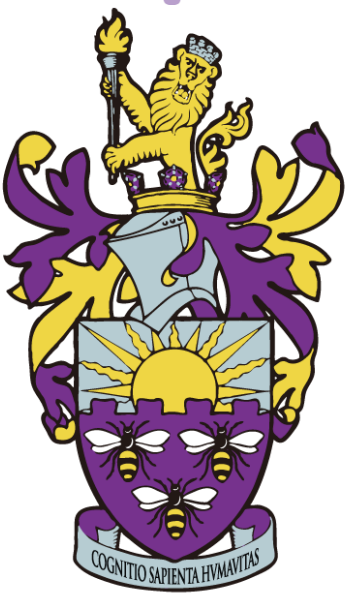
INFRARED SPECTRAL PATHOLOGY: THE INFLUENCE OF THE GLASS SUBSTRATE ON TISSUE CLASSIFICATION AND IMPLICATIONS FOR CLINICAL TRANSLATION

Jamie Lou (9424152)



SCHOOL OF CHEMICAL ENGINEERING AND ANALYTICAL SCIENCE
THE UNIVERSITY OF MANCHESTER 2017-2018
SUPERVISOR
Professor Peter Gardner

ACKNOWLEDGEMENTS

I would like to thank my supervisor Peter Gardner for his guidance and support throughout the project. The opportunity to be involved in leading cancer research has been truly inspiring and I have thoroughly enjoyed working with the Gardner group. I am grateful for everyone in the Gardner group who assisted me during my project. I would like to especially thank Nga-Tsing Tang and Jiayi Tang for guiding me through the MATLAB coding during my data analysis. Finally, I would like to thank the Christie Cancer Research Centre for preparing the samples and the patients who kindly donated their samples in aid of cancer research.

ABSTRACT

Globally, the number of people affected by cancer is increasing and the increasing number of biopsies required are putting immense strain on health care services and pathologists whilst resources are limited. This demands for a more efficient and effective way to manage these biopsies to improve the health care service and provide better support for patients.

Fourier-Transform Infrared Spectroscopy is an objective method with diagnostic capabilities proposed as a cancer pre-screening tool. However, infrared spectroscopy traditionally employs infrared inactive fluoride substrates as they are transparent across the useful infrared range. However, such substrates are unsuitable for use in clinics due to their cost and fragile nature and so cannot be implemented within an automated process. Glass has been proposed as an alternative given it being inexpensive, robust and familiarity amongst pathologists and clinics.

An issue surrounding glass is the limited window of wavenumbers where infrared is not absorbed, though studies by Bassan *et al.* determined that this window was sufficient to successfully classify tissue and differentiate between cell types. Other confounding factors such as the degree of staining have also been investigated with Pilling *et al.* concluding that degree of staining does not influence classification. Although promising, not all glass slides are alike and the variance in optical properties in glass slides has not yet been investigated. Therefore, before FT-IR can be introduced into clinics, the variations in optical properties of glass need to be examined and determine whether the differences, if any, affect the classification of tissue within the Random Forest algorithm.

The study examined 12 different plain glass substrates followed by sections of prostate tissue with benign prostate hyperplasia. The initial study on the plain glass found that there were some minor differences between the glass substrates, although following PCA, majority of the glass slides could not be separated with the exception of the Superfrost substrate. The root of the differences are potentially a result of manufacture processing techniques to produce the 'extra low-iron content' slide for improved transparency. PCA was also performed comparing adhesive coated slides against non-coated slides aiming to determine whether the adhesive coating would affect the spectra. The results concluded that the adhesive coating had no influence on the spectra with assumptions drawn that the coating is too thin to absorb any substantial infrared.

For the analysis of the tissue samples, a classifier was built using two operator annotations. Training was performed on the Academy and Fisherbrand slides whilst independent testing was done on the Superfrost slide. The training models provided reasonable AUC values (0.894, 0.851) and training accuracies averaging 77.5% for both operators. However, when compared to literature these results were sub-par in comparison. Assessing the performance using the independent test, a maximum AUC value of 0.525 and accuracies averaging 46% indicate poor performance. The findings of the study are inconclusive as there may be indications to suggest the Superfrost slide as an anomaly and unconfident operator annotations did not allow for a robust enough model to be built. The requirement for substrate standardisation is still undetermined and the use of FT-IR with glass substrates to complement pathologists is still very much possible. However, further studies are required using a wider range of glass substrates to fully determine the effects of glass on tissue classification. Future work studying effects of factors such as classifier parameters and coverslipping are needed before infrared spectral histopathology using glass substrates can be clinically translated.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Cancer caused an estimated 8.8 million deaths in 2015, contributing to approximately 1 in 6 of all deaths worldwide, with men and women mostly commonly diagnosed with prostate and breast cancer respectively. [1] Ageing populations, changes in lifestyles and advancements in healthcare are likely to increase the number of cancer diagnoses, calling for more effective and cheaper treatment to reduce strain on healthcare services with limited resources.

Early diagnosis is critical to the success rate of cancer treatment, with delayed diagnosis directly linked with reduced chances of survival. Prostate cancer has a 100% survival rate beyond 5 years if diagnosed at the earliest stage compared to less than 1 in 3 when diagnosed at the latest stage. Similarly, almost all women with breast cancer detected in its earliest stage survive beyond 5 years compared to only 3 in 20 at its latest stage. [2]

Currently, cancer diagnosis is done via histopathological imaging which requires a trained pathologist to manually inspect and analyse biopsy samples through a microscope. Often, this is a time consuming and labour-intensive process where typically only 20% of breast biopsy samples are cancerous [2]. Concerns are also raised due to its inherent subjectivity with research performed by M. Van den Bent on interobserver variations within histopathological diagnosis of glioma which is also supported by Ismail *et al.* studying observer variations in cervical intraepithelial neoplasia diagnosis. The findings indicate interobserver variations is a true concern with disagreements in gradings having potential clinical impact. The studies direct the causes of disparities to variations in sample quality, training and experience of pathologists, technicalities and ambiguous grading descriptions with subjective terms [3] [4]. In order to mitigate the short-comings of traditional pathology, an objective diagnostic tool is required.

Fourier-Transform infrared (FT-IR) spectroscopy has been proposed as a potential tool to improve cancer screenings. Infrared spectroscopy is an analytical technique which utilises infrared radiation to excite molecules and create vibrational transitions as the radiation is adsorbed. FT-IR has the potential to provide a fast, high throughput and objective method for classifying tissue. However, expensive substrates in the form of calcium/barium fluoride

which are traditionally used for spectroscopic imaging cannot be implemented in an automated process or in clinics due to their cost and fragility. Infrared spectroscopy operating in transflection mode using MirrIR slides (low electron/low-e) have been considered as an alternative considering its low cost, improved rigidity and improved absorbance. Despite its benefits, a problem arises with transflection measurements which would hinder its potential for clinical sampling as artefacts from a phenomenon known as the electric field standing wave (EFSW) and light dispersion could distort spectra and lead to incorrect classification and hence, a mistaken diagnosis. The presence of the phenomenon is currently debated between researchers with studies performed by Perez-Guaita *et al.* and Bassan *et al.* supporting the presence of the phenomenon. The former recognises that the artefacts are small and consistent whilst Bassan explicitly cautions the use of such substrates for transflection mode experiments [5] [6]. Furthermore, although the effect can be reduced by ensuring a consistent tissue sample thickness, achieving this is unrealistic due to variations in accuracy of microtomes, operator experience and methodologies between hospitals. The impact of the phenomenon is still not fully understood and therefore MirrIR is unsuitable for clinical use.

Conventional glass slides operating in transmission mode have been proposed as an alternative as they are much more affordable, considerably more robust and pathologists and clinics have an established history with using glass. Previously, glass would have been considered unsuitable for tissue classification as it is infrared active in the useful 'fingerprint' region where most biological information lies. Only until recently, Bassan *et al.* discovered a small window in the high wavenumber region of the infrared spectrum (3800-2500 cm^{-1}) which is not adsorbed by glass and has been used to successfully classify tissue [7]. Despite this, there are still uncertainties which need to be answered before introducing the technique as a diagnostic tool. Although most microscope slides are produced from soda-lime glass, not all glass slides are created equal and variations in chemical composition, structure and physical properties could impact the optical properties of the glass. Since exact compositions are unknown and trace impurities along with different processing techniques may have a significant effect on the spectra, this could ultimately affect classification. A classifier built without this consideration could lead to erroneous diagnosis resulting in over or undertreatment and be potentially fatal and such cases should be avoided at all costs. Since clinics source their microscopes slides from different

manufacturers, FT-IR spectroscopy using glass slides as a diagnostic tool cannot be clinically translated until the influence of glass on the infrared spectrum is determined.

The following report will aim to determine whether different glass substrates affect the spectra of tissue samples and whether the differences, if any, affect the final classification of the tissue within the Random Forest algorithm. The aim is to determine whether glass substrates are suitable for use in clinics and the implications towards clinical translation.

2 AIMS AND OBJECTIVES

The aim of the project is to determine whether tissue samples observed on glass slides from different manufacturers exhibit spectral variations which may affect the histological classification of the sample when assessed using the Random Forest algorithm. Spectral differences may occur due to variations in optical properties which arise from natural variations in the structure and/or the chemical composition of the glass.

At the end of the project, the following objectives should be accomplished:

- Perform FT-IR imaging on 12 different glass slides from 7 different manufacturers (Academy, Fisherbrand, Kindler, Sailbrand, VWR, Citoglas and Marienfeld)
- Complete FT-IR imaging of prostate tissue samples with benign prostatic hyperplasia (BPH) on each of the 12 slides.
- Process the raw data using MATLAB and perform principal component analysis (PCA) to the plain glass slides spectra.
- Apply the Random Forest algorithm to spectra obtained from the tissue samples to build a classifier to distinguish between epithelial and stromal cells.

Initially, plain glass substrates will be examined to determine whether the glass exhibit uniform spectra. The experiment will highlight whether additional qualities (such as a poly-L-lysine coating) or further processing (to reduce impurities) has an influence on the spectra and identify key variables that may separate one glass from another.

Tissue is comprised of various types of biological material and is considerably complex. As the classifier is highly sensitive, subtle changes in the spectra of tissue as a result of the substrate can have a significant impact in determining the histological class. The experiment aims to establish whether tissue samples across each slide produce a uniform

spectrum and whether the histological class can be determined with a high degree of accuracy, >95%. If successful, this will eventually be used to identify malignant cells using certain biological markers in suspected cancer patients.

The spectra for both plain glass and tissue samples are processed within MATLAB (The Mathworks, Natick, MA) and PCA is performed on the data sets. PCA enables differences to be emphasised by focusing on the main components which separate the data.

To build the classifier, a fraction of the data from the tissue samples will be used to train a classifier using the Random Forest algorithm, with the aim to distinguish between epithelial and stroma cells. The data not used for training will then be used to validate the classifier and subsequently an independent test will be performed on the model using untrained data. The data will then be processed and analysed with conclusions drawn from results.

The results gathered will be used to confirm the operability of FT-IR spectroscopy as a diagnostic tool which is integral to the advancement of introducing FT-IR spectroscopy in clinics and other medical applications.

3 METHODOLOGY

The experiment began with hyperspectral image collection, followed by data processing, PCA and finally building the classifier. Whilst obtaining the spectra, the sample was placed in a purge box using nitrogen to remove atmospheric water vapour and carbon dioxide. It was common practice to ensure the relative humidity reached 0% before measurements commenced as water vapour interferes with the spectra. The temperature was also recorded and remained between 22.5-23°C although this parameter and the difference is unlikely to influence the spectra.

3.1 PREPARATION OF SAMPLES

3.1.1 Glass slides

A selection of 12 glass slides (approx. 75x25x1mm) from different manufacturers (Academy, Fisherbrand, Sail Brand, Citoglas, Marienfeld, O.Kindler and VWR) will be analysed. The glass slides were prepared by cleaning the surface with propanol and wiped down with lint free tissue. Three randomly selected areas (given they were free of any non-

biological objects) were selected on each of the glass slides to be measured to allow for an average spectrum for each glass to be produced.

3.1.2 Tissue samples

Prostate biopsy tissue samples with BPH were obtained via transurethral resection of the prostate which were fixed in formalin and embedded with paraffin. The samples were provided by Cancer Research UK Manchester Institute (The Paterson Laboratory) with prior consent and ethical approval under the supervision of Trent Multi-Centre Research Ethics Committee (01/4/061). A poly-L-lysine coating was applied to each slide before the tissue was placed onto the substrate. The samples were prepared from a tissue block which was microtomed into 4 µm contiguous sections and each sample was floated onto each individual slide. The slides followed a process of dewaxing using xylene followed by rehydration using graded ethanol and finally stained with haematoxylin and eosin (H&E).

Pertex (CellPath, Newtown, Powys, Wales, UK), a xylene based drying medium along with a #1.5 histological cover slips (50x24x0.16 mm) were used preserve the samples. The processed slides air dried for a minimum of 24 hours and cleaned with propanol and lint free tissue before measuring.

3.2 SPECTRAL IMAGING

Hyperspectral imaging will be done with the aid of an Agilent Varian Cary 670 FT-IR spectrometer attached with a Varian-620 infrared microscope. The instrument is equipped with a liquid nitrogen cooled mercury cadmium telluride (MCT) 128x128 FPA detector. The microscope utilises a x15 magnification Cassegrain optic, producing a 704x704 µm field of view at a nominal pixel size of 5.5x5.5 µm.

When performing each experiment, due to time constraints, considerations for spectral resolution, number of scans and scan range need to be applied. The equipment is capable of measuring at a spectral resolution of 4 cm^{-1} although the difference between 4-5 cm^{-1} is considered minimal so 5 cm^{-1} was used to reduce acquisition time. Two levels of zero-filling were used so the data spacing between each data point is 1.929 cm^{-1}. This provided high quality images within a reasonable amount of time.

Generally, the higher the number of scans, the better the signal to noise ratio (SNR) although this will increase the time required for sampling. For the plain glass slides, 128

background scans and 16 sample scans were co-added. A range of sample scans were tested which concluded with 16 scans providing an optimum in quality and acquisition time. This may be considered a low number of sample scans although considering the plain glass slides were assumed to be homogenous and did not contain complex spectra seen in biological samples, this was acceptable.

