Classification of Tumor Signatures from Electrosurgical Vapors using Mass Spectrometry and Machine Learning: A Feasibility Study

Laura Connolly[1], Amoon Jamzad[2], Martin Kaufmann[1], Rachel Rubino[4], Alireza Sedghi[2], Tamas Ungi[1], Mark Asselin[1], Scott Yam[3], John Rudan[5], Christopher Nicol[4], Gabor Fichtinger[1], Parvin Mousavi[2]

1. Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Kingston, Canada
2. Medical Informatics Laboratory, School of Computing, Queen's University, Kingston, Canada
3. Department of Electrical and Computer Engineering, Queen's University, Kingston, Canada
4. Nicol Lab, Queen's University, Kingston, Canada
5. Department of Surgery, Queen's University, Kingston, Canada

ABSTRACT

PURPOSE: The iKnife is a new surgical tool designed to aid in tumor resection procedures by providing enriched chemical feedback about the tumor resection cavity from electrosurgical vapors. We build and compare machine learning classifiers that are capable of distinguishing primary cancer from surrounding tissue at different stages of tumor progression. In developing our classification framework, we implement feature reduction and recognition tools that will assist in the translation of xenograft studies to clinical application and compare these tools to standard linear methods that have been previously demonstrated.

METHODS: Two cohorts (n=6 each) of 12 week old female immunocompromised (R*ag2*−/−;*Il2rg*−/−) mice were injected with the same human breast adenocarcinoma (MDA-MB-231) cell line. At 4 and 6 weeks after cell injection, mice in each cohort were respectively euthanized, followed by iKnife burns performed on tumors and tissues prior to sample collection for future studies. A feature reduction technique that uses a neural network is compared to traditional linear analysis. For each method, we fit a classifier to distinguish primary cancer from surrounding tissue.

RESULTS: Both classifiers can distinguish primary cancer from metastasis and surrounding tissue. The classifier that uses a neural network achieves an accuracy of 96.8% and the classifier without the neural network achieves an accuracy of 96%.

CONCLUSIONS: The performance of these classifiers indicate that this device has the potential to offer real-time, intraoperative classification of tissue. This technology may be used to assist in intraoperative margin detection and inform surgical decisions to offer a better standard of care for cancer patients.

1. PURPOSE

The intelligent knife or the iKnife distributed by Waters (www.Waters.com) is a method of mass spectrometry that is designed to aid in tissue diagnosis and recognition. The device uses rapid evaporative ionization mass spectrometry (REIMS) to aerosolize tissue samples and provide enriched feedback about the chemical and biological profile of the tissue [1]. The iKnife sampling device integrates REIMS technology into surgical workflow by capturing the vapor that is released from high-frequency electric current (electrocautery) tools commonly used in tumor resection procedures. This device is capable of complimenting medical imaging to inform the surgeon about the margins of the resection without any histological analysis. To classify tissue samples intraoperatively, without the assistance of a mass spectrometry expert, the device will require a sophisticated machine learning database to distinguish between tissues from different cancer types and anatomical regions.

The chemical profiles of tissue acquired from this device have been demonstrated to show distinctive separation of different tissue types in preliminary ex-vivo and in-vivo studies [2][3]. This suggests that there is value in developing a robust machine learning algorithm for this device that can accurately separate cancerous from healthy tissue for a variety of different cancer types. As there are a limited number of medical centers with access to iKnife and the required resources to facilitate clinical studies, it is important to develop efficient protocols and data analysis workflows for training data. Xenograft models are a viable tool for collecting this data outside of slow-moving clinical studies; these models can be further beneficial if the classifier trained using xenograft data can also be translated effectively to clinical data.

In this paper, we demonstrate the feasibility of REIMS data collection and classification of primary cancer from its surrounding tissues in xenograft mice models of human breast cancer, at two different stages of cancer progression. We developed a data analysis framework that can be scaled up to intraoperative data collection. The raw data collected from the iKnife contains tens of thousands of ion peaks. We use linear feature reduction and classification to separate various tissue types. Although linear classification is sufficient for classifying homogeneous tissue samples from xenograft, we extend this classifier beyond our xenograft study to create a more general and versatile classifier. For this purpose we develop and compare a new framework that incorporates deep learning for future translation to clinical data.

