Translating Cancer Research to the Clinic

Shuter, Olivia Grace
Student ID: 201500148
sgoshute@liverpool.ac.uk
University of Liverpool

Prof. Peter Weightman
Supervisor
sc35@liverpool.ac.uk

Prof. Tim Greenshaw
Supervisor
green@liverpool.ac.uk

April 2025

I Declaration

I certify that this project is my own work, based on my personal study and/or research and that I have acknowledged all material and sources used in its preparation, whether they be books, articles, reports, lecture notes, and any other kind of document, electronic or personal communication. I confirm that I have read and understood the University's Academic Integrity Policy. I confirm that I have acted honestly, ethically and professionally in conduct leading to assessment for the programme of study. I confirm that I have not copied material from another source, nor colluded with any other student in the preparation and production of this work.

II Abstract

Accurate classification of oral epithelial lesions into benign tumour and malignant tumour categories is critical for timely diagnosis and improved patient outcomes. Currently, in the clinic, a predictive power for malignancy in these biopsies of 25-40% is seen when a histopathologist alone examines the biopsy. As a result, patients must attend painful biopsy procedures year-on-year when the statistics show that only 1 in 8 will become malignant. Fourier transform infrared (FTIR) imaging offers an alternative and potentially very powerful approach to characterising tissue. The fingerprint band (the wavenumber range from 1000–1800 cm^{-1}) of the FTIR spectrum is currently used in such studies, using tissue samples mounted on CaF_2 slides as standard microscope glass slides absorb strongly in the fingerprint band. This project investigates whether the functional band (from 2200 to 4000 cm^{-1}) accessible using glass slides, allows lesion classification as reliable as that obtainable using CaF_2. This project used FTIR data from 17 oral biopsies (10 malignant and 7 benign), and trained a multi-layer perceptron (MLP) on pixel-level spectra. On unseen data, the functional band achieved 99% $\pm$ 0.4% specificity and 77% $\pm$ 5.5% sensitivity (95% CI), compared to 99% $\pm$ 0.6% specificity and 89% $\pm$ 3.1% sensitivity when the fingerprint band is used, demonstrating only a modest performance trade-off. This is a significant performance improvement from the current ability of a histopathologist to predict the progression of the lesion (22%-40% accuracy). When tests were carried out using biopsies that were not represented in the training sample, for many samples the performance was similar to that using unseen data of samples that were present in the training dataset, though this was not the case for all unrepresented samples: further study is required here. It has also been shown for the first time that an MLP can select malignant regions beyond those identified by the histopathologist, a finding supported using unsupervised k-means clustering. These may represent areas where the additional information provided by the IR spectra allow improvements over purely manual annotation. Achieving near-equivalent classification using glass slides rather than CaF_2 simplifies the introduction of FTIR-based techniques in the existing histopathology workflow, facilitating clinical (and commercial) adoption and potentially reducing the number of painful biopsies patients with oral lesions must undergo while improving the accuracy of diagnoses.

Contents

III Glossary

FTIR: Fourier Transform Infrared

IR: Infrared

M: Malignant

B: Benign

MLP: Multi-Layer Perceptron

ML: Machine Learning

NN: Neural Network

CI: Confidence Interval

Epithelium: Layers of cells that line hollow organs and glands.

Malignant: A cancerous growth, characterised by abnormal cell division, uncontrolled growth, and the potential to spread to other parts of the body.

Benign: Growths or abnormalities on the skin that are non-cancerous and do not spread to other parts of the body.

Lesion: Area of abnormal or damaged tissue caused by injury, infection, or disease. A lesion can occur anywhere in or on the body, such as the skin, blood vessels, brain, and other organs. Examples of lesions include wounds, ulcers, abscesses, sores, cysts, and tumors. A lesion may be benign (not cancer) or malignant (cancer).

Stroma: The cells and tissues that support and give structure to organs, glands, or other tissues in the body. The stroma is mostly made up of connective tissue, blood vessels, lymphatic vessels, and nerves. It provides nutrients to the tissue or organ and removes waste and extra fluid.

Prognosis: The likely outcome or course of a disease; the chance of recovery or recurrence.

Artifact: In histopathology, an artifact is any structure or feature on a tissue slide that is not normally present in living tissue and is introduced during tissue preparation or processing (blood, tissue folding, air bubble for example).

H&E Staining: Short for Hematoxylin and Eosin staining, is a widely used histological technique to visualise cellular structures under microscope. It utilises two dyes: hematoxylin, which stains nuclei a purplish-blue, and eosin, which stains the cytoplasm pink.

Phenotype: The observable physical properties of an organism; these include the organism's appearance, development, and behaviour.

Dysplasia: The presence of abnormal cells within a tissue or organ. Dysplasia is not cancer, but it may sometimes become cancer.

Motivation

Cancer is an often fatal disease for patients and a burden on healthcare facilities worldwide. Unfortunately, the situation continues to worsen with exposure to carcinogens such as smoking, alcohol and ultra-processed foods increasing. Cases of oral cancer have increased by 30% over the last 20 years in the UK, a trend which is not expected to slow down [1]. Histopathological examination is crucial in cancer diagnosis, and the field now sees a transition towards adopting digital pathology and advanced computational techniques with the aim of diagnoses being more efficient, accurate and reliable.

FTIR-based classification of biopsies mounted on calcium fluoride (CaF_2) slides has previously been attempted, using machine learning analysis methods [2]. The need for CaF_2 slides is one of the barriers to the broad introduction of these techniques: pathology laboratories customarily prepare their tissue samples on glass slides. By demonstrating equivalent FTIR classification on glass, a clear path is unlocked for histopathologists to adopt the following workflow. Hopefully, this will result in fewer needless painful biopsies for patients with oral lesions while ensuring they receive timely treatment in cases where the lesions will transform into malignant (meaning the lesion is characterised by abnormal cell division, uncontrolled growth, and the potential to spread to other parts of the body) cancers.

1 Experimental Techniques

In this chapter the theory of the techniques used in this project will be explored. Firstly the physics of optics will be introduced, followed by a theoretical description of infrared (IR) imaging techniques. Finally, how these IR spectra are applied in a clinical setting and why the choice of substrate determines which spectral bands are accessible for analysis is discussed.

1.1 Electromagnetic Radiation

The electromagnetic spectrum consists of visible, radiowave, microwave, ultraviolet, infrared, X-ray and γ-ray regions. The interactions between these radiations and matter are described by classical and quantum theories. Maxwell's equations depict the nature of various radiations. In this theory, radiation is described as two perpendicular electric and magnetic fields. These fields oscillate at 90° angles to each other in single planes. The polarity change in the electric and magnetic vector results in a sinusoidal wave, evolving as a function of position and time.

The periodicity of the wave allows it to be described in terms of frequency, ν and wavelength, λ. The wave speed is a constant with a value of $c = 3.0 \times 10^8 \text{ms}^{-1}$ in a vacuum, hence the relationship:

$$c = \lambda\nu. \tag{1}$$

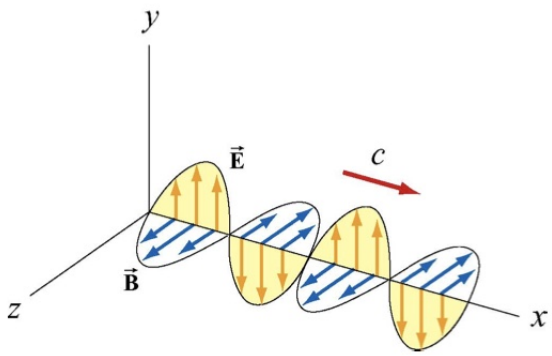


Figure 1: *Electromagnetic wave propagating in the x direction [3].*

The interaction between matter and EM radiation is described by the the refractive index of the material;

$$\underline{n} = n + i\kappa \tag{2}$$

where the real component represents the phase velocity of EM waves through a specified medium and the imaginary component evaluates the absorption of radiation in the same medium.

The plane wave propagates in the E and B fields according to Equations 3 and 4;

$$\mathrm{E}(z,t) = \mathrm{E}_0 \exp[\mathrm{i}(\underline{k}\mathrm{z} - \omega t)] \tag{3}$$

$$\mathrm{B}(z,t) = \mathrm{B}_0 \exp[\mathrm{i}(\underline{k}z - \omega t)] \tag{4}$$

where $\underline{k} = \frac{2\pi n}{\lambda}$. Substituting Equation 2 into Equation 3 and 4 gives the following relationship;

$$\mathrm{E}(z,t) = \exp\left(\frac{-2\pi\kappa z}{\lambda}\right) \cdot \mathrm{E}_0 \exp[\mathrm{i}(kz - \omega t)]. \tag{5}$$

The same process applies for the magnetic field. The relationship derived in Equation 5 proves that the imaginary component of the refractive index leads to absorption of EM radiation. This is shown by an increasing value of κ, providing a greater increase in exponential decay [4].

EM radiation interacts according to the wavelength when traversing a certain medium. This can be shown by Equation 6, which shows that the complex refractive index is a function of wavenumber $k = 1/\lambda$:

$$\tilde{n}(k) = n(k) + i\,\kappa(k)\,, \tag{6}$$

where k is the wavenumber in cm^{-1}.

Snell's Law (Equation 7) describes the relationship between the angle of incidence, θ_i, and refraction, θ_t, when incident light passes a boundary from one medium to another:

$$\frac{n_1(k)}{n_2(k)} = \frac{\sin\theta_t}{\sin\theta_i}\,. \tag{7}$$

1.2 Interactions of IR with matter

The way in which IR radiation interacts with matter is described in terms of changes in molecular dipoles. Consider a diatomic model, which has 3 degrees of translational freedom and two degrees of rotational freedom. A diatomic molecule has one degree of vibrational freedom, corresponding to the allowed stretch and compression of the molecule. Similarly, molecules containing many atoms will have $3N$ degrees of freedom. The differing vibrational states available to a molecule are also known as its vibrational modes [5].

To illustrate the frequency of vibrational modes, Hooke's Law (Equation 8) can be used and the system modelled as follows;

$$F = -kx. \tag{8}$$

Hooke's Law shows the relationship between displacement from the equilibrium, x, and the force, F, of a spring. The spring constant k acts as a scalar. On the quantum scales, the potential energy of this system can be calculated using Schrodinger's equation. A set of discrete energy levels are dictated by the following equation;

$$V = \left(v + \frac{1}{2}\right) h\nu. \tag{9}$$

The concept of reduced mass can be used to simplify equations by combining masses of the atoms;

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \tag{10}$$

where μ is the reduced mass. Now the relationship between force constant, the reduced mass and the frequency of absorption (considering wavenumber values for bond vibrational frequencies) can be stated;

$$\overline{\nu} = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}}. \tag{11}$$

Infrared radiation can only be absorbed when the incident radiation is of equal frequency to one of the vibrational modes of the molecule.

1.3 Fourier-Transform Infrared Spectrometers

1.3.1 Interferometer

The core of an FTIR spectrometer is a Michelson interferometer, which uses a broadband IR source whose emission is modulated by splitting the beam, sending one half to a fixed mirror and the other half to a linear translating mirror (known as a "moving" mirror), then recombining them. As the linear translating mirror changes position, the path difference Δ varies, so the constructive and destructive interference associated with each setting of the interferometer produces a beam with different combinations of wavenumbers that is used to interrogate the sample. Repeating this scan rapidly at tens of kilohertz produces a time domain signal (an interferogram) in which each point corresponds to a known Δ and therefore encodes contributions from all wavenumbers. A view of how every wavenumber maps into the interferogram (power as a function of mirror-path difference and wavelength) is found in Appendix A. A diagram of an interferometer is shown in Figure 2.

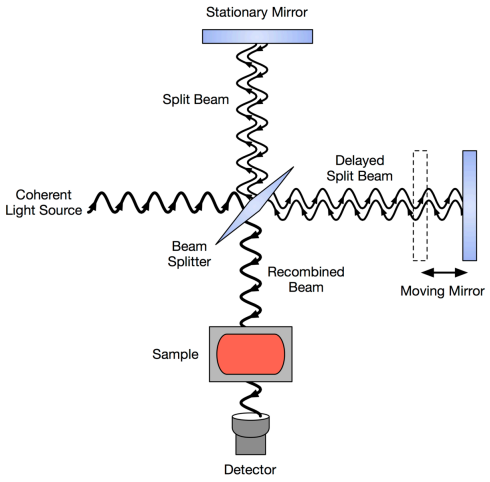


Figure 2: *Schematic of a Michelson interferometer used in FTIR. A broadband IR beam is split, reflected off a fixed and a linear translating mirror, then recombined and passed through the sample to produce an interferogram at the detector [6].*

1.3.2 Producing the FTIR spectrum

Fourier Transform Infrared Spectrometry (FTIR) exploits this interferogram by applying a mathematical Fourier transform to recover the underlying intensity as a function of wavenumber. Because the interferometer modulates *all* frequencies simultaneously, FTIR can collect the entire IR spectrum in one rapid scan, rather than stepping sequentially through each wavelength as in dispersive or tunable laser systems.

The intensity recorded at mirror position Δ is

$$I(\Delta) = \tfrac{1}{2} I_0\big(1 + \cos(2\pi\,\sigma\,\Delta)\big), \tag{12}$$

and integrating over the broadband source gives

$$I(\Delta) = \tfrac{1}{2}\int_0^\infty I(\sigma)\big[1 + \cos(2\pi\,\sigma\,\Delta)\big]\,d\sigma = I_{DC} + I_{AC}(\Delta), \tag{13}$$

where I_{AC} is the interferogram term of interest. Taking the Fourier transform then yields the final spectrum,

$$I(\sigma) = 2\int_{-\infty}^{\infty} I_{AC}(\Delta)\, e^{-\,i\pi\,\sigma\,\Delta}\,d\Delta. \tag{14}$$

A single wavenumber produces a cosine fringe in both field amplitude and detector power (see Appendix A). FTIR can simultaneously collect spectral data over a broad spectral range and, due to the nature of data acquisition, the complete IR spectrum of a sample can be obtained quickly.

The basic components of an FTIR spectrometer are;

To summarise, the Michelson interferometer provides a rapid broadband modulation which encodes all wavenumbers into a single interferogram. The subsequent Fourier transform disentangles each wavenumber's contribution, producing a high resolution spectrum, orders of magnitude faster than conventional scanning or dispersive instruments.

1.4 Application in a clinical setting

Infrared spectrometry has potential for clinical use because particular combinations of atoms and bonds in biomolecules absorb IR light at characteristic frequencies, where the natural oscillation frequency of each bond matches the incident radiation. By mapping these vibrational modes in tissue images, the molecular composition of the sample can be inferred. Additionally, by working on glass (instead of CaF_2), this technique can merge directly into existing pathology labs, making it far more likely to secure industry partnerships (where £10m+ in validation and marketing is at

stake) and thus cross the "valley of death" [7] from academic proof of concept to routine clinical use.

Since the middle of 20th century, infrared spectroscopy has been recognised as a non-destructive, label-free, sensitive and specific analytical method with many potential useful applications in different fields of biomedical research, in particular cancer research [8]. This is due to the nature of biological specimens; in general they all have a similar chemical structure at the scales studied with FTIR.