For the tissue samples, 256 background scans and 96 sample scans were co-added. Since the focus of the study was directly related to the spectrum of the tissue, it was paramount the spectra were of high quality and had an excellent SNR since the addition of noise could affect the classifier. Also, since the imaged area being examined was large, the time required to image the tissue compared to the time required for the extra scans was considered minimal. The tissue was mapped using individually scanned tiles which were then merged together to form the final mosaic, varying between 7x9 to 9x9 which consisted of 72 to 81 tiles. This process took approximately 3-4 minutes per tile.

The background scans for the plain glass slides were taken in air whilst the background for the tissue samples were taken on a clean area of the glass, away from the coverslip to avoid interference from the mounting medium, passing through the glass only.

The scan range was set between 3800-900 cm^{-1} which covers the entire mid-IR range where the molecules of biological interest lie. Although in the processing stages where the range is cut to 3800-2500 cm^{-1}, it is useful to measure the complete range for identifying potential anomalies or artefacts.

Throughout the experiments, the detector needs to be kept below 78K to reduce the amount of noise in the spectra which arise from the molecular vibrations of excited molecules; lowering the temperature decreases the energy of these molecules and hence reduces noise.

The experiment began by placing the sample onto the stage followed by calibrating the condenser and ensuring the optic was set to the x15 objective and the valve is in the open position to allow the light to pass through. The detector was then cooled with liquid nitrogen until the temperature fell below 78K. Once the detector was sufficiently cool, the lens needed to be focused which was done with an external control box and a live FPA image on the computer screen. The next step which concludes the setup was to find the

center-burst, which is also known as the zero-path-difference point and corresponds to the point where maximum interference is produced.

Once the humidity had reached 0%, a background scan was taken which takes a spectrum of the atmospheric components present within the purge box. This is automatically removed from the subsequent sample scans by the Agilent software. A new background scan was completed every time the purge box was opened as even small variations in atmospheric composition could alter the spectra. Once the background scan finished, the samples were ready to be measured. Using the external control box to move the stage and the live FPA imaging to find an area without any dirt or dust as non-biological material could influence the spectra. For the plain glass slides, the sampling step was repeated twice in random locations before sampling the next slide.

An experimental flow diagram detailing the experimental procedure is shown in Figure 1, the setup procedure was the same for both the plain glass and tissue samples.

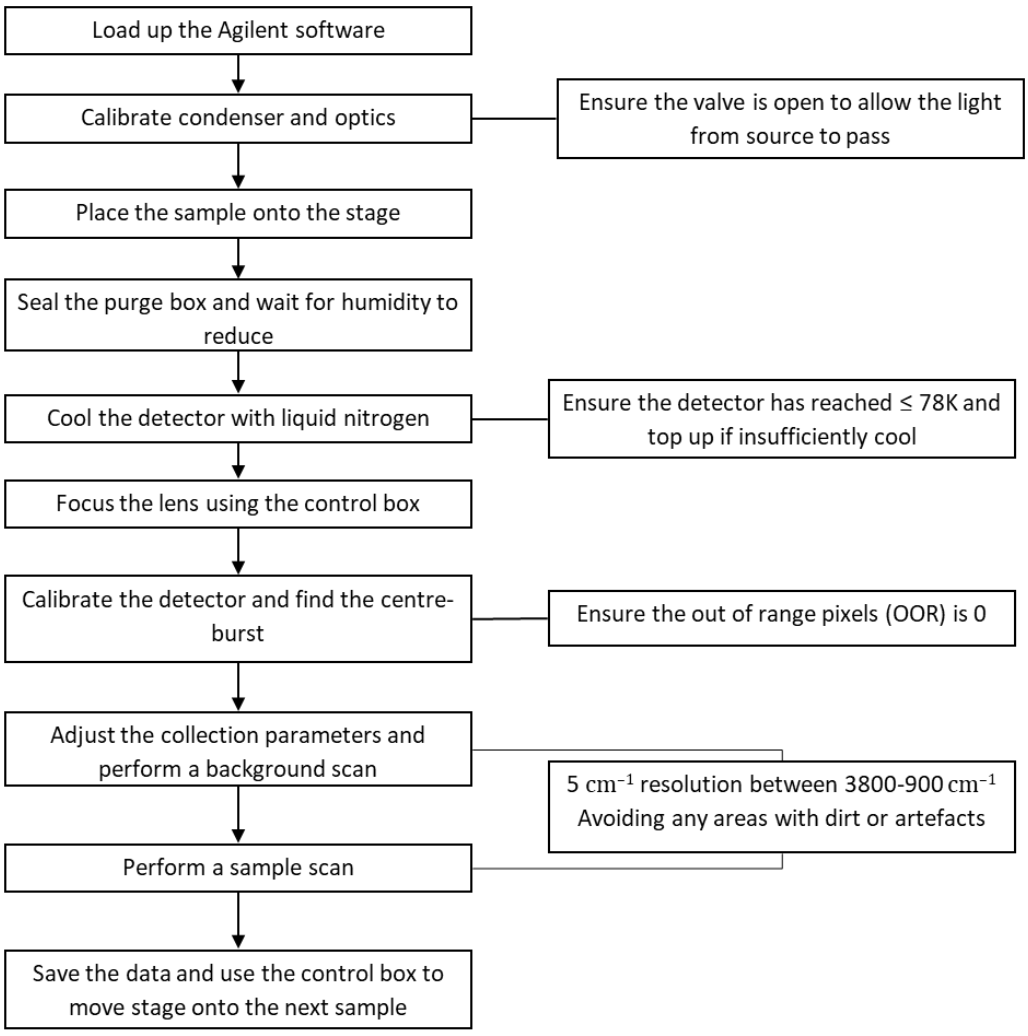


Figure 1 - Experimental flow sheet diagram

3.3 DATA PROCESSING

The raw data file (.seq/.dms format) containing the infrared hyperspectral images provided by the Agilent software required converting to a readable format before analysis was possible. The Centre for Hyperspectral Imaging (CHI) toolbox (Alex Henderson, University of Manchester, Manchester, UK) was used to perform this pre-processing step. Further details regarding the individual processing steps taken are detailed in the subsequent sections and theory behind PCA and Random Forest can be found in the project proposal.

3.3.1 Background removal & Happ-Genzel apodization

Background removal and Happ-Genzel apodization were automatically applied in the Agilent software. Background removal aims to remove the background spectra taken initially to remove spectra and noise that did not derive from the sample itself.

Apodization aims to reduce the ripples formed adjacent to a large peak in the form of smaller peaks and troughs which could be misconstrued as a functional group which is not actually present. A disadvantage of this method is that resolution and peak height are reduced. Although there are a variety of apodization functions available such as triangular and Gaussian, Happ-Genzel apodization was used as it provides the most suitable balance between resolution and suppression of ripples.

3.3.2 Noise reduction algorithm

Tissue samples are susceptible to low absorption due to several factors; operating in transmission mode, thickness of sample and relative size of sample. Coherent images with strong signal to noise ratio are required to identify changes in the spectrum based solely off biochemical variance and so therefore the noise must be reduced. Noise reduction is attained by performing PCA on the data set. By applying PCA, the significant components relating to the structural information are retained whilst the components related to noise are discarded. Figure 2 shows an example of a spectra pre and post noise reduction with the spectra of the noise removed on the right-hand side. The denoised spectra has been translated to enable differences to be discerned clearly. As the Figure shows, the spectra of the noise removed resembles noise and does not reflect any indication of biological material.

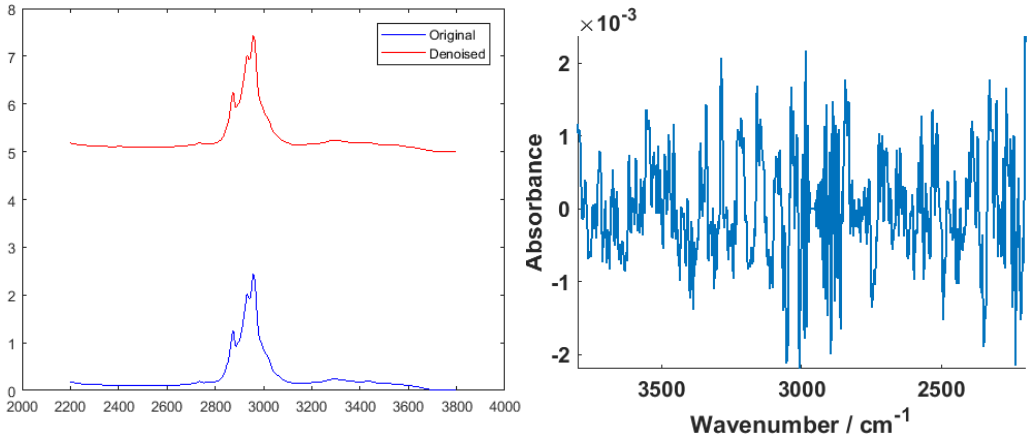


Figure 2 - Example of noise reduction, retaining 50 PC's in this example. Figure on the right shows the spectra of noise removed

3.3.3 Quality control

Tissue spectra underwent quality control prior to being analysed, spectra with intensities for the amide A band (3298 cm^{-1}) and the 3125 cm^{-1} trough outside the range of 0.02-1 were considered as non-biological material or noise and subsequently removed. This ensured that all the spectra in the data set resembled some form of biological material and removes spectra which exhibit scattering effects. This aims to improve the accuracy of the classifier as all the spectra should either correlate to stroma or epithelial. If training was built on non-biological data, the classifier might assign biomarkers which poorly reflect spectral features to differentiate between the two and lead to misclassification and reduced accuracy. This was particularly important since it was relatively difficult to completely avoid tissue free areas based upon the method of annotating the chemical image.

3.3.4 Savitzky-Golay 1st derivative

Savitzky-Golay, also known as digital smoothing polynomial filter is a form of digital data smoothing which is used on data with slow variable change and contains random noise. It fits the spectrum with a polynomial where the values are determined by replacing the raw value by a local average calculated by the surrounding points (19 points were used). This reduces the level of noise since the local 19 points measure roughly the same base value and strictly noise is removed through averaging [8]. By taking the first derivative, it enables peaks and troughs to be easily identified due to the sharp change in gradients at the maximum or minimum.

3.3.5 Vector normalisation (mean centred)

There is difficulty in achieving consistent sample thickness (particularly of the tissue and even the plain glass in some cases). Since differences in the thickness of samples will affect the absorption intensity as explained by the Beer-Lambert law, vector normalisation will be used to equalize differences in thickness to enable the variations in absorption to be fully appreciated. This emphasises the variations which are caused by biochemical changes as opposed to sample thickness. An example of pre and post-vector normalisation has been displayed in Figure 3.

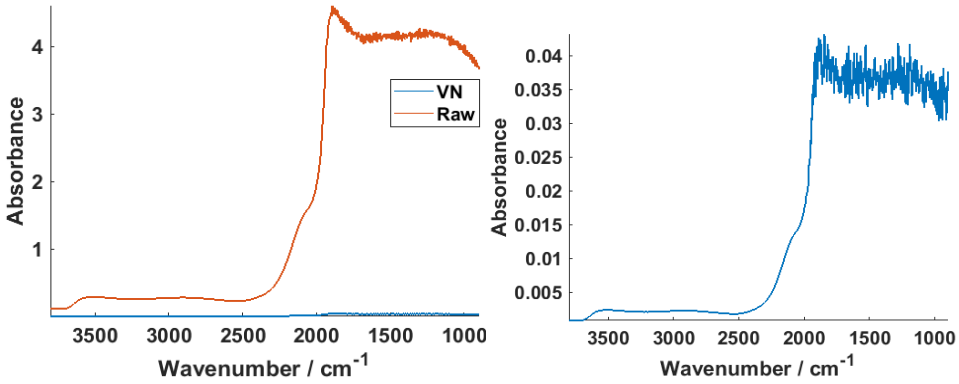


Figure 3 - Spectra of sample slide comparing raw data (orange) and vector normalised (blue). Figure on the right shows a zoomed in image of the normalised spectra

3.4 PLAIN GLASS SLIDES

Following the conversion to a readable format, the resulting Varian data cube produced a hyperspectral image with 16,384 spectra (128x128) at a corresponding 831 wavenumbers. The data for the three measured points on the glass were averaged to produce a single spectrum for each glass substrate. The data set underwent vector normalisation before absorbance was plotted against wavenumber.

3.5 TISSUE SAMPLES

A Nikon Eclipse 90i camera was used to image the tissue samples at 10x magnification, with an example of the imaged tissue sample shown in Figure 4. This was used as a reference guide to identify areas of epithelial cells and stroma during the annotations of the greyscale chemical images. The greyscale images were produced using the peak height of the amide A band (3298 cm^{-1}), normally the amide I band (1700-1600 cm^{-1}) will be used as it produces a better contrasting images although for glass this was out of range. On the chemical

images, pixels representing epithelial and stroma were annotated in green and red respectively.

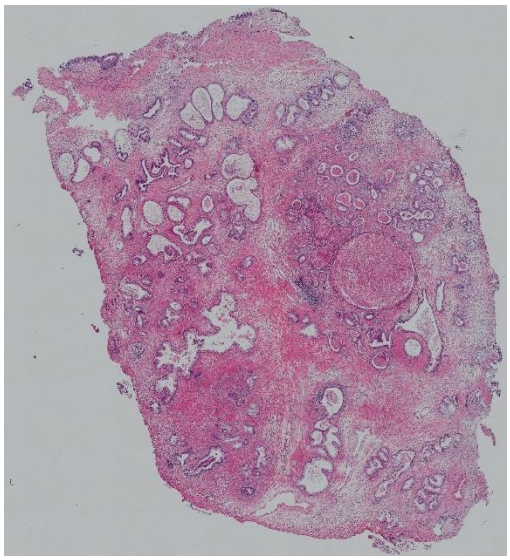


Figure 4 - Nikon brightfield image of tissue sample

Unfortunately, due to unforeseen circumstances with equipment and time constraints, the experiment was only performed on 3 slides from Academy, Fisherbrand and Superfrost (O.Kindler). Comparisons will be made against previous work to determine whether there are comparable classifier results as well as determining the effect of different operators on performance. Further analysis can be done by comparing annotations of the chemical images taken with a x15 objective and a x4 objective. The x4 objective has a considerably lower acquisition time as it can image up to 2.4x2.4 mm in a single tile compared to 700x700 µm, greatly increasing throughput. An additional benefit is reduced disk usage space (approx. 1/10 disk usage), minimising potential for file corruptions and faster processing [9].

