A Diffuse Reflection Broadband Spectroscopy Approach to Tissue Classification

David J. Morton[1]

[a]Queen's University, 99 University Ave, Kingston, Canada

ABSTRACT

Over 30% of breast conserving surgery (BCS) patients must undergo revision surgery due to cancerous tissue remaining in the cavity. We hypothesize that optical tissue characterization and classification can improve the success rate of BCS by giving the surgeon more information to work with intraoperatively. In this study, we assess the feasibility of diffuse reflection broadband spectroscopy to differentiate between Beef, Chicken, Pork, and Turkey samples. We recorded the optical reflectivity of various tissue samples in a range of 200 to 1000 nm. We then preprocessed this data and used an SVM, a KNN, and a 1D CNN to classify the resulting spectra. The KNN achieved an average F1 score of 95% and an accuracy of 95.5%. This outperformed the CNN and SVM which achieved 89% and 84% respectively. Through broadband spectroscopy, we demonstrate the feasibility of optical characterization for tissue classification. Applying this to breast conserving surgery can help inform the surgeon about the tissue composition of the resection cavity after initial tumor resection.

GITHUB LINK: https://github.com/16djm10/881FinalCode.git

Keywords: diffuse reflectance spectroscopy, tissue classification, breast cancer, margin assessment

1. INTRODUCTION

Breast cancer is the most common surgically treated cancer in women globally.[1] The standard recommendation in early-stage breast cancer is breast-conserving surgery. This process entails the complete removal of the malignant tumour, whilst minimizing the extraction of healthy tissue.[2] After the resection of the tumour there will exist a small cavity where the tumour once, the walls of which should consist entirely of healthy tissue. However, in 20-30% of cases, the patient is required to undergo revision surgery to remove cancerous tissue that was missed during surgery and left in the walls of the resection cavity.[3] Consequently, the patient has an increased risk of postoperative complications, impaired cosmesis, increased psychological distress, and increased cost. This problem is difficult to solve as cancer and healthy tissues are difficult to differentiate with the naked eye.

Thus any technological solution requires two main components. Tissue characterization, where we can use techniques to detect the differences between the tissues, and classification where the resulting differences are analyzed and determined to be healthy or cancerous. There are many ways to intraoperatively characterize tissues. However, some are more destructive than others. The iKnife[4] for example characterizes tissues by vaporizing the tissue and analyzing the chemical signature. An emerging noninvasive technique is characterizing tissue optically using low-intensity light. The development of a technique to intraoperatively and noninvasively scan the post-resection cavity for signs of malignant tissue will allow surgeons to be more confident that they've removed everything, resulting in a significantly lower chance of revision surgery.

Connolly et al.[5] offered a proof of concept to solve this problem that used a throughput broadband spectroscopy and ultrasound probe. This was performed on dual-layer phantoms consisting of beef and turkey to represent malignant and healthy tissue. Although this work is promising, the nature of throughput broadband spectroscopy is not feasible for intraoperative settings as the receiver would have to be inside the patient. Further, limited data were collected and a deep learning approach was not tested.

This work aims to further Connolly's idea and develop a freehand technique to use diffuse reflection broadband spectroscopy to differentiate between various tissue types. I hypothesize that this technique will produce strong optical features that can be reliably used to train a machine learning model to accurately differentiate between the various tissues.

To solve these problems, I propose the use of a broadband light source, diffuse reflection probe, and spectrometer to collect optical spectra data from various tissue samples. This technique looks to exploit various optical tissue properties such as reflectance, re-emittance, and electric permeability to characterize tissues and classify them using machine and deep learning techniques. I collect spectral data from 4 animal tissue samples consisting of beef, chicken, pork and turkey. Connolly et al.[5] record just one sample per tissue at just one wavelength. I have increased the spatial and frequency resolution, collecting a 6 by 6 grid of readings spatially, each containing 3600 frequency samples in the range of 200 to 1000 nm. These spectra are then processed using a variety of techniques followed by feeding them into various machine learning and deep learning architectures including KNN, SVM, and a custom-built 1D CNN.

