Fall_2023_Report

November 30, 2023

Thermal liquid biopsy (TLB) is a growing field of biochemistry that holds potential to be used as a tool to diagnose and monitor disease. Thermograms, which are the results of TLB, differ based on the health status of a patient, and can be used to train classification models to identify an illness. This report details the work that has been done with Lung Cancer (LC) Thermograms during the Fall 2023 semester. The goal of this work was to use machine learning to train classification models to identify LC diagnosis and stage using TLB. Table 1 shows the LC types and frequencies included in the dataset.

Table 1: Number of Lung Cancer Samples by Type

	Cancer Type	Count
0	Adenocarcinoma	72
1	Control	51
2	Squamous	46
3	SCLC	16
4	NOS	8
5	Large cell	6

Figure 1 shows the median thermogram for each LC type. It can be seen that the median curves present differently for each type of LC. Despite the differences in the median curves, variation in individual samples is high. Figure 2 shows the median thermogram curve, as well as the 5th and 95th quantiles observed. The variation of individual samples within each group suggests that classification using thermograms will be difficult.

0.1 Results of Pairwise Classifications

For all LC types with 10 or more thermograms, pairwise classifications were conducted. 1,000 iterations of bootstrap cross-validation (BSCV) was utilized to evaluate random forest model performance. For each bootstrapped sample, a grid of hyper parameter combinations was searched to find the best version. The parameters that were evaluated were the number of trees, the number of features, and tree depth. The hyper parameter states were grouped together, and the combination with the highest balanced accuracy was selected as the final model for each combination. Table 2 shows the results of each cancer type combination that were classified. Balanced accuracy was used as the primary metric for evaluating performance because it accounts for class imbalance. The balanced accuracy ranges from 0.51 - 0.67. The poor model performance for each cancer pair is not surprising given that the visualizations for each pair (Figures 3-8) have a large amount of overlap.

Table 2: Classification Results per Cancer Pair

	Pair	Weighted Accuracy	AUC
5	Control - SCLC	0.677083	0.799829
4	Squamous - Control	0.588179	0.622351
2	Adeno - Control	0.575928	0.619652
3	Squamous - SCLC	0.541017	0.598070
0	Adeno - SCLC	0.530405	0.593018
1	Adeno - Squamous	0.513628	0.524180

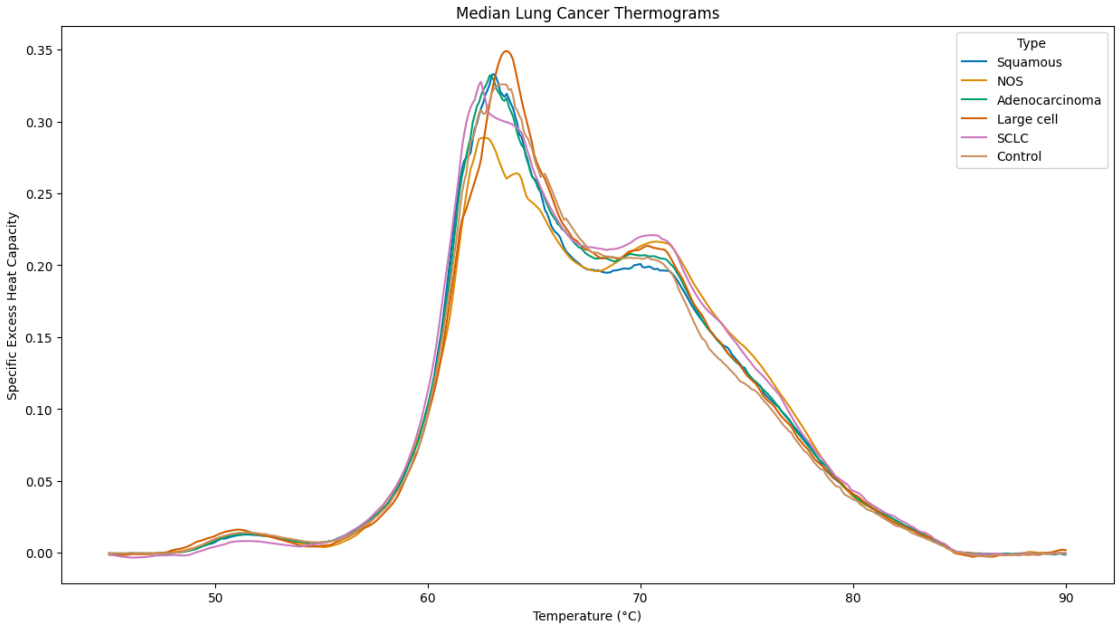

Figure 1: Median Lung Cancer Thermograms by type.

