Learning Patterns of Somatic Mutations and Medical Images in Human Cancer With Machine Learning

by

Rahul Mehta
B.S., University of Illinois at Urbana Champaign, 2011

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Bioengineering
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:
Prof. Muge Karaman, Advisor
Prof. Yang Dai, Chair
Prof. Jie Liang
Prof. Sujata Gaitonde, Department of Pathology
Prof. Yang Lu, Unversity of Texas MD Anderson Cancer Center

This dissertation is dedicated to my grandparents Bhagatram, Kaushalya, Vishwa, and Sarla Mehta.

ACKNOWLEDGMENT

This thesis is the culmination of a gratifying journey at UIC where my growth and development in graduate school was guided by both activities outside and inside the university. The completion of this thesis is due in no small part to my advisor Muge Karaman who took me on at the late stages of my graduate career. Muge has been a guiding hand to shaping the final versions of much of the research that is presented in this thesis regardless of how far it ventures away from her core areas.

I have been fortunate with the members of my committee and their patience during the writing of the my work. My final defense committee members Yang Dai, Jie Liang, Yang Lu, and Sujata Gaitonde have been very grateful with their time. I would also like to thank Dr. Yang Lu for allowing me to work him during my beginning years of graduate work and creating an avenue to collaborate with clinicians.

I would also like to thank Juan Carlos Niebles, Jie Li, and Fei Fei Li who took me on as an undergraduate student at the Beckman Institute at UIUC. I got to learn and experience the framework of the computer vision algorithms that have now exploded into our daily lives. Much of the work I did with them was the inspiration of ideas within this thesis.

Outside of the confines of graduate work I am eternally grateful to my doctors Dr. John Quigley and Dr. Leslie Jabine and nurses Janet Golick and Jessica Ittner for keeping me functional.

ACKNOWLEDGMENT (Continued)

I may have completed this thesis in a shorter amount of time, but it would have been infinitely more challenging without the love and support of my wife Jane Makin. Through her I have been able to extend my family to include Gwendolyn van Paasschen and Mike Palmer who have provided numerous welcome distractions. I would especially like to thank my parents Shekhar and Dolly Mehta for the encouragement to pursue graduate work. They have have taught me the value of exploring ideas and (maybe too much) optimism. I must of course thank my little brother Manu, who I have seen to grow from watching a repeat of Finding Nemo to editing this thesis.

CONTRIBUTIONS OF AUTHORS

Chapter 3 is a published manuscript in which I am a first author. Drs. Cai, Kumar, Knuttinen, Anderson, and Lu provided editing for the manuscript. Dr. Lu helped in creating the project idea and data collection. This chapter is based on the publication (1) (permissions available in Appendix A:

Mehta, Rahul, Kejia Cai, Nishant Kumar, M. Grace Knuttinen, Thomas M. Anderson, Hui Lu, and Yang Lu. "A lesion-based response prediction model using pretherapy PET/CT image features for Y90 radioembolization to hepatic malignancies." Technology in cancer research & treatment 16, no. 5 (2017): 620-629.

Chapter 4 is an unpublished manuscript in which I am a first author. Zheng Zhong and Drs. Karaman and Zhou helped in data processing and editing of the manuscript. The remaining authors helped obtained the data used for this manuscript. This chapter is based on (2):

Mehta, R., Karaman, M., Bu, Y., Zhong, Shiwei, W., Zhou, C., Weihong, H., Maosheng, X., Zhou, J. (2020). Discrimination of Malignant and Benign Breast Lesions Using Machine Learning on Multi-modal Diffusion MRI Parameters

Chapter 5 is a published manuscript in which I am a first author. Dr. Lu helped in processing and anotating parts of the medical images. Dr. Karaman helped in editing of the manuscript. This chapter is based on (3) (permissions available in Appendix B:

CONTRIBUTIONS OF AUTHORS (Continued)

Mehta, Rahul and Lu, Yang and Muge Karaman. "Mapping of Lesion Images to Somatic Mutations." Data and Text Mining in Biomedical Informatics (DTMBIO 2020), 2020, Galway, Ireland.

Chapter 6 is a published manuscript in which I am a first author. Dr. Karaman helped in editing of the manuscript. This chapter is based on (3) (permissions available in Appendix C:

Mehta, Rahul, and Muge Karaman. "Correlated Mixed Membership Modeling of Somatic Mutations." 2020 International Joint Conference on Neural Networks (IJCNN 2020), 2020, Glasgow, UK.

TABLE OF CONTENTS

CHAPTER		PAGE
1	**INTRODUCTION**	1
1.1	Motivation	1
1.2	Contributions	8
1.2.1	Learning from Medical Images	8
1.2.2	Learning From Variations of Somatic Mutation Profiles	11
1.3	Dissertation Overview	12
2	**BACKGROUND**	14
2.1	Medical Imaging	14
2.1.1	PET/CT	14
2.1.2	Diffusion Weighted Magnetic Resonance Imaging	16
2.2	Lesion Heterogeneity	20
2.3	Metastatic Cancer in the Liver	24
2.4	Breast Cancer	25
2.5	Somatic Mutations	26
2.6	Machine Learning	27
2.7	Medical Imaging Features	29
2.7.1	First and Second Order Statistics	30
2.7.2	Automated Feature Generation	31
2.8	Three Dimensional Object Features	32
2.9	Latent Variable Models	33
2.9.1	Overview	33
2.9.2	Decomposing Matrices with Latent Variable Models	36
2.9.3	Latent Feature Models	37
2.10	Modeling Choices for Latent Feature or Variable Models	38
2.10.1	De Finetti's Theorem	39
2.10.2	Completely Random Measure	40
2.10.3	Gamma Process	41
2.10.4	Learning Via Variational Inference	42
3	**A LESION BASED RESPONSE PREDICTION MODEL USING PRETHERAPY PET/CT IMAGE FEATURES FOR Y90 RADIOEMBOLIZATION TO HEPATIC MALIGNANCIES**	44
3.1	Motivation	44
3.2	Materials and Methods	47
3.2.1	Patients	47
3.2.2	Patient Treatment	48

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

3.2.3 Patient Imaging . . . 49
3.2.4 Clinical Evaluation of Treatment Responses . . . 49
3.2.5 Quantitative Analysis . . . 52
3.2.6 Predictive Performance and Statistical Analyses . . . 55
3.3 Results . . . 56
3.4 Discussion . . . 59
3.5 Conclusion . . . 64

4 DISCRIMINATION OF MALIGNANT AND BENIGN BREAST LESIONS USING MACHINE LEARNING ON MULTI-MODAL DIFFUSION MRI PARAMETERS . . . 65
4.1 Motivation . . . 65
4.2 Material and Methods . . . 68
4.2.1 Patient Characteristics . . . 68
4.2.2 Diffusion-weighted Image Acquisition . . . 69
4.2.3 Diffusion-weighted Image Analysis . . . 69
4.3 Data Analysis . . . 72
4.3.1 Pre-Processing . . . 72
4.3.2 Feature Extraction . . . 72
4.3.3 Feature Selection . . . 73
4.3.4 Statistical Analysis . . . 73
4.4 Results . . . 74
4.4.1 Diffusion Parameter Maps . . . 74
4.4.2 Comparison of Features . . . 74
4.4.3 Number of Histogram Bins Comparison . . . 76
4.4.4 Comparison of Classifier Performance . . . 76
4.5 Discussion . . . 80
4.6 Conclusion . . . 86

5 MAPPING OF LESION IMAGES TO SOMATIC MUTATIONS 87
5.1 Motivation . . . 87
5.2 Background . . . 91
5.2.1 Point Clouds . . . 91
5.2.2 Normalizing Flows . . . 92
5.2.3 High Dimensional Discrete Data . . . 93
5.2.4 Related Work . . . 95
5.3 Dataset . . . 96
5.3.1 Lesion as a Point Cloud . . . 96
5.3.2 Somatic Mutations Representation . . . 97
5.4 Model . . . 97
5.4.1 Lesion Point Cloud to Somatic Mutations . . . 99
5.4.2 Variational Objective . . . 101

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

5.4.3 Optimization of Shared Latent Space . . . 103
5.4.4 Implementation Details . . . 104
5.5 Experiments . . . 106
5.5.1 Comparison Metrics . . . 106
5.5.2 Results . . . 108
5.6 Discussion . . . 110
5.7 Conclusion . . . 114

6 CORRELATED MIXED MEMBERSHIP MODELING OF SOMATIC MUTATIONS . . . 117
6.1 Motivation . . . 117
6.2 Background . . . 121
6.2.1 Negative Binomial Distribution . . . 121
6.2.2 Mixed Membership Modeling . . . 122
6.2.3 Latent Space as a Correlated Random Measure . . . 124
6.3 Model . . . 125
6.3.1 Correlated Zero Inflated Negative Binomial Process . . . 125
6.3.2 Generative Structure . . . 126
6.3.3 Inference . . . 129
6.4 Experiments . . . 131
6.4.1 Comparison Metrics . . . 133
6.5 Results . . . 134
6.5.1 Quantitative . . . 134
6.5.2 Lung Adenocarcinoma Case Study . . . 138
6.5.3 Colon Cancer Case Study . . . 139
6.6 Discussion and Future Work . . . 139

7 CONCLUSIONS . . . 144
7.1 Contributions of Thesis . . . 144
7.2 Future Directions . . . 147

APPENDICES . . . 150
Appendix A . . . 151
Appendix B . . . 152
Appendix C . . . 154

REFERENCES . . . 155

VITA . . . 185

LIST OF TABLES

TABLE		PAGE
I	Characteristics of Eligible Patients for this Study	47
II	Combined Imaging Criteria for Assessment of Treatment Response.	48
III	Recall and Precision of the different feature groups.	56
IV	Various fused feature groups and the resulting Precision and Recall.	57
V	RMSE of LLOST Models	109
VI	PPV of LLOST_{B}	111
VII	CoZINB Inference	131
VIII	Preplexity CoZINB	134

LIST OF FIGURES

FIGURE		PAGE
1	Dissertation Overview	8
2	Example of Lesion Heterogeneity	22
3	A generic construction of the latent variable model	34
4	Gaussian Mixture Model	35
5	Example of Matrix Decomposition	37
6	FDT PET/CT of Two Patients	50
7	Liver Segmentation Model	51
8	ROC Plot of 3D Features	58
9	Average Visual Words	60
10	Specific Visual Words	61
11	DWI Model	71
12	DWI Map Comparison	75
13	Feature Comparison	77
14	Histogram Comparison Performance	78
15	ROC Comparison of Classifiers	79
16	Cross-Validation Performance	80
17	Test Performance	81
18	Sparsity and Point Clouds	94
19	LLOST Model Architecture	98
20	Invertible Neural Network	105
21	Perplexity and Boxplots of LLOST	107
22	Point prediction error for Tumor Mutation Load	115
23	TSNE of LLOST	116
24	Biological Influence CoZINB	123
25	CoZINB Plate Model	128
26	Precision@1 CoZINB	136
27	Specificity CoZINB	137
28	All Factors in CoZINB	141
29	LUAD TML Mutual Exclusivity	142
30	COAD TML Mutual Exclusivity	143

SUMMARY

Datasets generated at different stages of cancer diagnosis and prognosis has led to an era of precision medicine, where clinicians and researchers can now potentially create a tailored treatment plan for an individual patient. Whether datasets arise from magnetic resonance images, DNA microarray, or epigenetic markers, the thrust of the challenge is to create a treatment plan that comprehensively characterizes a patient's cancer. The prelude to this challenge is identifying the underlying structure–hidden patterns–of these datasets, and from there understanding how to utilize these patterns to accurately infer a patient's diagnosis and prognosis in cancer. Here we explore machine learning solutions that provide a powerful and elegant framework for improved characterization of patients by assessing their cancerous lesion.

We begin with analysis of cancerous lesions identified with fluorodeoxyglucose Positron emission tomography–computed tomography (FDG-PET-CT) and Diffusion Weighted Magnetic Resonance Imaging (DWI) for predicting a lesion's treatment response and classifying a lesion's histology , respectively. We treat metastatic liver cancer lesions obtained from FDG-PET-CT as 3D shapes and infer if a patient will respond or not-respond to a radiotherapy based on 3D shape features. Breast cancer lesions obtained from DWI are used to generate distinct parameters that represent tissue microstructure to differentiate if a lesion is benign or malignant. Situated within this framework, we illustrate the value of computational and statistical information available within a lesion image for other downstream tasks. We then move onto the intersection of medical imaging and genomics with a deep latent variable model that predicts somatic mutations from

medical images. Unlike traditional computational models that focus on a specific cancer or a specific set of mutations, we created a model that scaled to incorporate image and mutation data from all possible cancerous lesion types that are publicly available. A unique property of this model is that the lesion images are modeled as point clouds instead of three-dimensional images. Our approach studies the two different yet related datasets as two distinct latent probability distributions unified by one shared distribution. The shared distribution is implicitly encouraged to create a connection between the two domains, so that we can faithfully transfer information from one domain to another. This learning paradigm allows us to predict all possible somatic mutations within a patient, thereby potentially aiding clinicians to assess effective treatment solutions for a patient during the initial diagnosis.

As predicting the occurrence of somatic mutations only represents a minutia of the complexity of cancer biology, we propose a generative probabilistic latent variable model to determine co-occurrence patterns of somatic mutations. Whereas standard learning methodology uses heuristics and frequency for modeling somatic mutations, we created a data-driven dependent prior that enables us to specify a notion of similarity on both positive and negative correlations between somatic mutations. Our results showed biological processes, total number of somatic mutations, non-linear mutation-mutation interactions, and cancer type are all latent confounders that play an important role in influencing the co-occurrence patterns of somatic mutations

Together, our research demonstrates the value of correctly characterizing a cancerous lesion to generate patterns that provide diagnostic and prognostic insights of a patient's cancer.

CHAPTER 1

INTRODUCTION

1.1 Motivation

Through years of retrospective studies in cancer, a small subset of patients with cancer are characterized by a single clinical feature. For instance, melanoma patients have a phenotypic feature of abnormal lesions on the skin (4), breast cancer patients have a BRCA2 gene mutation as a genetic feature, and lung cancer patients have haemoptysis (blood in sputum) (5; 6; 7). Most scenarios of patient assessment for cancers, however, require extensive quantitative and qualitative knowledge of a patient's clinical features that ascribe to a type of cancer (8). Optimal diagnostic accuracy is achieved by correctly identifying the cancer type from a specific combination of clinical features (9). The discovery of these combinations, however, is an ongoing and complex problem from both a biological and clinical perspective. There are millions of cancer patients, and although many patients may share the same type or subtype of cancer, the patient's response to treatment, cancer progression, and mortality can all vary. This has led to a push in *precision medicine* (10) or *personalized medicine* (11) that uses computational models to stratify patients based on a patient population's demographics, histology, radiological images, or molecular datasets.

The popularity of computational models to improve diagnostic accuracy is a natural stepping stone in scaling current clinical protocols to the patient population and a specific subset of

patients. We can formalize this further as a study of patterns. Humans innately create patterns for better perception, understanding, and interpretation of the world (12). We often rely on a distinct or a combination of patterns to identify relationships among phenomena. For instance, weather *patterns* are used to predict the chance of precipitation. Mimicking human decision patterns (13) or creating computational patterns (14) has led to the generation of computational features for applications in classification, prediction (15), and generative (16) tasks. Similarly, the increasing availability of medical data in conjunction with powerful computation tools has resulted in intricate computational features that take into account the individual patient variability.

The generation of clinical features from medical images is of particular importance as medical imaging is omnipresent at initial and follow-up stages of patient assessment to noninvasively diagnose and monitor the cancer's progression. Clinicians and radiologists have used clinical features to create diagnostic protocols (17) that allow them to assess and monitor a lesion. Unfortunately, radiologists are still prone to have a high variance in sensitivity and specificity when assessing a lesion. For example, in breast, lung, and bladder cancer screenings, radiologists achieve a 74-90% sensitivity and a 60-76% specificity (18; 19; 20). Ideally, if multiple radiologists assessed an image, these statistics would increase (21). However, that is implausible.

With increasingly powerful computational models, researchers have created tools to extract computational features from medical images referred to as medical imaging features (22). Subsequently, imaging features often represent hidden patterns of a patient–patterns that are not directly discernible like clinical features. A natural step is to then deduce which of these imaging

features can be parlayed into further optimizing patient diagnosis. Indeed, numerous studies have shown the effectiveness of medical imaging features (23) to improve lesion characterization, and thereby a radiologist's ability to diagnose cancer or assess cancer progression. Many models concerning medical imaging features are built upon established models within computer vision such as facial recognition (comparable to lesion classification) (24) and segmentation (contour of a lesion) (25). This has led to computational tools such as computer-aided diagnosis (CAD) (22) that use medical imaging features and machine learning to improve cancer screening rates by reducing interobserver variability and increasing the sensitivity and specificity rates (22).

While these computational models have achieved a few state of the art results (22), they are still prone to low specificity (high rate of false positives) due to the mischaracterization of lesion heterogeneity from the imaging modality. As imaging features are reliant on the correct representation of lesions, small perturbations can dramatically impact imaging features and lead to decreased performance in machine learning models (26). For example, consider the problem of predicting the treatment response of a patient with pre- and post-treatment scans. In such a scenario, the imaging modality would ideally reflect lesion heterogeneity as a function of the distribution of the voxels within the lesion. Therefore if a lesion is heterogeneous, then different regions of the lesion would respond differently to the treatment, which should be observed in the distribution of the parameters (i.e. intensity) of the voxels. Alternatively, emerging imaging techniques for human cancers as well as computer vision models for 3D object feature generation (27) offer promise and flexibility in characterizing lesions. Thus with lesion characterization an

important implication for correct patient assessment, we investigate medical imaging features that better capture the distribution of voxels in Chapters 3 and 4.

Just as there is a push in computational modeling of patterns within medical images, there is also a push in computational modeling of generating treatments that are patient specific. Current and established targeted treatments are often based on the genetic profile of a patient population. The targeted treatments are designed to target somatic mutations, which are selected on the basis of mutated gene's frequency in a patient population. Lesion heterogeneity, however, decreases the effectiveness of many of the targeted treatments due to the complex genetic environment of a lesion. Cancer genes and therefore somatic mutations are pleiotropic as Wagner and Altenberg state (28): a mutation can affect a number of phenotypic characteristics of a cell. We can directly observe this phenomenon in the various types of subtypes of a distinct cancer type. For instance, consensus molecular subtype-2 (CMS2) in colon cancer shares many of the same somatic mutations as in other colon subtypes, but the permutation and the frequency of a set of mutations create a divergent path during cancer progression (29). Understanding the co-occurrence patterns of somatic mutations across cancer patients between cancers and within a cancer is one of the long-standing challenges in cancer biology.

A myriad of statistical models elucidate co-occurring mutations by following three basic patterns: two mutations occur randomly by following an expected mutation rate (30); mutual exclusivity (31), when the two mutations are rarely seen in the same tumor; and co-occurrence, when the two mutations occur together more frequently than expected by chance. A statistical model may further constrain the co-occurrence patterns by integrating known cellular interactions

(32) to reflect biological processes. This constraint simplifies another assumption of mutually exclusivity: two genes that are involved in the same biological process are biologically redundant. Therefore, the mutually exclusivity assumption defines that a mutation in one of the genes is sufficient to deregulate the affected biological process. Although these assumptions are intuitively and biologically sound, they do not always hold in practice. In Pritchard et. al (33) they discuss that there is no "established approach that explicitly models the generative process of mutual exclusivity patterns." Therefore co-occurrence patterns in somatic mutations cannot only be modeled by frequency patterns and known biological interactions. An ideal statistical model would incorporate fewer assumptions and aim to disentangle the sources of variability within each a cancer type.

Computational and statistical modeling challenges for identifying confounding variables in the somatic mutation dataset can be addressed by the confluence of neural networks and probabilistic modeling. Neural networks allow for possible non-linear generative processes(non-linear correlations between confounders), while the probabilistic model allows for a more expressive model to reflect prior knowledge about the dataset. Formally, this paradigm follows generative probabilistic models that have been successfully deployed to find population structure in genetics (34), language (35), and recommendation systems (36). Generative probabilistic models contain three parts: observed data (somatic mutation profiles), hidden structure (confounders), and a probability function. In our case, this hidden structure represents a set of co-occurring mutations and the probability function translates to the influence of each set of co-occurring mutations (factors) within the patient. Generraly, each factor is often interpreted as representing a known

biological process (36). The modeling challenge of generative probabilistic models is deeply intertwined with the choice of the priors for the probability function and hidden structure, which we further discuss in Chapter 2.9. One contribution to this thesis is the creation of a prior dubbed CoZINB that reflects cancer biology. We show that CoZINB infers the latent confounders to generate distinct and diverse co-occurrence patterns in somatic mutations for different cancers in Chapter 6.

In so far we have discussed two distinct datasets (imaging and somatic mutations) separately. Current studies have tried to bridge this gap with the creation of a nascent research topic referred to as *radiogenomics* or *imaging genomics*. These studies operate the assumption that lesion heterogeneity is reflected in medical imaging features through patterns of the lesion shape and contrast. Recent studies have shown correlations and predictive performance of medical imaging features to identify genetic features such as gene expression and somatic mutations with a combination of statistical and machine learning-based approaches (37). Given the large size of the datasets, many of these studies impost restrictions on the number of genetic features. This, however, can lead to biased results that favor dominant patterns within the patient population of cancer (38), thus limiting a model's capacity to reflect a lesion's complex biological environment.

Instead of limiting a model to correlate or predict a few genetic features, we would like our model to predict the full somatic mutation profile of a patient. This problem is similar to current daily activities we perform online such as personalized shopping recommendations on Amazon or personalized movie recommendations on Netflix. To examine every movie option available is impractical, so a recommendation algorithm should reflect the user history or users with similar

search patterns. The somatic mutation profiles and the corresponding lesion images we examine in this thesis, however, have less repeated patterns due the complexity of cancer biology that influences lesion heterogeneity. The two distinct datasets are also derived from two very different domains. There is no inherent relationship between a lesion obtained from an MRI image or a somatic mutation profile obtained from a gene sequencing test. Whereas, there is an intuitive relationship between movies regardless of their genres. In such cases, instead of examining the domains separately, we jointly model the distributions of each domain. There are two benefits of joint modeling: we can robustly create a distinct but separate representation of each domains with limited data, while also uncovering relationships among the two domains.

One of our contributions to this thesis builds such a model by using geometric imaging features and the theory of probabilistic models to create a joint deep latent variable model that predicts somatic mutations. A motivation of this model is to better capture potential mutations that are targeted by treatments, which we discuss further in Chapter 5.

The combination of challenges faced in characterizing lesions and genetic profiles can greatly benefit from further model development. We propose algorithms in this thesis that exploit features for patient assessment from three different perspectives in human cancers; predicting patient diagnosis and prognosis from medical images, the integration of medical images and somatic mutations to predict patient somatic mutation profiles, and inferring the co-occurrence patterns between somatic mutations.

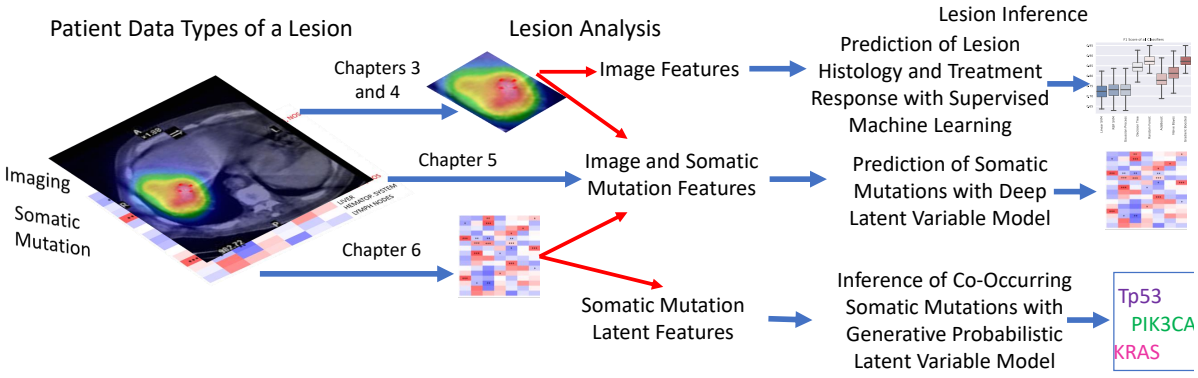


Figure 1: An overview of the dissertation for inference on cancerous lesions from three different perspectives: predicting patient diagnosis and prognosis from medical images, the integration of medical images and somatic mutations to predict patient somatic mutation profiles, and inferring the co-occurrence patterns between somatic mutations.

1.2 Contributions

1.2.1 Learning from Medical Images

A standard approach to stratify cancer patients is to grade the severity of a lesion. Clinicians often grade a lesion by identifying image patterns from various imaging modalities such as MRI or CT. For instance, the ground-glass opacification image pattern is the difference in contrast of voxels between adjacent and lesion tissue in a lung CT image. Researchers and clinicians have created a complete list of guidelines with specific imaging patterns such as in glioblastoma using MRI (VASARI Research Project (39)), breast cancer using mammograms (BI-RADS) (40), or lung cancer using CT (LUNG-RADS) (41). The goal of these guidelines is to create a standardized description for sharing patient diagnosis and prognosis between clinicians (42).

However, as discussed, the imaging features within the guidelines have low specificity due to inadequate capture of the lesion heterogeneity present in much of cancer population.

To better probe lesion heterogeneity, clinicians and researchers have implemented the imaging techniques F-fluorodeoxyglucose (FDG) Positron Emission/Computerized Tomography (FDG PET/CT) and Diffusion-Weighted Magnetic Resonance Imaging (DWI). These imaging techniques yield unique parameters from their corresponding images that reflect the tissue microstructure of a lesion such as vascular permeability, perfusion, and cellularity. Subsequently, the tissue microstucture is partly a reflection of lesion heterogeneity (43; 44). Liu et al. (45) showed that the apparent diffusion coefficient (ADC) of DWI in conjunction with a BI-RADS score from dynamic contrast MRI (DCE-MRI) for breast cancer patients significantly improved the diagnostic specificity. Hicks et al. (46) showed that FDG PET/CT in lung cancer improved staging diagnostic accuracy leading to a better assessment of treatment plans in patients. The versatility of FDG PET/CT and DWI have made these modalities appealing for clinical management of cancers (47; 48; 49).

The challenge of images from FDG/PET and DWI is how to quantitatively compare parameters extracted from the lesion images within a patient population. While the intuitive goal of comparing parameters is simple, formalizing this task is much more difficult. The standard approach follows using a single parameter such as the mean of the standard uptake value (SUV) in FDG PET/CT (50) or the mean of the ADC in DWI (51) to compare patient populations. A summary statistic of a single parameter, however, disregards the distribution of the voxels and can lead to false positives (52).

We, instead, exploit the distribution of the voxel parameters to elucidate quantitative imaging features within lesion images obtained from FDG PET-CT and DWI. Our approach is based on modeling the statistical properties shared between voxels using 3D shape features in PET/CT and statistical histograms in DWI. We show that 3D shape features are powerful in predicting treatment response of patients with metastatic liver cancer (Chapter 3). We also show that imaging features created using a statistical histogram have a robust performance of classifying breast lesion pathology (Chapter 4). These quantitative imaging features have the potential to improve precision medicine as it may expedite treatment and lead to faster decrease of cancer progression.

While, medical imaging identifies a patient's cancer type, a challenge in precision medicine is that a full treatment plan is generally not created until a biopsy of the lesion is performed (53; 54). As early treatment is crucial for decreasing cancer mortality (55), researchers and clinicians have defined imaging features of a cancer lesion's that correlate with functional, molecular, and metabolic information of a patient (56; 57; 58). These results suggest that imaging features can pose as a surrogate for biological information within a lesion.

Despite advancements of the previous studies, the correlations between imaging features and the biology of a tumor are often limited to providing information about the occurrence of specific genetic markers (38). Our approach instead posits we can exploit the structure of voxels in medical imaging to predict the entire somatic mutation profile of patients. There is another set of challenges that arise. First, cancer imaging datasets are often small, thereby limiting statistical power of conclusions. Second, selecting the appropriate combination of imaging features to

predict somatic mutations is not trivial. Third, scaling a small multi-class problem to 20,000 different classes (where each somatic mutation is a distinct class). The question now becomes how to select the correct subset of classes from an exponential search space. Indeed, as previous studies have done, one can limit the number of somatic mutations predicted, but this assumption can lead to false positives in downstream tasks such as treatment planning (38). As the cancer genome is still incomplete (59), such an assumption is inappropriate for capturing the possible dependent relationships that influence cancer biology.

In this thesis we develop **L**esion Point C**lo**ud to **S**omatic Mu**t**ations LLOST, a multidomain model in Chapter 5 that exploits the hidden structure of somatic mutations and medical imaging to predict somatic mutation profiles. LLOST represents lesions as point clouds instead of images to enable invariance to imaging techniques. The two different domains are jointly modeled with the utilization of variational autoencoders and invertible neural networks (60; 61). This modeling choice encourages the patients with the same cancer type to share similar patterns among the general patient population of multiple cancers. The prediction of somatic mutation profiles, is potentially useful for clinicians to guide them in creating a patient specific treatment plan.

1.2.2 Learning From Variations of Somatic Mutation Profiles

In our discussion of somatic mutations, we stated that current methodology surmises that cancer is mostly driven by a few highly frequently occurring somatic mutations and the remaining somatic mutations are passenger mutations. While, driver mutations may directly promote cancer angiogenesis, their co-occurrence patterns with passenger mutations have indirect effects that insofar have not been studied. Therefore, studying the complex patterns of somatic mutations

requires understanding the processes that influence how mutations are positively and negatively correlated with each other. We address these challenges in Chapter 6 by introducing a prior called the Correlated Zero Inflated Negative Binomial Process (CoZINB) in a generative probabilistic model. This specific choice of prior allows the dynamic creation of diverse sets of co-occurring somatic mutations that reflects cancer biology. We show that cancer type, the total number of mutations, biological processes, and mutational processes are all confounding variables that influence co-occurrence and mutual exclusivity patterns in somatic mutations.

1.3 Dissertation Overview

We begin by reviewing many of the imaging and statistical concepts that are utilized throughout this thesis. The second chapter starts with an overview of the imaging modalities used in 3 and 4 and current image feature generation techniques for two and three-dimensional images. This will serve as a backbone to the reasoning of why we chose specific imaging features for our tasks in Chapters 3 and 4. In Chapter 2, we also describe methods in generative probabilistic models and their use in modeling sparse high-dimensional data by presenting latent variables models that inspire the work in Chapters 5 and 6. Together, these concepts also enable examination of the importance of prior distributions (the distribution of the unknown parameters) and the posterior distribution (the updated parameters given the data), which led us to develop a unique prior distribution in Chapter 6. We conclude the chapter with Bayesian inference techniques, specifically amortized inference and variational inference.

An overview of the different perspectives we use to analyze cancerous lesions in this dissertation is shown in Figure Figure 1.

In Chapters 3 and 4, we present quantitative imaging features for use in FDG PET/CT and DWI, respectively. Chapter 3 demonstrates the use of quantitative imaging features as a tool for predicting treatment response in patients with metastatic liver cancer. This work is presented in (1). Chapter 4 demonstrates the use of quantitative imaging features as a tool to discriminate lesion pathology in breast lesions. This work is in the submission process and is presented in (2).

In Chapter 5, we address the limitation of the traditional paradigm of learning from medical images by optimizing the combination of quantitative features that predict all somatic mutations. We develop LLOST, a deep latent variable model built on dual Variational Autoencoders that consists of three latent spaces: two specific to the domain and a shared latent space. This model demonstrates the prediction of somatic mutation profiles from their corresponding images using publicly available data from The Cancer Genomic Archive and The Cancer Imaging Archive. Chapter 5 is presented in (3).

In Chapter 6, we present the Correlated Zero Inflated Negative Binomial Process (CoZINB) based on (62). The CoZINB model consists of parameters that reflect the total number of mutations, biological processes, and mutational processes, while also incorporating non-linear interaction between the parameters using neural networks. We use CoZINB in a probabilistic generative model to characterize the co-occurrence patterns of somatic mutations that better reflects cancer biology.

Lastly, in Chapter 7, we summarize our discussion and discuss future directions of our work.

CHAPTER 2

BACKGROUND

2.1 Medical Imaging

2.1.1 Positron Emission Tomography - Computed Tomography

Positron emission tomography (PET) is a non-invasive medical imaging technique based on the detection of positron emissions. The principle of PET is detecting 511 keV of energy after the collision of positrons with electrons present in neighboring tissue. The captured energy is reconstructed into an image that provides functional as well as metabolic information related to the tumor cells. The anatomical information provided by PET, however is low, so imaging systems have integrated PET with computed tomography (CT) or magnetic resonance imaging (MRI) to better assess cancer tissue.

A common radiotracer used to create the positron collisions is Fluorine 18 fluorodeoxyglucose (FDG). The tracer is transported into cells using the same transport enzymes as glucose and becomes trapped within the cells as it is not dephosphorylated due to the absence of an enzyme in cancer tissue. The uptake of FDG then represents the glyolic activity of tumor tissue, which is influenced by lesion histology, lesion aggressiveness, viability of malignant cells, the presence of

hypoxia, and local vascularisation. The conventional method to quantify FDG uptake in lesion tissue is through the standardized uptake value (SUV) defined as:

$$SUV = \frac{tissueactivity/ml}{administrateddoseofFDG} \times bodyweight$$

. Benign lesions have a declining uptake of over time, whereas malignant lesions have an increase in FDG uptake over time. The differences between uptake value of benign and malignant tissue is due to the underlying cellular structure (63). This property was exploited by Cottereau et al. (64) to correlate a high SUV with the over-expression of MYC and BCL2 genes that promote cellular proliferation, thereby affecting the cellular structure of the tissue.

The efficacy of FDG PET-CT is best observed in assessing the response of a tumor after therapy. The primary goal of non-surgical cancer therapy is the eradication of tumor cells, however, the complexity of tumor biology results in a variety of different responses. The standard diagnostic tool for measuring treatment response is decrease in tumor size (65) in imaging modalities such as MRI and CT, however residual masses may exist and can lead to recurrence of the cancer. In some cancers, decrease tumor size is also not indicative of treatment response (66). As FDG PET-CT is sensitive to the functional characteristics of tissue, the information derived from the SUV can be used to manage and modify treatment options. Martini N et. al used (67) FDG PET-CT to monitor therapy response in non-small cell lung cancer to determine the best time point for possible surgical intervention. Mikhaeel et al. (68) showed FDG PET-CT was more accurate in comparison to CT for assessing remission in patients after treatment of Non-

Hodgkin's lymphoma. The measurement of glycolic uptake has made FDG-PET an appealing tool to evaluate lesion heterogeneity and has created new possibilities to assess treatment options in patients.