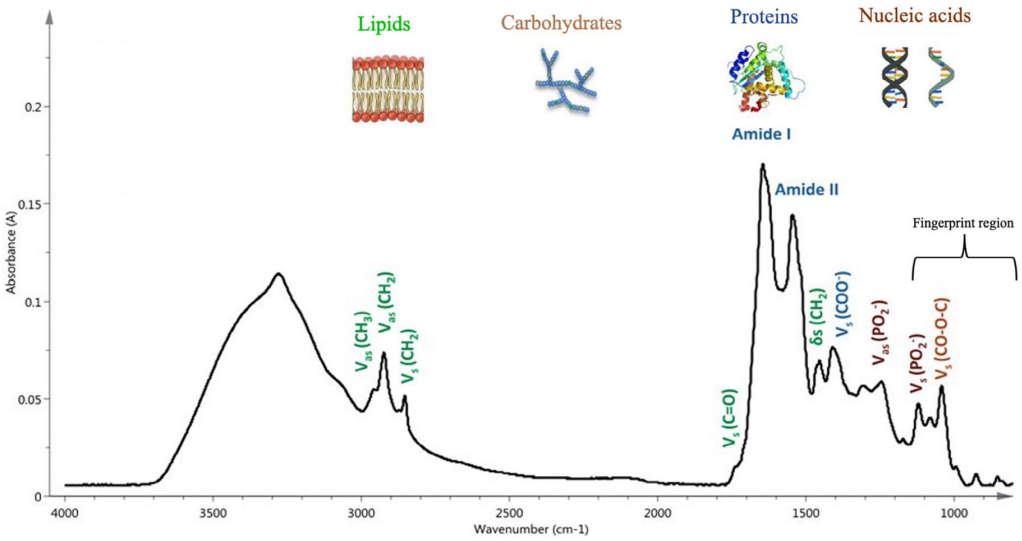
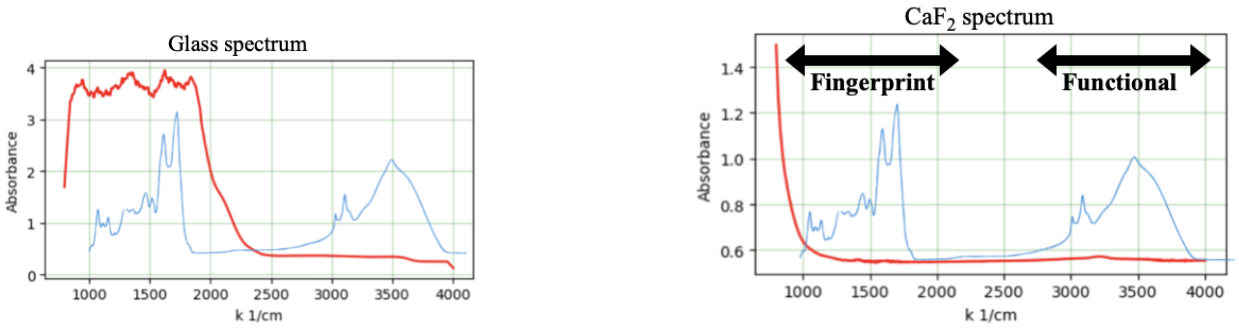


Figure 3: *A typical biological spectrum and corresponding biological components [9].*

A typical biological spectrum is shown in Figure 3. The amide I and amide II peaks shown are the dominant peaks, located at $k = 1640\,\mathrm{cm}^{-1}$ and $k = 1550\,\mathrm{cm}^{-1}$ respectively. These peaks arise from the contributions from proteins which have varying structural conformation. Proteins are found in all biological tissue and so their peaks are a key feature of these IR spectra. The various modes of CH_2 seen around $k = 2800\,\mathrm{cm}^{-1}$ in Figure 3 are also typically found in biological IR spectra, representing the lipids found. There is a plethora of information regarding the nature of a sample in a biological spectrum. This is both an advantage and disadvantage, due to the fact that a set of spectral profiles may only differ very slightly from one to the next. To solve this issue, analytical methods must be applied to extract the relevant information.

The reliability of the methods used to study biological samples is of high importance. The conclusions obtained from implementing these methods will eventually be used to determine whether tissue will become malignant, having real world consequences for patient diagnosis and prognosis. Hence, a robust approach and verification process is essential. The application of FTIR spectroscopy combined with analytical techniques underpins the following work.

1.4.1 Optical Transmission of Glass vs. Calcium Fluoride



(a) (b)

Figure 4: *(a) Optical absorbance of a standard microscope glass slide (red) overlaid with a representative FTIR tissue spectrum (blue) acquired on CaF_2; (b) absorbance of a CaF_2 substrate (red) with the same tissue spectrum (blue) and the* fingerprint *(1000–1800 cm^{-1}) and* functional *(2200–4000 cm^{-1}) regions indicated.*

Figures 4 and 5b demonstrate why CaF_2 is preferred as a medium to examine biopsies. The red trace shows that common microscope glass exhibits very high absorbance in the fingerprint band (1000–1800 cm^{-1}), obscuring the biochemical signatures found there. In contrast, CaF_2 has negligible absorbance across both the fingerprint and functional (2200–4000 cm^{-1}) regions, allowing the full spectral range of the tissue (blue trace) to be recorded. It is clear why this is preferred, but current histopathology practice utilises glass slides and the introduction of the expensive and hard to handle CaF_2 presents a significant hurdle. Since the fingerprint band carries key molecular information (such as protein, lipid and nucleic acid vibrations), its obstruction by glass poses a challenge.

The central question of this work is therefore: *can the functional band alone (accessible on the already used glass slides) provide sufficient contrast for accurate classification, matching the performance previously achieved on CaF_2 substrates* [2]?

Figure 5 illustrates the experimental results of an investigation into the effects of the type of glass used, since slides and two thicknesses of cover slip are readily available as an option for laboratory use. It is clear that the thinner cover slip (Figure 5d) restricts a smaller range of the FTIR window than the thicker glass substrates: the absorbance "drops off" at a lower k value than for the glass slide and thick cover slip, and it may be possible to recover part of the important fingerprint band as a result of this. (Note that in Figure 5, high absorbance measurements also show substantially increased uncertainty. This can be accounted for by photon counting statistics; when N photons are detected, the absolute error is $\propto \sqrt{N}$ and the relative error $\propto 1/\sqrt{N}$ which becomes large for small N or high absorbance.)

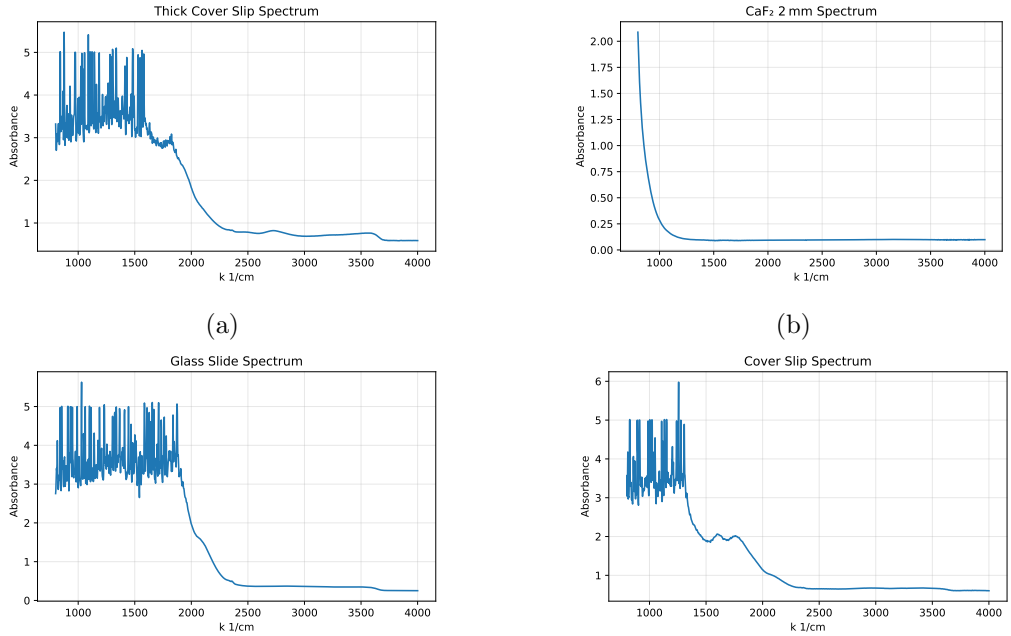


Figure 5: *Comparison of substrate transmission windows. Glass blocks nearly all signal below ∼2000 cm^{-1} (a,c,d), whereas CaF_2 remains transparent across both the fingerprint and functional bands (b).*

2 Analysis

The medical sector has, like many other sectors, been exposed to the applications of machine learning (ML). Here it can be particularly useful due to the complex and multidimensional structure of the data used. Without the advantages that an ML model provides, the patterns in such datasets would be difficult and time consuming to recognise. An ML analysis will be used to classify tissue as malignant (M) or benign (B) (benign growths or abnormalities are non-cancerous and do not spread to other parts of the body), and therefore predict patient outcomes. Whereas many prior FTIR studies have relied on unsupervised methods (for example cluster analysis [10], k-means clustering [11] and principal component analysis [12]) without an independent ground truth, here fully supervised learning is used. The biopsies that the model is trained on have confirmed outcomes and so the accuracy of the model can be concretely assessed. This section details the structure and workings of the ML pipeline and how this can allow for results more accurate than currently seen in a clinical setting to be obtained.

2.1 Preprocessing

The format of the data that is used can significantly influence the training of the model and hence its performance [13]. The raw data (the FTIR hyperspectral cubes) must be preprocessed into a compact feature matrix X and label vector y. The data for the ML model in this project is prepared as follows. See Appendix B for additional details of this process.

After the steps outlined in this section, the model inputs of $(X_{\text{train}}, y_{\text{train}})$ and $(X_{\text{test}}, y_{\text{test}})$ are compact, noise-reduced, scaled and normalised (if chosen) spectral feature matrices ready for multilayer perceptron (MLP) training and evaluation.

2.1.1 Spectral Normalisation

Each pixel's absorbance spectrum can be normalised to remove intensity-scale variations (for example due to varying sample thickness or scattering effects) whilst keeping the meaningful contrasts which allow the model to identify cell type. Three normalisation methods are considered in this work:

Global product normalisation: Each spectrum $\{A_i\}_{i=1}^{N}$ is divided by its geometric mean:

$$X_i \;=\; \frac{A_i}{\left(\prod_{j=1}^{N} A_j\right)^{1/N}} \;=\; \frac{A_i}{\exp\left[\frac{1}{N}\sum_{j=1}^{N}\log A_j\right]}\,.$$

By constructing $\prod_i X_i = 1$, the scaling removes multiplicative path-length effects whilst keeping relative peak ratios across the full region.

Local ratio normalisation: Normalise each consecutive pair of bands to highlight local transitions:

$$X_i \;=\; \frac{A_i}{A_{i-1}}\,, \quad i = 2, \ldots, N, \quad X_1 = 0.$$

Sharp spectral features are emphasised (for example individual vibrational modes) by cancelling broad baselines, but allows noise to be amplified when A_{i-1} is small.

Local difference normalisation: This method is an additive analogue of the local ratio method, computing 1st order differences:

$$X_i \;=\; A_i - A_{i-1}\,, \quad i = 2, \ldots, N, \quad X_1 = 0.$$

Slowly varying backgrounds are removed but the sign and magnitude of each transition are maintained.

In summary, each of these normalisation methods trades off preservation of absolute peak intensities against removing trivial scale effects. By comparing all three, it is ensured that the classifier can exploit both larger scale and subtle biochemical contrasts that are relevant for diagnosis in FTIR histopathology.

2.2 Machine Learning

Machine learning (ML) is a field of artificial intelligence in which algorithms learn to recognise patterns and make predictions from complex, multidimensional data without explicit programming. Biomedical FTIR datasets (such as hyperspectral maps of tissue biopsies) are ideally suited to ML, since they contain subtle spectral signatures that would be difficult and time-consuming to extract by hand. Deep learning methods (described in the following sections) employ stacked non-linear transformations to automatically refine informative features and detect faint differences across large volumes of spectra.

Most ML algorithms fall into two broad classes: supervised and unsupervised learning. In supervised learning, each input spectrum $\mathbf{x}_i$ is paired with a ground-truth label y_i, and the model learns a mapping $f : \mathbf{x} \to y$ by optimising a set of parameters during training. Classification algorithms are a subclass of supervised learning algorithms and use discrete labels (here plain, benign and malignant) to define decision boundaries that can be evaluated in terms of accuracy, sensitivity and specificity. By contrast, unsupervised learning (for example clustering) seeks intrinsic structure in $\{\mathbf{x}_i\}$ without reference to labels, making quantitative assessment more challenging when the ground truth is available.

In this work a supervised approach is adopted: each biopsy spectrum is associated with a confirmed diagnostic outcome, so that the learned classifier can be conclusively evaluated on held-out (unseen to the model) samples. The known labels y enable concrete measures of performance and ensure that the pipeline not only fits the training data but also generalises to new tissue specimens.

2.3 Neural Network and Deep Learning

Neural Networks (NN) are a type of ML model. Deep learning models are constructed by a stack of layers (the phrase "deep model" means multiple layers) with non-linear functions which have the capability to learn important features from input data [14]. They provide a powerful tool for supervised learning, where all data is labelled and the model has an answer to "aim" for. Consecutive layers, and nodes within those layers, can be added to form a Deep Network that represents increasingly intricate functions. These NN models can be thought of as mimicking how the brain works [15] as they are formed of nodes connected together, akin to the neurons in the brain.

2.4 Model Architecture

2.4.1 Multi-Layer Perceptron (MLP)

A **Multi-Layer Perceptron (MLP)** is a type of artificial NN widely used for classification and regression tasks. It belongs to the family of *feed-forward* neural networks, meaning that data flows in one direction (from the input layer, through one or more hidden layers, to the output layer) without any cycle or loop feature. An MLP consists of three main components:

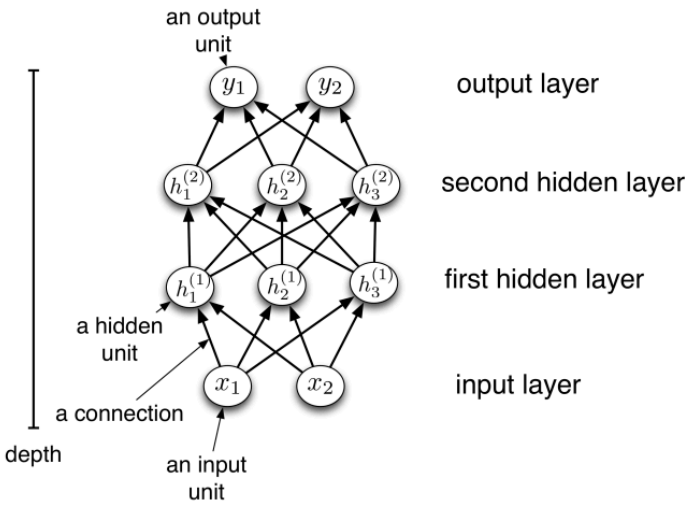


Figure 6: *Schematic representation of an MLP architecture, showing an input layer, two hidden layers, and an output layer. Each node in a layer is connected to every node in the subsequent layer.*

Input Layer: Receives the raw data. In the context of this project, each input node corresponds to a feature derived from the spectral data (for example transmission values at different wavenumbers).

Hidden Layers: One or more layers where the computation occurs. Each node in a hidden layer applies a weighted sum to its inputs, adds a bias term, and passes the result through a non-linear activation function (Section 2.4.5). These layers allow the network to learn complex, non-linear relationships within the data. A "hidden" layer refers to layers not directly exposed to input or output (hence it is "hidden" from the external environment).

Output Layer: The final prediction is calculated. For classification tasks, such as distinguishing between different types of tissue in this project, the output layer typically uses non-linear activation function to provide class probabilities.

The learning process involves adjusting the weights and biases associated with each connection in the network using a method called backpropagation (the algorithm which a NN uses to compute

the gradient of its loss with respect to each weight), in conjunction with an optimisation algorithm such as stochastic gradient descent (SGD) or Adam. The goal is to minimise a chosen loss function (Section 2.4.4), which quantifies the difference between the predicted and actual class labels during training and hence gives the user a metric of how well the model has performed.

An MLP is effective in this project, and in general when dealing with datasets where relationships between features are non-linear and not easily separable using simpler models. An example of such a model might be logistic regression, which imposes a straight line decision boundary in the feature space, but if some malignant and benign spectra overlap in an absorbance band then they would be incorrectly classified a lot of the time. An MLP can map curved, piecewise-linear boundaries that will capture these higher-order interactions between spectral bands, resulting in better discrimination on such data. An MLP is therefore a suitable choice for capturing these patterns.

2.4.2 K-means clustering

K-means clustering is a simple yet powerful *unsupervised* learning technique that seeks to partition N observations into K clusters by minimising the total variance within the cluster [16]. Starting from an initial set of K centroids, the algorithm alternates between assigning each spectrum to the nearest centroid and then repositioning each centroid to the mean of its assigned points. This method scales linearly with both the number of spectra and the number of clusters, making it well suited to high-dimensional FTIR data. In the context of tissue classification, k-means clustering can reveal intrinsic groups (such as plain, benign and malignant regions) based on the spectral signatures, providing an unsupervised baseline against which supervised classifiers (for example the MLP) can be compared.

2.4.3 Batch Standardisation

Large datasets with many different parameters, such as the ones considered in this project, make the task of the ML model quite difficult. Many ML algorithms function best when all features (input variables) are centered around 0 and have similar variances [17]. In particular, training can be adversely affected if the input variables have very different scales and/or variances. To resolve this issue, normalisation is required.

The features are standardised by removing the mean and scaling to unit variance. The standard score of a sample is calculated to be;

$$z = \frac{(x - \mu)}{\sigma} \tag{15}$$

where μ is the mean value of the sample data and σ is the standard deviation. This is also known

as 'Z-score' normalisation. Scikit-learn's [17] "StandardScaler" is used to implement this in the preprocessing of the data.