2. METHODS

2.1 Materials and experimental setup

The experimental setup consisted of the following devices and materials:

Optical data collection:

- 6 x silicon sample holders (6 by 4cm x 4 cm)
- 1 x Thorlabs stabilized tungsten light source
- 1 x Optical Probe
- 1 x Thorlabs spectrometer
- 1 x Thorlabs OSA software
- 1 x Mechanical gripper arm

Ultrasound data collection

- Clarius trans rectal ultrasound probe
- Ultrasound dampening pad
- Ultrasound Gel
- Ultrasound transmitting pad

The sample holder consisted of a 2 block by 3 block flexible silicone ice cube tray purchased from amazon. I originally planned to 3D print a 2x2 rigid holder but was advised against doing so as the material would be quite porous and not suitable for raw tissue. The light source produced a range of 300 – 2800 nm, with the range of 500 – 1800 maintaining 30% of the peak intensity. The spectrometer measured a range of 200 - 1000 nm. The OSA software was used to visualize and collect the spectrometry data. The mechanical arm was used to hold the probe in a fixed position and angle over the samples.

An image of the setup can be seen in Figure[1]

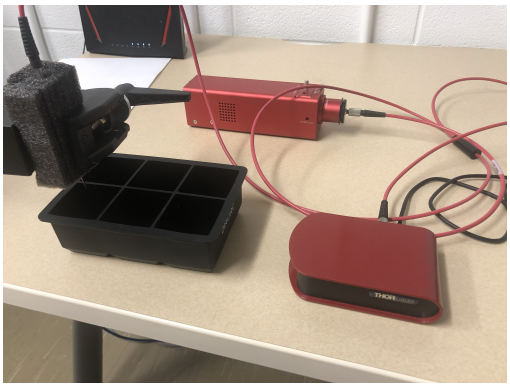

Figure 1. Image of the experimental setup

2.2 Data acquisition

The goal of the data acquisition process is optical characterization curves for the four tissues types in the range of 200 to 1000 nm. The optical signature consists of peaks at each wavelength representing the intensity of light that was reflected from the tissue back to the problem. Example data can be seen in Figure [4] .

To ensure the quality of the collected data, before the full data acquisition process I tested certain experimental parameters to see how they would affect the output. These parameters included how wet the surface was, the distance of the probe from the surface, and the ambient lighting in the room.

Surface dampness was tested to see if excess water created an extra boundary for light to reflect off of before being absorbed by the surface. Spectra were recorded when the samples were unaltered, when water was rubbed on the surface, and when the surface was patted with a paper towel. I found that excess water would cause a greater signal intensity while also distorting the shape slightly. Therefore, I removed excess moisture before sampling.

To test the effect of distance, and effectively beam width, I took spectra at 1mm, 5mm and 10mm. When the probe was at 1mm it tended to cap out the spectrometer's intensity range capability. However, this can be solved by reducing the capture interval. It was found that the distance from the sample did not change the shape of the spectra, but it did change the peak intensity level. Therefore, a distance of 5-10mm was used to keep the beam size reasonable.

Spectra were also taken with the ambient lights turned on or off. It did not appear to make a difference in the readings, but for consistency between ambient lighting sources, I turned the lights off for the experiment.

The tissue was allowed to warm to room temperature, cut into roughly 4cm x 4cm x 2cm chunks, placed in the tray and made as flat as possible. Due to the depth of the tray, foam blocks were used to raise the samples to the top of the container. The light source was allowed to warm up for 45 minutes to get up to peak intensity. The samples were placed under the optical probe, which is fixed 5-10mm above the samples and spectra were systematically taken in a 6 by 6 grid over the samples. Scans were performed in rows from left to right, top to bottom, leaving a 5mm margin to minimize interference from the walls of the container as seen in Figure [2]. Each sample was a 1-second time average to reduce signal noise. This was repeated for each tissue, resulting in 144 spectra in total consisting of approximately 3600 features each. Finally, a baseline reading was taken from a white sheet of paper to later subtract from each spectrum such that the differences could be analyzed to account for the light source intensity curve.