2.1.2 Diffusion Weighted Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is an imaging modality based on a patient being placed in a magnetic field through a magnet that surrounds a gradient coil, which further surrounds a radiofrequency (RF) coil. The protons in the body absorb and emit back resonance frequencies which are used to reconstruct and create 3-dimensional images of a region of interest. Contrast between tissue within the body is created as protons of different tissues return to an equilibrium state at varying rates. These rates are identified as the longitudinal relaxation time T1 and the transverse relaxation time T2. By manipulating scan parameters such as the sequence repetition time (TR) or the echo time (TE), the MRI can produce different levels of contrast referred to as T1-weighted MRI or T2-weighted MRI. Although invasive, patients can also be injected with gadolinium chelate to create contrast between tumor and normal tissue using dynamic contrast enhanced MRI (DCE-MRI) (69) to facilitate functional information of a lesion.

The T1 and T2 weighted images are often used for morphological information of a tissue. For example, fat has a shorter T2 than water and relaxes more readily than water so T2 weighted images correspond to lower intensity values of fat in comparison to T1 images. Since benign tissue often have higher levels of water, T2-weighted images correlated with benign morphology in breast cancers, especially cysts (70). Conversely T1-weighted images of a patient's brain can show tumor enhancement due to abnormally high-intensity areas with the increased fat level.

Upadhyay et al. (71) used a combination of T1 and T2 weighted images to help grade glioma, since T2 weighted images identify necrotic tissue, while T1 weighted images identified the fatty tissue.

To better improve sensitivity rates, clinicians have used DCE-MRI separately or conjointly with T1-weighted images for better lesion characterization. Similar to FDG PET/CT, cancerous tissue has a faster uptake of the contrast agent compared to normal tissue (72). This features reflect angiogenesis of the tumor and as the tumor microcirculation causes increased interstitial pressure to relieve the contrast back into the tissue (72). Hara et al. (73) showed that when DCE-MRI is used conjointly with T2 weight images the sensitivity of prostate cancer detection increased to 93%. While the increased performance of DCE-MRI has increased detection rates, the high variance of specificity in DCE-MRI is of a concern as it leads to false positives. This is because the rate of contrast uptake does accurately reflect the properties within the local tumor tissue (74). Lee et al. (75) estimated that for 65-year old women, around 7% of all invasive screen-detected cases will be overdiagnosed, whereas for 75-year old women around 13% will be overdiagnosed. This new understanding strongly implies the tissue microenvironment and therefore lesion heterogeneity is an importance factor to understand the underlying mechanism of cancer and to create effective treatment strategies (76).

MRI also has the ability to probe tissue microstructures by investigating the diffusion behaviour of water molecules within the tissue. The diffusion weighting signal is represented with a simplified mono-exponential model introduced by LeBihan et al. as

$$\frac{M}{M_0} = e^{-bD}$$

(77). The parameter b is known as the b-value or the diffusion-weighting strength and D is the diffusion coefficient. The b-value determines the degree of diffusion-weighting, where a large b-values generally indicates higher signal intensity due to restricted diffusion within the tissue. This property creates a contrast in images referred to as Diffusion Weighted Images (DWI) based on the unique diffusion properties of the microenviornment within tissue caused by macromolecules, fibers, and cells. The goal of DWI is appealing as the correct quantification of the tumor microenviornment would allow for clinicans to grade lesions at the voxel level, thus avoiding invasive biopsies, radiation as in FDG PET/CT, and contrast agents as in DCE-MRI.

In cancer imaging, DWI has been shown to be sensitive to cell density (cellularity) through its restriction of diffusion both in the intra- and extra-cellular spaces, which is important in analyzing the aggressiveness of the tumor as well as in treatment response (78). The most common quantitative marker to measure diffusion is known as the apparent diffusion coefficient (ADC) calculated using a mono-exponential function with two different b-values,

$$ADC = \frac{\ln \frac{M_2}{M_1}}{b_1 - b_2}$$

where M_1 and M_2 are the b-values at different signals of b_1 and b_2 respectively. Guo et al. (79) showed there is a significant statistical difference of the ADC value between malignant and benign breast lesions and recommended that ADC is a completementary tool for DCE-MRI to help decrease. Muraoka et al. (80) demonstrated a statistically significant lower mean ADC value in cancerous tissue of the pancreas, which was a result of collagenous fibers within the tissue.

Despite the various diffusion tissue properties that DWI probes, the information provided by the ADC is limited to probing the density of cells in a tissue (81). Advanced methods have been developed to quantify and assess the tissue microenvironment such as the intravoxel incoherent motion (IVIM) (82) and continuous time random-walk (CTRW) (83). The IVIM method fits a biexponential model as a function of increasing b-values to separate cellular and vascular components with three distinct parameters

$$\frac{S}{S_0} = fe^{((-b(D_{diff}+D_{perf})))} + (1-f)e^{(-bD_{diff})}$$

. The perfusion parameter (D_p) and microvascular volume fraction f probe the vascular components, and the diffusion parameter D_{diff} probes the diffusion within the tissue. With an optimal fitting, IVIM parameters can correlate with the tissue microstructure such as increased blood flow, which in turn can indicate aspects of tumor aggressiveness or malignancy. IVIM has gained traction in characterizing lesions with the growing use of DWI for predicting treatment response (84), grading lesion tissue (85), and differentiating between malignant and benign

tissue (86). CTRW fits a non-Gaussian multi-parametric function at increasing b-values to also create three parameters that represent the temporal (α) and spatial (β) cellular structure within a voxel and the anomalous diffusion coefficient (D_m) that is analogous to the ADC. What is beneficial about IVIM and CTRW models is that we can correlate the different parameters with micro-vascular and micro-structural properties of the tissue that in turn aid in assessing lesion malignancy. In Chapter 4 focus we examination of combination of the parameters from IVIM and CTRW models that optimally characterize a breast cancer lesion for differentiating a benign or malignant.

2.2 Lesion Heterogeneity

Lesion heterogeneity is a hallmark of cancer that has confounded researchers and clinicians in understanding the complexities of cancer. While genetic heterogeneity can refer to a diverse patient population, lesion heterogeneity reflects the cellular and molecular differences that define a cancer type as shown in Figure Figure 2. In his overview of cancer, Foulds (87) demonstrates that tumor angiogensis was not a simple linear process, but an evolution of cellular fitness through positive and negative selection. This is analogous to population genetics, where the study of the evolution of genes is measured through generations of positive and negative selection, cancer evolution instead is on a much smaller time scale. Heppler (88) expands on Foulds discussion, articulating that cells within a tumor that confer advantage are dependant on the microenvironment created by the organ or of regional differences in oxygen supply, acidity, nutrition supply, and to the bane of clinicians the therapy itself. Other non-genetic causes lesion heterogeneity include epigenetic changes (89), mutational processes (90), and differentiation

of stem cells (91). Therefore, lesion heterogeneity describes the existence of distinct cellular populations with specific phenotypic features.

Subclonal evolution within the cancer cells allows for multiple subtypes of a distinct cancer, and at a molecular level it is impossible for two patients to have the same genetic mutations that result in the cancer. The mechanisms that cause subclonal alterations are still actively researched, but researchers and clinicians have determined there is a need to investigate both synergistic and antagonistic interactions between subclones during tumour evolution. The motivating factor is that subclonal cell population often have a different genetic makeup and is the main reason why systemic treatments (chemotherapy, radiation, and surgery) for cancer are still the standard protocol as targeted treatment tends to confer resistances or leaves residual tissue that leads to cancer recurrence. Even systemic treatments fail due to selective pressure on cells that are resistant systemic therapy and potentially drive disease progression (92). An example of positive selection is observed in patients with EGFR mutated lung adenocarcinoma. These patients acquire resistance to due to positive selection of cells with the co-occurring mutation T790M in the EGFR gene (92) as the targeted treatment tyrosine kinase inhibitor (TKI) gefitinib targets exons 18 and 19 of the EGFR gene. Similarly, Sasaki et al. (93) showed patients treated with panitumumab for EGFR coloretcal tumors found an emergence of KRAS mutations during the course of therapy resulting in acquired resistance. Sasaki et al. showed that there were divergent evolving tumor cells that harborded KRAS mutations before treatment, and that under the selection pressure of anti-EGFR therapy, the cells with KRAS mutations rapidly expanded to cause tumor progression. Therefore, cancer recurrence and progression can be attributed to a

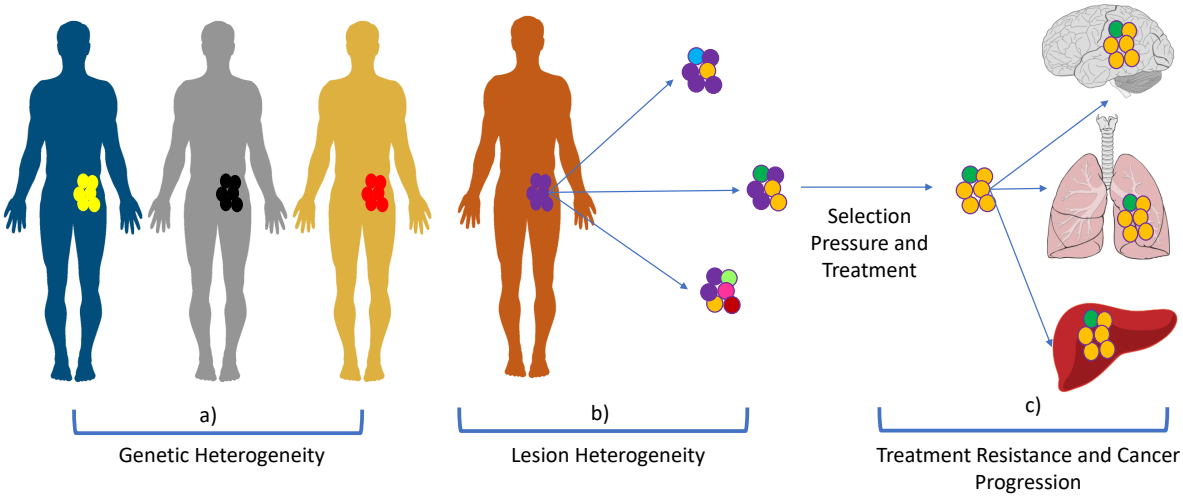


Figure 2: a) Patients rarely share the same genetic makeup a lesion making cancer specific treatments more difficult to create b) Lesions are also heterogenous, where different cells have different genetic profiles c) If a treatment only targets a specific cellular profile, selective pressure will increase for cells that are resistant to treatments, which can lead to cancer progression

small population of tumor cells with different mutations primary tumour lends and as a result causes the development of multiple mechanisms of resistance in the same patient under selection pressures from targeted therapies.

Since it is not yet feasible to genetically profile an entire tumor, researchers have relied on medical imaging or tumor mutation load (TML) to quantify tumor heterogeneity. TML is a measurement of the total number of mutations within a tumor and has been studies have suggested it as a potentially biomarker for immunotherapy such as Goodman et al. (94) who showed a higher TML can partially represent tumor heterogenity due to the overall characteristics of genetic environment of a tumor. Similarly, a multi-region sequencing study of Caucasian lung adenocarcinoma patients by Hanjani et al. (95) indicated high smoking-associated mutational

burden for clonal mutations correlated with low intratumoral heterogeneity. Unfortunately, TML is only one of the causal links to tumor heterogeneity and there is no established genetic marker that can quantify tumor heterogeneity.

Medical imaging can evaluate lesion heterogeneity by exploiting the properties of the imaging phenotype that reflect the inner organization of the tumor (96). Specifically, we assume that lesion heterogeneity is a confounder that influences the the tissue microstructure (structural heterogeneity) of a lesion captured by medical imaging (97). This structural heterogeneity is probed using metrics based on texture features, quantitative markers such as SUV in FDG PET/CT or parameter values in DWI. The texture and SUV metrics evaluate the spatial dissemination of cells within a tissue. Kwon et al. (98) created a metric to evaluate structural heterogeneity in oral cavity cancers as the derivative over the volume of a tumor as a function SUV that significantly correlated with patient prognosis. The parameters of DWI that probe cellularity in ADC or vascularity in IVIM can also correlate with structural heterogeneity in a lesion. For instance, Iima et al. (99) showed that the difference of the ADC at different b-values is higher in malignant areas of the lesion than benign. This potentially indicates the presence of cellular proliferation or cell apoptosis (81). In the context of lesion heterogneity, this can potentially correlate with the over-expression of oncogenes (100). While a one-to-one correspondence between a tumor's image voxel and a tumor's cell is currently unavailable, the analysis of tumors with medical imaging provides a non-invasive tool capturing the genetic landscape of a lesion as a function of the tissue microstructure (101).

2.3 Metastatic Cancer in the Liver

Cholangiocarcinoma (CCA) and Colorectal cancer (CRC) are the second and third most primary hepatic (liver) malignancies worldwide respectively (102). Tumors that arise due to CCA are associated with the "ductular epithelium of the biliary tree, either within the liver (intrahepatic cholangiocarcinoma) or more commonly from the extrahepatic bile ducts (extrahepatic cholangiocarcinoma)" (103). Malignant tumors in the liver from CRC, unlike other metastatic tumors (metastatic breast or lung cancers) arise due to the propensity of CRC tumors to spread to the liver. There is a reported rate of 70% of all patients with CRC develop some form of liver metastases (102). Patients who are suspected to have the cancers are diagnosed using computed tomography (CT) or the conjunction of CT and fluorine 18 fluorodeoxyglucose positron emitted tomography (FDG-PET). A prognosis of the patient is then made by identifying the number of tumors, satellite lesions, age, and primary tumor stage (104).

Surgical resection is the most effective treatment approach for both metastatic CCA and CRC. Unfortunately only a minority of patients (15%) are suitable for surgery (102) due to limited ability to preserve minimum functional liver function after surgery. Patients who could not have surgery to remove the tumor received systemic chemotherapy where the median survival time ranged from 12 to 20 months and 5-year survival rates of less than 5% (102). A greater survival rate is achieved through radiotherapy such as Yttrium-90 (Y90) microspheres that carries radiation to the tumor, thus preserving normal liver tissue. Current retrospective studies have reported 5 year survival rates to nearly 58% after resection of the liver once radiotherapy is completed due to decreased tumor size and/or number of tumors (105).

The management of the disease is monitored through CT or FDG PET/CT. The latter has shown to have superior performance in identifying residual tumors after treatment. Specifically, FDG PET/CT allows a clinician to monitor the tumor metabolism by measuring the glucose uptake (66) to identify local tumor ablation, which is not observed through CT. As early and subsequent treatment is key in increasing the survival rate of CCA or CRC, FDG PET/CT is a significant tool in assessing patient prognosis.

2.4 Breast Cancer

Breast cancer is a malignant tumor originated in the cells of the ductal tissue and is the second leading cause of cancer death in women (106). The average lifetime risk of developing breast cancer is 1 in 8 (106) and the associated 5-year breast cancer survival rate is 98% for local tumor, 84% for regional disease and 23% for metastatic tumors (106). The increased survival rate is attributed to early detection of breast cancer and is crucial for management of the disease (106). Most clinicians recommend annual self-breast examinations and annual bilateral screening of the breasts after the age of 40 (106). Unless a patient is at high-risk breast cancer susceptible due to a family of breast cancer and/or mutations in the BRCA1/2 genes, screening is done using mammograms. For screening at high-risk patients or once a lesion is established, the gold standard for prognosis is dynamic contrast enhanced MRI (DCE-MRI) that uses an injection of gadolinium chelate to create contrast between tumor and normal tissue.

The contrast enhancement patterns of DCE-MRI are related to tumor angiogenesis as Szabo et al., Mussuarkis et al., and Bone et al. found a significant correlation between tumor grade and rim-like enhancement patterns (107; 108; 109). From a machine learning perspective,

imaging features from breast lesion DCE-MRI images have been used to predict molecular subtype (110) and predict treatment response (111; 112). While these studies have achieved high accuracy performances, they are prone to false positives due to mischaracterization of tumor heterogeneity as discussed in Section 2.1.2. Moreover, like many cancers, breast cancer comprises of heterogeneous tumors with different characteristics that can lead to a variety of different treatment options. As early detection decreases mortality rate, better understanding of breast cancer lesions has opened up an avenue for DWI to better evaluate the tissue microenvironment.

Research studies have mainly evaluated the potential use of ADC and IVIM to differentiate between malignant and benign breast lesions using machine learning (113; 114) or statistical tests (115). The CTRW model has not yet been explored in breast lesions, but offer a promising approach to characterize lesions as it improved the diagnostic accuracy of differentiating low-grade and high-grade pediatric brain tumors (116). We therefore analyze breast cancer lesions using both IVIM and CTRW models as they probe different properties of the tissue microstructure.

2.5 Somatic Mutations

The mutations in cancer occur at vastly different scales, from a single base in a genome to entire chromosome arms, and are broadly classified as single nucleotide variants and small insertions and deletions (SNVs), copy number aberrations (CNAs), and structural variants (SVs).

At the smallest scale of mutations in cancer are single nucleotide variants (SNVs) where a single base or sequence is changed in the DNA sequence by insertion or deletion of a single or series of bases. If these mutations occur in the coding region of the genome, they can be further classified as synonymous or nonsynonymous. A synonymous mutation is one that does not change

the protein sequence encoded by the mutated region of DNA, while a nonsynonymous mutation does cause a protein sequence change. Nonsynonymous mutations are further classified by the type of protein sequence change they cause. Common classifications include missense mutations, which cause a protein sequence change and nonsense mutations, which insert a premature stop codon into the protein sequence. Both types of proteins result in changing the biological function of the protein (117).

Another type of mutations are copy number aberrations, which are a special type of structural variants. The copy number of a region of DNA in a healthy human cell is two, because the human genome is diploid (has two copies of each chromosome). Copy number aberrations (CNAs) occur when a region of DNA is either amplified (copied) or deleted, thus changing the copy number. Copy number aberrations are also often called using next-generation sequencing technology, although SNP arrays were popular until relatively recently as well. Because CNAs can span multiple genes, some of these methods perform target selection to identify the gene that is the most likely target of CNAs in a given region (118).

For the purposes of this thesis, we consider all somatic mutations on a gene equal and only consider the number of times a gene was mutated as an input to our methods in Chapters 5 and 6.

2.6 Machine Learning

This section introduces some basic concepts about machine learning. A machine learning model is classified as supervised, semi-supervised, reinforcement, or unsupervised learning

depending on the application. For this thesis we focus on supervised and unsupervised learning methods that were used in this dissertation.

Supervised learning consists of a training dataset with annotated data (segmentation labels, lesion histology) and the data's corresponding features. A common way to represent a training data is using two matrices, $\mathcal{D}_{train} = \{X_{NxM}, Y_{NxL}\}_N$, where X is the feature matrix with M features and Y is the target matrix with L targets (annotated data), respectively, and N is the number of samples. A parameterized function $f_\theta(X)$ maps X to the target Y using a learning objective. The learn objective is framed as a convex optimization problem to find the best parameter or parameters θ^* that minimizes a loss function $L(Y, Y_n^*)$ where Y_n^* is the predicted continuous or binary values of the nth sample. This loss function is visualized as decision boundaries between the possible targets, or more formally:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} L\left(f_\theta(X^{(n)}), Y^{(n)}\right). \tag{2.1}$$

For classification tasks, the target is denoted by a binary variable for classifying two classes or an one-hot vector $t \in \mathbb{R}^d$ over d predefined classes, where the lth column of Y_n is 1 if the target is the class-t and all the other columns are zero. For a regression task, such as linear regression the target is often a continuous value $y \in \mathbb{R}$. There are different distance metrics available as a loss function such as the most commonly used *squared L2* distance:

$$\text{squared}_{L_2}(Y^*, Y) = ||Y^* - Y||_2^2 = \sum_{n=1}^{N} (Y_n^* - Y_n)^2. \tag{2.2}$$

There are a number of frameworks for supervised learning such as Support Vector Machines (SVM), Gradient Boosted Learning (GB), and deep learning, which at present is one of the most popular frameworks. Each framework has an advantage or disadvantage and relies on the availability of training data as well as the application of interest.

Unsupervised machine learning models do not rely on the target variables and instead create decision boundaries based on the observational data itself X_{NXM}. The canonical example is clustering that categorizes observational data into groups based on a distance metric (another convex optimization problem). The popular clustering algorithm, k-means (119), partitions N samples of a dataset into a pre-defined k number of clusters, where each sample belongs to one of the k clusters. The distance metric is based on the center of cluster called centroids (initialized randomly) and each centroid is updated over a number of iterations based on the mean of all samples assigned to cluster k. For example, a researcher could use k-means clustering differentiate benign and malignant lesions based on the size of the lesions. Unsupervised learning is especially appealing in medical datasets, because target variables such as annotations of a lesion's histology are difficult to obtain.

2.7 Medical Imaging Features

Beside clinical imaging features such as volume, geometry, or ADC and SUV discussed previously, clinicians and researchers have utilized computational techniques to create unique imaging features. These features are based upon the foundation of classifying faces (120), which uses the difference in texture to differentiate between faces. Texture in this application refers to the evaluation of gray-level intensity and the position of voxels within a medical image.

2.7.1 First and Second Order Statistics

First-order statistics are defined as the distribution of the voxel intensity values in an image disregarding any spatial relationships between voxels. These statistics can be captured by using a histogram dividing the voxel intensities in an image into equally spaced bins and computing the proportion of voxels within each bin. From this histogram a practitioner can calculate the mean and variance statistics to describe the distribution of the voxels. Dhawan et al. (121) used a first order histogram features to quantify microcalcifications in breast cancer for differentiating between malignant and benign lesions. Likewise similar studies were performed for lesions from lung cancer using CT (122), liver cancer using CT (123), and glioblastoma using DCE-MRI (124).

A weakness of first-order histogram features is that they ignore the spatial distribution of voxels. As discussed before, lesion heterogeneity may only occur in certain areas of the lesion. Haralick et al. (125) improves upon first-order statistics by constructing a second-order histogram called the gray-level co-occurrence matrix (GLCM) that characterize both the spatial relationships and intensities between voxels in image regions. The GLCM matrix is constructed by calculating the summation of voxels consisting of various intensity combinations in an image over a defined window. The property of the window controls the neighborhood distance and connectivity, including the directionality of connectivity such that different GLCM matrices can be constructed to describe a certain image. There are several texture properties that can be extracted from the GLCM and is one of the reasons why it has become one of most popular tools for creating imaging features in medical imaging (126). Features that describe lesion

heterogeneity include change in entropy, homogeneity, and the maximum correlation coefficient. For example, Kim et al. showed lesions with higher entropy from MRI had a higher rates of mortality (127).

2.7.2 Automated Feature Generation

With the increasing availability of data, researchers are moving towards computational methods that do not rely on pre-defined mathematical models to create medical imaging features (128). A subsequent benefit of such models is increased reproducibility of imaging features at the cost of interpretability. So while one may define entropy from the GLCM as the randomness of neighboring pixels, there is no established definitions for automatic generated features. The most common technique is to use convolution neural networks (CNN) in a deep learning framework, often for prediction of a lesion as malignant or benign. Although CNN are not as interpertable, the imaging features they create are not dissimilar to texture features created by the GLCM (129). Specifically, the hypothesis is that CNNs combine low-level features (entropy) to increasingly complex shapes (geometry of lesion) until the lesion is correctly classified as benign on or malignant. The number of CNN layers in a deep learning framework represents the abstraction of the lesion, where LeCun et al. (130) describe that intermediate CNN layers recognise “parts of familiar objects, and subsequent layers [...] detect objects as combinations of these parts." This has led to an explosion of deep learning tools medical imaging in cancer where McKinney et al. (131) showed a deep learning framework in comparison with radiologists had a statistically significant increase in performance of diagnosing breast cancer in mammographs.

2.8 Three Dimensional Object Features

The representation of a three dimensional (3D) object with geometrical features is a heavily researched area in computer vision (132). Applications of this research has included pedestrian detection (133), pose estimation (134), 3D object classification (132), assisted driving (135), and 3D object retrieval (132). 3D regions of interest (ROI) obtained from MRI or CT scans are analogous to 3D objects. This similarity allow us to to examine the unique geometrical patterns of a 3D ROI from a perspective of 3D object classification.

Methods used to classify 3D objects and shapes are: feature-based methods, graph-based methods, and view-based methods. For this dissertation we focus on feature-based methods where the goal is to identify common salient features between lesions. Early work on 3D feature generation focused on global features and global feature distribution such as the distribution of volume and second-order moments (analogous to 2nd order texture features). Sadjadi et al. (136) calculated the moments of a 3D object to create an invariant feature descriptor immune to change in rotation of an object for face detection. Osada et al. created a histogram which is a probability distribution that measures the difference of the 3D object to a standard shapes such as a cube, sphere, or cylinders (137) for object detection. The global feature methods, however, have difficulty discriminating between 3D objects are of a similar shape (box versus a car) resulting in research towards local 3D feature methods.

Local features represent the 3D object as a mesh or a point cloud. Local feature methods then randomly select points on the surface of an object to create 3D shape feature descriptors. Shilane et al. (138) calculated spherical harmonic moments of randomly sampled points at four different

scales for classifying natural objects such as planes. Other local feature descriptors include the Heat Kernel Signature (HKS) (139) that uses a random walk to model the distribution of points on a surface to calculate correspondences between similarly shaped 3D objects. The common theme of these local 3D features is that they encode certain statistical properties of points, while being invariant to certain transformations. Unfortunately, as with texture features discussed in Chapter 2.7 it is not trivial to find the optimal feature combination for the application at hand.

With a large enough dataset, deep learning circumvents the feature selection problem by creating neural networks specific for point clouds or meshes. PointNet (140) is the first work that introduced an architecture to process a point cloud. Many variants modified the architecture of PointNet to to exploit spatial structure (141) or within a community of points (142). These methods remained computational efficient while having superior performance in classification tasks. This is because the architecture of these methods includes both local and global shape features that can discriminate between different regions of a point cloud (140).

2.9 Latent Variable Models

2.9.1 Overview

Latent variable models are a Bayesian unsupervised learning technique that creates structure in observational data through a practitioner's prior beliefs. The standard latent variable model seen in Figure Figure 3 consists of the observed data $x = (x_1, x_2, ..., x_n)$, the hidden structure (latent variable) z, the global hidden variables (latent variable) θ, and a joint probability distribution, $P(x) = \sum_z P(x, z, \theta)$, over the hidden variables and data using the law of total probability. By examining the model, we can observe how the data is generated through the

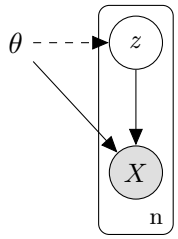
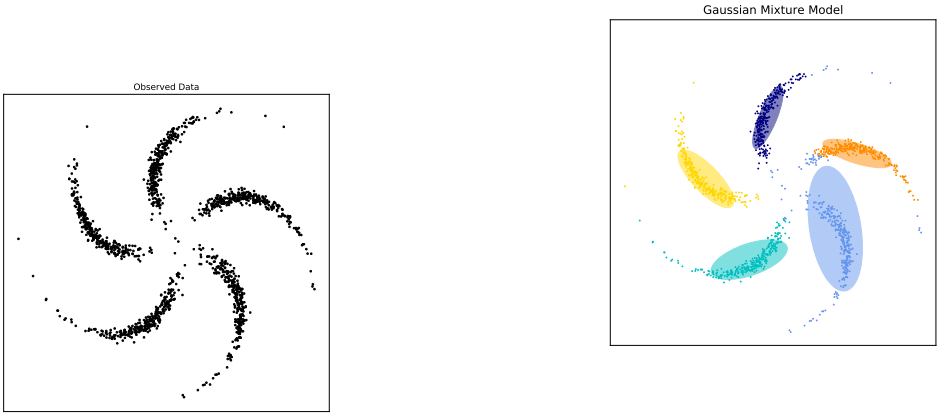


Figure 3: A generic construction of the latent variable model

factorization of the joint distribution with the product rule such that $P(x, z, \theta) = P(x|z, \theta)P(z, \theta)$. A practitioner's prior beliefs are reflected in how $P(z, \theta)$ is specified (i.e. the *generative* process) and is the foundation of many type of latent variable models not limited to probabilistic matrix factorization (143) , mixed-membership models (latent Dirichlet allocation (35)), mixture models (Gaussian mixture models (144)), and latent feature models (Indian buffet process (145)).

To further contextualize latent variable models, we examine the Gaussian mixture model (GMM) seen in Figure 4b. In the Gaussian mixture model each sample potentially comes from one of K latent classes which is explicitly encoded in the latent variable z_n that takes on of the K possible values. The factorization of the joint probability leads to $P(x_n, \theta_{z_n}) = \prod_{i=1}^{n} P(x_n|z_n, \theta)P(z_n|\theta)$ that describes the probability of the prior, $P(z)$, of selecting a class K and $P(x_i|z_i = k, \theta)$ denotes the likelihood of the nth sample belongs to one of the Gaussian distributions with the parameter θ i.e. Gaussian$(x_n; \mu_k, \sigma_k)$. What makes latent variable models more appealing than other unsupervised algorithms like k-means is that the latent variable corresponds to an intuitive assumption of how the observed data was generated or as the name



(a) (b)

Figure 4: Gaussian Mixture Models. a) Observed sample data with five different classes generated from a mixture of Gaussian distributions b) Using a Gaussian mixture model, a practitioner can fit the data with a potential set of class memberships that correspond to Gaussian distributions.

ascribed to them suggest the generative process of the underlying data. For the GMM the generative process is:

While a GMM is a simple and powerful model, the discrete prior limits a sample to belong to only one of the K possible classes. By relaxing this constraint such that a sample belongs to multiple classes while the the probability of the class assignments of a sample still sums to one is referred to as a mixed-membership model (35; 34). For example, from a mutation can belong to a number of biological process. This feature of mixed-membership models has attracted many studies to decompose biological datasets given that a biological process involves multiple actors

Algorithm 1 GMM Generative Process

for Each sample $j : 1, ..., J$ **do**
  Sample a discrete latent variable allocation$z_i \sim discrete_K(\pi)$
  Sample data point given the latent variable$z_i, x \sim P(x_i|\theta_{z_i})$
end for

in a coordinated formation. There are many subtle differences between these various models, but they all estimate the probability of a biological process being influenced by an biological such as a gene or protein as reviewed in Allison et al. (146).

2.9.2 Decomposing Matrices with Latent Variable Models

From another perspective latent variable models are similar to matrix decomposition or dimensionality reduction models that aim to reduce the size of the dataset X with n samples and D features . Features can be anything from parameters of the dataset such as the mean or variance, or the expression values of genes from a microarray dataset. One of the most common techniques for matrix decomposition is principal component analysis (147) that decomposes a data matrix into a latent representation using the singular value decomposition (SVD) $X = \Phi_\theta(Z)$. Analogous to latent variable models, Z (principal components) is the latent variable that describes the effect of a class k on sample n through a mapping function Φ with parameters θ. Since PCA cannot model prior beliefs about the dataset, Tipping et al. (148) extended PCA by allowing a practitioner to choose a prior on the latent variables. Therefore, each latent variable can now be interpreted as some type of generative process that influences the data. Using a Gaussian framework, we describe the generative process as:

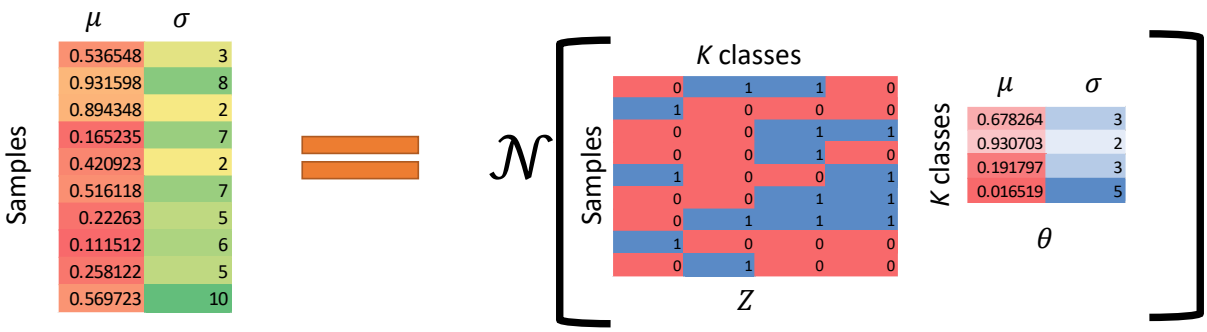


Figure 5: Hypothetical matrix decomposition of a dataset where the features are the mean, μ and variance σ of a population are decomposed in a binary latent matrix $\mathbf{Z}$ of dimensions $N \times K$. Then each class k has parameters of a Gaussian distribution depicted by the feature matrix θ with dimensions $K \times D$

$$P(z_{n,k}) \sim discrete(\pi_k) x_{n,d} \sim \mathcal{N}(\sum_{k=1}^{K} z_{n,k}, \theta_{k,d}) \tag{2.3}$$

If every column in θ is the mean, μ_k, of the $k$the class, then Z is a binary matrix where every column indicates if that sample z_n belongs to class k as seen in Figure Figure 5. There are a number of probabilistic matrix decompositions to infer greater complexity within the dataset such as the Variational Autoencoder (VAE) (60) that is used in this thesis. In the standard VAE the prior on the latent variable z is a normal distribution and the parameters, θ, are inferred using neural networks. Recent extensions to the VAE prior distributions have created more expressive models for exploring the hidden structure of gene expression (149) and protein structure (150).

2.9.3 Latent Feature Models

Mixture models and mixed-membership models are excellent models across a wide variety of data, but in some cases they are still too restrictive. The latent representation created by mixture

models implies that the total probability mass of a class assignment of a sample must sum to one. However, in a case of modeling patients' possible diseases, they can belong equivalently to having both cancer and diabetes. Latent feature models (145) circumvent this issue by creating a vector that indicates if a sample belongs to a specific class. In our example of a patient's disease, $x_n = [1, 1]$ corresponds to patient x_n having both cancer and diabetes. Latent feature models generalize the prior on the matrix Z in matrix decomposition by creating a prior that generates a matrix instead of individual elements in a matrix. More specifically, each row in a latent feature model still corresponds to a distinct sample, but now the columns correspond to distinct class and each sample may belong to a number of the classes. This is unlike probabilistic matrix factorization, where each row must sum to one. The trade-off between latent variable models and latent feature models is that the latter is more mathematically complex and computationally demanding due to the underlying generative process. As such there has been limited adaptation of latent feature models in genetic data, but researchers have showed their value such as Knowles et al. (151) identified unique gene expression patterns in drosophila using a sparse latent feature model.