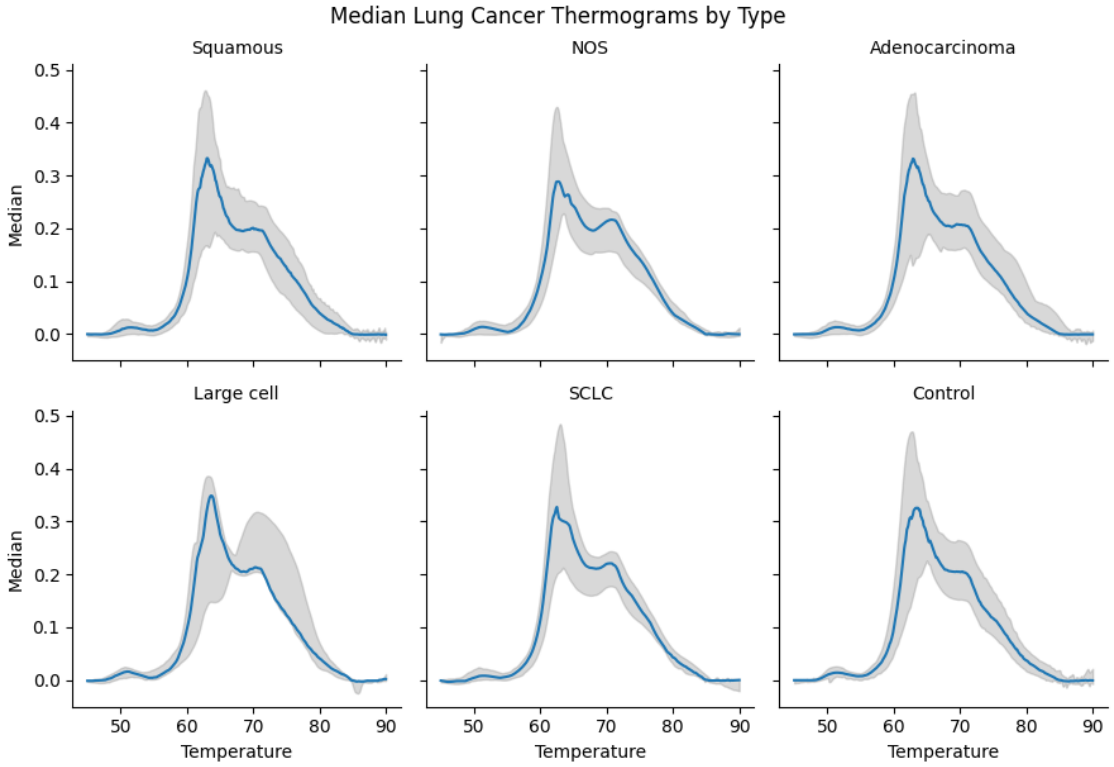

Figure 2: Lung Cancer Thermograms by Type with Quantile Ribbons

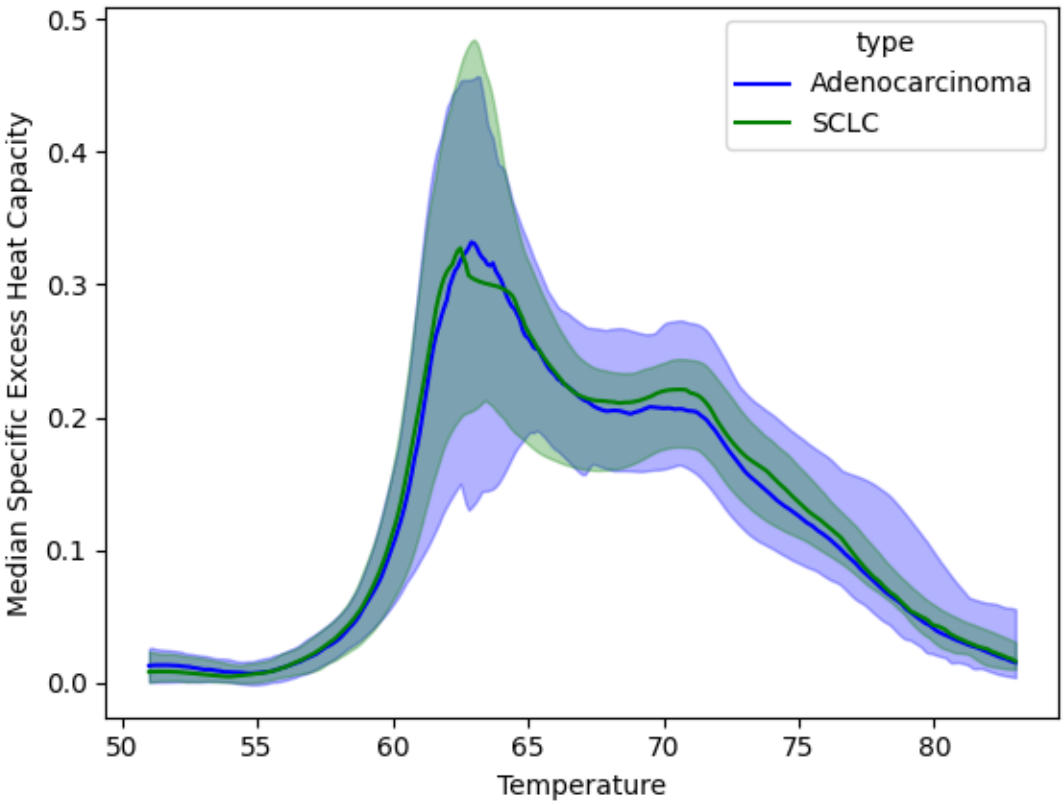