To capture the ultrasound data a damping pad was placed under the sample. US gel was then placed on the tissue surface. An ultrasound transmitting pad was placed on top of the sample followed by more gel. The transrectal probe was then held normal to the tissue surface. 5 readings were taken in dice configuration as seen in Figure [3]. Each reading took 5 seconds and the raw RF signal was recorded for later use. It is important to note that the US data was recorded for future steps of the project and was not included in the rest of the study.

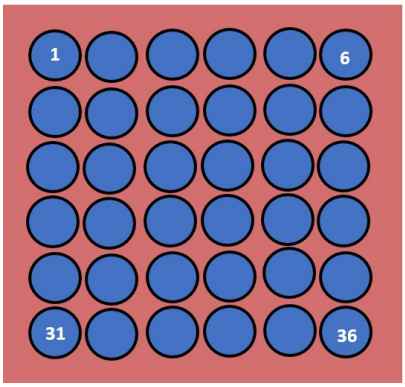

Figure 2. A visualization of the optical probe acquisition pattern

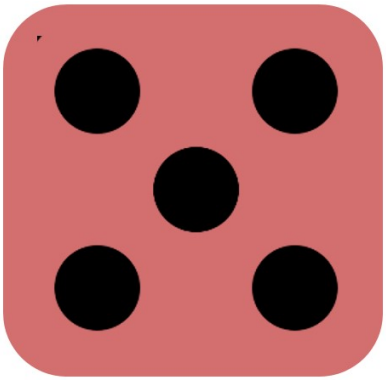

Figure 3. A visualization of the US probe acquisition pattern

2.3 Preprocessing

To improve the quality of the data for later classification various preprocessing steps were completed. The main preprocessing steps that I completed were: min-max normalization, dimensionality reduction, and data splitting.

The raw CSV files were imported into jupyter notebook, extraneous information was removed, and the raw data was stored into four 36x3647 numpy arrays representing the four tissue types. Due to variations in surface flatness, the probe did not maintain a constant height. To min-max normalization was performed to bring the peak intensities to 1. Since the light source itself does not generate light uniformly between 200 and 1000 nm, each tissue type had a common shape. To account for this the baseline spectrum was normalized and subtracted from each spectrum. The spectra were then min-max normalized again. This process resulted in each class showing much stronger visual differences.

All tissues have similar spectra outside of the visible spectrum of 400 to 700 nm. This is presumably due to the use of a white piece of paper for the baseline. This effectively only removed the shape in the range that the paper reflected, which may not include the ultraviolet (UV) and infrared (IR) range. This could be improved by using a mirror or getting the expected intensity curve and use the directly.

I then attempted to reduce the dimensionality of the data using Principle Component Analysis (PCA) to a more reasonable level such as 50 or 100 features. However, due to the very small size of the dataset, I could only reduce it to a size equal to the number of sample points of the smallest set (36). This significantly reduced the accuracy of the models and I decided not to perform this step.

The data from each class was then combined and randomly split into train, test, and validation sets at a ratio of 2:1:1. However, when using this data for testing I noticed that even the simplest machine learning models were achieving perfect accuracy and f1 scores. I determined that this was most likely due to the statistical similarity

