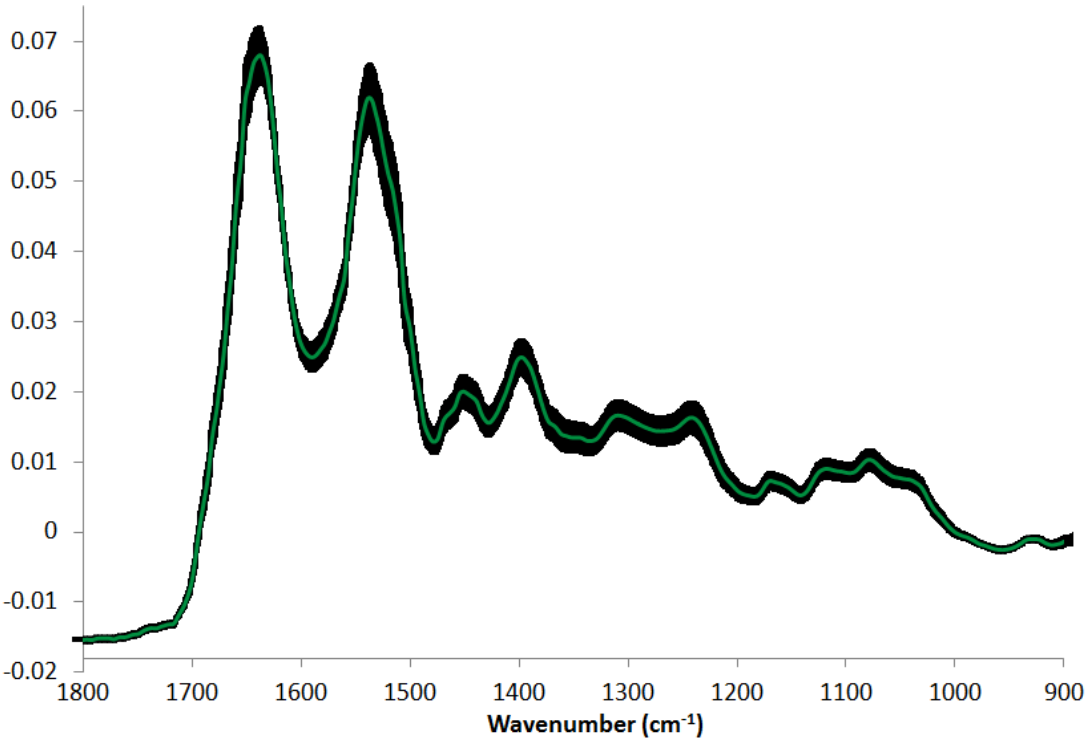


Figure5.1 - (A) Raw unprocessed data with error margin (STD) in the region of 1800–900 cm^{-1}.**(B)** Noise reduced (30 PCs) and vector normalised with error margin (STD) in the region of 1800–900cm^{-1}



ble 5.4 - The smallest and largest variance STDs related to Figure 5.2 (A), before d after pre-processing methods are applied to the data.

The largest variance between 3900-900 cm^{-1}in the raw (unprocessed) spectral data was in the fingerprint region at 1637.27 cm^{-1} (RSD: 11.94 %) (Figure 5.1(A)). The smallest variance in the raw data between 3900-900 cm^{-1} was at 3735.44 cm^{-1} and in the fingerprint region at 1792.51cm^{-1} (RSD: 15.19 %) (Figure 5.1(B)). Noise reduction (30 principal components) and vector normalization pre-processing methods were applied to the data to reduce the baseline and to smooth the data. The pre-processing methods significantly reduced the RSD and variance of the spectral data. The largest raw data RSD at 1637.27 cm^{-1} was reduced from 11.94 % to 2.35 % (pre-processed). The smallest spectral variance RSDs were reduced from 15.19 % to 6.26 % at 3735.44 cm^{-1} and from 11.94 % to 2.35 % at 1792.51 cm^{-1}. Table 5.1 displays the smallest and largest variances related to Figure 5.1 (A-B) before and after pre-processing methods were applied to the data.

	Smallest Variance Wavenumber (cm^{-1})/(RSD %)	**Largest Variance Wavenumber (cm^{-1})/(RSD %)**
Raw Data	1792.51 cm^{-1} (15.19)*(STD:0.0138)*	1637.27 cm^{-1} (11.94)*(STD: 0.042)*
After Pre-processing	1792.51 cm^{-1} (6.26)*(STD: 0.0004)*	1637.27 cm^{-1} (2.35)*(STD: 0.0043)*

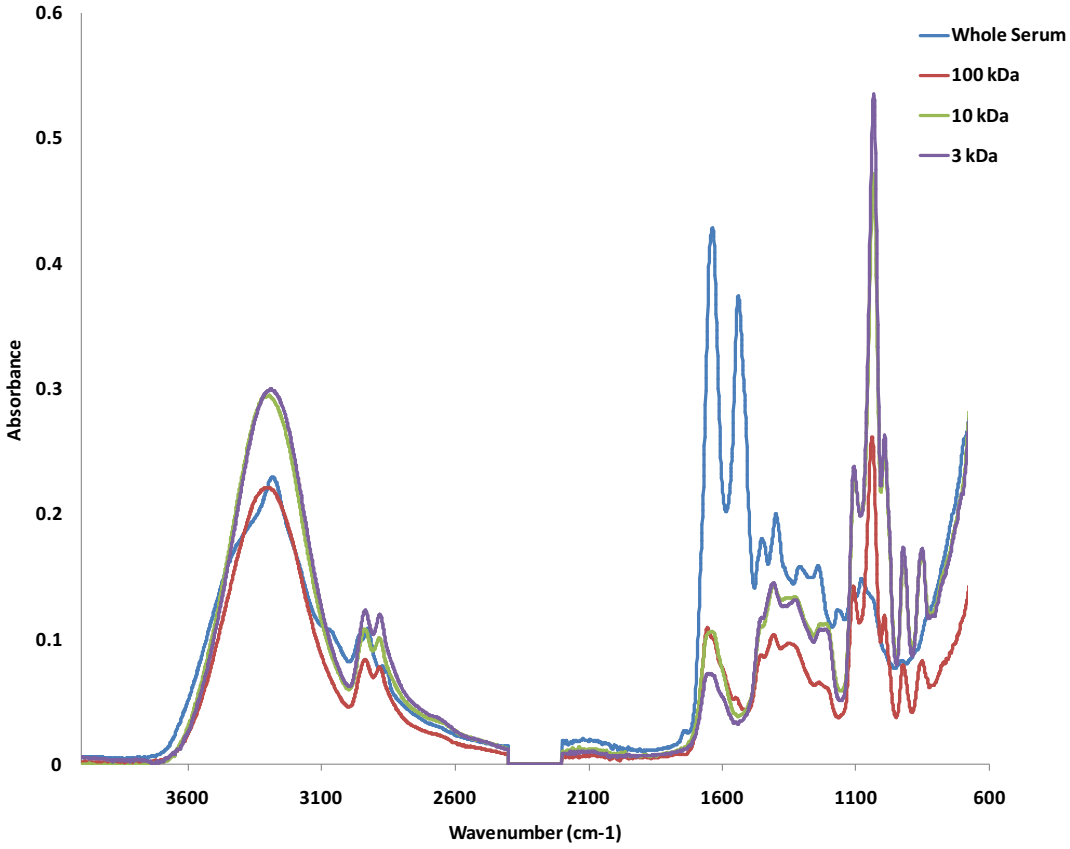
The RSD values of the raw spectra were low initially but were reduced further by implementing pre-processing methods. The reproducibility of spectral

igure 5.2 - Average spectra from whole serum (blue), 100 kDa (red), 10 kDa (green) nd 3 kDa (purple) serum filtrate samples. The fingerprint region varies from whole nd filtrate serum preparations due to the removal of biomolecules via filtration. he largest fingerprint absorbances derive from whole serum.

data using ATR-FTIR is high and exhibits minimal variance, especially after pre-processing.

5.4 Spectra Acquired from Serum Filtrate Samples

Following on from the discussed serum filtrate sample preparation in 5.2.3, spectra typical of each serum filtrate were plotted and are presented in Figure 5.2.



The spectra which are typical of each serum filtrate group are vastly different from one another due to the filtration step of whole serum. The region of 3600-600 cm^{-1} is different for each filtrate sample; however, it is the fingerprint region (1800-1000 cm^{-1}) that is of interest during MVA within this chapter. Following

the drying of the 1 µl volume of serum on the ATR crystal, the Amide I and Amide II bands between 1650-1450 cm^{-1} have strong absorbances, however, Figure 5.2 shows that the absorbance of the Amide bands is drastically reduced with all filtration samples, due to the filtration process removing biomolecules which contribute to these absorption peaks. A major difference between the spectra is at approx. 1025 cm^{-1}. The whole serum spectra does not exhibit a large and intense peak, tentatively assigned as contributions from carbohydrates (e.g. glucose), although as the blood serum is further filtered, these peaks become more intense. The 3 kDa spectrum displays the greatest absorbance of this peak, possibly due to blood serum containing approximately 4000 species of small proteins and peptides under 2.5 kDa [14]. Separating whole serum into sub-fractions reduces spectral complexity (Figure 5.2). Spectra collected from each of the filtration groups have the potential to increase the sensitivity of detection smaller and lower abundant molecules which may have otherwise been masked/lost due to absorbance peaks from higher mass proteins. Spectra from 100 kDa and 10 kDa shows Amide I and Amide II protein peaks suggestive that these samples contain concentrations of blood-related glycoproteins following filtration from whole serum. The absorbance bands of the Amide I and Amide II on the 3 kDa serum filtrate spectrum are weaker in comparison, suggestive that the filtration removed the proteins contributing to these bands.

5.5 ATR-FTIR Spectral Diagnostic Model

5.5.1 Pre-processing Selection Data

For each whole and filtration serum sample set, an identical approach was used to pre-process the spectral data, and to analyse using multivariate analysis methods. Firstly, to remove any bias from analysis models, the technical replicates from each sample were averaged so that each serum sample set contained three spectra from each patient; one average spectrum from each patient spot.

A visual quality test was performed to check for gross spectral error. To do this, the maximum absorbance point (~3290 cm^{-1}) of the O-H region was observed to have absorbance units between 0.17-0.23; Amide I and II to have absorbance units between 0.35-0.45 and the spectrum to have a smooth baseline with little or no noise contribution. The O-H and the Amide I and II absorbance's were selected for visual quality testing as they were consistently present on spectra of whole and filtrate serum samples. After 8 minutes when the narrowing point of the O-H band at ~3290 cm^{-1} and the absorbances of the Amide bands remain at a constant absorbance, the serum film had dried on the ATR crystal.

The fingerprint region (1800–1000 cm^{-1}) was selected for MVA. A PC based noise reduction, using the first 30 PCs of the data, was performed on the spectra to improve the signal-to-noise ratio. Following this, all spectra were vector normalised. PCA was performed on the pre-processed spectra, giving an unsupervised classification from which the loadings could be interpreted by showing the $\bar{\nu}$ regions which contribute to the separation.

Using LIBSVM code in Matlab, an automatic *n*-fold cross validation was performed (where *n*=3) on the training data to find the best values for the cost and gamma functions.

These values were then used to train the SVM in one-versus-rest mode using a randomly selected training set consisting of 2/3 of the patient-associated spectral data. The remainder of the data, making up the blind test set, was then projected into the model where an overall SVM classification accuracy (based on the true and predicted data class labels) was determined. The spectral data was split independently at a patient level into 1/3 blind and 2/3 training where all spectra from one patient was either in the train set or blind set.

The whole serum dataset had 3 various patient associated splits of blind and train data which were used for SVM analysis (*n*=3); the 100, 10 and 3 kDa serum filtrate data was split randomly into 2/3 train and 1/3 test and used for SVM (*n*=1).

Sensitivities and specificities were calculated for each SVM model and for each separate disease group. All spectra in the three different test and blind datasets were randomly assigned at a patient level, thus all 3 averaged patient spectra were either in the test or train. Pre-processing and MVA was carried out on the raw spectral data in Matlab using in-house written software.

5.6 Results of ATR-FTIR Spectral Diagnostic Model

5.6.1 Principal Component Analysis and Loadings

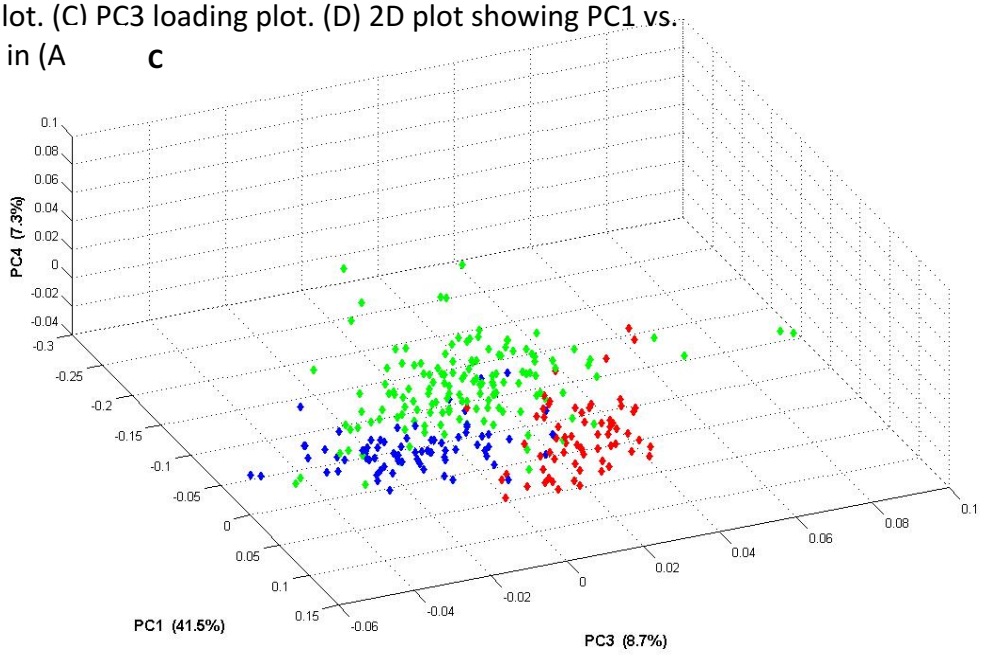
Figure 5.2 (A) shows a 3D PCA scatter plot displaying the separation between normal (non-cancer) (blue]), low-grade cancer (red) and high-grade cancer (green) patient spectral data collected from whole (unfiltered) serum. As discussed in chapter 4, the optimum sensitivities and specificities allowing for discrimination between disease state were achieved using whole serum samples, thus the spectral data acquired from whole serum in this chapter was investigated via PCA to obtain loadings to determine the biomolecular components contributing to the discrimination between brain tumour severities.

PC3 on the *x*-axis shows good separation between low-grade cancer and normal (non-cancer) data. PC2 on the *y*-axis shows good separation between high-grade cancer from low-grade and normal (non-cancer), however, there is still a significant overlap between patients. Towards the centre of Figure 5.4 (A) crossing both zero points on the *x*-axis and *y*-axis there are some patients who have been poorly classified. Section 4 of Chapter 4 discusses there being no separation in the non-cancer patient samples based on patient age and sex, thus the non-cancer group (blue) represents the non-cancerous group based on factors which include, but not limited to, biochemical components which differ from samples collected from diseased patients, or other factors such as BMI, medication use, lifestyle and activity etc. Low-grade patients with slow growing primary gliomas (WHO grade I and II) and tumours which are not rapidly infiltrating may be grouped closer to the

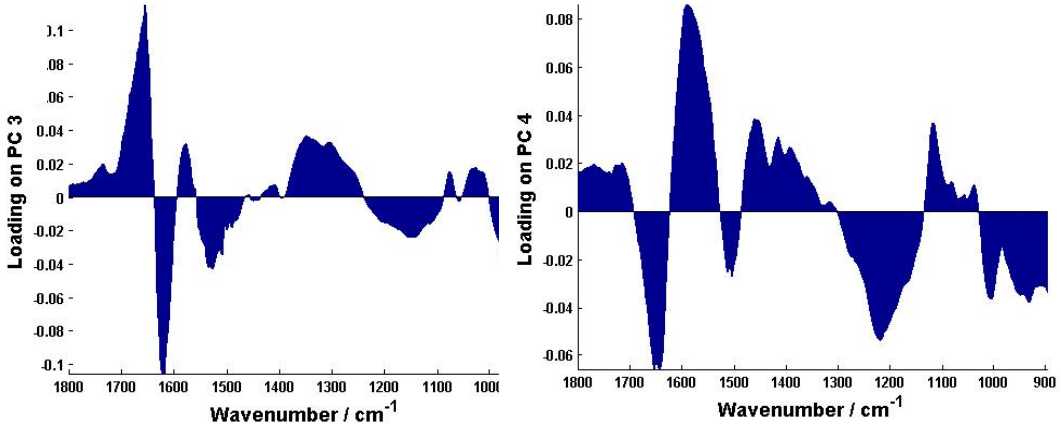
non-cancer group due to their biochemical profile being closer to that of a person without a disease, rather than a person with a high-grade and highly infiltrative brain tumour. Additionally, a low-grade patient (WHO grade II) who has an increasingly infiltrative and aggressive tumour could potentially be grouped with high-grade (WHO grade III and IV) patients. Some non-cancer patients were misclassified, via PCA (Figure 5.3), in the low and high-grade classes, possibly suggesting that the non-cancer patients have cancer or a disease. Non-cancer patient data was not available from the tissue bank, thus it was not possible to identify whether any patients had illnesses (non-cancer related) that may alter the biochemical composition of their blood sample. As was the case with all blood serum samples in this thesis, crucial patient information was unavailable from the tissue bank. It was not possible to link patient medication use, BMI etc. to their acquired blood serum spectra. Further analysis of this data involves supervised machine learning algorithms which allow for blind data to be projected into pre-defined spectral data classes. Spectral data collected from the three serum grades in this study have been shown to successfully separate into their different spectral groups using PCA. Figure 5.3 (B-C) are the loadings of PC1 and PC3. (D) shows a 2D plot of PC1 vs. PC3, as in the data presented in (A).

Figure 5.3 - (A) Three dimensional PCA scatter plot of normal (non-cancer) [blue], low-grade cancer [red] and high-grade cancer [green] from patient whole serum samples. (B) PC1 loading plot. (C) PC3 loading plot. (D) 2D plot showing PC1 vs. PC3 of the data presented in (A

C



D



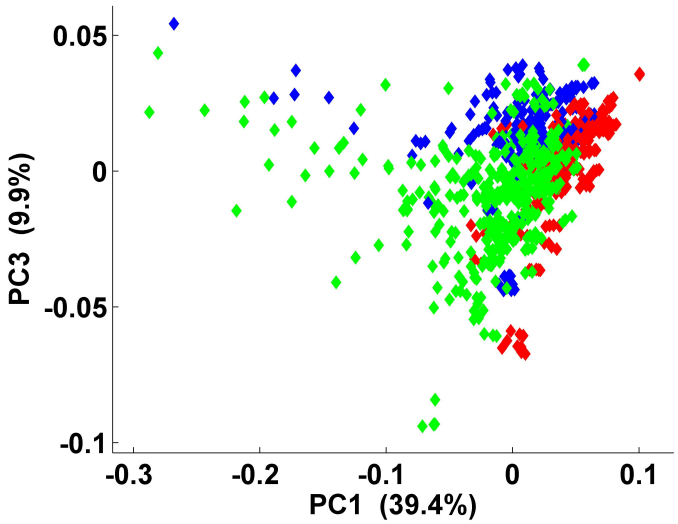


Table 5.5 shows the major spectral peaks and proposed biomolecular assignments responsible for PC1 loadings and Table 5.6 for PC3 loadings (1-4).

Table 5.5 - The major peaks and the proposed biomolecular assignments of PC1 loading (Figure 5.3(B)).

Direction Loading on PC 1 (B)	Wavenumber (cm^{-1})	Proposed Biomolecular Assignment
+ve	1698	Amide I
+ve	1513	Amide II
+ve	1445	CH_2 deformation of methylene group, lipids
+ve	1386	CH_3 deformation, lipids
-ve	1649	Amide I
-ve	1619	Amide I
-ve	1553	Amide II
-ve	1077	C-O stretch, deoxyribose/ribose, DNA, RNA

Table 5.6- The major peaks and the proposed biomolecular assignments of PC 3 loading (Figure 5.3(C)).

Direction of Loading on PC 3 (C)	Wavenumber (cm^{-1})	Proposed Biomolecular Assignment
+ve	1655	Amide I
+ve	1589	Amide II
+ve	1462	CH_2 deformation of methylene group (lipids)
+ve	1416	CH_3 asymmetric stretch (lipids)
+ve	1394	CH_3 deformation, lipids
+ve	1361	CH_3 (symmetric)
+ve	1119	C–O (antisymmetric), COH bend, lipids
+ve	1040	C–O, deoxyribose/ribose DNA, RNA
-ve	1644	Amide I
-ve	1513	C=C
-ve	1505	C=C
-ve	1220	C–C, C–H bend
-ve	1006	Phenylalanine (ring breathing)
-ve	935	C–C residue α-helix

The axes of the loading plots have positive and negative directions. The positive and negative peaks in the loading plots of PC1 and PC3 correspond to the spectral peaks that the PCA scatter plot is using to discriminate the groups within the patient spectral dataset. The major peaks of PC1 (Figure 5.2(B)) are –ve 1649 cm^{-1} which is assigned as Amide I, +ve 1513 cm^{-1} assigned as Amide II and –ve 1077 cm^{-1} assigned as C–O stretch (DNA/RNA), these positive and negative peaks separate the low-grade cancer (red) and non-cancer (blue) from the high-grade (green) data points.