Figure 3: Median Adenocarcinoma and Small Cell Lung Cancer Thermograms

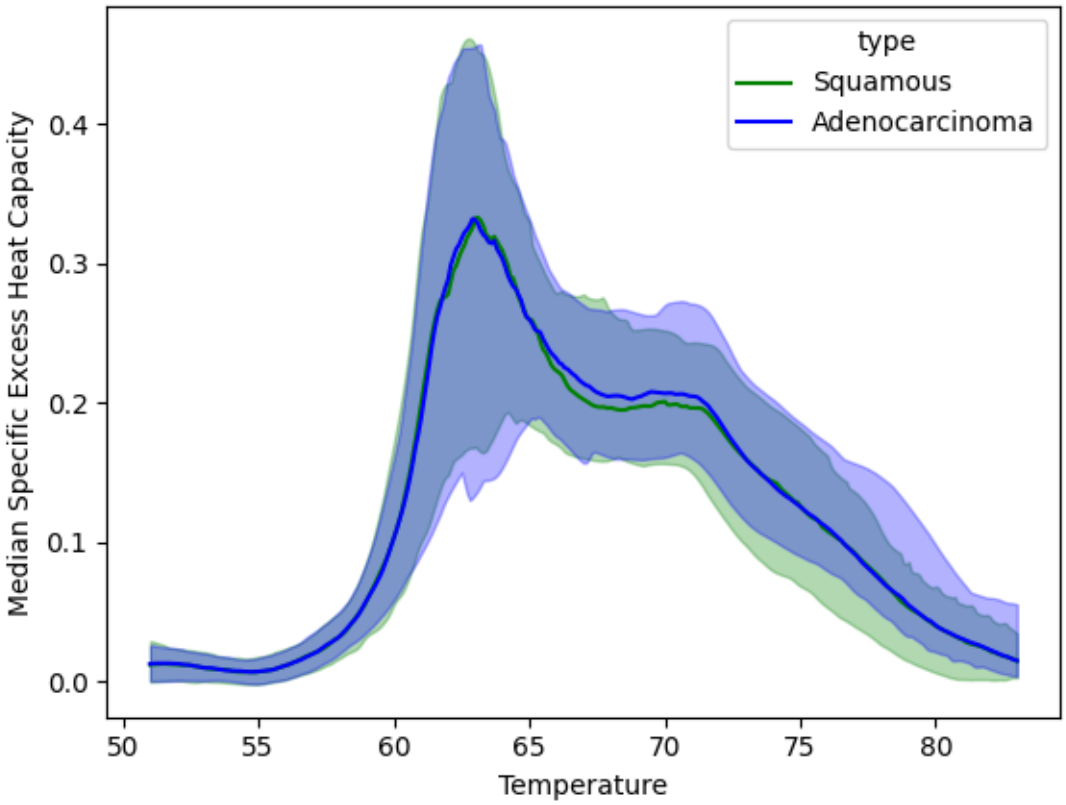

Figure 4: Median Squamous and