2. MATERIALS AND METHODS

2.1 Dataset

The dataset used in this study consists of two sets (n=6 each) of human breast cancer xenografts within R*ag2−/−;Il2rg−/−* mice ideally suited for these studies [4][5]. Female 12 week old mice were orthotopically injected in the abdominal mammary fat pad (Figure 1) with human MDA-MB-231 breast cancer cells (ATCC® HTB26™), established originally from the pleural effusion of a patient's metastatic mammary adenocarcinoma. Mice were euthanized at 4 and 6 weeks post injection to assess differences between early and late stage tumor progression using iKnife data of the tumor, surrounding tissue and metastasis at time of necropsy. Each mouse was subjected to an average of 27 burns of various tissue regions, including intraperitoneal (IP) membrane, primary breast tumor, lung, liver and kidney tissues (Figure 1).

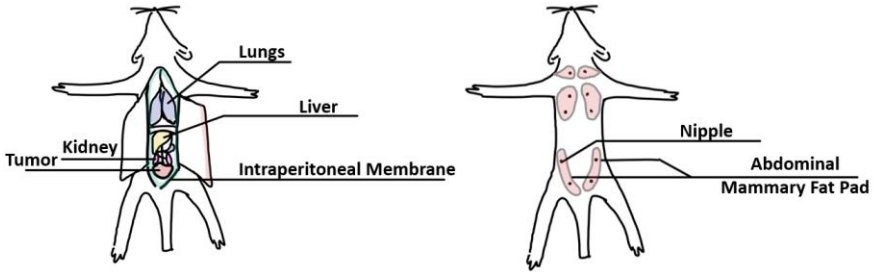

Figure 1: *Left*- location of electrocautery burns. *Right* – cell line injection site.

These areas were chosen because of their visible spectral distinction from one another (Figure 2) as well as their proximity to the injection site.

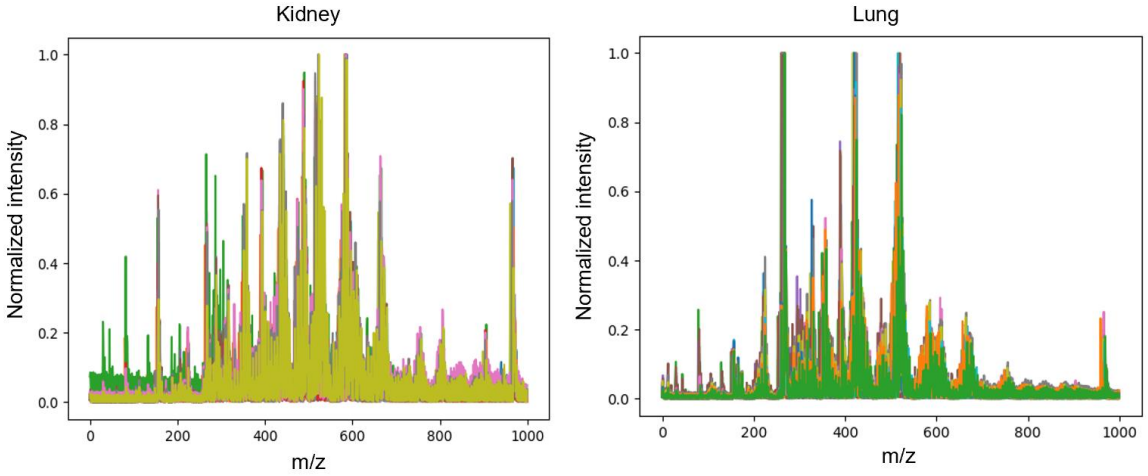

Figure 2: *Left* – Spectrum of multiple kidney samples. *Right* –Spectrum of multiple lung samples.

For each cohort, the breast cancer cells metastasized as anticipated to various tissues, including lung, liver and spleen. In the 6 week cohort, multiple highly visible macro-metastasis were noted in each organ, whereas metastasis in the 4 week cohort required more microscopic assessments (Figure 3). In this study, the aim was the separation of primary cancer from surrounding tissue and organs and the correlation between metastatic tissues at different stages of progression. To achieve this, metastasis bearing tissues are annotated here as the source organ or tissue from which they were derived. The spectra for each of these burns contains the corresponding intensity values for mass/charge ratios (m/z) in the 100-1200 range.