The major peaks of PC3 (Figure 5.3 (C)) are +ve 1655 cm^{-1} and -ve 1644 cm^{-1} both which are assigned Amide I and +ve 1589 cm^{-1} assigned as Amide II. The major +ve and -ve peaks separate the low-grade from the non-cancer PCA data points and -ve 1644 cm^{-1} (Amide I). Table 5.5 and 5.6 show the spectral assignments for PC1 and PC3 loadings respectively (Figure 5.2 (B-C)). Table 5 shows the proposed biomolecular assignments from the loadings of PC3.

5.6.2 ATR-FTIR Diagnostic Model SVM Results

ATR-FTIR spectra from all whole serum and serum filtrate aliquot samples were analysed to investigate sensitivities and specificities possible on patient and spectral levels using RBF-SVM analysis. Three different train and test spectral datasets were used to provide a range of sensitivities and specificities for whole serum (Table 5.7).

Table 5.7- Dataset 1 (*n* = 3) results presented with the ranges from datasets 2 and 3. Sensitivities and specificities for whole serum on patient and spectral levels for normal (non-cancer), low-grade and high-grade cancer.

	Normal (%)	**Normal Range (%)**	**Low (%)**	**Low Range (%)**	**High (%)**	**High Range (%)**	**Overall Average (%)**	**Overall Range (%)**
Patient Sensitivity	100	75.00 – 100	87.50	87.50 – 87.50	93.75	92.86 – 93.75	**93.75**	**75.00 – 100**
Patient Specificity	95.83	95.40 – 100	100	95.45 – 100	93.75	87.50 – 93.75	**96.53**	**87.50 – 100**
Spectra Sensitivity	95.83	78.26 – 95.83	86.36	85.00 – 91.67	95.65	92.86 – 95.65	**92.61**	**78.26 – 95.83**
Spectra Specificity	97.06	95.45 – 100	100	95.45 – 100	91.30	86.36 – 91.30	**96.12**	**86.36 – 100**

One train and test spectral dataset was used to achieve sensitivities and specificities for the filtrate aliquots. The whole serum SVM diagnostic model achieved 93.75 % sensitivity and 96.53 % specificity (overall average) on a patient level. The whole serum dataset also achieved 92.61 % sensitivity and 96.12 % specificity on a spectral level. The best whole serum RBF-SVM diagnostic model misclassified three patients in the blind dataset; one low-grade patient and two high-grade patients. The blind dataset consisted of one average spectrum per 1 µL of patient serum analysed resulting in three spectra per patient in the dataset. If two or more spectra from the three averaged for each patient were classified as either high-grade, low-grade or non-cancer, then this is regarded as the diagnosis.

In comparison to the whole serum SVM-RBF results, the 100 kDa (Table 5.8), 10 kDa (Table 5.9) and 3 kDa (Table 5.10) serum filtrate aliquots did not achieve sensitivities or specificities as high. The 100 kDa diagnostic model achieved 69.05% sensitivity and 85.87 % specificity on a patient level and the 10 kDa diagnostic model achieved a sensitivity and specificity on a patient level with 79.68 % and 88.75 % respectively.

Table 5.8- Sensitivities and specificities for 100 kDa serum filtrate samples on patient and spectral levels for normal (non-cancer), low-grade and high-grade cancer with an overall average.

	Normal (%)	**Low-Grade (%)**	**High-Grade (%)**	**Overall Average (%)**
Patient sensitivity	50	57.14	100	**69.05**
Patient specificity	95.45	95.45	66.7	**85.87**
Spectra sensitivity	54.17	61.90	93.75	**69.94**
Spectra specificity	94.12	94.12	67.44	**85.39**

Table 5.9- Sensitivities and specificities for 10 kDa serum filtrate samples on patient and spectral levels for normal (non-cancer), low-grade and high-grade cancer with an overall average.

	Normal (%)	**Low-Grade (%)**	**High-Grade (%)**	**Overall Average (%)**
Patient sensitivity	85.71	70.00	83.33	**79.68**
Patient specificity	85.00	100	81.25	**88.75**
Spectra sensitivity	75.00	66.67	78.38	**73.35**
Spectra specificity	80.33	98.25	78.72	**85.77**

Table 5.10 - Sensitivities and specificities for 3 kDa serum filtrate samples on patient and spectral levels for normal (non-cancer), low-grade and high-grade cancer with an overall average.

	Normal (%)	Low-Grade (%)	High-Grade (%)	Overall Average (%)
Patient sensitivity	62.50	66.67	66.67	**65.28**
Patient specificity	73.68	100	68.67	**80.81**
Spectra sensitivity	70.83	69.23	68.57	**69.54**
Spectra specificity	76.36	98.04	74.47	**82.96**

The 3 kDa dataset achieved the lowest sensitivities and specificities and on a patient level achieved 65.28 and 80.81 % respectively (Table 5.9).

The results presented in tables 5.7-5.10 show that whole serum provides the greatest sensitivities and specificities compared to the serum filtrate preparations when distinguishing between cancer *vs*. non-cancer, as well as brain tumour severities. Whole serum contains a plethora of biochemical molecules which allow for optimum disease classifications (Table 5.7), thus the combination of the cytokines and angiogenesis factors, such as follistatin, interleukin-8, leptin and angiopoietin, to name but a few, is favourable, especially with the filtration step not being required. The 10 and 3 kDa filtrate preparations do not provide as high sensitivities and specificities, this suggests that the biomolecules involved with the discrimination between the various grades of brain tumour are not as discriminatory compared to larger molecular weighted biomolecules (e.g. 100 kDa filtrate preparations).

Table 5.11 shows the optimum cost and gamma values for whole serum and serum filtrate sample SVM analysis where training accuracy is the correctly predicted spectra from the training set and total accuracy is the correctly predicted spectra from the blind set. Three different train and test spectral datasets were used to provide a range of sensitivities and specificities for whole serum (*n*=3). One train and test spectral dataset was used to achieve sensitivities and specificities for the filtrate aliquots (*n*=1).

Table 5.11 - The optimum cost and gamma values for the whole serum and serum filtrate aliquot samples.

	Whole Serum	Whole Serum (2)	Whole Serum (3)	100 kDa Data	10 kDa Data	3 kDa Data
Optimal Cost (C)	22.63	32	22.63	2048	2048	2048
Optimal Gamma (γ)	4	5.66	8	0.85	16	16
Training Accuracy (%)	85.86	87.96	86.46	72.58	90.80	78.53
SVM Total Accuracy (%)	96.88	86.46	91.58	79.57	79.76	75.64

Whole serum contains a range of biochemical components such as cytokines and angiogenesis factors which are redundant proteins that are secreted with growth and are involved in regulating and determining the nature of an immune response, such as with the development of cancer (5). Chapter 4 of this thesis demonstrates how cytokines and angiogenesis factors differ between serum from glioma and non-cancer patients.

The whole serum RBF-SVM diagnostic model achieved the greatest patient and spectral sensitivities and specificities as a result of the IR spectral dataset

exhibiting bands characteristic of all serum components such as cytokines and angiogenesis factors which have varied molecular weights (6). Follistatin has a molecular weight of 35-70 kDa (7); interleukin-8 a molecular weight of 8.4 kDa (8); angiopoietin 55 kDa (9); leptin 16 kDa (10) and PECAM-1 130 kDa (11). The cytokines and angiogenesis factors deemed to be significant when distinguishing between non-cancer and GBM blood serum (discussed previously in Chapter 4) were interleukin-8, angiopoietin, leptin, PECAM-1, PDGF, HGF and follistatin.

Serum filtrate aliquots produced from whole serum samples achieved lower sensitivities and specificities as a result of filtering and removing serum biomolecular components above the filter cut-off range.

Petrich *et al*. reported the significant differences in IR spectra of dried serum before and after filtration. It was revealed that whole serum has more spectral regions which allows for the discrimination between serum samples from patients with acute myocardial infarction (AMI) and serum samples from non-AMI patients compared to 100 kDa and 10 kDa serum filtrates. The molecules responsible for the significant differences have a molecular mass greater than 100 kDa and 10 kDa due to decreasing sensitivities and specificities with greater sample filtration, thus whole serum spectra has more areas of significance when differentiating disease states (12).

5.7 Discussion

The results presented in this chapter show that the combination of SVM analysis and ATR-FTIR spectroscopic profiles of the molecular vibrations within 1

μlserum films on an ATR-FTIR crystal can be used to diagnose and differentiate between high and low-grade brain tumours from non-cancer with high sensitivities and specificities.

The raw spectra collected from 1 μl serum spots after 8 minutes of drying have a low standard deviation which is further reduced with noise reduction and vector normalisation pre-processing methods. The variance between the spectral data after pre-processing is as little as 0.0004 at 1792.51cm^{-1}. The reproducibility of spectral data using ATR-FTIR is high and exhibits minimal variance especially after pre-processing.

This chapter has demonstrated that the combination of mid-IR spectral data collected from whole serum samples from high-grade, low-grade and normal (non-cancer) patients and RBF-SVM analysis has the ability to be implemented within the clinical environment as a rapid, reagent-free and cost-effective diagnostic regime. The analysed serum filtrate aliquots in this chapter (100, 10 and 3 kDa) have been shown to not achieve as high sensitivities and specificities as whole serum; filtration removes key biomolecules which exhibit spectral bands that are involved in the MVA diagnostic process. The loadings from PC1 and PC3 (Figure 5.3(B-C)) show the peaks which are responsible for the separation between high-grade, low-grade and non-cancer spectral datasets. The major peaks from PC1 show that Amide I, Amide II and C–O stretch (DNA/RNA) bands are responsible for the separation between non-cancer and low-grade cancer from high-grade cancer PCA data points. The major peaks from PC2 shows that Amide I and Amide II are responsible for the separation of low-grade cancer from non-cancer PCA data points.

Serum filtrate aliquot samples in this chapter did not achieve as high sensitivities and specificities in comparison to whole serum, thus the filtration serum preparation step can be avoided allowing for an ever faster spectroscopic diagnostic process. Bonnier *et al*. found that careful washing protocols needs to be implemented prior to using the commercially available filters, as used to prepare the serum filtrate samples in this chapter. Contaminants deriving from glycerine were found to be present in the IR spectra of serum prior to the careful washing of the filter before use (13). ATR-FTIR spectroscopy is useful for diagnostic applications, however, unless a multi-ATR crystal device is created for high-throughput analysis then the practicality of collecting spectral data from each sample one-by-one becomes a concerning time-consuming issue.

The method described within this chapter has the ability to diagnose gliomas (high-grade and low-grade versus non-cancer serum) from whole patient serum ATR-FTIR spectral data to sensitivities and specificities as high as 100 and 95.83 % respectively.

References

[1] Baker M J., Clarke C., Demoulin D., Nicholson J M., Lyng F M., Byrne H J., Hart C A., Brown M D., Clarke N W., Gardner P. *An investigation of the RWPE prostate derived family of cell lines using FTIR spectroscopy*. Analyst, 2010. 135, 887-894

[2] Lyng F M., Faolain E O., Conroy J., Meade A D., Knief P., Duffy B., Hunter M B., Byrne J M., Kelehan P., Byrne H J. *Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool*. Experimental and Molecular Pathology, 2007. 82(2), 121-129

[3] Maziak D E., Do M T., Shamji F M., Sundaresan S R., Perkins D G., Wong P T. *Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: an exploratory study*. The International Journal of Cancer Epidemiology, Detection and Prevention, 2007. 31(3), 244-253

[4] Meurens M., Wallon J., Tong J., Noel H., Haot J. *Breast cancer detection by Fourier transform infrared spectrometry*. Vibrational Spectroscopy, 1996. 10(2), 341-346

[5] Borish L C., Steinke J W. *Cytokines and chemokines*. Journal of Allergy and Clinical Immunology, 2003. 111(2), S460-475

[6] Grinwald L R. *Chemokine research trends*. Nova Science Pub Inc., 2007

[7] Santa Cruz Biotechnology. *Follistatin (H-114): sc-30194* [Online] http://datasheets.scbt.com/sc-30194.pdf [Accessed on 10th July 2015]

[8] Sigma-Aldrich. *Interleukin-8 human* [Online] http://www.sigmaaldrich.com/catalog/product/sigma/i1645 [Accessed on 10th July 2015]

[9] abcam. *Anti-angiopoietin 1 antibody (ab8451)* [Online] http://www.abcam.com/angiopoietin-1-antibody-ab8451.html [Accessed on 11th July 2015]

[10] Liu C., Liu X J., Barry G., Ling N., Maki R A., De Souza E B. *Expression and characterisation of a putative high affinity human soluble leptin receptor*. Endocrinology, 1997. 138(8), 3548-3554

[11] Santa Cruz Biotechnology. *PECAM-1 (M-20): sc-1506* [Online] http://datasheets.scbt.com/sc-1506.pdf [Accessed on 10th July 2015]

[12] Petrich W., Lewandrowski K B., Muhlestein J B., Hammond M E., Januzzi J L., Lewandrowski E L., Pearson R R., Dolenko B., Fruh J., Haass M., Hirschl M M., Kohler W., Mischler R., Mocks J., Ordonez-Llanons J., Quarder O., Somorjai R., Staib A., Sylven C., Werner G., Zerback R. *Potential of mid-infrared spectroscopy to aid the triage of patients with acute chest pain*. Analyst, 2009. 134(6), 1092-1098

[13] Bonnier F., Petitjean F., Baker M J., Byrne H J. *Improved protocols for vibrational spectroscopic analysis of body fluids*. Journal of Biophotonics, 2014. 7(3-4)

[14] Merrell, K., Southwick, K., Graves, S. W., Esplin, M. S., Lewis, N. E., Thulin, C. D. *Analysis of low-abundance, low molecular-weight serum proteins using mass spectroscopy*. Journal of Biomolecular Techniques, 2004. 15(4), 238-248

Chapter 6

BRAIN TUMOUR DIFFERENTIATION: STRATIFIED SERUM DIAGNOSTICS VIA ATR-FTIR SPECTROSCOPY

6.1 Introduction and Aims

In the previous chapter, ATR-FTIR spectroscopy was demonstrated to differentiate between brain tumour severities (high-grade and low-grade glioma) using whole serum and serum filtrate aliquots with high sensitivities and specificities. In summary of the study to discover which fraction of serum was most discriminatory between brain tumour severities, it was found that whole serum provided greater sensitivities and specificities in comparison to serum filtrate aliquots.

The research in this chapter aims to expand on the previously conducted research to, for the first time, detect cancer *vs.* non-cancer, metastatic cancer *vs.* organ confined, brain cancer severity and the organ of origin of metastatic disease from the same serum sample enabling stratified diagnostics depending upon the clinical question asked. Furthermore, this chapter investigates feature extraction (FE) fed SVM analysis to maximize classification accuracy based on the most discriminatory features of the projects spectral dataset.The work presented within this chapter has been submitted to Neuro-oncology. Manuscript title: *Brain Tumour Differentiation: Rapid Stratified Serum Diagnostics via Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy*.

able 6.1 - Total subject number of tumour grade, age range, mean age and ender of patient samples

6.2 Experimental Details

Blood samples were collected from a total of 433 patients. Table 6.1 provides demographic information based on all patient groups. All serum samples were collected from the Walton Centre Research Tissue Bank (Liverpool, England) and BTNW Tissue Bank (Preston, England) with full ethical approval (BTNW and Walton Centre ethical application #1108).

Tumour Grade	Number of Subjects	Age Range/ Mean Age	Gender
Non-cancer	122	16-89 44.77 years	64 male 58 female
All Cancer	311	19-82 57.77 years	133 male 178 female
Glioma	87	19-81 49.90 years	52 male 35 female
Low-Grade Glioma	23	19-60 38.35 years	11 male 12 female
High-Grade Glioma	64	25-81 61.44 years	41 male 23 female
Meningioma	47	24-78 55.98 years	13 male 34 female
Metastasis	177	25-82 59.45 years	68 male 109 female
Lung Metastasis	84	25-82 59.32 years	36 male 48 female
Breast Metastasis	36	27-76 50.92 years	0 male 36 female
Melanoma Metastasis	25	25-80 56.00 years	14 male 11 female

Non-cancer control serum samples were collected from individuals who presented no symptoms of cancer at Royal Preston Hospital NHS Trust, UK blood donation event as well as those presenting to the clinic for elective surgery. The average age is 57.77 and 44.77 years for the cancer and non-cancer patient sample sets, respectively. Removing all teenage patients (*n*=2; 1 non-cancer diagnosis, 1 low-grade diagnosis) increases the non-cancer patient sample set average age to 60.32 years and cancer to 57.89 years. Where possible each tumour grade was age and sex matched. The whole metastatic tumour subset has an average age of 59.45 years - this can further be split up into lung metastasis with an average age of 59.32; breast metastasis subset average age of 50.92 years and melanoma metastasis subset an average age of 56.00 years. Patient information, e.g. BMI, drug use, diet, lifestyle, etc. was unavailable from the tissue bank from which the blood serum samples were collected. Patients who have a blood sample collected pre-operatively are highly likely to be on an array of medication for cancer or have been treated with chemotherapy/radiotherapy, thus impacting upon the blood serum sample acquired. In contrast, a patient who donated a non-cancerous sample will not have the physiological changes induced by the intake of cancer medications and therapies. To summarise, cancer patients are likely to be on pain killers, steroids and have different regional physiological functions in comparison to those patients which have not been diagnosed with cancer. Appendix 2 shows expanded patient demographic information for all 433 individual patients.

6.2.1 Instrumentation

All spectra from all patient serum samples in Table 6.1 were collected using an Agilent Cary 600 Series FTIR spectrometer with a PIKE Technologies MIRacle^{TM} single reflection ATR configured with a diamond crystal plate. 1 µl volumes of human serum were pipetted onto the ATR-FTIR crystal using an Eppendorf Research Plus 0.5ul-10.0ul pipette. After spectral collection from each 1 µl dried serum spot, Virkon Surface Disinfectant (Fisher Scientific) and 99.5 % absolute ethanol (Thermo Scientific) were used consecutively to wash the serum film off the Di crystal.

6.3 ATR-FTIR Diagnostic Model

As was the case in chapters 4 and 5, all whole serum samples were thawed prior to spectral collection at room temperature for up to 30 minutes. Spectra were collected in a random order within the serum sample sets. For each sample, a 1 µl serum spot was pipetted onto the ATR-FTIR crystal and allowed to dry for 8 minutes, at which time 3 spectra were collected. Prior to spectral collection, a background absorption spectrum was collected (for atmospheric correction) before the 1 µl of serum was pipetted onto the ATR-FTIR crystal. A single background was collected per sample replicate. Spectra were acquired in the range of 4000 - 600 cm^{-1}, at a resolution of 4 cm^{-1} and averaged over 32 co-added scans. In total, 3897 ATR-FTIR spectra were collected from all serum samples. To the best of my knowledge, this is the largest study on mid-infrared spectroscopy in relation to cancer research with a 433 patient cohort.

All serum samples were collected from the Walton Centre Research Tissue Bank (Liverpool, England) and Brain Tumour North West (BTNW) Tissue Bank (Preston, England) with full ethical approval (BTNW and Walton Centre ethical application #1108). All patient blood samples were collected pre-operatively and left to clot at room temperature for 30-120 minutes from blood draw. Centrifugation of blood samples at 1200x*g* for 10 minutes was performed to separate the clot, followed by the serum being aliquoted into cryovials. All samples were snap frozen using liquid nitrogen and were stored at a temperature of -80°C. Table 5.2 shows patient demographic data of the 23 low-grade patients.

6.4 Data Handling and Analysis

Initially Agilent's Resolutions-Pro FTIR software was used for data handling after which the spectra were imported for further analysis and processing into Matlab™ using in-house written and open source protocols.

For all spectra acquired, the fingerprint region (1800-1000 cm^{-1}) was selected for MVA. A PCA noise reduction, using the first 50 PCs, of the data was performed on the spectra to improve the SNR ratio. Following noise reduction, all spectra were vector normalised. Using LIBSVM and in-house written protocols in MATLAB™, an *n*-fold CV was performed (*n*=5) on the training data to determine the optimum values for the cost and gamma functions. Table 6.2 shows the optimum cost and gamma functions for each stratum (e.g. cancer *vs.* non-cancer).

le 6.2 - Optimal cost and gamma values for each stratum with a mean cross
dation accuracy

	Cancer *vs.* Non-cancer	**Metastasis *vs.* Brain**	**Glioma *vs.* Meningioma**	**HGG *vs.* LGG**	**Lung *vs.* Skin *vs.* Breast**
Optimal Cost (C)	128	128	128	128	32
Optimal Gamma (γ)	128	128	128	128	128
Mean Cross-Validation Accuracy (%)	94.74	94.07	95.37	95.60	93.10

Optimum cost and gamma values were used to train the SVM in a one-versus-rest mode using a randomly selected training set consisting of 2/3 of the patient associated spectral data. The remainder of the data (1/3) was used to create the blind set which was then projected into the model, and confusion matrices were calculated giving an overall SVM classification based on the true and predicted data class labels. For each stratum, 525 combinations of 2/3 training and 1/3 blind were performed based upon patient membership, thus, all spectra from one patient was either in the train set or the blind set. Sensitivities and specificities were calculated for each combination in order to understand the effect of patient membership in training and blind sets based upon sensitivity and specificity. Sensitivities and specificities were calculated using equations 4.1 - 4.2 in Chapter 4.