Each coloured in pixel from the annotated chemical image representing an individual spectrum was imported into MATLAB and separated into epithelial and stroma. Quality testing was performed on the spectra based off the amide A intensity to remove areas free of tissue or spectra which suffered from high levels of scattering, the range of intensities was set between 0.02-1. Further processing involved retaining only the 3700-3125 cm^{-1} range as outside this region there is generally little or no diagnostic information and removes spectra associated with the mounting medium.

Table 1 - Number of epithelial and stroma spectra imported from each slide, O1 represents operator 1.

Class	Academy	Fisherbrand	SuperFrost
Epithelial O1	40,953	44,937	46,907
Stroma O1	21,524	15,540	21,892
Epithelial O2	57,642	55,572	53,588
Stroma O2	39,077	29,889	35,660

The number of spectra imported from the chemical images for each class and slide are listed in Table 1. In order to build a non-biased classifier, training spectra from both classes must be equal [10]. Therefore, random under-sampling was used where spectra were randomly selected from the from the larger class to meet this requirement. For the training, 90% of the selected spectra were randomly allocated for training whilst the remaining 10% were used for validation.

Noise reduction was then applied to the data set retaining 50 principal components, this value was chosen as it was the value at which the spectra of the noise removed did not resemble any biological information and was strictly noise. This avoids removing excess biological information if too little PC were used and retaining too much noise if too many. Following this step, vector normalisation and the 1st derivative Savitzky-Golay smoothing was applied.

The number of trees in the forest were set to 500 trees with the number of variables set to 30 where the variables are randomly selected wavenumbers which provide the best separation of the nodes to enable a higher classification accuracy. Node size was set to one to allow the trees to grow as large as possible at the expense of speed of training. The number of trees were determined using an out-of-bag error method. This method determines the optimum number of trees before the classification error plateaus and the error does not decrease with increasing number of trees. Figure 5 details this, with the ideal number of trees set at a liberal 500.

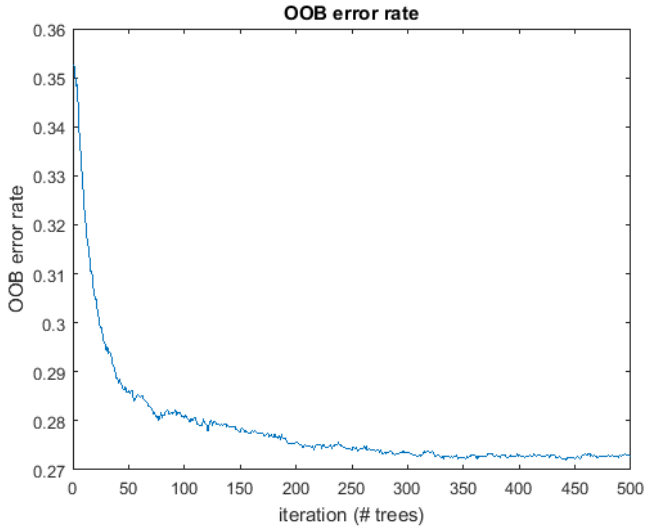


Figure 5 - Out of bag error rate for number of tree determination

No acceptance threshold was set for the validation of the classifier, so the spectra were classified as either epithelial or stroma even if the trees are a 49-51% split.

After validation, the classifier performance was inspected using an independent test with the aim to determine whether it can predict the class on the other substrate. The processing procedures leading up to this step was repeated on the untrained slide, however, since the model was built using an unbiased approach, all the spectra from the independent test was used for testing and there was no need to randomly sample the data. The order of spectra being classified was randomised to prevent the model from predicting the number of classes present in the data set.

4 RESULTS AND DISCUSSION

Table 2 –Label & manufacturer details for the glass substrates

Label	Manufacturer (Cat. No)	Description	Material, with composition (if known)
A	Fisherbrand (1238-3118)	76x26mmx1.0-1.2mm, ground edges, 90°	Soda-lime glass
B	Academy (N/A 131)	76x26mmx1.0-1.2mm, ground edges, 45°	Clear glass
C	VWR (100408)	76x26x1 mm, cut edges	Sheet glass
D	Kindler SuperFrost (8037/1)	26 x 76x1 mm, ground edges, 45°	Extra white soda lime silicate glass with very low iron content
E	Citoglas (10127105P)	75x25mmx1.0-1.2mm, single frosted, ground edges, 45°	Extra-white glass/pure white glass
F	Citoglas (10127105P-G)		Soda-lime glass, 70-73% SiO_2, 13-15% Na_2O, 7-12% CaO, 1-4.5% MgO, 1-2% Al_2O_3, 0-0.3% SO_3, 0.05-0.2% Fe_2O_3
G	Citoglas (1A5105)		
H	Sail Brand (7105)	25.4 x 76.2mmx1.0-1.2mm, ground edges, single frosted	Clear glass
I	Sail Brand (7101)	25.4 x 76.2mmx1.0-1.2mm, ground edges	Clear glass
J	Marienfeld Histobond (0810001)	26mmx76mmx1mm, ground edges, twin frosted	Soda-lime glass adhesive coated, SiO3 72.3 %, Al2O3 0.5 %, Fe2O3 <0.02 % Na2O 13.3 %, CaO 8.8 %, K2O 0.4 %, MgO 4.3 %
K	Marienfeld Histobond + (0810401)	26mmx76mmx1mm, ground edges, single frosted	
L	Marienfeld Histobond S+ (0810501)		

4.1 GLASS SLIDES

The mean spectra for each of the plain glass substrates were plotted and compared with one another as shown in Figure 6. The graph provides an initial preview of any obvious differences and the effect of different glass on the spectra. This will aid future evaluations when analysing the tissue samples.

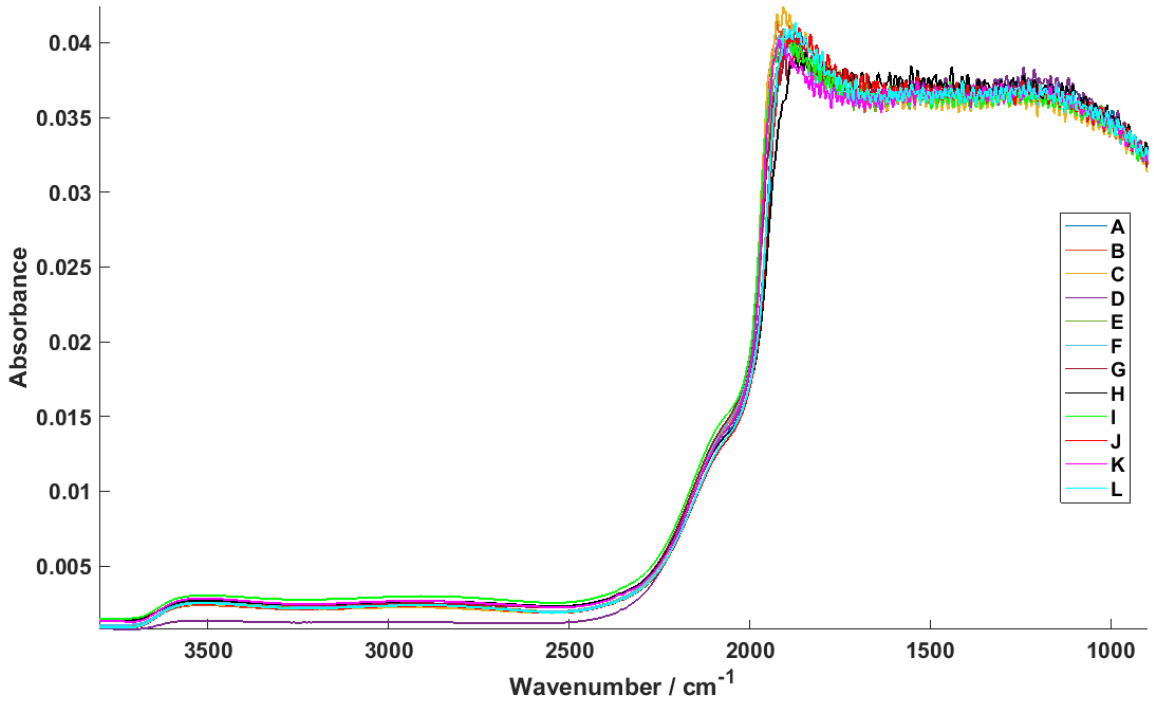


Figure 6 - Vector normalised spectra of the 12 glass substrates A-L

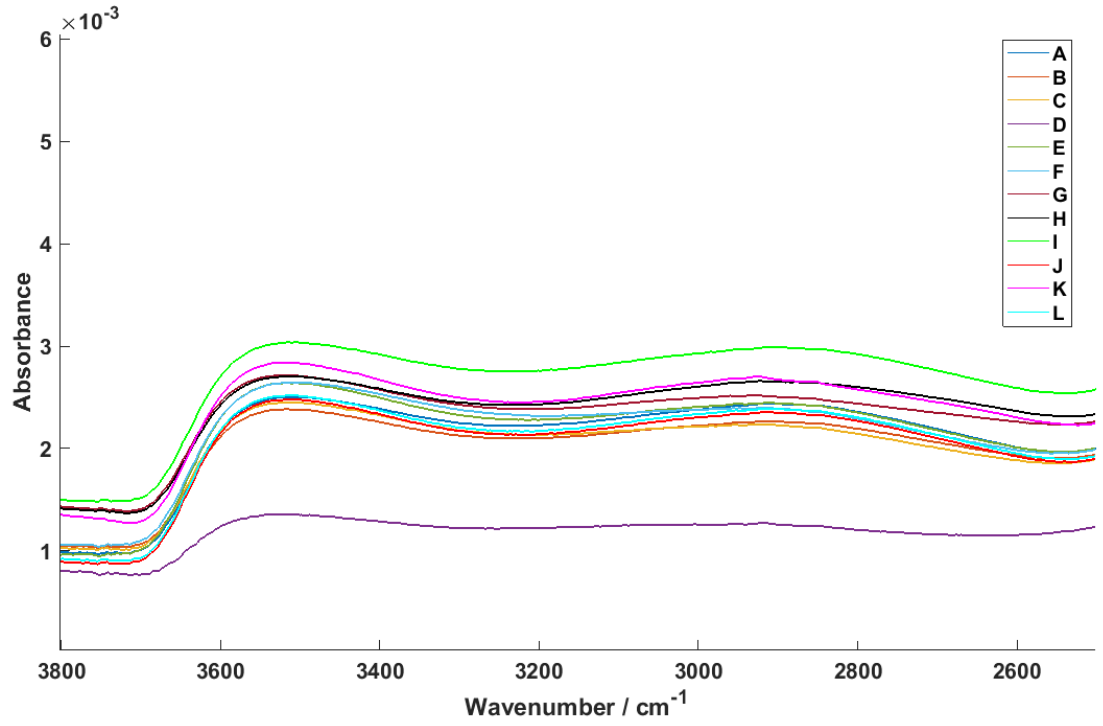


Figure 7 - Zoomed in spectra of plain glass slides, highlighting region above 2500 cm^{-1}

As expected, the absorbance for the spectra is much higher in the lower wavenumber ("fingerprint region") between 2500-900 cm^{-1} where the infrared radiation is absorbed seen in Figure 6. The absorption begins to rise exponentially and appears to peak at roughly 1800 cm^{-1}. The focus however, is on the higher wave number region with minimal absorption of IR enabling the range between 3800-2500 cm^{-1} to be used for classification.

Within this range of wavenumbers, the absorbance does not surpass 0.003 with the exception of slide D with a lower absorbance that only reaches a maximum of 0.0015. The spectra appear to be fairly similar in terms of the location of peaks and peak shape, a broad peak appearing at approximately 3540 cm^{-1} and a very broad peak at 2850 cm^{-1}.

Silica glass can be divided into four types, I-IV, which separates them depending on the production method, amount of metallic impurities and OH content measured in ppm [11]. As previously stated, the spectra appear to have two broad peaks at 3540 and 2850 cm^{-1} which represent the presence of OH-groups. The peak at 3540 cm^{-1} correspond to oscillations of O-H groups and 2850 cm^{-1} to bonded OH-groups interacting with strong hydrogen bridges to other non-bridging oxygen [12] [11]. The latter is only present in mixed silicate glass which is to be expected since most microscope slides are made from mixed silicate, 'soda-lime' glass as opposed to pure silicate. The peak at 2850 cm^{-1} appears to be much broader and are marginally distinguishable for slides D, G and H and remain relatively flat past 2800 cm^{-1} compared to the other slides which seem to have a small smooth decline in absorbance. This minor decrease in absorption suggests that these slides could have a lower abundance of OH groups which Bruckner refers to as impurities which weaken the glass network. The presence of hydrogen bonding causes shrinkage in the glass network which leads to an increase in the refractive index and density with increasing OH-content [11]. This may affect the transmission of infrared and suggest that a purer glass with less hydrogen bonding may have superior transmission.

Furthermore, this plateau is seen in the spectra produced from slide D, indicating that these slides might also have further processing applied to produce a glass with a 'very low iron content' as stated from the manufacturer to increase transparency and remove the greenish tint compared to conventional soda-lime glass. The addition of trace elements is not uncommon as they provide specific properties to the glass and modify the optical properties such as light transmission, UV absorption and colour etc. The effects of these arise from the interactions between different ions in various oxidation states which are a function of the redox state of the glass. A prime example is the iron content in glass which adds a green tint to some glass with higher concentrations of ferric oxide (Fe_2O_3). Low iron content glasses (<10ppm) provide enhanced light transmission compared to standard 'clear' glass between 300-2500 nm which would further explain the differences seen for slide D [13].

An ATR FT-IR transmittance spectra of Iron (III) oxide which is commonly present in soda-lime glass is shown in Figure 8. The transmittance between the interested range is only a maximum of 57% suggesting a lower iron content could benefit the infrared transmittance [14].