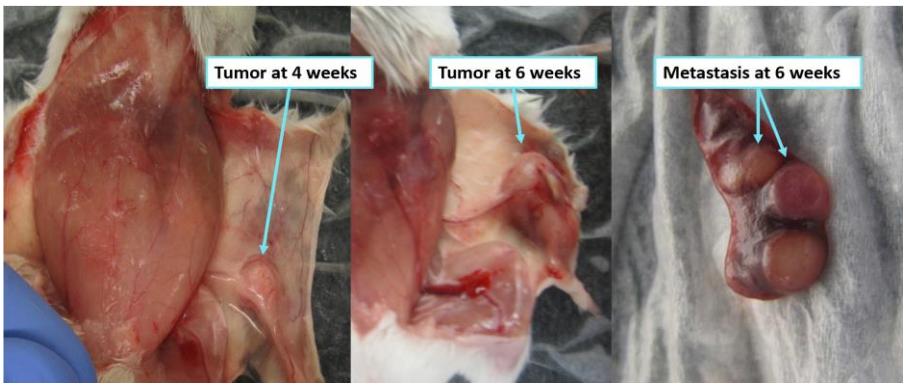

Figure 3: *Left* – Tumor progression after 4 weeks. *Middle* – Tumor progression after 6 weeks. *Right* – Metastasis on Spleen after 6 weeks.

The 6 week cohort contains a total of 6 mice and 145 burns and the 4 week cohort, containing 6 mice and 125 burns. The distribution of burns and tissue type is summarized in Table 1.

Table 1*:* The number of burns taken at each anatomical location for each cohort.

	Tumor	**Lung**	**Liver**	**Kidney**	**IP Membrane**
4 Week Cohort	18	23	40	19	25
6 Week Cohort	31	21	57	17	19

The data was partitioned into 3 separate training and 3 separate test sets for further analysis. We used multiple partitions because of the variability between samples of the same tissue type taken at week 4 versus week 6 which was likely caused by varying metastasis progression (Figure 4). The first partition was made up of samples from the 6 week cohort for training and the 4 week cohort for testing. The second partition was made up of samples from the 4 week cohort for training and the 6 week cohort for testing. Finally, the third partition was made using a combination of samples from both the 4 and 6 week cohort for training and a separate combination of both cohorts for testing. In all partitions the training and testing samples were distinctly separate.

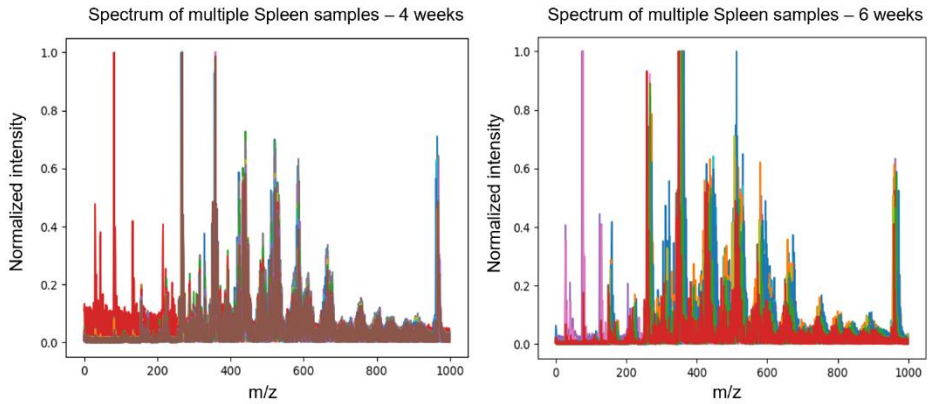

Figure 4: *Left* – Spectrum of multiple spleen samples at 4 weeks**.** *Right* **–** Spectrum of multiple spleen samples at 6 weeks.

2.2 Preprocessing

The spectral data was preprocessed before further analysis. The preprocessing steps consisted of normalization, smoothing, calibration and sub-band reduction (Figure 5). Lockmass correction was performed to ensure the data was calibrated correctly. The peaks of the spectra were shifted left or right according to the location of the lockmass peak – an ion of a known m/z ratio that is used to calibrate the system. The 600-900 m/z range was used for further classification as the predominant phospholipid and triglyceride peaks are within this range [6]. Sub-band reduction was performed to reduce the 50000 peaks in this region to 1000 by taking the maximum peak intensity for increments of approximately 50 peaks.