2.10 Modeling Choices for Latent Feature or Variable Models

As we have discussed, the choice of priors in a latent feature or variable model defines the interpertability and expressive of the model. In our example of the GMM, we used a simple discrete distribution that was represented as a K-dimensional one hot vector (only one class is selected). Alternatively, one can stack priors to create a hierarchy such that latent variables share parameters with latent variables higher up in the hierarchy akin to hierarchical clustering. A

more popular approach is use a stochastic process such as the Dirichlet Process (DP) (152) that is a non-parametric Bayesian prior to support all discrete distributions of any dismensionality. A DP-GMM allows a mixture model to have an unbounded number of classes and removes the difficulty of determining the optimal number of classes in a generic GMM (153). The success of Dirichlet process has led to other priors for Bayesian nonparametric in latent feature models that we describe and use in this thesis.

2.10.1 De Finetti's Theorem

Models built upon Bayesian non-parameterics exploit the properties of exchangeability and we discuss the concept for motivating the priors we use in this thesis.

Definition 2.10.1. A finite set of random variables $x_1, ..., x_n$ is *exchangeable* if every permutation of $x_1, ..., x_n$ has the same joint distribution as every other permutation of the random variables (154)

This definition simply states that the order of the features is inconsequential to the underlying generative process that describes the data.

For Bayesian non-parameterics this definition is extended for an infinite number of features called De Finetti's Theorem.

Definition 2.10.2. An infinite set of random variables $x_1, ..., x_{\text{inf}}$ is *infinitely exchangeable* if and only if there is a random probability measure P with respect to $x_1, ..., x_{\text{inf}}$ are conditionally individual and identically distributed with distribution $P(\theta)$ has the same joint distribution as every other permutation of the random variables (154)

Therefore a latent matrix is exchangeable then there is a distribution P known as the De Finetti mixing distribution such that, conditional on θ drawn from the probability measure P $x_1, ..., x_{\text{inf}}$ are conditionally individual and identically distributed. The priors we use in this thesis are based on the probability measures created using the completely random measure.

2.10.2 Completely Random Measure

A *random measure* M is a stochastic process that is indexed on a measurable space A. A completely random measure (CRM) (155) is a random measure ρ such that collection of finite disjoint sets $A_1, ..., A_n \in A$ are independent random variables $\rho(A_1), ..., \rho_{A_n}$. Since De Finetti's theorem from Definition 2.10.1 describes that any infinitely exchangeable sequence can be described as a mixture of i.i.d. distributions, a CRM can be used as the De Finetti mixing distribution. Kingman et al. (155) showed that a CRM can be decomposed into three parts: the existence of a feature (atom a), the weight (w), and an ordinary component deterministic component. What is of particular interest to practitioners is the atom and the weight described as:

$$M(A) = \sum_{n}^{\infty} w_n \delta_{a_n}(A) \tag{2.4}$$

Kingman et al. (155) further characterizes this as a joint measure $\nu(dw, da)$ and the goal is to identify ν. In latent feature and variable models researchers have identified ν as the Levy process which has independent and stationary events with mean measure $\nu(dw)$. A Poisson process is an example of a Levy process as every interval in $t - s$ is independently Poisson distributed with mean $\lambda(t - s)$ i.e. $\nu(dw) = \lambda(t - s)$. This leaves $nu(da)$, which is referred to as the base measure

or more formally $H(A)$. H is some kind of probability distribution on A. For example, in a non-parametric GMM the base measure is the Dirichlet Process (156) to allow for an unbounded number of possible classes. Griffths et al. (145) created the Indian Buffet process (IBP) that is a distribution over exchangeable binary matrices using a CRM where th. The CRM that creates the IBP has the Levy measure of the Bernoulli process and the base measure of the Beta process (157) to generate a infinite collection of k classes in a binary matrix. Other CRM include the gamma process (152), poisson process (155), and the beta process (158).

2.10.3 Gamma Process

The Gamma Process (GP) is a CRM we use often in thesis. The GP has a levy measure $\nu(dw) = c\frac{e^{-cw}}{w}$ where c is a concentration parameter. We can show that $M \sim GP(H, c)$ is a CRM since the random measure $M(A_n)$ on any set $A_n \in A$ are disjoint gamma distributions. We can therefore write $M(A_n) \sim Gamma(H(A_n), c)$ such that $H(A_n)$ is the shape and c is the rate (152). The gamma process has an infinite number of atoms but we can still have a finite sum if the base measure is finite ($H(A) < \infty$) because we know the mean of the gamma distribution, $E[M(A)] = \frac{H(A)}{c}$. The advantage of using a Gamma Process as a prior for a latent feature model is that the matrix is integer-valued instead of a binary vector since a Gamma distribution generates weights that lie in the space of $[0, \text{inf}]$. Since much of the genetic data is count based, a GP is an appealing prior on latent feature models to model genetic data. As a note, the famous Dirichlet Process is constructed by normalizing the gamma process (152). The Dirichlet Process, however, is not a CRM since not every set is disjoint due to the fact that weights are constrained to sum to 1 (i.e. weights are normalized).

2.10.4 Learning Via Variational Inference

The Bayesian part of latent variable models comes from decomposing the joint distribution using Baye's rule:

$$P(\theta, z|x) = \frac{P(x|z,\theta)p(z,\theta)}{p(x)} \tag{2.5}$$

The posterior distribution $P(\theta, z|x)$ is what allows one to explore the data and interpret the hidden structure. The posterior inference, however, is generally intractable with the exception of simple models due to the normalizing constant of $p(x)$. A common technique is to utilize Markov Chain Monte Carlo (MCMC) (159) that creates a Markov Chain over the latent variables until the chain converges to a stationary distribution. Unfortunately as the complexity of a model increases or the dataset size, the computational burden increases when using MCMC.

The algorithms utilized in this thesis follows the paradigm of variational inference to approximate the posterior distribution. Variational inferences posits a family of variational distribution $q(z, \theta)$ such that a scoring function minimizes the distance between the variational distribution and the true posterior. Whereas MCMC creates samples from the Markov chain to converge to a posterior distribution, variational inference transforms the posterior inference problem as an optimization problem by exploiting Jensen's inequality:

$$\begin{aligned}
\log p(x) &= \log \int_{z,\theta} p(x,z,\theta) dz d\theta \\
&= \log \int_{z,\theta} q(z,\theta) \frac{p(x|z,\theta)}{q(z,\theta)} dz d\theta \\
&\leq \int_{z,\theta} q(z,\theta) \log \frac{p(x|z,\theta)}{q(z,\theta)} dz d\theta \\
&= \mathbb{E}_q[\log(p(x|z,\theta)] - \mathbb{E}_q[\log q(z,\theta)]
\end{aligned}$$

This results in a log-likelihood of $p(x)$ that we are interested in optimizing or more commonly referred to as the evidence lower bound (ELBO), since it is a lower bound on the data itself $p(x)$.

CHAPTER 3

A LESION BASED RESPONSE PREDICTION MODEL USING PRETHERAPY PET/CT IMAGE FEATURES FOR Y90 RADIOEMBOLIZATION TO HEPATIC MALIGNANCIES

This chapter is published as Mehta, Rahul, Kejia Cai, Nishant Kumar, M. Grace Knuttinen, Thomas M. Anderson, Hui Lu, and Yang Lu. "A lesion-based response prediction model using pretherapy PET/CT image features for Y90 radioembolization to hepatic malignancies." Technology in cancer research & treatment 16, no. 5 (2017): 620-629.

3.1 Motivation

Yttrium-90 radioemboliszation (Y90-RE) is recommended for unresectable, chemorefractory liver-dominant primary or metastatic hepatic disease with a life expectancy of 3 months or longer (160). Accumulated studies demonstrate that Y90-RE can improve overall outcome of disease progression, from being unresectable to resectable or covert incurable disease to transplantable and potentially curable in patients with colorectal carcinoma (CRC) liver metastasis (161), neuroendocrine liver metastasis (162), primary hepatocellular carcinoma (163), or intrahepatic cholangiocarcinoma (164). If a patient, however, is unresponsive to the Y90-RE, the expensive and technical demanding treatment will result in unnecessary risk and cost to the patient.

Clinical decisions regarding the treatment response of Y90-RE rely on imaging assessment at various stages of the treatment process and is pivotal for patient management. As patients

who are referred for Y90-RE generally have advanced, unresectable, and chemo-refractory liver malignancies, fluorine-18-2-fluoro-2-deoxy-D-glucose positron emission tomography and computed tomography (FDG PET/CT) has been widely used in clinical practice for preprocedural workup. FDG PET/CT can depict tumor metabolic activity of the tumors as well as certain anatomic features including tumor size and lesion density. European Organization for Research and Treatment of Cancer (EORTC) (165) criteria and PET Response Criteria in Solid Tumors (PERCIST) use the change in tumor standardized uptake value (SUV) to determine tumor response and subsequently analyze patient outcome (166). CT assessments can use size-based revised Response Evaluation Criteria in Solid Tumors (RECIST 1.1) (167) or tumor necrosis based Choi criteria (65). Recent reports have shown that FDG PET/CT were useful in predicting overall survival after Y90-RE in patients with FDG-avid liver metastases (168; 169), or primary liver cancers (170) especially cholangiocarcinomas (171).

The objective of this study is to create a model that predicts patient treatment response from the pretherapy FDG PET/CT scan alone. Since FDG uptake has been found to be correlated with microenviromental tumor characteristics, such as hypoxia, cell proliferation and blood flow, we hypothesize baseline FDG PET/CT scans contain unique imaging features that are shared across the patient population (172). Toward this goal, several studies have evaluated the effectiveness using conventional methods such as the maximum of the baseline SUV (SUVmax) for analyzing the prediction of therapy response. However, the predictive power in these studies remains considerably low, thus limiting SUVmax as a prospective predictive tool for patient treatment options (173; 174).

In contrast, machine learning and pattern classification techniques trained on imaging features based on the texture of a tumor have yielded better results in predicting patient response (175). However, the predictive performance of these features are still limited by the mischarcterization of lesion heterogeneity and the influence of lesion transformations such as orientation or scale. For instance, a small tumor may still be a malignant and a tumor may only be a partial responder due to heterogeneity.. The ideal scenario to compare lesions would be able to use imaging markers that are dependent only on the lesion morphology as described by the distribution of the voxel properties within the lesion.

To address the above challenge, we build a model that extracts invariant texture and shape imaging features from lesions to identify common imaging features. We assume each lesion is a mixture of these imaging feature and the contribution of a distinct feature can be used to differentiate between responders and non-responders. A Multinomial Naive Bayes classifier is utilized to investigate the accuracy of the imaging features. The classifier is trained on the extracted imaging features of the tumor from pre-therapy FDG PET/CT and we test if the classifier can predict treatment response of a separate distinct lesion. The framework is general and can be applied to classify other lesions and the corresponding treatment response. Our results show that the computerized machine vision predication algorithm based on extracted pretherapy FDG PET/CT imaging features is able to predict a patient as a responder or non-responder to Y90-RE with an Area Under the Curve (AUC) of .848 and a sensitivity of 0.791.

Patient Characteristics	Values
Age,years	
Median	62.5
Mean	63
Range	30-77
Sex	
Male	6
Female	6
Tumor volume, mm^3	
Mean ± standard deviation	54.9 ± 11.8
Range	30.4-91.5

TABLE I: Characteristics of Eligible Patients for this Study

3.2 Materials and Methods

3.2.1 Patients

This retrospective study has been approved by University of Illinois Hospital and Health Sciences System Institutional Review Board. A total of 173 Y90-RE procedures were performed between January 2011 and June 2014 in our institution. Patients who underwent both pre- and post-therapy FDG PET/CT scans with at least 3 months clinical follow up were included. For patients who received multiple consecutive Y90-RE therapies, we treated each patient's follow-up scan as a pre-therapy scan for the next Y90-RE treatment. There are were a total of 12 patients (8 with liver metastases from colon cancer and 4 patients with cholangiocarcinoma) with 30 Y90-RE procedures that met our selection criteria (Table Table I). Patients with complete and partial response are were combined into responders (R) while stable and progression of disease are were combined into non-responders (NR). On lesion-based analyses (Table Table II), our data

Tumor Response	EORTC (165) and PERCIST (166) Based PET criteria	Choi criteria (65) based on tumor necrosis	RECIST1.1 (167)-Based Size Criteria
Complete Response (CR)	Resolution of metabolic active lesions	Complete necrosis	Disappearance of all lesions no new lesions
Partial response (PR)	25% and 0.8U decrease of SUVmax of the most intense lesions	More than 15% decrease in tumor density	30% decrease in sum diameters of the target lesions
Progression disease (PD)	Greater than 25% increase in SUVmax or more than 20% increase in extent	Not applicable	20% increase of sum diameter of the target lesions
Stable disease (SD)	Increase of less than 25% or decrease of less than 15% of SUVmax or no visible increase in extent	Not applicable	<30% decrease, or <20% increase of sum diameter or the target lesions

TABLE II: Combined Imaging Criteria for Assessment of Treatment Response.

set was decided determined to have 17 responders and 13 non-responders. Of the 17 responders, 4 cases were from 3 patients with cholangiocarcinoma, and remaining 13 cases were from 8 patients with metastatic colon cancer. Out of the 13 non-responders, 4 cases were from 3 patients with cholangiocarcinoma, and the remaining 9 cases were from 7 patients with metastatic colon cancer. Representative FDG PET/CT cases of responder and non-responder for Y90-RE. are shown in Figure Figure 6.

3.2.2 Patient Treatment

All patients underwent hepatic arteriography with Tc-99m macroaggregated albumin (MAA) to detect extrahepatic shunting 7 to 14 days before prior to Y90-RE. To correct for flow into extrahepatic organs, we used coil embolization of shunting vessels or catheter positioning. All patients had less than 20% lung shunt fraction, or lung exposure more than 30 Gy in a single Y90-RE procedure, or more than 50 Gy in multiple Y90- RE sessions that are contraindicated to Y90-RE.

Y90-RE procedures were performed according to the published protocols (176). The lobar branch of the hepatic artery supplying the tumors was used to supply the Y90 resin microspheres (SIR-Spheres; Sirtex Medical, Lane Cove, Australia). The Y-90 dose was prescribed based on a published body surface area (BSA) method (160). The tumor and total liver volume ratio was calculated using CT.

3.2.3 Patient Imaging

The 18F-FDG PET/CT scan examinations werewas performed within 2 weeks before treatment and, at 4 weeks and 3 months after treatment. 18F-FDG PET/CT was performed on a GE Discovery 690 PET/CT scanner (GE Medical Systems, Milwaukee, WI) using a standard protocol. Following Miller et al. (177), patients fasted for at least 4 hours before scanning and had a blood glucose level <200mg/dl at the time of FDG injection. Dedicated PET /CT scans from the skull base to the upper thighs were obtained 60-90 minutess after IV injection of 10-13mCi of FDG. with CT parameters of: 120 kV, 120 mAs, pitch 0.813, 16 x x1.5mm collimation; and. PET parameters: of 3 minutes bed/position. Additionally, FDG PET/CT examinations were performed within 2 weeks before treatment, at least 30 days after treatment, and at 3 months intervals subs as recommended by Miller et al. (177).

3.2.4 Clinical Evaluation of Treatment Responses

The FDG PET/CT imaging studies scans were analyzed retrospectively on a dedicated AW PACS workstation (GE Medical Systems, Milwaukee, WI). Lesion-based treatment response on FDG PET/CT waswere clinically evaluated according to a combined criteria: (PET- based EORTC (165) and PERCIST (166) criteria, tumor necrosis based Choi criteria (65) and tumor

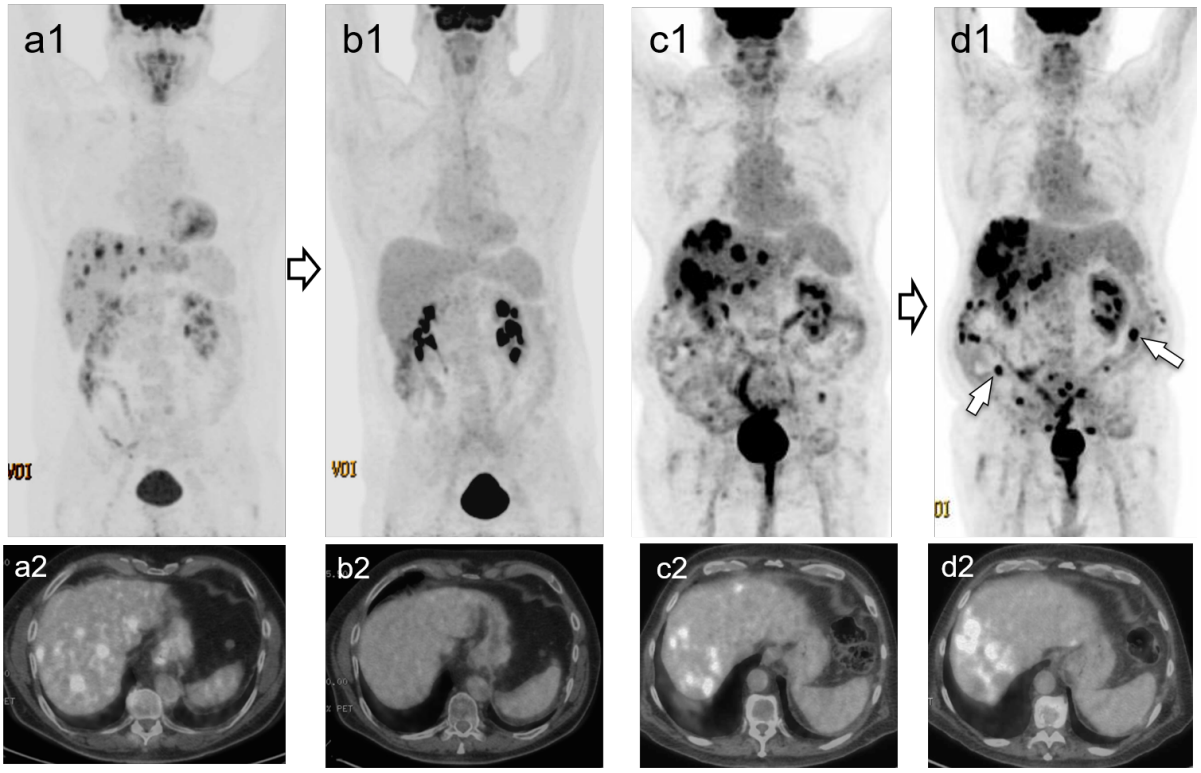


Figure 6: Representative FDG PET/CT cases of responder and nonresponder for Y90-RE. **Patient 1** with liver metastases from colon cancer is a responder to Y90-RE with (a1) Pretherapy FDG PET/CT (a2) Fusion FDG PET/CT image showing multiple hypermetabolic bilobar liver metastases (b1) Posttherapy FDG PET/CT MIP CT (b2) Posttherapy FDG PET/CT image demonstrating resolution of hypermetabolic liver metastases. **Patient 2** with liver metastases from colon cancer is a nonresponder to Y90-RE with (c1) Pretherapy FDG PET MIP image (c2) Fusion FDG PET/CT image showing multiple hypermetabolic bilobar liver metastases (d1) Posttherapy FDG PET/CT image demonstrating progression of hypermetabolic liver metastases (d2) Additional multiple new hypermetabolic intraperitoneal metastases also developed (d1, arrows).

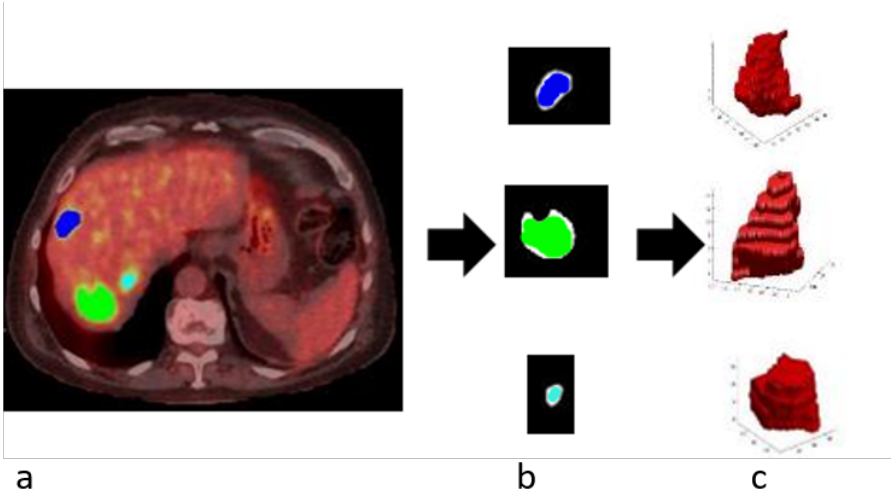


Figure 7: (a) Representative 3 FDG PET/CT avid liver metastases with different morphology in a patient with metastatic colon cancer each marked with a distinct color. (b) The region of interests (ROIs) of the 3 liver metastatic tumors are extracted using spatial fuzzy clustering and then used to interpolate into 3D space. (c) The 3D rendering of the lesions.

size based RECIST1.1 (167), (Table Table II), with the consensus among physicians. Lesions density, size and metabolic lesion volume changes were correlated with the changes in SUVmax on FDG-PET. In the conflicting cases, the criteria with the best response were used to determine patient response. For example, SUV change within 25% (stable on PERCIST), but decreases in size by 30% (partial response on RECIST), or necrosis by 50% (partial response on necrosis criteria), was considered partial response. We considered a patient as a complete responder based on a resolution of FDG avidity and/or complete necrosis of treated lesions regardless of changes in lesion size. Cancer progression was attributed to an occurrence of new lesions in the treated liver lobe regardless of changes in SUV, necrosis or lesion size.

3.2.5 Quantitative Analysis

Three-dimensional lesion extraction. For every FDG PET/CT slice, the lesion cross-section was separated using the Spatial Fuzzy C-Means Clustering algorithm (SFCM) (178; 179). This technique automatically assigned assigns image voxels to a cluster based on their relative distance to each other and the correlation between their intensities. The subsequently optimized clusters based on both spatial and intensity characteristics filters out lesions with partial volume effect (PVE) or voxels of noise (179). The three-dimensional (3D) representation of an extracted lesion was then rendered by interpolating image slices along the aligned cross-section directions using cubic spline interpolation (Figure Figure 7). To improve computational efficiency, lesions were stored in a 41x41x41 cube as this e were excluded. This semi-automated algorithm allowed us to analyze much more information than manual contouring, while being effective in segmenting lesions in PET/CT(21).. Once a lesion was extracted from FDG PET/CT image slices, we interpolated relative to the total number of slices containing the lesion and created a 3D representation of the lesion (Figure 2). Every lesion was set to a resolution of 41x41x41, as this cube size was size is able to incorporate all the various lesion sizes (although matrix sizes were kept the same, larger lesions took much more space of the cube than smaller lesions).

Image feature extraction techniques. As any lesion can be described as a 3D object in the disease process, we use these 3D computer vision methods to find the common patterns within the lesions. These methods have found success in comparing natural 3D shapes such as buildings (180) as well as shape analysis of proteins (181). In the computer-aided FDG PET/CT image analyses, we used 3D Spherical Gabor filter (3DSG) as the texture feature to find probe

the spatial interrelationships and arrangement of the basic elements of the tumor (182), 3D Zernike Descriptor (3DZD) to find extract co-occurring geometric patterns (181), and Wave Kernel Signature (WKS) for the shape features (183). Details on each algorithm are described as follows.

A 3DSG filter is the product of a 3D Gaussian and a complex exponential function representing a sinusoidal plane. We created multiple filters by changing the shape σ, rotation (θ, ϕ), and central frequency (F) of the Gaussian. A representation pf 3D lesion was obtained by convolving the original 3D lesion, *I(x,y,z)* with each filter. As a result of the convolution, each representation describes how many voxels does a cycle of periodically repeating intensity variations occurs in the 3D lesion. The parameters used for creating the filters were based on obtaining a set of Gabor filters that could acquire the largest variation within the voxels which were determined to be .125 cycle/voxel for the highest frequency and .03125 cycle/voxel for the lowest frequency. Given these criteria, the following parameters that achieve the best performance are:

$$F = \pi 2^{\frac{i+2}{2}}, \theta = k\pi/8, \phi = j\pi/8]$$

Where $i = 2, 3, 4, 5, 6, j = 0, ..7$, and $k = 0, \ldots, 7$. In addition we set the shape of the envelope by setting $\sigma = \frac{1}{F}$. The resulting filter bank contained 320 3D Gabor kernels (5x8x8) of size 41x41x41 that were convolved with the original 3D image of the lesion, I(x,y,z). To obtain rotation and scale invariant features we calculated the 3D Discrete Fourier Transform (DFT) of the mean and standard deviations of the convolved images as the magnitude of the DFT has

been shown to be shift-invariant. Each lesion was then represented with a feature vector of size 1x640.

Wave Kernel Signature (WKS). The WKS feature descriptor provides robust analysis to non-isometric perturbations of the surface (183). This is imperative in our work as tumors have no predefined shape and are anisotropic (grow in many directions). Furthermore, the key advantage of WKS is that it is able to differentiate between the small structures between shapes that also appear often within lesions. The main goal is to find correspondences between similarly shaped lesions that also have the same treatment response. Specifically, each signature describes a unique property of the lesion. We adopted parameters that are found the original implementation of the WKS to generate 100 signatures, resulting in a feature vector dimension 1x100.

3-Dimensional Zernike Descriptor (3DZD). The advantage of 3DZD is its ability to describe non-spherical like shapes, a property that also describes tumor shapes (181). The 3DZD was achieved by first converting the original image Cartesian coordinates, $I(x, y, z)$, into a set of spherical coordinates $I(r, \theta, \phi)$, and then evaluating the Zernike Polynomial at that point with parameters n, m, and, l. We use the reported optimal parameters (184) of n, m, and l as 20, 14, and -14 respectively, with a resulting feature vector size of 1x122.

Spatially Sensitive Bag of Features (SSBoF) and Visual Words. Given the large size of the feature vectors in comparison to the size of our dataset and the lack of interpretability of the features, we used SSBoF to summarize the features into meaningful information. The SSBof aggregates the shape features by clustering a pair of features such each cluster encodes

the spatial distance between two shape features. We can intuit this as a shape feature for a box should be much closer in space to a rectangle than a circle. Each cluster is referred to as a "visual word" and frequency pattern of how often a visual word occurs represents a 3D lesion (185). Therefore each 3D lesion is represented by a histogram, where each bin signifies a visual word and the counts indicate how often a specific visual word occurs in the 3D lesion.

We created a SSBoF vector for each of the original feature descriptors (3DSG, WKS, and 3DZD) by finding the the optimal number of number of visual words (codebook) that represent a 3D lesion and provide the best predictive performance. We also created a SSBoF vector that combines all the original feature descriptors to determine if performance can be improved. To remove any potential bias caused by our assumption that a patient's follow-up FDG PET/CT scan can act as a pretherapy scan for subsequent Y90-RE treatment, we added a visual word to all histograms denoting if a patient had received a previously received 90-RE therapy.

3.2.6 Predictive Performance and Statistical Analyses

We used a Multinomial Naïve Bayes (MNB) classifier to predict if patients are responders or non-responders (186). The MNB classifier is an extension of the Naive Bayes (NB) classifier that uses a multinomial distribution for each of the features to remove the conditional Independence assumption used in a NB classifier. A multinomial distribution is more appealing in our model since our feature vector is a count of the visual words that described each lesion. Our training set was 22 FDG PET/CT scans, with 13 responder cases and 9 non-responder cases; our test set size was 8 with 4 responder cases and 4 non-responder cases. In order to overcome the small size dataset size, we used bagging to create multiple training sets by sampling the training set with

Features	Recall	Precision
SUVmax (S)	0.485	0.760
Tumor Volume (V)	0.583	.631
3D Gray Level CoCurrence Matrix (C)	0.53	0.77
3D Spherical Gabor (G)	0.716	0.795
WKS (W)	0.763	0.833
3D Zernike Descriptor (Z)	0.611	0.811

TABLE III: Recall and Precision of the different feature groups.

replacement (187). A MNB classifier was then trained on each new training set and we predicted a patient as a responder or non-responder based on averaging the output class probabilities computed by each MNB.

Performance of the classifier was evaluated using a precision/recall characteristic. Precision, or positive predictive value, is the number of correct predicted responders (true positives) divided by the total number of predicted responders (sum of true and false positives). Recall, or sensitivity, is the number of correct predicted responders (true positives) divided by the total number of clinical defined responders (sum of true positives and false negatives). We also used the Kruskal-Wallis test to assess if there was a significant difference in the performance of the classifier when using different feature descriptors, and the Dunn test to compare between the responder and non-responder groups.

3.3 Results

The experiments were geared towards decreasing unnecessary procedures and patient risk, so we chose parameters that would minimize the number of false responders, i.e., false positive rate

Fused Features	Recall	Precision
G + W + Z	0.791	0.839
G+ W + Z + S	0.758	0.763
G + W + Z + C	0.683	0.717
G + W + Z + V	0.821	0.844

TABLE IV: Various fused feature groups and the resulting Precision and Recall.

(FPR). Given this condition, we found the optimal size of our codebook being to be 17, 8, and 11 for 3DSG, WKS, and 3DZD respectively ($p < .05$). The precision and recall results were are listed in Table Table III. The WKS feature descriptor provided the best recall (0.763) and precision (0.833) in comparison to 3DZD ($p < .05$), and marginal improvement in comparison to 3DSG. The receiver operating characteristic (ROC) curve, show similar predictive performance between WKS, 3DSG, and 3DZD in Figure Figure 8. Therefore measuring the spatial distributions of voxels in terms of frequency and shape allows us predict patient treatment response.

To test whether the new imaging features enrich lesion analysis, we compared our imaging features with SUVmax, 3D Gray Level Co-occurrence Matrix (3DGLCM), and tumor volume. Both SUVmax and 3DGLCM ROC Figure Figure 8 is similar to previous research (175; 188). Furthermore Table Table III, shows tumor volume or SUVmax as a sole predictors resulted a recall of 0.485 and 0.583, and a precision of 0.76 and 0.631 respectively, significantly lower than ($p < .05$) those of 3DZD, 3DSG and WKS. Similarly, features from the 3DGLCM had a recall of 0.53 and a precision of 0.77, indicating that the contrast between voxel intensity does not accuractely capture lesion heterogenity.

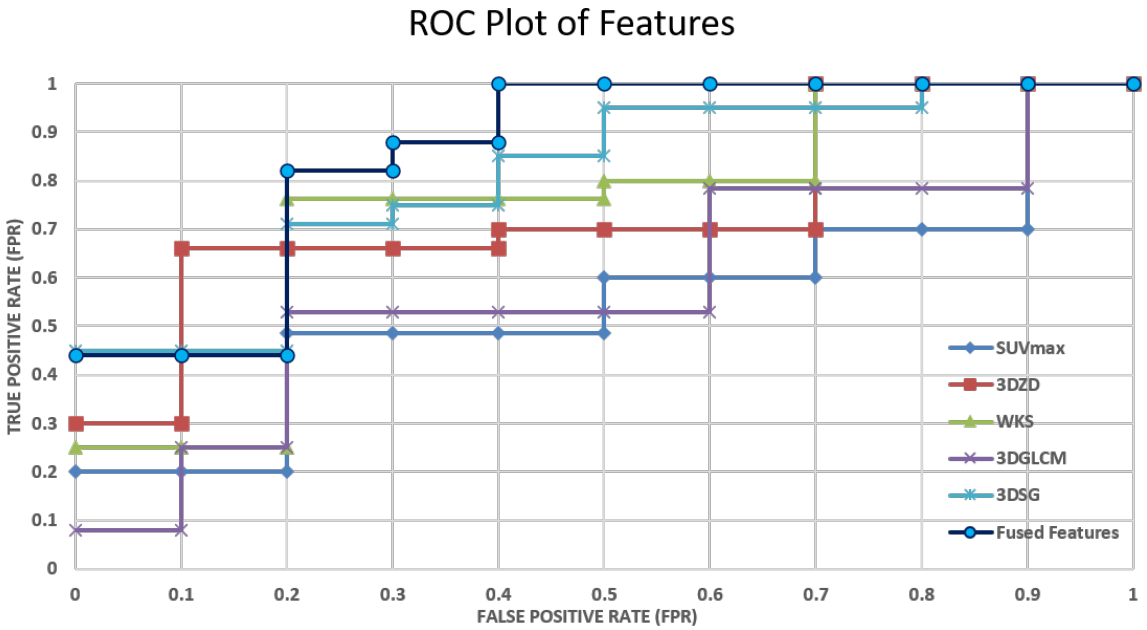


Figure 8: ROC Plot of the distinct features: SUVmax, 3DZD, WKS, 3DGLCM, and 3DSG, as well as the best before fused features of 3DSG, WKS, 3DZD, and volume.

We hypothesized combinations of the various features would improve the results. Using the same conditions as before we obtained an optimum codebook size of 14 when combining When combing 3DZD, 3DSG, and WKS as one feature descriptor (G+W+Z). The resulting descriptor is also concatenated with SUVmax, 3DGLCM, and tumor volume (TableTable IV). We omitted the models with different combination features that did not show additional performance gains.

With the addition of tumor volume as a feature (G+W+Z+V), there was a significant increase ($p < .05$) in recall (0.821) comparing with that of G+W+Z (recall 0.791). Adding the 3DGLCM or SUVmax as a feature did not offer more insight into patient response and decreased predictive power. The difference between the various combinations of features illustrates the

value of volume as a distinct global feature to improve the performance. The analysis of the visual words can provide important information regarding the most distinctive features that are shared among different the non-respond and responder patient groups.

We analyzed the visual words of the 3DSG, 3DZD, and WKS to determine which visual word is shared among the different patient groups (Figures Figure 9,Figure 10). The average visual words histograms for responders and non-responders were computed using the corresponding histograms obtained from the model. We then assessed the visual word that is distinctive among the two patient groups. We found that the 10th bin for 3DSG, 5th bin for 3DZD, and 2nd and 7th bin for WKS were the most pronounced in responders ($p < 0.05$). In the case of non-responders, the 4th bin for 3DSG, 3rd bin for 3DZD, and 1st and 8th bin for WKS were found to be most pronounced ($p < 0.05$). These results suggest that there are common imaging patterns that differentiate the responder and non-responder patient groups.