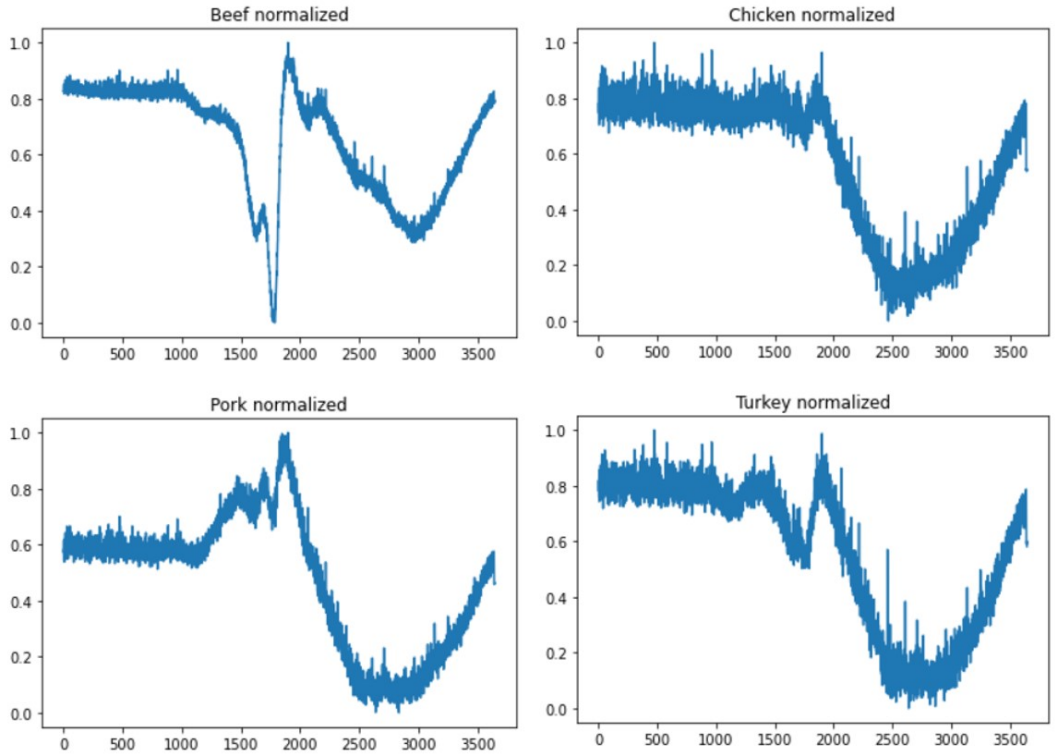

Figure 4. Example normalized spectrum for each tissue type

between the sets. Since most of the variation came from the spatial distance between readings, I devised a more robust data splitting method. Rather than splitting data randomly, I instead split it spatially into quadrants as seen in Figure [5]. I then conducted a leave-one-out cross-validation where 2 of the quadrants were used for training, 1 for validation, and 1 for testing. This ensured that spatially similar readings were not shared between sets. This resulted in a decrease in performance by the models, indicating that this was indeed an issue that needed to be addressed.

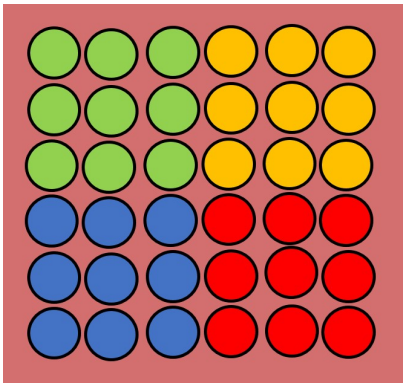

Figure 5. Showing quadrants

2.4 Model Description

I used three approaches to classify the tissues, 2 machine learning approaches and 1 deep learning approach. The architectures I used for classification are a Support Vector Machine (SVM), a K-Nearest-Neighbours (kNN)

from the sklearn library, and a custom-designed 1D CNN using the Pytorch library. The hyperparameters I tested are as follows. The default SVM hyperparameters were used. The main KNN hyperparameter of the KNN is the value of k which denotes the number of nearest neighbours used in the classification. The value of k was swept from 1 to 100 with the default value of 3 offering good results. The CNN I created had a large number of hyperparameters. The hyperparameters that I held constant were the number of convolutional and fully connected layers at 3 and 2 respectively, the activation function was kept as ReLU, and softmax was used at the end. The main hyperparameters that I tested were the kernel size, the depth of each convolutional layer, the number of nodes in the fully connected layers, the stride and length of the max-pooling layers, the learning rate, and the batch size. I tried kernel sizes of 3 and 5, feature depths of (6,12,24), (16,32,64), and (32,64,128), fully connected layers of size 20, 64 and 256, max-pooling strides of 2 and 4, max-pooling lengths of 2 and 4, a learning rate of 1e-2, 1e-3,1e-4, and a batch size of 1, 5, 10. The results of which will be discussed later.