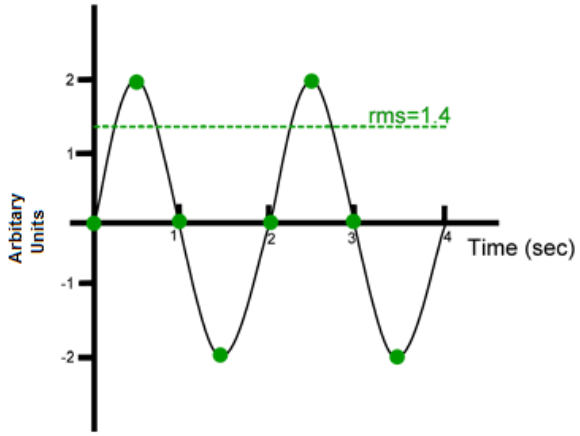
6.5 Feature Extraction (FE)

The main function of FE is to elucidate and rank the relevant discriminatory spectral information from recorded ATR-FTIR data. Per stratum, all pre-processed spectral data was variably ranked with information gain using the top 30 % of all variables. Variable ranking highlights the wavenumber variables that are most salient between the spectral classes. In the case of the cancer *vs.* non-cancer stratum, 130 wavenumber variables associated to 6 wavenumber regions were selected (Table 6.4). Following variable ranking, the 6 wavenumber regions were selected on a 2D plot of the mean spectrum, upon which FE was performed. FE was performed whereby spectral descriptors such as root-mean-square (RMS) energy, peak kurtosis, peak skew, peak centroid, as well as peak frequency and peak amplitude can be extracted from each user selected spectral band, thus the relevant spectral band shapes involved in the discrimination between classes are able to be captured.

The feature information is ranked and scored in descending order to describe how each feature of the model explains the difference between the groups of recorded spectral data. The most discriminatory features highlighted during feature extraction were then used for a feature based SVM (FE-SVM). Using LIBSVM and in house written protocols in MATLAB™, an *n*-fold cross validation was performed (*n*=5) on the cancer *vs.* non-cancer spectral training data to determine the optimum values for the cost and gamma functions; 128 was optimal for cost and gamma respectively. FE-SVM was performed using all 130 spectral features followed by the top 30 and top 2 features for the cancer *vs.* non-cancer data set.

Figure 6.1 - Example of root mean square (RMS) energy - calculated by square rooting the sum of the average of each peak intensity. Adapted from (1).

RMS energy is calculated by the square root of the average of all squared intensity points of a spectrum (Figure 6.1) (1).



Peak kurtosis is a measure of whether spectral data are peaked or flat in relation to the normal distribution, as illustrated in Figure 6.2 (2). Positive kurtosis spectra exhibit distinct peaks near to the mean and decline rapidly; in contrast, negative kurtosis spectra have flat tops near to the mean rather than a sharp peak (3).

Figure 6.3 - Example of peak skewness. Normally distributed data (symmetrically distributed) has a skewness of zero, whereas the value of a peaks skew indicates which direction it is in. Adapted from (4).

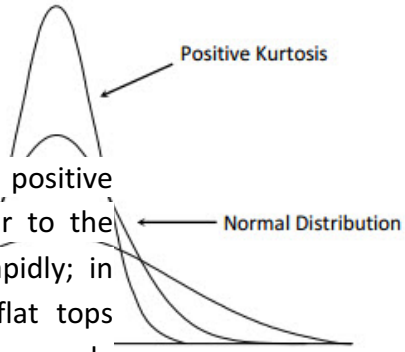
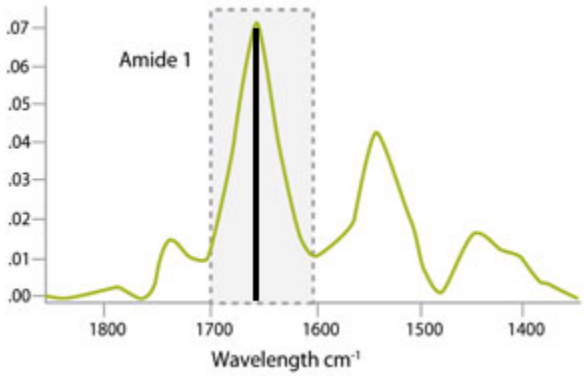


Figure 6.2 - Example of peak kurtosis - positive kurtosis spectra exhibit distinct peaks near to the mean (normal distribution) and decline rapidly; in contrast, negative kurtosis spectra have flat tops near to the mean rather than a sharp peak. Replicated from (2).

Skewness is a measure of the symmetry, or lack of symmetry, of a peak. Normally distributed data has a skewness of zero. Negative values for skewness represent data which are skewed to the left and positive to the right. Data which is skewed to the left, for example, has a longer tail on the left in comparison to the tail on the right of the peak (3). Figure 6.3 illustrates the skewness of a peak (4).

Figure 6.4 - Peak centroid of the Amide I band.
Adapted from (6).

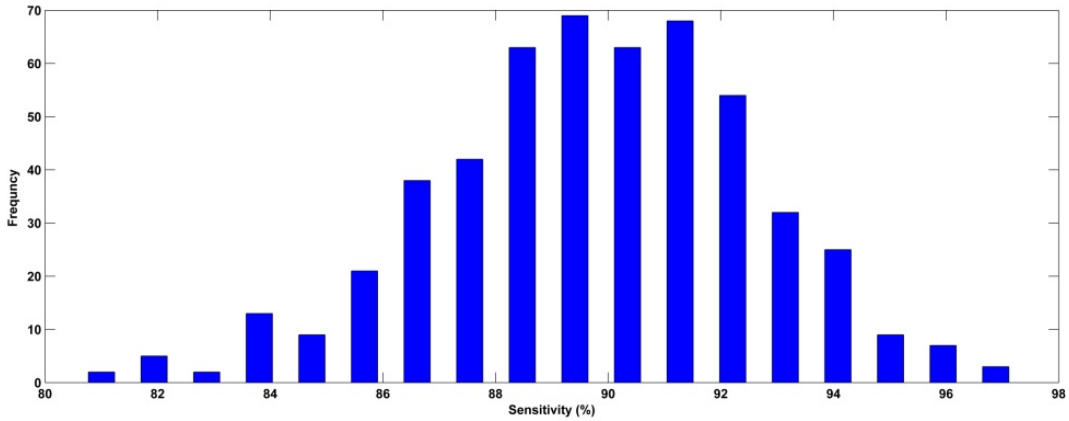
Peak centroiding finds the centre of the user highlighted peak to basically represent a weighted-average of the peak (5). Figure 6.4 illustrates the centroid of the Amide I band (6).

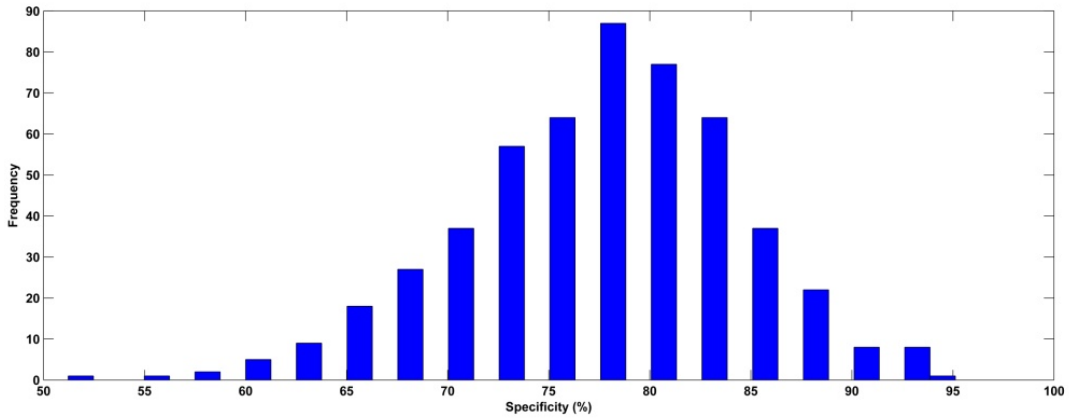


6.6 Results from ATR-FTIR Diagnostic Model

ATR-FTIR spectra from 433 patients (3897 spectra) were analysed to investigate sensitivities and specificities possible on a patient level. 525 iterations with different training and blind spectral datasets (split 1/3 blind and 2/3 training on a patient basis) were used to analyse the power of the RBF-SVM analysis and to understand the variance within a biospectroscopic dataset. Figures 6.5 - 6.6 show of the range of sensitivities and specificities achieved for the cancer *vs.* non-cancer stratum.

Figure 6.6 - Histogram showing the specificity results for 525 iterations of SVM conducted using the fingerprint region from 1800-1000 cm^{-1}.
SVM conducted using the fingerprint region from 1800-1000 cm^{-1}.





The sensitivity and specificity range for cancer *vs.* non-cancer is 81 – 97 % and 51 – 95% respectively. Sensitivities and specificities range from;

- 46 – 80 % and 60 – 93 % for metastatic cancer *vs.* brain cancer respectively (Appendix 4 - Figure 1);
- 48 – 100 % and 31 – 100 % for glioma *vs.* meningioma respectively (Appendix 4 - Figure 2);
- 50 – 100 % and 2 – 100 % for HGG *vs.* LGG respectively (Appendix 4 - Figure 3);

Table 6.3 - Mean, mode and optimum sensitivities (sens) and specificities (spec) obtained for each stratum

- 28 – 95 % and 68 – 98 % for the metastatic origin stratum (Appendix 4 - Figure 4 - breast; Appendix 4 - Figure 5 - skin; Appendix 4 - Figure 6 - lung).

These ranges of sensitivities and specificities show the diagnostic potential of the 525 iterative approach for the discussed datasets. The ranges of sensitivities and specificities demonstrate the ability of spectral diagnoses with various splits of train datasets. It is shown (Figure 6.5-6.6) that the cancer *vs*. non-cancer iterative approach had a mode of 89.4 % and 78.0 % for sensitivity and specificity, respectively.

Table 6.3 shows the mean, mode and optimum sensitivities and specificities for each stratum. The optimum sensitivity and specificity columns refer to the sensitivity and specificity that best describes the sample set based upon disease grouping.

Model	Optimum Sens (%)	Optimum Spec (%)	Mean Sens (%)	Mean Spec (%)	Mode Sens (%)	Mode Spec (%)
Cancer *vs.* Non-Cancer	97.1	95.1	**89.8**	**77.5**	89.4	78.0
Metastatic Cancer *vs.* Brain Cancer	80.0	93.2	**79.7**	**64.0**	64.4	80.0
Glioma *vs.* Meningioma	100	100	**81.1**	**66.7**	82.1	75.0
HGG vs. LGG	100	100	**80.9**	**48.5**	85.0	50.0

ble 6.3 - *continued…*

Metastatic Model	Optimum Sens (%)	Optimum Spec (%)	Mean Sens (%)	Mean Spec (%)	Mode Sens (%)	Mode Spec (%)
Metastatic Lung	95.4	95.9	**79.0**	**85.7**	81.4	84.9
Metastatic Skin	84.4	94.4	**63.9**	**82.0**	64.4	80.3
Metastatic Breast	78.6	98.9	**51.4**	**90.1**	50.0	90.9
Metastatic Model Mean	86.3	98.3	**64.8**	**86.0**	65.3	85.4

The optimum, mode and mean sensitivities and specificities observed for all strata range from 48.5 % to 100.0 % respectively, with the optimum sensitivities and specificities ranging from 86.3 % - 100 %. The cancer *vs.* non-cancer stratum achieved a mean sensitivity and specificity of 89.8 % and 77.5 % respectively, metastatic cancer *vs.* brain cancer of 79.7 % and 64.0 % respectively, glioma *vs.* meningioma of 66.7 % and 82.1 % respectively, high grade glioma *vs.* low grade glioma of 80.9 % and 48.5 % respectively and the origin of metastasis of 64.8 % and 86.9 % respectively.

These results show the power of ATR-FTIR spectroscopy to diagnose disease states based upon a stratified approach; however variance still exists in the spectral datasets due to the selection of patient populations in the training and blind set. For each stratum, sensitivity and specificity variance exists between classification model iterations. This shows that certain patient partitions provide better classification for

the remaining blind patient data set. A reason for this is redundant data maximising the spectral variance within a group within the data variables of the spectral fingerprint region i.e. patient data containing higher intra-group spectral variance partitioned together to form the training set would produce poorer classification models.

6.7 Results of Stratified Diagnostics via FE

To maximise classification accuracies the most discriminatory features of a spectrum can be extracted and ranked based on their similarity of a desired dataset, thus assigning scores on the feature's ability to discriminate between classes, maximising inter-group differences (7). Unless otherwise stated, all tabulated spectral assignments were taken from (8-13).

As discussed previously, the spectral features used to discriminate between the pre-defined classes are the peak centroid (measure of the peak's central point), peak skew (measure of asymmetry in the peak's shape), peak kurtosis (a measure of the shape of a RMS peak relating peaked vs. flat-topped), peak amplitude and RMS energy. These features were extracted from pre-defined sub-bands of each spectrum (user highlighted in the SpecToolbox); features were then ranked, using the information gain, based upon the resulting score.

Following FE and variable ranking, the most discriminatory characteristics of the spectrum (from 1800 to 900 cm^{-1}) were extracted. Table 6.4 displays the most discriminatory regions with proposed biomolecular assignments highlighting

Table 6.4 - The discriminatory wavenumber regions with assigned biomolecular assignments

spectral components relating to proteins, lipids, carbohydrates and nuclear material.

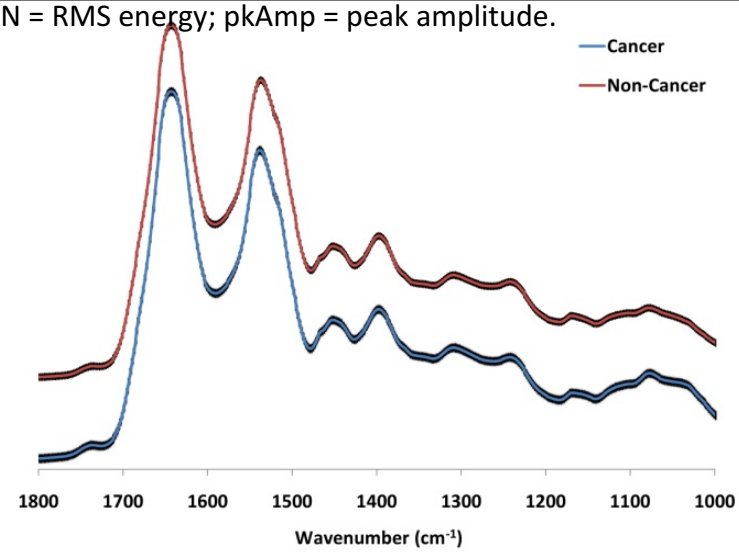
Wavenumber Region(cm^{-1})	**Spectral Assignments**
1008 - 1230	C-O, deoxyribose/ribose, DNA, RNA (PO_2^-), C-C, C-H bend
1315 - 1384	CH_3/CH deformation of lipids
1380 - 1465	CH_3 lipids/proteins
1460 - 1590	Amide II
1600 - 1706	Amide I
1700 - 1799	C=O of lipids

The features observed for the 2-class strata - enabling classification of cancer *vs.* non-cancer, metastatic *vs.* brain cancer, glioma *vs.* meningioma and high-grade glioma *vs.* low-grade glioma - focus on the detection of primary brain cancer originate from the Amide I (8) and Amide (9), C-O stretch of lipids/proteins(10), CH_2 of lipids/proteins (13) and contributions from nuclear materials (DNA & RNA via PO_2^- stretches) wavenumber regions (10). The top 10 features for each 2-class stratum are displayed in Figures 6.7-6.10, with a mean spectrum for each individual disease state.

These wavenumber regions have been described previously in research by Gajjar *et al.* observing discrimination between brain cancer states using tissue spectroscopy (11). This research highlighted the Amide I (1655 cm^{-1}), Amide II (1547 and 1582 cm^{-1}), carbohydrate (1173 cm^{-1}), glycogen (1014 cm^{-1}) and phosphate regions as describing the majority of the difference between IR spectra of tissue origination between non-cancerous patients and tumour types. Growth factor proteins are known to be excessively produced in tissue of high-grade brain tumours where proliferation, angiogenesis and metastasis is occurring (14), therefore, the spectral response to the presence of these proteins can be used to differentiate between brain cancer and non-cancer disease state tissue specimens. During carcinogenesis, proliferating cells within tissue uptake glycogen and amino acids to produce nucleic acids, consequently the increased proliferation occurring in cancerous tissues leads to adapted metabolic needs and increased concentrations of the described molecules, which are detectable via IR spectroscopy (15). An example of the molecules within blood serum allowing for differentiation is interleukin-8 (IL-8) which is expressed in cancer and has been deemed as significant with tumour severity (Chapter 4). (16). IL-8 is a key protein involved in neoangiogenesis - a process whereby cancer creates new blood vessels to supply developing tumours with nutrients and ensures growth. Many tumours produce growth factors, such as interleukin-8, that stimulate angiogenesis thus inducing surrounding healthy cells to secrete the same factors. The presence of IL-8, among other cytokines and angiogenesis factors, in patients with gliomas is known (16-17), in addition, our Top 10 spectral assignments (Figure 6.7-6.10 and Table 6.6-6.8)

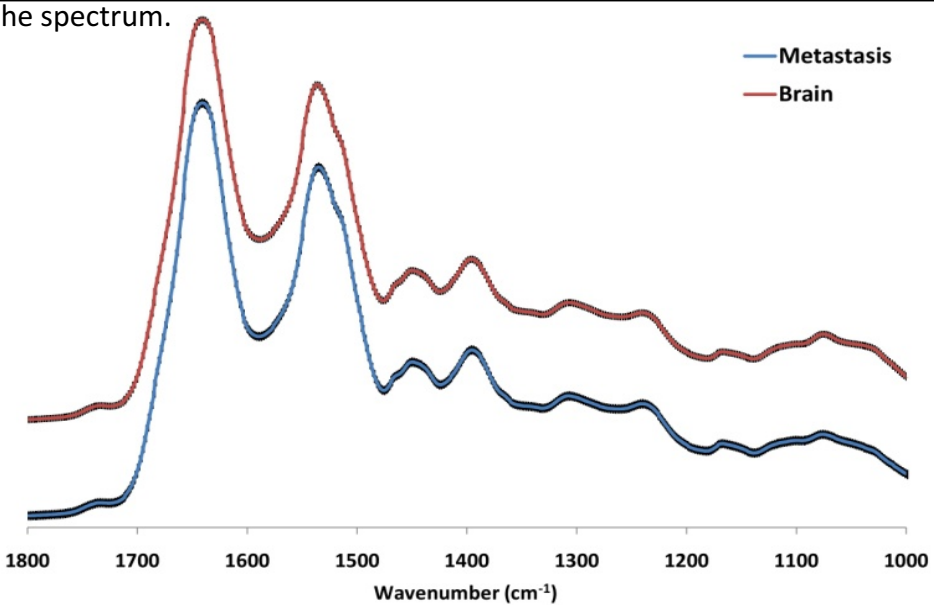
allowing for discrimination between cancer *vs.* non-cancer to primary-site metastatic brain cancer include Amide I and Amide II, thus pointing such biomolecules towards being involved in the discrimination. Supporting this, Tables 5.5-5.6 in Chapter 5, concerned with the discrimination between brain tumour severities, found that the loadings from PCA show the major peaks allowing for discrimination between low-grade glioma and non-cancer from high-grade glioma include Amide I (–ve 1649 cm^{-1}) and Amide II (+ve 1513 cm^{-1}). The major peaks allowing for discrimination between low-grade glioma and non-cancer include Amide I (+ve 1655 cm^{-1} and -ve 1644 cm^{-1}) and Amide II (+ve 1589 cm^{-1}). During tumour development, tumour cells release their nucleic acids into the blood circulation via apoptosis and necrotic cell death, resulting in circulating DNA and RNA in blood samples from cancer patients. Moredechai *et al*. reports that DNA/RNA spectral ratios are higher in malignant samples than in non-diseased samples (18), additionally; this research has found DNA/RNA spectral bands to be involved in the discrimination between disease states.

ure 6.7 – Top ten extracted features, based upon feature extraction score, and
erage spectrum from each disease state for cancer *vs.* non-cancer. Spectra
fset for ease of visualisation) display standard deviation error bars presented
a cloud around the spectrum. RMSN = RMS energy; pkAmp = peak amplitude.



