Spatial Analysis of Tumor Heterogeneity Using Machine Learning Techniques
(A paper to be published at the IEEE MASS conference.)

Abstract—The treatment and study of cancer are in part hindered by cells and tissue of the same cancer type exhibiting differences from one another. This *tumor heterogeneity* is thus an important characteristic worth better understanding and analyzing. In the past, this analysis has been mostly carried out manually by clinicians and researchers. However, with advances in algorithms and computational resources, we can analyze tumor samples using statistical methods and machine learning techniques. Our work features an automated pipeline for analyzing the spatial gene expression of tumor tissue samples. For the task of segmenting tissue regions into tumor, non-tumor, and hepatocyte regions, our models (logistic regression, support vector machine, and random forest classifier) achieve over 90 percent accuracy on all tests. We find these results to be encouraging for future research in spatial analysis of tumor heterogeneity using similar methods.

Index Terms—machine learning, computational biology, can- cer, tumor heterogeneity, spatial gene expression