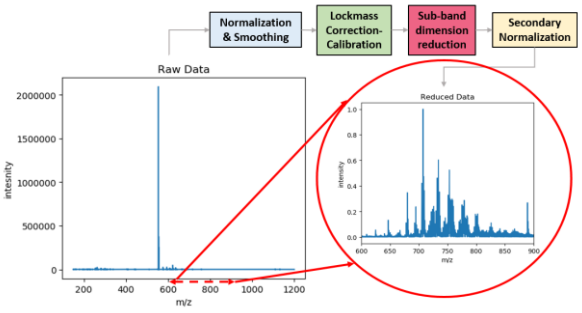

Figure 5: Summary of preprocessing steps.

2.3 Data Augmentation

Our dataset was limited to only 270 samples with an uneven distribution of tissue types. To improve this sample size, and reduce the impact of this imbalance, a larger and more balanced dataset is necessary. In the absence of further mouse samples, we augment the spectral data. We perform this by manipulating the slope of the spectra and offsetting the peaks with simulated noise. Figure 6 describes our proposed method of data augmentation used for this study. We added Gaussian noise to low frequency band isolated from the FFT of the data to simulate intensity fluctuations that are unrelated to unique tissue patterns. Gaussian noise is also added to offset peak positions and simulate instrumentation drift over time.

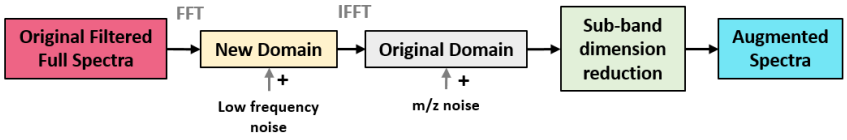

Figure 6: Data augmentation process.

Our method resulted in augmented spectra with the same general features as the original spectra but varying intensities and peak locations (Figure 7).

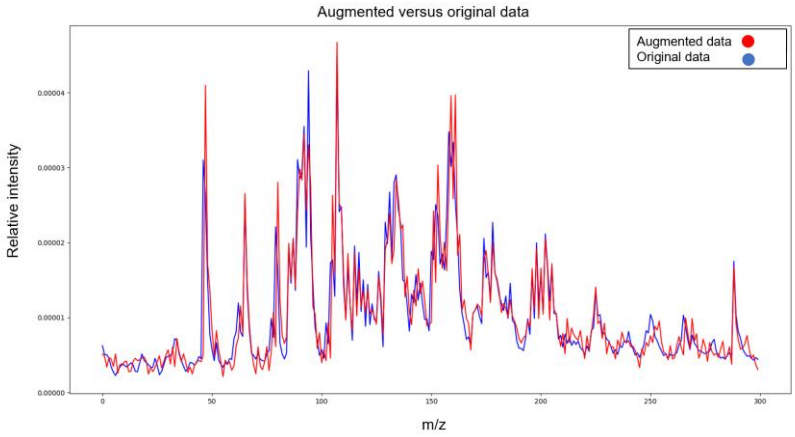

Figure 7: Augmented versus original spectra.

2.4 Feature Reduction and Classification

Linear discriminant analysis (LDA) and principal component analysis (PCA) are standard methods of feature reduction that look for linear data trends to identify correlation in large data. These methods have already been demonstrated for class separation of REIMS data [2]. A shortcoming of these methods is that they will not consider non-linear trends that may exist in the data [7]. Non-linear modeling approaches including deep learning are known to improve the accuracy of cancer classification problems and offer a more scalable approach when dealing with clinical data that is noisy in nature [7]. We developed our data analysis and classification framework with a neural network to identify non-linear trends. We compare the performance of the linear approaches with non-linear modeling as we believe moving forward to intraoperative data will require dealing with noisy and highly heterogeneous measurements.