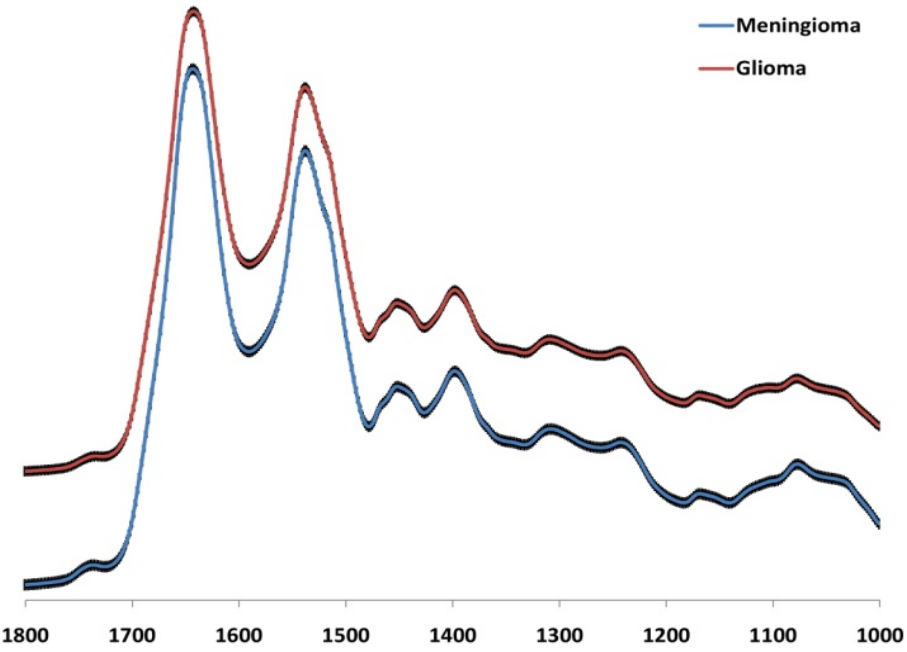
FE Number	Type	Feature Wavenumber (cm^{-1})	Brief Assignment	FE Score
1	RMSN	1242-1176 *vs.* 1115-1020	C-OH groups, PO_2^-, RNA/DNA *vs.* PO_2^- stretch of nucleic acids, RNA/DNA	0.4415
2	Skew	1242-1176 *vs.* 1537-1483	C-OH groups, PO_2^-, RNA/DNA *vs.* CH_2 of lipids, Amide II of proteins	0.4195
3	pAmp	1242-1176 *vs.* 1115-1020	C-OH groups, PO_2^-, RNA/DNA *vs.* PO_2^- stretch of nucleic acid, RNA/DNA	0.4169
4	RMSN	1537-1483 *vs.* 1115-1020	CH_2 of lipids, Amide II of proteins *vs.* PO_2^- stretch of nucleic acids, RNA/DNA	0.3997
5	RMSN	1537-1483 *vs.* 1635-1599	CH_2 of lipids, Amide II of proteins *vs.* Amide I of	0.3881
6	Skew	1635-1599 *vs.* 1537-1483	Amide I of β-pleated sheet *vs.* CH_2 of lipids, Amide II of proteins	0.3864
7	RMSN	1537-1483	CH_2 of lipids, Amide II of proteins	0.3863
8	Centroid	1115-1020 *vs.* 1585-1543	PO_2^- stretch of nucleic acids, RNA/DNA *vs.* Amide II	0.3845
9	Centroid	1585-1543 *vs.* 1115-1020	Amide II δ N-H (60%), *v* C-N (40%), α-helix *vs.* PO_2^- stretch of nucleic acids, RNA/DNA	0.3537
10	Skew	1537-1483 *vs.* 1242-1176	CH_2 of lipids, Amide II of proteins *vs.* C-OH groups, PO_2^-, RNA/DNA	0.3448

Figure 6.8 - Top ten extracted features, based upon feature extraction score, and average spectrum from each disease state for metastatic cancer *vs.* brain cancer. Spectra (offset for ease of visualisation) display standard deviation error bars presented as a cloud around the spectrum.



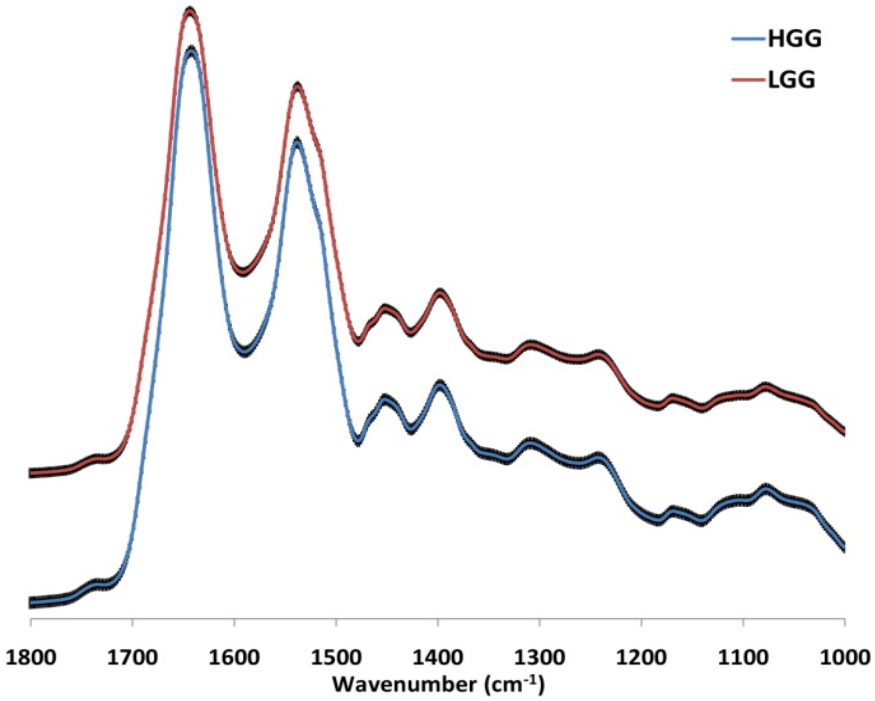
FE Number	Type	Feature Wavenumber (cm^{-1})	Brief Assignment	FE Score
1	RMSN	1589-1558 vs. 1787-1699	Amide II δ N-H (60%), *v* C-N (40%), α-helix *vs.* C=O of lipids, RNA/DNA	0.2738
2	RMSN	1378-1317 vs. 1787-1699	COO^-, CH_3 of lipids/proteins *vs.* C=O of lipids, RNA/DNA	0.2627
3	RMSN	1787-1699	C=O of lipids, RNA/DNA	0.2546
4	RMSN	1699-1637 vs. 1787-1699	Amide I of proteins *vs.* C=O of lipids, RNA/DNA	0.2433
5	RMSN	1787-1699 vs. 1590-1558	C=O of lipids, RNA/DNA *vs.* Amide II δ N-H	0.2343
6	RMSN	1079-1025 vs. 1787-1699	C-O of glycogen *vs.* C=O of lipids, RNA/DNA	0.2256
7	Centroid	1787-1699	C=O of lipids, RNA/DNA	0.2189
8	RMSN	1787-1699 vs. 1699-1637	C=O of lipids, RNA/DNA *vs.* Amide I of proteins	0.2158
9	RMSN	1378-1317	COO^-, CH_3 of lipids/proteins, Amide III	0.2114
10	RMSN	1787-1699 vs. 1378-1317	C=O of lipids, RNA/DNA*vs.* COO^-, CH_3 of lipids/proteins, Amide III	0.1938

re 6.9 - Top ten extracted features, based upon feature extraction score, and age spectrum from each disease state for glioma *vs*. meningioma. Spectra (offset ease of visualisation) display standard deviation error bars presented as a cloud nd the spectrum.



FE Number	Type	Feature Wavenumber (cm^{-1})	Brief Assignment	FE Score
1	RMSN	1459-1369 vs. 1591-1556	CH_2/CH_3 of proteins/lipids , COO^- of fatty acids *vs.* Amide II	0.4466
2	RMSN	1591-1556 vs. 1459-1369	Amide II *vs.* CH_2/CH_3 of proteins/lipids	0.4408
3	RMSN	1085-1029 vs. 1459-1369	PO_2^- stretch of nucleic acids, RNA/DNA *vs.* CH_2/CH_3 of proteins/lipids	0.4407
4	RMSN	1459-1369	CH_2/CH_3 of proteins/lipids	0.4310
5	RMSN	1459-1369 vs. 1672-1612	CH_2/CH_3 of proteins/lipids *vs.* Amide II	0.4166
6	RMSN	1672-1612 vs. 1459-1369	Amide II *vs.* CH_2/CH_3 of proteins/lipids	0.4045
7	RMSN vs. pkAmp	1459-1369 vs. 1591-1556	CH_2/CH_3 of proteins/lipids *vs.* Amide II	0.3856
8	RMSN	1459-1369 vs. 1758-1670	CH_2/CH_3 of proteins/lipids *vs.* Amide I, C=O of nucleic acids/lipids	0.3776
9	RMSN	1758-1670 vs. 1672-1612	Amide I, C=O of nucleic acids/lipids *vs.* Amide II	0.3738
10	RMSN	1672-1612 vs. 1758-1670	Amide II *vs* Amide I, C=O of nucleic acids/lipids	0.3683

Figure 6.10 - Top ten extracted features, based upon feature extraction score, and average spectrum from each disease state for HGG *vs*. LGG. Spectra (offset for ease of visualisation) display standard deviation error bars presented as a cloud around the spectrum.



FE Number	Type	Feature Wavenumber (cm^{-1})	Brief Assignment	FE Score
1	Skew	1704-1600	Amide I	0.3359
2	RMSN	1380-1328 vs. 1490-1452	COO^-, CH_3 of lipids/proteins *vs.* CH_2 of lipids/CH_3 of lipids/proteins	0.3219
3	Skew	1704 -1600 vs. 1787-1702	Amide I *vs.* C=O of lipids, RNA/DNA	0.3208
4	Skew	1787-1702 vs. 1704-1600	C=O of lipids, RNA/DNA *vs.* Amide I	0.3096
5	RMSN	1380-1328 vs. 1201-1172	COO^-, CH_3 of lipids/proteins *vs.* C=O of proteins	0.3010
6	Centroid	1704-1600 vs. 1787-1702	Amide I *vs.* C=O of lipids, RNA/DNA	0.3005
7	Centroid	1787-1702 vs. 1704-1600	C=O of lipids, RNA/DNA *vs.* Amide I of proteins	0.3005
8	Centroid	1704-1600 vs. 1201-1172	Amide *vs.* C=O of proteins	0.2998
9	Centroid	1201-1172 vs. 1704-1600	C=O of proteins *vs.* Amide I	0.2998
10	RMSN	1380-1328 vs. 1787-1702	COO^-, CH_3 of lipids/proteins *vs.* C=O of lipids, RNA/DNA	0.2993

The features observed for the metastatic stratum enabling discrimination between the organs of origin (lung *vs.* melanoma *vs.* breast) originate from vibrations of C-O, C=O and C-H associated with lipids and protein macromolecules. Additionally, contributions associated with nucleic material (DNA & RNA via PO_2^-) and minimal contributions from the Amide wavenumber regions allow for the discrimination between the metastatic subtypes (Top 10 features tabulated in 6.5 - 6.7 for each primary site. interlekin-8, leptin and cell-free serum DNA secreted during vasculogenesis and angiogenesis have a positive correlation with the development and metastasis of cancer, thus enabling for discrimination between the metastatic stratum (19).

Figure 6.11 displays an average spectrum of each disease state. This is consistent with research performed by Gazi *et al*. (12) when utilising FTIR microscopy to investigate discrimination of metastatic prostate cancer tissue and organ confined prostate cancer. Gazi *et al*. show increases in biomolecular intensities of carbohydrate, phosphate and lipid intensities between organ confined prostate cancer and prostate cancer bone metastatic tissue specimens.

Krafft *et al*. (20) highlights spectral features at 1026, 1080 and 1153 cm^{-1} as molecular markers for brain metastases of the primary tumour renal cell carcinoma. Krafft *et al.* also found that the intensity at 1735 cm^{-1} (assigned to the carbonyl vibrations (C=O) of ester groups) was indicative of brain metastases of breast cancer; an increase in Amide II intensity and broadening of the Amide I low wavenumber shoulder near 1625 cm^{-1} for brain metastases of lung cancer and an intensity minimum near 1400 cm^{-1} for brain metastases of colorectal cancer when

Figure 6.11 - Average spectrum from each disease state (lung, melanoma (skin) and breast) for metastatic disease origin. Spectra (offset for ease of visualisation) display standard deviation error bars presented as a cloud around the spectrum.

performing IR spectroscopic imaging of brain tissue. The similar wavenumber regions observed for the tissue spectroscopic studies, as compared to serum based spectroscopic studies, provide corroborating evidence of the power of the analysis as serum biochemical profiles are understood to reflect the tissue status.

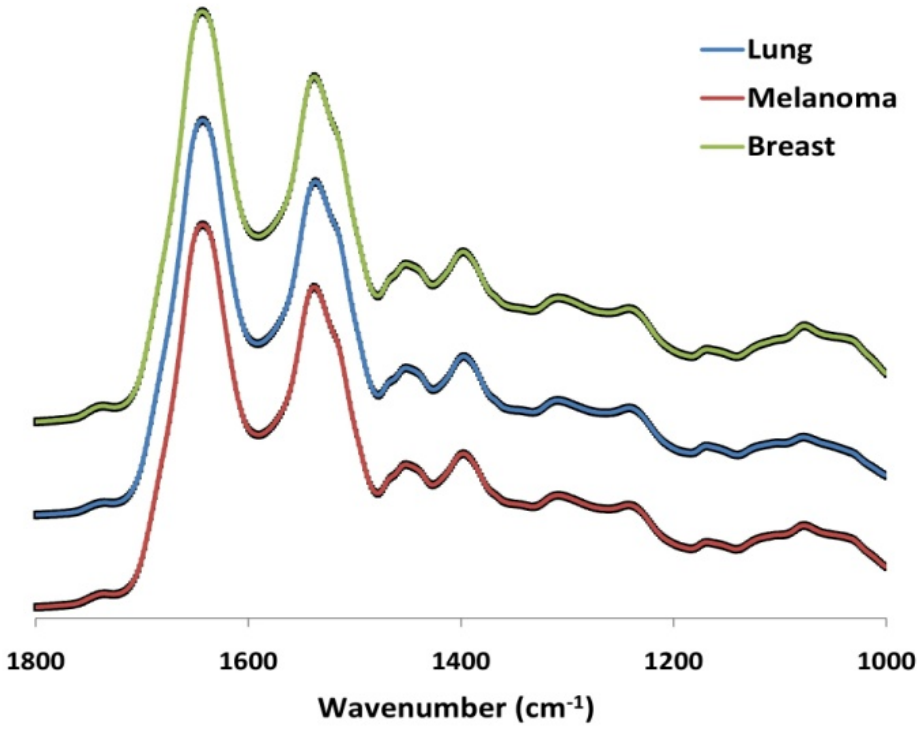


Table 6.6 - Top ten extracted features, based upon feature extraction score, for lung cancer metastatic origin.

Table 6.5 - Top ten extracted features, based upon feature extraction score, for melanoma (skin) cancer metastatic origin.

FE Number	Type	Feature Wavenumber (cm^{-1})	Brief Assignment	FE Score
1	Skew	1095-1012 vs. 1542-1475	C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA *vs.* CH_2/CH_3 of lipids, Amide II	0.4902
2	Skew	1095-1012	C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA	0.4719
3	Skew	1095-1012 vs. 1797-1749	C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA *vs.* C=O of lipids	0.4703
4	Skew	1095-1012 vs. 1631-1606	C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA *vs.* Amide I	0.4649
5	Skew	1095-1012 vs. 1224-1112	C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA *vs.* C-O of glycogen, PO_2^- of nucleic acids	0.4608
6	RMSN	1095-1012 vs. 1631-1606	C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA *vs.* Amide I	0.4584
7	Kurtosis	1095-1012 vs. 1224-1112	C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA *vs.* C-O of glycogen, PO_2^- of nucleic acids	0.4436
8	RMSN	1542-1475 vs. 1797-1749	CH_2/CH_3 of lipids, Amide II *vs.* C=O of lipids	0.4402
9	Kurtosis	1631-1606 vs. 1095-1012	Amide I *vs.* C-O of glycogen, PO_2^- stretch of nucleic acids. RNA/DNA	0.4293
10	Kurtosis	1224-1112 vs. 1095-1012	C-O of glycogen, PO_2^- of nucleic acids *vs.* C-O of glycogen, PO_2^- stretch of nucleic acids, RNA/DNA	0.4084

FE Number	Type	Feature Wavenumber (cm^{-1})	Brief Assignment	FE Score
1	Kurtosis	1187-1106 vs. 1103-1012	C-O of proteins, RNA/DNA *vs.* C-O of glycogen,PO_2^- of nucleic acids, RNA/DNA	0.4967
2	Skew	1103-1012 vs. 1625-1604	C-O of glycogen,PO_2^- of nucleic acids, RNA/DNA *vs.* Amide I	0.4941
3	RMSN	1103-1012 vs. 1625-1604	C-O of glycogen,PO_2^- of nucleic acids, RNA/DNA *vs.* Amide I	0.4790
4	Kurtosis	1103-1012 vs. 1187-1106	C-O of glycogen,PO_2^- of nucleic acids, RNA/DNA *vs.* C-O of proteins, RNA/DNA	0.4740
5	Skew	1103-1012	C-O of glycogen,PO_2^- of nucleic acids, RNA/DNA	0.4683
6	Skew	1799-1776 vs. 1103-1012	C=O of lipids *vs.* C-O of glycogen,PO_2^- of nucleic acids ,RNA/DNA	0.4674
7	Skew	1103-1012 vs. 1625-1604	C-O of glycogen,PO_2^- of nucleic acids, RNA/DNA *vs.* Amide I	0.4627
8	Kurtosis	1540-1492 vs. 1103-1012	Amide II δ N-H (60%), ν C-N (40%), β-structure *vs.* C-O of glycogen, PO_2^- of nucleic acids, RNA/DNA	0.4566
9	Kurtosis	1625-1604 vs. 1103-1012	Amide I *vs.* C-O of glycogen, PO_2^- of nucleic acids, RNA/DNA	0.4410
10	Skew	1103-1012 vs. 1799-1776	C-O of glycogen, PO_2^- of nucleic acids, RNA/DNA *vs.* C=O of lipids	0.4405

Table 6.7 - Top ten extracted features, based upon feature extraction score, for breast cancermetastatic origin.

FE Number	Type	Feature Wavenumber (cm^{-1})	Brief Assignment	FE Score
1	Skew	1089-1012 vs. 1550-1465	C-O of proteins, RNA/DNA *vs.* CH_2 of lipids, Amide II	0.4923
2	Kurtosis	1089-1012 vs. 1089-1012	C-O of proteins, RNA/DNA *vs.* C-O of proteins, RNA/DNA	0.4832
3	Skew	1089-1012 vs. 1236-1207	C-O of proteins, RNA/DNA *vs.* PO_2^- of nucleic acids	0.4805
4	Skew	1089-1012	C-O of proteins, RNA/DNA	0.4744
5	Kurtosis	1089-1012 vs. 1089-1012	C-O of proteins, RNA/DNA *vs.* C-O of proteins, RNA/DNA	0.4545
6	Skew	1089-1012 vs. 1637-1610	C-O of proteins, RNA/DNA *vs.* Amide I	0.4544
7	RMSN	1089-1012 vs. 1637-1610	C-O of proteins, RNA/DNA *vs.* Amide I	0.4299
8	Skew	1089-1012 vs. 1089-1012	C-O of proteins, RNA/DNA *vs.* C-O of proteins, RNA/DNA	0.4238
9	RMSN	1089-1012 vs. 1236-1207	C-O of proteins, RNA/DNA *vs.* PO_2^- of nucleic acids	0.4195
10	Kurtosis	1637-1610 vs. 1089-1012	Amide I *vs.* C-O of proteins, RNA/DNA	0.4114

In order to examine the ability of feature extraction to improve the diagnostic capability of stratified serum diagnostics a 525 iteration feature-based SVM was performed using all of the 130 features discovered during the feature extraction process, the top 30 features and the top 2 features for the cancer vs. non-cancer stratum, based on a variable ranking process. The top 30 features are displayed in Table 6.8. All 130 features are tabulated in Appendix 3.

Table 6.8 - Top 30 spectral features (type and wavenumber) selected by variable ranking from the cancer *vs.* non-cancer stratum

FE Number	Type	Wavenumber (cm^{-1})	FE Score
1	RMSN	1176-1242 vs. 1020-1115	0.4512
2	Skew	1176-1242 vs. 1483-1537	0.4376
3	Peak Amplitude	1176-1242 vs. 1020-1115	0.4292
4	RMSN	1483-1537 vs. 1020-1115	0.4288
5	RMSN	1483-1537 vs. 1599-1635	0.4163
6	Skew	1599-1635 vs. 1483-1537	0.4141
7	RMSN	1483-1537	0.4032
8	Centroid	1020-1115 vs. 1543-1585	0.3985
9	Centroid	1543-1585 vs. 1020-1115	0.3982
10	Skew	1483-1537 vs. 1176-1242	0.3937
11	Centroid	1020-1115 vs. 1483-1537	0.3918
12	Centroid	1483-1537 vs. 1020-1115	0.3914
13	RMSN	1599-1635 vs. 1483-1537	0.3866
14	RMSN	1020-1115 vs. 1176-1242	0.3811
15	Centroid	1020-1115	0.3708
16	Centroid	1020-1115 vs. 1641-1693	0.3700
17	Centroid	1641-1693 vs. 1020-1115	0.3697
18	RMSN	1483-1537 vs. 1543-1585	0.3667
19	Peak Amplitude	1483-1537 vs. 1020-1115	0.3614
20	RMSN	1020-1115 vs. 1483-1537	0.3478
21	Peak Amplitude	1020-1115 vs. 1176-1242	0.3473
22	Centroid	1176-1242 vs. 1020-1115	0.3445
23	Centroid	1020-1115 vs. 1176-1242	0.3443
24	Skew	1543-1585 vs. 1483-1537	0.3350
25	Skew	1483-1537	0.3325
26	Centroid	1020-1115 vs. 1599-1635	0.3313
27	Centroid	1599-1635 vs. 1020-1115	0.3312
28	Skew	1020-1115	0.3273
29	Skew	1020-1115 vs. 1483-1537	0.3258
30	Skew	1020-1115 vs. 1641-1693	0.3255

Figure 6.12 - 6.14 displays the histograms showing the sensitivity and specificities achieved when analysing 525 iterations of a 130 feature fed SVM (Figure 6.18), 30 feature fed SVM (Figure 6.19) and 2 feature fed SVM (Figure 6.20) for the cancer *vs.* non-cancer stratum.