3.4 Discussion

To have the ability to predict the outcome of a treatment based on pretherapy FDG PET/CT may avoid unnecessary patient risks and expensive, invasive procedures, along with the potential to provide precision treatments. Our study focuses on using pretherapy FDG PET/CT data to establish a Y90-RE treatment response prediction model. Imaging analysis of a patient's lesion characteristics is important for tumor response evaluation after therapy. Exploration of commonly used image features such as lesion size, tissue density, and lesion SUVmax on FDG PET/CT in predicting tumor response have found them to be limiting. Meanwhile, computer-aided imaging analyses in evaluation of tumor response have been showing promising results. Tixier et al.

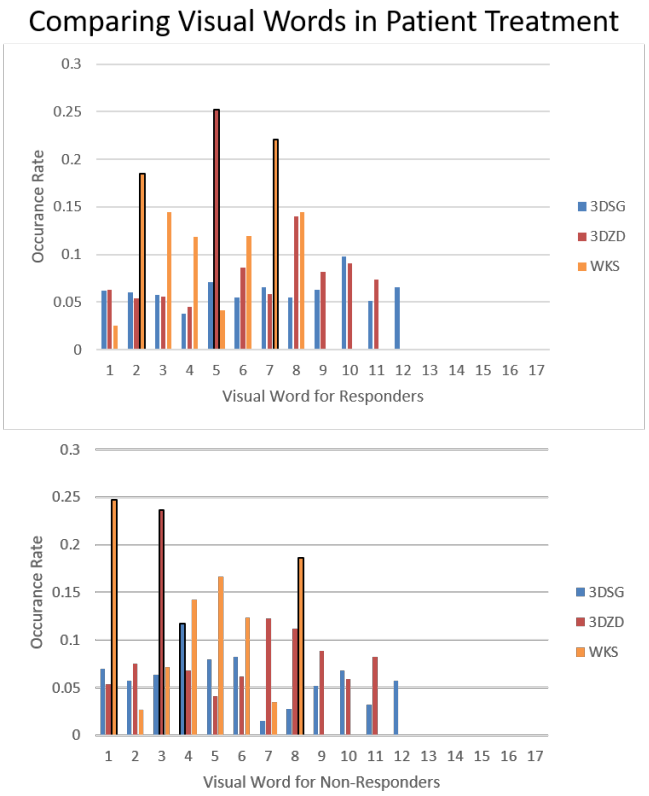


Figure 9: Average visual words for responders (top) and nonresponders (bottom): 10th bin for 3D spherical Gabor filter (3DSG), 5th bin for 3D Zernike descriptor (3DZD), and 2nd and 7th bin for wave kernel signature (WKS) were the most pronounced in responders, whereas the 4th bin for 3DSG, 3rd bin for 3DZD, and 1st and 8th bin for WKS were found to be most pronounced for nonresponders. These visual words can be used to determine patients who are more likely to respond to treatment.

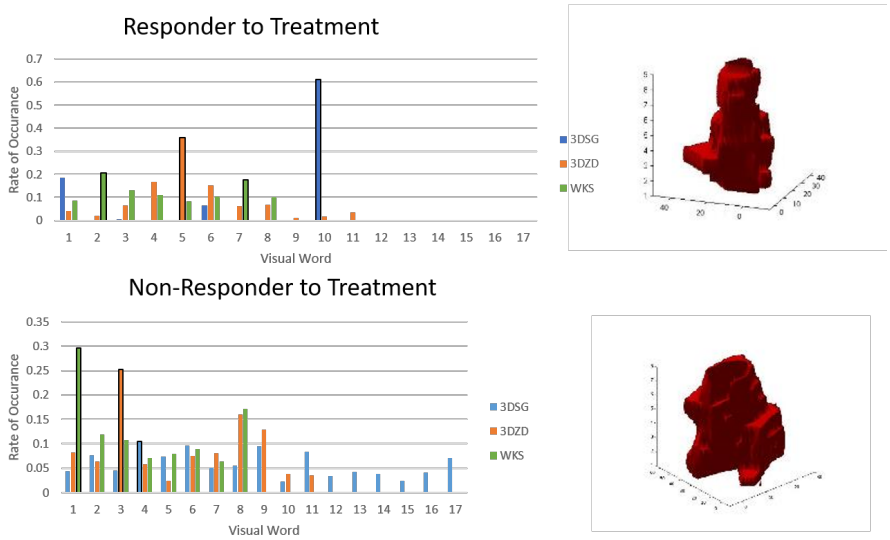


Figure 10: The most common visual words (left column) for the corresponding largest tumor lesions (right column) in the representative responder (upper row) and nonresponder (lower row). In the case of 3D spherical Gabor filter, the high occurrence of the 10th bin indicates a likely patient response, as shown in responder 2 (upper row); the high occurrence of the fourth bin describes a tumor that will not respond, as shown in nonresponder 3 (lower row).

(188) used local homogeneity as a texture feature to identify patients who would respond to chemoradiotherapy with an Area Under the Curve (AUC) of .7 in comparison to SUV, which that had an AUC of .59. Similarly, Tan et al. (175) showed a tumor with a greater score in homogeneity is more likely to respond to therapy.

The preliminary success of using textures features in tumor imaging analyses indicates there is a large amount of information in a radiological image that is not discernaible by conventional tools. We demonstrate that a predictive algorithm based on 3D imaging features extracted from pre-therapy FDG-PET/CT scans is able to predict patients as responders or non-responders of Y90-RE. We also showed that SUVmax as an imaging feature has little predictive power. Upon further analyses of individual image features (Table Table III), the commonly used clinical image features, SUVmax and tumor volume, were associated with the lowest predication recall and precision. This is probably due to the fact these are global features and do not fully capture the distribution of voxels within a lesion. Specifically these features do not reflect the heterogeneity within a lesion (180). Moreover, we found texture features derived from 3DGLCM also had low prediction power because it is reliant on voxel intensity and is unable to detect subtle differences between tumors. The WKS feature descriptor in particular provided the highest recall and precision ($p < .05$) in predicting a patient's response to Y90-RE as it describes the irregular shape of tumors. This finding is consistent with Waclaw et al. (189) that a distribution the shape of a tumor effect it's response to a treatment.

When combining different feature descriptor groups, we found that the best group of features in improving optimal precision and recall is the combination of the 3D invariant features (3DSG,

3DZD and WKS) and tumor volume ($p < .05$). As each feature adds a unique element to the classifier, it is understandable that the performance of the feature group is superior to individual features alone. Of particular interest is the feature of tumor volume. When used alone in the predication model, tumor volume had a low predicative power. However, in combination with the invariant imaging features, the performance increased significantly. This is probably likely due to the factor invariant images ignore scale when finding patterns among lesions, yet the volume of a tumor is still a critical factor in determining the prognosis of a patient. Therefore, an improved model should modify the imaging features so they are not invariant to volume.

We envision that, insofar as tumor heterogeneity is concerned, there is certain tumor homogeneity, or common features, that are related with underlying gene profiling within the tumor and tumor microenviroment. Beyond the potential to stratify patients further with respect to imaging features, as we observ in Figures Figure 9 and Figure 10, unique visual words are associated with patient response to Y90-RE. Therefore, it is possible to select patients that have the potential to respond to response to Y90-RE based on their pretherapy FDG PET/CT imaging features.

A major limitation of this study is the small number of cases. A although, we attempted to overcome the limitation of a small dataset by using a bootstrapped model, a larger dataset would surely will sure provide us with a more robust learning schema as itand could improve performance. Upon validating our model further on a larger dataset, we hope to provide a disease- and treatment-specific prediction model based on pretherapy FDG PET/CT image features.

3.5 Conclusion

In summary, we developed a model that predicts Y90-RE therapy response in patients with primary and secondary liver cancers, based on lesion's invariant texture and shape imaging features extracted from pretherapy FDG PET/CT scans. Our approach utilized utilizes computer vision techniques of 3D Spherical Gabor, Wave Kernel Signature, and 3D Zernike Descriptors to find regions of tumors that were similar locally even if they have global differences such as volume. By using a Spatially Sensitive Bag of Features to describe imaging features as visual words, we were able to find unique image features that were common within responders and non-responders. We showed the benefit of using the invariant techniques that were resistant to change in scale, transform, and rotation in assessing Y90-RE treatment response in comparison to routine clinical used image features such as SUVmax and tumor volume. The model improved improves when combining our image features with tumor volume, indicating imaging features invariant to scale are not necessary when describing lesions. While our model will need further validation on a large dataset, the proposed method is general and can be potentially applied to any lesion from a different disease model.

CHAPTER 4

DISCRIMINATION OF MALIGNANT AND BENIGN BREAST LESIONS USING MACHINE LEARNING ON MULTI-MODAL DIFFUSION MRI PARAMETERS

4.1 Motivation

Breast cancer is the second cause of female cancer death in the US1. An accurate characterization of breast lesions is important for efficient risk assessment and optimized treatment planning. Magnetic resonance imaging (MRI) techniques have been used to evaluate suspicious breast lesions due their increased sensitivity over ultrasound and mammography2. Dynamic contrast-enhanced MRI (DCE-MRI), for example, is the primary MRI technique used for breast cancer diagnosis (190; 191; 192). Although the reported sensitivity of DCE-MRI is as high as 94%-100%6, its specificity has been found to be varying from 37% to 97%6–10. This variable specificity is a major limitation of DCE-MRI as a false positive diagnosis results in unnecessary biopsies. Moreover, the use of exogenous contrast in DCE-MRI is potentially problematic in patients with compromised kidney functions, cardiac failure, respiratory disorders, or pregnancy11.

With its sensitivity to probe underlying tissue microstructure, the use of diffusion weighted magnetic resonance imaging (DWI) has increased for characterizing breast lesions (193; 194; 195) using the apparent diffusion coefficient (ADC). Numerous studies have reported lower ADC values in malignant breast lesions due to increased cellularity compared to the benign lesions

(194; 196). For differentiation between malignant and benign breast lesions, ADC has been shown to achieve a sensitivity of 85–95%7–10,16 and a specificity ranging from 50% to 90%7–10,16. To improve diagnostic performance, clinicians have used DWI with ADC as an adjunct technique to the DCE-MRI (197). While the conjunction of the two MRI techniques has improved breast lesion characterization, the ADC has not been fully established as an imaging marker for breast lesion characterization18. This is partly because the mono-exponential model, from which ADC is derived, assumes that the diffusion displacements in a homogeneous medium follow a Gaussian distribution. This assumption, however, does not accurately represent the complex phenotypic and functionally distinct cell populations observed in breast tissue. Specifically, the complexity of breast tissue is further pronounced in lesions, benign or malignant, due to hypercellularity, angiogenesis, and other factors (198). As such, a simple ADC, which has been associated with tissue cellularity (199), cannot adequately reflect the complex water diffusion process in biological tissues (200).

While cellularity is an important measure for tissue characterization, breast tissue has many other properties, such as vascularity and heterogeneity. Recent studies have indicated that these properties can be probed by utilizing a specific portion of the b-value "spectrum" in DWI (200). For example, by utilizing relatively lower b-values (0-200; and 800 s/mm2), intravoxel incoherent motion (IVIM) model can reveal tissue cellularity and micro-vascularity through diffusion coefficient (D_{diff}), pseudo-diffusion coefficient (D_{perf}), and perfusion fraction (f) (see Chapter 2) (199). Researchers have reported the IVIM model to produce potential imaging markers that provide information on the functional properties of breast cancers without a contrast

agent (201; 202; 203; 201; 204; 205; 206; 207). Recent studies have also shown the feasibility and the added value of IVIM parameters with conventional ADC for differentiation between benign and malignant breast lesions (208; 209). Other studies have recently proposed a high b-value ($>$ 3000 s/mm2) non-Gaussian model, continuous-time random-walk (CTRW) model (210; 211) that recognizes the intra-voxel diffusion heterogeneity in time and space (212; 213; 214). The CTRW model introduces two new parameters related to temporal (α) and spatial (β) intravoxel tissue heterogeneity and an anomalous diffusion coefficient, D_m. Prior research has shown that the CTRW model or its predecessor fractional order calculus model (212; 213) is sensitive to tissue microstructural changes in the diseases of brain such as adult (215) and pediatric brain tumors (210; 216) and of body such as gastrointestinal stromal tumor (217). Within breast imaging, however, only a few studies have used a high-b-value DWI model (218; 219). Moreover, the use of a full b-value spectrum to probe a variety of tissue properties such as cellularity, vascularity, and heterogeneity, has not been explored for breast tissue characterization.

Irrespective of the model used to fit the diffusion signal, the most common approach for lesion characterization is to calculate the mean or median values of the estimated parameters using a region-of-interest (ROI)-based analysis. These values are then used in conjunction with a statistical model such as logistic regression, to determine whether a lesion is benign or malignant (220; 221). While simple, this approach artificially homogenizes the voxel-level functional information revealed from heterogenous tumor tissue (222). The shortcomings of using only median or mean values have led to recent diffusion MRI studies that use machine learning (113; 223) and quantitative features (224) obtained from the ADC or IVIM model

parameter maps to determine molecular subtype (113) or breast lesion malignancy (201; 225). These studies have shown that a histogram analysis can provide additional information by parsing the statistical distribution of the DWI parameters. The quantitative features obtained from the parameters can ameliorate the characterization of the specific tissue properties (i.e. vascularity, heterogeneity, and/or cellularity), probed by the IVIM and CTRW models. The challenge is to distinguish which of the extracted quantitative features offer the best discriminative performance.

In this study, we investigate the histogram-based quantitative markers obtained from the parameters of a low-b-value IVIM model and a high-b-value CTRW model for differentiating malignant and benign breast lesions using a machine learning paradigm. We identify the benefits of using a multi-parametric approach for breast lesion characterization by determining the top quantitative features using area-under the curve (AUC), F1-score, and accuracy obtained from a variety of machine learning models such as decision trees (DT), support vector machine (SVM), and gradient boosted (GB) classifiers. We demonstrate that a combined set of parameters from the IVIM and CTRW models translate to a robust predictive performance for discrimination between benign and malignant breast lesions.

4.2 Material and Methods

4.2.1 Patient Characteristics

The institutional review board approved this retrospective study, and written informed consent was obtained from all participating patients. Between May 2017 and March 2018, 31 patients (age range: 27 to 86) with a total of 40 pathologically verified breast lesions were enrolled. The inclusion criteria were as follows: (1) aged 18 years or older, (2) no previous neoadjuvant

chemotherapy or radiotherapy, and (3) ACR BI-RADS 4 or 5 (suspiciously malignant or highly suggestive of malignancy) and ACR BI-RADS 2 or 3 (benign or probably benign) lesions detected on breast MRI validated with a tissue biopsy.

4.2.2 Diffusion-weighted Image Acquisition

MRI was performed on a 3T scanner (GE Healthcare, Discovery MR750) with an 8-channel breast coil (Invivo Corp., Gainesville, FL) DWI with 11 b-values of 01, 501, 1001, 3002, 5002, 8002, 11004, 15004, 20006, 25006, and 30008 s/mm2 (subscripts denoting the number of averages) was carried out using a single-shot spin-echo echo planar imaging (EPI) sequence. Other image acquisition parameters were as follows: TR/TE = 7000/78 ms, slice thickness = 5 mm, FOV = 32 cm x 32 cm, and image matrix size = 256 x 256. Following Karaman et al. (83) trace-weighted images are generated by applying the Stejskal-Tanner diffusion gradient along the x, y, and z direction at each non-zero b-value to minimize the effect of diffusion anisotropy.

4.2.3 Diffusion-weighted Image Analysis

We first analyzed the multi-b-value diffusion-weighted (DW) images with the CTRW model using the following diffusion-attenuated MR signal:

$$\frac{S}{S_0} = E_\alpha(-(bD_\mathrm{m})^\beta) \tag{4.1}$$

, where $\mathrm{D_m}$ (in mm2/ms) is an anomalous diffusion coefficient, the parameters α and β (unitless) relate to temporal and spatial diffusion heterogeneities, respectively, and E_α is a Mittag-Leffler

function (210). We also analyzed the DW images using the IVIM model which has the following diffusion-attenuated MR signal (226):

$$\frac{S}{S_0} = fe^{((-b(D_{diff}+D_{perf})))} + (1-f)e^{(-bD_{diff})} \tag{4.2}$$

where f is the perfusion fraction, D_{diff} is diffusion coefficient in mm2/s, and D_{perf} is the pseudo-diffusion coefficient. The CTRW model parameters, D_m, α, and β, and the IVIM model parameters, f, D_{diff}, and D_{perf}, were estimated by fitting Equations Equation 4.1 and Equation 4.2 respectively. Each fit used a nonlinear least-squares estimation with an iterative Levenberg-Marquardt method in Matlab (MathWorks, Inc., Natick, MA) on a voxel-by-voxel basis, where thereafter a noise filtering and Rician noise correction was performed (212; 227). The nonlinear fitting proceeded with two major steps for the CTRW model fitting: (a) estimating D_m by a mono-exponential model using low b-values ($b \leq 1100$ s/mm2) diffusion images and (b) simultaneously estimating α and β from all images (b-values = 0–3000 s/mm2) after fixing each voxel's D_m at its estimated value (83). For the IVIM model fitting, a "segmented" approach was employed in three steps (228): (a) estimating D_{diff} by a mono-exponential model using the diffusion images at mid-range b-values (200-800 s/mm2) with the assumption that the pseudo-diffusion is negligible in this regime, (b) extrapolating the mono-exponential fit to b=0 to estimate f, and (c) constraining D_{diff} and f in the bi-exponential fit for Equation Equation 4.2 to obtain D_{perf}.

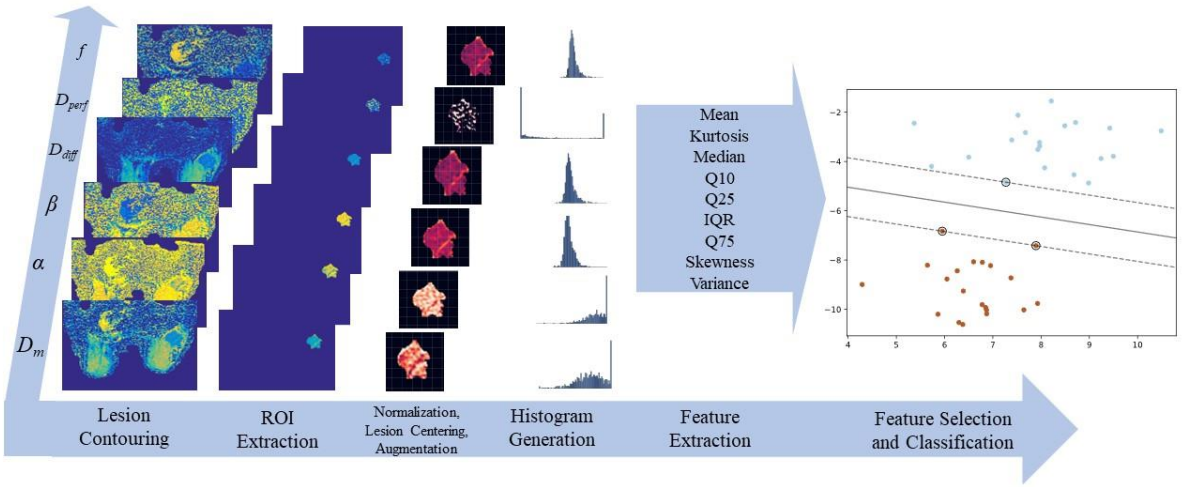


Figure 11: Workflow of our model from left to right with corresponding parameter maps and histograms from starting from D_m to f, bottom to top.

4.3 Data Analysis

Our full model pipeline is shown in Figure Figure 11.

4.3.1 Pre-Processing

A single slice region of interest (ROIs) were manually drawn by a radiologist including only the solid region of the tumor. The ROIs were then extracted from the individual maps of the CTRW parameters, D_m, α, and β; IVIM parameters, f, D_{diff}, and D_{perf}; and max-min normalized into a range of [0, 1] as seen in Figure 2. The parameter ROIs were then randomly cropped, flipped, and rotated to prevent overfitting55 and increase the effective size of the dataset from 40 samples to 120 samples.

4.3.2 Feature Extraction

A histogram of the voxel values within the ROI was generated for each diffusion parameter, D_m, α, β, D_{diff}, D_{perf}, and f to calculate histogram features. These features included: kurtosis, skewness, variance, mean, median, interquartile, 10% quantile, 25% quantile, and 75% quantile values of the histogram. Since there are nine features for each of the six parameters, we have a total of 54 quantitative features for the machine-learning analysis.

A key parameter in generating the histograms is setting the number of the bins or the bin width. We investigated the impact of the bin width on the predictive performance of breast lesion differentiation by using various number of bins: 40, 60, 80, 100, and 120. Various machine-learning classifiers, as explained in more detail in Statistical Analysis section, were then trained by using the distinct set of features generated by a histogram with a specific number of bins.

For each classifier, the optimal number of bins was chosen by determining which bin size yielded the best predictive performance.

4.3.3 Feature Selection

A recursive feature selection was performed by using the Boruta algorithm (229) combined with a modified two-stage multiple testing methodology process. At every iteration, feature importance was determined for all features with respective to the class classification (benign or malignant). The Benjamin Hochberg FDR test was used to test whether a feature performs better than expected by random. This was followed by a Bonferroni correction to account for multiple iterations of using the same set of features.

4.3.4 Statistical Analysis

The ability of the top features to predict a lesion as benign or malignant was compared using Kernel Methods, Ensemble Methods, and Naive Bayes. Kernel methods included SVM (230) and Gaussian Processes (GP) (231). Ensemble methods included techniques that make predictions based on either bagging (i.e. training individual parallel smaller models such that each model is trained on a subset of the data) such as DT (232) and Random Forest (233) (RF); or boosting (i.e. training sequential individual models such that the next model minimizes the prediction loss from the previous model) such as GB (234) or adaBoost (235) (AB). The respective parameters of each classifier are given in our code .

We split our training set and an independent test set to a ratio of 80/20, respectively. Additionally, we constricted the test set to not include any augmented samples so that the classifier only tests on unseen samples. The classifier parameters are fine-tuned using grid search

with a stratified cross-validation of 10 repeated 5-folds such that it splits the training set into five folds ten times to minimize variance and bias (236). To account for a slightly imbalanced dataset, we under-sampled the majority class during each grid search so that every fold contains an equal representation of each class.

Optimization of parameters including histogram bin width was performed by maximizing the ROC curves during the cross-validation stage. To compare models in the testing stage we used AUC, F1-Score, and accuracy metrics65 with 10000 bootstrapped samples to confidence intervals. A comparison of quantitative features was done using a Mann-Whitney U-test to generate a p-value.

4.4 Results

4.4.1 Diffusion Parameter Maps

Figures Figure 12a-Figure 12c show the maps of CTRW model parameters D_m, α, and β while Figures Figure 12d-Figure 12f show the maps of IVIM model parameters D_{diff}, D_{perf}, and f from one representative malignant (left column) and benign patient (right column). The malignant lesion exhibited lower values in all parameters. The difference was more prominent in the CTRW parameters D_m and α and IVIM parameters D_{diff} and f. The malignant lesion also showed increased variance of the parameters within the ROI in comparison to benign lesion.

4.4.2 Comparison of Features

The assessment of the top 18 features with 95% confidence intervals of their relative importance are shown in Figure 3a. Of those features we examined, the top 8 features that carry the most weight in classification were determined as the median of β (βmedian), skewness of β ($\beta_{skewness}$),

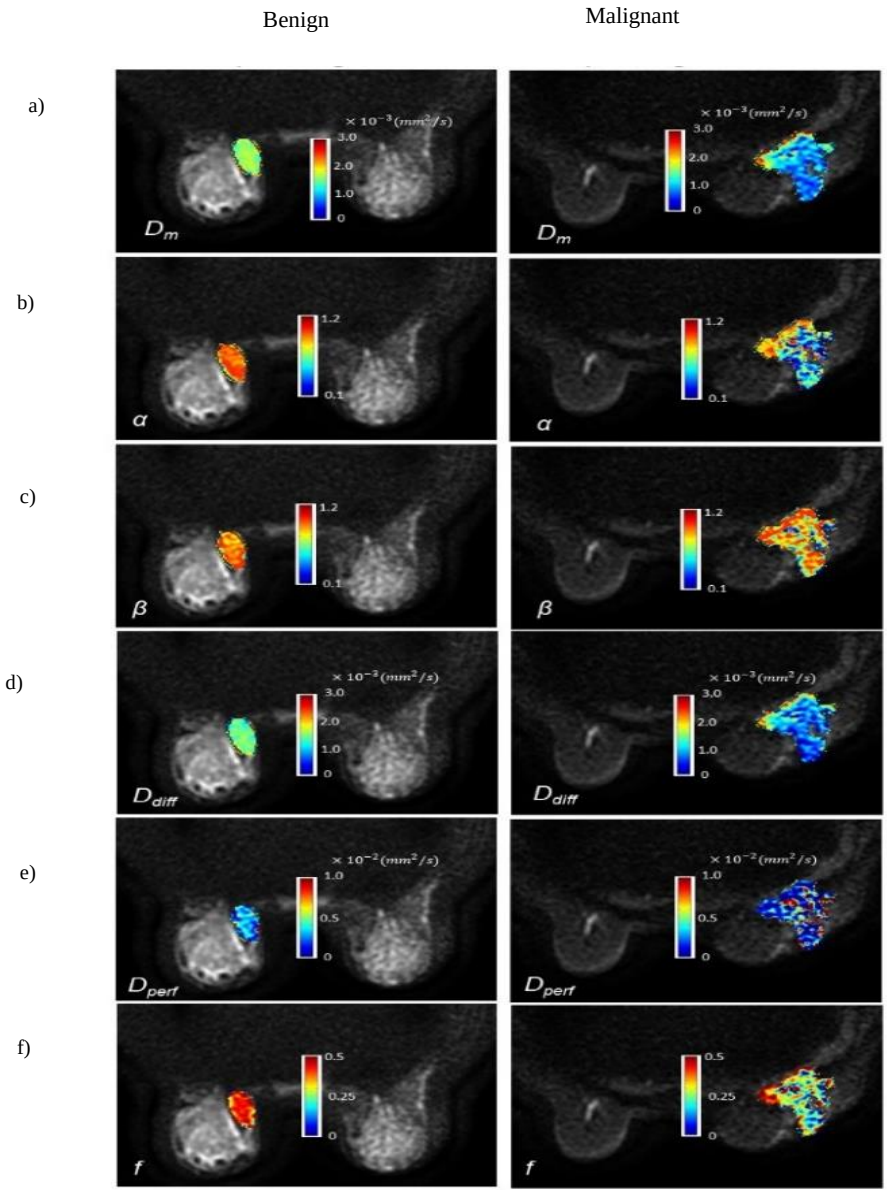


Figure 12: Diffusion parameter maps from a benign patient (left column) and malignant patient (right column). The CTRW parameter maps are given in a) D_m, b) α, and c) β; IVIM parameter maps are in d) Ddiif, e) D_{perf}, and f) f.

mean of β (β_{mean}), third quartile of f (f_{Q3}) , third quartile of $\mathrm{D_{diff}}$ (D_{diff}^{Q3}), kurtosis of $\mathrm{D_{perf}}$ ($D_{perf}^{kurtosis}$), third quartile of $\mathrm{D_m}$ (D_{pm}^{Q3}), and median of $\mathrm{D_m}$ (D_{m}^{median}). The boxplots and the descriptive statistics (mean and standard deviations) of these top features for both the benign and malignant groups are given in Figure Figure 13b. These top features were found to be statistically significantly different between the malignant and benign lesions (p-value$<$ .01) with respect to the Mann-Whitney-U test. In addition, although (β_{mean} and ($\beta_{kurtosis}$ are relatively the same in terms of feature importance (Figure Figure 13a), we found classifier performance increased with (β_{mean} and decreased with ($\beta_{kurtosis}$.

4.4.3 Number of Histogram Bins Comparison

Figures Figure 14a-h show the ROC curves generated by using a various number of histogram bins for linear SVM, radial basis function (RBF) SVM, GP, DT, RF, AB, GB, and NB classifiers, respectively. Varying the number of histogram bins yielded no significant impact on predictive performance, besides the GP classifier as seen in Figure Figure 14c. For the remaining classifiers, we determined the optimal number of histogram bins as the one that minimizes the variance within the ROC curves, which was found to be 80 for all classifiers with a p-value $<$ .05 with respect to the Mann-Whitney-U test. In the case of the GP classifier, features are extracted from a histogram that uses 40 bins.

4.4.4 Comparison of Classifier Performance

A comparison of the machine learning classifiers: linear SVM, RBF SVM, GP, DT, RF, AdaBoost, and NB, and GB, in the cross-validation stage is shown in Figures 5 and 6a-6c. The comparison between the ROC curves given in Figure Figure 15 shows that the GB classifier

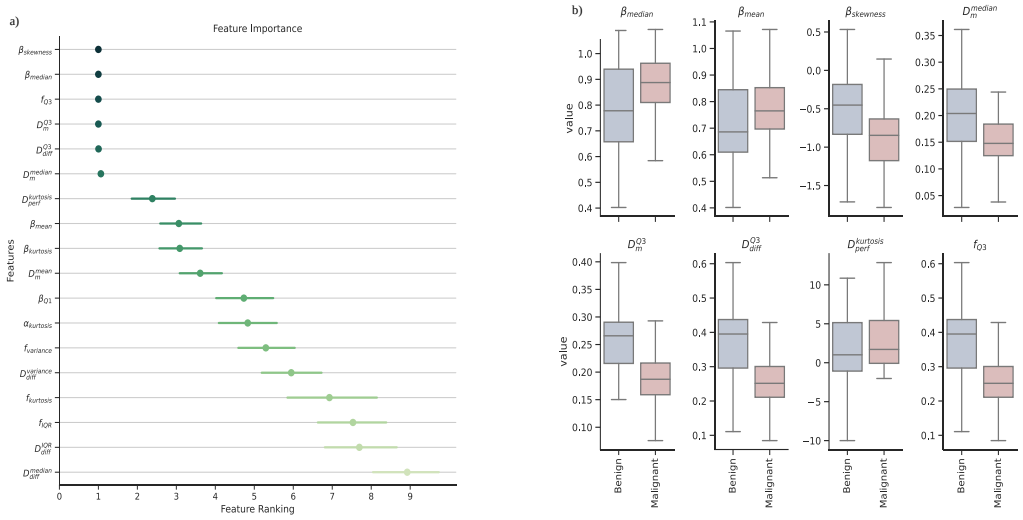


Figure 13: a) Feature importance plot of the top 18 features by feature rank with their respective confidence intervals. b) Paired box plots of the malignant and benign groups and the top 8 features selected from the feature importance paradigm.

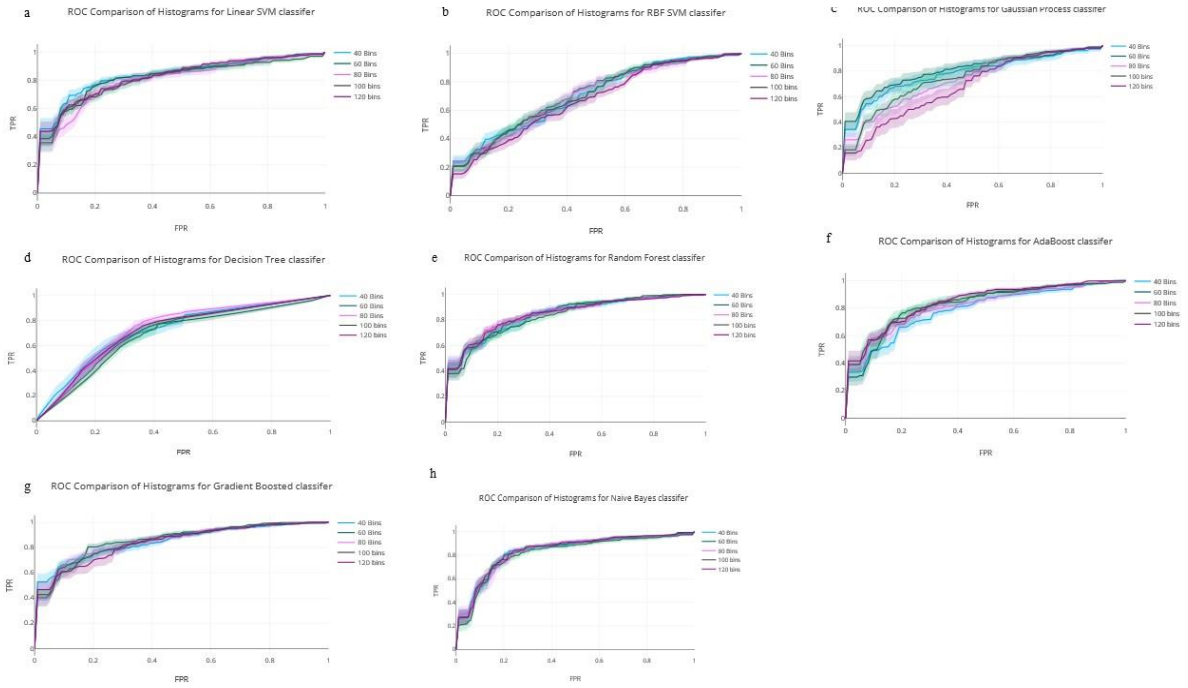


Figure 14: Comparison of histogram bin widths across the ROC of all classifiers: a) Linear SVM, b) RBG SVM, c) Gaussian Process, d) Decision Tree, e) Random Forest, f) AdaBoost, g) Gradient Boosted, and h) Naïve Bayes.