The architecture I used can be seen in Figure [6]

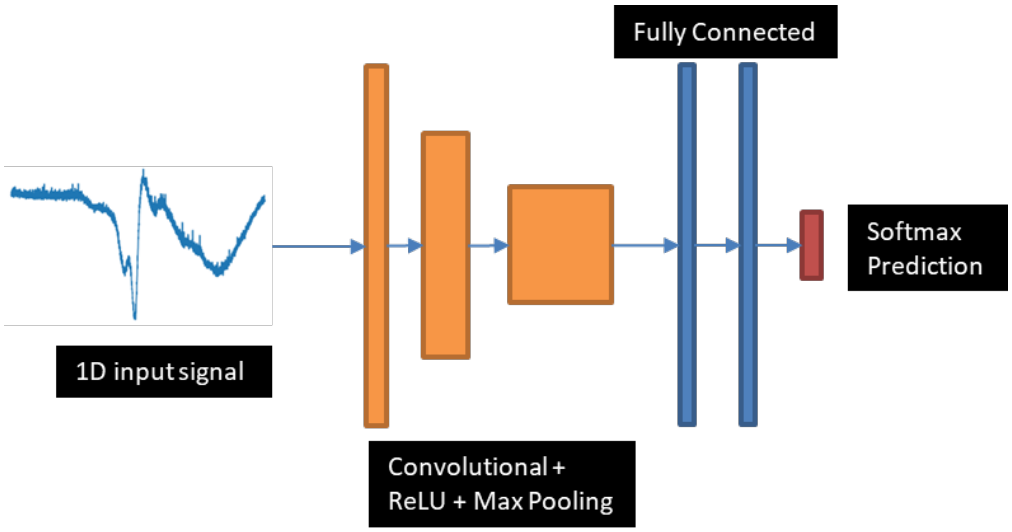

Figure 6. An example of the CNN structure used for experimentation

For training of the machine learning models I used 4-fold cross-validation where 2 of the quadrants were used to train, 1 was used for validation, and the final quadrant was used for testing. I then averaged over the 12 resulting models where. The CNN model trained considerably slower so I only did 3-fold cross-validation, reserving a single quadrant for testing. The resulting 3 models were averaged.

3. RESULTS

After training all models the accuracy and f1 score were calculated on the test sets. For SVM and KNN they were again averaged over all 12 permutations. For CNN they were only calculated on 1 permutation due to the time it took to train. The CNN results varied sporadically, achieving scores of 75% to 100% on identical architectures for the same datasets after retraining thus it was not possible to find specific performance metrics to compare. Therefore, a relative and qualitative analysis was performed. The kernel size of 3 showed significant improvement over a size of 5. A feature depth of (6,12,24) for the CNN layers never achieved an f1 of about 0.4 with a feature depth of (32,64,128) offering the best performance. A fully connected node size of 256, a max-pooling stride of 4, and a max-pooling size of 4 also offered significant performance improvements.

The final CNN model can be seen in Table [1].

As for training hyperparameters, the learning rate that allowed the model to learn was 1e-3 and a batch size of 1 was chosen as increasing the batch size delayed the convergence without providing any benefit.

Table 1. The final 1D CNN architecture

1D CNN	Value
Convolutional layers	3
Feature depth	(32,64,128)
Kernel size	3
Fully connected layers	3
Number of FC nodes	256
Max pooling stride	4
Max pooling size	4
Activation	ReLU

The final results can be seen in Table [2]. The KNN offered the best performance, achieving an average accuracy and f1 score of 0.956 and 0.954 respectively. This was followed by the CNN I made which produced an average accuracy and f1 score of 0.889 and 0.895 respectively. Finally, the SVM produced the worst results of 0.840 and 0.794 for accuracy and F1 score. It is worth noting that although the CNN model would sporadically achieve accuracy and F1 values of 1, showing the potential of the solution.

Table 2. Final accuracy and F1 score for each model

Model	Accuracy	F1 score
KNN	0.956	0.954
SVM	0.840	0.794
1D CNN	0.889	0.895

.

4. DISCUSSION

In this study, the diffuse optical reflection properties of tissues in the range of 200 to 1000 nm were successfully used to differentiate between samples of beef, chicken, pork, and turkey. Figure [4] shows clear differences in the spectra of each tissue type. The KNN model implemented achieved an average F1 score of 95% while training on just 72 spectra. This indicates that broadband spectroscopy or optical characterization is a viable technique for tissue classification. These results are promising for extending this technique to situations where strictly visual differences are not obvious such as in healthy vs cancerous tissues.