The range of sensitivities and specificities from FE-SVM achieved higher percentages and occurred over a smaller range, when compared to the SVM analysis of data from the full spectral fingerprint region (shown in Figures 6.5 - 6.6). Sensitivities and specificities achieved for the cancer *vs.* non-cancer dataset from FE-SVM range from:

- 82 – 98 % and 66 – 97 % respectively for the 130 feature fed SVM;
- 81 – 98 % and 66 – 95 % for the top 30 feature fed SVM respectively;
- 81 – 96 % and 51 – 95 % for the top 2 feature fed SVM respectively;

compared to 81 – 97 % and 51 –95 % respectively for the cancer *vs.* non-cancer fingerprint region SVM shown in section 6.6.

The range of sensitivities and specificities was reduced with the use of feature extraction, thus the FE-fed SVM method improved the ability of disease identification with ATR-FTIR spectroscopy. The results obtained from fingerprint region analysis *vs.* the top 2 features FE-SVM are comparable to one another. Including all 130 features allows for greater classification of blind patients.

ations from the top 130 spectral features of the cancer *vs*. non-cancer dataset
m Table 6.9).

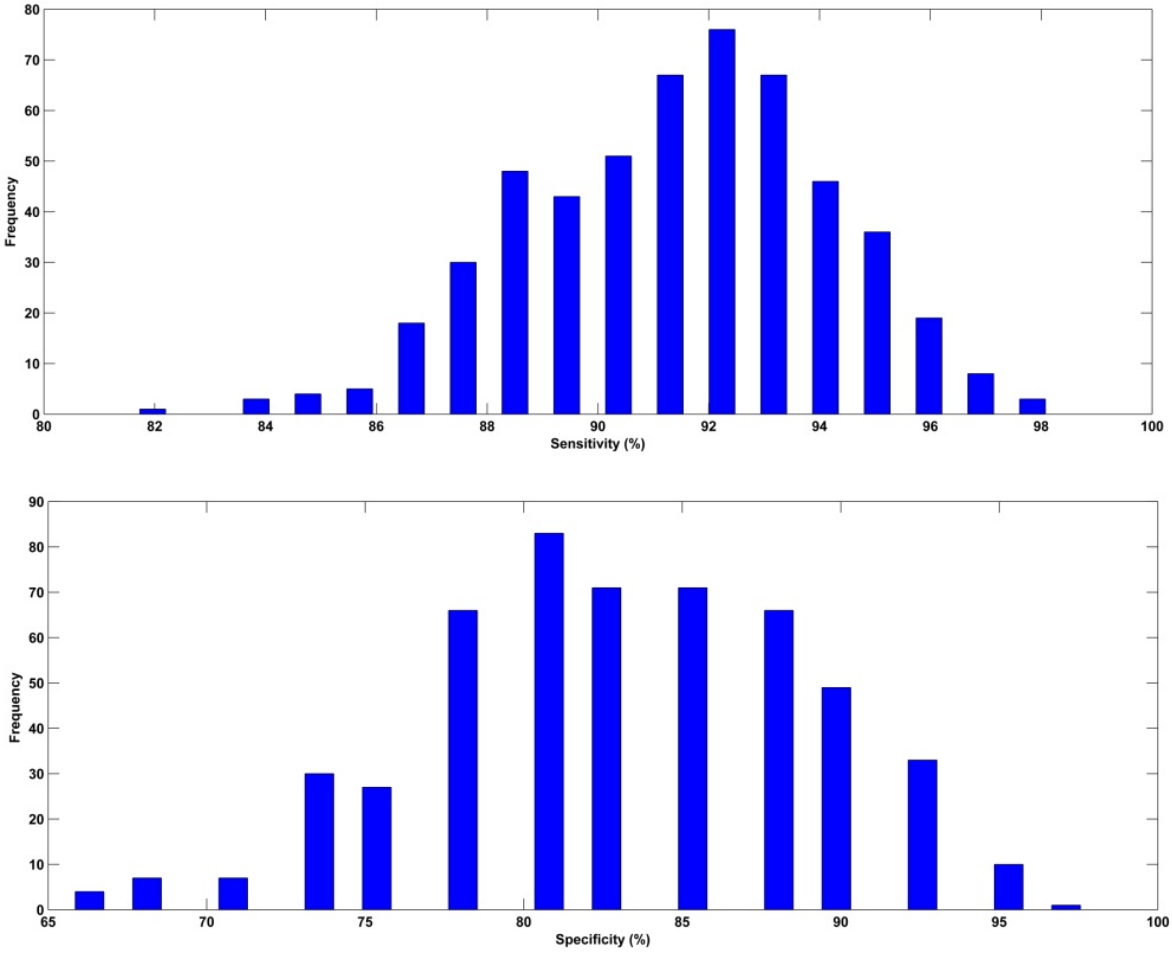
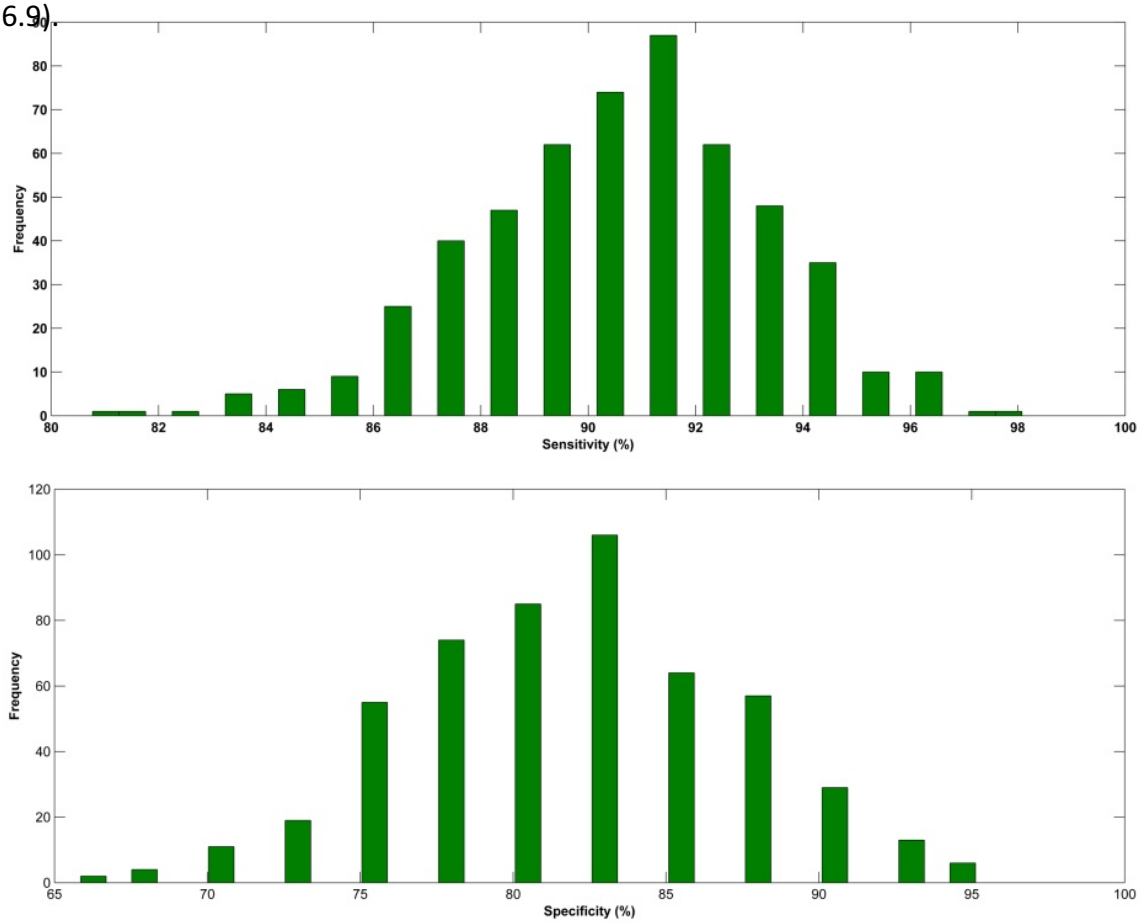
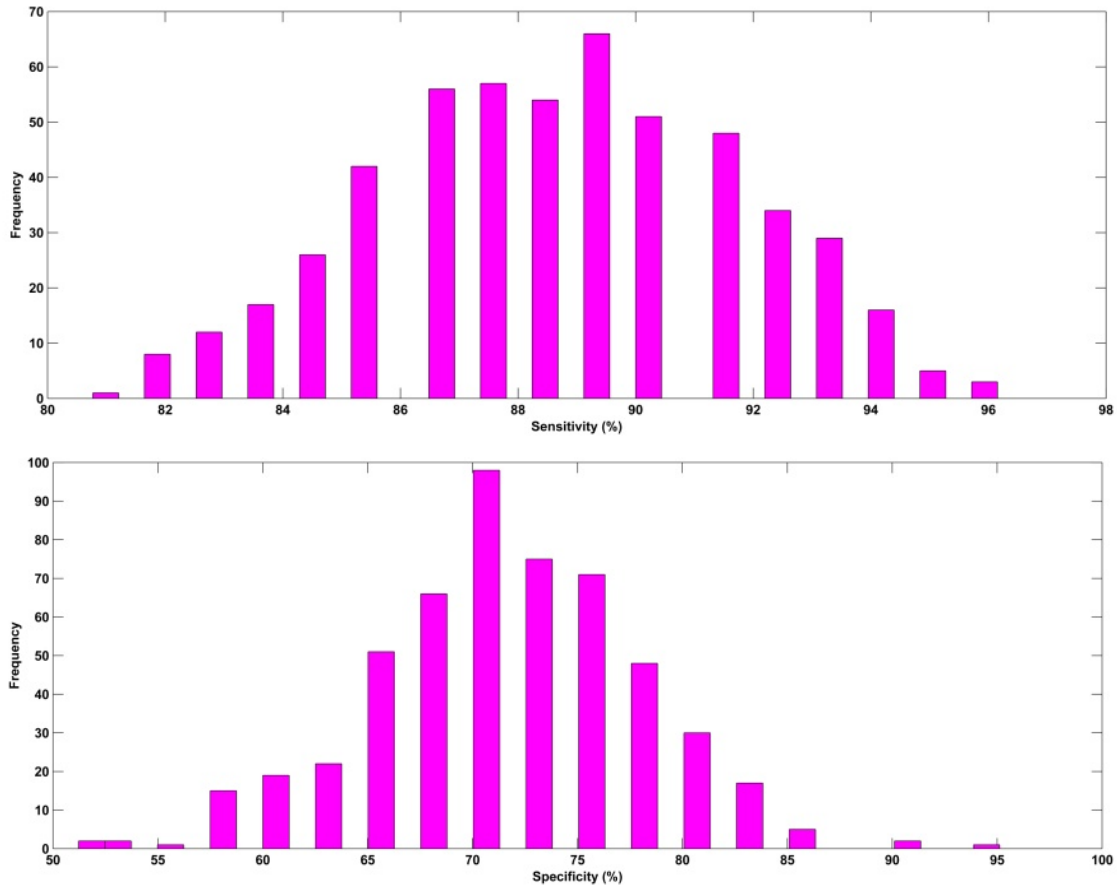


Figure 6.13 Histogram showing the range of sensitivities and specificities over 525 iterations from the top 30 spectral features of the cancer *vs*. non-cancer dataset (from Table 6.9).



525 iterations from the top 2 spectral features of the cancer *vs*. non-cancer dataset (from Table 6.9).



The mode sensitivity and specificity for the full fingerprint region SVM of the cancer *vs.* non-cancer stratum was 89.4 % and 78.0 % respectively, compared to the mode sensitivities and specificities of 92.3 % and 80.5 % when using 130 spectral features, thus FE-SVM improved upon the initial results within this chapter.

The top 30 features achieved 91.3 % and 82.9 % when using 30 features and 89.4 % and 70.7 % when using 2 spectral features. To best present the results of the FE-SVM analysis, Table 6.9 displays the results of the top 130, top 30 and top 2 spectral features of the cancer *vs.* non-cancer dataset.

able 6.9 - Optimum, mean and mode sensitivities and specificities for the ancer *vs.* non-cancer stratum using 130, 30 and 2 spectral features

Model	Optimum Sens (%)	Optimum Spec (%)	Mean Sens (%)	Mean Spec (%)	Mode Sens (%)	Mode Spec (%)
All 130 Features	98.1	97.6	**91.5**	**83.0**	92.3	80.5
Top 30 Features	98.1	95.1	**90.6**	**81.9**	91.3	82.9
Top 2 Features	96.2	95.1	**88.7**	**77.7**	89.4	70.7

The mean sensitivity and specificity for the feature extracted models follows the same trend with all 130 features achieving 91.5 % sensitivity and 83.0 % specificity, 30 features achieving 90.6 % sensitivity and 81.9 % specificity and 2 features achieving 88.7 % sensitivity and 77.7 % specificity. Whole serum analysis with all 130 relevant spectral features achieved the greatest diagnostic ability. The mean sensitivities and specificities shown on table 6.9 show a decreasing trend due to the removal of features from the analysis.

The mean sensitivities and specificities achieved using full fingerprint region SVM are similar to those that can be achieved using the top 2 spectral features of 89.8 % sensitivity and 77.5 % specificity. The top 2 spectral features that describe the differences between the cancer *vs.* non-cancer disease groupings are RMS energy of C-OH groups (10), PO_2^-, RNA/DNA (8) (1176-1242 cm^{-1}) *vs.* vibrations PO_2^-

stretch of nucleic acids (8), RNA/DNA (8) (1020-1115 cm^{-1}) and the skew of the C-OH groups of lipids/proteins, DNA/RNA (1176-1242 cm^{-1}) *vs.* CH_2 of lipids/proteins (1483-1537 cm^{-1}) (8), thus the ability of just 2 spectral feature descriptors is as powerful as the full fingerprint region analysis. As discussed in 6.7, during the process of proliferation cells uptake glycogen and amino acids to produce nucleic acids for new cells growth, thus the metabolic needs of the cells are altered (15). The features for cancer *vs.* non-cancer described within this chapter represent the biological processes of cancer progression. For example, the C-OH group of lipids/proteins are related to the abundance of lipid production as a consequence of cancer progression. Likewise, the PO_2^- stretching of nucleic acids is related to tumour development. During the progression of a tumour, the tumour cells release their nucleic acids into the blood stream; hence the biochemistry of tissue is reflected in the blood which can be examined using ATR-FTIR spectroscopy. The process of the tumour cells releasing their cellular components into the blood circulation occurs by apoptosis and necrotic cell death, resulting in high levels of circulating DNA and RNA in blood samples of patients with cancer (21).

6.8 Discussion

The results presented in this chapter were obtained via 525 iterations with different training and blind spectral datasets (split 1/3 blind and 2/3 training on a patient basis) to assess the power of the RBF-SVM analysis. The FE portion of this chapter shows that greater sensitivities and specificities are achieved with the

(whole serum) cancer *vs.* non-cancer dataset during FE-SVM. Fingerprint region FE-SVM sensitivities and specificities of the cancer *vs*. non-cancer dataset range from 81-97 % and 51-95 % respectively, compared to the increased sensitivities and specificities obtained with 130 features of 82-98 % and 66-97 % respectively.

The organ of origin in metastatic brain tumour patients has also been investigated via 525 fingerprint region SVM iterations. The optimum, mode and mean sensitivities and specificities observed for all strata range from 48.5 - 100 % respectively.

Features are ranked in order of how representative they are of the original data, thus a reduction in the diagnostic ability from 2 spectral features, compared to all 130 or top 30, is not surprising due to the reduction in spectral information available during feature fed SVM. The features for the cancer *vs*. non-cancer dataset, highlighted by variable ranking, represent the biological processes of cancer progression (15).

The ability to select and rank spectral features enables the extraction of data that describes the differences within the disease groupings without addition of added variance based upon other contributing factors from the patients and enables biochemical differences, via spectral peaks, to be observed whereas a full spectral SVM does not.

In addition, the selection of spectral features, based upon the collection of the full FTIR spectrum, allows for targeting of the most discriminatory regions during a sparse frequency collection approach (22) and reduction in the processing

power required for classification of disease states providing a quicker and more efficient spectroscopic diagnostic process.

References

[1] Discovery of Sound in the Sea. *Introduction to signal levels* [Online] http://www.dosits.org/science/advancedtopics/signallevels/ [Accessed on 10th August 2015]

[2] Stack. *Imgur*Exchange [Online] http://i.stack.imgur.com/KBQLN.jpg [Accessed on 11th August 2015]

[3] National Institute of Standards and Technology. *Measures of skewness and kurtosis* [Online] http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm [Accessed on 15th August 2015]

[4] Penn State University. *Distributions* [Online] https://onlinecourses.science.psu.edu/stat100/sites/onlinecourses.science.psu.edu.stat100/files/lesson03/distributions.gif [Accessed on 15th August 2015]

[5] OFC Berkeley. *Tutorial Three: Loops and Conditionals* [Online] https://www.ocf.berkeley.edu/~ipasha/python/tutorial-three-loops.pdf [Accessed on 15th August 2015]

[6] Strug I., Utzat C., Nadler T. *IR-based protein & peptide quantitation* GEN, 2012. 32(19)

[7] Vicinanza D., Baker M J.,Stables R., Clemens G.*Assisted differentiated stem cell classification in infrared spectroscopy using auditory feedback.* In: 20th International Conference on Auditory Display (ICAD–2014), 2014-06-22 - 2014-06-25, New York.

[8] Meurens M., Wallon J., Tong J., Noel H., Haot J. *Breast cancer detection by Fourier transform infrared spectrometry*. Vibrational Spectroscopy, 1996. 10(2), 341-346

[9] Naumann D. *FT-infrared and FT-Raman spectroscopy in biomedical research*. Applied Spectroscopy Reviews, 2001. 36(2-3)

[10] Maziak D E., Do M T., Shamji F M., Sundaresan S R., Perkins D G., Wong P T. *Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: an exploratory study*. The International Journal of Cancer Epidemiology, Detection and Prevention, 2007. 31(3), 244-253

[11] Gajjar K., Heppenstall L D., Pang W., Ashton K M., Trevisan J., Patel I I., Llabjani V., Stringfellow H F., Martin-Hirsch P L., Dawson T., Martin F L. *Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis*. Analytical Methods, 2013. 1(5), 89-102

[12] Gazi E., Baker M J., Dwyer J., Lockyer N P., Gardner P., Shanks J H., Reeve R S., Hart C A., Clarke N W., Brown M D. *A correlation of FTIR spectra derived from prostate cancer biopsies with Gleason grade and tumour stage*. European Urology, 2006. 50(4), 750-761

[13] Meade A D., Lyng F M., Knief P., Byrne H J. *Growth substrate induced functional changes elucidated by FTIR and Raman spectroscopy in in-vivo- cultured human keratinocytes*. Analytical and Bioanalytical Chemistry, 2007. 387(5), 1717-1718

[14] Catuogno S., Esposito C L., Quintavalle C., Condorelli, G., Francis V., Cerchia L. *Nucleic acids in human glioma treatment: innovative approaches and recent results.*Journal of Signal Transduction, 2012. Article ID: 735135, 1-11

[15] Munoz-Pinedo C., Mjiyad N El., Ricci J-E. *Cancer metabolism: current perspectives and future directions*. Cell Death & Disease, 2012. 3(1), e248

[16] Wartofsky, L., Van Nostrand, D. *Thyroid cancer - a comprehensive guide to clinical management*. Springer, 2016

[17] Stockhammer, G., Obwegeser, A., Kostron, H., Muigg, A., Felber, S., Maier, H., Slavc, I., Gunsilius, E., Gastl, G. *Vascular endothelial growth factor (VEGF) is elevated in brain tumour cysts and correlates to tumour progression.* Acta Neuropathology, 2000. 100(1), 101-105

[18] Mordechai, S., Sahu, R. K., Hammody, Z., Mark, S., Kantarovich, K., Guterman, H., Podshyvalov, A., Goldstein, J., Argov, S. *Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma*. Journal of Microscopy, 2003. 215(1), 86-91

[19] Elishmali, Y. I., Khaddour, H., Sarkissyan, M., Wu, Y., Vadgama, J. V. *The clinical utilization of circulating cell free DNA (CCFDNA) in blood of cancer patients).*International Journal of Molecular Science, 2013. 14(9), 19825-18958

[20] Krafft C., Shapoval L., Sobottka S B., Schackert G., Salzer R. *Identification of primary tumors of brain metastasis by infrared spectroscopic imaging and linear discriminant analysis*. Technology in Cancer Research & Treatment, 2006. 5(3), 291-298

[21] Schwazenbach, H. *Circulating nucleic acids as biomarkers in breast cancer*. Breast Cancer Research, 2013. 15(5), 211

[22] Clemens, G., Bird, B., Weida, M., Rowlette, J., Baker, M. J. *Quantum cascade laser-based mid-infrared spectrochemical imaging of tissues and biofluids*. Spectroscopy Europe/Asia, 2014. 26(4), 14-20

[23] Hands, J. R. Clemens, G., Walker, C., Dawson, T., Jenkinson, M., Davis, C., Ashton, K., Brodbelt, A., Baker, M. J. *Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples.* Journal of Biophotonics, 2014. 7(3-4); 189-199

Chapter 7

Overall Conclusions

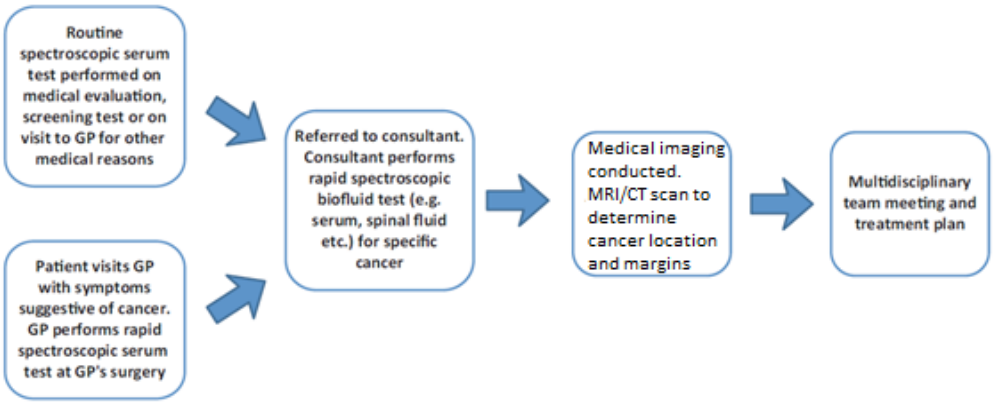
7.1 Conclusion

The early diagnosis of brain cancer is widely understood to improve mortality and morbidity rates via the early intervention of therapy, hence the development of this rapid spectroscopic diagnostic regime. The development of ATR-FTIR spectroscopy as a technique for the rapid diagnosis of primary and metastatic brain tumours has been presented within this thesis; in addition, to the best of my knowledge, this is the largest study on mid-infrared spectroscopy in relation to cancer research with a 433 patient cohort consisting of 3897 ATR-FTIR spectra.