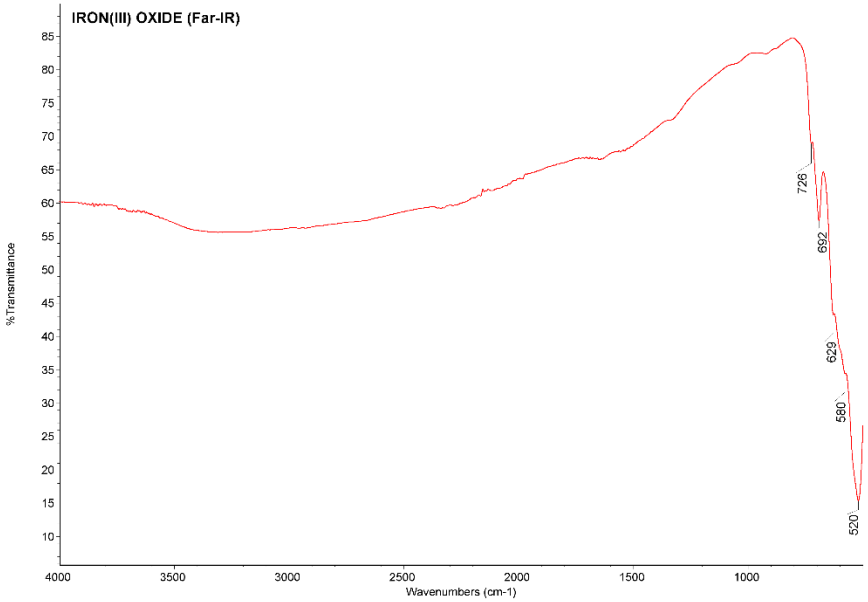


Figure 8 - ATR FT-IR spectra of Iron (III) oxide in the mid-IR range [12]

Albeit, the differences in absorbance is unlikely to affect classification as it measures the abundance of a functional group present. A greater importance is placed on distortions in the shape/position of the peaks as this could potentially misdirect the classifier by misidentifying functional groups. Furthermore, the absorbance is relatively low values compared to values expected in biological material found in tissue samples. Therefore, it is unlikely the small variations in absorbance within this range will obscure any peaks when measuring tissue.

Since product compositions and technical data is unavailable, it is not possible to determine the true underlying factor behind the differences. Furthermore, determining the differences between the spectra of glass is highly complex and could be related to a combination of chemical composition, structure and/or anomalous to the particular batch tested, where further testing and perhaps analysis using different methods would be required but is beyond the scope of the project.

It is important to note that the composition of the glass slides from the same manufacturers and perhaps even different batches will not be consistent. The composition quoted from Citoglas claims their glass slides range between 70-73% SiO_2, 13-15% Na_2O, 7-12% CaO, 1-

4.5% MgO, 1-2% Al_2O_3, 0-0.3% SO_3 and 0.05-0.2% Fe_2O_3 indicating that there are natural variations in the product due to varying composition of the raw material. It is unlikely to expect glass with consistent chemical composition as external factors such as location of the plant and environmental factors can also have influence.

PCA was performed to further identify any variations in the spectra not easily observable and determine whether clustering would occur between glass with the same manufacturer or even with groups of manufacturers. The PCA plot is shown in Figure 9 uses 50 random spectra from each run to create a plot with 150 data points for each slide. This is done to prevent overcrowding of the plot and avoid obscuring points behind one another.

The plot indicates that the slides appear to be very similar as the two principal components which separate the data points are both <1%, indicating the PCA is unable to find large degree of differences and separation is minimal. However, this PCA plot encompasses the whole range and from the loading plot shown in Figure 9, the variances in the region of interest is masked due to being overshadowed by the saturation of absorbance in the lower wavenumber range. A new PCA plot shown in Figure 10 was created which only included the higher wavenumber range which enabled a better separation to occur.

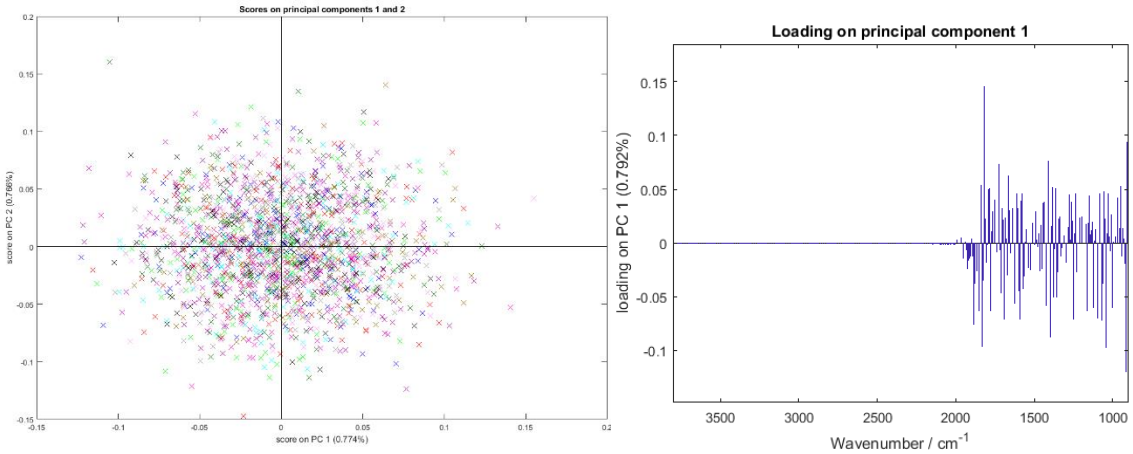


Figure 9 - Scores plot of PC12 for all slides (left), loadings plot on PC1 of all glass slides (right)

In Figure 10, the scores plot on principal component 1&2 indicate that the PCA is unable to distinguish between the different substrates apart from slide D. There is distinct clustering around the origin and multiple overlap amongst points, without any separation and the variation seen in PC1 is shared amongst all glass. This implies that all the glass slides are alike in terms of the spectra, hinting that the glass substrate would not have a great impact on tissue classification. This is also supported by the scores plot on PC13, the

absence of distinct clustering or groups imply that there are relatively little or no differences to separate the substrates where even slide D is considered 'normal'. Further analysis was completed by removing slide D from the data set as it was treated as an anomalous result, by doing so, it aimed to determine whether further separation could be seen in the PCA. However, this was ineffective and had little effect on further separating the data. The scores and loading plots for the data set without slide D can be found in the appendix.

By analysing the loading plot for PC2, it's possible to determine which wavenumbers separate slide D from the others. Loading plot for PC2 was the only applicable plot since only the scores plot involving PC2 show separation, whilst PC13 cannot separate D since similar variation is seen in all 12 slides. The clustering of slide D points in the positive y-axis of PC2 are mainly due to wavenumbers 3800-3600 cm^{-1} and 2750-2500 cm^{-1}. Interestingly, the PCA proposes that the main features that separate slide D is the initial value at 3800 cm^{-1} and the sharpness of the rise towards the initial peak. Usually, OH-stretching is responsible for the absorbance close to that range near 3600 cm^{-1} however, rarely reaches above 3700 cm^{-1} and so the difference of origin on the y-axis is unsettled. The variance seen in PC2 only accounts for 0.729% so whether this difference is enough to influence the chemical spectra is yet to be determined.

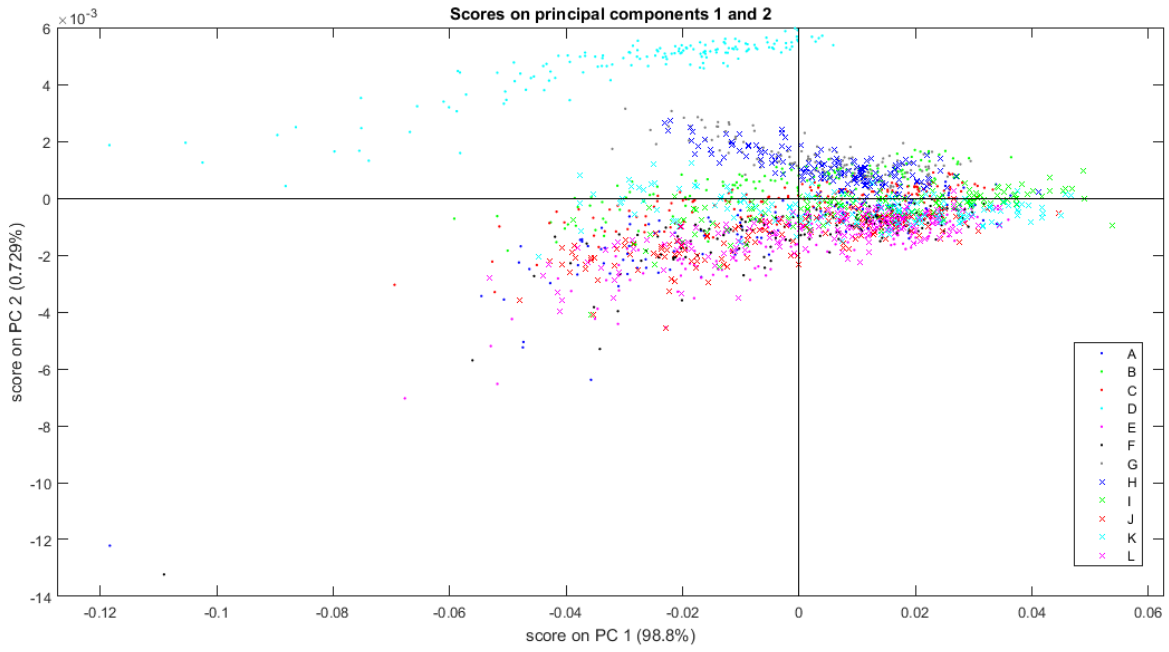


Figure 10 - Scores plot on PC12 of all 12 slides

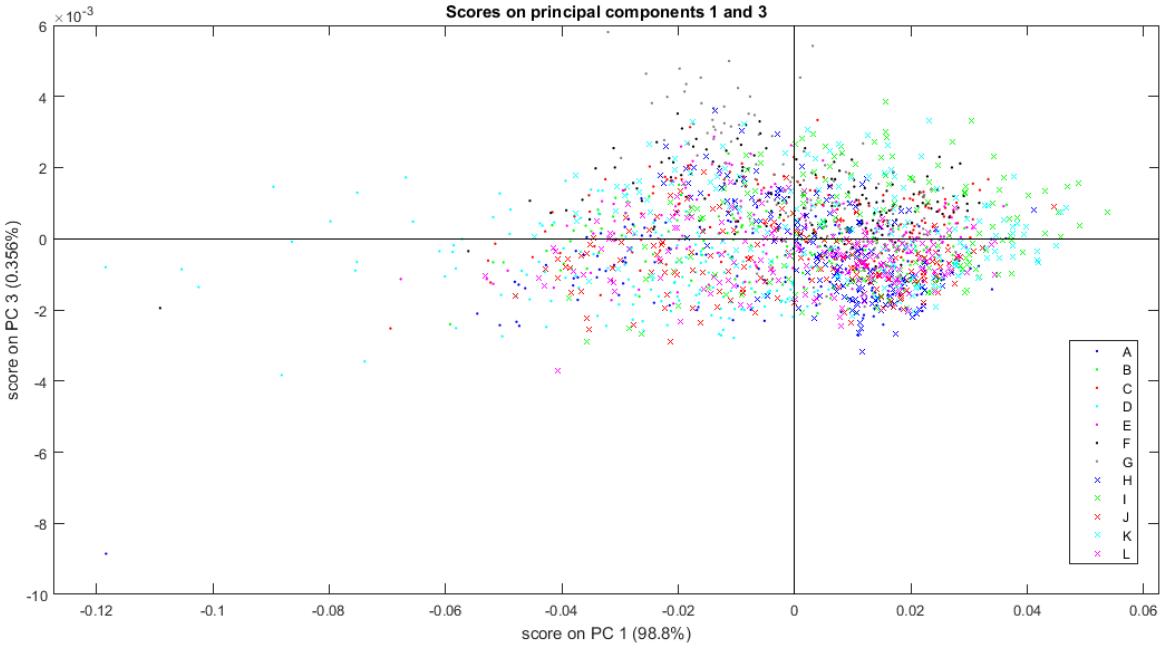


Figure 11 - Scores plot for PC13

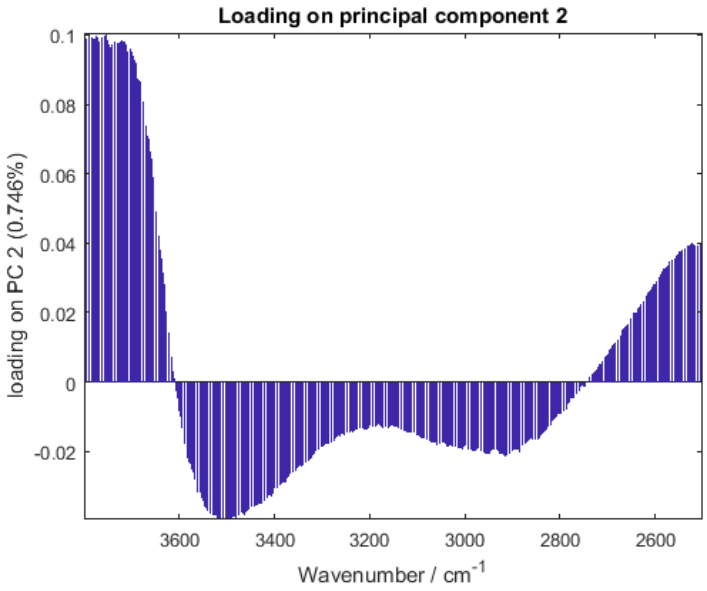


Figure 12 - Loading plot for PC2

4.1.1 PCA of slides, charge coated (JKL) against non-coated slides (ABCEFGHI)

On first assumptions, it was assumed that the slides with the charged coating could potentially exhibit some spectral differences (based upon different allocated spectral features) and so coated slides and non-coated slides were compared in a PCA plot shown in Figure 14.

The charged coated slides were grouped into a single data set and 350 points were randomly selected, the same applied for the non-charge coated slides excluding slide D as it was considered as an anomaly. Although the analysis was unable to distinguish between the coated and non-coated, it seems that the coated slides are more alike with one another compared to the charged slides. The charged slides have a greater variance with less clustering compared to the non-charged slides, particularly in when looking at PC3. Granted, PC3 only contributes to 0.202% of the variance so may seem to skew the distribution more. A potential factor behind this is since the glass used for coated slides are assumed to be the same, the PCA is only able to identify the "considerably higher positive charge than the regular HistoBond+ slides" referring to the charge coating on the Histobond S+ slides. Whilst the PCA could not separate the charge coated slides from one another, it may have been able to realise there was a small difference due to the difference in charge densities of the coatings.

Although the adhesive coating used by Marienfeld is not specified, poly-L-lysine is typically used as an adhesive coating. Poly-L-lysine is a bio-compatible cationic polymer which is coated onto glass as an adhesive for biological cells and DNA since these are known to carry a net negative charge. The structure of poly-L-lysine is shown in Figure 13.