Adenocarcinoma Thermograms

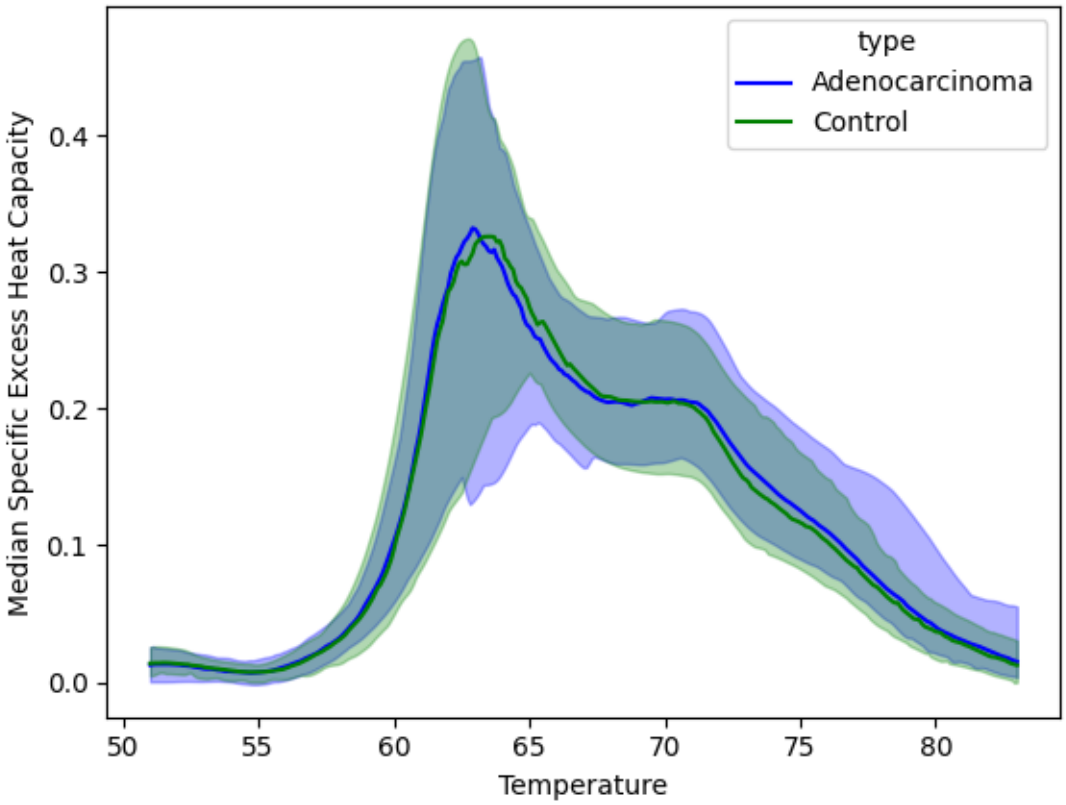

Figure 5: Median Adenocarcinoma and Control (Healthy) Thermograms