The diagnostic methodology discussed throughout this thesis has the potential to be intertwined with current healthcare protocols to benefit patient outcomes through early cancer diagnosis. Blood is routinely collected from patients for diagnostic and monitoring purposes, creating no requirement for dedicated sample collection for objective spectral diagnoses. Whole serum ATR-FTIR spectroscopy involves no sample preparation and is cost-effective due to the minimal use of consumables (to remove the dried serum film from IRE), thus it is a beneficial diagnostic tool with little financial burden on healthcare budgets. A multi-ATR crystal device created for high-throughput analysis would allow the regime to be more practical for use in a clinical setting rather than analysing each patient one-

Figure 1.12 - A schematic showing the potential use of ATR-FTIR spectroscopy in a clinical setting. Replicated from (1).

by-one. Figure 1.12 shows a schematic showing the potential use of ATR-FTIR spectroscopy in the clinical setting.



The initial aim of this project was to determine whether cancer could be differentiated from non-cancerous state serum samples. Prior to the spectral study described in Chapter 4, our collaborator at the University of Central Lancashire (Preston, UK) investigated the biomolecules which allow for the discrimination between GBM and non-cancer patient serum samples. Follistatin achieved the highest sensitivity and specificity, although angiopoietin, interleukin-8 and leptin were among the biomolecules which allowed for discrimination between cancer and non-cancer. The aforementioned biomolecules have various molecular weights, thus it was decided to conduct a whole (unfiltered) serum study alongside a 10 kDa filtrate serum sample study (derived from the whole sample) to investigate whether improved sensitivities and specificities could be achieved via ATR-FTIR spectroscopy. The major differences between the spectra acquired from whole and 10 kDa serum filtrate samples was the presence of the Amide II (1637 cm^{-1}) and lipid CH_2 bands

(1307 cm^{-1}) in whole serum and their absence in the filtrate sample spectra. Patient level sensitivities and specificities for whole serum were 87.5 and 100 % and for 10 kDa serum filtrate 78.9 and 88.9 % respectively via ATR-FTIR. Whole (unfiltered) serum samples achieved greater sensitivities and specificities in comparison to the 10 kDa filtrate serum samples. The relatively high sensitivities and specificities from the filtrate samples gave reason to indicate that sub-fraction biomolecular components of whole serum may provide significant diagnostic targets.

An important consideration for serum analysis via ATR-FTIR is the drying of the sample on the crystal allowing for intimate contact to occur. Following 70 repeated drying experiments, ranging from 0-32 minutes, it was determined that 8 minutes was a sufficient length of time for drying of a 1 µl spot of serum at ambient room temperature.

The existing patient spectral dataset was expanded upon to not only differentiate between cancer (GBM) *vs*. non-cancer, but to include low-grade patients (e.g. astrocytoma) to allow for brain tumour severities to be determined (LGG *vs*. HGG *vs*. non-cancer) via serum ATR-FTIR spectroscopy. This portion of the project included low-grade patients and further sample filtration analysis. All whole serum samples were analysed along with their 100 kDa, 10 kDa and 3 kDa counterparts to determine which filtrate fraction allowed for optimal brain tumour severity determination. The highest patient-level sensitivities and specificities were achieved from whole serum with a general decreasing trend occurring from 100 kDa to 3 kDa filtrate samples. The method had shown the ability to diagnose gliomas from whole patient serum ATR-FTIR spectroscopy to sensitivities and specificities as

high as 100 and 95.83 % respectively. The loadings from PCA show that the major peaks responsible for discrimination between non-cancer and LGG from HGG are Amide I, Amide II and C-O stretching of DNA/RNA. Again, the Amide I and Amide II bands were responsible for the discrimination of LGG from non-cancer.

Following the investigation into determining which fraction of serum was most discriminatory for cancer diagnoses, the research project was expanded to, for the first time, detect cancer *vs.* non-cancer, glioma *vs.* meningioma, metastatic cancer *vs.* organ confined, brain cancer severity and the organ of origin of metastatic disease from whole serum samples. The aforementioned list of clinical questions ranging from cancer *vs*. non-cancer to metastatic disease origin can be determined using the same serum sample, thus enabling stratified diagnostics depending upon the clinical question asked. Chapter 6 demonstrates the successful discrimination of the discussed clinical questions with high sensitivities and specificities using the fingerprint region (1800-1000 cm^{-1}) of the collected spectra over 525 test and train set RBF-SVM iterations. Optimum sensitivities and specificities between 80.0 and 100 % were achieved for all strata.

In addition, the FE performed in Chapter 6 identified the salient spectral information for the cancer *vs.* non-cancer model which allows for targeting the most discriminatory regions during spectral collection, thus reducing collection times. Chapter 6 examined the ability of FE to improve our diagnostic ability by extracting discriminatory features of the original spectral data. Increased sensitivities and specificities of 82-98 and 66-97 % respectively were achieved for the top 130 FE-fed

SVM compared to 81-97 % and 51-95 % respectively for the cancer *vs.* non-cancerfingerprint region RBF-SVM.

This thesis demonstrates the ability to reduce the time to diagnosis based upon a relatively non-invasive diagnostic test that would provide rapid patient entry to the clinical process (Figure 7.1), profiling of at-risk population cohorts, as well as enabling close clinical follow-up throughout resulting in a reduction in mortality and morbidity rates whilst increasing the efficiency of the healthcare system.

References

[1] Dorling K M., Baker M J. *Highlighting attenuated total reflection Fourier transform infrared spectroscopy for rapid serum analysis*. Trends of Biotechnology, 2013. 31(6), 327-328

Chapter 8

Potential for Future Work

8.1 Future Work

The direction for future biofluid/serum ATR-FTIR spectroscopy will move towards the development of a multi-ATR crystal device to allow for high-throughput analysis to be performed in a clinical setting. Currently, each patient serum sample is analysed individually, hence the practicality of a one-by-one diagnostic regime in a clinical environment is not ideal.

This thesis has focused on the development of ATR-FTIR spectroscopy for brain tumour diagnoses via blood serum; however, the regime presented within this thesis could be adapted for use with other biofluids, such as urine, cerebral spinal fluid (CSF) and tears etc. Furthermore, other disease targets such as urinary system tumours or renal failure could be investigated using patient urine samples.

An interesting area to investigate, having real clinical relevance, would be to develop the regime with patients biofluids obtained from those who only have breast, lung or skin cancers to predict the likelihood of them developing a metastatic tumour to another site within the body.

ATR-FTIR spectrometers are available in hand-held versions allowing for the proposed regime to be used in the field or in situations where a bench-top FTIR spectrometer is not available/appropriate. Future research is likely to involve the

use of these hand-held instruments for spectral diagnoses as they are portable and easily translatable into the clinic.

The future of mid-infrared spectroscopy and spectral diagnosis will include QCLs for discrete frequency collection of biological samples. QCLs have recently been integrated within a microscope allowing for simpler instrumentation. Faster spectral collection times are achieved when collecting discrete target wavelengths rather than collecting the more time consuming full IR spectrum.

Aiming to test the robustness of the spectral models created throughout this PhD, towards the end of the project, 161 serum samples were obtained from patients with various grades of brain tumours and provided to us by our collaborator at Nottingham City University Hospital (NCUH), UK. All received serum samples were anonymised and the brain tumour grade of each sample was unknown. Spectral collection was performed using the same methods described throughout this thesis. The aim of this multi-centre collaboration was to receive serum samples obtained from patients with various brain tumour grades where the site of analysis was uninformed on individual patient sample diagnoses. Due to time constraints, the collected spectral data remains to be analysed, however, the goal of the analysis was to determine a spectral diagnosis for each blind serum sample. The spectral diagnoses for all patients were then to be reported back to NCUH for comparison against the patient's histopathological diagnoses, allowing for sensitivities and specificities to be calculated representing our spectral diagnostic performance.

Appendix 1

Blood sample patient demographics from samples used in Chapter 4. Appendix 2 includes these patient numbers, along with all other patient information used throughout this thesis.

Patient Number	Gender	Age at Sample Collection	Diagnosis
572	F	26	Non-Cancer
575	F	67	Non-Cancer
576	M	81	Non-Cancer
579	F	61	Non-Cancer
580	F	80	Non-Cancer
582	F	74	Non-Cancer
583	M	39	Non-Cancer
585	M	34	Non-Cancer
586	M	60	Non-Cancer
587	F	65	Non-Cancer
589	M	45	Non-Cancer
591	M	85	Non-Cancer
592	F	67	Non-Cancer
593	M	84	Non-Cancer
594	M	57	Non-Cancer
595	M	39	Non-Cancer
598	F	68	Non-Cancer
600	M	68	Non-Cancer
603	M	87	Non-Cancer
606	M	37	Non-Cancer
607	F	56	Non-Cancer
608	F	47	Non-Cancer
609	M	55	Non-Cancer
610	M	57	Non-Cancer
616	M	65	Non-Cancer
11001	M	62	Glioblastoma multiforme
11004	M	55	Glioblastoma multiforme
11005	M	66	Glioblastoma multiforme
11012	M	75	Glioblastoma multiforme
11014	M	44	Glioblastoma multiforme

11026	M	73	Glioblastoma multiforme
11028	F	65	Glioblastoma multiforme
11029	M	66	Glioblastoma multiforme
11030	F	50	Glioblastoma multiforme
11032	F	50	Glioblastoma multiforme
11034	M	58	Glioblastoma multiforme
11050	F	72	Glioblastoma multiforme
11053	M	62	Glioblastoma multiforme
11057	M	41	Glioblastoma multiforme
11073	F	63	Glioblastoma multiforme
11094	F	44	Glioblastoma multiforme
11097	M	45	Glioblastoma multiforme
11101	F	69	Glioblastoma multiforme
11120	F	51	Glioblastoma multiforme
11121	F	79	Glioblastoma multiforme
11123	M	70	Glioblastoma multiforme
11124	M	75	Glioblastoma multiforme
11137	F	66	Glioblastoma multiforme
11142	M	76	Glioblastoma multiforme
11145	M	55	Glioblastoma multiforme
11154	M	67	Glioblastoma multiforme
11155	F	70	Glioblastoma multiforme
11156	M	57	Glioblastoma multiforme
11163	M	68	Glioblastoma multiforme
11181	M	68	Glioblastoma multiforme
11188	M	70	Glioblastoma multiforme
11197	F	61	Glioblastoma multiforme
11205	M	73	Glioblastoma multiforme
11212	F	65	Glioblastoma multiforme
11268	F	74	Glioblastoma multiforme
12009	M	50	Glioblastoma multiforme
12012	M	25	Glioblastoma multiforme
12014	F	64	Glioblastoma multiforme
12019	F	39	Glioblastoma multiforme
12025	M	41	Glioblastoma multiforme
12028	F	72	Glioblastoma multiforme
12030	F	69	Glioblastoma multiforme
12032	M	64	Glioblastoma multiforme
12042	F	55	Glioblastoma multiforme

12046	F	53	Glioblastoma multiforme
12052	M	65	Glioblastoma multiforme
12055	M	25	Glioblastoma multiforme
12057	M	66	Glioblastoma multiforme
12063	M	56	Glioblastoma multiforme

Appendix 2

Expanded patient demographic information for all 433 individual patients.

Patient Number	Sex	Patient Age at Sample Collection	Diagnosis
Non-Cancer Patients			
572	F	26	Normal
575	F	67	Normal
576	M	81	Normal
579	F	61	Normal
580	F	80	Normal
582	F	74	Normal
583	M	39	Normal
585	M	34	Normal
586	M	60	Normal
587	F	65	Normal
589	M	45	Normal
591	M	85	Normal
592	F	67	Normal
593	M	84	Normal
594	M	57	Normal
595	M	39	Normal
598	F	68	Normal
600	M	68	Normal
603	M	87	Normal
606	M	37	Normal
607	F	56	Normal
608	F	47	Normal
609	M	55	Normal
610	M	57	Normal
616	M	65	Normal
933	M	41	Normal

1115	M	89	Normal
1116	F	64	Normal
1127	F	44	Normal
1128	F	43	Normal
1133	M	76	Normal
1136	F	72	Normal
1367	F	34	Normal
1368	M	46	Normal
1369	F	36	Normal
1370	F	52	Normal
1371	F	49	Normal
1372	F	41	Normal
1373	F	41	Normal
1374	F	40	Normal
1375	M	23	Normal
1376	M	53	Normal
1390	M	70	Normal
1391	M	71	Normal
1392	M	34	Normal
1393	F	71	Normal
1394	M	57	Normal
1395	M	66	Normal
1396	M	21	Normal
1397	M	32	Normal
1398	M	40	Normal
1399	F	24	Normal
1400	M	39	Normal
1401	M	65	Normal
1402	F	25	Normal
1403	F	67	Normal
1475	M	22	Normal
1476	M	41	Normal
1477	F	50	Normal
1478	M	47	Normal
1479	M	16	Normal
1480	F	39	Normal
1481	F	28	Normal

1482	M	52	Normal
1483	F	53	Normal
1484	M	56	Normal
1485	M	51	Normal
1486	M	56	Normal
1487	F	53	Normal
1488	M	54	Normal
1489	F	49	Normal
1490	F	23	Normal
1491	F	31	Normal
1492	M	31	Normal
1493	M	24	Normal
1494	M	31	Normal
1495	F	33	Normal
1496	F	37	Normal
1497	M	47	Normal
1498	F	62	Normal
1499	M	unknown	Normal
1500	F	unknown	Normal
1501	M	49	Normal
1502	M	28	Normal
1503	M	27	Normal
1504	M	28	Normal
1505	F	32	Normal
1506	F	32	Normal
1507	F	58	Normal
1508	F	51	Normal
1509	F	24	Normal
1510	M	58	Normal
1511	un- kno-wn	57	Normal
1512	M	25	Normal
1513	M	56	Normal
1514	F	25	Normal
1515	F	57	Normal
1516	F	43	Normal
1517	F	51	Normal

1518	M	32	Normal
1519	F	35	Normal
1520	F	27	Normal
1521	M	49	Normal
1522	M	unknown	Normal
1523	M	unknown	Normal
1524	M	54	Normal
1525	F	33	Normal
1526	M	29	Normal
1527	F	60	Normal
1528	M	52	Normal
1529	F	51	Normal
1530	un-kno-wn	24	Normal
1531	M	25	Normal
1533	F	53	Normal
1534	M	57	Normal
1535	M	29	Normal
1536	M	23	Normal
1537	M	24	Normal
1538	F	23	Normal
1539	F	50	Normal
1540	F	25	Normal
1541	F	50	Normal
Low-Grade Glioma Patients			
932	F	59	Oligoastrocytoma
1043	M	21	Ganglioglioma
1044	M	29	Infiltrating edge of glioma
1046	F	19	LGG diagnosis only
1047	F	54	Ependymoma
1048	F	35	LGG diagnosis only
1049	M	29	LGG diagnosis only
11006	F	34	Astrocytoma
11017	M	40	Astrocytoma
11020	M	50	Astrocytoma
11084	F	30	Astrocytoma

11109	M	46	Oligodendroglioma
12001	F	60	Astrocytoma
12060	F	53	Oligoastrocytoma
12100	M	37	Oligoastrocytoma
12103	F	31	Oligodendroglioma
12118	M	20	Astrocytoma
12121	M	59	Oligodendroglioma
12136	M	41	Oligodendroglioma
12156	F	44	Astrocytoma
12164	F	20	Astrocytoma
12324	M	43	Oligodendroglioma
12329	F	28	Oligoastrocytoma
High-Grade Glioma Patients			
935	F	61	Glioblastoma multiforme
1110	M	33	Glioblastoma multiforme
1138	M	40	Oligoastrocytoma III
1150	M	57	Gliosarcoma IV
1175	M	75	Glioblastoma multiforme
1314	M	70	Glioma III
11001	M	62	Glioblastoma multiforme
11004	M	55	Glioblastoma multiforme
11005	M	66	Glioblastoma multiforme
11012	M	75	Glioblastoma multiforme
11014	M	44	Glioblastoma multiforme
11026	M	73	Glioblastoma multiforme
11028	F	65	Glioblastoma multiforme
11029	M	66	Glioblastoma multiforme
11030	F	50	Glioblastoma multiforme
11032	F	50	Glioblastoma multiforme
11034	M	58	Glioblastoma multiforme
11050	F	72	Glioblastoma multiforme
11053	M	62	Glioblastoma multiforme
11057	M	41	Glioblastoma multiforme
11073	F	63	Glioblastoma multiforme
11094	F	44	Glioblastoma multiforme
11097	M	45	Glioblastoma multiforme

11101	F	69	Glioblastoma multiforme
11120	F	51	Glioblastoma multiforme
11121	F	79	Glioblastoma multiforme
11123	M	70	Glioblastoma multiforme
11124	M	75	Glioblastoma multiforme
11137	F	66	Glioblastoma multiforme
11142	M	76	Glioblastoma multiforme
11145	M	55	Glioblastoma multiforme
11154	M	67	Glioblastoma multiforme
11155	F	70	Glioblastoma multiforme
11156	M	57	Glioblastoma multiforme
11163	M	68	Glioblastoma multiforme
11181	M	68	Glioblastoma multiforme
11188	M	70	Glioblastoma multiforme
11197	F	61	Glioblastoma multiforme
11205	M	73	Glioblastoma multiforme
11212	F	65	Glioblastoma multiforme
11268	F	74	Glioblastoma multiforme
12009	M	50	Glioblastoma multiforme
12012	M	25	Glioblastoma multiforme
12014	F	64	Glioblastoma multiforme
12019	F	39	Glioblastoma multiforme
12025	M	41	Glioblastoma multiforme
12028	F	72	Glioblastoma multiforme
12030	F	69	Glioblastoma multiforme
12032	M	64	Glioblastoma multiforme
12042	F	55	Glioblastoma multiforme
12046	F	53	Glioblastoma multiforme
12052	M	65	Glioblastoma multiforme
12055	M	25	Glioblastoma multiforme
12057	M	66	Glioblastoma multiforme
12063	M	56	Glioblastoma multiforme
451	M	75	Glioblastoma multiforme
499	M	73	Glioblastoma multiforme
549	F	67	Glioblastoma multiforme

556	F	81	Glioblastoma multiforme
615	M	68	Glioblastoma multiforme
733	M	74	Glioblastoma multiforme
763	M	67	Glioblastoma multiforme
774	M	72	Glioblastoma multiforme
769	M	70	Glioblastoma multiforme
Meningioma Patients			
11066	F	68	Meningioma
11068	F	71	Meningioma
11070	F	56	Meningioma
11090	F	34	Meningioma
11111	M	60	Meningioma
11127	F	49	Meningioma
11173	F	69	Meningioma
11191	M	69	Meningioma
11215	F	44	Meningioma
11214	F	69	Meningioma
12043	M	49	Meningioma
12056	F	57	Meningioma
12058	M	73	Meningioma
12077	F	67	Meningioma
12085	F	58	Meningioma
12115	F	64	Meningioma
12127	F	61	Meningioma
12157	F	73	Meningioma
12163	F	66	Meningioma
12261	F	68	Meningioma
12326	M	39	Meningioma
12396	F	59	Meningioma
12418	F	67	Meningioma
12441	M	53	Meningioma
12449	M	51	Meningioma
12452	M	75	Meningioma
12470	F	62	Meningioma
13028	F	43	Meningioma
13039	F	35	Meningioma