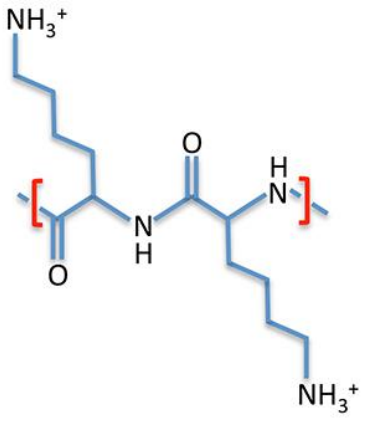


Figure 13 - Poly-L-lysine structure [15]

Solid poly-L-lysine has a chemical spectrum within the interested range, with N-H stretching producing three bands between 3270-3030 cm^{-1}, it also produces a band at the amide I (carbonyl C=O stretching of the peptide group) location within 1600-1700 cm^{-1} although these bands are removed when the range is cut [16].

Although a relatively simple process to apply, a reproducible layer of poly-L-lysine may not be easily achievable with fluctuating absorptions, stabilities and low coverage of the coated layer. Furthermore, the poly-L-lysine may have different densities and different manufacturers/products may use various grades that can also affect the spectra [15]. There are also different secondary structures which poly-L-lysine can have as they adopt either an α-helix, β-sheet and random coil formation. A paper by Polzi *et al.* investigating the infrared spectra of these secondary structures in aqueous solution indicated that there were differences in the amide I band of the peptides and proteins with peak shifts related to structural changes [17]. Mauerer *et al.* investigated the changes of amide I bands of poly-L-lysine upon spray-drying and found that the configuration changed to anti-parallel β-sheet as it is the most energetically favourable state. This is a result of a combination of stronger H-bonds but also steric effects [18]. Miritic *et al.* also concluded that solvent, pH value and temperature are also factors in the conformation of poly-L-lysine [19]. Therefore, it would not be unusual to expect shifts or changes in the N-H region determined by configuration of the polymer which was investigated by Rozenberg *et al.* [16].

It is unlikely that a consistent coating can be achieved on all slides since there are variations of poly-L-lysine (including the use of poly-D-lysine) and different methodologies practiced across different clinics. Also, the method of application onto the slide may vary.

Although, with the aforementioned factors in mind, the Marienfeld brand slides which have an adhesive coating applied, it appears this does not make a difference as PCA is unable to separate between charge coated and non-coated slides. An assumption made as to why the coating did not have an effect on the spectrum is that the coating is simply too thin to readily absorb significant amounts of infrared. Typically, the surface coverage of poly-L-lysine is expected to be within the nanometre range which is below sensitivity of the equipment used. Further analysis with a greater number of adhesive slides using a range of manufacturers and adhesives could support this conclusion.

As a precautionary step, all the glass slides had a poly-L-lysine coating applied prior to having the tissue sample cut onto it. This has two added benefits, one being better adhesion of tissue onto the slide and two, removing the coating as a potential variable.

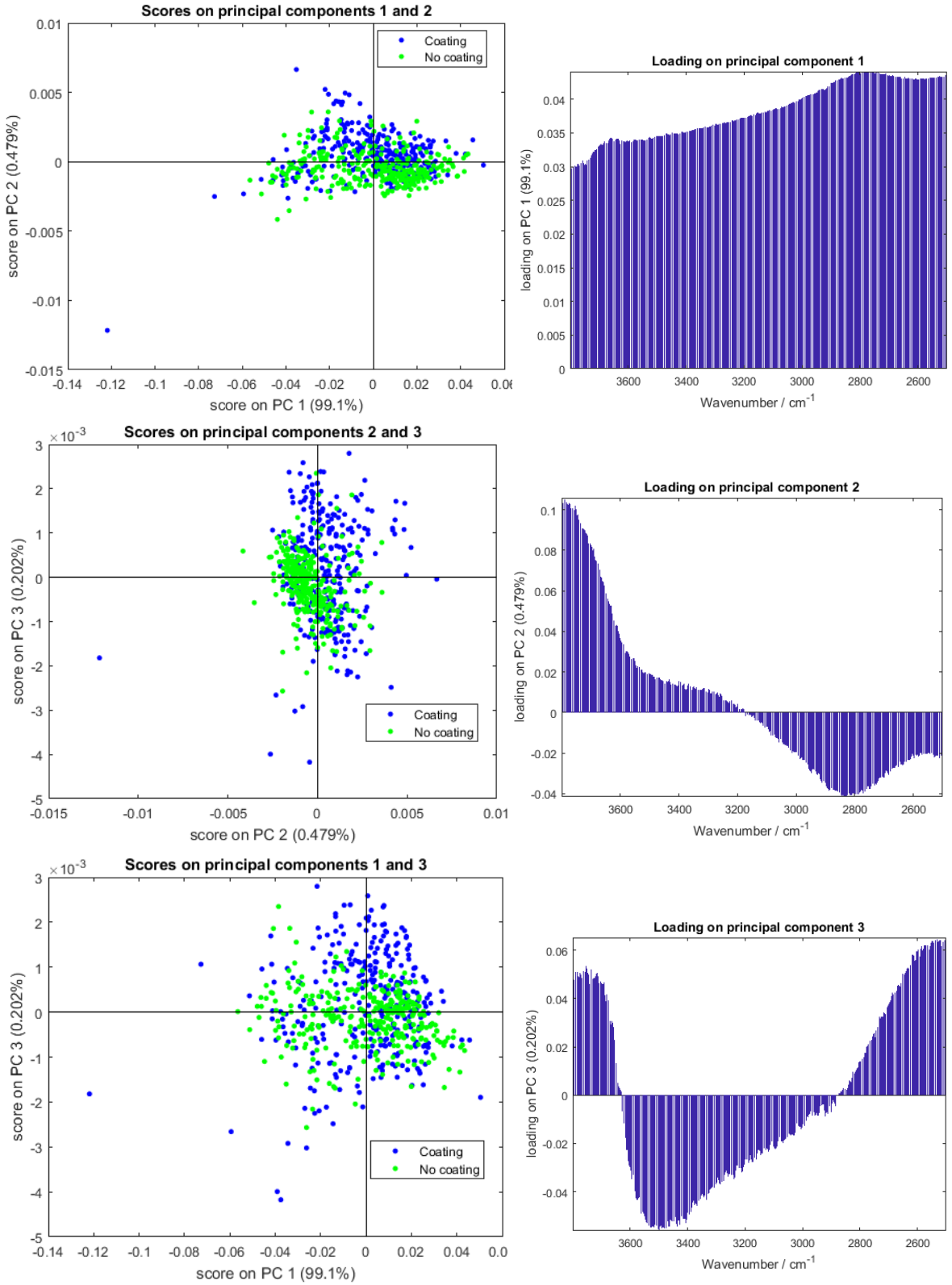


Figure 14 - Scores and loading plots of charge coated slides (blue) against non-coated (green).

From the analysed results, it appears that there are subtle differences in glass slides which originate from chemical composition or structure. Although it was seen that the adhesive coating did not affect the spectra, potentially due to the thickness of the layer being too thin to have any significant absorbance. The impact of these differences to more complex spectra seen in biological material and how it may affect the classification will need to be investigated.

4.2 TISSUE SAMPLES

The mean spectra of the prostate tissue on the Academy slide is shown in Figure 15, revealing three sharp peaks at 2874 cm^{-1}, 2934 cm^{-1}, and 2959 cm^{-1} within the lipid region occurring from the mounting medium [10]. The spectra indicate that the sample were successfully dewaxed as the three main bands associated with paraffin in the C–H stretching region occurring at 2846 cm^{-1}, 2917 cm^{-1}, and 2954 cm^{-1} are inconsistent with the bands present [20]. Distinct differences in biological material absorbance seen at the amide A band suggest that the amide A band is best suited for differentiating different histological classes [10]. This is supported by clear differences in spectra at the amide A peak seen in Figure 15.

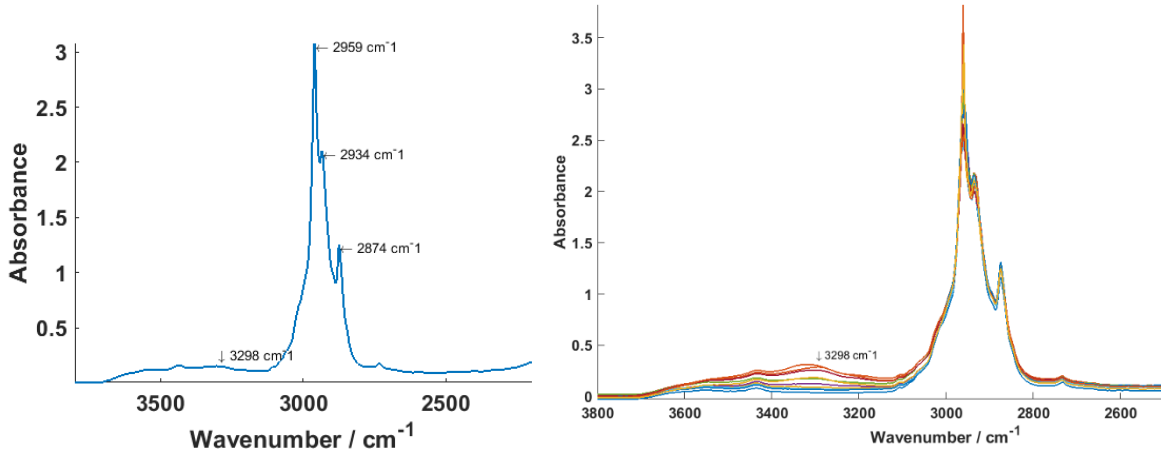


Figure 15 - Mean spectra of raw prostate tissue sample on academy slide (left) and 10 randomly selected spectra plotted to highlight differences in amide A band (right).

4.2.1 Comparing operator annotations with classifier performance

A Nikon brightfield image was taken of each prostate tissue sample and a false greyscale chemical image was produced from the spectra using the amide A band at 3298 cm^{-1}. The chemical images produced were a good representation of the H&E stained samples with excellent agreement in the morphology. The high contrast images enabled key regions such

as glandular epithelium and strands of stroma to be clearly distinguished which allowed for a good point of reference for the annotations.

In Figure 16, there are annotations completed by two operators, where each pixel represents one spectrum. The annotations were completed without prior trained experience and brief knowledge in distinguishing between epithelial and stromal cells. By comparing the two annotations, there are good agreements in epithelial cells highlighted in green, with mostly the same areas shaded and multiple areas of overlap. The stroma cells however, are less alike with only a few similar regions shaded. On average across the three slides, operator 2 had imported 10,000 more epithelial spectra and 15,000 more stroma spectra compared to operator 1.

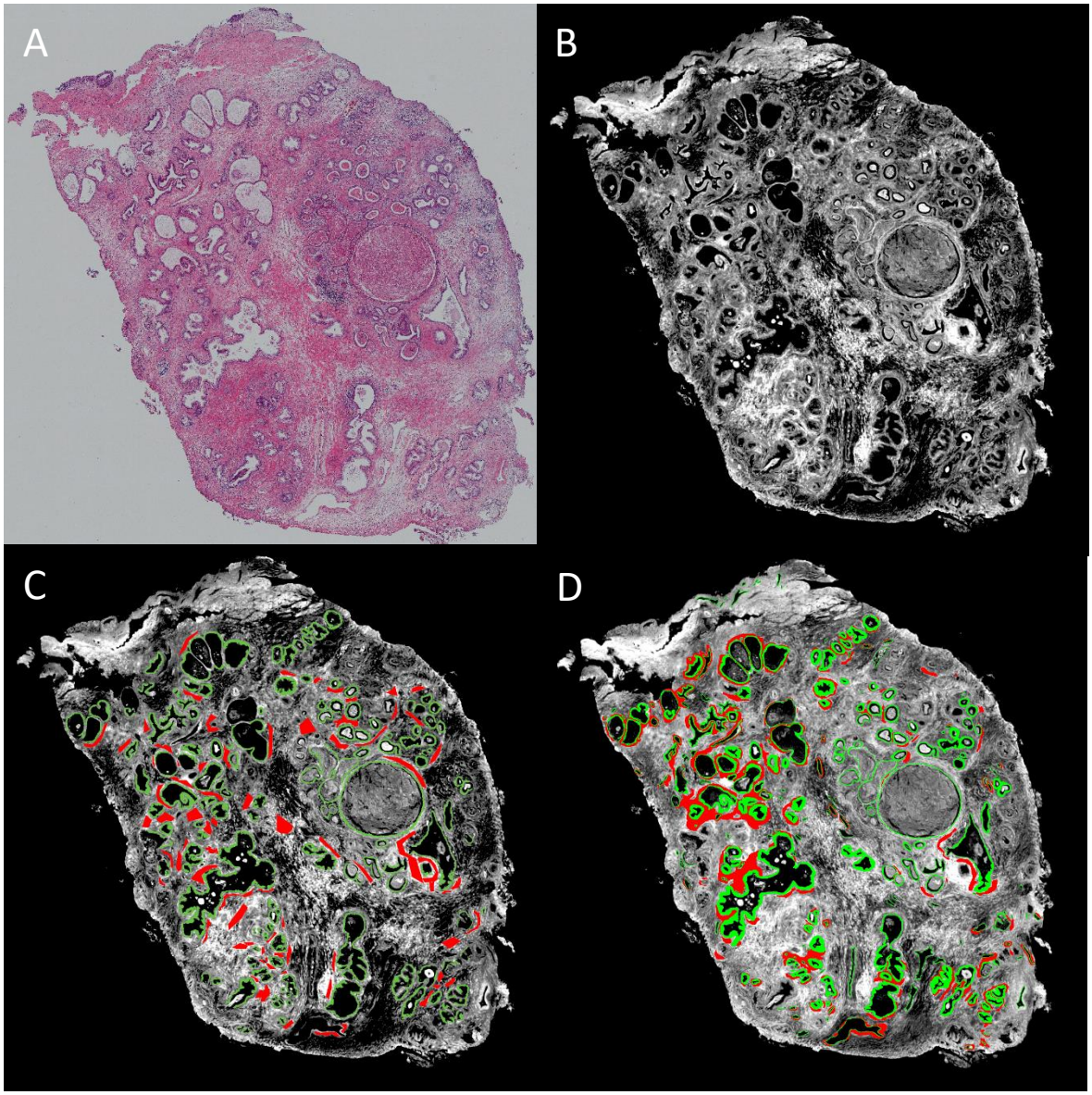
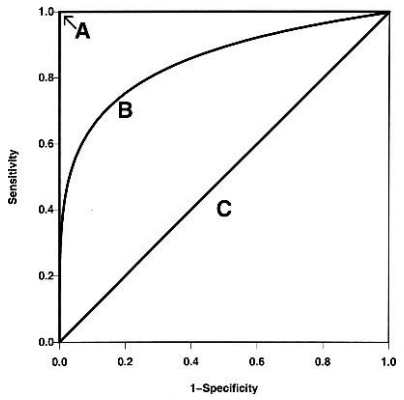


Figure 16 - A) Nikon brightfield image of prostate tissue on Academy slide b) Chemical image of A) produced from the amide A band C) Annotations on chemical image by operator 1 D) Annotations on chemical image by operator 2

The effect of the glass substrate on tissue classification was determined by building a classifier using two of the glass substrates (Academy and Fisherbrand) for training and performing an independent test using the final slide (Superfrost). The classifier was built using 90% of the data set for training and the remaining 10% for validation with the aim to classify between epithelial and stroma cells. The performance of the two classifiers built from different operator annotations will compared, key parameters such as variables, node size and trees will remain the same to ensure for a fair comparison.

