Deep learning segmentation of kidney tissue microarrays using infrared spectral imaging

Connor McKeown, Jordan Langlois, Zach Caterer, Dr. Rahul Gomes, Dr. Michael Walsh
Department of Computer Science, Department of Materials Science and Biomedical Engineering

University of Wisconsin Eau Claire

Introduction

Renal Cell Carcinoma

- Renal function is an essential marker in the classification of renal disease and **clinical symptoms of renal failure develop** when there is **15% renal function.**
- In this study, we used infrared spectroscopic (IR) imaging to investigate biomolecular markers from renal transplant biopsies.
- These images are used for the **classification of regions of fibrosis** from biopsies containing renal cell carcinoma (chromophobe and oncocytoma) and the prediction of fibrotic proliferation using biochemical signatures.

Infrared Spectroscopic (IR) Diagnosis

- IR spectroscopy is a diagnostic approach utilizing **human tissue to label biochemical signatures.** Images are captured in several hundred wavelengths in the infrared region of the electromagnetic spectrum giving researchers access to more information than traditional RGB images captured by a microscope.
- While images captured in several bands are great for disease diagnosis, it poses **significant challenges for manual cell review** by a pathologist.

Project Goals

- Apply **feature selection** to remove data with less importance ands reduce dimensionality.
- Apply a **deep learning model** on filtered dataset for identification of fibrosis

The Dataset

Our dataset contains **767 bands** of spectral data that has been extracted from a mosaic of 36 kidney cancer tumors cores. The image of all cores is shown in Figure 1. The tumors contain the **Chromophobe and Oncocytoma** variants of kidney cancer.

In our preprocessing stage, regions of interest (ROIs) were identified as cancer that would allow us to analyze the specific data in a region that can be classified as either Chromophobe or Oncocytoma by the researcher.

Software used for Analysis

TensorFlow Python Scikit-learn ENVI

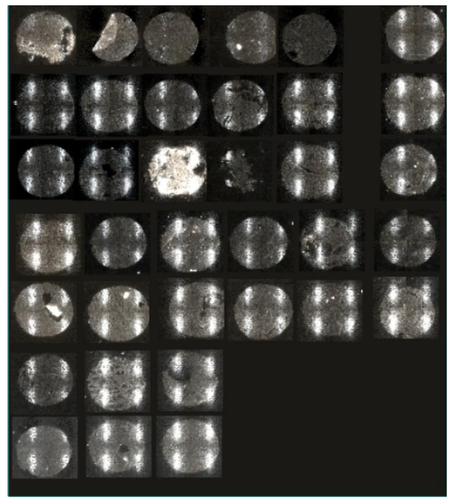

Fig 1. An image of the mosaic of 36 cores. This image shows grayscale representation of one band.

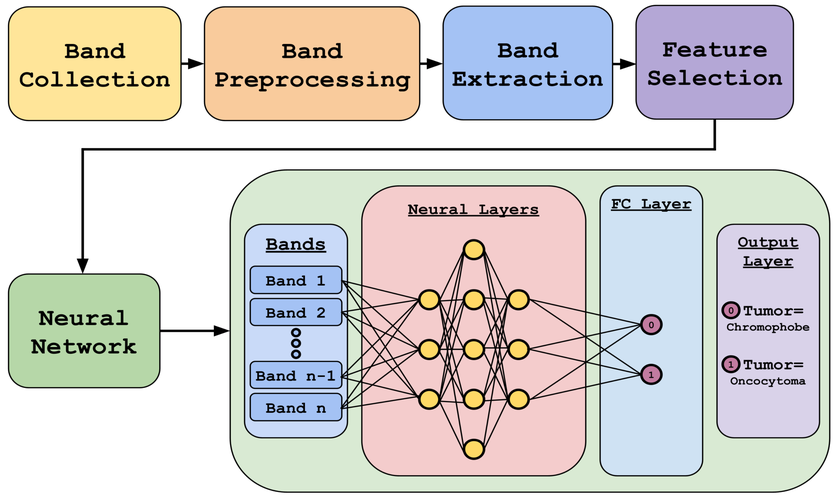

Fig 2. Visual representation of our proposed model.