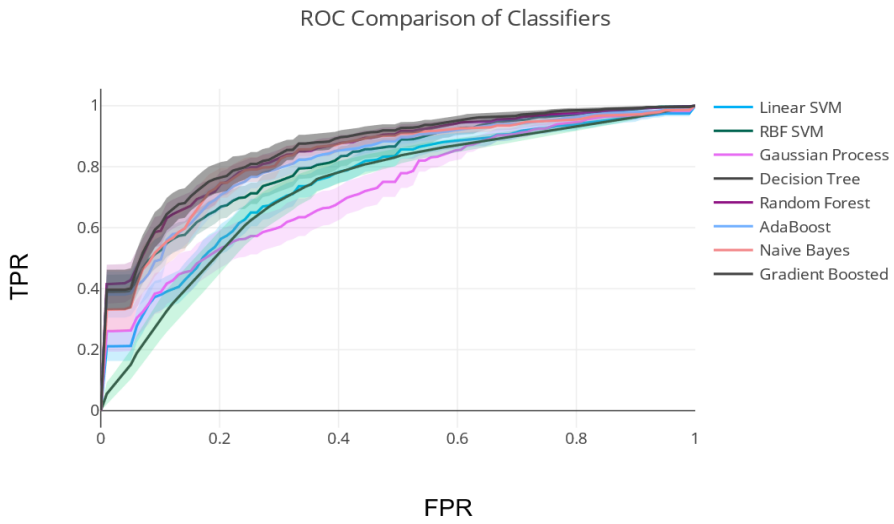


Figure 15: The ROC curves of the classifiers for differentiation between the benign and malignant breast lesions in the cross-validation stage. The curves were generated by using the optimal number of histogram bins for each classifier. The colored area denotes the standard error.

performs better than the other classifiers as it converges faster towards an optimal tradeoff between the true positive ratio (TPR) and the false positive ratio (FPR). The swarm plots given in Figures Figure 16a-c describe the performance metrics of all the classifiers. The GB and RF classifiers provided the most robust performance for at a p-value $< .05$ with respect to the Mann-Whitney-U test across all the metrics.

Similar results were observed in the performance metrics of the classifiers obtained during the testing stage. The box plots of AUC, accuracy, and F1 score of all classifiers are given in Figures Figure 17a-c. The GB classifier provided the best performance with the highest mean

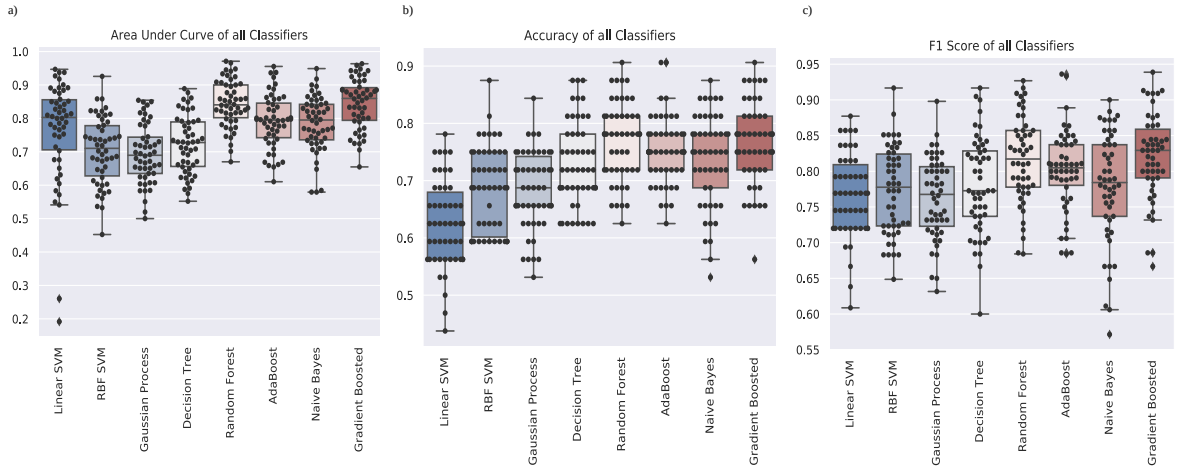


Figure 16: The swarm plots of AUC (a), accuracy (b), and F1 Score (c) metrics of each classifier during the cross-validation stage.

AUC of 0.942 (p-value$<$ .05) with a 95% confidence interval of [0.904, 0.981]. Additionally, the GB classifier produced a mean accuracy of 0.833 with a 95% confidence interval of [0.8, 0.9] and a mean F1 score of 0.872 with a 95% confidence interval of [0.824, 0.967].

4.5 Discussion

We have shown that a combination of quantitative features from the parameters of the CTRW and IVIM diffusion models obtained from a full b-value spectrum can accurately discriminate malignant and benign breast lesions with a machine-learning-based approach. Specifically, we have demonstrated that a combination of the model parameters aid in the discrimination of breast

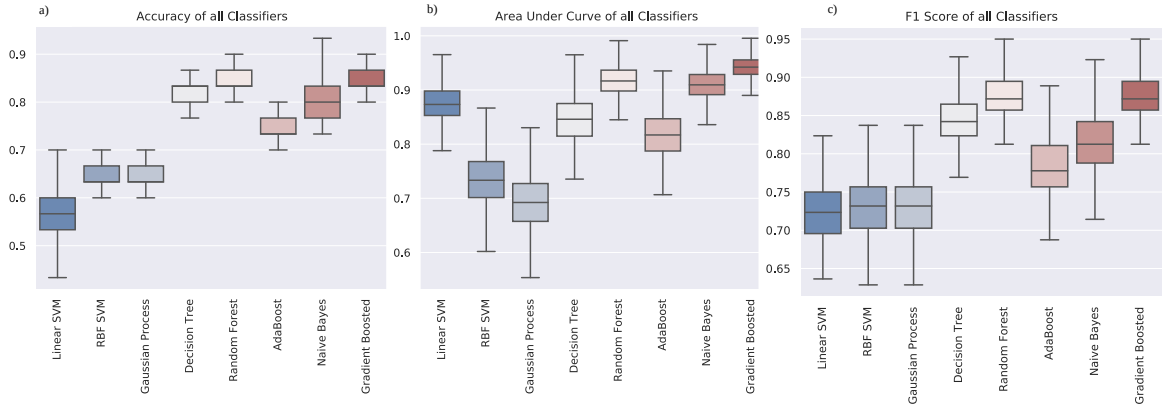


Figure 17: The box plots AUC (a), accuracy (b), and F1 Score (c) metrics of each classifier during the testing stage.

lesions with their sensitivity to probe a specific tissue property such as vascularity, heterogeneity, or cellularity. Altogether, a multi-modal DWI approach highlights the importance of capturing the distinct statistical distributions of the voxels in the parameter maps.

To use the CTRW and IVIM parameters for classification, we relied on quantitative features generated from statistical histograms, since traditional quantitative features are not applicable as the voxel values are analogous to multivariate continuous distributions rather than voxel intensity66. In total, we extracted nine reproducible quantitative features generated for each of the six parameter maps. Our analysis showed that there is a relationship between the quantitative features β (βmedian), skewness of β ($\beta_{skewness}$), mean of β (β_{mean}), third quartile of f (f_{Q3}) , third quartile of D_{diff} (D_{diff}^{Q3}), kurtosis of D_{perf} ($D_{perf}^{kurtosis}$), third quartile of D_{m} (D_{pm}^{Q3}), and median of D_{m} (D_{m}^{median}) and lesion histology such that the model can differentiate between benign and malignant lesions. We showed that the combination of quantitative features with a GB classifier provides a robust classification performance of lesion malignancy with a mean accuracy of 0.833. As lesion heterogeneity greatly influences patient diagnosis (237), a comprehensive analysis of multi-parametric information provided by the IVIM and CTRW models offers a unique perspective to the characterization of underlying breast tissue as well as a powerful framework for discrimination of lesion malignancy.

The most important quantitative feature for differentiating malignant and benign lesions in our study was the third quartile value of the CTRW model's diffusion coefficient parameter, D_{m}, which was found to be lower in malignant lesions than the benign ones. This outcome is consistent with previous studies(237; 194) which showed that malignant breast lesions exhibit lower ADC

values due to the decreased water diffusion. Our study, however, differs from earlier studies in its model assumption, use of higher b-values, and evaluation of multiple features obtained from diffusion coefficient for discriminating benign and malignant lesions. Our results showed that the third quartile of D_m, which may have a higher sensitivity to detect focal regions of higher cellularity, overperformed the mean D_m, that was not among the top performer features. The other two CTRW model parameters, α and β, on the other hand, have not been investigated in breast before. Mean, median, and skewness of the spatial heterogeneity parameter, β, have been found to be sensitive to changes in malignant breast lesions. β's sensitivity to intra-voxel tissue heterogeneity, which is known to increase in malignant tissue, has been reported in previous studies using CTRW and its predecessor fractional order calculus model (210; 115). Although quantitative features from β were among the top performers, the exact biophysical underpinning for the favorable performance of β is unknown; and should be investigated in future studies.

Our analysis of the IVIM model showed that kurtosis of the pseudo-diffusion coefficient, D_{perf}, the third quartile of the perfusion fraction, f, and the third quartile of diffusion coefficient, D_{diff} were among the top performing features to discriminate benign and malignant lesions. The features from f and D_{diff} were significantly higher in benign lesions (p-value $<$ .05), whereas features from D_{perf} was significantly higher in malignant lesions (p-value $<$ .10). Similar studies have reported the use of features from f and D_{diff} to discriminate breast lesions (115). Unlike our results, these studies showed features generated from f were higher in malignant cases. Our results are the reverse, and this is attributed to the max-min normalization process in our model. Specifically, the area of the malignant lesions in our dataset is significantly larger (p-value $<$ .01)

and f map parameter values in the malignant lesions are concentrated to have lower values (i.e. high skewness). Therefore, the max-min normalization further increases the influences of the lower values in the f parameter maps in the malignant cases. Before normalization, the values in the f parameter maps in our study were consistent with established literature (113; 115). The underlying biological implications from the IVIM parameter maps indicate that only certain areas of a lesion are malignant, and as a result have higher vascularity. The different regions of a lesion having different properties recapitulates cancer causes lesion heterogeneity. Moreover, this posits that size of a lesion is a significant confounder that influences the distribution of values within the parameter maps.

It is worth noting that no feature alone performed better than random when using the GB classifier (or any other classifier), which indicates that a combination of quantitative features provides a superior characterization of a lesion. This is consistent with previous studies which showed improved model performance with a combination of quantitative features from DWI model parameters (113; 114; 115) in comparison to a single feature. Further research to examine the distribution at the individual voxel level as distinct features themselves could elucidate even more the nuance differences between the parameter maps and lesions. Hence, analysis of diffusion parameter maps offers the potential to represent breast lesions from another prescriptive and further subtype patients, which can be of an interest in precision medicine.

With computational power easily accessible, we used multiple machine learning algorithms with the extracted quantitative features. Vidic et. al (113) used a RBF SVM classifier with a combination of four features from IVIM model and ADC to classify breast lesions with an

accuracy 0.96. The RBF SVM classifier using our quantitative features from the IVIM and CTRW models performed poorly with a mean accuracy of only 0.563. The training set accuracy, however, was higher with a mean of 0.734, indicating that the SVM classifier is prone to overfitting with these quantitative features. The difference in our results and Vidic et. al (113) is suspected to be caused by batch effects due to acquisition and fitting strategies. Our results showed that the GB and RF classifiers are the most robust in classification performance as they provide the highest metrics for F1-score and AUC. Moreover, neither of the classifiers underfit or overfit as the error in the independent test set is similar to the error in the cross-validation set.

We must acknowledge the following study limitations. First, this study was a retrospective analysis of patients at a single institution with a small dataset. We circumvented this issue in this study by augmentation of our dataset through random transformations; and removed the chance of overfitting. Using a larger population would increase the statistical power of quantitative features as imaging biomarkers. Although our quantitative features are reproducible, we are uncertain if the GB classifier can account for variations in machines and imaging. This problem again can be solved by correcting for batch effects as well as acquisition and fitting protocols being pushed towards a more standardized process (115). Finally, our study only evaluated lesion features from a single representative slice. Although the CTRW model probes intra-voxel tissue heterogeneity, single-slice analysis may miss important features because of intra-lesion heterogeneity. A comparison of the CTRW model parameters between a volumetric lesion and a slice would validate the ability of the parameters to probe lesion heterogeneity. Similarly,

a histogram analysis of the lesion volume for all model parameters may potentially improve discrimination between malignant and benign lesions.

4.6 Conclusion

In conclusion, we showed that IVIM and CTRW model parameters can characterize the underlying tissue in a lesion. Specifically, our approach showed that the quantitative features obtained through modeling the IVIM and CTRW model parameters as histogram distributions can differentiate malignant and benign lesions. This showed the benefit of combining quantitative features that potentially explain lesion heterogeneity in contrast to using a single feature such as the mean or median. Finally, when these quantitative features are used in conjunction with a GBM classifier, they predicted lesion malignancy with a mean accuracy of 0.833. Therefore, a combining lesion characterization from multiparametric DWI models parameters can be a powerful tool for discriminating breast lesions.

CHAPTER 5

MAPPING OF LESION IMAGES TO SOMATIC MUTATIONS

This work is published in Mehta, Rahul and Lu, Yang and Muge Karaman. "Mapping of Lesion Images to Somatic Mutations." Data and Text Mining in Biomedical Informatics (DTMBIO 2020), 2020, Galway, Ireland.

5.1 Motivation

Targeted cancer therapies, those based on identifying the genetic makeup of a legion, are often more effective and can incur fewer side effects than traditional therapies (238). Unfortunately, the efficacy of these treatments is highly dependent on early detection and treatment and so delayed turnaround time of genetic analyses can result in catastrophic consequences for survival rate. To bypass the time-consuming process of biopsy and genetic analysis, computational models can exploit quantitative and qualitative information derived from medical images of lesions to obtain cancer imaging features[1] in order to predict the specific genetic markers present in the individual patient. These imaging features can then be leveraged to project patient prognosis and determine targeted treatment options.

A design challenge in computational models that predict genetic markers from imaging features is the the inherent imaging and genetic heterogeneity of cancerous lesions. Current models overcome this limitation by enforcing imaging features to discriminate between a specific

[1]For the remainder of the paper we refer them as imaging features

set or cluster of genetic markers (239; 56; 240). This design choice, however, incurs false positives in downstream tasks for patient prognosis (241). While a model may correctly predict the presence of a genetic marker, it is often the confluence of many genes that influence cancer progression (242). For example, patients recommended for EGFR TKI targeted therapy in Non-Small Cell Lung Cancer (NSCLC) became resistant to the treatment due to a distinct secondary mutation in the EGFR gene, EGFR-T790M (243). As cancerous lesions are from a combination of genetic and epigenetic changes within a patient, imaging features must predict all possible genetic markers (244) to accurately assess the correct treatment options for a patient.

The idea of exploiting computational cancer imaging features from lesion images for patient analysis goes back decades (245). The general framework for the task of prediction using imaging features is posited as: given a black box model (e.g., a neural network) that converts a medical image into a set of imaging features, we can map the imaging features onto a function (e.g., Softmax) which learns parameters that optimizes the prediction accuracy of a set of labels of interest. Following the same paradigm, we would like our model to map the image of a lesion onto a patient's full somatic mutation profile. The learning task is unfortunately, impeded by the overwhelming size of the output space, i.e., the somatic mutation profile.

Our goal of mapping a lesion image to a somatic mutation profile is analogous to many methods in domain mapping (246) where a high-dimensional input is mapped to a high-dimensional output as in text translation, image to image translation, and image captioning. Current approaches in these applications have leveraged deep latent variable models (247; 248; 61) to enforce a shared latent space (249) or a cyclical structure (250). The main objective in these models is

to minimize the loss when translating between domains such that model can recover the data from the original domain, i.e. domain A to domain B back to domain A. The key advantage of deep latent variable models is that the neural architecture can conserve complex correlations among different domains by constraining the loss functions. For example, in image captioning, the neural network architecture can conserve the shape features of a mountain across populations (height, peak, width), while also changing these correlations respective to mutable semantic characteristics such as "snowy" or "volcano." We can apply the same concepts in the biological domain, however, we must consider two specific challenges:

1) Many models featuring cancer imaging features use a single slice of a lesion, however, we cannot follow the same assumption since the exact location of the lesion biopsies are unknown. Additionally, lesion images come from multiple imaging modalities such as computerized tomography (CT) or magnetic resonance imaging (MRI). Our model must incorporate all of the lesion slices during inference, while also remaining invariant to the imaging domain.

2) Somatic mutation datasets, like most genetic datasets, are discrete and high-dimensional. Complexity is further increased due to sparsity, that is, the data are characterized by few frequently occurring mutations and a long-tail of rare mutations. Therefore, the model must use a function that mitigates the underfitting of the data.

In this paper, we construct, Lesion Point Cloud to Somatic Mutations, LLOST, with dual variational autoencoders (VAE) (60) where each encoder/decoder architecture represents the domains of interest: the lesion image and the somatic mutations. To include the challenges of the biological domain, each VAE has it's own architecture. For the lesion VAE, we use a

point-cloud encoder/decoder architecture, which allows us to use a compact representation of a volumetric lesion. For the somatic mutation VAE, we use a Negative-Binomial likelihood to model the sparsity and high dimensionality of the dataset. The two VAEs are coupled together with a single invertible neural network conditioned on the cancer type to unify the two domains in a shared latent space. Each VAE also consists of their own domain specific latent space with conditional normalizing flows priors as a way to model the complexity of the two very different distributions (251; 252). The main idea is to use one domain to generate the other by using a series of normalizing flows conditioned on features learned in the shared latent space. Hence, by virtue of the LLOST's framework, we can transfer features from the lesion domain via the shared latent space to create a mutation specific latent space from the conditional prior. The concatenation of the two latent spaces then generate a prediction of how many times each distinct gene is mutated i.e., a full somatic mutation profile of a patient.

Our choice of utilizing somatic mutation profiles in comparison to other cancer genetic markers is that the somatic mutation profiles provide two immediate applications for clinicians. One use for clinicians is the identification of co-occurring somatic mutations of interest for early targeted treatment such as immunotherapies (253). The other is to predict a patient's tumor mutational load (TML), a sum of the total number of mutations in a lesion, which current research has proposed as a potential biomarker for determining patient prognosis and sensitivity to targeted treatments (254).

To analyze LLOST, we use the somatic mutation and lesion imaging data from The Cancer Genomic Archive and The Cancer Imaging Archive, respectively, for inference and prediction.

We analyze the predicted somatic mutation profiles using perplexity and distance measurements. A ramification of our work is a better assessment of treatment options for patients at earlier stages of diagnosis, as well as the possibility of further downstream tasks such as prognosis of patients using tumor mutational load.

5.2 Background

In this section we discuss the necessary background on the components of our model: point cloud data, high dimensional discrete data, and normalizing flows. We use uppercase notation to describe the lesion and mutation domains, M and I, respectively. We use bold lowercase notation to describe latent space parameters, where the subscripts depict a specific domain (M,I) or sample (n), and unbolded lowercase notation to denote values assumed by the variables.

5.2.1 Point Clouds

Point clouds have gained traction in the analysis of 3D objects with the increasing availability of 3D sensors and acquisition technologies. Point clouds are particularly amenable in describing arbitrary shapes as each point is simply a (x, y, z) coordinate in Euclidean space, thereby offering a compact representation of surface geometry. Recent deep learning models are built upon the nascent ideas introduced in PointNet (140), which operate directly on the point cloud, and therefore follow a more data-driven approach of extracting features. Several studies have extended PointNet for different applications in shape completions, 3D segmentation, and 3D classification, and point cloud generation (255). The synergistic component within these models is the feature vector that aggregates global and local features of the point cloud. Depending on

the neural network architecture, the feature vectors can include descriptions of shape, volume, surface topology, object geometry, and the relationship between individual points.

5.2.2 Normalizing Flows

Normalizing flows (NF) (251; 248) is a type of likelihood based deep latent variable model that aims to map a simple base density, $p(\epsilon)$, to a complex density, $p(z)$, through several invertible parametric transformations with tractable Jacobians. An example of such a density is:

$$p_{\boldsymbol{z}}(\boldsymbol{z};\theta) = p_{\epsilon}(f_{\theta}(\boldsymbol{z}))\left|\det\frac{df_{\theta}(\boldsymbol{z})}{dz}\right| \tag{5.1}$$

where $f : \mathcal{R}^D \mapsto \mathcal{R}^D$ is an invertible function with parameters θ.

Due to their simplicity in inference using maximum likelihood estimation, NF have become a popular choice to enrich the VAE posterior for image generation (252) or as a prior in VAE for discrete sequences (256).

Another advantage of NF is it's inherent invertible construction that allows training in the forward and the reverse directions. Recently, (257) exploited this mechanism for inverse sampling problems when both distributions are arbitrarily complex. The model, however, must be able to sample from both distributions and evaluate the forward and and backward processes. These requirements are easily satisfied using neural network architecture based on current literature such as in RealNVP (251) and Autoregressive Flows (258). Since our data comes from two complex distributions, NF provide an intuitive method to move between distributions and uncover distinct properties of the latent space.

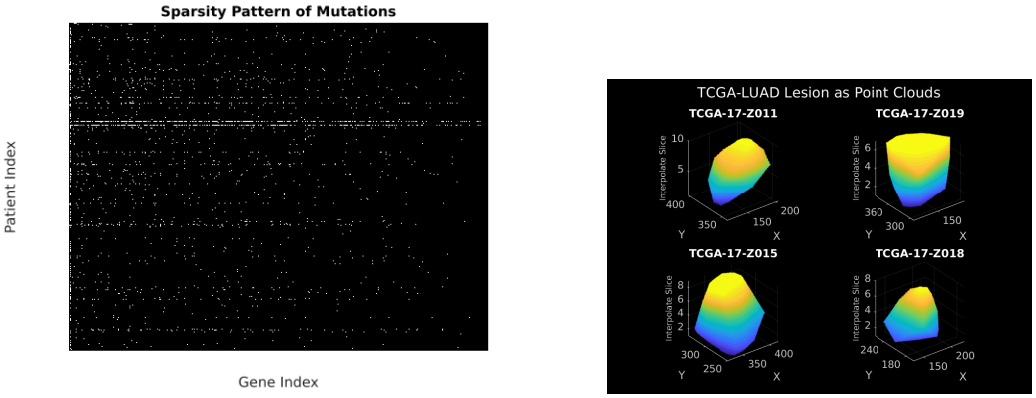
5.2.3 High Dimensional Discrete Data

Parameter estimation of discrete count data is often done using latent variable models (LVM) $p(x) = \int_z p(x|z)p(z)dz$ where $p(x)$ is the distribution of the original data, $p(z)$ is the prior for the latent distribution and $p(x|z)$ is some probability distribution. A caveat of LVMs is that the practitioner must determine which distributions to use for $p(z)$ and $p(x|z)$ to correctly capture the underlying distribution of the data (259).

Thanks to numerous genomics studies (260), the Negative-binomial (NB) distribution has emerged as the probability distribution of choice to estimate count data often seen in genetic datasets. We exploit the inference methods of VAE to learn the parameters of the NB distribution, (261) by rewriting the VAE generative process as:

$$\begin{aligned} \mathbf{z_M} \sim \mathcal{N}(0,1); \mathbf{r_M} \sim exp(f_{\theta^r}(\mathbf{z_M})); \\ \mathbf{p_M} \sim \frac{1}{1+exp(-f_{\theta^p}(\mathbf{z_M}))}; M_n \sim NB(\mathbf{r_M}, \mathbf{p_M}) \end{aligned} \tag{5.2}$$

That is, we draw a Gaussian random vector $\boldsymbol{z_M}$, pass it through a neural network architecture parameterized by f_{θ^r} to generate the vector for the dispersion parameter, $\boldsymbol{r_M}$, via an exponential function. Similarly, the vector for the mean parameter, $\mathbf{p_M}$, is generated by passing $\boldsymbol{z_M}$ through a neural network architecture parameterized by f_{θ^p} via a sigmoid function. Finally, we can generate a sample M_n from the NB distribution with parameters $\boldsymbol{p_M}$ and $\boldsymbol{r_M}$. We can also use the the NB parameters to generate binary labels, i.e., we would instead predict if a specific somatic mutation occurs in a patient rather than the number of times it occurs. The likelihood follows a Bernoulli distribution:



(a) (b)

Figure 18: a) Shows the amount of sparsity in the TCGA Pan Cancer dataset, where the total number of genes is 21900 (column) and number of patients is 10295 (rows). Each white dot represents a the existence of a somatic mutation. b) Shows examples of lung cancer lesions from different patients as points clouds respective of the axial plane. The points were generated by interpolating along with z-axis, with respect to the individual lesion slices.

$$M_n^b \sim Bernoulli(1 - (1 - \mathbf{p_M})^{\mathbf{r_M}}) \tag{5.3}$$

As discussed earlier, an important design choice in LVMs is the choice of the distribution of the prior. In the previous example of equation Equation 5.2 we used a VAE with a univariate Gaussian prior for the latent space, but that can become incongruous as datasets become more complex (259). There are several priors available to us to increase the effectiveness of LVMs. What we need acknowledge, is the ability to perform efficient inference and the flexibility to learn the distributions of two very different datasets.

5.2.4 Related Work

Early techniques for multimodal learning or domain mapping framed the learning paradigm as information retrieval, finding the best domain match from a pool of embeddings in a shared latent space (262). These are now superceded by deep latent variable models (DLVM) such as VAE or General Adversial Networks (247) (GAN) that scale to larger datasets, but they still employ a strategy of sharing a latent space. A popular choice for multimodal learning in DLVMs is the conditional variational autoencoder (CVAE) where a shared latent space is generated by concatenating the observed labels with the latent space.

A limitation of CVAE is that the learned latent space tends to encourage a distribution that encompasses dominant patterns of the data. Since labels are not available during test time, a CVAE model will only predict labels that approximate the original training distribution. Many authors circumvent this issue by modifications of CVAE via modification of the prior for the latent space (263; 258). While the results of various CVAE models are impressive, sharing the latent space removes domain specific features (264) and leads to limited diversity in the latent structure. To counteract this issue, (265) created a separate shared latent space along with a domain specific latent space to generate a more diverse latent structure for image captioning. Similarly, (266) used a separate embedding layer that better model semantically similar texts. Inspired by these models, LLOST creates a distinct shared latent space, which is augmented by conditioning it on the cancer type.

5.3 Dataset

We use the Pan Cancer Dataset from TCGA (267), which contains the unique somatic mutation profiles for 10295 patients. We convert this into a matrix format where each row is a patient and each columns is a unique gene so that the total number of columns is 21332. A snapshot of the matrix is shown in Figure 18a. We then find the corresponding lesion image(s) of each patient from TCIA database (from any modality). If a cancer type had less than 10 patient samples we did not include it in our dataset. The final dataset consists of 1342 patients from 18 cancer types each with a lesion image and a corresponding somatic mutation profile as shown in Table Table VI.

5.3.1 Lesion as a Point Cloud

A key decision in building our dataset concerned the context of how our model could extract diverse and specific information from the lesion image regardless of cancer type. There is no explicit way to compare lesions as they are are highly irregular and heterogeneous among samples. We propose to use point clouds as a way to model the intrinsic antisotropic and heterogeneous nature of lesions. Point clouds offer a rich interpretation of lesions as discussed in Section 5.2.1, and have several other attractive properties that pertain to the medical imaging dataset. In particular, since point clouds are a set of un-ordered points, they are modality independent (CT or MRI) compared to a set of pixels where each pixel intensity depends on the imaging modality. Furthermore, a point cloud lesion decreases the computational footprint since each lesion is now represented as a 2D matrix in comparison to a volumetric lesion which is a 3D matrix.

To create the lesion point clouds, we first extract the lesions from individual slices using the segmentation labels provided by TCIA. If segmentation labels were not available, a radiologist delineated the lesions within the slice. Since lesions come from a different number of scanners and modalities, we transform the lesions onto their real-world coordinates using the information stored within their respective DICOM files. As shown in Figure 18b, the voxels within a lesion volume are then interpolated to create a point cloud following the model in (268). Finally, for a specific lesion point cloud I_n, each point in I_n is a set of coordinates from $\mathbb{R}^3$ uniformly sampled from the surface of the full lesion point cloud.

5.3.2 Somatic Mutations Representation

Somatic mutations are represented as a count matrix, M, where each row in M is a patient, and each column is a distinct gene. A single element in the matrix indicates the number of times a gene is mutated. This is analogous to Bag of Words for text datasets (269). In our case, the vocabulary, V, is the set of all the genes (21332), N is the total number of samples, and M_n is a vector of counts of the n^{th} sample. A binarized version of matrix, where each vector, M_n, indicates the occurrence of a matrix is shown in Figure 18a.

5.4 Model

Given the above machinery, we represent our full dataset as N patients, where each patient n is associated with a point cloud $\mathbf{I_n}$, a count vector $\mathbf{M_n}$, and one hot vector label, y identifying cancer type. Our goal is to learn the conditional distribution $p(M|I, y, \boldsymbol{z})$ where $\boldsymbol{z}$ is a stochastic latent variable.

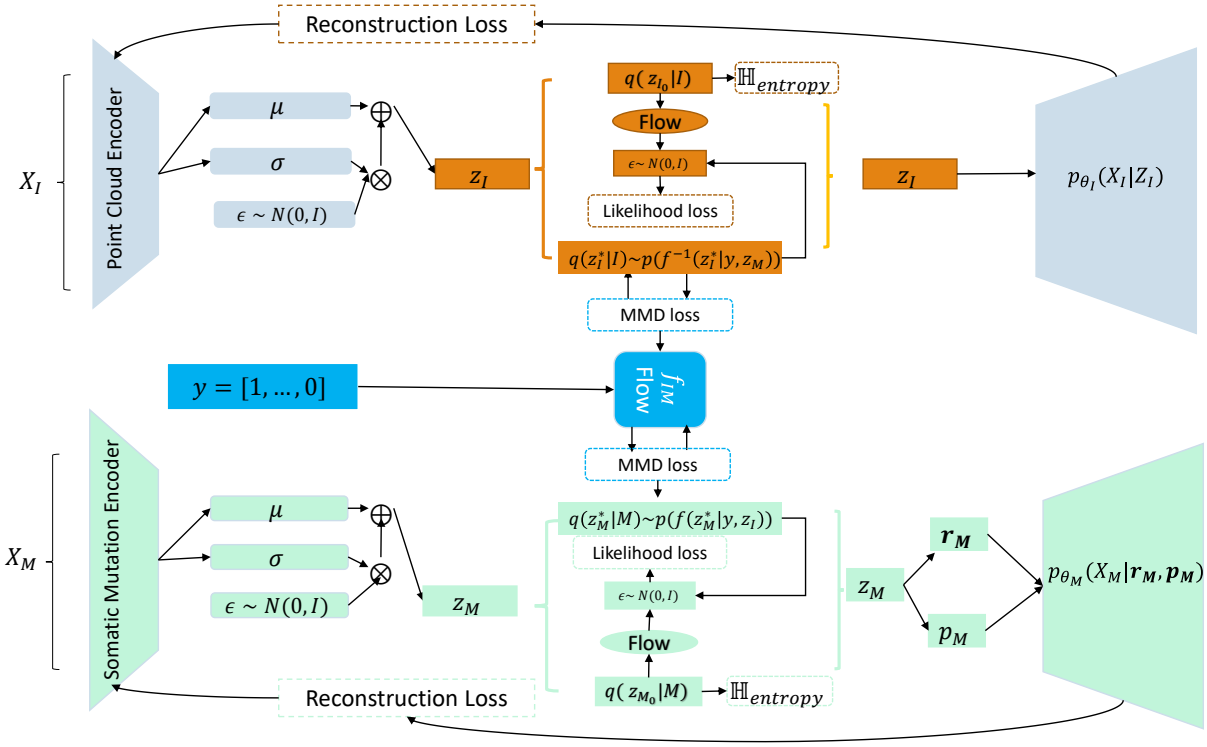


Figure 19: Model Architecture of LLOST. During training, the approximate posterior distribution of the domain specific embedding tries to match the true posterior with a learnable prior conditioned on the shared latent spaces. The shared latent space is trained by matching the distribution of the domain, so that it maps shared embeddings to domain specific embeddings. The model is trained bidirectionally to maximize the ELBO, which is a sum of reconstruction loss, the KL divergence of the conditional NF, and the MMD loss of the shared latent space. For clarity we drop the subscripts referring to the individual neural network parameters.

5.4.1 Lesion Point Cloud to Somatic Mutations

The Lesion Point Cloud to Somatic Mutation (LLOST) learns the conditional distribution $p(M|I, y, \boldsymbol{z})$ using two VAEs for each domain, X_I (lesion) and X_M (mutation). Each VAE embeds the data into a lower dimensional latent space $\boldsymbol{z_I}$ and $\boldsymbol{z_M}$, respectively. Instead of using a single latent for each domain, we propose to use two domain specific latent spaces and a shared latent space. We denote $\boldsymbol{z_{M_0}}$ for the mutation specific latent space, $\boldsymbol{z_{I_0}}$ for the lesion specific latent space, $\boldsymbol{z_M^*}$ for the shared mutation latent space, and $\boldsymbol{z_I^*}$ for the shared lesion latent space. So $z_I = [\boldsymbol{z_I^*}, \boldsymbol{z_{I_0}}]$ and $z_M = [\boldsymbol{z_M^*}, \boldsymbol{z_{M_0}}]$. The conditional distribution of the domains is then $p(M, I|y, \boldsymbol{z_{M_0}}, \boldsymbol{z_M^*}, \boldsymbol{z_{I_0}}, \boldsymbol{z_I^*})$ and is learned by approximating the latent variables using a variational posterior $q(\boldsymbol{z_M^*}, \boldsymbol{z_I^*}, \boldsymbol{z_{I_0}}, \boldsymbol{z_{M_0}}|M, I, y)$.

To learn our conditional distribution we must first define our priors. Whereas the original VAE uses a fixed standard Gaussian prior for the latent space, we instead use a trainable prior via a conditional NF to model the complexity of the distributions of each domain (256). Having a trainable prior allows us to adapt the shape of the prior based on the data, and avoids fewer modeling assumptions (248; 270). Using Equation Equation 5.1, our conditional NF prior for the image domain is then constructed as $p_{\eta_I}(\boldsymbol{z_{I_0}}|\boldsymbol{z_I^*}) = p_\epsilon(f_{\eta_I}(\boldsymbol{z_{I_0}}|\boldsymbol{z_I^*}))\left|\det\frac{df_{\eta_I}}{dz}\right|$, where η_I are the parameters of the neural network. This formulation allows the model to push the prior towards matching the approximate posterior during training. Therefore during test time, LLOST samples from the trained prior $p_{\eta_I}(\boldsymbol{z_{I_0}}|\boldsymbol{z_I^*})$ and the shared latent space $p(z *_{I_0} |y)$, unlike CVAE that samples from the variational posterior for prediction tasks.