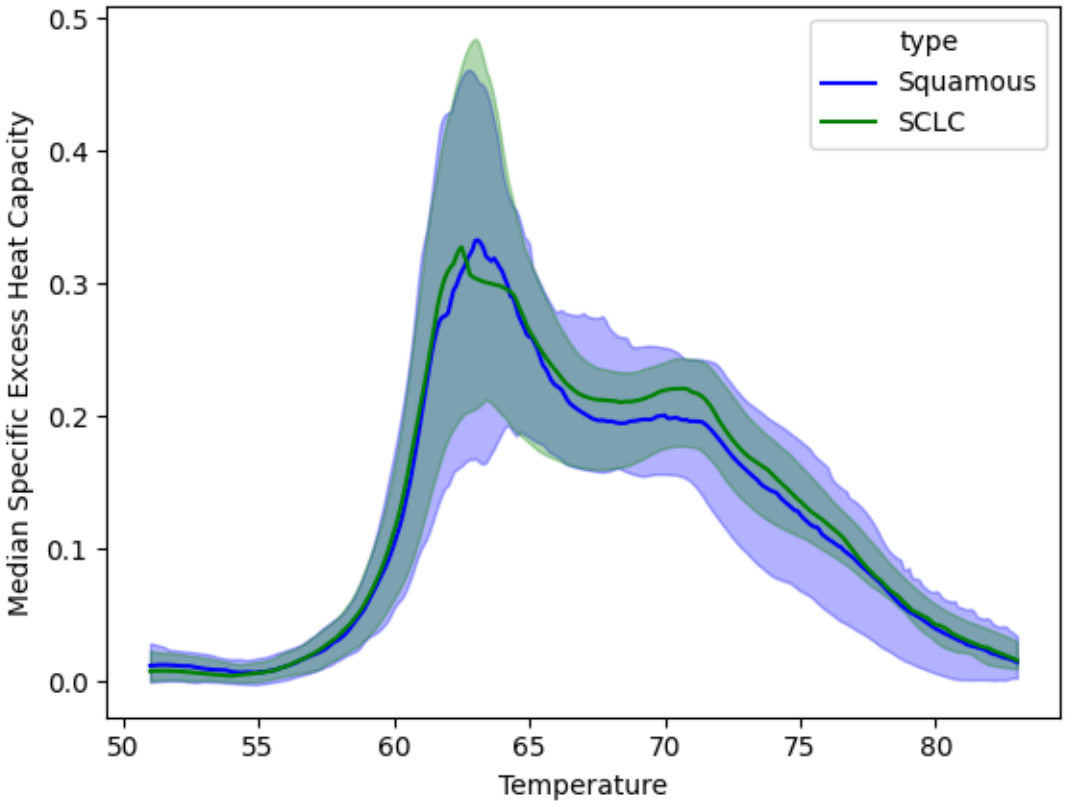

Figure 6: Median Squamous and Small Cell Lung Cancer Thermograms

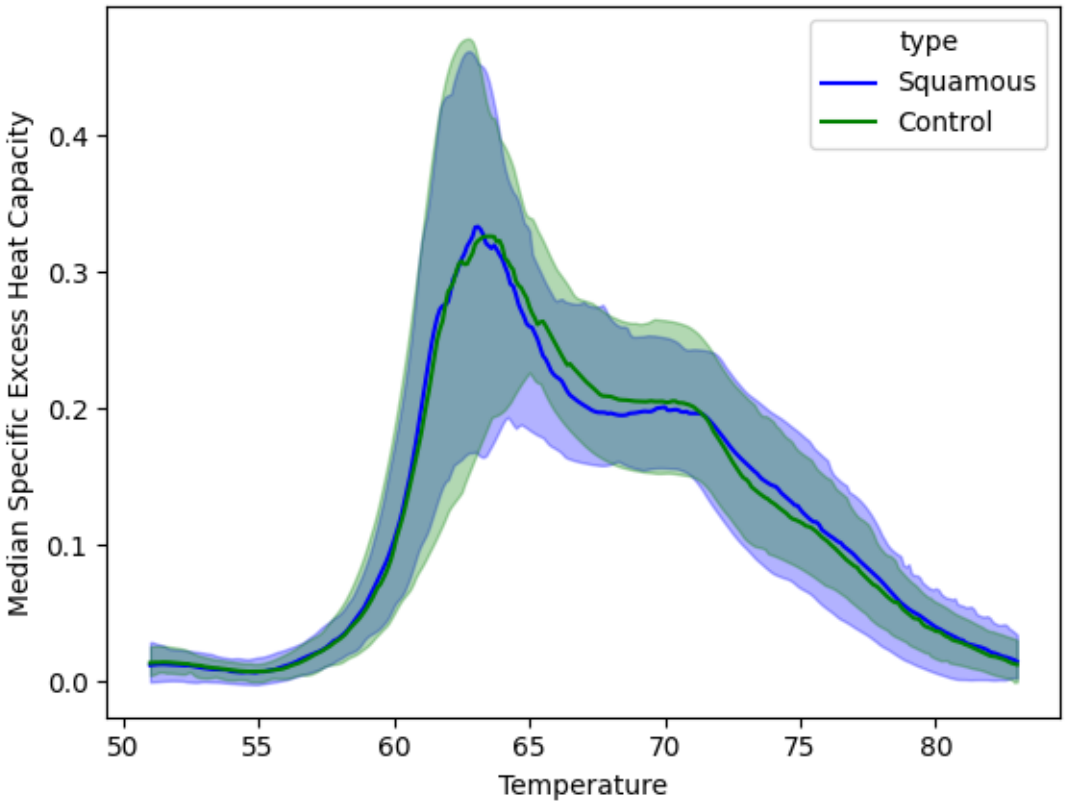