Proposed Architecture

Feature Selection

To reduce dimensionality, an **ensemble learning** model called ExtraTreesClassifier was applied to all 767 bands to identify their importance. Based on the results, we selected the **top 150 bands** which returned the **highest importance** associated with model prediction. The importance values of these 150 bands are shown in Figure 3.

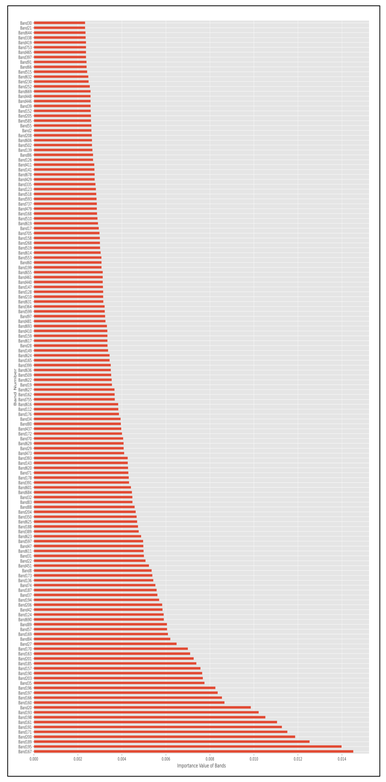

Fig 3. Importance of 150 bands.

Neural Network

- A basic neural network model was used to perform **binary classification** between two forms of cancer.
- The neural network consisted of one input layer, seven dense layers, and one output layer.
- Model was trained for **100 epochs.**
- **65%** of the data was used for training and **35%** was used for validation.
- The entire approach is shown in Figure 2.

Acknowledgements

The research has been made possible through funding from student-faculty collaborative research by Office of Research and Sponsored Programs and the Blugold Fellowship Program.

Results

- When the model was trained on all 767 bands, training data accuracy reached 100% but validation data accuracy was only 54.4%. This shows that the model was **overfitting** the training data.
- Using selected 150 features **significantly improved** validation accuracy from 54.4% to 81.8%. The model training accuracy was 94.4%.
- **Using the reduced number of features creates a better cancer prediction model.** The model training results are shown in Figure 4. Results are shown in Table 1

Accuracy Comparison	Training Accuracy	Validation Accuracy
All Features	100%	54.40%
150 Features	94.40%	81.80%

Table 1. Final accuracies after 100 epochs.

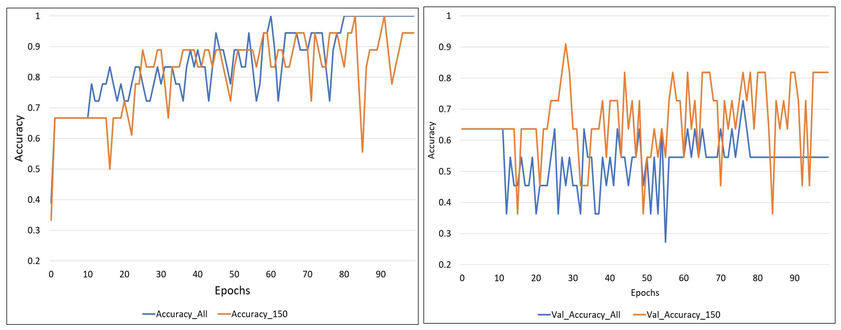

Fig 4. Accuracy graphs (training+validation) for a basic neural network on all datasets (left) and feature selected dataset (right) on our dataset that ran for 100 epochs.

Future work

- **Explore the spectral signatures** of these 150 bands and why they are so closely associated with cancer.
- **Tune** the deep learning model by using more cores.
- Apply **bootstrapping** for making the model more robust.

We thank the Office of Research and Sponsored Programs for supporting this research, and Learning & Technology Services for printing this poster.