The framework we developed consists of a neural network with two branches: an autoencoder for sophisticated dimension reduction and a classifier for prediction. This structure was chosen so that the decisions made at each layer of the neural network are chemically intuitive and can be used to understand classification outliers. An autoencoder is an unsupervised neural network designed to deconstruct data to a few distinct features and then reconstruct the original data from those features. The central layer of this network is the encoding layer which can be used as a reduced and informative representation of the input data, for our data this layer is a representation of the most significant ion peaks. Previous studies show the unique shape of asymmetrical autoencoders outperform symmetrical autoencoders by reducing predictive error during reconstruction [8]. An asymmetrical autoencoder model was used as the first branch in our network to decrease the dimensionality of the training space and improve classification accuracy. Each layer of the autoencoder was activated using the ReLU activation function and regularized according to Kullback-Liebler (KL) divergence [9]. A dropout rate of 0.5 was implemented after each layer to prevent overfitting. The autoencoder was optimized to minimize the mean squared error using gradient descent with the Adam update rule. The second branch of our network is a classifier that is trained at the same time so the network is forced to extract the most useful features for both reconstruction and classification (Figure 8). The final layer of this branch of the network was activated using the Soft max activation function and optimized to minimize the categorical cross entropy. The loss weights for each branch of the network are set to 0.1 and 10 for the autoencoder and classification branch respectively. Using our method of augmentation, the training data was augmented 6 times to increase the number of samples in the training set. As a baseline, we use PCA dimensionality reduction and LDA classification to validate our model. PCA was used to reduce our input data from 1000 to 150 principal components and an LDA classifier was fit to the augmented training set for classification.

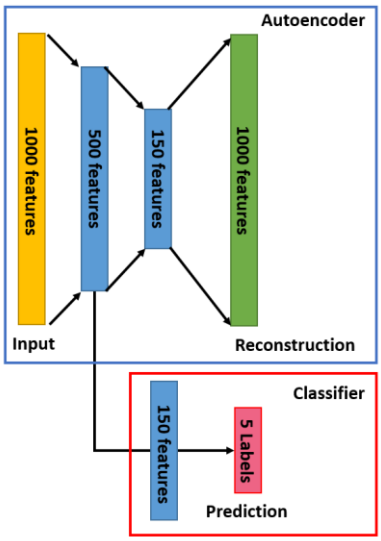

Figure 8: Neural network structure.

3. RESULTS AND DISCUSSION

Each classifier was used to predict the classes for each sample in the test set. Table 2 describes the combinations of samples used for the training and testing sets and the accuracy achieved with each method.

Table 2: Training and testing set specifications and results achieved using linear and non-linear classification techniques.

Training Set	Testing Set	PCA + LDA	NN
6 week cohort	4 week cohort	96%	96.8%
4 week cohort	6 week cohort	84.8%	91.7%
Both cohorts	Both cohorts	88.7%	92.2%

The confusion matrices obtained using the 6 week cohort as the training set and the 4 week cohort as the testing set are shown in Figure 9 below.

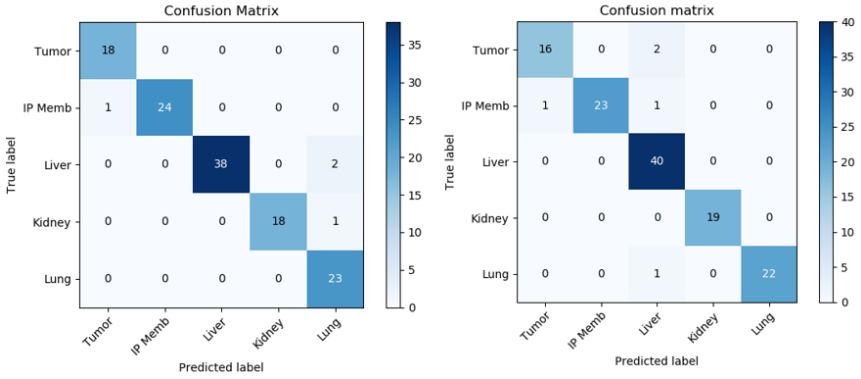

Figure 9: Results obtained using the 6 week cohort samples as the training set and the 4 week cohort as the test set. *Left*- classification results with the neural network (96.8%). *Right*- classification results without the neural network (96%).

Using our non-linear method, the classifier only made 4 incorrect predictions with respect to liver, IP membrane and kidney samples. The same confusion matrices were generated using the 4 week cohort as the training set and the 6 week cohort as the testing set (Figure 10).