We take advantage of the conditional NF architecture again in the shared latent space, where the base distribution is instead the latent space $q(\boldsymbol{z}^*_{\boldsymbol{M}}|M)$. The shared latent space is modeled as a single invertible neural network, $f_{\theta_{IM}}$ (61; 257) conditioned on the cancer type label, y. When optimized correctly, the encoded distribution $q(\boldsymbol{z}^*_{\boldsymbol{M}}|M)$ will match the distribution of the shared latent space $p(f_{\theta_{IM}}(\boldsymbol{z}^*_{\boldsymbol{I}}|y))$. Intuitively, we are taking advantage of the invertibility of the NF architecture to let the cancer label type guide the distribution of the image domain to the distribution of the mutation or vice versa.

We can then predict a somatic mutation profile using these steps:

1. Input the lesion point cloud into X_I to generate $\boldsymbol{z_I}$, which we use to create $\boldsymbol{z^*_I}$
2. The network of the shared latent space $f^{-1}_{\theta_{IM}}$ given the cancer type label is then used to map $\boldsymbol{z^*_I}$ to $\boldsymbol{z^*_M}$
3. Generate domain specific latent space $\boldsymbol{z_{M_0}}$ conditioned on $\boldsymbol{z^*_M}$ (from the above step) via the trained conditional NF prior $p(\boldsymbol{z_{M_0}}|\boldsymbol{z^*_M})$
4. Generate the parameters $\boldsymbol{r_M}$ and $\boldsymbol{p_M}$ of the NB likelihood using $\boldsymbol{z_M} = [\boldsymbol{z_{M_0}}, \boldsymbol{z^*_M}]$ from the previous steps
5. Predict the mutation profile $M_n \sim NB(\boldsymbol{r_M}, \boldsymbol{p_M})$

Since the two domains are independent with respect to the latent spaces, we can summarize the generative model above as a joint distribution and factorize:

$$\begin{gathered} p(M, I, \boldsymbol{y}, \boldsymbol{z_{I_0}}, \boldsymbol{z_{M_0}}, \boldsymbol{z_I^*}, \boldsymbol{z_M^*}) = \\ p_{\theta_M}(M|\boldsymbol{z_{M_0}}, \boldsymbol{z_M^*}) p_{\theta_I}(I|\boldsymbol{z_{I_0}}, \boldsymbol{z_I^*}) \\ p_{\eta_I}(\boldsymbol{z_{I_0}}|\boldsymbol{z_I^*}) p_{\eta_M}(\boldsymbol{z_{M_0}}|\boldsymbol{z_M^*}) p_{\theta_{IM}}(\boldsymbol{z_I^*}|\boldsymbol{y}) p_{\theta_{IM}}(\boldsymbol{z_M^*}|\boldsymbol{y}) \end{gathered} \tag{5.4}$$

where $p(\boldsymbol{z_{I_0}}|\boldsymbol{z_I^*})$ and $p(\boldsymbol{z_{M_0}}|\boldsymbol{z_M^*})$ are the conditional NF priors. The distributions, $p(\boldsymbol{z_I^*}|\boldsymbol{y})$ and $p(\boldsymbol{z_M^*}|\boldsymbol{y})$, are of the shared latent space. The subscripts η and θ indicate the distinct network parameters for each domain. Similarly, we can factorize the approximated latent posterior:

$$\begin{gathered} q(\boldsymbol{z_M^*}, \boldsymbol{z_I^*}, \boldsymbol{z_{I_0}}, \boldsymbol{z_{M_0}}|M, I) = \\ q_{\phi_M}(\boldsymbol{z_{M_0}}|M) q_{\phi_I}(\boldsymbol{z_{I_0}}|M) \\ q_{\phi_M}(\boldsymbol{z_M^*}|M) q_{\phi_I}(\boldsymbol{z_I^*}|I) \end{gathered} \tag{5.5}$$

Note that the cancer type label, y, is not involved in the latent posterior approximation since it is not involved in the encoder architecture of the model.

5.4.2 Variational Objective

We first consider the standard learning objective of the VAE with a Gaussian prior and a latent space z_M. This objective is optimized by maximizing the Evidence Lower Bound (ELBO) with encoder and decoder parameters θ and ϕ respectively:

$$ELBO(\theta, \phi) = \mathbb{E}_{p^*(M)} \mathbb{E}_{q_\phi(\boldsymbol{z_M}|M)}[\log_\theta p(M|\boldsymbol{z_M}) + \log p(\boldsymbol{z_M}) - \log q_\phi(\boldsymbol{z_M}|M)]. \tag{5.6}$$

The first term is the reconstruction loss of the original data, $p^*(M)$. The last two terms act as a latent space regualizer, which is also referred to as the Kullback Libler (KL) divergence between the approximate posterior distribution $q_\phi(\boldsymbol{z}|M)$ and the prior $p(z)$. We extend this to our model and the resulting posterior factorization is:

$$\begin{aligned} ELBO(\boldsymbol{\Omega}) = \\ &\mathbb{E}_{p_{f_{IM}}(\boldsymbol{z^*_M}|y) q_{\phi_M}(\boldsymbol{z_{M_0}}|M)}[\log_{\theta^r, \theta^p}(M|\boldsymbol{z_{M_0}}, \boldsymbol{z^*_M})] \\ &\mathbb{E}_{p_{g_{IM}}(\boldsymbol{z^*_I}|y) q_{\phi_I}(\boldsymbol{z_{I_0}}|I)}[\log_{\theta^I}(I|\boldsymbol{z_{I_0}}, \boldsymbol{z^*_I})] \\ &+\mathcal{L}_{\boldsymbol{z^*_M}} + \mathcal{L}_{\boldsymbol{z^*_I}} \\ &-KL[q_{\phi_M}(\boldsymbol{z_{M_0}}|M, \boldsymbol{z^*_M})|p_{\eta_M}(\boldsymbol{z_{M_0}}|\boldsymbol{z^*_M})] \\ &-KL[q_{\phi_I}(\boldsymbol{z_{I_0}}|I, \boldsymbol{z^*_I})|p_{\eta_I}(\boldsymbol{z_{I_0}}|\boldsymbol{z^*_I})] \end{aligned} \tag{5.7}$$

Thus, the overall objective can be considered a hybrid of the standard VAE objective and maximum likelihood estimation (MLE) with respect to the neural network parameters $\boldsymbol{\Omega}$. The first two terms in Equation Equation 5.7 are the reconstruction errors of the original data. The middle two terms are losses of matching the shared latent space with latent space from the encoder discussed in Section 5.4.3. Lastly, the KL divergence of the domain specific latent space is a MLE with respect to the conditional NF prior with an entropy regularizer as shown in

Equation Equation 5.8. Using Equation Equation 5.1 we can rewrite the KL divergence for the lesion domain as:

$$\begin{aligned} &KL[q_{\phi_I}(\boldsymbol{z_{I_0}}|I, \boldsymbol{z_I^*})|p_{\eta_I}(\boldsymbol{z_{I_0}}|\boldsymbol{z_I^*})] = \\ &-\mathbb{E}_{q_{\phi_I}(\boldsymbol{z_{I_0}}|I)}[p_\epsilon(f_{\eta_I}(\boldsymbol{z_{I_0}}|\boldsymbol{z_I^*})) \\ &+\log(\det(\frac{df_{\eta_I}}{dz}))] + \mathcal{H}(q_{\phi_I}(\boldsymbol{z_{I_0}}|I)). \end{aligned} \tag{5.8}$$

The last term, $\mathcal{H}$, is the entropy of the approximate posterior distribution from the encoder. The first and second terms are the log-likelihood of z_{M_0} under the prior distribution modeled by the conditional NF. Note that we can simply sample from $q_{\phi_I}(\boldsymbol{z_{I_0}}|I)$ to calculate the entropy.

5.4.3 Optimization of Shared Latent Space

To encourage sharing of information, we want the shared latent space, $p(\boldsymbol{z_I^*}|y)$, to match the generated latent space $q(\boldsymbol{z_I^*}|I)$ after the flow transformation $f^{-1}_{\theta_{IM}}$. Instead of using the KL divergence to optimize the loss, we use the Maximum Mean Discrepancy (MMD). The MMD divergence confers two advantages. First, we can explicitly fit the distribution of $\boldsymbol{z_I^*}$ without any assumptions about a prior. Secondly, if convergence is reached, we can sample from the shared latent space after the respective flow transformation and disregard the need to generate random samples from a fixed prior during test time. The only requirement is that the size of $|\boldsymbol{z_M^*}|$ matches the size the $|\boldsymbol{z_I^*}|$. This optimization procedure is similar to ones used in Adversarial Autoencoders (271), where the divergence function is instead the Jensen Shannon divergence, which scales to higher dimensions. Since we can control the size of the shared latent space,

we are not limited by the scaling issues of MMD. We achieved best results with the inverse multiquadtratic kernel:

$$\mathcal{L}_{\boldsymbol{z}^*_M} = k(\boldsymbol{z}^*_{\boldsymbol{M}}, \tilde{\boldsymbol{z}^*_{\boldsymbol{M}}}) = 1/(1 + \left\|(\boldsymbol{z}^*_{\boldsymbol{M}} - \tilde{\boldsymbol{z}^*_{\boldsymbol{M}}})/h\right\|^2) \tag{5.9}$$

where $\tilde{\boldsymbol{z}^*_{\boldsymbol{M}}}$ is the distribution from the forward process of f_{IM}. This is also applied in the reverse direction to calculate $\mathcal{L}_{\boldsymbol{z}^*_{\boldsymbol{I}}}$. We can conceptualize this as mapping the shared latent space to domain specific embeddings of the somatic mutation profile.

5.4.4 Implementation Details

Figure Figure 19 shows the architecture and distinct latent spaces of LLOST. We adopt the encoder-decoder structure proposed in (272) for the point cloud VAE with the size of the latent space $\boldsymbol{z_I}$ as 512. The encoder-decoder structure of the VAE for somatic mutations is a set of symmetric multilayer perceptron (MLP) with dimensions [1000, 500, 300] when using a NB-likelihood and [800, 500] for the Bernoulli likelihood, where the last dimension indicates size of the embedding space, $\boldsymbol{z_M}$.

Our model architecture for the latent spaces follows the NF architecture of realNVP (251), where one flow with a single affine coupling block for brevity is shown in Figure Figure 20. The input to the coupling layer is the partition of the input $z = [z_1, z_2]$. The output of one coupling layer is then $y_2 = z_2$ and $y_1 = z_1 \otimes exp(s(z_2, c)) + t(z_2, c))$ where functions s and t are neural networks, and c is the specific condition. Conditioning c onto s and t does not affect invertiblility, since the s and t are never inverted, therefore the functions can utilize any neural network

architecture. We use 12 flow steps each with two affine coupling blocks for domain specific latent space, and we use 24 flow steps with 3 coupling blocks for the shared latent space.

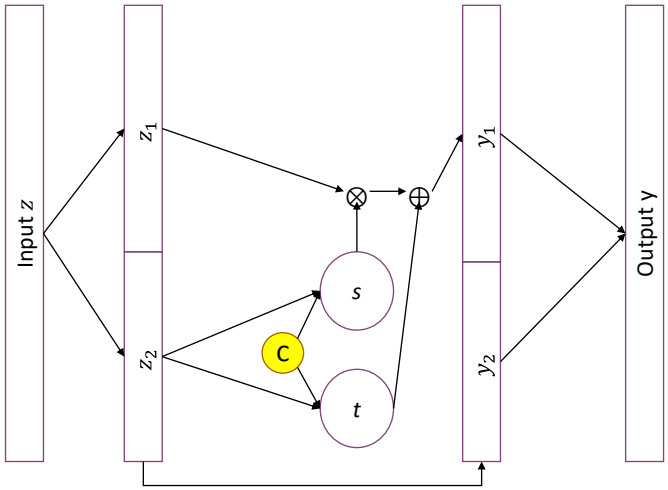


Figure 20: An example of a conditional coupling block where s and t are neural networks. Here c is the condition that is simply concatenated to each neural network.

To learn the parameters of the encoder, decoders, and the latent space we optimize the negative of Equation Equation 5.7 via bidirectional training. Before we update any parameters, we calculate the loss of the invertible networks in the forward and reverse directions given a sample from both domains. This technique encourages both domains to influence the parameters of the shared latent space and the learnable priors.

5.5 Experiments

We use 70% of the dataset to train, 15% to validate, and 15% to test our model for accurate reconstruction and prediction of a somatic mutation profile.

We compare two versions of LLOST, one with a NB likelihood, LLOST_{NB}, and another with the Bernoulli likelihood, LLOST_{B}, along with the three baseline CVAE models. For the baseline models, CVAE_r, uses the ResNet (273) architecture to extract imaging features from a single lesion slice with a NB likelihood decoder. The CVAE_p and CVAE_{pb} models both use the point cloud architecture from our model, but CVAE_p has an NB likelihood decoder and CVAE_{pb} has a Bernoulli likelihood decoder. We then follow the CVAE methodology by concatenating both the somatic mutation profile matrix and the cancer type matrix with the output of the encoders and decoders. The baseline models use a NF prior for the latent space. Testing is done by generating the latent embeddings for the lesion and then concatenating it with random samples from the latent space.

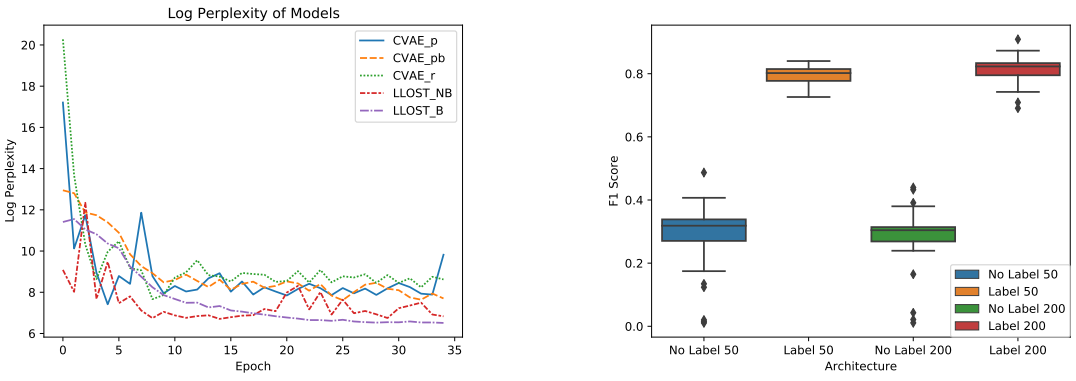
Our experiments on the shared latent space are done via an ablation study, where we remove the conditioning by the cancer type label. We also study the effect of increasing the width of the shared latent space.

5.5.1 Comparison Metrics

We use log perplexity, $-\frac{1}{N}\Sigma_n^N \frac{1}{L_n} \log p(M_n|z_{m_0}, z_m^*)$, to compare the goodness-of-fit of our models and the baseline models evaluated over 35 epochs. The perplexity metric is a function of the reconstruction error, $\log p(M_n|z_{m_0}, z_m^*)$, total number of mutations in a sample, L_n, and the number of the samples, N.

As our goal is the prediction of the count of somatic mutations, we use the root mean squared error (RMSE) and TML to measure the prediction error of the models with NB likelihoods. The RMSE shows the average error for the count of mutations for each distinct gene. TML is a relatively recent biomarker used for determining sensitivity to targeted treatments (274) and is simply the sum of all mutations within a sample. We calculate the prediction error of TML using a simple point estimate, $\text{TML}^* - \text{TML}$, where $*$ indicates our predicted value.

We use the F1-score and Positive Predictive Value (PPV) metrics to assess the Bernoulli likelihood model. Specifically, we use the F1-score to compare the experiments on the shared latent space. The PPV metric is used to determine the predictive accuracy of LLOST_B for specific cancer types. To calculate this metric we use 15% of the samples of a specific cancer type for testing, while the remaining samples are used for training.



(a) (b)

Figure 21: a) Log Perplexity (lower is better) of CVAE_p, CVAE_pb, CVAE_r, LLOST_r, and LLOST_B, viewed as a function of epochs b) A comparison of the F1 score during testing with two sizes of z_L^* or z_M^* and no label in the shared latent space.

5.5.2 Results

Figure 21a displays the overall log-perplexity scores over 35 epochs, where after the models are asymptotic and in the case of CVAE_p, the model begins to collapse. We observe that LLOST has the best performance in predicting the original somatic mutation profile. Specifically, LLOST_B is the best in recreating the original distribution of the somatic mutation profile. This also shows that learning the occurrence of a specific mutation is much easier compared to learning the count of a specific mutation, since CVAE_pb also out-performs the other CVAE baselines, which have a NB likelihood. When comparing the CVAE_p and CVAE_r, we see that the point cloud features provide additional information that aids the model to predict the counts of each mutation. The key insight is that the shared latent space along with domain specific embeddings allow for better predictive performance of the somatic mutation profiles.

In Table Table V we show the RMSE of the models with NB likelihoods. Once again, the baseline models do not have enough representation power in their latent space to accurately describe the underlying distributions of the two domains. As the CVAE is simply a concatenation of the encoder with the somatic mutation matrix, the approximate posterior distribution cannot scale to incorporate such a high dimensional dataset. Although LLOST_NB still has high RMSE, there is over a 50% increase in performance of predicting the TML. This supports our hypothesis that learning distinct and shared latent spaces for very different domains is beneficial for high dimensional prediction tasks.

We examine why the LLOST_NB has a high RMSE in Figure Figure 22. We observe the NB likelihood favors over-counting mutations. The average prediction error of TML is 54 in samples

Model	RSME
CVAE_r	855.45
CVAE_p	731.37
LLOST_NB	315.15

TABLE V: RSME of Models with the NB likelihood

with less than 400 TML. When qualitatively assessing the predicted somatic mutation profiles, we observe that genes with the most frequent mutations are predicted to occur almost at least once such as TP53, KRAS, and BRCA1. This pattern continues and is abundantly apparent when TML is larger than 1000, such that commonly occurring mutations occur at least once.

In Figure 21b we observe that there is no significant difference between the sizes of the shared latent space for predicting the somatic mutation profiles using LLOST_B, with 50 having .79 and 200 having 0.81 F1-scores. More precisely, this shows LLOST_B has a lower false positive rate in comparison to LLOST_NB, so it is not as influenced by the frequency of commonly occurring mutations. This is attributed to parameterization of the Bernoulli likelihood via the NB parameters. Since frequently occurring mutations have a lower variance, they will be only assigned a probability when a mutation is actually present in a somatic mutation profile.

Figure 21b also shows an ablation study of the cancer type label, y in LLOST_B. Without the label, the performance of LLOST_B degrades significantly. We suspect that the label encourages the distribution from the lesion domain to match the structure of the distribution of the mutation domain, akin to supervised machine learning models. Without the label, the decoder is most

likely using lesion specific features, which to the decoder is essentially noise and therefore the output distribution is also noisy.

We observe in Table Table VI the PPV of individual cancer types using LLOST_{B}. As expected the cancers with a higher number of samples are better able to predict a somatic mutation profile from a corresponding medical images of a lesion. This table also demonstrates that LLOST_{B} does not overfit or underfit, since the PPV varies across all cancer types. PPV scores for some cancers with lower samples (LUAD, LUSC, COAD, UCEC, and CESC) are higher than expected. This highlights that LLOST_{B} can learn the underlying structure of the somatic mutation profiles through the latent representations of domain specific and shared features.

Figure Figure 23 displays the shared latent space in the forward direction. We see that some of the cancers overlap such as LUAD with LUSC and UCEC with COAD. This is also evident in cancer biology as LUAD and LUSC share similar somatic mutations, where LUSC tends to have a higher TML (260). Similarly, cancer biology indicates UCEC and COAD also share somatic mutations (260). This shows LLOST_{B} can map features from the lesion domain and orient them to match the features of the somatic mutation domain.

5.6 Discussion

Overall, our results indicate that a distinct shared latent space greatly improves the learning of two very different datasets, especially when one dataset is sparse and high-dimensional. Here we further investigate some of our design choices.

LLOST_{NB} generally overestimates the count of somatic mutations. We hypothesize this is because of an imbalanced dataset where cancer types with larger samples propagate the phenom

Cancer Type	Total Samples (15%)	PPV
Bladder Cancer (BLCA)	119 (18)	0.805 ± 0.117
Breast Cancer (BRCA)	139 (21)	0.834 ± 0.319
Cervical Squamous Cell Carcinoma (CESC)	54 (8)	0.726 ± 0.172
Colorectal Cancer (COAD)	21 (3)	0.710 ± 0.067
Esophageal Carcinoma (ESCA)	16 (2)	0.513 ± 0.196
Glioblastoma Multiforme (GBM)	91 (14)	0.645 ± 0.043
Head and Neck Cancer (HNSC)	159 (24)	0.840 ± 0.124
Kidney Chromophobe (KICH)	15 (2)	0.264 ± 0.015
Kidney Renal Clear Cell Carcinoma (KIRC)	186 (28)	0.829 ± 0.175
Kidney Renal Papillary Cell Carcinoma (KIRP)	34 (5)	0.565 ± 0.293
Brain Lower Grade Glioma (LGG)	110 (17)	0.732 ± 0.023
Liver Hepatocellular Carcinoma (LIHC)	95 (14)	0.776 ± 0.149
Lung Adenocarcinoma (LUAD)	41 (6)	0.719 ± 0.271
Lung Squamous Cell Carcinoma (LUSC)	36 (5)	0.735 ± 0.043
Ovarian Serous Cystadenocarcinoma (OV)	103 (15)	0.749 ± 0.386
Prostate Adenocarcinoma (PRAD)	14 (2)	0.394 ± 0.039
Stomach Adenocarcinoma (STAD)	46 (7)	0.672 ± 0.086
Uterine Corpus Endometrial Carcinoma (UCEC)	63 (10)	0.713 ± 0.261

TABLE VI: Positive predictive value of predicting if a distinct Somatic Mutation occurs in the specific cancer types. The number of test samples for each cancer type is indicated in Column 2 within the parentheses. The PPV is the mean Positive Predicted Value with standard deviation indicated by $\pm$ using bootstrapping with a 1000 runs.

of cross-excitation (275). Specifically, if two mutations co-occur together frequently, then the presence of one will excite the other if one is not present. One method of decreasing this phenom is to increase the dataset size, so that the conditional probabilities of co-occurrance decrease. Another avenue, is to change the likelihood to a Zero-Inflate Negative Binomial distribution, which has a superior performance with datasets containing many zeros (absence of mutations).

A major benefit of our model is the use of a distinct shared latent space as it allows us to move from the imaging domain to the mutation domain. In our original hypothesis, we stated we can use domain specific features along with partial correspondence from the shared latent space to improve prediction accuracy. We observe this in the high PPV score of COAD even with limited samples. This is most likely is due to the shared somatic mutations between COAD, UCEC, and BLCA, as reported in (260), thereby highlighting the influence of domain specific features.

As discussed previously, the correlations between domains is observed in LUAD and LUSC, which both have less than 40 samples. Although, LUAD and LUSC share very similar somatic mutation profiles, the geometric properties from the lesion domain in conjunction with the cancer type are correlated with features from the mutation domain to allow for enough discrimination between the two samples. We also observe this again in the uterine based cancers where there is significant overlap between UCEC and COAD due to their shared geometrical properties.

With further optimization, it is possible to use LLOST_{B} to aid clinicians, since somatic mutations themselves are a strong genetic marker for patient prognosis, subtyping, and treatment planning. For example, BRAF is a somatic mutation targeted in metastatic colon cancer, (276),

which our model predicted in COAD with a F1 score of 0.760[1]. At initial diagnosis, the cancer may not seem metastatic, but with the aid of, LLOST_B, the indication of the BRAF mutation could suggest the clinicians to focus on specific treatment plans that target aggressive types of COAD.

By predicting a full mutation profile, LLOST_B can also aid clinicians in determining ineffective treatment plans. For example, the occurrence of TRIM27 and EGFR together in LUAD is associated with poor response to anticancer therapy in EGFR-mutated lung cancers (243). Among the LUAD samples, our model predicts TRIM27 and EGFR with an F1 score of .791 in the full mutation profile, thereby these patients could be recommended to receive a different treatment.

There are a number of avenues we can take on improving this model. A short term goal is to determine a more effective optimization strategy for the shared latent space, such as including a loss to model reconstruction of the cancer type label in the shared latent space. A clear extension to this model is incorporation of established radiological and imaging texture features or other genetic domains such as RNA-SEQ. Neither of these are trivial extensions as it requires a significant change to neural network architecture as we must account for the unordered nature of point clouds and manipulation of the shared latent space. We also recognize, that due to the resolution of medical imaging modalities, achieving a high predictive power of somatic mutations

---

[1]Note this is not the overall F1 score for all mutations in COAD.

would be very difficult; regardless, we hypothesize our model may at least aid clinicians in quickly determining an effective treatment plan.

5.7 Conclusion

We presented, LLOST, a deep latent variable model for the prediction of a somatic mutation profiles of patients based on their corresponding image using dual variational autoencoders joined together with a separate latent space. We have shown that it is possible to predict somatic mutation profiles without reducing the dimensionalality of the dataset. We also showed that shared correlations between the imaging and mutation domain help predict mutation profiles when there are a limited amount of samples in the training set. The LLOST_{B} in specific has several attractive features: point cloud representation of a lesion, sharing of a latent space across two significantly different domains, a flexible training objective, and prediction of distinct somatic mutations.

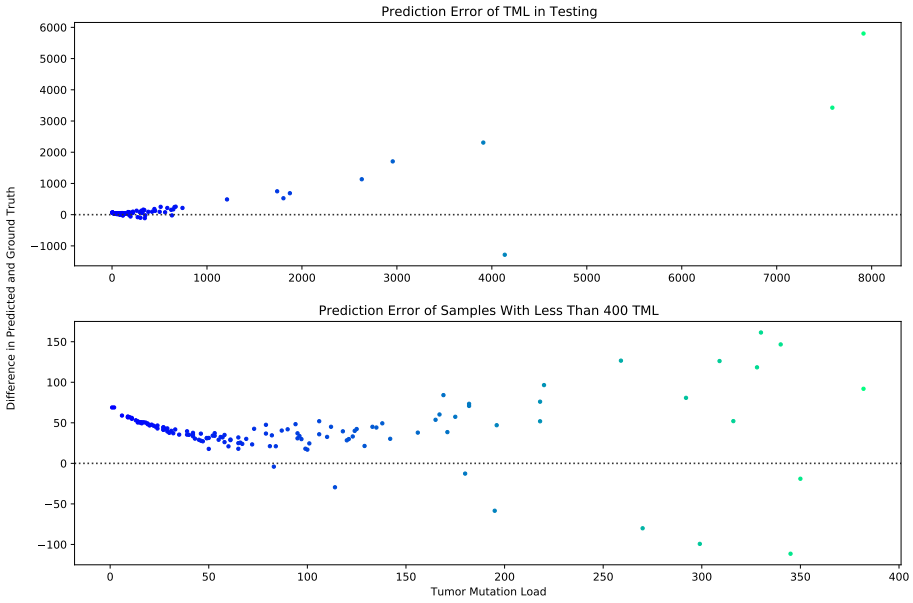


Figure 22: Point prediction error of TML. The top plot shows the point estimate error in predicting the TML using LLOST_{NB} in the test samples. The bottom plot is a zoomed in of the top plot, where samples with less than 400 TML is reported. X-axis is the expected TML. Y-axis is the difference in TML of predicted and expected.

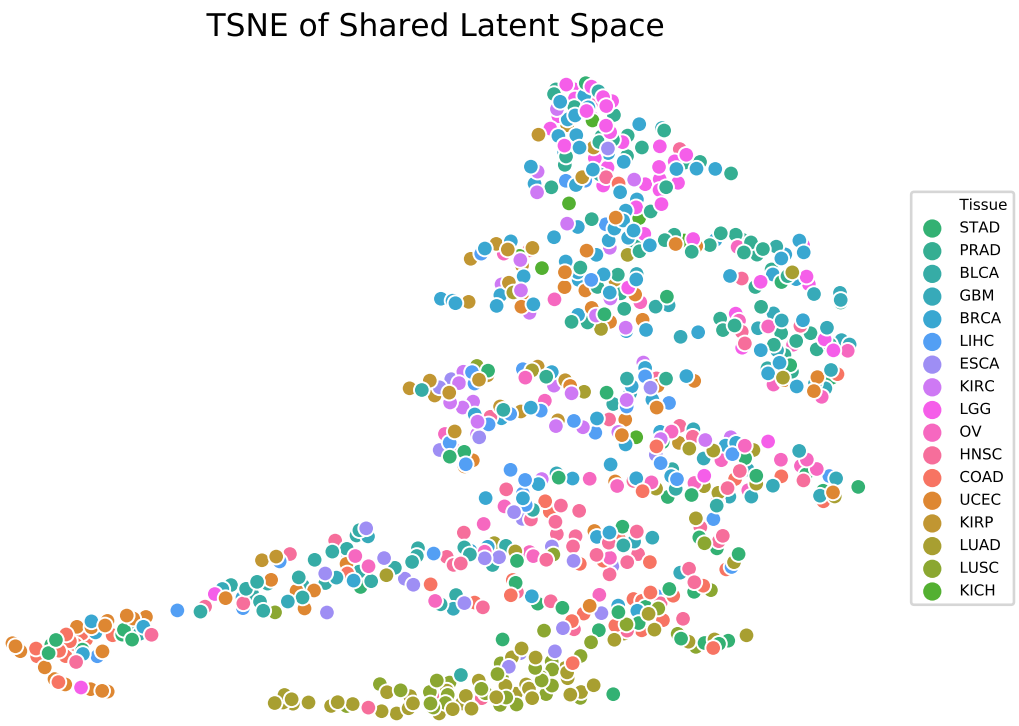


Figure 23: A TSNE of the shared latent space in the forward direction after a test batch of lesions images. Best viewed digitally.

CHAPTER 6

CORRELATED MIXED MEMBERSHIP MODELING OF SOMATIC MUTATIONS

This work is published in Mehta, Rahul, and Muge Karaman. "Correlated Mixed Membership Modeling of Somatic Mutations." 2020 International Joint Conference on Neural Networks (IJCNN 2020), 2020, Glasgow, UK.

6.1 Motivation

Discerning the relationships between somatic mutations in cancers is the foundation for targeted treatment and patient subtyping. Since somatic mutations in cancer genomes are often heterogeneous and sparse, where two patients with the same cancer may share only one mutation among thousands, models summarize the high-dimensional interactions into a simpler form. This requires a model that incorporates multiple confounding variables to determine relationships between somatic mutations. Based on current literature (277; 278) mutually exclusive and co-occurring mutations are influenced by non-linear relationships between gene mutation frequencies, biological processes, cancer (sub)types, total number of mutations in a tumor (TML), and positive selection for mutations. The combination of multiple confounding variables and the inherent sparsity of somatic mutation data poses a challenge to understand the underlying co-dependencies between mutations.

Statistical and computational models that try to discover relationships between somatic mutations often decompose a patient's mutation profile into a set of higher-level structures that closely resemble known biological processes. This approach (279; 280) generally follows a random walk on an existing biological interaction network. This networks can be modeled as a graph, $G = (V, E)$, where each vertex, V is a gene, and the edge, E denotes the interaction among genes. The network is then modified into a weighted graph, with edge weights representing probability of interactions and vertex weights corresponding to the frequency of a mutation in a gene. A walk is then simulated by starting at a mutated gene and moving to another gene based on the probabilities of edge and vertex weights. The end result is a smaller subnetwork called a functional network that represents an altered biological process. While functional networks have been validated to discover some aberrant genes and pathways, they often result in false positives due to the inherent assumptions made.

The most common compendium of interaction networks widely used to generate functional networks is the Kyoto Encyclopedia of Genes and Genomes (KEGG) (281). The KEGG interaction networks specify genetic pathways, which are complex graphical networks with directed and undirected edges connecting genes based on their physical and biochemical properties. The genetic pathways are then ascribed to specific biological processes. For example, the biological process of cell apoptosis (cell death) is controlled by two known genetic pathways compromising of a multitude of different genes. The networks within the KEGG database, however, are diverse and recapitulate a disease free patient. Functional networks therefore, assume the interaction networks are also cancer-relevant and disease-specific. As a result, functional networks are

generalized to a common patient population and struggle to discriminate between different cancer types (282).

The second assumption is how functional networks take advantage of mutual exclusivity in somatic mutations. The process of mutual exclusivity in somatic mutations describes how mutations do not occur together if they are in the same genetic pathway (31). In functional networks, accounting for mutual exclusivity corresponds to the frequency of a mutation, which is the weight of a vertex V_i in the graph. Theory, however suggests that there are multiple confounding factors that cause mutual exclusivity (277). For example, the mutual exclusivity of mutations in the TP53 and MUC16 genes are better explained by cancer type and TML in cororectal cancer (277).

The last assumption is of preprocessing somatic mutation data to a limited number of mutations. Although preprocessing is done in numerous studies beyond genetics, there is no gold standard for somatic mutations. Furthermore, due to the high dimensionalality of somatic mutation data and the limited number of samples, preprocessing will be biased towards frequently occurring mutations (283). So while a model can identify novel co-occurring mutations within the functional networks, it may only reflect the model's preference for a specific paradigm. For example, HotNet2 (279) removes samples with more than 400 somatic mutations, however, overdispersion of a gene and TML are both significant factors that influence the relationship between mutations.

From a machine learning perspective we can describe the problems faced by functional networks as overfitting. This is elucidated by the false positives produced from functional

networks. Specifically, functional networks memorize the parameters of the interaction network instead of learning the parameters of the somatic mutation dataset. So while, functional networks may reproduce valid biological processes, they do not necessarily capture cancer-relevant relationships between mutations. For example, functional networks validated targeted treatment drugs erlotinib or gefitinib for mutations in the EGFR gene. Since the EGFR is omnipresent in genetic pathways and as a mutation, functional networks constrain a patient to be mainly influenced by the EGFR mutation. From the previous example, these treatments, however, are only temporarily effective and a relapse often occurs due to the presence of a co-occurring mutation in the same genetic pathway that had equal cancerous potential (284). So, although functional networks correctly identified the EGFR mutation, it is the interplay between many mutations that influences cancer biology.

In this paper, we propose to exploit the inherent latent structure of a somatic mutation dataset with a generative probabilistic model. Instead of relying on interaction networks *a priori*, we use a prior based on a correlated random measure (CoRM). The CoRM enables the model to specify a notion of similarity on the possible latent distributions of the somatic mutations. In our case, we would like the latent distributions to correspond to two unique characteristics of the somatic mutation data: mutual exclusivity and cancer-related biological processes. Specifically, the CoRM assigns probabilities to particular configurations of latent distributions via a Zero Inflated Negative Binomial Process (ZINB) and enforces mutual exclusivity via a correlation structure through the conjunction of the Beta-Bernoulli Process and neural networks.