The performance of the classifier will be measured using a Receiver Operator Curve (ROC) which evaluates the relationship between true positives (TP) (sensitivity) and false positive rate (FP) (1-specificity). The equations relating sensitivity and specificity to the TP, FP, true negative (TN), and false negative (FN) are shown in Figure 17.

An ideal test with perfect accuracy would have an Area Under Curve (AUC) equal to 1, where the sensitivity is at 1 and 1-Specificity at 0 represented by line A in Figure 17. A classifier with poor performance will have an AUC <<1, a poor classifier with no diagnostic benefit is indicated by line C with line B representing a reasonable test with an AUC between the two at 0.85.



$$\text{Sensitivity} = \frac{TP}{P} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{N} + \frac{TN}{TN + FP}$$

Figure 17 - ROC curve example with equations describing the axes [21]

The ROC curves for the training the models for both operators are shown on the left-hand side of Figure 18 and Figure 19. The ROC curve for operator 1 and operator 2 was produced using approximately 32,000 and 65,000 spectra from each class respectively. Comparing the ROC curves, the two operators appear to be very similar. From the graphs, both classes were identified with roughly the same accuracy and AUC. Interestingly, the independent test performed better on the model with the lower accuracy, although the difference was within 5% and both performed poorly.

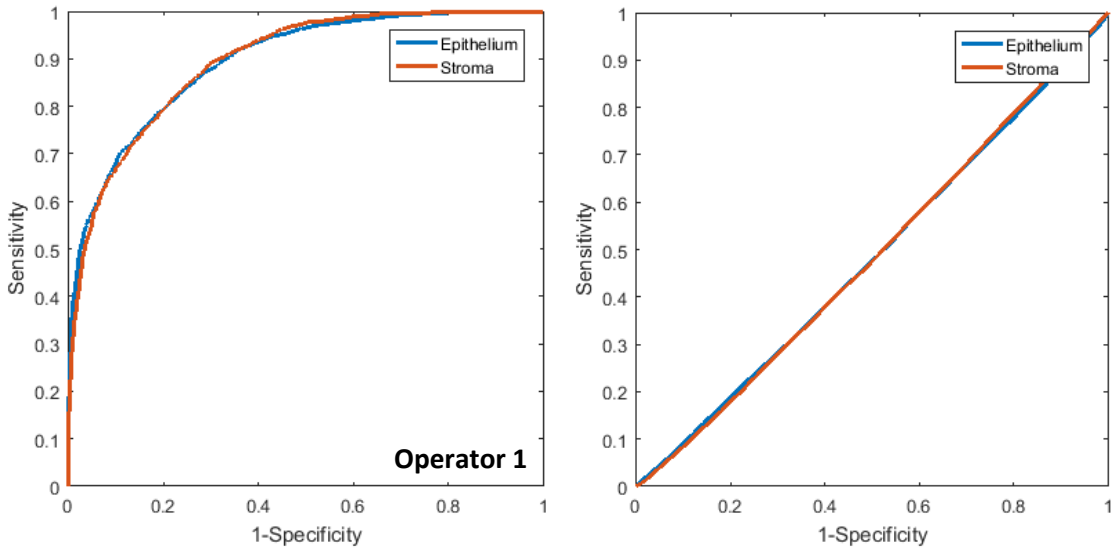


Figure 18 - Operator 1 ROC curves for training (left) and independent testing (right)

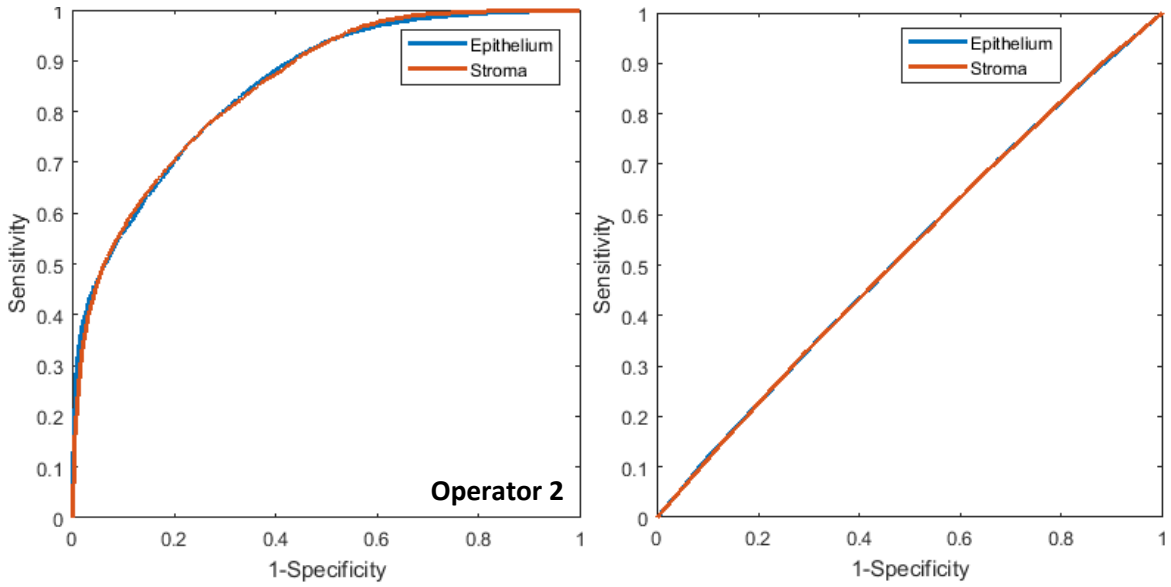


Figure 19 - Operator 2 ROC curves for training (left) and independent testing (right)

The AUC values achieved for epithelium and stroma for operator 1 are both 0.894 and operator 2 performed slightly worse with epithelium and stroma AUC values of 0.851. Whilst these are reasonable values, compared to similar work these are noticeably poor. Pilling *et al.* consistently achieved AUC values on four classes, epithelium, stroma, blood and concretion above 0.99 in training [10].

The trained model was then used to class spectra from the Superfrost slide as an independent test. Operator 1 and operator 2's database of spectra contained roughly 60,000 and 75,000 spectra respectively. The model performed poorly with an overall AUC value of 0.482 and 0.525 for operator 1 and 2 respectively suggesting the classifier is of little diagnostic benefit. The results are pale in comparison to Pilling's results of epithelium (0.986), stroma (0.981), blood (0.986) and concretion (0.998) from a test set of 449,591

spectra from 80 independent patients. Similarly, Bassan *et al.* was able to achieve values of epithelium (0.9845), stroma (0.9956), blood (0.9922) and necrosis (0.9932) using 141,343 spectra from 69 cores [7]. Fernandez *et al.* was also able to correctly predict across 10 classes using 260 tissue microarray (TMA) cores from 40 patients with a minimum accuracy of 90%, excluding nerve spectra, the value rises to 94% [22].

The confusion matrix for training and independent testing is presented in Table 3 and Table 4 which show the accuracy prediction of each class, with the correctly predicted accuracies highlighted in bold. Likewise, the AUC values mentioned above the percentage of each class correctly classified is substandard compared to previous studies. Although the training accuracies were not provided, they may be inferred from the remarkably high AUC values which imply a value close to ~99% accuracy for training. The independent test for Pilling and Bassan's work both achieved >90% and >97% respectively. Interestingly, the stroma prediction for operator 2's model was higher than in the training, potentially caused by misclassification of epithelium rather than good stroma prediction. Although stroma produced a reasonable result, the poor epithelium prediction drastically reduces the overall accuracy.

Table 3 – Confusion matrix with training accuracies (%) for each class

		Predicted Class			
		Operator 1		**Operator 2**	
		Epithelium	Stroma	Epithelium	Stroma
True Class	Epithelium	**81.9**	18.1	**77.3**	22.7
	Stroma	22.8	**77.2**	26.5	**73.5**
	Overall	**79.6**		**75.4**	

Table 4 - Confusion matrix with independent test accuracies (%) for each class

		Predicted Class			
		Operator 1		**Operator 2**	
		Epithelium	Stroma	Epithelium	Stroma
True Class	Epithelium	**38.9**	61.1	**27.6**	72.4
	Stroma	41.1	**58.9**	24.6	**75.3**
	Overall	**45.3**		**46.8**	

Considering the number of spectra used in previous work, the results are unexpectedly poor. During training Pilling only used 586 spectra for each of the four classes whilst Bassan's database of spectra contained 3792 epithelial, 8088 stromal, 838 blood, and 1801 necrosis [10] [7]. Fernandez also used 171,000 pixels (spectra) across 10 classes [22].

Although for Fernandez's case it could be argued a lower number of spectra were sufficient due to the use of BaF_2 slides. Using BaF_2 enabled a greater range to be used and hence was able to obtain more biological information from the lower wavenumber region. Considering the use of a minimum of 30,000 spectra in each model for training each class, one would expect the classifier to perform equally as well or even better.

Furthermore, the number of patients used within the study may be argued in relation to the performance of the classifiers. Previous work used a number of patients for training and testing which would provide a wide database of epithelial and stroma spectra. However, the factor of interpatient variability albeit, effects considered minimal, can be excluded in this study as it used a contiguous tissue block from a single patient. This would slightly enhance the results although not enough to explain the large difference in accuracies.

A classifier built using different parameters within Random Forest may also be a source for inconsistent results. For example, the number of trees used in Pilling's paper was only 200 and the determination behind the use of 200 was unjustified. Furthermore, only retaining 15 PC's during the noise reduction is questionable. The spectra of the noise removed faintly resembled the spectra of biological material opposed to noise when retaining 15 PC's was attempted. The use of more trees and more PC's in this study is expected to boost performance, although this was not the case.

Additionally, Bassan used different parameters during hyperspectral image collection. Bassan measured at a spectral resolution of 8 cm^{-1} using 32 and 8 co-added background and sample scans compared to 5 cm^{-1} using 256 and 96 in this study (same as Pilling). The difference between using 5-8 cm^{-1} spectral resolutions has not yet been determined however, reducing spectral resolution could miss important biological information between wavelengths. Furthermore, so few scans could impact the SNR although the determining the effect of tissue classification using different collection parameters will need to be explored in future work.

Previous work also opted for an acceptance threshold value varying from 60%-95% compared to no threshold in this study. The effect of including an acceptance threshold is it removes spectra classified with low statistical confidence and excludes spectra casted with roughly equal number of votes. Furthermore, the effects of the glass coverslip and its ability to modify the transmitted light has not been studied. Since a visual examination of

the slides was not possible, this factor cannot be ruled out for causing low diagnostic accuracy. Future work will need to include a wide range of different coverslips and assess the impact on infrared transmittance from improper application or deformation.

The poor results were likely owed to lack of experience and training amongst the operators and difficulty accurately identifying pixels of epithelium and stroma. Earlier work completed benefitted from the support of a trained pathologist, highlighting the argument of interobserver variability, which in this case the difference is at two extremes. Although, hopefully amongst trained pathologists the difference and its effect on the classifier would be minimal.

Surprisingly, prediction accuracies for stroma were higher than epithelial, unlike results from earlier studies where accurate epithelial recognition was observed across areas of tissue with neoplasia, BPH and atrophy, resulting in an AUC of 0.994 for epithelium. The AUC results for stroma sub-types were 0.983, 0.961 and 0.894 for fibrous stroma, mixed stroma and smooth muscle respectively [10]. The basis for the difference is the greater difficulty separating stroma from similar cell types such as cells composed of blood vessels. Also, generally epithelial cells are easier to identify on the chemical image. The difference likely stems from the annotations as there was difficulty in annotating epithelial cells particularly at the edges where annotations could overlap non-epithelial material, compared to stroma where large areas of stroma are usually found together and is easier to avoid non-stroma.

It is difficult to clarify whether the results achieved were a result of the differences due to the glass substrate or were an anomalous result since the classifier was only trained on Academy and Fisherbrand slides and independently tested on Superfrost. The Superfrost slide was treated as an anomaly in the earlier analysis involving the plain glass and so it would have been advantageous to compare the results from the independent test using different slides. Furthermore, annotations completed by a trained pathologist would enable higher training accuracies which would further dictate whether poor classification was due to the glass substrate. Since the training accuracies for the models were not entirely confident, it made it difficult to attain high classification on the independent test regardless of whether the glass had an effect on the spectra or not. An improvement to the method used could be increasing the number of spectra within the database by including more slides during training and testing to allow for a more robust model to be built.

4.2.2 Comparing classifier performance with a x15 objective and a x4 objective

The spectra acquired using a x15 objective used to produce the chemical images seen take a considerably longer period compared to using a x4 objective. Hence, it would be beneficial if similar classifier performance could be achieved using the x4 objective to reduce acquisition time and optimise throughput. The trade-off for faster acquisition time is spatial resolution. Due to time constraints, re-imaging and re-annotating the sample was not possible so the x4 objective had to be produced digitally using resampling methods within MATLAB. The dimensions of the x15 objective are 19.2x19.2 µm whilst the x4 is 5.5x5.5 µm. Since 5.5 is not a factor of 19.2, a rounded value of 4 had to be used. The resulting image is seen in Figure 20, every 4 pixels in both x and y direction were merged together to form 1 pixel and produce the x4 image.