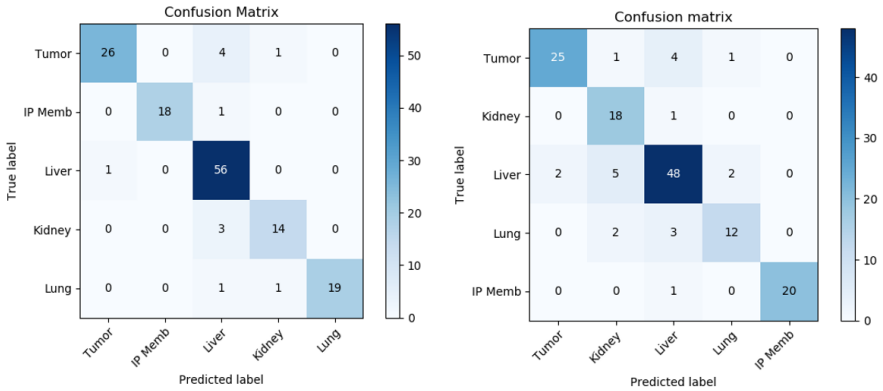

Figure 10: Results obtained using the 4 week cohort samples as the training set and the 6 week cohort as the test set. *Left*- classification results with the neural network (91.7%). *Right*- classification results without the neural network (84.8%).

Our non-linear method again outperforms traditional LDA and PCA, but the classifier incorrectly predicts a more distributed set of tissue types. Finally, the confusion matrices were generated using a combination of both cohorts for training and for testing (Figure 11).

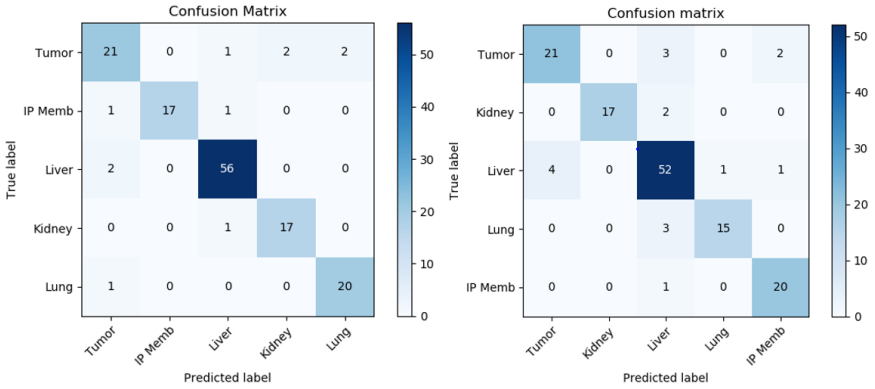

Figure 11: Results obtained using the 4 week and 6 week cohort samples as the training set and a combination of 4 week and 6 week samples as the test set. *Left*- classification results with the neural network (92.2%). *Right*- classification results without the neural network (88.7%).

Again the non-linear classifier outperforms traditional LDA and PCA but mislabeled tissue types are more random and distributed.

To better understand the misclassifications made using our non-linear network, the autoencoder reconstruction of one of these mislabeled samples was investigated. When the 6 week cohort was used for training set and the 4 week cohort was used for testing set, a liver sample was mislabeled as a lung sample (Figure 12).

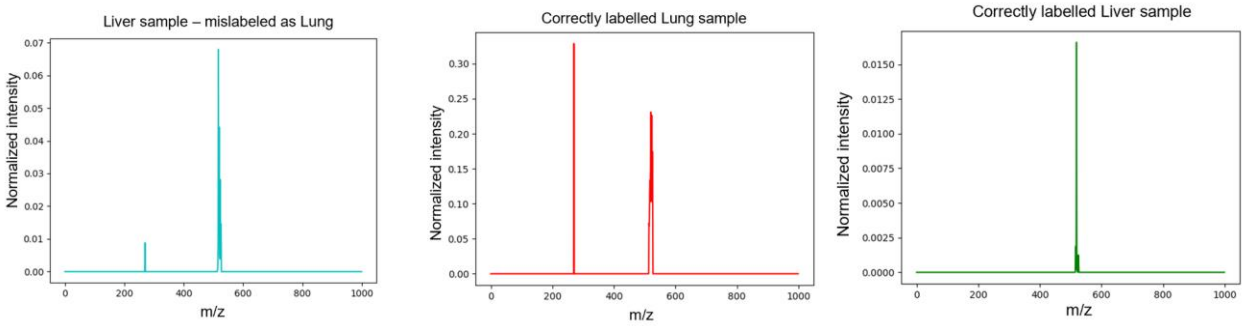

Figure 12: *Left* – Autoencoder reconstruction of mislabeled liver sample (labelled as lung). *Middle* – Autoencoder reconstruction of correctly labelled lung sample. *Right* – Autoencoder reconstruction of correctly labelled liver sample.