Figure 7: Median Squamous and Control Thermograms

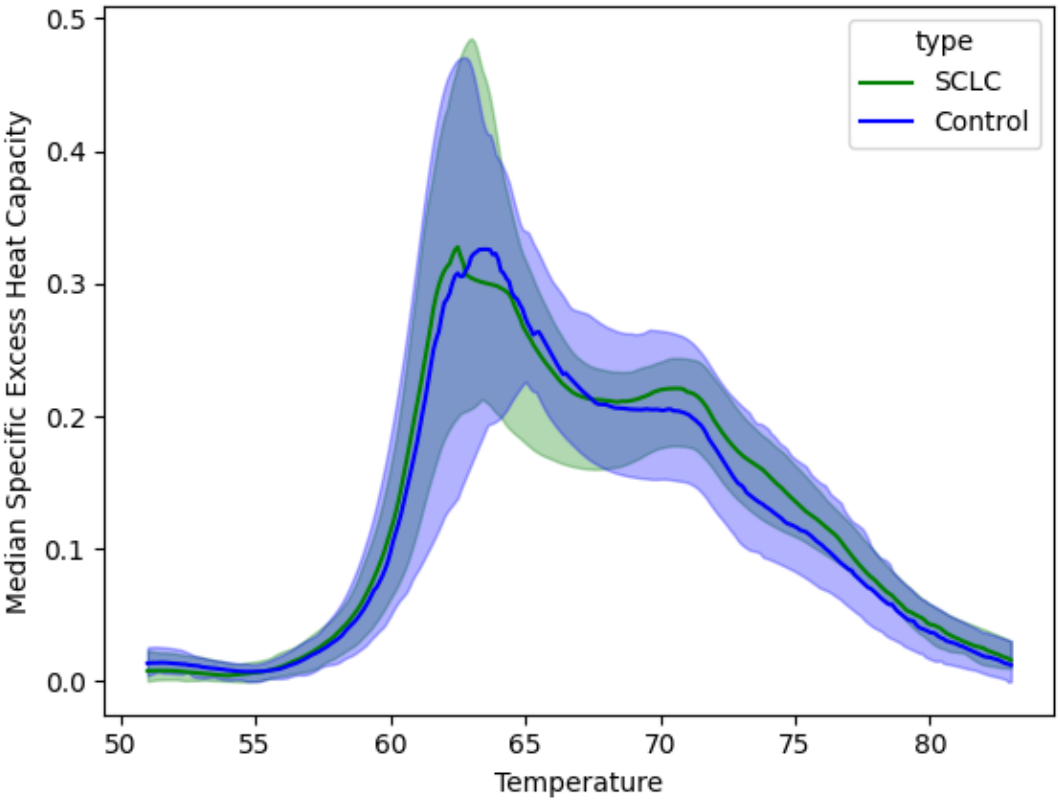

Figure 8: Median Small Cell Lung Cancer and Control (Healthy) Thermograms