As Figure 20 shows, the difference in resolution is significant. In the x4 objective it is much more difficult to identify and distinguish the morphology of the tissue. This is reflected in the training and independent test ROC curves shown in Figure 21. In both the training and independent test, the AUC were 0.660 and 0.490 respectively. The results indicate that using a x4 objective is not viable, the resulting image quality is low and resampling the spectra loses a large amount of information required for high accuracy classification. Additionally, the number of spectra used questions the robustness of the classifier. Prior to resampling around 30,000 spectra for each class were used, however, after sampling the number reduced to 139. Similarly, the number of spectra used in the independent test also decreased dramatically, falling from 60,000 to 88. Therefore, although a x4 objective can be used to define margins, a x15 objective is needed to resolve details needed for annotations.

Furthermore, an issue with using this method is that the annotations were completed on the x15 objective and rendered down. The same annotations may not be achievable with the lower spatial resolution and completing annotations to a high degree of accuracy on the x4 chemical image may prove difficult. The resampling shown was performed on operator 1 annotations, however similar results were achieved when re-applied to operator 2 annotations.

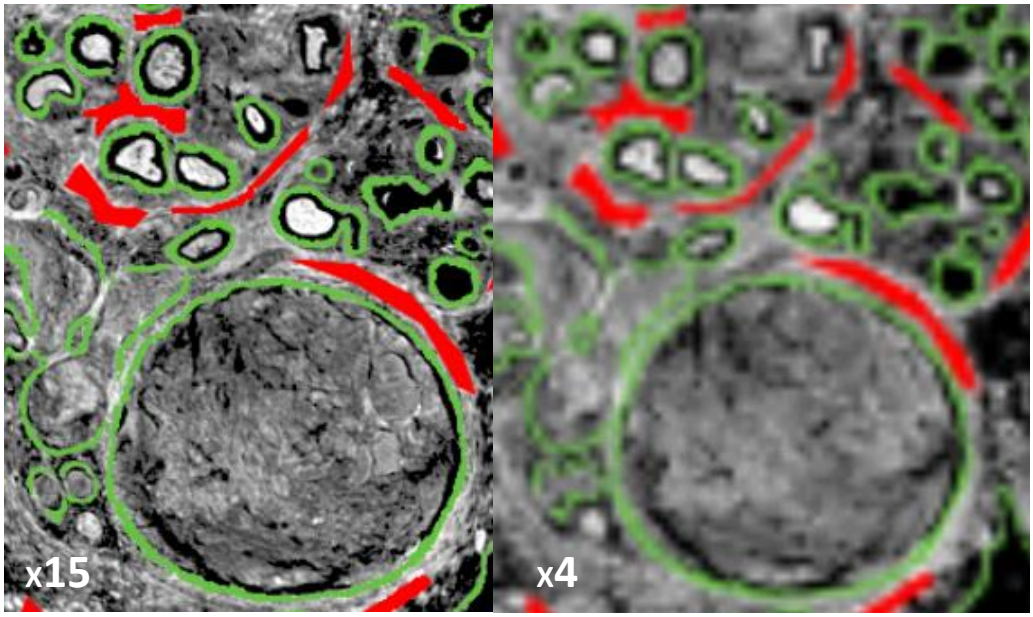


Figure 20 - Comparison of chemical images using a x15 objective and digitally produced x4 objective

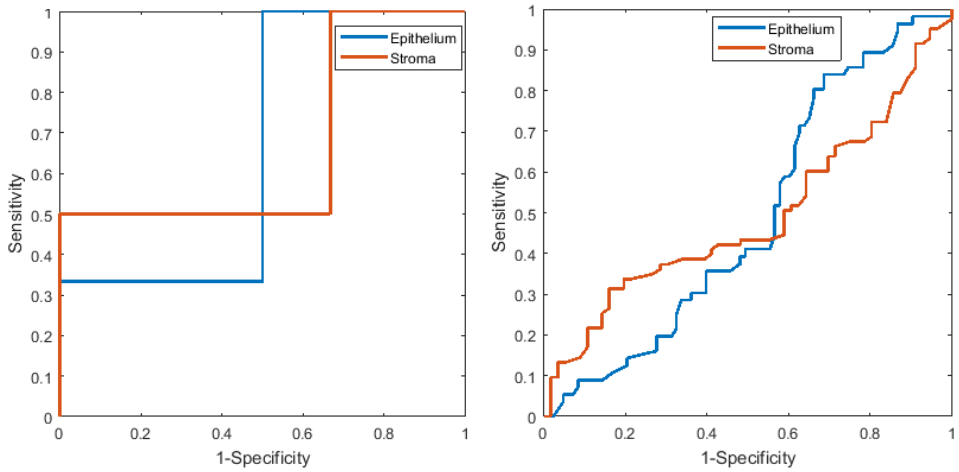


Figure 21 - ROC curves for training (left) and independent test (right) using a x4 objective

5 CONCLUSION AND FUTURE WORK

A selection of 12 plain glass substrates from 5 different manufacturers: Academy, Fisherbrand, Sail Brand, Citoglas, Marienfeld, Kindler and VWR were measured using a FT-IR spectrometer to assess the differences in spectra amongst glass. After PCA, the majority of the glass slides could not be separated with the exception of the slide from Kindler (slide D). The factors behind the differences are unclear although a combination of glass purity with the reduced iron content present in this specific glass seem to attribute to an enhanced infrared transmission. Further analysis was performed aiming to determine whether glass slides with adhesive coatings can be distinguished from non-coated slides. The results from this analysis could not divide the two with an assumption that the coating is too thin to have any impression on the absorbance. Although confident in the results gathered, further analysis can be done using a supervised method of analysis namely, CVA. It may be possible to see further separation amongst the glass slides with labelled data which would enhance and support the work seen in this study. Furthermore, if future results similar to slide D are replicated in other batches or slides, it may be useful to fully understand the chemical composition of the glass and structure using X-ray diffraction [23] [24]. Since only some manufacturers provided technical product specifications, it made it difficult to pinpoint the root cause of the differences and assumptions were made based on manufacturer descriptions.

The glass slides (Academy, Fisherbrand and Superfrost) were investigated to determine the effect of the glass substrate on tissue classification and the results were inconsistent compared to values from literature. The whole tissue section was imaged using up to 81 tiles producing a database of over a million spectra each. A chemical image was then produced using the amide A band at 3298 cm^{-1} and compared with brightfield images for annotations of epithelial and stromal cells completed by two operators. Using the Random Forest algorithm, Academy and Fisherbrand slides were used to build the training model using a minimum of 60,000 spectra for each training model. The Superfrost slide was used for independent testing and the minimum number of spectra classified in each test was 60,000. The AUC results for both training sets were 0.894 and 0.851 and the independent test performed with limited success, achieving an AUC of 0.482 and 0.525. Although the performance of the independent test was significantly poor, determining whether the glass substrate was at the heart of the poor performance could not be justified. The limited data

set and potential anomalous slide alongside the operator annotations did not allow for a robust training model to be built, decreasing the validity of the experiment.

In order to provide clarity on the issue, ideally a trained pathologist should assist with the annotations as there are multiple sub-types of epithelial and stroma which are all liable to have varying spectra from one another. In addition to epithelial and stroma, blood and concretion present may have been incorrectly annotated as well. If a pathologist was not available, it would have been useful to have a tool to identify or determine which spectra specifically were classed with roughly same number of trees or highlight areas of the chemical image with low accuracy. This would then enable a review of the annotations and allow for corrections to be made. In addition, developing computational methods to overlay the brightfield image onto the chemical image and directly annotate on the detailed microscope image may have provided more accurate results.

There are also other limitations involved within the study, above all the most important factor being the limited number of glass slides used during the tissue classification. Since the Superfrost slide was excluded out as an anomaly in the earlier PCA, it would have been significant in determining whether this anomaly had an effect on tissue classification. A robust model built from a wide range of glass substrates providing an AUC closer to 1 would have been able to identify this. If every independent test consistently achieved similar values and the Superfrost slide performed poorly again, it would conclude that the glass substrate does influence tissue classification and subsequently would be impractical for use in clinics.

Additionally, other forms of classification and regression analysis could have been considered, the use of support vector machines (SVM) or back propagation neural networks could have been used to cross-validate the Random Forest results. Liu *et al.* studies comparing the classifiers for electronic tongue data classification has shown that Random Forest outperform other classification method with average correct rate of 99.07% [25]. However, a more relevant study by Statnikov *et al.* compared Random Forest with SVM for micro-array breast cancer classification and found the latter to excel. [26] Statnikov concluded that Random Forest did not exhibit “best of class performance” contrary to studies performed by Diaz-Uriarte *et al.* [27] as Statnikov was able to identify flaws in the methodology which skew the results in favour of Random Forest. The most appropriate classifier is still under debate although it may be valuable to consider both to

identify the best suited classifier for the particular study. Besides, there are constantly new algorithms being coded which have varying training time, prediction time and accuracy. Since accuracy is prioritised, algorithms such as Xgboost and neural networks which are regarded amongst the community as high accuracy classifiers could be explored in future studies.

Whilst on the topic of computational methods, classification accuracy could be improved by including the use of non-linear spectral feature extraction methods such as neighbourhood preserving embedding (NPE) and supervised NPE (SNPE). Non-linear methods are an effective way at identifying the characteristic features of the spectra and provide a better representation of spectral features that are less sensitive to outliers. Lee *et al.* compared feature-extracted variables using PCA, NPE and SNPE of normal, adenoma and cancer stomach and colon sections [28]. Lee was able to demonstrate different spectral features selected by each method with the scores plot indicating that PCA struggled to discriminate between classes whilst NPE and SNPE had indicative signs of good discrimination. As the non-linear methods are more reflective of the spectral features present in the data set, it is unsurprising that a classifier built using NPE/SNPE outperforms the classifier using PCA. Although, Lee admits that the study is not enough to generalize superiority of non-linear methods, applying these in future studies could provide useful insight to the best method to achieve highest accuracy.

With the aforementioned in mind, the benefits of glass slides are numerous, given that they are readily available, robust, cost effective and established amongst pathologists. A major benefit of using infrared histopathology is the high throughput with training and classification achieved rapidly, only limited by speed of data acquisition. The whole tissue section was imaged in this study which took roughly 5-8 hours although this can be minimised by using TMA cores or a 2x2 mosaic which require a considerably shorter acquisition time. Another benefit is that it can be done unsupervised and so workflow will not be disrupted, future work to optimise collection parameters and data acquisition will further improve throughput.

5.1 FUTURE WORK

This study indicates that there is potential for using FT-IR with glass substrates for tissue classification and supports the prospect of identify malignant tissue for cancer diagnosis. With refinement in the methodology, standardising annotations and optimising classifier parameters, an automated cancer diagnostic tool is on the horizon. This study only focused on epithelial and stromal cells using assigned biomarkers however, the study could be expanded to consider other spectral biomarkers for performance. A study by Dukor *et al.* has identified the use of the extracellular matrix for pathological diagnosis, particularly carcinoma [29]. Dukor was able to recognise the baseline slope of a 1280 cm^{-1} band differs between cancerous and normal connective tissue, indicating that there are useful spectral markers outside of conventionally used cells. The use of extracellular matrix in spectroscopic analysis could be extended onto glass and complement cellular diagnosis or for validation.

In order to strengthen the outlook of using of glass substrates, ensuring that the mounting medium or paraffin do not produce any wavelengths or bands within the region used for classification is paramount. Although the main peaks for mounting medium are removed prior to classification, it is unknown as to whether there are still bands remaining between 3700-3125 cm^{-1} which could potentially affect the spectra.

Further analysis on different types of tissue should also be considered, so far only prostate and breast tissue has been examined. Other structurally complex tissue will also need to undergo testing to ensure that tissue types with similar absorptions do not overlap and be mis-assigned as one another. In future, it may also be possible to grade cancer and assign different Gleason scores to determine the stage of cancer on glass. Gazi *et al.* has used infrared spectra of prostate cancer tissue on BaF_2 to relate the tissue biochemistry with the Gleason score assigned by a pathologist. If this could be replicated on glass substrates, it would be an extremely useful complement to pathologists, improving histological assessment and offering better support for patients [30].

The prospects for FT-IR spectroscopy is plentiful, future developments and gained understanding of the spectra of cancerous cells or cells suffering from diseases are a step towards spectral pathology. Work has already been completed using FT-IR and Raman to study drug cell interactions and allow multivariate data to highlight details such as mode of action or side effects. If this could work could be developed on glass, it may provide aid to

drug development and more effective treatment [31]. The uses of FT-IR are also not limited to medical applications and has proven useful in the environmental industry as well as chemicals. FT-IR has been applied to forensic soil analysis as well as identifying components used in paint binders. If information was available in the high wavenumber range the switchover to glass could be considered.

Although this study was unable to replicate the high classification accuracies seen in other work, continuing work on glass substrates is still effective to bridging the gap towards an automated pre-screening tool for cancer. Ongoing research will help reduce preventable deaths related to cancer and help streamline diagnosis, allowing health care services to offer cancer screenings from an earlier age. Future work using a wider range of tissue samples, glass substrates and exploring the effects of the mounting medium and coverslips will need to be considered before infrared spectral pathology can be practiced in clinics.

6 REFERENCES

[1] Who.int. (2018). [online] Available at: http://www.who.int/cancer [Accessed 14 Jan. 2018]., [Online].

[2] Cancer Research UK. (2018). Cancer Research UK. [online] Available at: http://www.cancerresearchuk.org [Accessed 16 Jan. 2018]., [Online].

[3] M. van den Bent, Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective, Acta Neuropathologica, vol. 120, no. 3, pp. 297-304, 2010

[4] S. Ismail, A. Colclough, J. Dinnen, D. Eakins, D. Evans, E. Gradwell, J. O'Sullivan, J. Summerell and R. Newcombe, Observer variation in histopathological diagnosis and grading of cervical Intraepithelial neoplasia, Obstetrical & Gynecological survey, v.

[5] D. Perez-Guaita, P. Heraud, K. Marzec, M. de la Guardia, M. Kiupel and B. Wood, Comparison of transflection and transmission FTIR imaging measurements performed on differentially fixed tissue sections, The Analyst, vol. 140, no. 7, pp. 2376-2382, 2015

[6] P. Bassan, J. Lee, A. Sachdeva, J. Pissardini, K. Dorling, J. Fletcher, A. Henderson and P. Gardner, The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications

[7] P. Bassan, J. Mellor, J. Shapiro, K. Williams, M. Lisanti and P. Gardner, Transmission FT-IR Chemical Imaging on Glass Substrates: Applications in Infrared Spectral Histopathology, Analytical Chemistry, vol. 86, no. 3, pp. 1648-1653, 2014

[8] Savitzky-Golay Smoothing Filters. (1990). Computers in Physics.