2.4.4 Loss Functions

The loss function represents the deviation between the MLP's predictions and the desired targets (ground truth results). The goal of training procedure is to minimise the loss function, hence obtaining the most accurate predictions possible. This, as mentioned above, is achieved by optimising the weights in the MLP. Mathematically this can be represented as;

$$\mathbb{E}[\Delta(\mathbf{y}_i, g(f(\mathbf{y}_i))] \tag{16}$$

where g and f are functions received from hidden layers, and $\mathbf{y}$ is a target variable. In summary, the model tries to find the weights to learn to represent an approximation of the function. The goal is to find the weights within the structure to obtain the smallest difference according to some metric $\Delta(\cdot)$ between y_i and y_i [18].

2.4.5 Activation Functions

An activation function (AF) of a node is a function that calculates the output of the node (based on its inputs and the weights on individual inputs) [19]. There are a variety of different AFs, each of which are appropriate for use in different ML scenarios. In general, they should aim to add non-linearity to improve training convergence, and should not increase the complexity of the model. The distribution of data should be retained to ensure better training of the model [20]. Non-linearity is important to allow the model to learn complex relationships in the data. A linear activation, where no "transform" is applied at all, would be simple for a model to train but would not learn any complex mapping functions. The non-linear activation then enables non-linear patterns and variations in the data to be detected.

Rectified Linear Unit, or ReLU [21], has become a widely used AF. Its function can be illustrated as follows;

$$\text{ReLu}(x) = \max(0, x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

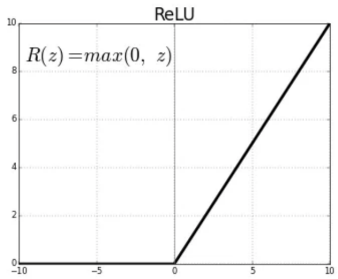


Figure 7: *Graphical representation of ReLU activation function. It has a piece-wise linear nature, characterised by a flat response for negative input values (0 gradient) and a linear increase for positive input values (gradient of 1) [21].*

2.4.6 Dropout Regularisation

Deep NNs with many parameters are known to be prone to overfitting (imagine you are learning maths and practice with the same questions every day, you'd become very good at solving these specific problems, but if given a new problem, you would struggle. This is analogous to overfitting in NNs). Overfitting is most likely seen when the training set is limited or highly correlated (as can often be the case with FTIR pixel spectra). Dropout [22] is a simple yet powerful regularisation technique that mitigates this risk by randomly "dropping" a fraction of the hidden nodes during each training update. During backpropagation gradients flow only through the "surviving" nodes, preventing any one node from becoming overly specialised.

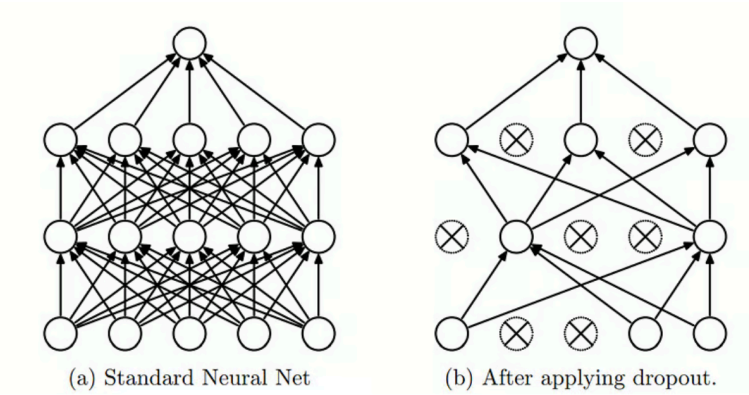


Figure 8: *Illustration of a fully-connected layer before and after applying dropout. Left: standard dense connections. Right: a subset of nodes are temporarily deactivated encouraging a more robust, distributed representation [23].*

By randomly omitting nodes at each update, dropout effectively trains a group of "thinned out" sub-networks and averages their predictions at test time, where all nodes are restored but activations are scaled by the dropout probability. Using this method will allow exploration of whether there is overfitting to particular patterns, which would in effect degrade the generalisation across biopsies.

2.4.7 Model Evaluation

To evaluate the performance on the testing dataset, the concepts of sensitivity and specificity are used. A testing method within this context should be able to detect positive (confirmed disease/malignant) and negative (confirmed healthy/benign) patients. If these instances are not detected correctly, then there is the potential for highly undesirable real world consequences. In the case that many positive cases are missed, patients will go with their disease undetected and will not receive appropriate treatment. Conversely, a patient being incorrectly identified as having the disease will undergo unnecessary painful biopsies, experience needless anxiety and incur additional healthcare costs.

$$TPR = \frac{TP}{TP + FN} \tag{17}$$

$$TNR = \frac{TN}{TN + FP} \tag{18}$$

True Positive Rate (TPR), or sensitivity (Equation 17), quantifies a test's ability to correctly diagnose a positive case. A high sensitivity will rarely fail to find a true positive case. True Negative Rate (TNR), or specificity (Equation 18) quantifies a test's ability to correctly reject a positive diagnosis.

To further illustrate this concept, a 100% sensitive test with a low specificity value means that a true negative case will have a high probability of being diagnosed as positive (or having the disease). A 100% specific test with low sensitivity means that a true positive case has a high probability of being diagnosed as negative (or not having the disease). Clearly, it is important to consider both of these metrics. A confusion matrix, shown in Appendix B (Figure 16), is a useful visual aid to see how well the model has performed.

In order to ensure an unbiased estimate of the classifier's performance, the data is split at the biopsy level rather than at the pixel level. Let

$$X \in \mathbb{R}^{N \times 16}, \quad y \in \{0, 1, 2\}^N$$

define the full feature matrix and label vector over all pixels. We then choose a subset of whole biopsies for training (with all of their pixels), hold out one (or more) entire biopsies for testing, and never mix pixels from the same slide between the two sets. By holding out entire biopsies from training, the network cannot rely on slide-specific features (overfitting and not "learning"). The reported metrics therefore reflect genuine generalisation to new specimens.

Generalisation To test generalisation across different biopsies, one biopsy's pixels are used for training and a second held out biopsy is used for testing. After training, the model is run back on

the held-out pixel set to obtain class predictions. The log-loss and overall accuracy is recorded and the process is repeated over all biopsy pairs to build a full "loss-matrix" that quantifies transfer performance across the 17 biopsies.

Error Analysis

The statistical uncertainty values are calculated as described in Appendix B. Note that these values represent only the *statistical* (sampling) variability across biopsies. This approach is used to calculate the error on the quoted TPR and TNR values in Section 3. Several additional sources of error are not considered in the calculated confidence intervals (CIs). Typical reproducibility of absorbance measurements on the FTIR equipment is thought to be of the order of 2–3%, arising from a combination of factors such as source-intensity fluctuations, detector noise and temperature and humidity changes [24]. A Monte Carlo simulation of the FTIR equipment would be ideal to enable a more thorough analysis of uncertainty.

A concerning source of systematic error is inter-observer agreement on individual dysplasia features of oral epithelial biopsies. Frequent disagreement was found between histopathologists on key structural criteria [25], with "poor to moderate" agreement seen regarding key features of the biopsies. Such a variability in grading these lesions can result in a measurable mismatch of around 5-10% at the pixel level, which can be treated as a bias term in an error calculation. Additionally, sample preparation may likely present a source of systematic error. Variations in section thickness, mounting medium, or cutting angle could introduce further uncertainty at the level of a few per cent. Additionally, normalisation and preprocessing can bias the spectral features used by the MLP, affected by choice of baseline correction or smoothing parameters for instance. Under the usual assumption that these systematic errors are independent of the sampling variability, one may form a rough total uncertainty estimate by combining in quadrature:

$$\sigma_{\text{total}} \approx \sqrt{\left(\text{MOE}_{95\%}\right)^2 + \sum_j \sigma_{\text{sys,j}}^2},$$

where each $\sigma_{\text{sys,j}}$ is the estimated standard deviation of a systematic effect and MOE is the margin of error at the 95% confidence interval. In practice, however, such estimates are often based on manufacturer specifications or smaller ancillary studies, and are reported separately from the statistical CIs.

2.5 Method

After preprocessing, the pipeline trains an MLP NN classifier for the task of identifying the tissue imaged in a pixel as being malignant (M), benign (B) or plain (no lesion observed). The workings of the MLP are described in Section 2.4. After training, the model is tested on unseen biopsies and calculates metrics such as log-loss, TPR and TNR. The overall pipeline consists of four stages: data preparation, feature construction, model training, and model evaluation.

First, two Boolean masks (true/false) are built. One is a tissue vs background mask from averaging absorbance over the full spectral range and thresholding at the 10^{th} percentile. The other is an MB-mask (malignant vs benign), by marking pixels flagged in the biopsy metadata by a histopathologist.

A sample of wavenumbers is selected between a chosen range (depending on which band is desired for the model), and for each pixel inside the tissue mask the normalised spectral features are extracted, giving a 3D array of shape (rows, cols, 16). The label map is constructed;

$$y_{r,c} = b_{\text{type}} \times \text{NT\_mask}_{r,c} \ \in \ \{0, 1, 2\},$$

so that plain pixels are assigned 0, benign pixels 1, and malignant pixels are 2. Finally, the tissue pixels are flattened into a 2D feature matrix;

$$X \in \mathbb{R}^{N \times 16} \quad \text{and} \quad y \in \{0, 1, 2\}^N,$$

and each feature is standardised (Section 2.4.3) to zero mean and unit variance. Each sample fed to the classifier is a valid tissue pixel with a consistent feature vector and a clear integer label.

Model architecture and training: To investigate general model performance, a Scikit-learn MLPClassifier [17] with four hidden layers of sixty nodes each and ReLU activation functions is used. The output layer consists of three logits (one per class: 0, 1, 2 for plain, benign and malignant), and the network is trained by stochastic gradient descent (learning rate=0.001, momentum=0.9, L2 penalty α=1×10^{-5}) to minimise the cross-entropy loss. The model is fit for two iterations through the dataset of up to one thousand epochs (a complete pass through the training data) each, saving the state after each pass for incremental improvement. Each epoch processes all N pixels in smaller batches of 128 samples, updating weights via backpropagation and accumulating an epoch loss curve for monitoring the training process.

To investigate varying dropout layers, a PyTorch [26] MLP with four hidden layers each containing sixty nodes, a chosen activation function, and varying dropout layers (see Section 2.4.6) is built. PyTorch is used in this case as Scikit-learn's built-in MLP Classifier does not easily support adding dropout layers (it would require nontrivial backend manipulation), whereas PyTorch makes it simple to insert and configure dropout anywhere in the network. The final output layer has three possible outputs, one per class.

3 Results and Discussion

The aim of this project was to investigate whether using the functional band, and therefore glass, can be used to correctly classify lesions as malignant or benign. The sample size consists of 17 biopsies, which is small but indicates a positive result. This result is important nonetheless as the experiment is novel and it was not known *a priori* whether glass would prove to be usable. Even a single failure in this sample size would be sufficient to negate the idea and so this preliminary result demonstrates the need to perform larger, more detailed studies. If FTIR techniques on glass can

match CaF_2 performance, then clinicians could immediately harness this in their standard workflow and reduce painful repeat biopsies (the majority of which never become malignant) and catching the minority that do, earlier and with less invasive testing. For comparison in this section, one M and one B sample will be shown.

In the clinic, an accuracy of 25–40% is seen, derived from the Kaplan–Meier (survival) estimates of 5-year malignant transformation rates in lesions as graded or assessed by the pathologist. In a study by Ho et al. [27], "severe dysplasia" (the highest grade assigned by the pathologist) transforms in 33% of cases by five years, while lesions deemed to be "non-homogeneous" on initial clinical inspection transform in 38% of cases. Moderate dysplasia shows 24% at five years, and mild dysplasia only 16%. These statistics show that even the most indicative histopathological grades or appearances only allow a histopathologist to predict progression to cancer in roughly 25–40% of cases. **Can ML analysis improve upon these predictions?**

In the first instance, Sci-kit learn's MLP was used to build the model. An investigation into the best normalisation technique for the data was carried out. As mentioned in Section 2.1, the preprocessing of the data is a key part of the analysis pipeline. On a model with the same architecture only the normalisation scheme is changed in each instance and the results are reported in Table 1.

Table 1: Investigation of different normalisation schemes (2sf)

Normalisation	**TPR**	**TNR**
None	0.78 ± 0.067	0.98 ± 0.0080
Global product	0.77 ± 0.053	0.98 ± 0.0060
Local difference	0.63 ± 0.053	0.98 ± 0.0090
Local ratio	0.14 ± 0.047	0.96 ± 0.013

Table 1 shows that the chosen normalisation scheme of the data has a measurable impact on the model's ability to classify pixels correctly. Due to the superior performance in this stage of the analysis, the global product method of normalisation will be used for the rest of the analysis. Significantly worse results were seen for every normalisation with no scaling applied (as described in Section 2.4.3), so all samples are scaled. No measurable difference was seen when training the model with the "Adam" vs SGD optimiser. This could be explained the fact that the loss trend is relatively smooth, and the network is shallow enough that both optimisers converge to essentially the same minimum. Although Adam adapts per-parameter learning rates and SGD uses a single global step size (often with momentum, a technique used to smooth out the optimisation process), here the gradients are not sparse or very noisy, and the curvature of the cross-entropy surface doesn't vary massively between coordinates. With the chosen learning rate, batch size and training duration, both algorithms make very similar steps toward the optimum and so achieve indistinguishable test losses.

Table 2: Fingerprint vs Functional band on optimised architecture

Region of Spectrum	**TPR**	**TNR**
Fingerprint	0.89 ± 0.031	0.99 ± 0.0040
Functional	0.77 ± 0.055	0.99 ± 0.0062

Table 2 quantifies how well the model is able to detect M or B pixels on unseen testing data and therefore one can make conclusions on the predictive power of the model. The specificity (TNR) for both fingerprint and functional band is very high, which can be understood by the nature of the sample. The vast majority of the pixels in the FTIR dataset are characterised as morphologically "normal" epithelium, whose FTIR spectra are highly consistent in nature. For the model it is simple to recognise the single, dominant "healthy" signature than in the converse case of detecting a subtly variant disease phenotypes (the observable characteristics of the cancer). The model almost never mistakes a benign pixel for a malignant one. Additionally, the technique of the binary mask in training effectively balances the rarer malignant pixels against the abundant benign background which reinforces the penalty on the model for a false positive.

A notable difference in TPR, or sensitivity, value can be seen between the two regions. In the fingerprint band (Section 1.4) there is an information rich section of vibrational bands such as the amide I and II, phosphate and sugar ring modes and lipid bending, all of which will shift subtly yet reproducibly when tissue becomes malignant in nature. The MLP "learns" from the complex, distinctive patterns to give a TPR of 0.89 ± 0.031.

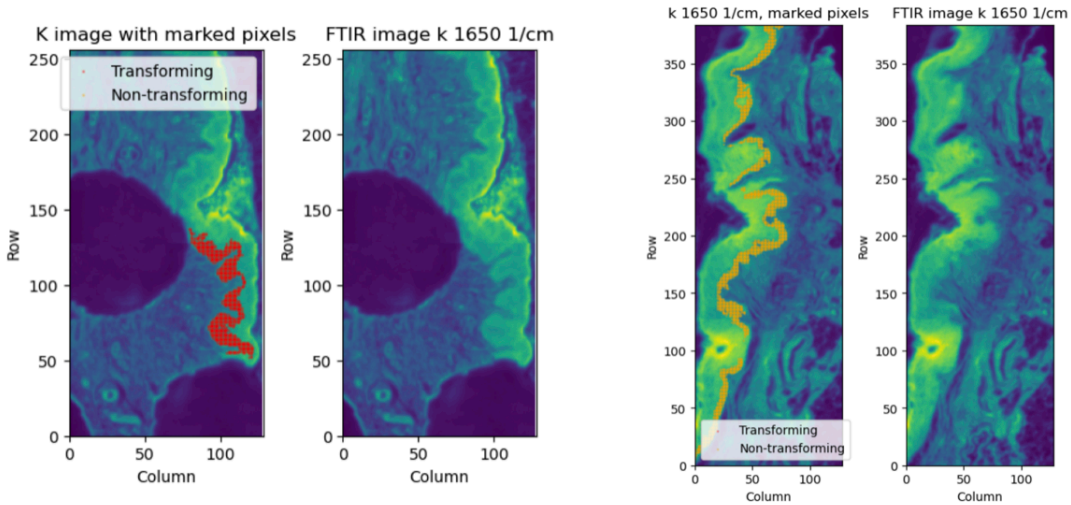


Figure 9: *FTIR images of (left) a biopsy containing a malignant region (coloured red) and (right) a biopsy containing a benign region (coloured yellow).*

The FTIR images of 2 samples are shown in Figure 9, illustrating how histopathologist annotations overlay the raw FTIR absorbance maps at a chosen wavenumber. On the left, the panel shows a

biopsy known to contain malignant tissue. A region of high absorbance (bright green/yellow) is indicative of protein-rich (due to the dominant amide I band, see Figure 3), potentially cancerous structures (confirmed cancerous in this work). In contrast, pixels labelled "non-transforming" (benign) lie predominantly in the darker background, corresponding to healthy stroma.