On the right in Figure 12 is a correctly labelled autoencoder reconstruction of the spectra for a liver sample. This reconstruction is consistent with what we observe for healthy liver tissue. On the left in Figure 12 is the reconstruction of an incorrectly labelled liver sample demonstrating an obvious qualitative difference from the healthy liver sample. The second peak in the spectrum occurring at approximately 270 m/z is unexpected for liver tissue. This spectra more closely resembles the reconstruction in the middle in Figure 12 of a correctly labelled lung sample. As demonstrated earlier, the training set in this partition (the 6 week cohort) had macro-metastasis on the lungs and liver whereas the testing set (4 week cohort) did not. Therefore, this incorrect classification may actually indicate that there is reason to revisit this mislabeled sample because there may be presence of early metastasis on the liver at 4 weeks.

These misclassifications can be used, alongside the correct classifications, to better understand the metastasis progression of tissue samples around the primary tumor at progressing stages of the disease. A more robust algorithm can be developed for this purpose by introducing a non-tumor bearing control group of mice and investigating metastasis and healthy tissue samples more closely.

4. CONCLUSIONS

The classifier is shown to accurately distinguish metastasis from surrounding tissue for cancer at different stages of tumor progression. The non-linear classifier outperforms the linear classifier for all three dataset partitions. The performance distinction of linear versus non-linear classification techniques is small (an average of 3.7% higher) for this data which is not unexpected because our dataset is homogeneous and the classes we are identifying are linearly separable. The addition of neural networks extends our classifier for more general usage and new datasets. The implementation of the autoencoder in our non-linear method also indicates that our network can be used to better understand the level of metastasis progression in various tissue regions. Future work may include a closer investigation into metastatic and non-metastatic tissue and comparing neural networks and linear models for the translation of xenograft models to clinical studies. Our results indicate that tissue classification is possible with the use of the iKnife and machine learning classification. This technology can be applied intraoperatively to inform margin decisions during tumor resections.

5. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

We demonstrate the feasibility of REIMS data collection and classification of primary cancer from its surrounding tissues in xenograft mice models of human breast cancer, at two different stages of cancer progression. Although linear classification is sufficient for classifying homogeneous tissue samples from xenograft, we extend this classifier beyond our homogeneous study to create a more general and versatile classifier.

REFERENCES

[1] J. Balog *et al.*, "Intraoperative tissue identification using rapid evaporative ionization mass spectrometry," *Sci. Transl. Med.*, 2013.

[2] E. R. St John *et al.*, "Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery," *Breast Cancer Res.*, vol. 19, no. 1, p. 59, May 2017.

[3] J. Alexander *et al.*, "A novel methodology for in vivo endoscopic phenotyping of colorectal cancer based on real-time analysis of the mucosal lipidome: a prospective observational study of the iKnife," *Surg. Endosc.*, vol. 31, no. 3, pp. 1361–1370, Mar. 2017.

[4] S. E. Le Dévédec *et al.*, "An improved model to study tumor cell autonomous metastasis programs using MTLn3 cells and the Rag2-/- γc-/- mouse," *Clin. Exp. Metastasis*, vol. 26, no. 7, pp. 673–684, Oct. 2009.

[5] J. Ahn *et al.*, "Fer protein-tyrosine kinase promotes lung adenocarcinoma cell invasion and tumor metastasis," *Mol. Cancer Res.*, vol. 11, no. 8, pp. 952–963, Aug. 2013.

[6] J. Balog *et al.*, "In Vivo Endoscopic Tissue Identification by Rapid Evaporative Ionization Mass Spectrometry (REIMS)," *Angew. Chemie Int. Ed.*, vol. 54, no. 38, pp. 11059–11062, Sep. 2015.

[7] R. Fakoor *et al.*, "Using deep learning to enhance cancer diagnosis and classification," *Proc. ICML Work. Role Mach. Learn. Transform. Healthc.*, vol. 28, 2013.

[8] A. Majumdar and A. Tripathi, "Asymmetric stacked autoencoder," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 911–918.

[9] S. A. Thomas *et al.*, "Dimensionality reduction of mass spectrometry imaging data using autoencoders," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1–7.