[9] C. Beleites, and M. Kansiz, (2014). FTIR Microscopic Imaging of Large Samples with 4x and 15x Infrared Objectives: A Case Study of a Carcinoma Tissue Section. Agilent. [Online].

[10] M. J. Pilling, A. Henderson, J. Shanks, M. Brown, N. Clarke and P. Gardner, Infrared spectral histopathology using haematoxylin and eosin (H&E) stained glass slides: a major step forward towards clinical translation, Analyst, vol. 122, no. 1, pp. 1258-1.

[11] R.Brückner, (1970). Properties and structure of vitreous silica. I. Journal of Non-Crystalline Solids, 5(2), pp.123-175.

[12] Www2.lbl.gov. (2018). Characteristic IR Band Positions. [online] Available at: http://www2.lbl.gov/mmartin/bl1.4/IRbands.html [Accessed 22 Apr. 2018]. [Online].

[13] The status of Flat Soda Lime Silicate Glass and its raw materials under REACH. (2015). Glass for Europe. [Online].

[14] Lisa.chem.ut.ee. (2018). Iron(III) oxide – Database of ATR-FT-IR spectra of various materials. [online] Available at: http://lisa.chem.ut.ee/IR_spectra/paint/fillers/ironiii-oxide/ [Accessed 20 Apr. 2018]. [Online].

[15] Proteinslides.com. (2018). Polylysine surfaces | MicroSurfaces, Inc.. [online] Available at: http://www.proteinslides.com/polylysine [Accessed 21 Feb. 2018]. [Online].

[16] M.Rozenberg, and G.Shoham, (2007). FTIR spectra of solid poly-l-lysine in the stretching NH mode range. Biophysical Chemistry, 125(1), pp.166-171.

[17] L. Polzi, I. Daidone, A. Amadei, A Theoretical Reappraisal of Polylysine in the Investigation of Secondary Structure Sensitivity of Infrared Spectra, J Phys Chem B. 2012 Mar 15;116(10):3353-60. doi: 10.1021/jp211063x.

[18] A.Mauerer, and G Lee,. (2006). Changes in the amide I FT-IR bands of poly-l-lysine on spray-drying from α-helix, β-sheet or random coil conformations. European Journal of Pharmaceutics and Biopharmaceutics, 62(2), pp.131-142.

[19] A.Mirtič, and J.Grdadolnik, (2013). The structure of poly-l-lysine in different solvents. Biophysical Chemistry, 175-176, pp.47-53.

[20] C. Hughes, L. Gaunt, M. Brown, N. W. Clarke and P. Gardner, Assessment of paraffin removal from prostate FFPE sections using transmission mode FTIR-FPA imaging. Anal Methods 2014; 6: 1028–1035.

[21] K.H. Zou, A. J. O'Malley and L. Mauri, Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models, Circulation. 2007;115:654-657, February 5, 2007.

[22] D. C. Fernandez, R. Bhargava, S. M. Hewitt and I. W. Levin, Infrared Spectroscopic Imaging for Histopathologic Recognition, Nature Biotechnology, vol. 23, no. 4, pp. 469-474, 2005.

[23] B. E.Warren, and J.Biscce, (1938), The Structure of Silica Glass by X-Ray Diffraction Studies, Journal of the American Ceramic Society, 21: 49-54. doi:10.1111/j.1151-2916.1938.tb15742.x.

[24] R.H. Matjie, Zhongsheng Li, Colin R. Ward, David French,Volume 87, Issue 6,2008,Pages 857-869,ISSN 0016-2361,https://doi.org/10.1016/j.fuel.2007.05.050

[25] M.Liu, M.Wang, J.Wang, D.Li, Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar,Sensors and Sensors Actuators B 133 970–80.

[26] A. Statnikov, L. Wang, C.F. Aliferis, A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification, BMC Bioinformatics, 9 (319) (2008), pp. 1-10.

[27] R. Diaz-Uriarte, S. Alvarez de Andres: Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006, 7: 3.

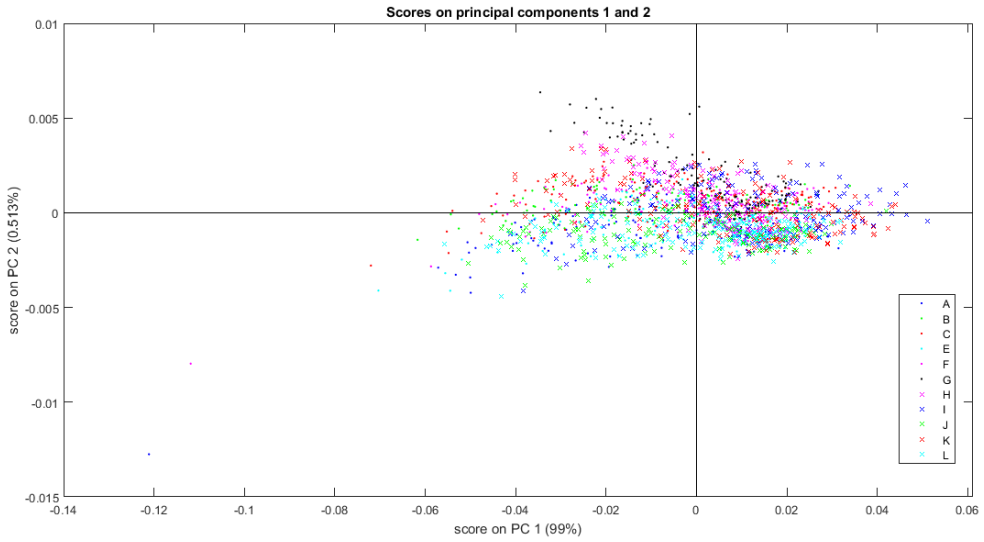
[28] S. Lee, K. Kim, H. Lee, CH. Jun, H. Chung, JJ. Park (2013) Improving the classification accuracy for IR spectroscopic diagnosis of stomach and colon malignancy using non-linear spectral feature extraction methods. Analyst 138(14):4076–4082.

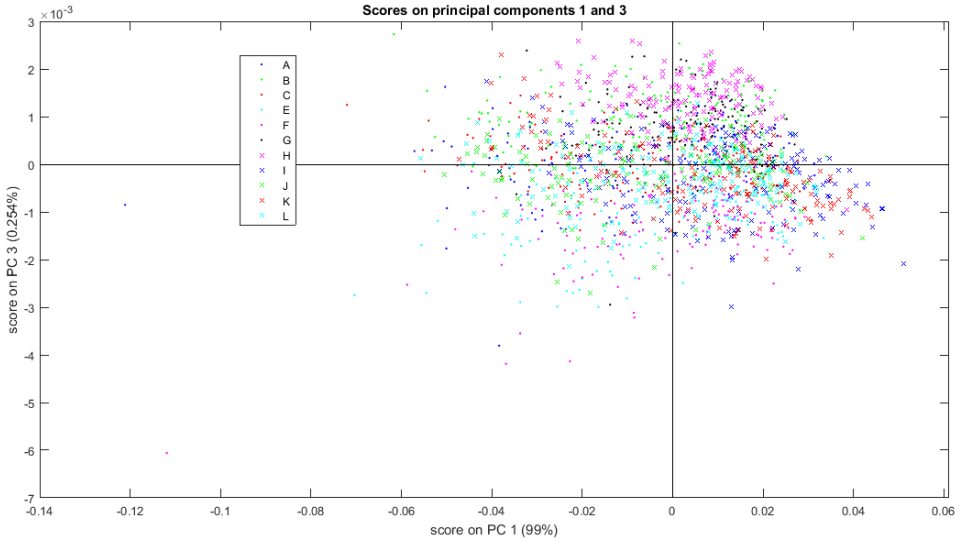
[29] R.K Dukor et al., A Method for Analysis of Clinical Tissue Samples Using FTIR Microspectroscopic imaging, Spectroscopy of Biological Molecules, New Direction 1999, pp. 471-472.

[30] E. Gazi, M. Baker, J. Dwyer, N. P. Lockyer, P. Gardner, J. H. Shanks, R. S. Reeve, C. A. Hart, N. W. Clarke and M. D. Brown, “A correlation of FTIR spectra derived from prostate cancer biopsies with Gleason grade and tumour stage,” European Urology, vol. 50, no. 4, pp. 750-761, 2006.

[31] L.E. Jamieson, H.J. Byrne. 2016. Vibrational spectroscopy as a tool for studying drug-cell interaction: could high throughput vibrational spectroscopic screening improve drug development?Vib. Spectrosc. In press. http://dx.doi.org/10.1016/j.vibspec.2016.09.003

7 APPENDIX





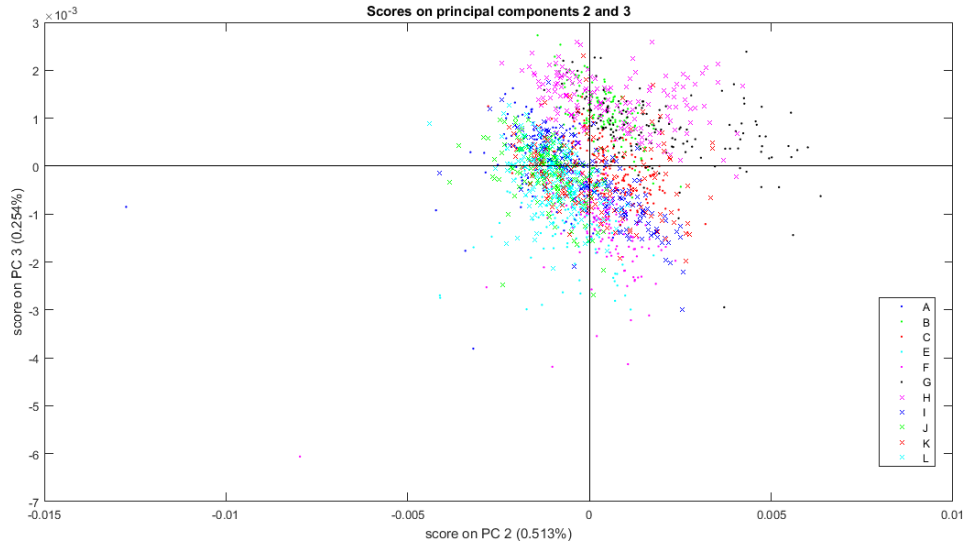


Figure 22 - Scores plot (PC12, PC13, PC23) of glass slides except slide D

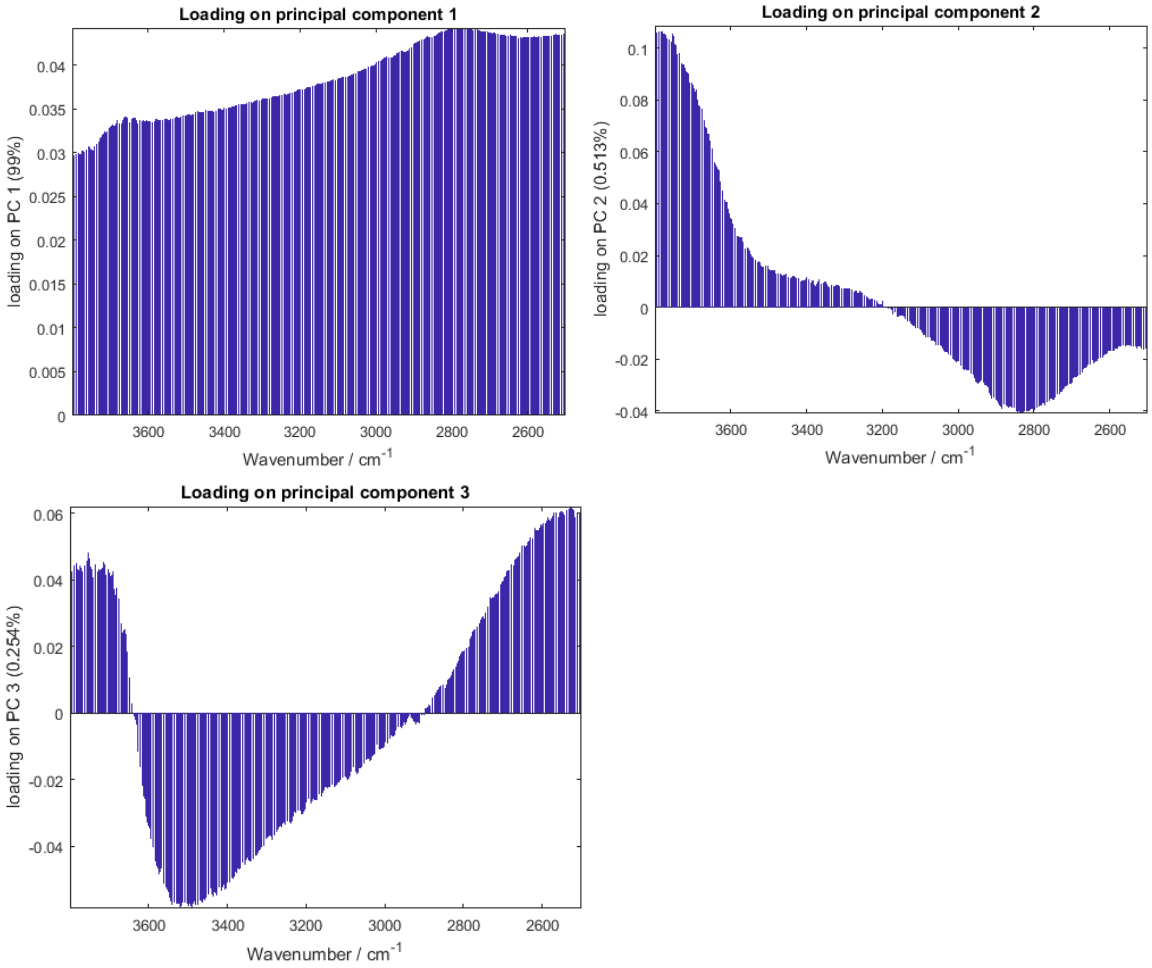


Figure 23 - Loading plots (PC1-3) for scores plot without slide D