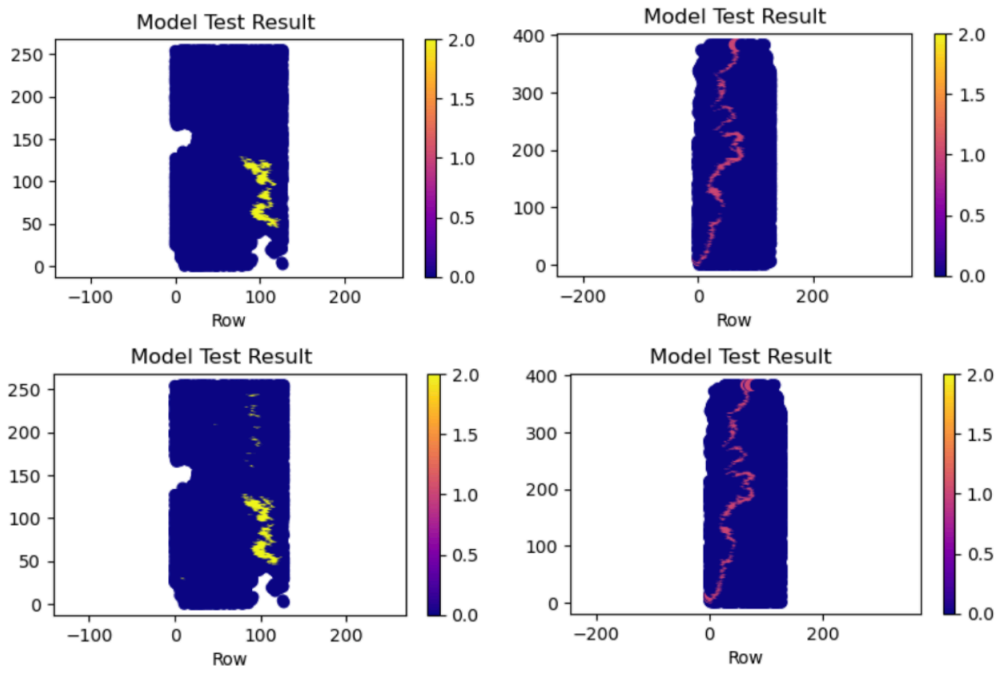


Figure 10: *Pixel MLP classification maps on two held-out biopsies, comparing fingerprint (top row) and functional (bottom row) bands. Each panel shows coordinates of every tissue pixel, coloured by the predicted class (0 = plain, 1 = benign, 2 = malignant). Top left/right use only fingerprint band inputs, and bottom left/right use only functional band inputs.*

Figure 10 compares the spatial distribution of pixel class predictions on two independent biopsies, when the MLP is given fingerprint band inputs (top left/right) versus functional band inputs (lower left/right). In each panel, tissue pixels predicted as background (class 0) are shown in dark blue, benign in pink (class 1), and malignant in yellow (class 2). When trained with the fingerprint band (top left/right), the model highlights malignant tissue which closely follows the histopathologist's marks. This reflects the high sensitivity calculated (TPR $\sim 0.89 \pm 0.031$) observed when the information-rich fingerprint band is available. In contrast, the functional band maps (lower left/right) still recover the main malignant area but in a narrower, more fragmented manner. Fewer pixels are labelled malignant and isolated false negatives become apparent, consistent with the lower TPR (0.77 ± 0.0055) in this band. Importantly, in both cases the plain tissue (healthy stroma) is almost never misclassified as malignant, demonstrating very high specificity across the slide. These results illustrate that although the functional band alone can identify the primary cancerous regions, access to the fingerprint bands yields more complete and spatially accurate margins.

To assess how well MLP generalises across independent tissue specimens, a train i, test j experiment

is performed: for each of the biopsies, a network is trained on a single biopsy and evaluated on each of the remaining biopsies independently, recording the cross-entropy test loss in a 17×17 "loss matrix" (Figure 11).

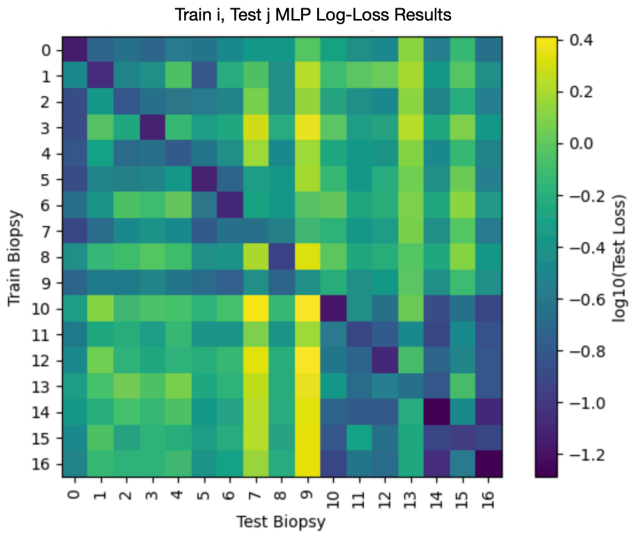


Figure 11: *Test loss for the MLP shown as a 17x17 matrix. Rows denote the training biopsy and columns denote the test biopsy; darker cells indicate lower loss (better generalisation).*

Figure 11 summarises how well a single biopsy trained MLP transfers to all other held-out biopsies (where the global product normalisation scheme is used). As expected, the darkest cells lie along the diagonal (training and testing on the same biopsy), corresponding to the lowest log10-loss (around 1.2), demonstrating that it can reproduce individual spectra well. Log10-loss is used for better interpretability of the results as values are of similar magnitude. Interestingly, many off-diagonal entries also remain relatively dark (log10-loss below -0.5), showing that the network has learned spectral features that generalise across multiple patients. However, there are clear lighter "bands" and isolated paler cells where the loss rises above zero, indicating specific biopsy pairs for which learned patterns do not transfer cleanly. Importantly, the fact that many held-out slides still show low loss values in the functional band shows that this method could reliably transfer to new patient samples without retraining. This is an essential requirement for any diagnostics tool facing potential commercial and regulatory scrutiny.

Certain rows (for example biopsy 7) and columns (for example biopsy 14–16) display consistently lower loss values across nearly all pairings, which could suggest that those samples have robust spectral signatures that the model finds especially reproducible. Overall, this result confirms that, even with just 17 biopsy slides, the MLP captures predominantly generalisable features of malignant versus benign tissue. The exact reason for the similarity or difference shown between pairs is unknown. A possible explanation is that features dependent on the biopsy preparation are being identified (for example in the properties of the paraffin used to prepare the biopsy). Another possible reason is that there are fundamental features of the biopsies on a cellular level which the model is detecting. This outcome is very interesting, and could suggest that there are 'signatures' of

cancer which we are currently unaware. This outcome presents motivation for future investigation in this area, and discovery of a generalisable signature of cancer would be groundbreaking.

Dropout Layer Results Different instances of dropout regularisation (Section 2.4.6) are investigated to give insight into whether the model developed may be overfitting.

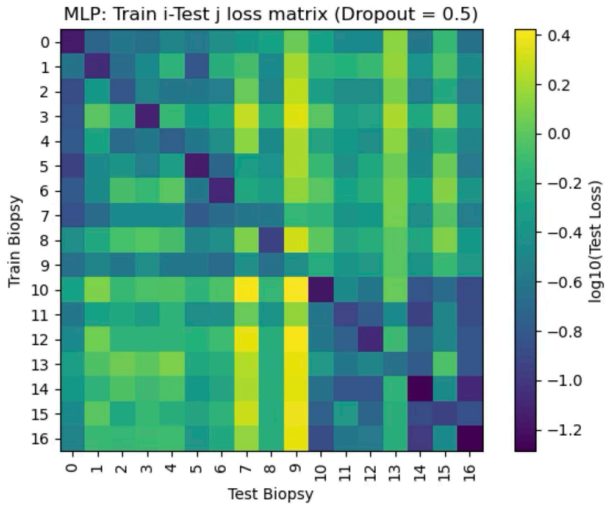


Figure 12: *Test loss for the MLP shown as a 17x17 matrix. A dropout of 0.5 is applied to the model.*

As can be seen by comparing Figure 11 and 12, the results are very similar when dropout layer is added. This suggests that the dominant source of variation is a genuine difference in the biopsy makeup. The model's limiting factor regarding cross-biopsy generalisation lies in the inter-slide heterogeneity of the tissue and preparation, not in the tendency of the architecture to overfit. Another result with a substantial dropout layer of 0.8 is shown in Appendix C to prove that little difference is seen, even with a dropout value considered to be very high.

A crucial point to make concerning this project is that the target values assigned to the training and test data are not 100% accurate. As discussed in Section 2.4.7, a source of error is inter-observer variability between histopathologists, and even for a particularly accurate histopathologist, malignant areas in a biopsy may be missed or even mis-labelled. It has been suggested in this work that the MLP model can outperform a histopathologist's predictions and preliminary results in this project hint that the ML models have potentially found areas missed by a histopathologist.

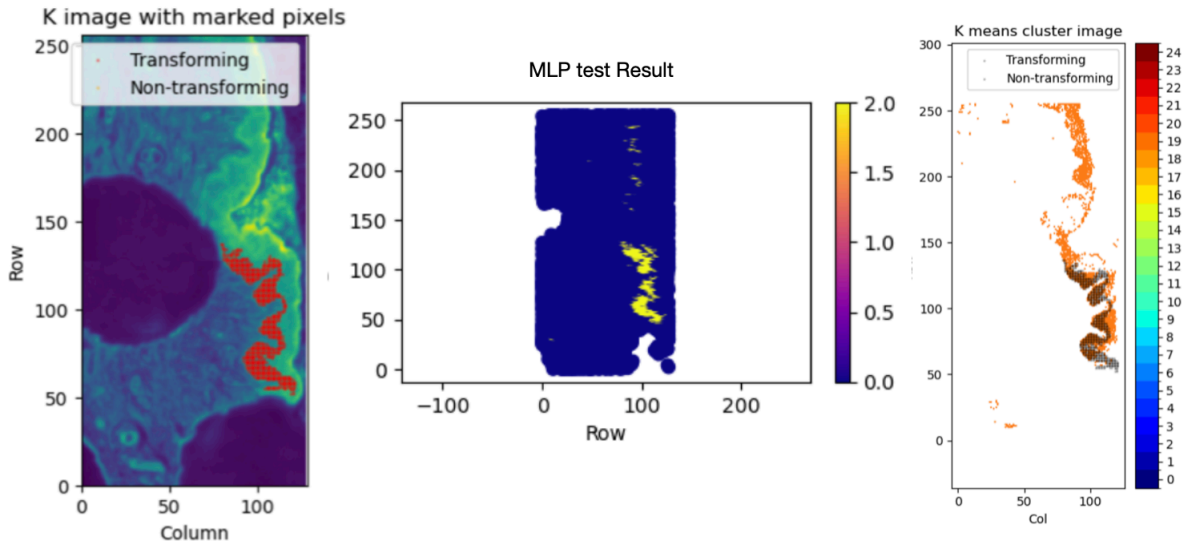


Figure 13: *Comparison of histopathologist annotation, MLP prediction, and unsupervised clustering on the same M biopsy.* Left: *FTIR absorbance map overlaid with the pathologist's malignant pixel marks.* Centre: *MLP classification on the held-out slide, showing a contiguous region of malignant predictions (yellow) that closely follows,but in places extends beyond the histopathologist's annotation.* Right: *K-means clustering (25 clusters) of the full FTIR spectrum, with cluster labels plotted as a colourbar and the same histopathologist marks overlaid. Clusters 18–24 (orange–brown) capture the same morphology of tissue, including portions unmarked by the pathologist.*

Figure 13 shows three views of the same malignant biopsy region to illustrate how the MLP can highlight potentially malignant areas that were possibly overlooked during manual annotation. On the left, the histopathologist's red dots denote the "transforming" (malignant) pixels; yet this hand drawn mask is subject to the inter and intra-observer variability. In the centre, the MLP's pixelwise predictions (yellow) closely follow the histopathologist's annotation but they also extend along the tissue ridge where no red marks appear. The network may be using the IR information to identify malignant regions not apparent in the H&E stained optical image available to the pathologist (see Figure 18 in Appendix D). Interestingly, the overlooked regions were also identified as potentially malignant using an unsupervised learning technique. K-means clustering (Section 2.4.2) was applied to eleven wavenumbers chosen across the functional and fingerprint regions with 25 clusters being sought. As can be seen in the right panel in Figure 14, cluster 19 forms a band which overlaps with both the histopathologist's selected region and the MLP's extended prediction. Together, these novel results might imply that the FTIR-ML pipeline can generalise beyond imperfect ground-truth masks and may reveal cancerous regions that a single histopathologist misses. This is certainly motivation for further investigation and validation in future work.

4 Conclusion and Outlook

The results shown here demonstrate that tissue samples on standard glass slides rather than CaF_2 allow the application of FTIR and machine learning methods in the identification of malignant lesions in oral dysplasia. This will ease clinical adoption of FTIR-based techniques, as it requires less change to existing histopathology workflows. For patients, this could translate into fewer painful repeat biopsies for lesions that will not progress to become cancerous, and earlier intervention for those that will.

In this project it has been demonstrated for the first time that, even when restricting analysis to the functional band of the infrared spectrum, an MLP can distinguish malignant from benign tissue with high specificity (TNR$\sim$0.99 $\pm$ 0.0060) and sensitivity of TPR$\sim$0.77 $\pm$ 0.050, a performance greatly improved than currently seen in the clinic from histopathologist observation alone. By systematically comparing normalisation schemes (Table 1), global product scaling was identified as the optimal preprocessing step, and it was confirmed that z-scaling (batch normalisation) is essential to avoid heavily degraded performance. When comparing the fingerprint band with the functional band on the same optimised network architecture (Table 2), it is seen that that the fingerprint inputs yield only a modest improvement in sensitivity (TPR$\sim$0.89 $\pm$ 0.031) over the functional inputs, while both achieve near-perfect specificity. This suggests that the functional band alone retains sufficient biochemical contrast to detect malignant transformation, thereby validating the use of standard glass substrates for FTIR histopathology. It is important to note that, although the architecture is described as optimised in this work, ideally a much more extensive study of the ideal parameters would be carried out, and no network is ever completely optimised. By making the step to validating glass slide FTIR, the need is removed for specialised CaF_2 substrates which dramatically lowers barriers to clinical roll-out. This makes it far more attractive to diagnostic users, thereby shortening the path from the laboratory to bedside and sparing patients needless repeat biopsies or delayed detection.

To investigate how well a network trained on one biopsy performs on another (and thus investigate the cross-sample generalisation ability of the network), a 17x17 loss matrix is constructed (Figure 11). Each entry, m_{ij}, shows the test loss when the MLP is trained solely on biopsy i and evaluated solely on biopsy j. Low off-diagonal values identify biopsy pairs whose spectral signatures are most similar, while high values reveal pairs that differ significantly. Figure 11 reveals a clear trade off between same biopsy accuracy (the lowest losses on the matrix diagonal) and cross biopsy generalisation (off-diagonal entries). Many different biopsy pairs produce low test losses, indicating potential underlying features that merit further study.

An important aspect of the issue with training models on this data has been highlighted in this report. The training data, which should be as clean and free of error as possible, is subject to the potential sources of error made by the histopathologist. The current "Gold Standard" for histopathology (see Figure 20 in Appendix D) is, unfortunately, very far from perfect and so by construction is the training data used for this model. Inter-observer and intra-observer agreement in the context of reporting on grading the lesions concerned with oral cancer has been reputedly considered unreliable.