13059	F	50	Meningioma
13062	F	72	Meningioma
13066	M	37	Meningioma
13067	F	70	Meningioma
13077	F	41	Meningioma
13080	F	44	Meningioma
13089	F	59	Meningioma
13090	F	48	Meningioma
13095	F	27	Meningioma
13099	F	37	Meningioma
13112	M	64	Meningioma
13121	M	78	Meningioma
13126	F	35	Meningioma
13133	F	78	Meningioma
13140	M	24	Meningioma
13158	M	44	Meningioma
13162	F	48	Meningioma
13212	F	66	Meningioma
Metastatic Patients			**Organ of Origin**
358	M	74	Melanoma
456	M	70	Renal
509	M	80	Melanoma
517	F	54	Lung
553	F	72	Renal
560	F	43	Breast
562	F	57	Lung
567	F	63	Melanoma
666	F	61	Lung
694	M	59	Lung
697	F	56	Renal
700	F	50	Colorectal
712	F	39	Breast
722	F	63	Breast
725	F	39	Breast
745	M	50	Lung
756	F	54	Breast

795	F	58	Bowel
827	F	55	Breast
837	F	68	Lung
841	F	36	Melanoma
847	M	63	Lung
853	F	55	Bowel
865	F	63	Lung
866	M	82	Lung
884	F	65	Lung
888	M	63	Colorectal
893	M	75	Lung
901	M	63	Melanoma
912	F	50	Breast
943	F	64	Breast
947	F	64	Lung
952	M	58	Lung
972	F	50	Breast
1001	F	67	Lung
1011	F	68	Lung
1012	F	52	Breast
1025	F	77	Rectal
1031	M	65	Lung
1040	F	51	Breast
1041	M	59	Renal
1053	M	53	Lung
1056	F	54	Melanoma
1057	M	63	Lung
1060	M	63	Lung
1070	M	71	Lung
1081	M	67	Melanoma
1082	M	71	Lung
1092	M	63	Lung
1103	F	39	Lung
1111	M	63	Melanoma
1113	F	69	Lung
1117	F	75	Lung
1144	F	67	Breast

1148	F	67	Breast
1149	M	53	Bowel
1158	F	55	Breast
1184	F	59	Lung
1311	F	74	Lung
1326	M	73	Renal
1327	M	55	Renal
1330	F	77	Lung
1338	F	71	Lung
1348	F	81	Lung
1354	F	40	Lung
1382	F	44	Breast
1431	M	54	Melanoma
1438	F	80	Lung
1443	F	52	Breast
11011	F	42	Melanoma
11019	M	59	Non-small cell lung cancer (NSCLC) Adenocarcinoma
11027	F	64	Breast
11037	M	67	NSCLCAdenocarcinoma
11043	F	59	Lung
11047	F	67	Lung
11049	M	59	NSCLC Squamous cell
11052	F	79	NSCLC Adenocarcinoma
11054	F	61	NSCLC Adenocarcinoma
11059	M	68	Unknown primary, brain metastasis is adenocarcinoma
11075	F	74	Melanoma
11076	F	60	Lung
11080	F	29	Colon
11087	M	71	NSCLC Adenocarcinoma
11089	M	73	NSCLC, brain metastasis is large cell neuroendocrine
11103	F	64	Ovarian
11104	M	65	Renal
11108	F	55	Breast

11110	M	66	NSCLC
11146	M	59	NSCLC Adenocarcinoma
11152	F	44	Breast
11161	M	64	NSCLC Adenocarcinoma
11164	F	25	Melanoma
11187	F	50	Adenocarcinoma bowel
11193	M	41	Melanoma
11213	F	60	NSCLC Adenocarcinoma
12011	M	62	Renal
12037	F	52	Lung
12041	F	64	NSCLC Adenocarcinoma
12044	F	36	Melanoma
12051	F	48	Breast
12071	M	49	Melanoma
12074	F	76	Breast
12098	F	44	Breast
12104	F	59	NSCLC Adenocarcinoma
12133	M	56	NSCLC Adenocarcinoma
12135	F	27	Breast
12138	F	44	Breast
12145	F	64	NSCLC Adenocarcinoma
12148	M	65	Renal
12149	M	56	NSCLC Adenocarcinoma
12158	F	73	NSCLC Adenocarcinoma
12160	F	63	Breast
12166	F	74	NSCLC Adenocarcinoma
12179	F	70	NSCLC Adenocarcinoma
12180	F	41	Ovarian
12212	M	65	NSCLC Adenocarcinoma
12262	F	67	NSCLC Adenocarcinoma
12264	F	69	NSCLC Adenocarcinoma
12313	M	65	Bowel
12321	F	39	Breast
12381	F	65	Renal
12382	F	58	NSCLC Adenocarcinoma
12385	F	66	NSCLC Adenocarcinoma
12397	F	62	Renal

12403	M	66	Melanoma
12416	F	78	Lung
12438	F	40	Breast
12448	F	52	Breast
12460	M	53	NSCLC Adenocarcinoma
13010	F	65	NSCLC Adenocarcinoma
13022	F	58	Breast
13035	F	39	Melanoma
13042	F	44	Melanoma
13050	F	66	NSCLC Adenocarcinoma
13051	F	53	Melanoma
13060	M	74	NSCLC Adenocarcinoma
13065	M	58	Colon
13072	M	70	Melanoma
13075	F	49	Breast
13079	F	66	NSCLC Adenocarcinoma
13101	M	26	Testicular
13107	M	66	NSCLC Adenocarcinoma
13110	F	44	Breast
13113	M	66	Lung
13130	F	57	Breast
13176	M	67	Melanoma
13196	M	60	NSCLC Adenocarcinoma
13218	F	39	Breast
13222	F	unknown	Non-small cell lung cancer (NSCLC)
13226	F	66	NSCLC Adenocarcinoma
13244	F	60	NSCLC Adenocarcinoma
13245	F	62	Endomaterial
13272	M	75	Melanoma
13276	F	67	NSCLC Adenocarcinoma
13281	M	66	NSCLC Adenocarcinoma
13288	M	68	NSCLC Adenocarcinoma
13302	F	35	Melanoma
13305	M	73	Melanoma
13306	F	57	Breast
13311	F	53	Lung

13326	M	46	Lung
13401	M	68	Lung
13405	F	73	Lung
13426	F	39	Breast
13431	M	72	Large-cell Neuroendrocrine (Lung)
13492	M	68	Lung
13510	F	72	Colon
13624	F	69	Lung
13626	F	59	Lung
13666	M	76	Lung (Atypical Carcinoid)
13667	M	38	NSCLC Adenocarcinoma
13668	M	63	NSCLC Adenocarcinoma
13669	M	69	Metastatic carcinoma from an unknown primary
13670	F	49	Breast
13671	M	53	Renal
13672	M	57	Melanoma
13673	M	42	Papillary Carcinoma Metastasis (Renal)

Appendix 3

Table 6.9 - All 130 spectral features (type and wavenumber) selected by variable ranking from the cancer *vs.* non-cancer stratum

All 130 spectral features of the cancer *vs*. non-cancer dataset acquired by the SpecToolbox

FE Number	Type	Wavenumber (cm^{-1})	FE Score
1	RMSN	1176-1242 vs. 1020-1115	0.4512
2	Skew	1176-1242 vs. 1483-1537	0.4376
3	Peak Amplitude	1176-1242 vs. 1020-1115	0.4292
4	RMSN	1483-1537 vs. 1020-1115	0.4288
5	RMSN	1483-1537 vs. 1599-1635	0.4163
6	Skew	1599-1635 vs. 1483-1537	0.4141
7	RMSN	1483-1537	0.4032
8	Centroid	1020-1115 vs. 1543-1585	0.3985
9	Centroid	1543-1585 vs. 1020-1115	0.3982
10	Skew	1483-1537 vs. 1176-1242	0.3937
11	Centroid	1020-1115 vs. 1483-1537	0.3918
12	Centroid	1483-1537 vs. 1020-1115	0.3914
13	RMSN	1599-1635 vs. 1483-1537	0.3866
14	RMSN	1020-1115 vs. 1176-1242	0.3811
15	Centroid	1020-1115	0.3708
16	Centroid	1020-1115 vs. 1641-1693	0.3700
17	Centroid	1641-1693 vs. 1020-1115	0.3697
18	RMSN	1483-1537 vs. 1543-1585	0.3667
19	Peak Amplitude	1483-1537 vs. 1020-1115	0.3614
20	RMSN	1020-1115 vs. 1483-1537	0.3478
21	Peak Amplitude	1020-1115 vs. 1176-1242	0.3473
22	Centroid	1176-1242 vs. 1020-1115	0.3445
23	Centroid	1020-1115 vs. 1176-1242	0.3443
24	Skew	1543-1585 vs. 1483-1537	0.3350
25	Skew	1483-1537	0.3325
26	Centroid	1020-1115 vs. 1599-1635	0.3313
27	Centroid	1599-1635 vs. 1020-1115	0.3312
28	Skew	1020-1115	0.3273
29	Skew	1020-1115 vs. 1483-1537	0.3258

30	Skew	1020-1115 vs. 1641-1693	0.3255
31	Peak Amplitude	1543-1585 vs. 1020-1115	0.3243
32	Skew	1641-1693 vs. 1483-1537	0.3238
33	RMSN	1543-1585 vs. 1483-1537	0.3233
34	Peak Amplitude	1483-1537 vs. 1543-1585	0.3226
35	Skew	1020-1115 vs. 1543-1585	0.3221
36	RMSN	1020-1115	0.3129
37	Skew	1020-1115 vs. 1176-1242	0.3117
38	Peak Amplitude	1599-1635 vs. 1020-1115	0.3102
39	Peak Amplitude	1543-1585 vs. 1483-1537	0.3098
40	Centroid	1599-1635 vs. 1483-1537	0.3062
41	Centroid	1483-1537 vs. 1599-1635	0.3061
42	Peak Amplitude	1020-1115 vs. 1483-1537	0.3008
43	RMSN	1599-1635 vs. 1020-1115	0.2916
44	Peak Amplitude	1641-1693 vs. 1020-1115	0.2871
45	Peak Amplitude	1020-1115 vs. 1543-1585	0.2852
46	Peak Amplitude	1020-1115	0.2788
47	RMSN	1020 -1115 vs. 1641-1693	0.2783
48	Peak Amplitude	1020 -1115 vs. 1641-1693	0.2783
49	RMSN	1543-1585 vs. 1020-1115	0.2777
50	Peak Amplitude	1020 -1115 vs. 1599-1635	0.2772
51	RMSN	1020 -1115 vs.1543-1585	0.2688
52	RMSN	1020 -1115 vs. 1599-1635	0.2640
53	Centroid	1599-1635	0.2624
54	RMSN	1641-1693 vs. 1020 -1115	0.2600
55	Skew	1483-1537 vs. 1543-1585	0.2541
56	Centroid	1543-1585 vs. 1599-1635	0.2498
57	Centroid	1599-1635 vs. 1543-1585	0.2498
58	Centroid	1543-1585 vs. 1641-1693	0.2493
59	Centroid	1641-1693 vs.1543-1585	0.2493
60	RMSN	1176-1242 vs. 1483-1537	0.2466
61	RMSN	1483-1537 vs. 1176-1424	0.2432
62	Skew	1483-1537 vs. 1641-1693	0.2405
63	Skew	1483-1537 vs. 1599-1635	0.2228
64	Centroid	1543-1585	0.2212
65	RMSN	1483-1537 vs. 1641-1693	0.1827
66	Skew	1020-1115 vs. 1599-1635	0.1826
67	Kurtosis	1599-1635 vs. 1176-1242	0.1795
68	Kurtosis	1176-1242 vs. 1599-1635	0.1733

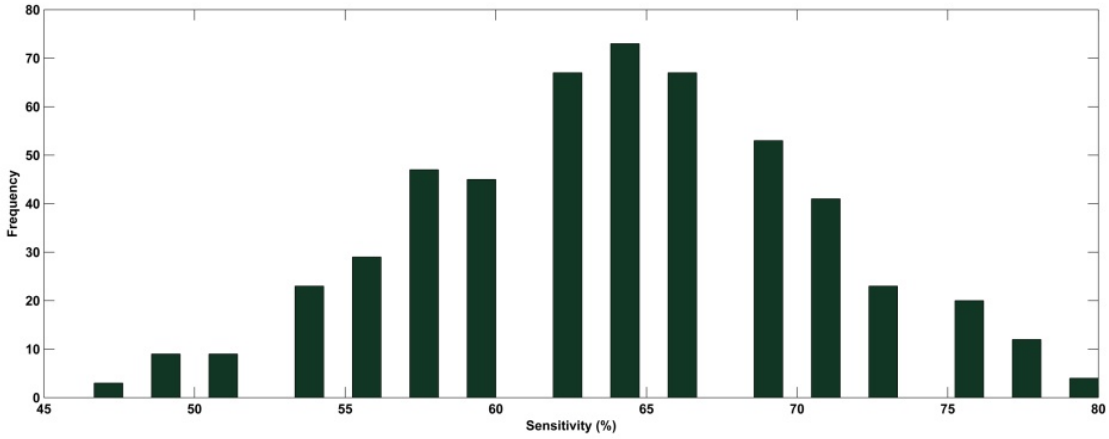
69	Centroid	1641-1693 vs. 1599-1635	0.1730
70	Centroid	1599-1635 vs. 1641-1693	0.1730
71	RMSN	1543-1585 vs. 1641-1693	0.1719
72	Peak Amplitude	1483-1537 vs. 1599-1635	0.1648
73	Kurtosis	1483-1537 vs. 1176-1242	0.1638
74	Peak Amplitude	1599-1635 vs. 1483-1537	0.1596
75	Centroid	1176-1242 vs. 1599-1635	0.1570
76	Centroid	1599-1635 vs. 1176-1242	0.1568
77	Kurtosis	1599-1635 vs. 1483-1537	0.1558
78	RMSN	1641-1693 vs.1543-1585	0.1557
79	Centroid	1641-1693 vs.1483-1537	0.1533
80	Centroid	1483-1537 vs. 1641-1693	0.1533
81	Peak Frequency	1483-1537	0.1525
82	Centroid	1176-1242 vs. 1483-1537	0.1484
83	Centroid	1483-1537 vs. 1176-1242	0.1483
84	Kurtosis	1483-1537	0.1479
85	Centroid	1483-1537	0.1400
86	RMSN	1641-1693 vs. 1483-1537	0.1372
87	Kurtosis	1543-1585 vs. 1176-1242	0.1341
88	Peak Frequency	1020-1115 vs. 1483-1537	0.1266
89	Peak Amplitude	1176-1242 vs. 1483-1537	0.1251
90	RMSN	1599-1635 vs. 1641-1693	0.1248
91	Kurtosis	1543-1585	0.1228
92	Centroid	1176-1242 vs. 1543-1585	0.1221
93	Centroid	1543-1585 vs. 1176-1242	0.1221
94	Peak Frequency	1020-1115 vs. 1543-1585	0.1213
95	Peak Frequency	1020-1115	0.1201
96	Kurtosis	1176-1242 vs. 1543-1585	0.1173
97	Peak Frequency	1020-1115 vs. 1599-1635	0.1122
98	Kurtosis	1641-1693 vs. 1176-1242	0.1115
99	Peak Amplitude	1483-1537 vs. 1176-1242	0.1090
100	Peak Frequency	1599-1635 vs. 1483-1537	0.1073
101	Kurtosis	1543-1585 vs. 1483-1537	0.1072
102	RMSN	1641-1693 vs. 1599-1635	0.1061
103	Kurtosis	1176-1242 vs. 1641-1693	0.1057
104	Kurtosis	1641-1693 vs. 1543-1585	0.1033
105	Kurtosis	1641-1693	0.1024
106	Kurtosis	1641-1693 vs. 1483-1537	0.0978
107	Kurtosis	1176-1242	0.0969

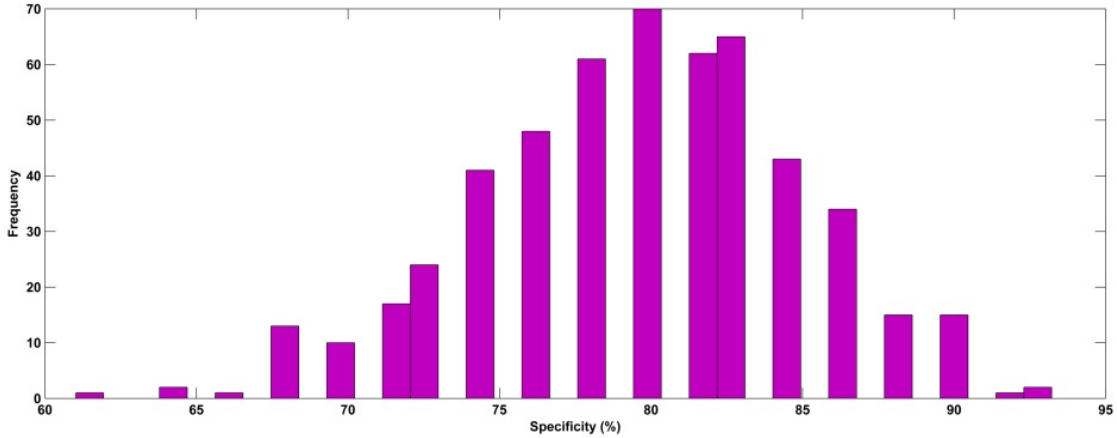
108	Peak Frequency	1483-1537 vs. 1020-1115	0.0961
109	Peak Frequency	1543-1585 vs. 1020-1115	0.09574
110	Peak Frequency	1599-1635 vs. 1020-1115	0.0950
111	Peak Frequency	1176-1242 vs. 1020-1115	0.0946
112	Peak Frequency	1483-1537 vs. 1599-1635	0.0944
113	Kurtosis	1483-1537 vs. 1641-1693	0.0927
114	Peak Frequency	1641-1693 vs. 1020-1115	0.0883
115	RMSN	1543-1585	0.0878
116	Kurtosis	1020-1115 vs. 1543-1585	0.0862
117	Skew	1599-1635 vs. 1176-1242	0.0851
118	Peak Amplitude	1483-1537	0.0805
119	Peak Amplitude	1483-1537 vs. 1641-1693	0.0789
120	RMSN	1599-1635	0.0755
121	RMSN	1641-1693	0.0754
122	Kurtosis	1020-1115 vs. 1483-1537	0.0712
123	Peak Amplitude	1641-1693 vs. 1483-1537	0.0695
124	Kurtosis	1020-1115 vs. 1599-1635	0.0676
125	Kurtosis	1543-1585 vs. 1599-1635	0.0662
126	Kurtosis	1599-1635 vs. 1543-1585	0.0647
127	Kurtosis	1020-1115 vs. 1641-1693	0.0637
128	Skew	1599-1635	0.0632
129	Kurtosis	1543-1585 vs.1599-1635	0.0631
130	Centroid	1641-1693	0.0615

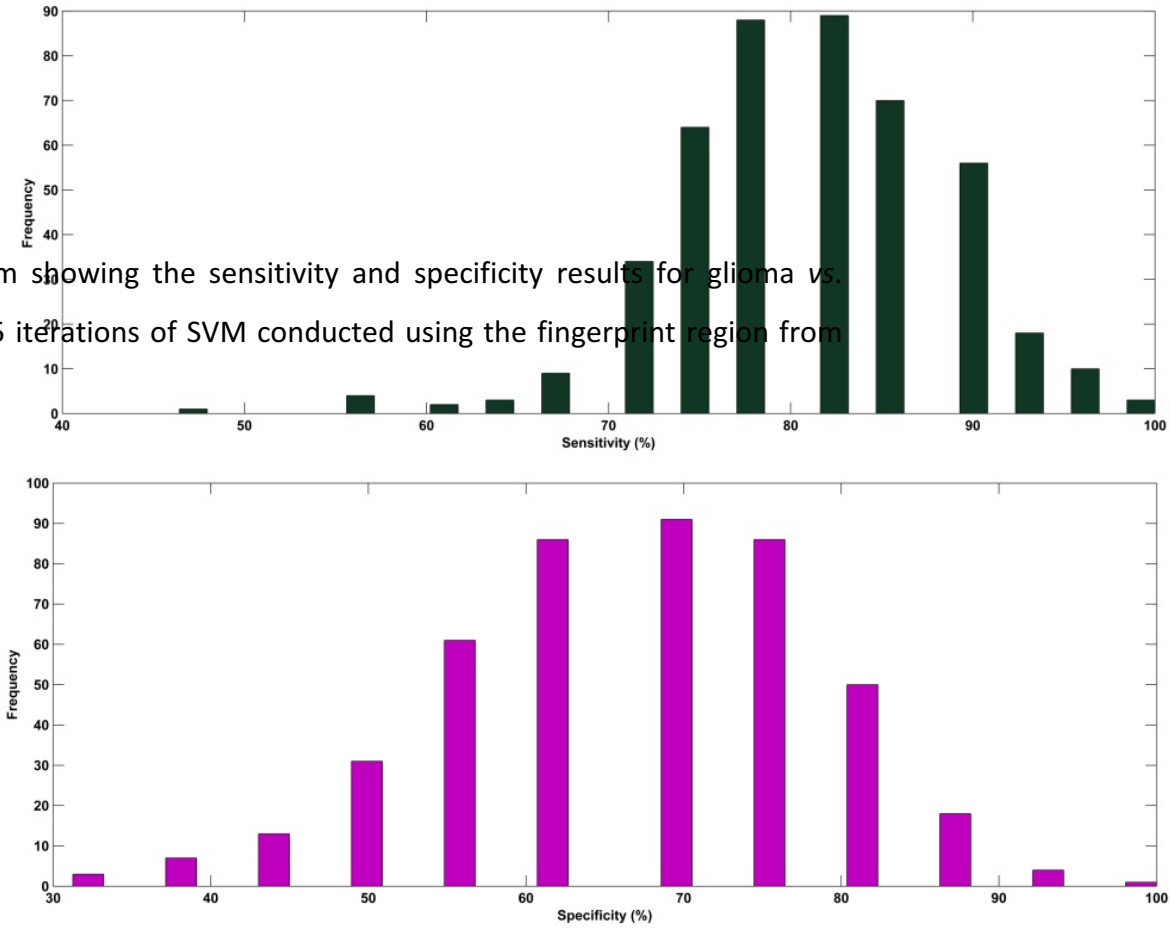
gure 1 -Histogram showing the sensitivity and specificity results for metastatic ncer *vs*. brain cancer for 525 iterations of SVM conducted using the fingerprint gion from 1800-1000 cm^{-1}.