One limitation of this study is the lack of variability in the dataset. All spectra were taken from a 4cm x 4cm area of tissue which is not statistically representative of the generalized tissues. This can be addressed by collecting tissue samples from different butchers, at different locations in the animal, and sampling a large surface area. This will drastically increase the size and robustness of the dataset, which should make it more generalized and increase performance on independent testing. A larger dataset will also allow for, and potentially necessitate, the implementation of more complex deep learning models to account for the increase in nonlinearities within the classes.

Data augmentation is another way to further increase the robustness of the model. A way to test which areas of the spectrum are most important to the decision-making of the model would be to split the spectrum into ultraviolet, visible, and infrared. This would ensure that the model is not too reliant on the visible spectrum such that a cheap camera could be used instead. It is hypothesized that a simple webcam could be used for colour differentiation and classification in this experiment due to the prevalent visible differences. However, the

extension goal of this project is to characterize cancer vs healthy tissue where the visible differences are minimal. Using just the non-visible portions of the spectrum would be a could way to test the feasibility of this technique for applications with minimal physical differences such as in cancer detection.

An interesting extension would be to implement a feature importance network such that we could see precisely the wavelengths that are most important in the decision-making process. From this, we could tell the portion of the spectrum that carries the most weight and we could fine-tune our data to focus there. On the same note, it would be interesting to implement an attention-based deep learning method to potentially improve the performance of the CNN network to better filter out the unimportant features.

In the future, we plan to integrate this technique with the previously recorded raw RF ultrasound data to introduce structural information. This has been shown in Connolly et. al. to be a promising technique to extract sub-surface features which allow for the classification of tissues obscured by other tissues. This is especially important in breast tumour resection as a positive margin may be missed if cancer branches into and is subsequently surrounded by healthy tissue.

5. CONCLUSION

Using the measured spectra from a variety of tissue types we can conclude that diffuse reflection broadband spectroscopy is an effective means to optically characterize tissue. The resultant spectra from 4 different tissues were able to be easily classified with a simple KNN classifier achieving an average f1 score of 95.4% across 12 dataset permutations. Extending these results to breast conserving surgery, diffuse reflection broadband spectroscopy data can be used to effectively inform the surgeon about the presence of tumours in the post resection cavity. Combining this approach with US data has the potential to further inform the surgeon by allowing for more accurate detection of cancer hidden under healthy tissue. Overall, this information has the potential to drastically reduce the rate of revision surgery, increasing the emotional, physical, and financial health of those diagnosed with surgically treatable breast cancer.

REFERENCES

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: a cancer journal for clinicians* **71**, 209–249 (May 2021).

[2] Hargreaves, A. C., Mohamed, M., and Audisio, R. A., "Intra-operative guidance: Methods for achieving negative margins in breast conserving surgery," *Journal of Surgical Oncology* **110**(1), 21–25 (2014). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jso.23645.

[3] Black, D. M., Hunt, K. K., and Mittendorf, E. A., "Long term outcomes reporting the safety of breast conserving therapy compared to mastectomy: 20-year results of EORTC 10801," *Gland Surgery* **2**, 12023–12123 (Aug. 2013). Publisher: AME Publishing Company.

[4] Tzafetas, M., Mitra, A., Paraskevaidi, M., Bodai, Z., Kalliala, I., Bowden, S., Lathouras, K., Rosini, F., Szasz, M., Savage, A., Manoli, E., Balog, J., McKenzie, J., Lyons, D., Bennett, P., MacIntyre, D., Ghaem-Maghami, S., Takats, Z., and Kyrgiou, M., "The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease," **117**(13), 7338–7346. Publisher: Proceedings of the National Academy of Sciences.

[5] Connolly, L., Jamzad, A., Nikniazi, A., Poushimin, R., Nunzi, J. M., et al., "Feasibility of combined optical and acoustic imaging for surgical cavity scanning," in [*Proc. of SPIE Vol*], **12034**, 120341H–1 (2022).