As discussed there are areas of improvement and additional robustness for the work carried out in this project. In a broader sense, the field of histopathology is constantly evolving, especially in the adoptation of ML methods with the intent of improving the entire process of current histopathology practices. A recent review of the current landscape with ML is given in [28]. An identified area of uncertainty when working with the model in this work, and touched upon in the aforementioned review, is the origin of trends seen in Figure 11 and 12 across biopsies. The similarities identified may either be due to the way the slide is prepared (the process of which is outlined in Appendix D) or could signify trends at the cellular level. Methods of slide preparation presents an issue in itself in the quest to make the model generalisable to any biopsy, due to the variability of the numerous stages involved which result in a different H&E stained image (ultimately the training data). To further investigate whether these trends are due to slide preparation or cellular level differences (which would provide strong motivation for further work indicating potential signatures of malignant cancer) it would be sensible to attempt to remove the effect of the staining process. A recent study [29] quantified the effect of "stain colour augmentation" and "stain colour normalisation" as a method of mitigation for this variability, showing promising effect on the ability of a Convolutional NN to generalise on an unseen stain variation. In future work, similar techniques could be adopted to reduce the effect of staining on the ML pipeline of this work.

Important ethical concerns beyond the data quality and model performance should not go unaccounted for in work such as this project, where the outcomes of the ML could potentially change lives. There is the ethical risk of amplifying existing biases in the "ground truth" annotations provided by histopathologists. Since the MLP is trained on pixel-level labels derived from a single pathologist's review, any systematic tendencies or idiosyncrasies in their grading can be learned and perpetuated by the model. In an ideal workflow, multiple independent pathologists would annotate each slide mitigating the risk of individual bias. More broadly, patient privacy and autonomy must be safeguarded when handling sensitive clinical data. As highlighted in [30], the integration of AI into oncology must result in clear guidelines to protect patient confidentiality and to ensure transparency in how decisions are made. Models should be audited regularly for demographic or feature space biases, and clinicians must retain the final decision making authority to respect patient's autonomy. Acknowledging and actively mitigating these ethical challenges means we can move confidently towards AI tools that truly improve patient outcome without undermining trust or fairness in care.

Ultimately, this could shift oral cancer diagnosis with machine learning analysis from an academic exercise into a cost-effective, minimally invasive clinical reality which reduces patient anxiety, pain, and healthcare stress while ensuring that the genuine high-risk cases are caught early.

References

[1] Cancer Research UK. Cancer incidence statistics. Technical report, Cancer Research UK, 2018.

[2] Safaa Jedani, Cassio Lima, Caroline Smith, Philip Gunning, Richard Shaw, Steve Barrett, Asterios Triantafyllou, Janet Risk, Royston Goodacre, and Peter Weightman. An optical photothermal infrared investigation of lymph nodal metastases of oral squamous cell carcinoma. *Scientific Reports*, 14, 07 2024.

[3] MIT. Maxwell's equations and electromagnetic waves. Available at https://web.mit.edu/8.02t/www/802TEAL3D/visualizations/coursenotes/modules/guide13.pdf (26/10/2024).

[4] Barnaby Ellis. *Infrared Spectroscopic Techniques Predictive Modelling Applied to Oral Cancer Diagnostics*. PhD thesis, University of Liverpool, 2022.

[5] Barbara H. Stuart. *Infrared Spectrocopy: Fundamentals and Applications*. John Wiley and Sons, Ltd, 2004.

[6] Wikimedia Commons. File:ftir interferometer.png — wikimedia commons, the free media repository, 2020.

[7] Alessandro Rossini. Bridging the technological valley of death. https://alessandrorossini.org/bridging-the-technological-valley-of-death/, 2025.

[8] D L. Woernley. Infrared absorption curves for normal and neoplastic tissues and related biological substances. *Cancer Research*, 12:516–23, 07 1952.

[9] Madeha Alkelani and Ntandoyenkosi Buthelezi. Advancements in medical research: Exploring fourier transform infrared (ftir) spectroscopy for tissue, cell, and hair sample analysis. *Skin Research and Technology*, 30, 06 2024.

[10] Simona Sabbatini, C. Conti, Corrado Rubini, Vito Librando, Giorgio Tosi, and Elisabetta Giorgini. Infrared microspectroscopy of oral squamous cell carcinoma: Spectral signatures of cancer grading. *Vibrational Spectroscopy*, 68:196–203, 09 2013.

[11] Johannes Pallua, C Pezzei, Bettina Zelger, Georg Schaefer, L Bittner, Verena Huck-Pezzei, S Schoenbichler, H Hahn, Anita Kloss-Brandstätter, F Kloss, Guenther Bonn, and Christian Huck. Fourier transform infrared imaging analysis in discrimination studies of squamous cell carcinoma. *The Analyst*, 137:3965–74, 07 2012.

[12] C. Conti, Elisabetta Giorgini, Tiziana Pieramici, C. Rubini, and G. Tosi. Ft-ir microscopy imaging on oral cavity tumours, ii. *Journal of Molecular Structure*, 744-747:187–193, 06 2005.

[13] Sedir Mohammed, Lukas Budach, Moritz Feuerpfeil, Nina Ihde, Andrea Nathansen, Nele Noack, Hendrik Patzlaff, Felix Naumann, and Hazar Harmouch. The effects of data quality on machine learning performance, 2024.

[14] Yoshua Bengio an Goodfellow and Aaron Courville. *Deep Learning*. MIT Press, Cambridge MA, 2016.

[15] Prajeesh Prathap. Feed-forward and recurrent neural networks: The future of machine learning. *Computational Intelligence and Neuroscience,*, 2021.

[16] Analytics Vidhya. A comprehensive guide to k-means clustering in python. https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/, 2019.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[18] Umberto Michelucci. An introduction to autoencoders, 2022.

[19] Duch Wlodzislaw and Norbert Jankowski. Survey of neural transfer functions. *Neural Computing Surveys*, 2:163–212, 11 1999.

[20] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. Activation functions in deep learning: A comprehensive survey and benchmark, 2022.

[21] Siddharth Sharma, Simone Sharma, and Anidhya Athaiya. Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, 04:310–316, 05 2020.

[22] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[23] Utsav Raj. Dropping the knowledge bomb: Understanding dropout layers in deep learning. Medium Blog, Nov 2023.

[24] Elizabeth Ashley, Emanuele Cauda, Lauren Chubb, Donald Tuchman, and Elaine Rubinstein. Performance comparison of four portable ftir instruments for direct-on-filter measurement of respirable crystalline silica. *Annals of work exposures and health*, 64, 04 2020.

[25] O. Kujan, A. Khattab, R. J. Oliver, S. A. Roberts, N. Thakker, and P. Sloan. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. *Oral Oncology*, 43(3):224–231, March 2007. Epub 2006 Aug 22.

[26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.

[27] M. W. Ho, J. M. Risk, J. A. Woolgar, E. A. Field, J. K. Field, J. C. Steele, B. P. Rajlawat, A. Triantafyllo, S. N. Rogers, D. Lowe, and R. J. Shaw. The clinical determinants of malignant transformation in oral epithelial dysplasia. *Oral Oncology*, 48:969–976, 2012.

[28] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, 16, 09 2017.

[29] David Tellez, Geert Litjens, Peter Bandi, Wouter Bulten, John-Melle Bokhorst, Francesco Ciompi, and Jeroen van der Laak. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology, 02 2019.

[30] Bahareh Far. Artificial intelligence ethics in precision oncology: balancing advancements in technology with patient privacy and autonomy. *Exploration of Targeted Anti-tumor Therapy*, 4:685–689, 08 2023.

[31] Katie Hanna, Anna-Lena Asiedu, Thomas Theurer, David Muirhead, Valerie Speirs, Yara Oweis, and Rasha Abu-Eid. Advances in raman spectroscopy for characterising oral cancer and oral potentially malignant disorders. *Expert Reviews in Molecular Medicine*, 26:e25, 2024.

A Simulated Interferometer

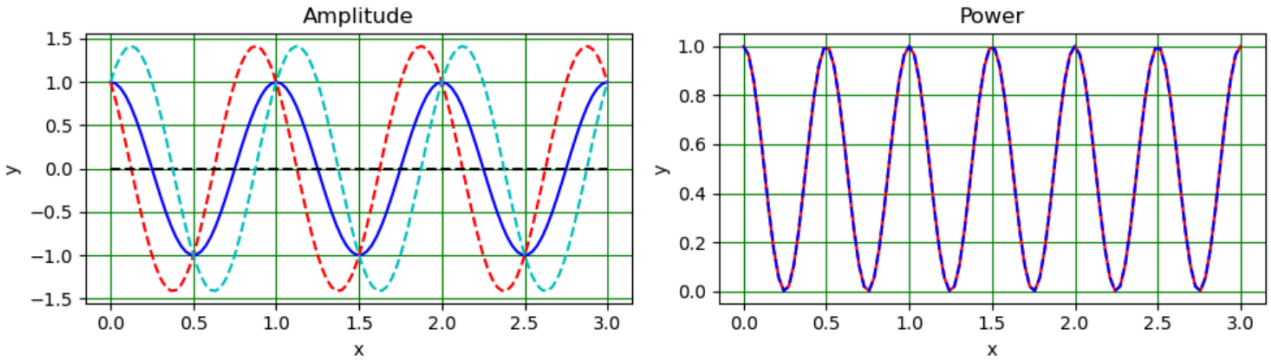


Figure 14: *Ideal interferogram at one wavenumber σ: (left) field amplitude $I(\Delta) = \frac{1}{2}I_0[1 + \cos(2\pi\sigma\Delta)]$, (right) corresponding detector power $P(\Delta) \propto [1 + \cos(2\pi\sigma\Delta)]/2$.*

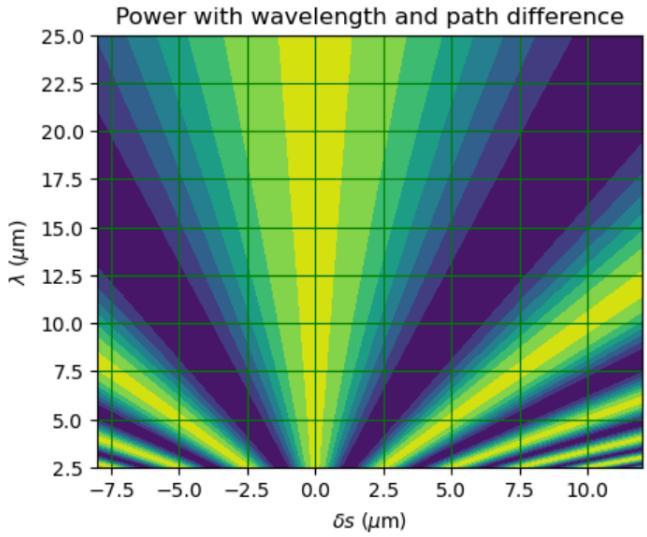


Figure 15: *Contour plot of interferometer power $P(\delta s, \lambda) \propto \big[1 + \cos(2\pi\,\delta s/\lambda)\big]/2$ across the mid-IR band. Each "V"-shaped arm corresponds to a single wavelength's fringe envelope; the central vertical ridge at $\delta s = 0$ denoting zero path difference.*

B Data Preparation and Error Analysis

The spectral data is prepared for the MLP as follows:

1. **Load raw spectra and metadata** For each biopsy (of which there are 16 consisting of 8 confirmed M and 8 confirmed B) we read in the 3D FTIR cube and the accompanying wavenumber array.

2. **Define a tissue mask** Reduce each pixel's full spectrum to its mean absorbance,

$$\overline{A}(r,c) \;=\; \frac{1}{N_k}\sum_{k} \mathrm{FTIR}[r,c,k],$$

   then threshold at a chosen quantile q to obtain $\texttt{in\_tissue}[r,c]$. This removes the background to "focus" the model.

3. **Label M vs. B pixels:** A Boolean mask is constructed by marking every pixel whose coordinates appear in the ground truth tables for M or B regions:

$$B_{\mathrm{mask}} = \mathrm{np.zeros}((n_{\mathrm{rows}}, n_{\mathrm{cols}}), \mathrm{bool}),$$

$$M_{\mathrm{mask}}[\mathrm{row_M}, \mathrm{col_M}] = \mathrm{True} \quad B_{\mathrm{mask}}[\mathrm{row_B}, \mathrm{col_B}] = \mathrm{True}.$$

4. **Select spectral features:** Rather than using all wavenumbers, a subset is chosen (tailored to the regions of interest discussed in Section 1.4.1) by finding their indices in the raw data array. This yields a vector $\texttt{key\_inds}$ of length n_{keys}.

5. **Compute final features with chosen normalisation method:** For example the "local difference" transform at each pixel;

$$X_{r,c,i} \;=\; \mathrm{FTIR}[r,c,\texttt{key\_inds}[\mathrm{i}+1]] \;-\; \mathrm{FTIR}[r,c,\texttt{key\_inds}[\mathrm{i}]]$$

$$(i = 0, \ldots, n_{\mathrm{keys}} - 2).$$

   Here;

$$y_D[r,c] = \mathrm{b\_type} \times \texttt{NT\_mask}[r,c] \in \{0,1,2\}$$

   This multiplication therefore assigns 0 to unmarked (background) pixels and "b-type" (1 for benign or 2 for malignant) to marked pixels, yielding labels in 0, 1, 2.

6. **Flatten to sample matrix:** Only the $\texttt{in\_tissue}$ pixels are extracted, stacking their n_{keys}-dimensional vectors into $X = \{X_D[r,c,:] \mid \texttt{in\_tissue}[r,c]\}$ and similarly flatten y:

$$\mathrm{rows,cols} = \mathrm{np.where(intissue)}$$

$$X_{\mathrm{flat}} = X_D[\mathrm{rows}, \mathrm{cols}, :] \quad y_{\mathrm{flat}} = y_D[\mathrm{rows}, \mathrm{cols}]$$

Statistical Error Analysis calculations: Let x_i denote the measured rate (TPR or TNR) for biopsy i, $i = 1, \ldots, n$ with $n = 17$ biopsies. The sample mean and the sample standard deviation;

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

The standard error of the mean (SEM) is

$$\mathrm{SEM} = \frac{s}{\sqrt{n}},$$

which quantifies the precision with which $\bar{x}$ estimates the true underlying rate. To form a two tailed 95 % confidence interval (CI), the Student's t–distribution critical value is used $t_{0.025,\,n-1}$ (here $t \approx 2.12$ for $n - 1 = 16$), giving a margin of error and CI;

$$\mathrm{MOE}_{95\%} = t_{0.025,\,n-1}\,\mathrm{SEM}, \qquad \mathrm{CI}_{95\%}: \quad \bar{x} \pm \mathrm{MOE}_{95\%}.$$

Confusion matrix to visualise model evaluation:

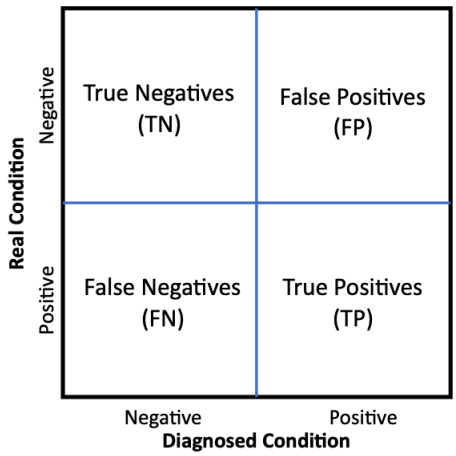
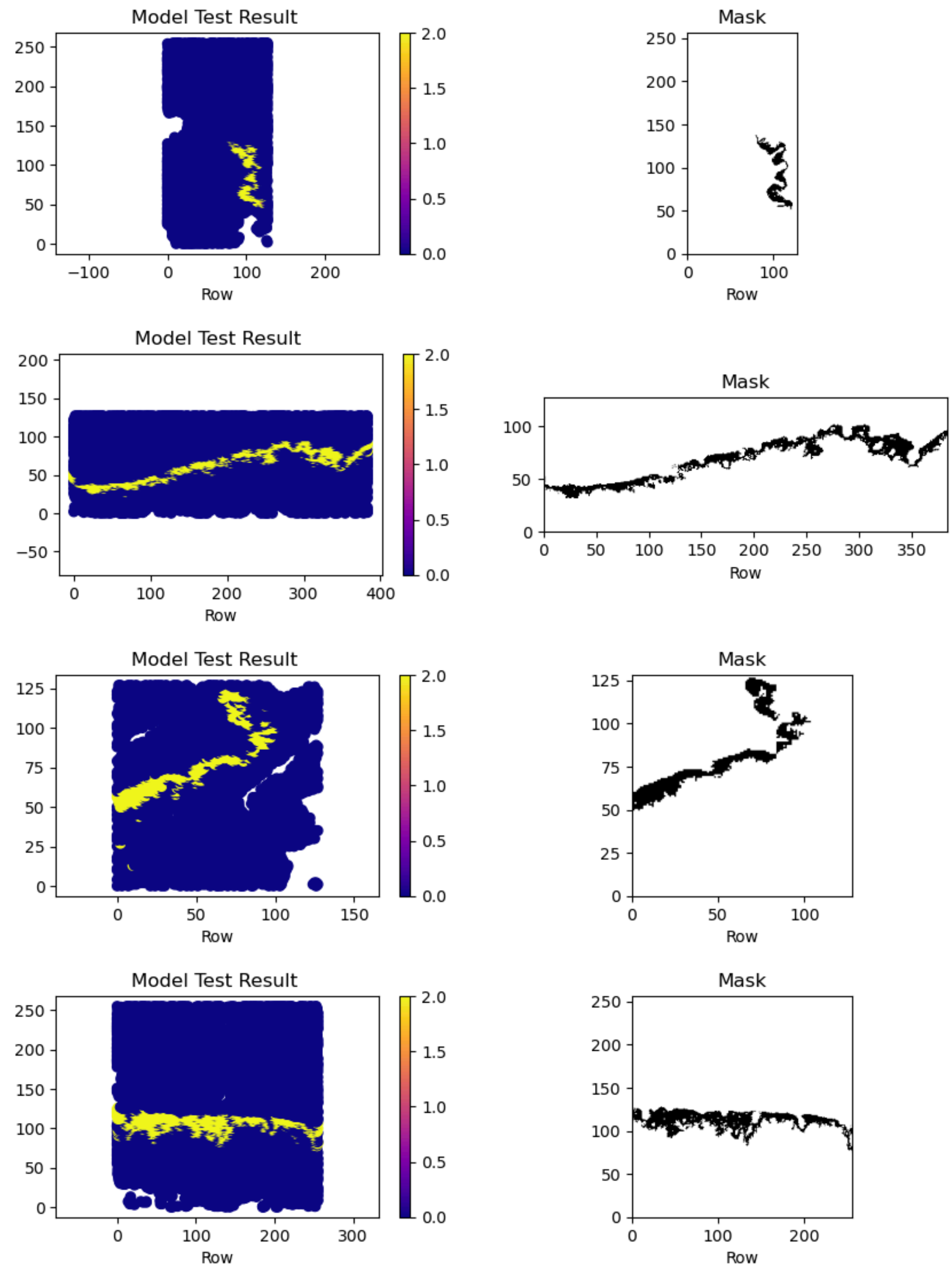
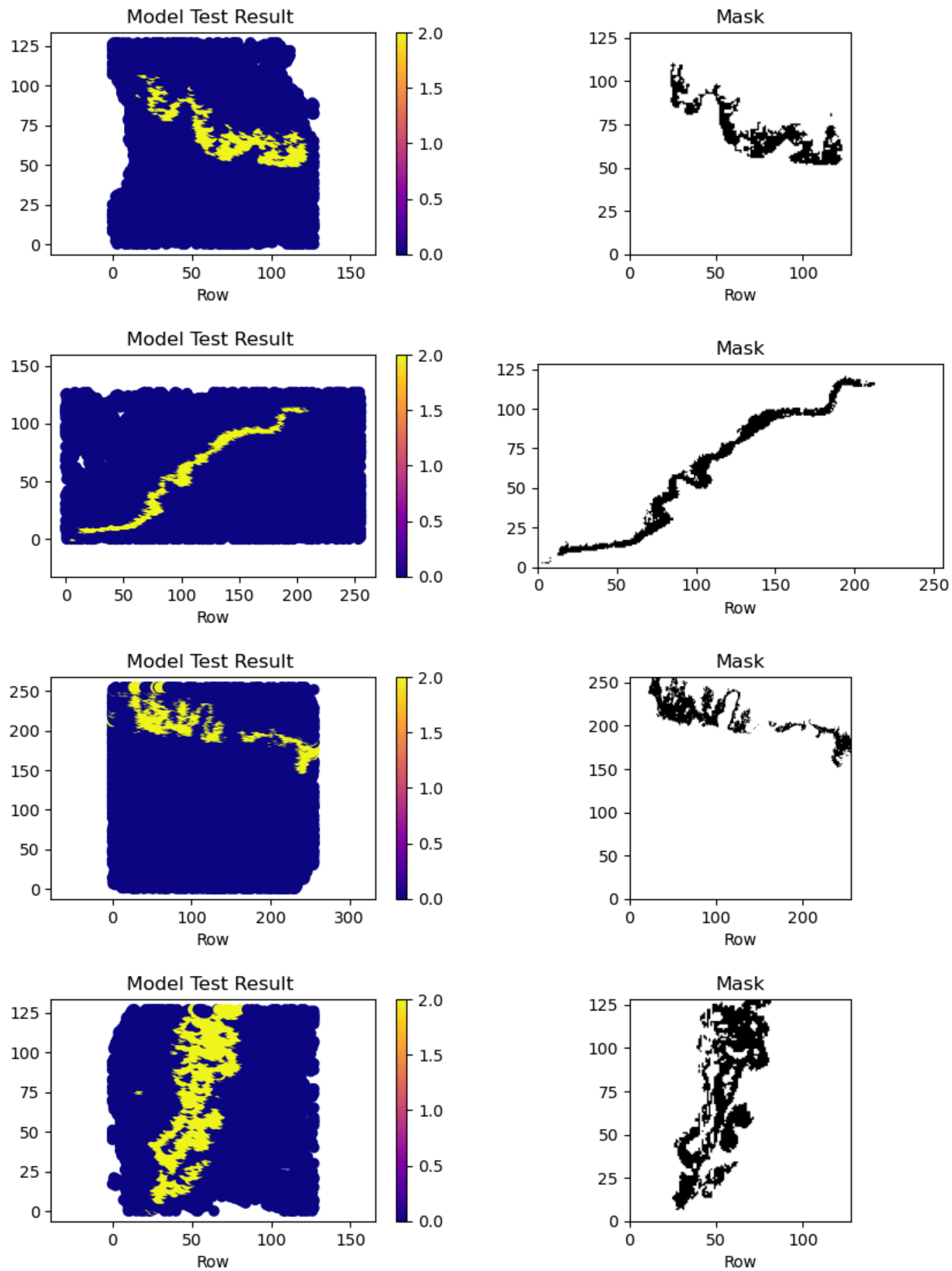


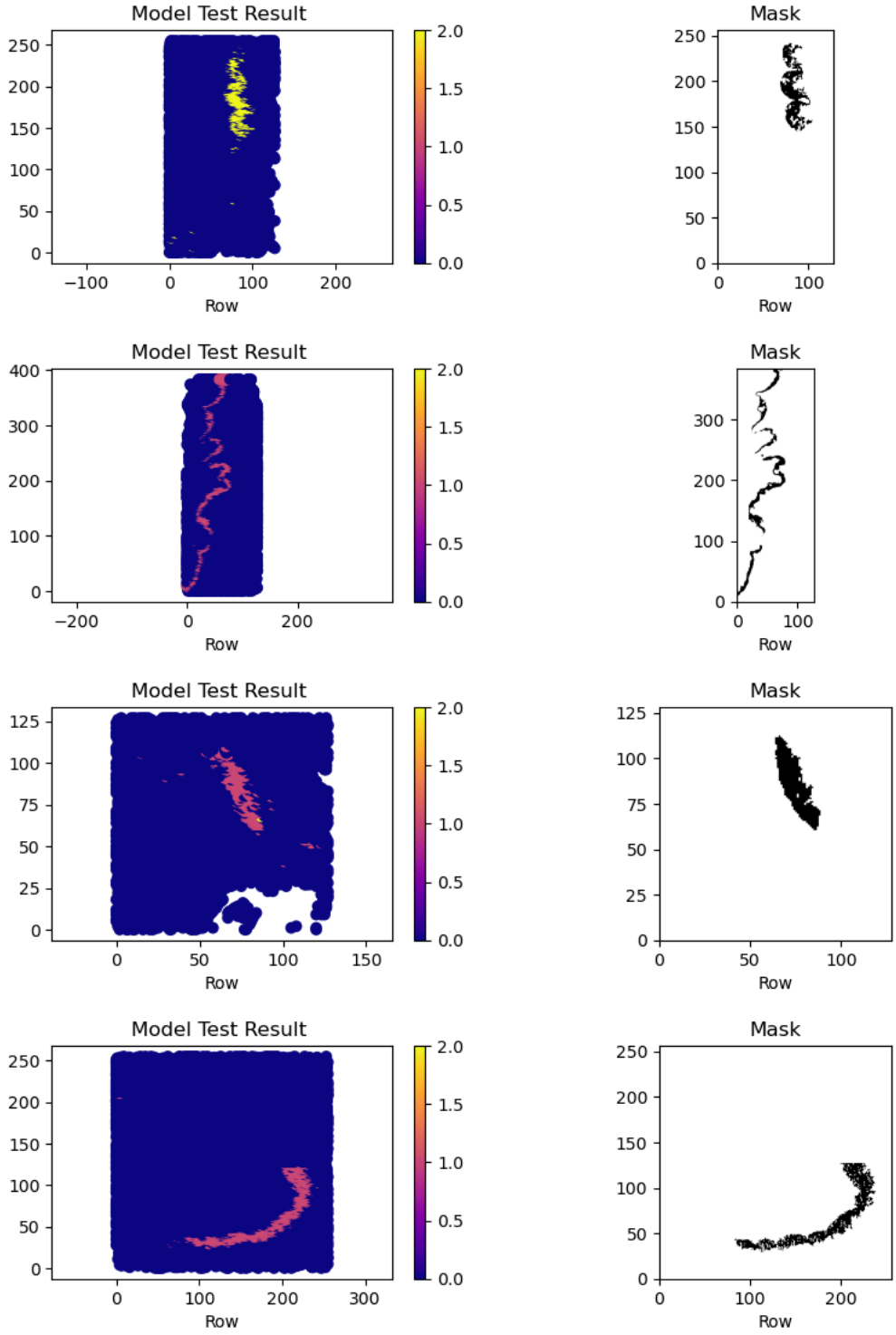
Figure 16: *Confusion matrix.*

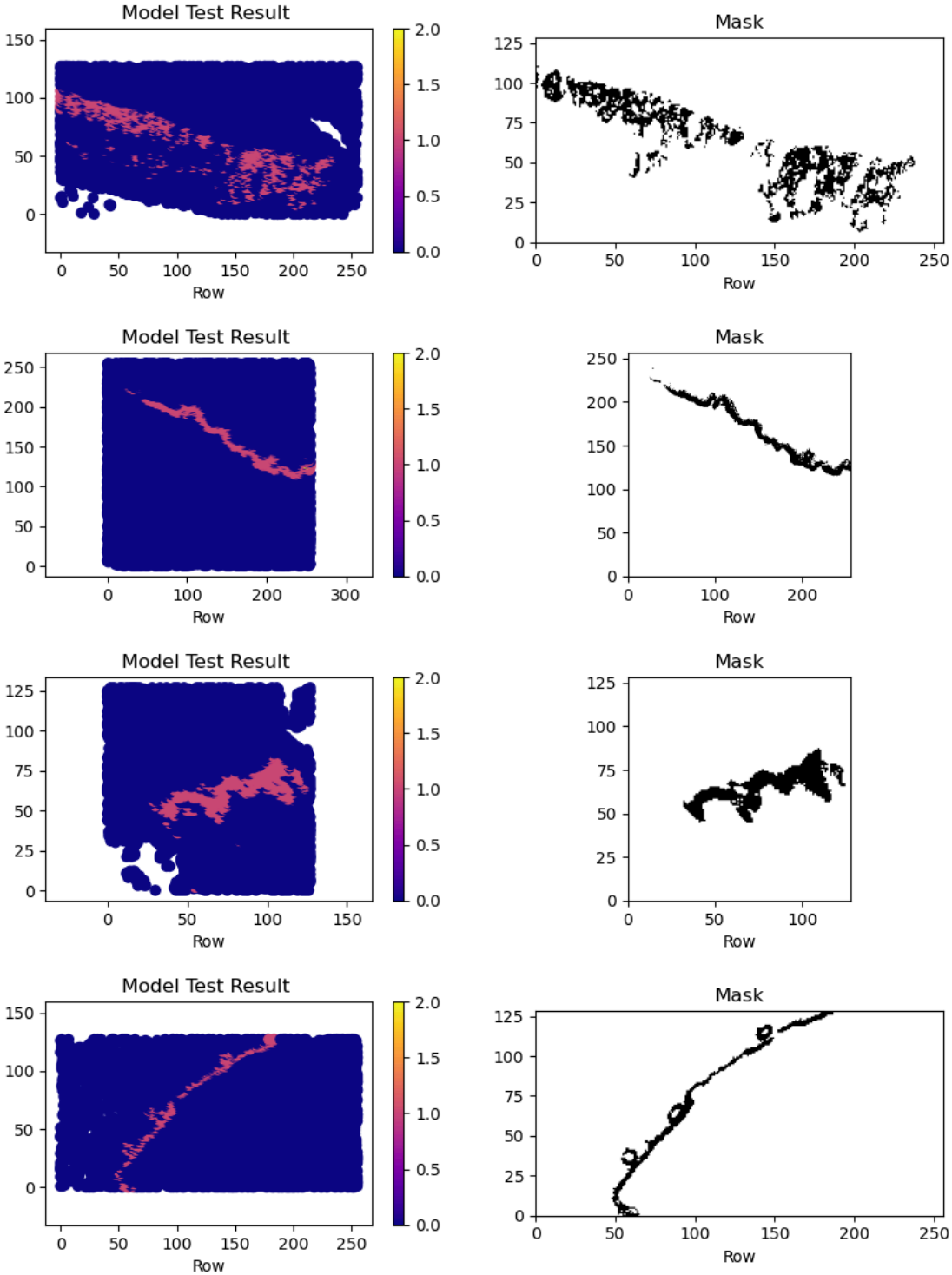
C Full MLP Test Results

Results of all 17 biopsies from the MLP model in the fingerprint and functional band:

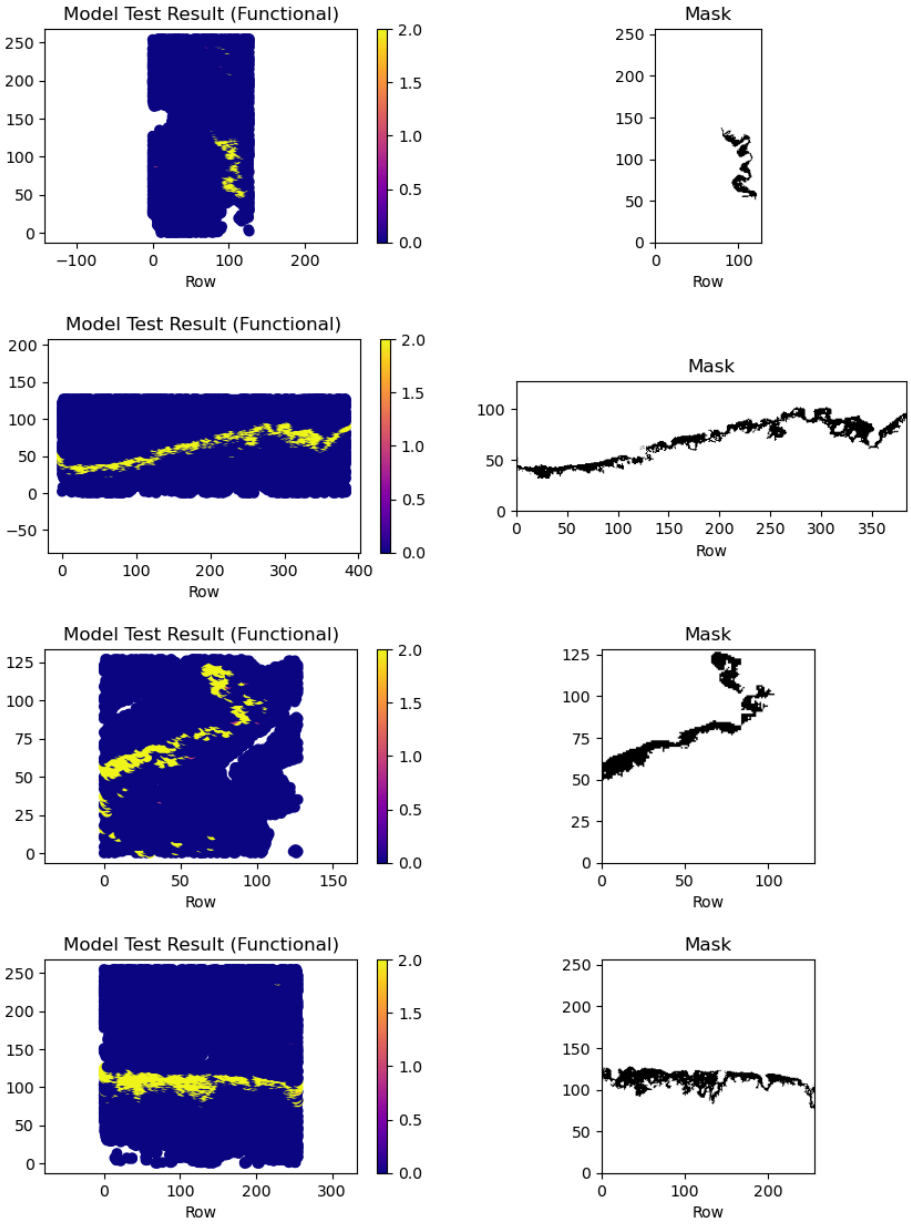

Model Test Result
Mask
Row
Model Test Result
Mask
Row
Model Test Result
Mask
Row
Model Test Result
Mask
Row