Appendix 4

Sensitivities and specificities of the strata discussed in Chapter 6







ıre 2 - Histogram showing the sensitivity and specificity results for glioma *vs.* ıingioma for 525 iterations of SVM conducted using the fingerprint region from 0-1000 cm^{-1}.

Figure 3 - Histogram showing the sensitivity and specificity results for HGG *vs.* LGG for 525 iterations of SVM conducted using the fingerprint region from 1800-1000 cm^{-1}.

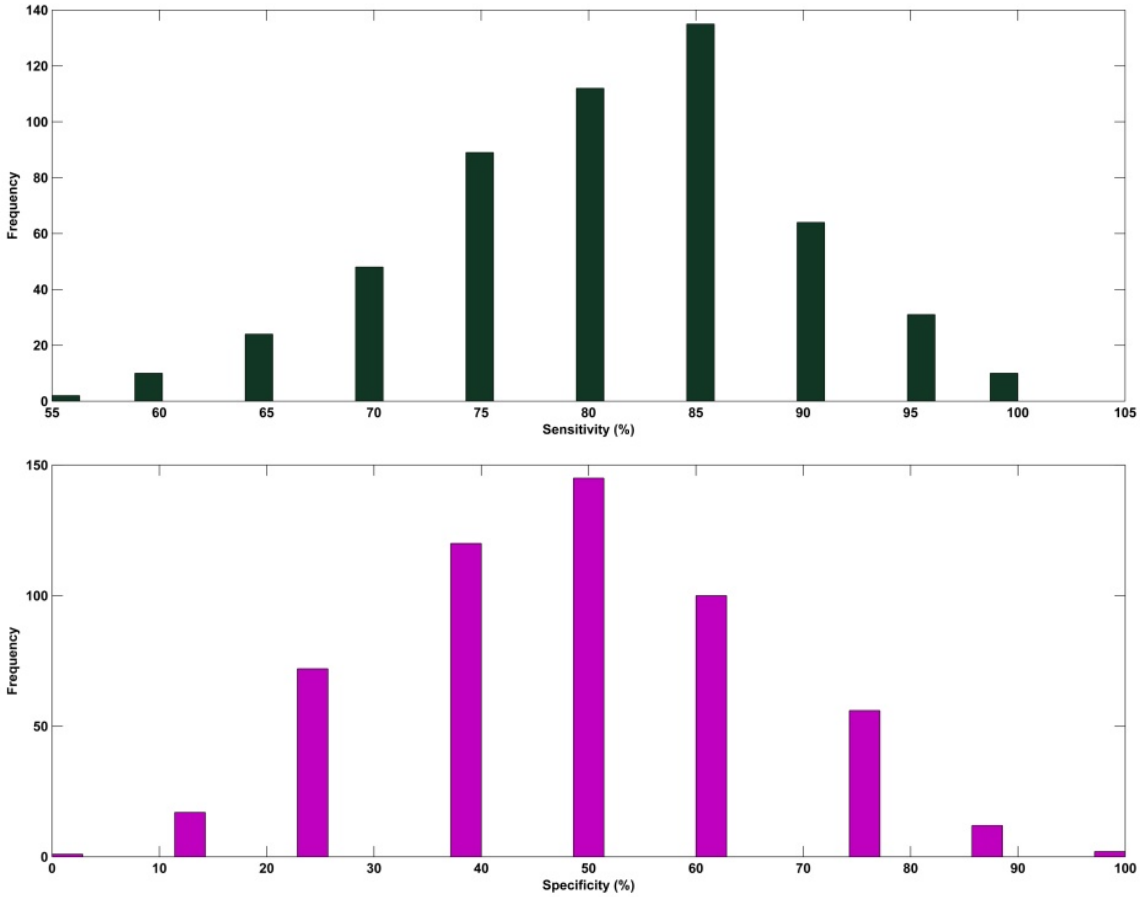


Figure 4 Histogram showing the sensitivity and specificity results for breast metastatic origin for 525 iterations of SVM conducted using the fingerprint region from 1800-1000 cm^{-1}.

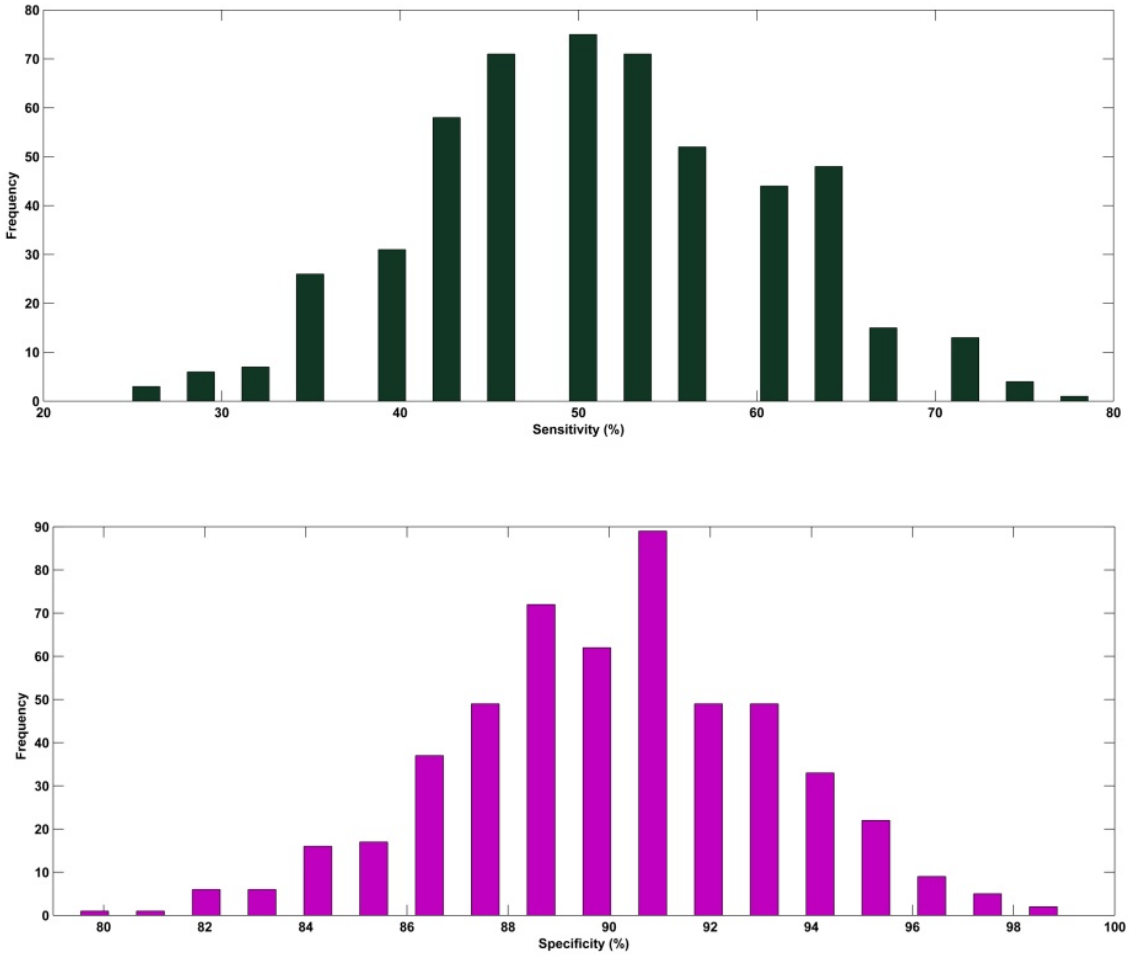


Figure 5 - Histogram showing the sensitivity and specificity results for skin melanoma) metastatic origin for 525 iterations of SVM conducted using the fingerprint region from 1800-1000 cm^{-1}.

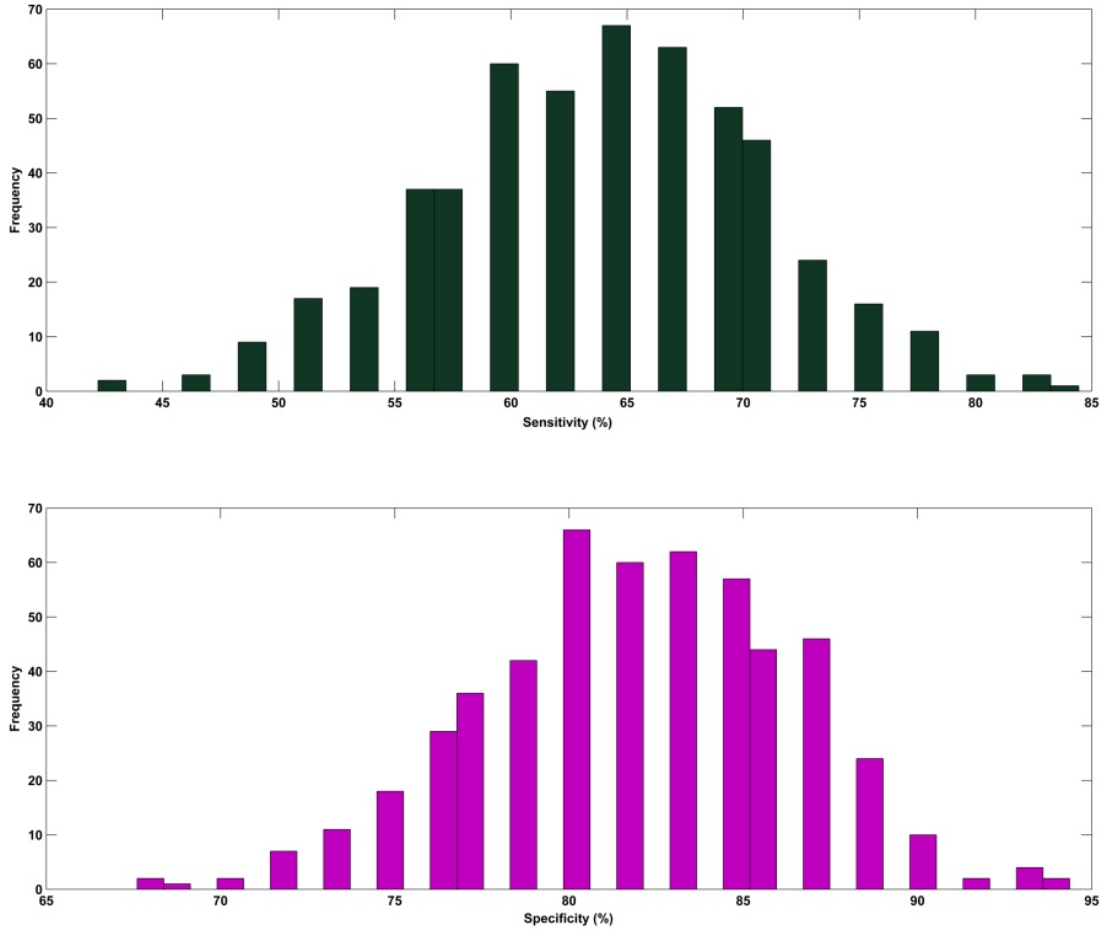
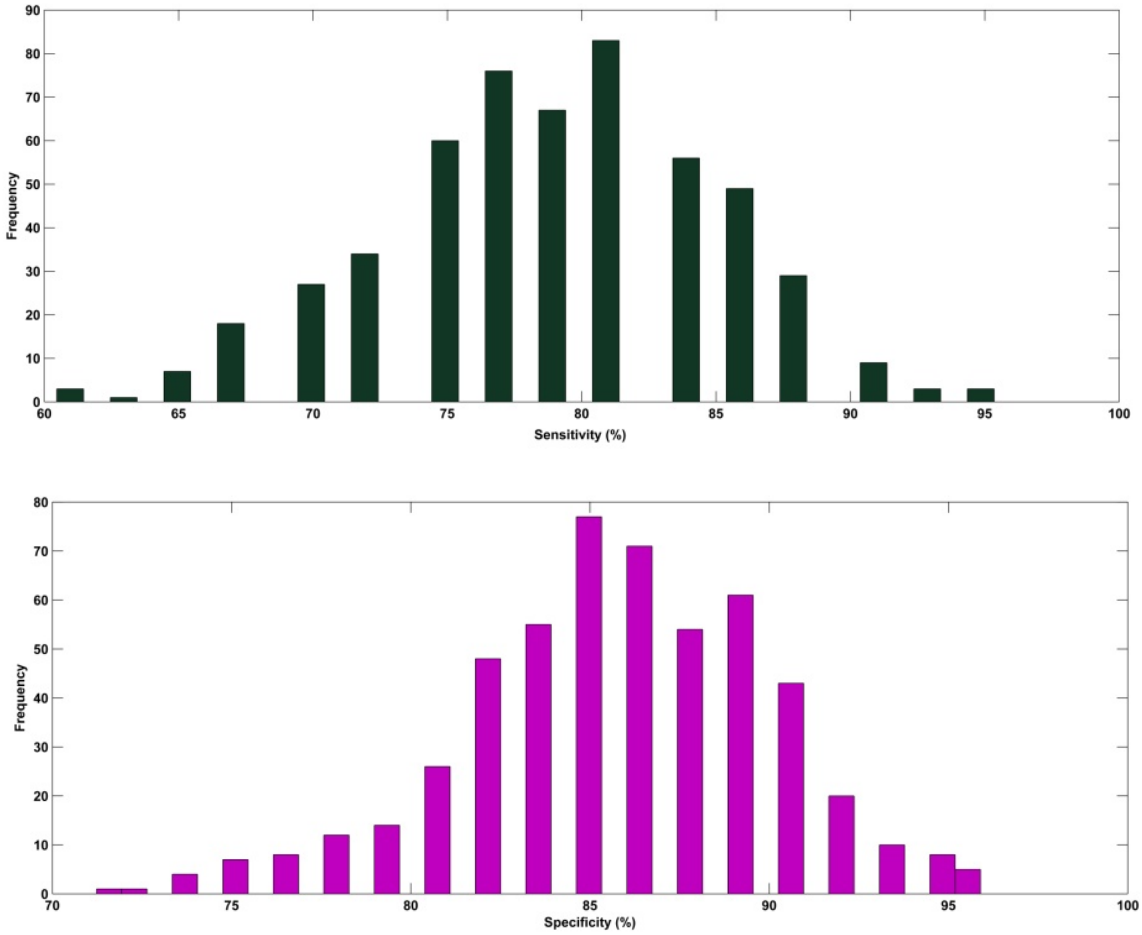


Figure 6- Histogram showing the sensitivity and specificity results for lung metastatic origin for 525 iterations of SVM conducted using the fingerprint region from 1800-1000 cm^{-1}.



Appendix 5

AWARDS AND PRESENTATIONS

Awards and Honours

- Royal Society of Chemistry – 2015 Analytical Twitter Poster Conference – Runner-up prize
- National Health Service (NHS) – 2014 Research and Innovation Showcase – 1st place poster prize
- 2014 Federation of Analytical Chemistry and Spectroscopy Societies (FACSS) Student Award – *awarded in September 2014 at the SciX conference, NV, USA*
- National Health Service (NHS) – 2013 Research and Innovation Showcase – Runner-up prize
- British Neuro-oncology Society Annual Meeting (BNOS) 2013 – Best Clinical Poster prize

Oral Presentations

9) Illuminating the Future of Cancer Diagnosis via Serum ATR-FTIR Spectroscopy
Hands, J .R., Clemens, G., Baker, M. J. *PITTCON Conference & Expo 2015*, March 2015

8) Illuminating the Future of Cancer Diagnosis via Serum ATR-FTIR Spectroscopy
Hands, J .R., Clemens, G., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *Federation of Analytical Chemistry and Spectroscopy Societies (FACSS) presents SciX2014 (The Great Scientific Exchange) Conference*, September 2014

7) Illuminating the Future of Cancer Diagnosis via Serum ATR-FTIR Spectroscopy
Hands, J .R., Clemens, G., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Stables, R., Baker, M. J. *SPEC 2014 – Shedding New Light on Disease*, August 2014

6) The Rapid Diagnosis of Gliomas via Serum Spectroscopy
Hands, J .R., Dorling, K. M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *British Neuropathological Society 115th Meeting*, Institute of Child Health, University College London, March 2014

5) Illuminating Cancer Diagnosis: The Rapid Orthogonal Diagnosis of Cancer
Hands, J .R., Dorling, K. M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *Changing Attitudes to Cancer*, University of Central Lancashire, September 2013

4) Rapid Diagnosis of Gliomas using Serum ATR-FTIR and Raman Spectroscopy
Hands, J .R., Dorling, K. M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *International Conference on Advanced Vibrational Spectroscopy 7*, August 2013

3) Illuminating Cancer Diagnosis: The Rapid Orthogonal Diagnosis of Gliomas
Hands, J .R., Abel, P. M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *British Neuro-oncology Society Annual Meeting*, July 2013

2) Illuminating Cancer: Rapid Diagnosis of Gliomas via Serum Spectroscopy
Hands, J .R., Abel, P. M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J*. Graduate Research School Annual Research Conference 2013* - University of Central Lancashire, July 2013

1) The Use of ATR-FTIR and Raman to Diagnose Brain Cancer from Serum
Hands, J. R., Baker, M. J, *SPEC 2012*, Chiang Mai, November 2012

POSTER PRESENTATIONS

7) Rapid Stratified Serum Spectroscopic Diagnostics
Hands, J .R., Clemens, G., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Stables, R., Baker, M. J. *International Conference on Advanced Vibrational Spectroscopy (ICAVS8)* – 2015

6) Rapid Brain Cancer Diagnosis: A Stratified Approach via Serum Spectroscopy
Hands, J .R., Clemens, G., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *Clinical Raman and Infrared Spectroscopy Network Summer School*, July 2015

5) Rapid Brain Cancer Diagnosis: A Stratified Approach via Serum Spectroscopy
Hands, J .R., Clemens, G., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *Clinical Raman and Infrared Spectroscopy Network Conference,* Exeter, April

4) Light Fantastic: Rapid Diagnosis of Gliomas via Stratified Serum Spectroscopy
Hands, J .R., Clemens, G., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J.*NHS Healthcare Roadshow*, Royal Preston Hospital, November 2014

3) Investigating the Use of the Spero Quantum Cascade Laser (QCL) Infrared Micro-spectrometer for the Rapid Screenings of Cancerous Blood Serum
Clemens, G., Bird, B., Weida, M., Hands, J., Barre, M., Baker, M. J. *Federation of Analytical Chemistry and Spectroscopy Societies (FACSS) presents SciX2014 (The Great Scientific Exchange) Conference*, September 2014

2) Light Fantastic: Rapid Diagnosis of Gliomas via Serum Spectroscopy
Hands, J .R., Dorling, K., Abel, P., M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *NHS Healthcare Roadshow*, Royal Preston Hospital, November 2013

1) Light Fantastic: Rapid Diagnosis of Gliomas via Serum Spectroscopy
Hands, J .R., Dorling, K. M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J. *FTIR Spectroscopy in Microbiological and Medical Diagnostics Workshop*, Robert-Koch Institute, October 2013

Appendix 6

Investigating the Rapid Diagnosis of Gliomas from Serum Samples using Infrared Spectroscopy and Cytokine and Angiogenesis Factors

Hands, J. R., Abel, P., Ashton, K., Dawson, T., Davis, C., Lea, R. W., McIntosh, A. J., Baker, M. J.

Analytical and Bioanalytical Chemistry, 405(23), 2013, 7347-7355

Appendix 7

Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) Spectral Discrimination of Brain Tumour Severity from Serum Samples

Hands, J .R., Dorling, K., Abel, P., M., Lea, R. W., Ashton, K. M., Dawson, T., Jenkinson, M. D., Brodbelt, A., Davis, C., Walker, C., Baker, M. J.

Journal of Biophotonics, 7(3-4), 2014, 189-199