The main contributions of this work are as follows. We propose the Correlated Zero Inflated Binomial Process, CoZINB, a novel generative latent variable model with an implicit dependency structure and latent parameters that represent the sparsity of a somatic mutation dataset. Therefore, CoZINB avoids biases from frequency and interaction networks that confound functional network methods. The CoZINB is used to factorize a somatic mutation profile to infer positively and negatively correlated latent factors[1] that consist of co-occuring mutations and represent biological processes. We create a computationally efficient inference scheme with the confluence of stochastic variational inference and amortized variational inference (285; 286; 287). Our experiments on the pan cancer dataset from The Cancer Genomic Archive (TCGA) verify the benefits of our model to uncover co-occurring mutations while maintaining rules of mutual exclusivity and cancer specific processes.

6.2 Background

6.2.1 Negative Binomial Distribution

The dispersion parameter r of the Negative Binomial (NB) distribution allows models self-excitation (275), an amenable property in biology, such that if a distinct cancer mutation occurs in a tumor, it is likely to occur again in the tumor. Another intuitive way to express the NB distribution is to model it as a Poisson-Gamma mixture distribution. The NB distribution is then a weighted mixture of Poisson distributions, where the rate parameter is an unknown gamma distribution. Thus, an NB distribution is analogous to an overdispersed Poisson distribution.

---

[1]There are a number of ways to describe sets based on the current nomenclature (clusters, factors, and topics). For the remainder of the paper we refer to the latent space of our model as factors.

Research in differential expression analysis (288; 289), regression, and clustering of single-cell data (149) has shown superior performance of NB distribution.

6.2.2 Mixed Membership Modeling

Algorithm 2 Latent Dirichlet Allocation a Mixed Membership Model

for Each cluster $k : 1, ..., K$ **do**
  Sample $\phi_k \sim Dirichlet_M(\eta_1, \ldots, \eta_m)$
end for
for Each sample $j : 1, ..., J$ **do**
  Sample $\theta_j \sim Dirichlet_K(\alpha_1, \ldots, \alpha_K)$
  for Each data point x_j, m in sample j **do**
    Sample Cluster $z_{j,m} \sim Discrete_K(\theta_j)$
    Sample data point $x_{j,m} \sim Discrete_M(\phi_{z_{j,m}})$
  end for
end for

Mixed-membership modeling is essentially a soft clustering problem. In contrast to hard clustering where a data point is assigned to one and only one cluster, each data point in a mixed-membership model is associated with a number of clusters. It is the interaction among these clusters that gives rise to an observed data point. In the context of cancer biology, mixed-membership models allow us to remove the restrictive assumption that mutations are mutually exclusive across clusters (290).

The generative process of LDA follows Algorithm 2. What specifically matters, is how choosing the priors: the factor score matrix, $\boldsymbol{\theta}$ and the factor loading matrix, $\boldsymbol{\phi}$ for different

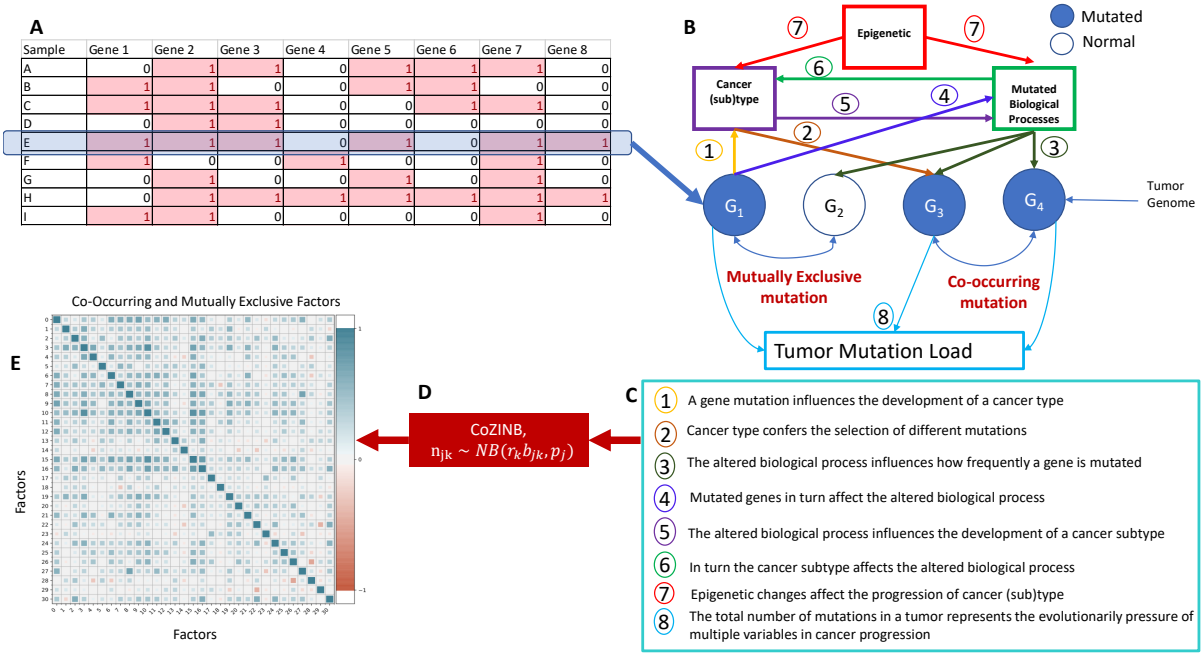


Figure 24: CoZINB models multiple cofounding variables to determine co-occurring mutations and mutually exclusive mutations. (a) Is an example of a set of binarized somatic mutation profiles represented as a matrix, where 1 indicates that a gene is mutated (b-c) There is a complex relationship between cancer, biological processes, and the development of mutations. (277) (d) These relationships are modeled via a correlated Beta-Bernoulli Process and an independent Gamma process as seen in Equations Equation 6.2 and Equation 6.4, where n_{jk} is the number of times the k_{th} factor appears in sample j. (e) CoZINB learns a diverse latent structure of somatic mutation profiles, with positively and negatively correlated factors that contain a set of co-occurring mutations

generative models. For example, if we change the priors in Algorithm 2 to Gamma distributions and the likelihood to a Poisson distribution we can obtain Poisson Factor Analysis (PFA) (275). Another extension are hierarchical mixed-membership models, which increase model capability by stacking stochastic processes such that algorithm infers the number possible clusters within a dataset (291).

Though powerful, mixed-membership models are limited as they do not explicitly model the correlations between the mixing proportions of any two clusters. Using science articles as an example, the LDA topic model cannot capture that the presence of a topic on cells is more positively correlated with a topic on cancer in comparison to a topic on astronomy. Recent innovations in joint modeling of correlation and mixture proportions has resulted in improved prediction and interpretability of the topics as seen in (292) and an hierarchical extensions in (285; 287; 286). Continuing in the context of topic modeling, these methods inject a Gaussian covariance into the $\boldsymbol{\theta}$ such that the factor score matrix takes into account the correlation between topics. Our model follows this format, where the bottom stochastic process is based on correlated random measures.

6.2.3 Latent Space as a Correlated Random Measure

Homogenous completely random measure (CRM) (155) is built upon a joint measure as $v(da, dw) = H(da)v_w(dw)$ where $H(da)$ is the base measure and $v_w(dw)$ is the mean measure (levy measure). As an example, the Gamma Process (ΓP) has a mean measure of $v(da, dw) = H(da)e^{-cw}/wdw$. Intuitively, we can think of the base measure as defining the existence of factors and the rate measure assigning a weight to each factor. This definition allows us to have an infinite number of factors, but a finite mass, and is a foundation of many of the non-parametric Bayesian statistical models.

A correlated random measure (CoRM) is created by augmenting a CRM into a higher space to include locations such that the mean measure is now $v(da, dw, dl) = H(da)v_w(dw)v_l(dl)$, where $v_l(dl)$ is a vector of locations in $\mathbb{R}^d$. We can then draw correlated weights x via a transformation

function as $x \sim T(\cdot|w, F(l))$, given the uncorrelated weights, w, and a random kernel function on the locations $F(l)$. Based on Proposition 1 from (286) the transformed levy measure is now $\tilde{v} = v(da, dw, dl)p(dx|F(l), w)$ The transformation distribution, like a homogeneous CRM, is restricted to have finite mass to guarantee generation of useful statistical models.

6.3 Model

6.3.1 Correlated Zero Inflated Negative Binomial Process

The Correlated Zero-Inflated Negative Binomial Process (CoZINB) is a correlated hierarchical Bayesian model for learning the latent representations of somatic mutation profiles. We take advantage of neural networks for the random function $F(l)$ to model non-linear correlations of somatic mutations. The transformation distribution based on the Beta-Bernoulli Process (BeBP) is then used to enforce sparsity of the globally shared latent factors. We describe this is as a draw from the CoZINB, $X_j \sim NB(RB_j, p)$ with a mean measure $\tilde{v}$:

$$H(da)v_l(dl)\alpha w^{-1}(1-w)^{w-1}p(dx|F(l), w)R_0(dr)dw$$

so B_j is a draw from the transformation distribution $p(dx|F(l), w)$ and R is an independent ΓP. By deriving a latent representation with an unsupervised correlation structure, we use the inherent information shared across the somatic mutation profiles to ensure we capture mutual exclusivity between factors.

As shown in Figure Figure 24, by using CoZINB we can model 1) the excessive amount of mutations that do not occur in a patient, 2) non-linear interactions of somatic mutations 3)

implicit cancer subtype based on the factors 4) total number of mutations in a somatic mutation profile. We will show in our experiments the latent factors follow a pattern of co-occurrence that is also observed in cancer biology.

6.3.2 Generative Structure

We represent our dataset as a bag of words commonly seen in text corpus, $X \in \mathbb{N}^{J \times M}$, where J and M are the number of patients and distinct number of mutations, respectively. As in many factor models we also introduce an additional latent variable, z_{jm}, indicating the factor assignment for mutation m. The mutation profile of each patient is realized as a mixture of latent factors ($\phi_{z_{jm}}$) shared by all patients. Specifically, we model a distinct mutation count in a patient as:

$$\begin{aligned} n_{jmk} &= \sum_{i=1}^{N_j} \delta(z_{ji} = k, m_{ji} = m) \\ n_{jk} &\sim NB(r_k b_{jk}, p_j). \end{aligned} \tag{6.1}$$

From a biological perspective n_{jk} is the number of times a k biological process occurs in patient j. The count n_{jmk} corresponds to the contribution of underlying biological process k to the count of mutation m in sample j. The total number of mutations in patient j is then N_j. The shape parameter r_k captures the popularity of the biological process k across all patients. The probability parameter p_j models the patient specific somatic mutation profile. Specifically, p_j accounts for heterogeneity among the patient population. To enable sparsity and correlations within the latent factors, we introduce Bernoulli latent variables b_{jk}. When $b_{jk} = 1$, the latent factor k is present in patient j, otherwise irrelevant. The correlations within the binary random

variables are produced via a transformation of the BeBP as discussed in 6.2.3. A transformed distribution follows (286) as

$$\begin{aligned} b_{jk} &\sim Bernoulli(\sigma(\sigma^{-1}(\pi_k) + F(\dot{)})) \\ \pi_k &\sim Beta(\alpha/K, \alpha(1-1/K)) \end{aligned} \tag{6.2}$$

where σ is the sigmoid function. As a result, the proposed model assigns positive probability only to a subset of factors, based on the correlation structured created in $F(\dot{)}$.

Inspired by Variational Autoencoders (VAE) (60) to model nonlinear correlations we use a deep neural network to create a kernel $F(\dot{)}$ is:

$$\begin{aligned} f(h_j, l_k) &\sim N(u_f(h_j, l_k), \sigma_f^2(h_j, l_k)) \\ h_j &\sim N(0, aI), l_k \sim N(0, bI). \end{aligned} \tag{6.3}$$

where h_j is a patient specific vector as the output of an inference network of a VAE. The locations, l_k, are then concatenated with h_j as in input the decoder of a VAE to create the kernel. Unlike the standard VAE decoder which aims to recreate the input of a encoder, the decoder in our model supplies prior information to $F(\dot{)}$. This paradigm of generating the kernel is similar to Empirical Bayes (293), where we use the structure of the data as a prior to generate $F(\dot{)}$.

Building upon the above equations, the full generative structure of the CoZINB topic model follows the paradigm of the Gamma-Poisson construction of a NB process:

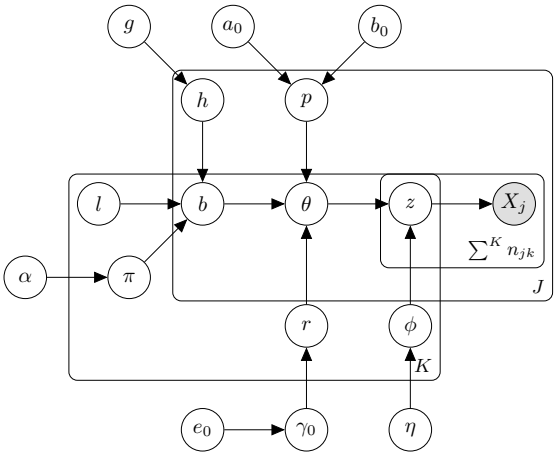


Figure 25: Correlated Zero Inflated Negative Binomial Process Topic Model

$$
\begin{aligned}
x_{jm} \sim Discrete(\phi_{z_{jm}}), z_{jm,k} \sim \phi_{mk}\theta_{jk} \\
\phi_k \sim Dirichlet(\eta_0 \mathbf{1}_M), N_j = \sum_{k=1}^{K} n_{jk}, \\
n_{jk} \sim Poisson(\theta_{jk})\theta_{jk} \sim Gamma(r_k b_{jk}, p_j), \\
r_k \sim Gamma(\gamma_0, 1/\alpha), \gamma_0 \sim Gamma(e_0, 1/f_0), \\
p_j \sim Beta(a_0, b_0), L_{jk} \sim CRT(n_{jk}, r_k b_{jk}), \\
L_k^{'} \sim CRT(\sum_{j=1}^{J} L_{jk}, \gamma_0)
\end{aligned}
\tag{6.4}
$$

and b_{jk} follows Equation Equation 6.2. To aid in inference of r_k, we introduce a data augmentation latent variable, L_{jk} based on the Chinese Restaurant Table as in (275). The full model plate is shown in Figure Figure 25.

6.3.3 Inference

Given the set of observed mutation profiles, X and their corresponding mutation counts m, our goal is to infer the factors z_{jm}, the factor score matrix, θ_{jk}, the factor loading matrix, ϕ_{mk}, and the factor locations l_k. For a biological analogue, θ_{jk} is the proportion of the biological process k in sample j, ϕ_{mk} is the proportion of each mutation in factor k, and l_k is correlation between a set of co-occurring mutations. Denote $n_{jk} = \sum^{N_j} \delta(z_{jm}) = k$, where K is some finite, but large, truncation level, we posit the fully factorize variational scheme:

$$\begin{aligned} q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}, Z, R, P, l, L, h) = \prod^{K} q(l_k) q(r_k) q(\pi_k) \\ q(\phi_k) q(\gamma_0) \prod^{J} \Bigg[q(h_j|X) q(p_j) \prod^{K} q(b_{jk}) q(\theta_{jk}) q(L_{jk}) \\ q(L_k^{'}) \prod_{m=1}^{N_j} q(z_{jm}) \Bigg] \end{aligned} \tag{6.5}$$

The variational distribution for each latent variable is associated with it's own variational parameter(s) as follows

$$q(l_k) = \delta_{\tilde{l_k}}, q(\gamma_0) = \delta_{\tilde{\gamma_0}},$$

$$q(\phi_k) = Dirichlet(\tilde{\eta})$$

$$q(\pi_k) = Beta(\tau_{k1}, \tau_{k2})$$

$$q(r_k) = Gamma(\tilde{r_{k1}}, \tilde{r_{k2}})$$

$$q(p_j) = Beta(\tilde{a_j}, \tilde{b_j})$$

$$q(b_{jk}) = Bernoulli(\nu_{jk})$$

$$q(\theta_{jk}) = Gamma(\tilde{\theta_{jk1}}, \tilde{\theta_{jk2}})$$

$$q(L_{jk}) = \delta_{\tilde{L_{jk}}}$$

$$q(L_k^{'}) = \delta_{\tilde{L_k^{'}}}$$

$$q(z_{jm}) = Multinomial(\psi_{jm})$$

We let $q(h_j|X_j) = \delta_{g(X_j)}$ where g is the inference network of a VAE.

To update the variational parameters we can use stochastic variational inference (SVI) that assumes we subsample $J_t \in J$ patients at every iteration t and optimize the noisy variational objective:

$$\begin{gathered} L_t = \mathbb{E}[\ln p(l, r, \phi, \gamma_0, \pi)] + \\ \frac{J}{J_t} \sum_{j \in J_t} \mathbb{E}[\ln h_j, \theta_j, z_j, p_j, b_j, L_j, X_j] + H[q(\phi]] + \\ \frac{J}{J_t} \sum_{j \in J_t} H[q(\theta_j, z_j, b_j, p_j)] \end{gathered} \tag{6.6}$$

At each iteration t, we update the local variables, z, θ, b, and, p using closed form equations, while the remaining variables, with the exception of ϕ are updated via stochastic gradient descent via a decreasing step size (294). The complete conditional updates are summarized in Table 1. [1]

Latent Variable	Update Type	Variational Update	Variational Parameter
ϕ_k	Closed	$\eta + \sum_{J \in J_t} \sum_{n=1}^{N_j} \psi_{jm}(k) * \mathbf{1}(X_{jm} = \tilde{m})$	$\eta_{k\tilde{m}}$
π_k	Closed	$\alpha/K + \sum^J \nu_{jk}$	τ_{k1}
π_k	Closed	$\alpha(1 - 1/K) + J - \sum^J \nu_{jk}$	τ_{k2}
r_k	Closed	$\gamma_0 + \sum^J \mathbb{E}[L_{jk}]$	r_{k1}
r_k	Closed	$\alpha - \sum^J \mathbb{E}[b_{jk}] \ln 1 - p_j$	r_{k2}
p_j	Closed	$a_0 + N_j$	$\tilde{a}$
p_j	Closed	$b_0 + \sum^K \mathbb{E}[r_k] \mathbb{E}[b_{jk}]$	$\tilde{b}$
b_{jk}	Closed	$\mathbb{E}[\pi\tilde{}_{k2}](\mathbb{E}[\pi\tilde{}_{k2}] + 1 - \mathbb{E}[\pi_{k1}])^{-1}$	ν_{jk}
θ_{jk}	Closed	$\mathbb{E}[r_k] \mathbb{E}[b_{jk}] + \sum^M \psi_j m(k)$	$\tilde{\theta_{jk1}}$
θ_{jk}	Closed	$\mathbb{E}[\ p_j]$	$\tilde{\theta_{jk2}}$
z_{jm}	Closed	$\log \theta_{jk} + \log \phi_{km}$	ψ_{jm}
γ_0	Gradient Ascent	$\sum^K \nabla_{\gamma_0} \mathbb{E}[p(\gamma_0)] + \sum^J \sum^K \nabla_{\gamma_0} \mathbb{E}\left[p(L_k^{'} \vert L_{jk}, \gamma_0\right)]$	δ_{γ_0}
l_k	Gradient Ascent	$\sum^K \nabla_l \mathbb{E}[\ln p(l_k)] + \sum^J \sum^K \nabla_l \mathbb{E}[p(b_{jk} \vert \pi_k, h_n, l_k]$	δ_{l_k}
h_n	Gradient Ascent	$\sum^J \nabla_h \mathbb{E}[\ln p(h_j)] + \sum^J \sum^K \nabla_h \mathbb{E}[p(b_{jk} \vert \pi_k, h_n, l_k]$	δ_{h_j}
L_{jk}	Gradient Ascent	$\sum^J \sum^K \nabla_L \mathbb{E}[p(L_{jk} \vert n_{jk}, b_{jk}, r_k]$	$\delta_{L_{jk}}$

TABLE VII: A summary of the update information for the Variational Parameters.

6.4 Experiments

To show the importance of modeling sparsity we consider two latent variable models as comparisons:

---

[1] Note that $\mathbb{E}[\tilde{\pi_{k2}}] = E[\pi_{k1}](1 - p_j)^{\mathbb{E}[r_k]}$.

PRME is a hierarchical mixed membership model based on the Dirichlet Process and the correlations are induced through neural networks with a Gaussian kernel, so

$$\theta_{jk} \sim Gamma(r_k, f(h_j, l_k))$$

. Where r_k is a stick-breaking Dirichlet Process instead of a Gamma Process as in CoZINB.

Prod-LDA (295) is an extension of LDA to deep generative models. It replaces the Gaussian Prior in a VAE with a Dirichlet distributed prior to better learn the latent structure of text data.

To compare CoZINB's latent factors with functional networks, we use the HotNet2 model.

Hotnet2 is used in a discriminate approach to compare functional networks with CoZINB. We train two distinct linear support vector machine (SVM) classifiers (296) with the subnetworks learned by Hotnet2 (279) and the factors from CoZINB as features. The SVMs are used to predict the cancer type of a patient given the features. Additionally we compare the correlation between functional networks and biological process and the correlation between factors from CoZINB and biological processes

Architecture encoder and decoder follow simple multilayer preceptrons (MLP) with dimensions shown in Table Table VIII. The hyperparameters for PRME and CoZINB are set as $a = 1, b = 1, \alpha = 1, \eta = 0.2, a_0 = .001, b_0 = .001, e_0 = .001, f_0 = .001$, $M = 21332$, and a truncation level K as 100 for all models. All gradient updates are done via Adam (297) with a learning rate of $1e-3$.

We use the pan cancer dataset from the TCGA consisting of 10296 tumor samples (J) and 21332 distinct mutations (M). For training the models we select 70% of the dataset, setting 20% for validation of parameters, and 10% for testing. The only pre-processing we do is remove 'abParts,' which is not represented in any database.

6.4.1 Comparison Metrics

Per-heldout-word perplexity

$$exp(\frac{1}{|X_{Test}|}\sum_{m\in X_{Test}} \ln p(X_{jm}|X_{j,Train}))$$

is used to measure the utilization of the latent dimensions of the two other latent variable models.

Precision@1 measures the influence of the factors on determining the correct count of one distinct (@1) held-out mutation as TML increases. This is calculated by

$$\frac{1}{N}\sum_{1}^{N}\delta(\tilde{X_{jm}} = X_{jm}),$$

where $\tilde{X_{jm}}$ is the predict count of mutation m.

Specificity is used to compare the discriminative power of the functional networks and the latent factors of CoZINB in predicting the correct cancer type Lung Adenocarcinoma (LUAD), Colon Cancer (COAD), Ovarianc Cancer (READ), Stomach Cancer, and Breast Cancer (BRCA).

Spearman-rho is used to assess how well the factors produced from CoZINB match with biological processes.

Qualitatively we analyze the biological implications of the factors obtained by CoZINB for a subset of the TCGA dataset in LUAD and COAD. We chose these subsets as there are an established number of studies identifying co-occurring and mutual exclusive somatic mutations within these cancers.

6.5 Results

6.5.1 Quantitative

	Perplexity@H	Encoder Sizes	Decoder Sizes
Prod-LDA	@0 1986.30 @100 1091.33 @200 1297.19 @400 1691.76	$[M \times 100]$ $[100 \times 100]$ $[100 \times 100]$	$[100 \times 100]$ $[100 \times 100]$ $[100 \times M]$
PRME	@0 1151.96 @100 721.42 @200 997.19 @400 1091.76	$[M \times 1000]$ $[1000 \times 1000]$ $[1000x1000]$ $[1000 \times 40]$	$[40 \times 80]$ $[80 \times 80]$ $[80 \times 80]$ $[80 \times 2]$
CoZINB	@0 779.69 @100 721.77 @200 699.35 @400 709.34	$[M \times 1000]$ $[1000 \times 1000]$ $[1000 \times 1000]$ $[1000 \times 40]$	$[40 \times 80]$ $[80 \times 80]$ $[80 \times 80]$ $[80 \times 2]$

TABLE VIII: Perplexity of the held-out test Pan Cancer dataset and architecture of Prod-LDA, PRME, and CoZINB. The @H indicates if the models were trained with only the top frequently occurring mutations i.e., @100 is a training set with only top 100 frequently mutations, @0 = no mutations were held out

As Table Table VIII shows, CoZINB performs better than PRME and Prod-LDA when there are a excessive number of zeros in the dataset. Specifically, we show the importance of using

a prior that can incorporate sparsity, by only using the top occurring mutations at levels of 100, 200, and 400 during. As we limit the amount of mutations the models need to learn, the perplexity of all models begin to improve. The sparsity of mutations, therefore poses a significant challenge in the learning paradigm. CoZINB achieved the best performance due to the use of selector variables b_{jk}. Specifically the selector variables assigns a finite number of latent factors for each sample and an explicit zero mass to the rest. Any random signals then will receive a zero probability, wheres the baseline models will assign some small probability to the latent factor resulting in a poor local optima.

In Figure Figure 26 we observe the influence of the CoZINB shape and probability parameters. The PRME limits the distribution of the factor score matrix, θ_{jk}, to the frequency of latent factors. In comparison, CoZINB models the count data as the frequency of the latent factors in r_k and the probability of a count of somatic mutations in p_j. More precisely, the expected count of mutation m in sample j is proportional to $\frac{p_j}{1-p_j}$ and the factor score matrix.

Our assertion that functional networks via Hotnet2 are not specific to cancer types and overfit to the structure of interaction networks is shown in Figure Figure 27. There is a significant variance in predicting cancer type when using the functional networks, especially for LUAD. The poor performance for HotNet2 in LUAD is likely due to the similarities of mutation profiles of LUAD and LUSC cancers, where LUSC generally has a higher TML. In contrast, CoZINB factors preserved cancer type information, especially in BRCA where it achieves an average specificity of 0.83. This suggests that interaction networks, while valuable, have much less influence in determining the relationship between mutations.

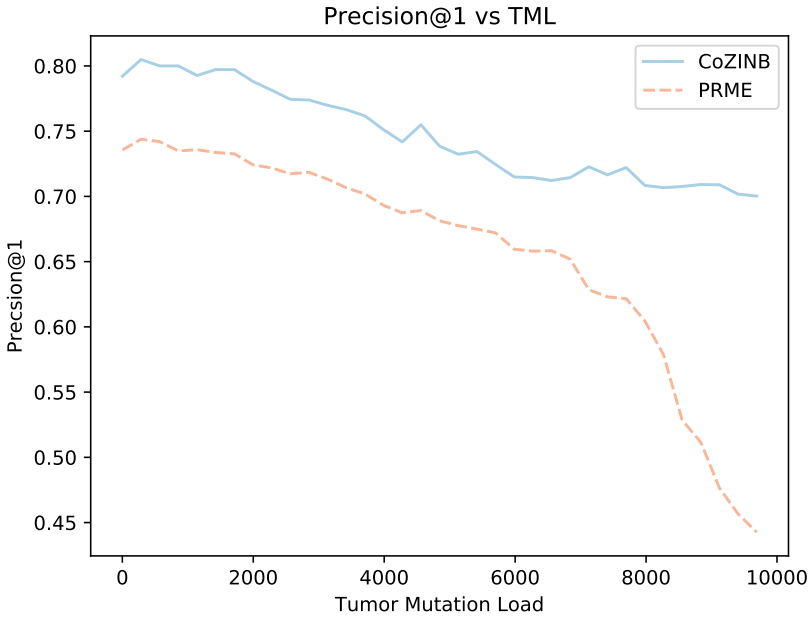


Figure 26: Precision@1 for CoZINB is higher than PRME as it takes account into the sparsity of the dataset through the Negative-Binomial process

In Figure Figure 28 we show the interplay of latent factors in LUAD, COAD, READ, and BRCA. Each bar represents how many times a latent factors occurs in a tissue. We can see there are distinct factors that only correspond to specific cancers. We use the spearman-rho coefficient to assess if the latent factors 11, 12, 39, 40, 47, and 52 from BRCA represent a biological process. We compare it against the spearman-rho coefficient between functional networks of BRCA and biological processes. Using the existing database provided by KEGG, CoZINB has a ρ of 0.67 for the latent factors of BRCA, while HotNet2 is significantly more correlated with a $\rho$of 0.78. This not unexpected since HotNet2 is built on the prior information from interaction networks.

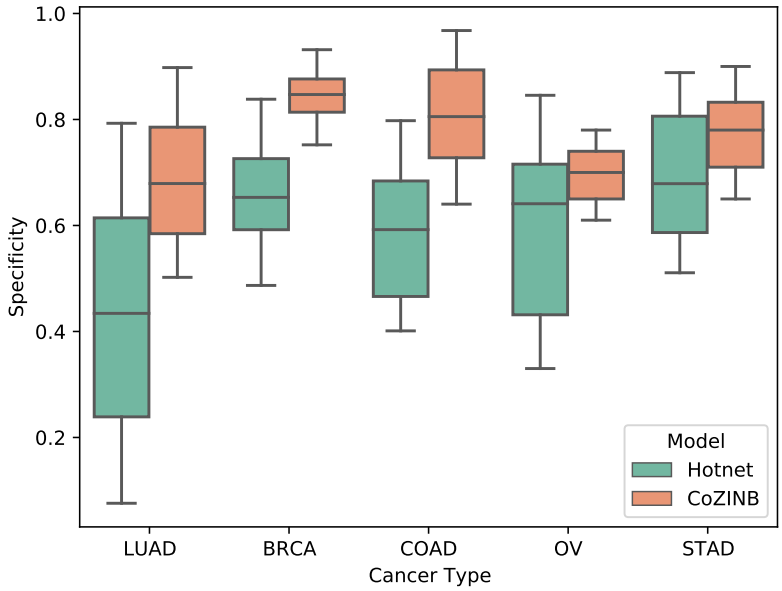


Figure 27: Number of times a factor occurs for all the patients with Lung Adenocarcinoma (LUAD), Colon Cancer (COAD), OV (Ovarian Cancer), STAD (Stomach Cancer),and Breast Cancer (BRCA).

A key insight from Figures Figure 26, Figure 27, and Figure 28 is that we observe CoZINB's robust grouping of frequently occurring mutations into a few factors. The frequently occurring mutations incur a large penalty due to the ZINB process and the optimization procedure prevents the probability mass of frequently occurring mutations to spread across factors. Thereby, the remaining factors are more diverse and allow for better discriminative power between TML and cancer type. We see this again in the case studies below, where frequently occurring mutations are grouped into Factors 1-4.

6.5.2 Lung Adenocarcinoma Case Study

A well studied mutual exclusive set of somatic mutations in LUAD are KRAS and EGFR where in a significant fraction of LUAD patients, these two mutations are rarely if-ever observed together in the same tumor (298). The general assumption of limited co-occurrence patterns between EGFR and KRAS is that there is no selective pressure to favor cells with both mutations over cells with a single mutation since both mutations activate similar biological processes This behavior is reproduced in CoZINB, where in the most common factors 0-4, KRAS and EGFR are at the opposite spectrum, with a mean distance of 7185.

The factors that are more unique to LUAD, Figure 28 in our model are identified by factors 27 with top 5 mutations as: [DNAH5 PCLO ANK3 TTN TP53BP2] and 37 which include the mutations: [USH2A TTN TSHZ2 DNAH3 MUC16]. To confirm this computationally, we use a simple Random Forest (RF) for classification of cancer type across all patients using the counts the top performing factors as features. With a significance of $p < 0.05$ RF gives the ranks factors 27 and 37 as the most important features for LUAD, thereby, verifying our model can also implicitly determine patient cancer type.

Figure Figure 29 shows the unique factors in every patient with LUAD as TML increases. Specifically LUAD with high TML will have some form of the background somatic mutations captured in factors 0 to 3, however as, TML begins to increase, occurrences of factors that contain mutually exclusive mutations will appear. Examining the top underlying somatic mutations of Factor 30 (MUC16, FLG, LRP1B, ZFHX4 CSMD3, RYR) and Factor 1 (PIK3CA, PTEN, TP53, MUC16, MAP3K1,), we observe mutual exclusivity of the MUC16 and TP53 genes as discussed

in section 5.1. CoZINB also identifies the known pattern of mutual exclusivity of the EGFR, KRAS, and BRAF genes in LUAD. CoZINB, additionally shows that Factor 7 with both KRAS and BRAF are, however, co-occurring at higher TML. Cisowskia et al. (299) hypothesized that a BRAF mutated lesion that confers an additional KRAS mutation hyperactivates a mutated biological process that further promotes mutations on both BRAF and KRAS.

6.5.3 Colon Cancer Case Study

Figure Figure 28 shows Factor 6 is unique to the 408 COAD tumor samples within the dataset, with top 5 mutations: [PMVK, BAZ2B, STARD3, SESTD1, KLHL4]. Using the same methodology as before, a RF classifier ranks factors 6 and 41 as the most significant ($p < .05$) features for determining if a patient has Colon Cancer. Moreover, these specific mutations are also known to commonly occur together as passenger mutations (300).

COAD also has specific factors associated with increasing TML in Factors 8 and 21 ($p < .01$) with mutations: [MUC16, DNAH5, TTN, PCLO, ANK3] and [BRCA2, CTNNAL1, DDX52, NAH10, AK9]. Of note are the occurrences of MUC16, PCLO, and BRCA2, which are all hypermutated in colon cancers. Similar to LUAD we can see Factor 8 is mutually exclusive with factors that include the TP53 mutation with increasing TML.

6.6 Discussion and Future Work

The CoZINB, unlike existing methods for assessing co-occurring somatic mutations, is unsupervised and infers a latent structure in a sparse and complex dataset. A major concept shown in our model is that we can probabilistically capture mutual exclusivity between somatic mutations as a non-linear transformation of multiple latent variables. We also show that to

correctly represent cancer biology and patterns of mutual exclusivity and co-occurrance in somatic mutations, a model needs to incorporate tumor mutational load, cancer type, and non-linear mutation correlations as confounding variables.

Although, we argued against the use of interaction networks, they might have a place in a semi-supervised framework. For example, it would be useful to understand the causal relationship of somatic mutations and unaltered genes, analogous to link-prediction in community network detection. This learning paradigm, however, is a complex combinatoral problem, considering the possible exponential search space of somatic mutation interactions.