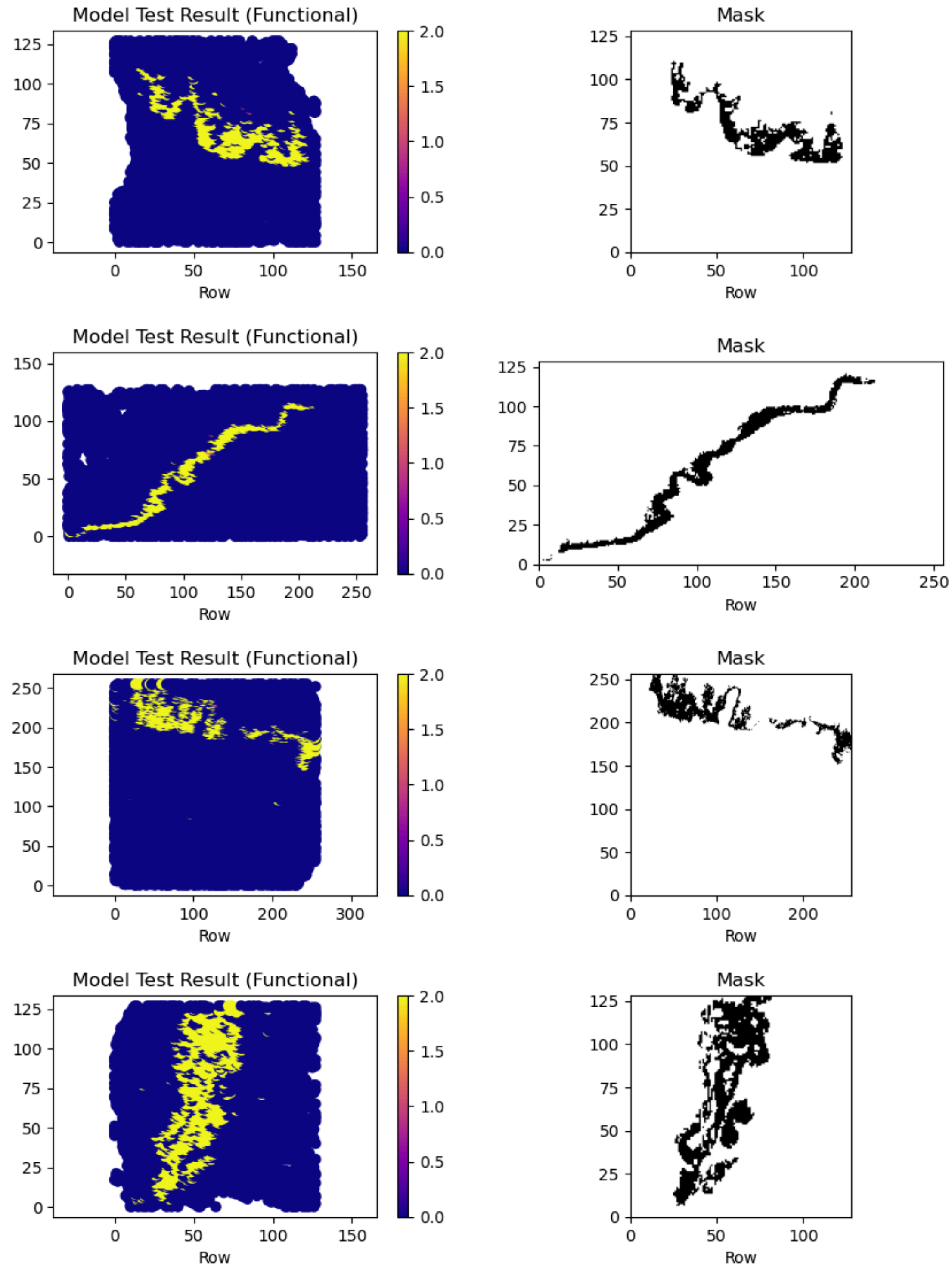

Model Test Result
Mask
Row
Model Test Result
Mask
Row
Model Test Result
Mask
Row
Model Test Result
Mask
Row
2.0
1.5
1.0
0.5
0.0

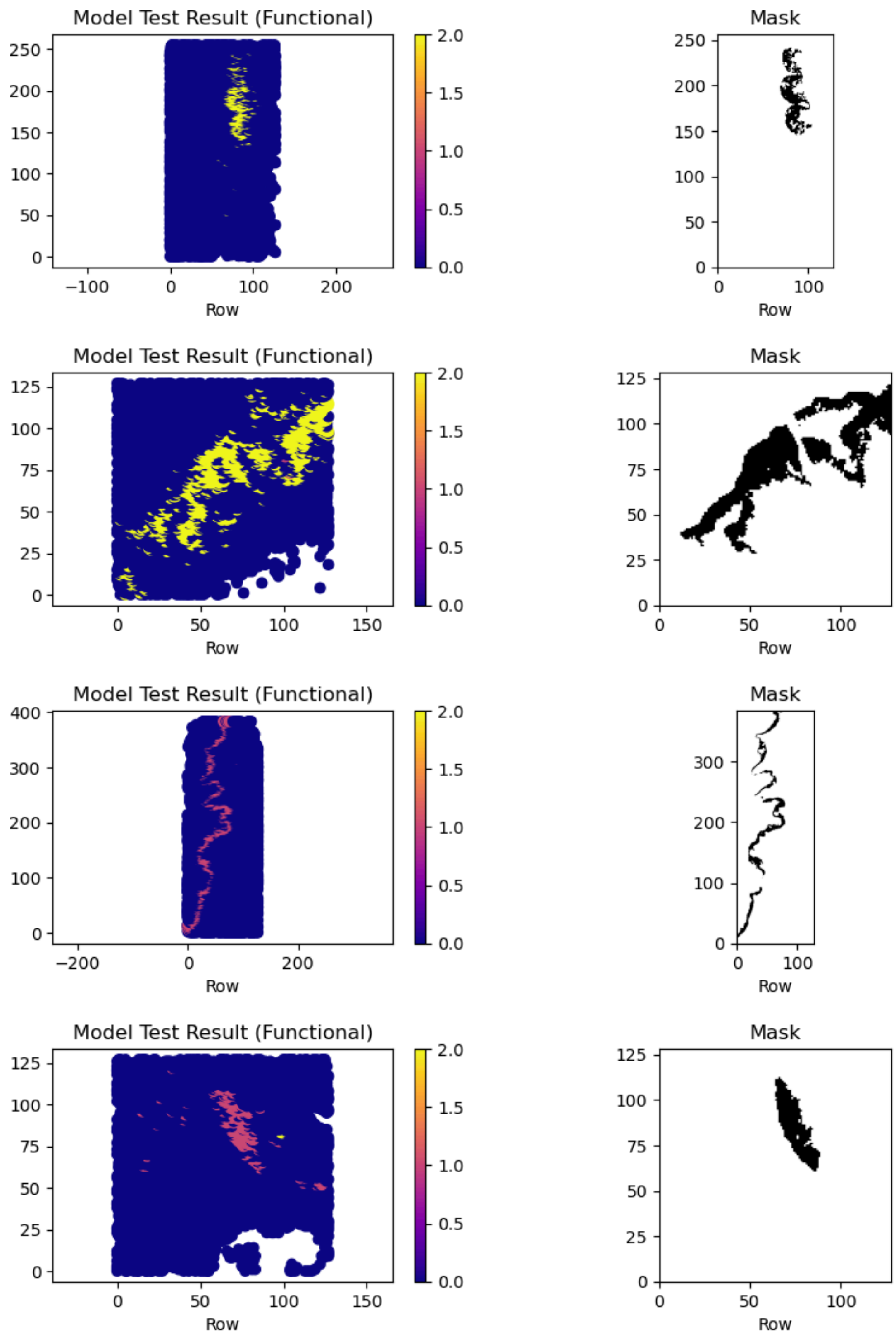

Model Test Result
Row
Mask
Row
Model Test Result
Row
Mask
Row
Model Test Result
Row
Mask
Row
Model Test Result
Row
Mask
Row

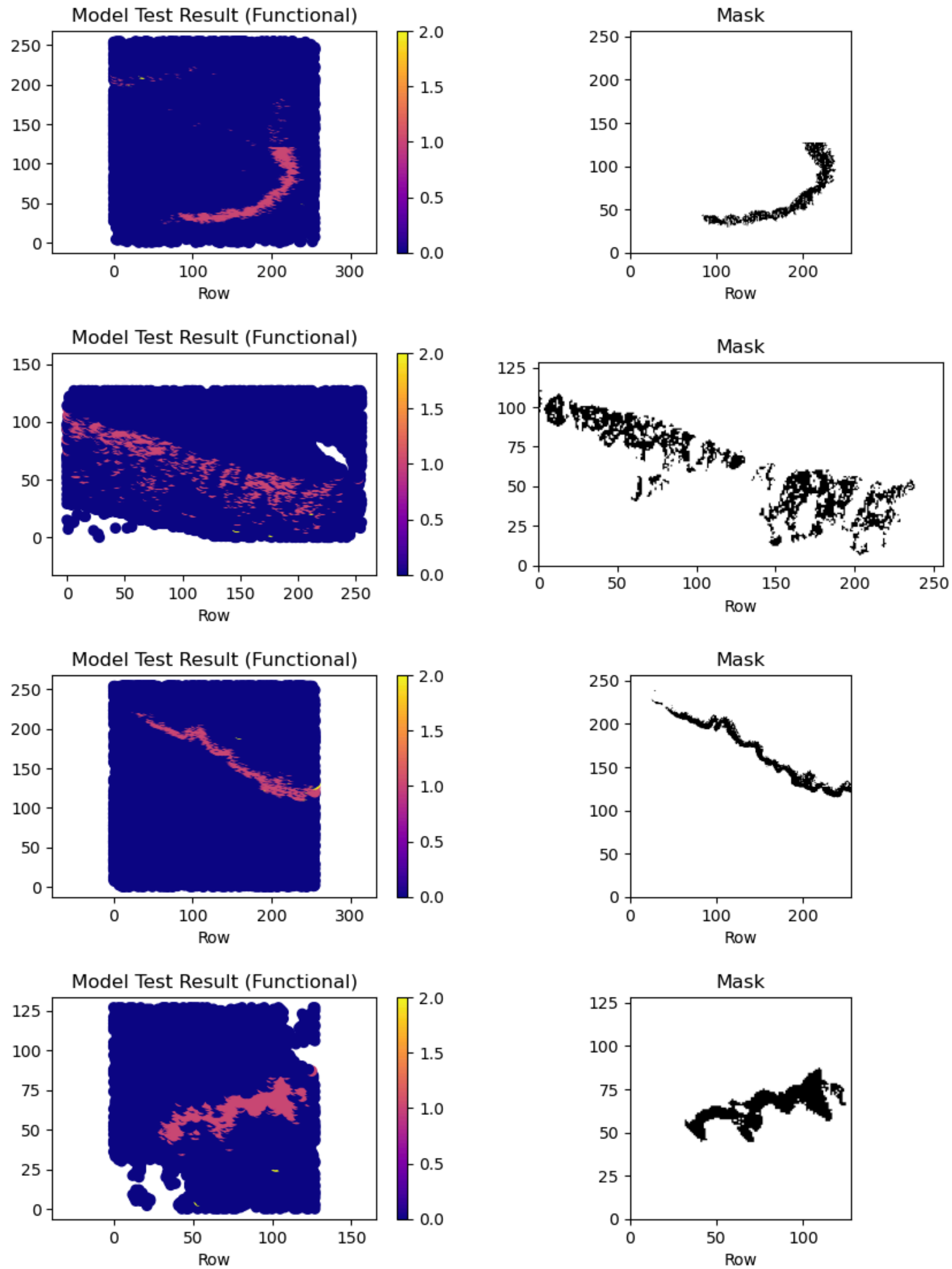

Model Test Result
Row
Mask
Row
Model Test Result
Row
Mask
Row
Model Test Result
Row
Mask
Row
Model Test Result
Row
Mask
Row

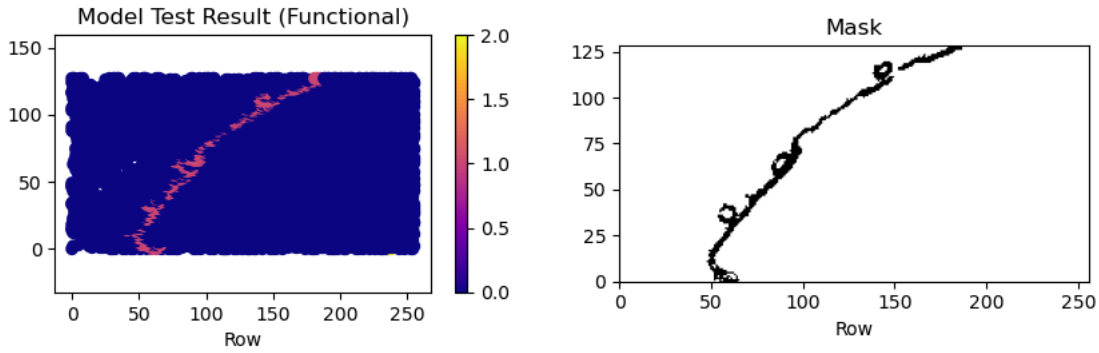
Full model results for the MLP in the functional region:




Model Test Result (Functional)
Mask
Row
Model Test Result (Functional)
Mask
Row
Model Test Result (Functional)
Mask
Row
Model Test Result (Functional)
Mask
Row


Model Test Result (Functional)
250
200
150
100
50
0
−100
0
100
200
Row
2.0
1.5
1.0
0.5
0.0
Mask
250
200
150
100
50
0
0
100
Row
Model Test Result (Functional)
125
100
75
50
25
0
0
50
100
150
Row
2.0
1.5
1.0
0.5
0.0
Mask
125
100
75
50
25
0
0
50
100
Row
Model Test Result (Functional)
400
300
200
100
0
−200
0
200
Row
2.0
1.5
1.0
0.5
0.0
Mask
300
200
100
0
0
100
Row
Model Test Result (Functional)
125
100
75
50
25
0
0
50
100
150
Row
2.0
1.5
1.0
0.5
0.0
Mask
125
100
75
50
25
0
0
50
100
Row


Model Test Result (Functional)
Mask
Row
Model Test Result (Functional)
Mask
Row
Model Test Result (Functional)
Mask
Row
Model Test Result (Functional)
Mask
Row



Dropout Layer Results:

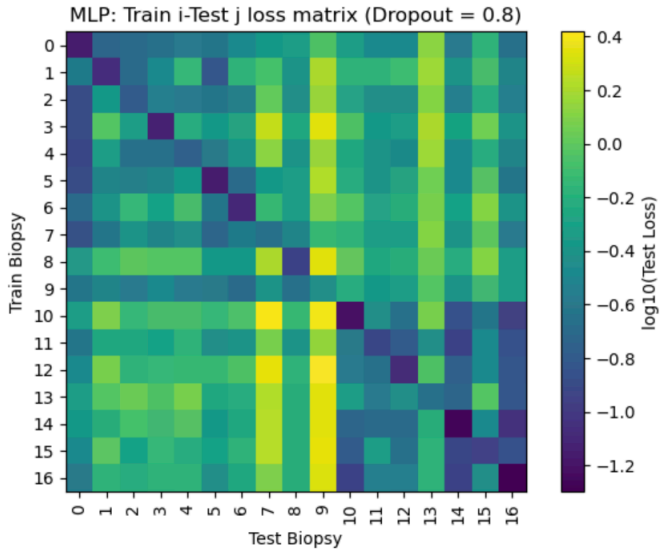


Figure 17: *Train i-Test j results with a dropout layer of 0.8 applied.*

D Supplementary Context and Code

Link to code: https://github.com/livshu/MPhys-Project-Code

Description of the project: This project explores a way to help clinical professionals treat oral cancers more accurately, and hopefully with fewer painful repeat biopsies, by using imaging techniques and artificial intelligence. Currently, tissue samples are examined under a microscope by a histopathologist (someone who studies changes in tissues caused by disease), but studies show they can only predict which early lesions will turn cancerous (since they with either remain benign or become cancerous) about 25–40% of the time. When imaged using infrared techniques, very small differences in the spectra can reveal the chemistry inside each cell. Many of these spectra are fed into a simple neural network "brain" (called a multi-layer perceptron) that learns to tell benign from malignant tissue. Even without access to the region of the spectrum which calcium fluoride allows us to see (calcium fluoride is not currently used in clinical practice but it allows us to see a region of the spectrum called the fingerprint band, which provides the artificial intelligence with much more information), the method on glass slides already used in the clinic (where we can no longer access the information-rich fingerprint region) achieves ∼99% accuracy at spotting healthy tissue and 77% accuracy at identifying cancerous changes. In tests on biopsies unseen to the algorithm, it can also highlight potentially cancerous regions the pathologist missed. These results suggest that combining infrared imaging techniques with machine learning could give patients clearer answers faster and significantly improve their outcomes and experience.

Contextual Figures:

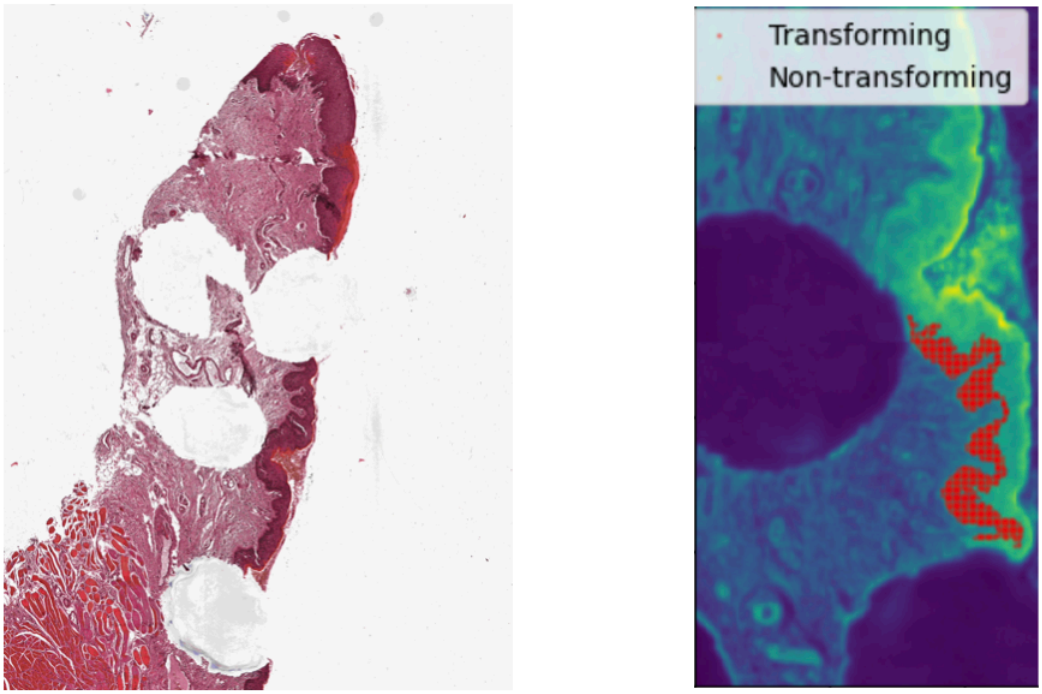


Figure 18: *Left: Histopathology image of an excised oral biopsy, stained with haematoxylin & eosin (H&E), showing the epithelial structures. Right: FTIR absorbance map for the same region, overlaid with the pathologist's annotations.*

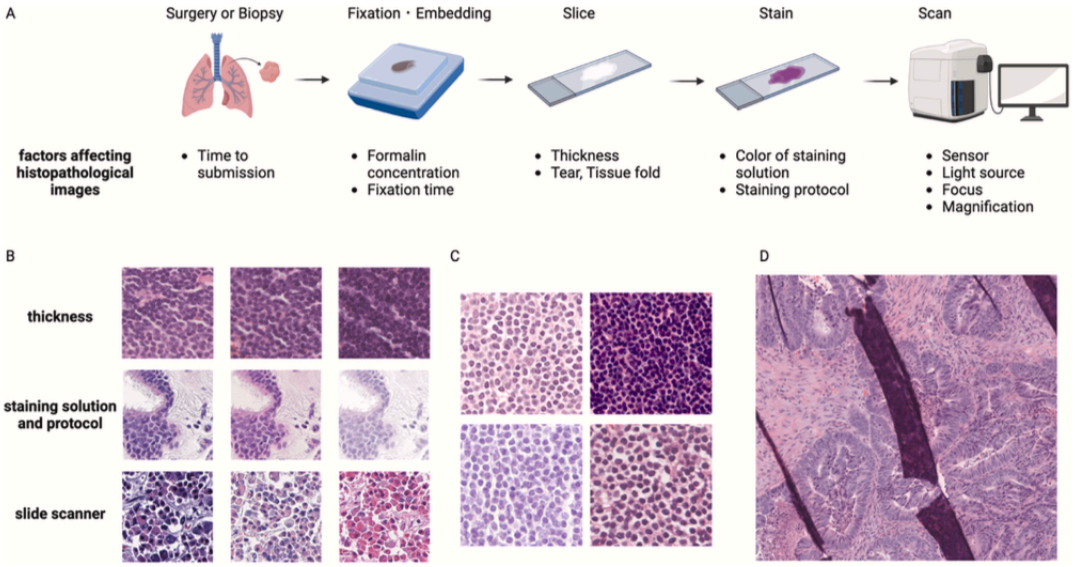


Figure 19: *Variation in pathology images and artifacts. A) Factors which affect the quality of the image. B) Examples of images produced by factors such as tissue thickness, stain solution and protocol and slide scanners. C) Examples of the same structures from different institutions. D) A tissue fold artifact. Diagram taken from [28]*

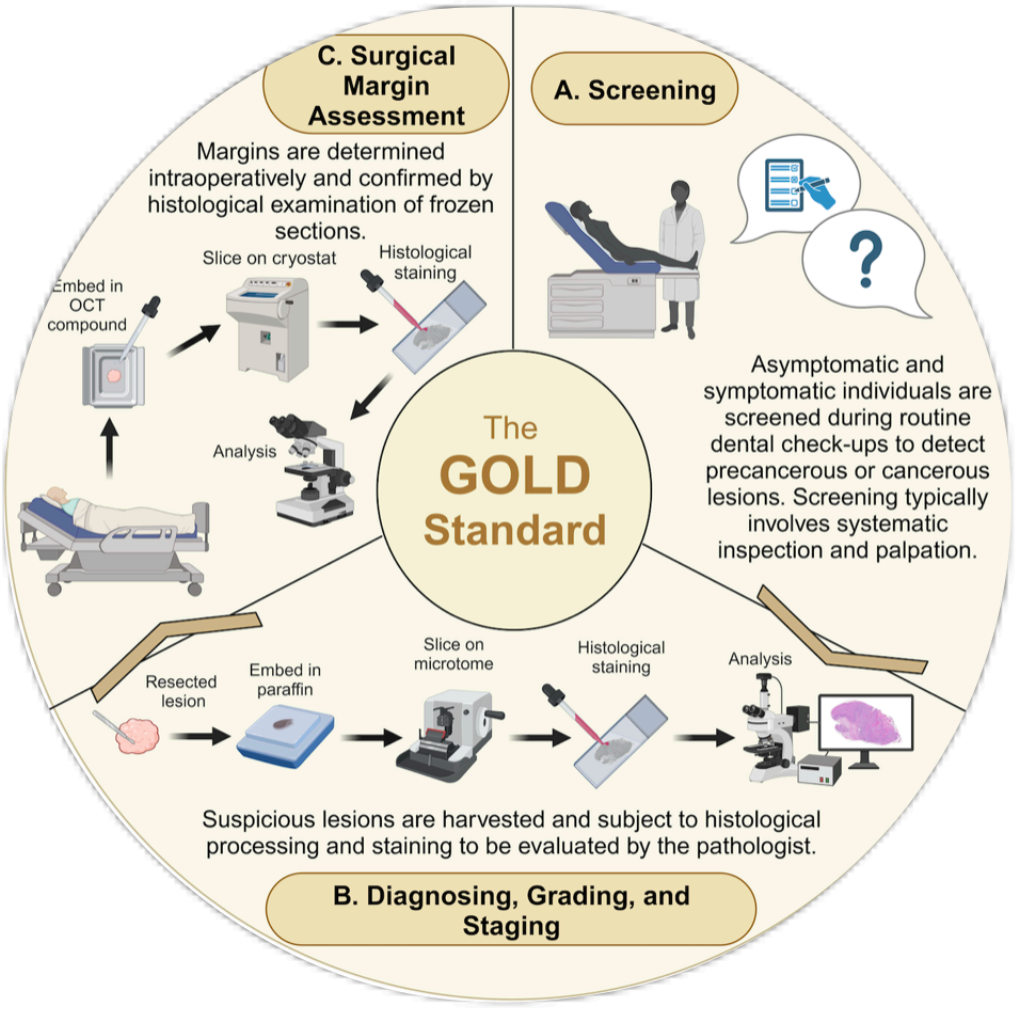


Figure 20: *(A) The current standard for screening is via conventional visual oral examination under a bright light to detect abnormal oral findings and involves systematic inspection and palpation of the oral cavity and regional lymph nodes. (B) To achieve a definitive diagnosis, a tissue biopsy followed by histological assessment is the current gold-standard for oral cancer and OPMD diagnosis. This is invasive and prone to interpretative disparity amongst pathologists. (C) When oral cancer is diagnosed, surgery is the main treatment. Currently, this is labour intensive, time-consuming and subjective. Figure taken from [31].*

E Risk Assessment and Project Plan

UNIVERSITY OF LIVERPOOL

RISK ASSESSMENT FORM

School/Department: Physical Sciences /Dept of Physics	**Building: Oliver Lodge/Home Room 225 (Space Ref 02/051)**
Task: MPhys Project with student Olivia Shuter: To perform a comparative study of Cancerous tissue on Glass and CaF_2 substrates using Fourier Transform Infrared Spectroscopy	
Persons who can be adversely affected by the activity: **Caroline Smith, Paul Harrison, Paul Unsworth, James Ingham and Olivia Shuter (OS)**	

Section 1: Is there potential for one or more of the issues below to lead to injury/ill health (tick relevant boxes)

People and animals/Behaviour hazards

Allergies		Too few people		Horseplay		Repetitive action		Farm animals	
Disabilities		Too many people		Violence/aggression		Standing for long periods		Small animals	
Poor training		Non-employees		Stress		Fatigue		Physical size, strength, shape	
Poor supervision		Illness/disease		Pregnancy/expectant mothers		Awkward body postures		Potential for human error	
Lack of experience	√	Lack of insurance		Static body postures		Lack of or poor communication		Taking short cuts	
Children		Rushing		Lack of mental ability		Language difficulties		Vulnerable adult group	

What controls measures are in place or need to be introduced to address the issues identified?

Identified hazards	What controls are currently planned or in place to ensure that the hazard identified does not lead to injury or ill-health?	RISK SCORE			Is there anything more that you can do to reduce the risk score in addition to what is already planned or in place?	RESIDUAL RISK SCORE		
		L	C	R		L	C	R

Lack of experience	Training will be provided to student (OS) who will not be allowed to start work unsupervised	2	2	4		2	2	4

L = likelihood; C = consequence; R = overall risk rating

Section 2: Common Workplace hazards. Is there potential for one or more of the issues below to lead to injury/ill health (tick relevant boxes)

Fall from height		Poor lighting		Portable tools		Fire hazards		Chemicals	√	Asbestos	
Falling objects		Poor heating or ventilation		Powered/moving machinery		Vehicles		Biological agents	√	Explosives	
Slips, trips, falls		Poor space design		Lifting equipment		Radiation sources		Waste materials	√	Genetic modification work	
Manual handling		Poor welfare facilities		Pressure vessels		Lasers		Nanotechnology		Magnetic devices	
Display screen equipment	√	Electrical equipment	√	Noise or vibration		Confined spaces		Gases		Extraction systems	
Temperature extremes		Sharps	√	Drones		Cryogenics	√	Legionella	√	Robotics	
Home working		Poor signage		Overseas work		Overnight experiments		Unusual events		Community visits	
Late/lone working	√	Lack of/poor selection of PPE		Night work		Long hours		Weather extremes		Diving	

What controls measures are in place or need to be introduced to address the issues identified?

Identified hazards	**What controls are currently planned or in place to ensure that the hazard identified does not lead to injury or ill-health?**	**RISK SCORE**			**Is there anything more that you can do to reduce the risk score in addition to what is already planned or in place?**	**RESIDUAL RISK SCORE**		
		L	**C**	**R**		**L**	**C**	**R**
Electrical equipment	There are no exposed electrical conductors, and no unauthorised people are allowed to remove equipment covers. All equipment is powered from 240V mains and is PAT tested.	2	4	8		2	4	8

Cryogenics	Causes burns in contact with skin so the use of gloves and eye protection is required.	2	2	4		2	2	4
Display Screen Equipment	Adjustable height chair is provided to ensure correct working height and adequate breaks will be taken.	1	1	1		1	1	1
Lone working	Undergraduates are not permitted to work unsupervised. Keys are required to access the lab and are not issued.	2	2	4		2	2	4
Sharps	Cut resistant gloves are available for handling broken glass before disposing in glass waste bin. A sharps waste bin is provided for the correct disposal of sharps.	2	2	4		2	2	4
Chemicals	Lab coat, safety glasses and gloves to be worn. MSDS must be consulted before using and specific COSHH assessments must be completed when necessary.	3	2	6		3	2	6
Biological Agents	Tissue is fixed prior to bringing into the department. Wash hands after handling. Wear appropriate PPE to handle human material. Only members of staff that have completed the HTA training are allowed to handle specimens.	2	2	4		2	2	4
Waste Materials	The appropriate waste streams are available.	2	2	4		2	2	4

Legionella	Taps are flushed weekly as per University guidelines.	3	3	9		3	3	9

L = likelihood; C = consequence; R = overall risk rating

Section 3: Additional hazards: are there further hazards NOT IDENTIFIED ABOVE that need to be considered and what controls are in place or needed? (list below)

Identified hazards	What controls are currently planned or in place to ensure that the hazard identified does not lead to injury or ill-health?	RISK SCORE			Is there anything more that you can do to reduce the risk score in addition to what is already planned or in place?	RESIDUAL RISK SCORE		
		L	C	R		L	C	R

Section 4: Emergency arrangements (List any additional controls that are required to deal with the potential emergency situation)

Emergency situation	Additional control required

Risk assessor (signature) Caroline Smith ...Date..17/10/2024.....Authorised by (signature) ...Date..17/10/2024.....

Gantt Chart for project timeline:

PHYS-498 MPhys Project	start	end	11%
Project Initiation	**26/09/24**	**18/10/24**	**100%**
Apply for projects	26/09	02/10	100%
Assigned to project	03/10	03/10	100%
Begin project	07/10	11/10	100%
Prepare project proposal	10/10	18/10	100%
Discuss ethics and risk with supervisors	10/10	18/10	100%
Submission of proposal	18/10	18/10	100%
Data	**15/10/24**	**01/04/25**	**3%**
Collect data	15/10	21/11	10%
Review preliminary data	22/11	31/12	0%
Review project	03/02	06/02	0%
Analyse data	03/02	01/04	0%
Presentation	**21/10/24**	**24/11/24**	**0%**
Prepare presentation	21/10	09/11	0%
Develop skills	02/11	20/11	0%
Review and practice	10/11	24/11	0%
Complete presentation	24/11	24/11	0%
Poster (S02 dates TBC)	**17/01/25**	**22/02/25**	**0%**
Prepare poster draft	17/01	31/01	0%
Finalise and practice poster	01/02	22/02	0%
Poster presentation	22/02	22/02	0%
Report Writing	**07/10/24**	**28/04/25**	**7%**
Review literature	07/10	01/11	25%
Write introduction	16/10	15/11	10%
Write main body of report	08/02	10/03	0%
Finalise results and write up	12/03	11/04	0%
Finalise report	13/04	28/04	0%
Final report submission	28/04	28/04	0%