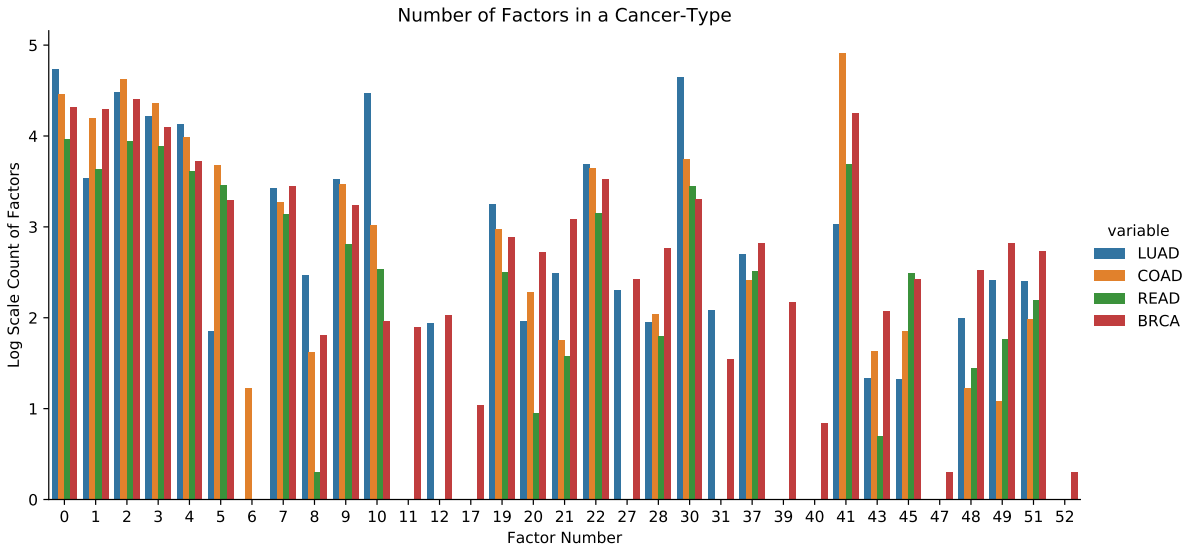


Figure 28: Number of times a factor occurs for all the patients with Lung Adenocarcinoma (LUAD), Colon Cancer (COAD), Renal Cancer (READ), and Breast Cancer (BRCA).

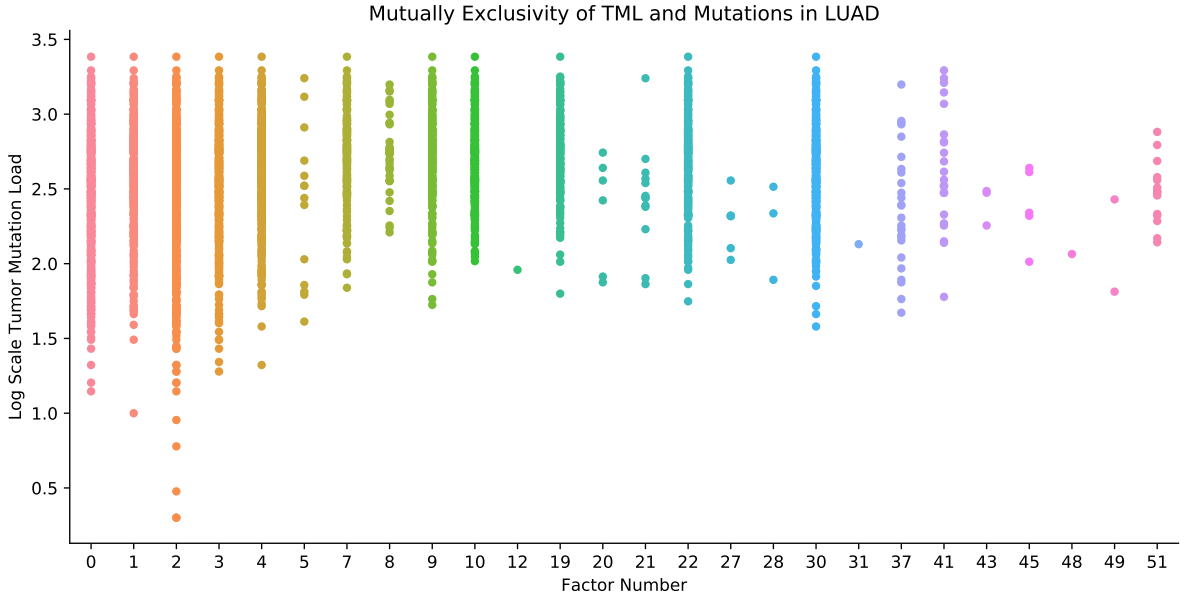


Figure 29: A pattern of mutual exclusivity in LUAD. The X-axis represents a factor, while the Y-axis the total number of mutations in a tumor. As tumor mutation load increases, more unique factors appear such as Factor 7 which has both KRAS and BRAF which are known to be mutually exclusive, but can co-occur together in some subtypes.

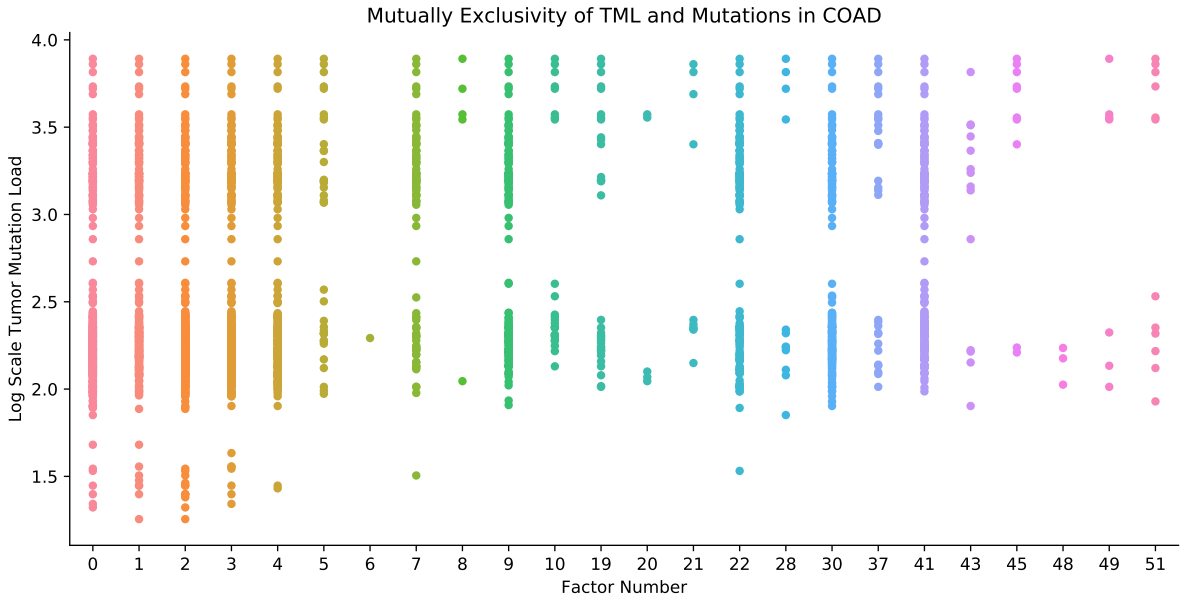


Figure 30: A pattern of mutual exclusivity in COAD. The X-axis represents a factor, while the Y-axis the total number of mutations in a tumor. As tumor mutation load increases, more unique factors appear, specifically Factors 8 and 21 that contain MUC16 and BRCA2 respectively.

CHAPTER 7

CONCLUSIONS

In this dissertation, we applied the tools of machine learning and probabilistic latent variable models to improve the characterization of cancerous lesions. The core challenge of modeling cancerous lesions stems from heterogeneity. Within a patient population of the same cancer, genetic heterogeneity leads to a unique genetic profile of the cancer for each patient. While, clinicians and researchers have tackled the challenge of genetic heterogeneity with *precision medicine* it is hampered by heterogeneity within a lesion itself. The underlying biology of a lesion, where each cell potentially has a distinct mutation, is complex and it is often difficult to identify the phenomena that influences the state of the lesion. This dissertation details the use of machine learning and probabilistic latent variable models to probe patterns that account for lesion heterogeneity for improved differentiation and characterization of lesions.

7.1 Contributions of Thesis

In Chapter 3, we motivated 3D shape features as a tool for exploiting the spatial distribution of voxels in FDG PET/CT images for predicting a patient's treatment response in metastatic liver cancer. We posited that delineation of a lesion explained by the uptake of FDG is potentially controlled by the structure of the lesion tissue, and that non-responding or responding parts of a lesion share similar shape features across the patient population. Formally, structural heterogeneity of lesion tissue is a consequence of lesion heterogeneity. With shape features that

are invariant to deformations and transformations, we captured correspondences between similar patients regardless of artifacts created by the imaging modality or the volume of a lesion. We investigated the performance of shape features and showed that shape features offered better characterization of a lesion's treatment response in comparison to texture or radiological features. We also showed that while invariance to volume improves correspondence between shape features, volume is a critical adjunct measurement to differentiate between responders and non-responders of treatment.

We continued with theme of the distribution of voxels in Chapter 4 by using a histogram to study how the probabilistic distribution of voxels in DWI parameters can differentiate benign and malignant lesions in breast cancer. The IWIM parameters $\mathrm{D_p}$, f,and $\mathrm{D_{diff}}$ probe tissue microstructure and vascularity. The CTRW parameters α and β probe the intra-cellular space within a voxel, while $\mathrm{D_m}$ probes the cellular density within a tissue. These parameters therefore exhibit aspects that reflect cancer progression such as cellular proliferation which has properties of increased cell density and blood flow. Using the kurtosis, skewness, variance, mean, median, interquartile, 10% quantile, 25% quantile, and 75% quantile values of the histogram for each parameter, we determined which metric had the most significance in identifying lesion histology. We presented that the Gradient Boosted Classifier with β (βmedian), skewness of β ($\beta_{skewness}$), mean of β (β_{mean}), third quartile of f (f_{Q3}) , third quartile of $\mathrm{D_{diff}}$ (D^{Q3}_{diff}), kurtosis of $\mathrm{D_{perf}}$ ($D^{kurtosis}_{perf}$), third quartile of $\mathrm{D_m}$ (D^{Q3}_{pm}), and median of $\mathrm{D_m}$ (D^{median}_{m}) as features gave the best performance to differentiate benign and malignant lesions in breast cancer with an F1-score of 0.872.

We developed a deep generative latent variable model LLOST in Chapter 5 with two distinct deep latent variable models (DLVM) to predict somatic mutations from the lesion's 3D image. One DLVM leverages point clouds to create modality invariance of lesion's 3D image, while retaining shape and geometrical properties to extract imaging features. The other DLVM used a Negative Binomial likelihood to capture the sparsity that is present in somatic mutations. The DLVMs are then coupled with an invertible neural network that allows for transfer from one domain to the other domain with a Maximum Mean Discrepancy loss function. Moreover, each DLVM uses Normalizing Flows as a prior to model the unique distributions of each of the domains. We trained and tested LLOST on data from The Cancer Genomic Archive (TCGA) and The Cancer Imaging Archive. Our results showed that we can predict somatic mutations from a lesion's 3D image even if our training dataset consists of a only a few cases a cancer due to correspondence between the image features and co-occurrance patterns of somatic mutations. We further demonstrate the versatility of LLOST by demonstrating the prediction of the entire somatic mutation profile, in comparison to current approaches that predict only a specific subset of somatic mutations.

Using a correlated Bayesian non-parametric model dubbed the CoZINB in Chapter 6, we capture a complex web of variables that reflects the co-occurrence patterns of somatic mutations in cancer. In particular by applying CoZINB to pan-cancer somatic mutation dataset from the TCGA, we demonstrated that augmenting the latent spaces with a latent space that represented non-linear correlations across a set of somatic mutations shows co-occurrence patterns of somatic mutations are influenced by the total number of mutations (TML), biological processes, and

mutational processes. We showed that it is important to take account genetic and lesion heterogeneity across a patient population of cancer and restrict prior knowledge, otherwise a model is prone to under-fitting and over generalizes co-occurring somatic mutations. Specifically, CoZINB creates sets of co-occurring somatic mutations that are distinct to specific a cancer type. We also discussed the biological interpretation of co-occurrence patterns with a perspective from TML in lung (LUAD) and colon (COAD) adenocarcinoma. We showed that TP53 and MUC16 genes in COAD are not mutually exclusive due to biologically redundancy, but due to a increased mutational load ($p - value < 0.01$). Hence, if a treatment only targets the TP53 gene, the MUC16 mutation will still confer a selective advantage to cancer progression leading to metastasis. IN LUAD, KRAS and BRAF are generally known to be mutually exclusive mutations but we showed that both mutations co-occur at higher TML. The co-occurrence pattern is not due to biological redundancy, but due to an underlying biological process suspected to be cell cycle arrest and senescence, which leads to a hyper-mutations and increased TML.

7.2 Future Directions

There are a variety of challenges in characterizing cancerous lesions, and we discuss future work that would seek to create a bridge between medical imaging and the genomic profile of a patient.

Firstly, Chapters 3 and 4 introduced imaging biomarkers that correlated with patient treatment response or lesion histology using a small dataset. Statistical power needs to be increased by increasing the dataset size before considering these imaging biomarkers as an effective diagnostic tool. Another goal is understand the biological mechanisms that influence the

tissue microstructure that is reflected in the imaging features. This would require collaborations of pathologists such that the lesion image is co-registered with a lesion's tissue sample, which is not trivial. Helfrich et al. (301) applied such a paradigm to show low apparent diffusion coefficient (ADC) values reflect the dense cell growth pattern associated with high ratios of malignant glands in prostate cancer.

The lesion point cloud to somatic mutations (LLOST) model of Chapter 5 could readily be extended to capture other aspects of the genomic profile, such as DNA methylation and gene expression. From an exploratory perspective, it would also be insightful to capture the non-linear correlations between genetic data that influence the geometry and shape of a lesion. For instance, a mutation in the EGFR gene often leads to low expression of various negative regulators for EGFR (302). By extending LLOST to include gene expression data as another domain, we would be able predict somatic mutations and gene expression data. The LLOST framework could potentially benefit from a structured prior for the genetic data, such as the one explored in Chapter 6 to evaluate lesion heterogeneity. We envision that this would directly aid clinicians for determining patients that would not respond to treatment due to presence of secondary mutations that also drive cancer.

There is a key trade-off in utilizing prior information in generative probabilistic latent variable models. The sets of co-occurring mutations created by the Correlated Zero Inflated Negative Binomial Process (CoZINB) in Chapter 6 do not always represent biological processes, which impedes biological validation. Extending CoZINB to mix weak prior information such as an interaction network, with the assumption that the interaction network is incomplete would

possibly allow for a more biological interpretation of results. Such an approach falls under the framework of semi-supervised machine learning which is an active area of development. For example, CoZINB can be extended with the idea that somatic mutations appear as a graph with evolving interactions between sets of somatic mutations as well as the arrival and departure dynamics of individual somatic mutations in different sets during cancer progression.

The value of building machine learning models hinges on creating patterns within medical images and genetic data and beyond that mirror the underlying biological process. This requires the collaboration of multiple fields, which could occur as a result of recent advances in obtaining and sharing of medical data. We therefore anticipate that machine learning and latent variable models will provide new ways to ameliorate unique characterization of patients.

APPENDICES

Appendix A

COPYRIGHT OF CHAPTER 3

Appendix B

COPYRIGHT OF CHAPTER 5

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you must follow the requirements listed below:

Textual Material Using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line.

In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Full-Text Article

If you are using the entire IEEE copyright owned article, the following IEEE copyright/ credit notice should be placed prominently in the references. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]

Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line. In placing the thesis on the author's university website, please display the following message in a prominent place on the website:

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of University of Illinois at Chicago's products or services.

Appendix B (Continued)

Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.htmltolearnhowtoobtainaLicensefromRightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Appendix C

COPYRIGHT OF CHAPTER 6

Authors can include partial or complete papers of their own (and no fee is expected) in a dissertation as long as citations and DOI pointers to the Versions of Record in the ACM Digital Library are included. Authors can use any portion of their own work in presentations and in the classroom (and no fee is expected).

REFERENCES

1. Mehta, R., Cai, K., Kumar, N., Knuttinen, M. G., Anderson, T. M., Lu, H., and Lu, Y.: A lesion-based response prediction model using pretherapy pet/ct image features for y90 radioembolization to hepatic malignancies. Technology in cancer research & treatment, 16(5):620–629, 2017.

2. Mehta, R., Karaman, M., Bu, Y., Zhong, Z., Shiwei, W., Zhou, C., Weihong, H., Maosheng, X., and Zhou, J. X.: Discrimination of malignant and benign breast lesions using machine learning on multi-modal diffusion mri parameters. 2020.

3. Mehta, R., Lu, Y., and Karaman, M.: Mapping of lesion images to somatic mutations. In Data and Text Mining in Biomedical Informatics, 2020.

4. Clark Jr, W. H., Elder, D. E., Guerry IV, D., Epstein, M. N., Greene, M. H., and Van Horn, M.: A study of tumor progression: the precursor lesions of superficial spreading and nodular melanoma. Human pathology, 15(12):1147–1165, 1984.

5. Thorlacius, S., Olafsdottir, G., Tryggvadottir, L., Neuhausen, S., Jonasson, J. G., Tavtigian, S. V., Tulinius, H., Ögmundsdottir, H. M., and Eyfjörd, J. E.: A single brca2 mutation in male and female breast cancer families from iceland with varied cancer phenotypes. Nature genetics, 13(1):117–119, 1996.

6. Consortium, B. C. L.: Cancer risks in brca2 mutation carriers. Journal of the National Cancer Institute, 91(15):1310–1316, 1999.

7. Hamilton, W., Peters, T. J., Round, A., and Sharp, D.: What are the clinical features of lung cancer before the diagnosis is made? a population based case-control study. Thorax, 60(12):1059–1065, 2005.

8. Hamilton, W.: The caper studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. British journal of cancer, 101(S2):S80–S86, 2009.

9. Balogh, E. P., Miller, B. T., and Ball, J. R.: The Diagnostic Process. Dec 2015.

10. Collins, F. S. and Varmus, H.: A new initiative on precision medicine. New England journal of medicine, 372(9):793–795, 2015.

11. Hamburg, M. A. and Collins, F. S.: The path to personalized medicine. New England Journal of Medicine, 363(4):301–304, 2010.

12. Stevens, P.: Patterns in Nature. Atlantic Monthly Press book. Little, Brown, 1974.

13. Ripley, B. D.: Pattern recognition and neural networks. Cambridge university press, 2007.

14. Webb, A. R.: Statistical pattern recognition. John Wiley & Sons, 2003.

15. Schwenker, F. and Trentin, E.: Pattern classification and clustering: A review of partially supervised learning approaches. Pattern Recognition Letters, 37:4–14, 2014.

16. Bishop, C. M.: Pattern recognition and machine learning. springer, 2006.

17. Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., et al.: New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). European journal of cancer, 45(2):228–247, 2009.

18. Miglioretti, D. L., Smith-Bindman, R., Abraham, L., Brenner, R. J., Carney, P. A., Bowles, E. J. A., Buist, D. S., and Elmore, J. G.: Radiologist characteristics associated with interpretive performance of diagnostic mammography. Journal of the National Cancer Institute, 99(24):1854–1863, 2007.

19. Al Mohammad, B., Hillis, S., Reed, W., Alakhras, M., and Brennan, P.: Radiologist performance in the detection of lung cancer using ct. Clinical radiology, 74(1):67–75, 2019.

20. Lu, Y.-Y., Chen, J.-H., Liang, J.-A., Wang, H.-Y., Lin, C.-C., Lin, W.-Y., and Kao, C.-H.: Clinical value of fdg pet or pet/ct in urinary bladder cancer: a systemic review and meta-analysis. European journal of radiology, 81(9):2411–2416, 2012.

21. Duijm, L., Louwman, M., Groenewoud, J., Van De Poll-Franse, L., Fracheboud, J., and Coebergh, J. W.: Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. British journal of cancer, 100(6):901–907, 2009.

22. Giger, M. L.: Machine learning in medical imaging. Journal of the American College of Radiology, 15(3):512–520, 2018.

23. Erickson, B. J., Korfiatis, P., Akkus, Z., and Kline, T. L.: Machine learning for medical imaging. Radiographics, 37(2):505–515, 2017.

24. Lyons, M., Akamatsu, S., Kamachi, M., and Gyoba, J.: Coding facial expressions with gabor wavelets. In Proceedings Third IEEE international conference on automatic face and gesture recognition, pages 200–205. IEEE, 1998.

25. Long, J., Shelhamer, E., and Darrell, T.: Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3431–3440, 2015.

26. Moosavi-Dezfooli, S.-M., Fawzi, A., and Frossard, P.: Deepfool: a simple and accurate method to fool deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2574–2582, 2016.

27. Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., and Yu, F.: ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.

28. Wagner, G. P. and Altenberg, L.: Perspective: complex adaptations and the evolution of evolvability. Evolution, 50(3):967–976, 1996.

29. Guinney, J., Dienstmann, R., Wang, X., De Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al.: The consensus molecular subtypes of colorectal cancer. Nature medicine, 21(11):1350–1356, 2015.

30. Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., et al.: Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature, 499(7457):214–218, 2013.

31. Yeang, C.-H., McCormick, F., and Levine, A.: Combinatorial patterns of somatic gene mutations in cancer. The FASEB journal, 22(8):2605–2622, 2008.

32. Ciriello, G., Cerami, E., Sander, C., and Schultz, N.: Mutual exclusivity analysis identifies oncogenic network modules. Genome research, 22(2):398–406, 2012.

33. Boyle, E. A., Li, Y. I., and Pritchard, J. K.: An expanded view of complex traits: from polygenic to omnigenic. Cell, 169(7):1177–1186, 2017.

34. Pritchard, J. K., Stephens, M., and Donnelly, P.: Inference of population structure using multilocus genotype data. Genetics, 155(2):945–959, 2000.

35. Blei, D. M., Ng, A. Y., and Jordan, M. I.: Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022, 2003.

36. Gopalan, P. K., Charlin, L., and Blei, D.: Content-based recommendations with poisson factorization. In Advances in Neural Information Processing Systems, pages 3176–3184, 2014.

37. Aerts, H. J.: The potential of radiomic-based phenotyping in precision medicine: a review. JAMA oncology, 2(12):1636–1642, 2016.

38. Mazurowski, M. A.: Radiogenomics: what it is and why it is important. Journal of the American College of Radiology, 12(8):862–866, 2015.

39. Cancer imaging archive wiki.

40. Orel, S. G., Kay, N., Reynolds, C., and Sullivan, D. C.: Bi-rads categorization as a predictor of malignancy. Radiology, 211(3):845–850, 1999.

41. Manos, D., Seely, J. M., Taylor, J., Borgaonkar, J., Roberts, H. C., and Mayo, J. R.: The lung reporting and data system (lu-rads): a proposal for computed tomography screening. Canadian Association of Radiologists Journal, 65(2):121–134, 2014.

42. of Radiology, A. C. et al.: Breast imaging reporting and data system. BI-RADS, 2003.

43. Lin, G., Keshari, K. R., and Park, J. M.: Cancer metabolism and tumor heterogeneity: imaging perspectives using mr imaging and spectroscopy. Contrast Media & Molecular Imaging, 2017.

44. Hu, L. S., Ning, S., Eschbacher, J. M., Baxter, L. C., Gaw, N., Ranjbar, S., Plasencia, J., Dueck, A. C., Peng, S., Smith, K. A., et al.: Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. Neuro-oncology, 19(1):128–137, 2017.

45. Liu, D., Ba, Z., Ni, X., Wang, L., Yu, D., and Ma, X.: Apparent diffusion coefficient to subdivide breast imaging reporting and data system magnetic resonance imaging (bi-rads-mri) category 4 lesions. Medical science monitor: international medical journal of experimental and clinical research, 24:2180, 2018.

46. Hicks, R. J., Kalff, V., MacManus, M. P., Ware, R. E., Hogg, A., McKenzie, A. F., Matthews, J. P., and Ball, D. L.: 18f-fdg pet provides high-impact and powerful prognostic stratification in staging newly diagnosed non–small cell lung cancer. Journal of Nuclear Medicine, 42(11):1596–1604, 2001.

47. Joye, I., Deroose, C. M., Vandecaveye, V., and Haustermans, K.: The role of diffusion-weighted mri and 18f-fdg pet/ct in the prediction of pathologic complete response after radiochemotherapy for rectal cancer: a systematic review. Radiotherapy and oncology, 113(2):158–165, 2014.

48. Baliyan, V., Das, C. J., Sharma, R., and Gupta, A. K.: Diffusion weighted imaging: technique and applications. World journal of radiology, 8(9):785, 2016.

49. Petrillo, A., Fusco, R., Petrillo, M., Granata, V., Delrio, P., Bianco, F., Pecori, B., Botti, G., Tatangelo, F., Caracò, C., et al.: Standardized index of shape (dce-mri) and standardized uptake value (pet/ct): two quantitative approaches to discriminate chemo-radiotherapy locally advanced rectal cancer responders under a functional profile. Oncotarget, 8(5):8143, 2017.

50. Kinahan, P. E. and Fletcher, J. W.: Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. In Seminars in Ultrasound, CT and MRI, volume 31, pages 496–505. Elsevier, 2010.

51. Kwee, T. C., Galbán, C. J., Tsien, C., Junck, L., Sundgren, P. C., Ivancevic, M. K., Johnson, T. D., Meyer, C. R., Rehemtulla, A., Ross, B. D., et al.: Comparison of apparent diffusion coefficients and distributed diffusion coefficients in high-grade gliomas. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 31(3):531–537, 2010.

52. Long, N. M. and Smith, C. S.: Causes and imaging features of false positives and false negatives on 18 f-pet/ct in oncologic imaging. Insights into imaging, 2(6):679–698, 2011.

53. Bruening, W., Fontanarosa, J., Tipton, K., Treadwell, J. R., Launders, J., and Schoelles, K.: Systematic review: comparative effectiveness of core-needle and open surgical biopsy

to diagnose breast lesions. Annals of internal medicine, 152(4):238–246, 2010.

54. of Clinical Oncology, A. S.: Biopsy, 2019.

55. Group, E. B. C. T. C. et al.: Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. The Lancet, 366(9503):2087–2106, 2005.

56. Gevaert, O., Xu, J., Hoang, C. D., Leung, A. N., Xu, Y., Quon, A., Rubin, D. L., Napel, S., and Plevritis, S. K.: Non–small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. Radiology, 264(2):387–396, 2012.

57. Yamamoto, S., Maki, D. D., Korn, R. L., and Kuo, M. D.: Radiogenomic analysis of breast cancer using mri: a preliminary study to define the landscape. American Journal of Roentgenology, 199(3):654–663, 2012.

58. Diehn, M., Nardini, C., Wang, D. S., McGovern, S., Jayaraman, M., Liang, Y., Aldape, K., Cha, S., and Kuo, M. D.: Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proceedings of the National Academy of Sciences, 105(13):5213–5218, 2008.

59. Chin, L. and Gray, J. W.: Translating insights from the cancer genome into clinical practice. Nature, 452(7187):553–563, 2008.

60. Kingma, D. P. and Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.

61. Dinh, L., Krueger, D., and Bengio, Y.: Nice: Non-linear independent components estimation. arXiv preprint arXiv:1410.8516, 2014.

62. Mehta, R. and Karaman, M.: Correlated mixed membership modeling of somatic mutation profiles. In International Joint Conference on Neural Networks. IEEE, 2020.

63. Noy, A., Schöder, H., Gönen, M., Weissler, M., Ertelt, K., Cohler, C., Portlock, C., Hamlin, P., and Yeung, H.: The majority of transformed lymphomas have high standardized uptake values (suvs) on positron emission tomography (pet) scanning similar to diffuse large b-cell lymphoma (dlbcl). Annals of Oncology, 20(3):508–512, 2009.

64. Cottereau, A.-S., Lanic, H., Mareschal, S., Meignan, M., Vera, P., Tilly, H., Jardin, F., and Becker, S.: Molecular profile and fdg-pet/ct total metabolic tumor volume improve risk classification at diagnosis for patients with diffuse large b-cell lymphoma. Clinical Cancer Research, 22(15):3801–3809, 2016.

65. Choi, H., Charnsangavej, C., Faria, S., Macapinlac, H., Burgess, M., Patel, SR, C., and L.L.: Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J Clin Oncol, 25:1753–1759.

66. de Geus-Oei, L.-F., Vriens, D., van Laarhoven, H. W., van der Graaf, W. T., and Oyen, W. J.: Monitoring and predicting response to therapy with 18f-fdg pet in colorectal cancer: a systematic review. Journal of Nuclear Medicine, 50(Suppl 1):43S–54S, 2009.

67. Martini, N., Kris, M. G., Flehinger, B. J., Gralla, R. J., Bains, M. S., Burt, M. E., Heelan, R., McCormack, P. M., Pisters, K. M., Rigas, J. R., et al.: Preoperative chemotherapy for stage iiia (n2) lung cancer: the sloan-kettering experience with 136 patients. The Annals of thoracic surgery, 55(6):1365–1374, 1993.

68. Mikhaeel, N., Timothy, A., O'doherty, M., Hain, S., and Maisey, M.: 18-fdg-pet as a prognostic indicator in the treatment of aggressive non-hodgkin's lymphoma-comparison with ct. Leukemia & lymphoma, 39(5-6):543–553, 2000.

69. Gordon, Y., Partovi, S., Müller-Eschner, M., Amarteifio, E., Bäuerle, T., Weber, M.-A., Kauczor, H.-U., and Rengier, F.: Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. Cardiovascular diagnosis and therapy, 4(2):147, 2014.

70. Trimboli, R. M., Verardi, N., Cartia, F., Carbonaro, L. A., and Sardanelli, F.: Breast cancer detection using double reading of unenhanced mri including t1-weighted, t2-weighted stir, and diffusion-weighted imaging: a proof of concept study. American Journal of Roentgenology, 203(3):674–681, 2014.

71. Upadhyay, N. and Waldman, A.: Conventional mri evaluation of gliomas. The British journal of radiology, 84(special_issue_2):S107–S111, 2011.

72. Verma, S., Turkbey, B., Muradyan, N., Rajesh, A., Cornud, F., Haider, M. A., Choyke, P. L., and Harisinghani, M.: Overview of dynamic contrast-enhanced mri in

prostate cancer diagnosis and management. American Journal of Roentgenology, 198(6):1277–1288, 2012.

73. Hara, N., Okuizumi, M., Koike, H., Kawaguchi, M., and Bilim, V.: Dynamic contrast-enhanced magnetic resonance imaging (dce-mri) is a useful modality for the precise detection and staging of early prostate cancer. The Prostate, 62(2):140–147, 2005.

74. Yang, X. and Knopp, M. V.: Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: a review. BioMed Research International, 2011.

75. Lee, S. J., Li, X., Huang, H., and Zelen, M.: The dana-farber cisnet model for breast cancer screening strategies: An update. Medical Decision Making, 38(1_suppl):44S–53S, 2018.

76. Dorrius, M. D., Jansen-van der Weide, M. C., van Ooijen, P. M., Pijnappel, R. M., and Oudkerk, M.: Computer-aided detection in breast mri: a systematic review and meta-analysis. European radiology, 21(8):1600–1608, 2011.

77. Le Bihan, D. and Breton, E.: Imagerie de diffusion in vivo par résonance magnétique nucléaire. Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre, 301(15):1109–1112, 1985.

78. Koh, D.-M. and Collins, D. J.: Diffusion-weighted mri in the body: applications and challenges in oncology. American Journal of Roentgenology, 188(6):1622–1635, 2007.

79. Guo, Y., Cai, Y.-Q., Cai, Z.-L., Gao, Y.-G., An, N.-Y., Ma, L., Mahankali, S., and Gao, J.-H.: Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 16(2):172–178, 2002.

80. Muraoka, N., Uematsu, H., Kimura, H., Imamura, Y., Fujiwara, Y., Murakami, M., Yamaguchi, A., and Itoh, H.: Apparent diffusion coefficient in pancreatic cancer: characterization and histopathological correlations. Journal of Magnetic Resonance Imaging, 27(6):1302–1308, 2008.

81. Chen, L., Liu, M., Bao, J., Xia, Y., Zhang, J., Zhang, L., Huang, X., and Wang, J.: The correlation between apparent diffusion coefficient and tumor cellularity in patients: a meta-analysis. PloS one, 8(11):e79008, 2013.

82. Le Bihan, D., Poupon, C., Amadon, A., and Lethimonnier, F.: Artifacts and pitfalls in diffusion mri. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 24(3):478–488, 2006.

83. Karaman, M. M., Wang, H., Sui, Y., Engelhard, H. H., Li, Y., and Zhou, X. J.: A fractional motion diffusion model for grading pediatric brain tumors. NeuroImage: Clinical, 12:707–714, 2016.

84. Hauser, T., Essig, M., Jensen, A., Laun, F. B., Münter, M., Maier-Hein, K. H., and Stieltjes, B.: Prediction of treatment response in head and neck carcinomas using ivim-dwi: evaluation of lymph node metastasis. European journal of radiology, 83(5):783–787, 2014.

85. Zhang, Y.-D., Wang, Q., Wu, C.-J., Wang, X.-N., Zhang, J., Liu, H., Liu, X.-S., and Shi, H.-B.: The histogram analysis of diffusion-weighted intravoxel incoherent motion (ivim) imaging for differentiating the gleason grade of prostate cancer. European radiology, 25(4):994–1004, 2015.

86. Iima, M., Nobashi, T., Imai, H., Koyasu, S., Saga, T., Nakamoto, Y., Kataoka, M., Yamamoto, A., Matsuda, T., and Togashi, K.: Effects of diffusion time on non-gaussian diffusion and intravoxel incoherent motion (ivim) mri parameters in breast cancer and hepatocellular carcinoma xenograft models. Acta radiologica open, 7(1):2058460117751565, 2018.

87. Foulds, L.: The natural history of cancer. Journal of chronic diseases, 8(1):2–37, 1958.

88. Heppner, G. H.: Tumor heterogeneity. Cancer research, 44(6):2259–2265, 1984.

89. Easwaran, H., Tsai, H.-C., and Baylin, S. B.: Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Molecular cell, 54(5):716–727, 2014.

90. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al.: Signatures of mutational processes in human cancer. Nature, 500(7463):415–421, 2013.

91. Ayob, A. Z. and Ramasamy, T. S.: Cancer stem cells as key drivers of tumour progression. Journal of biomedical science, 25(1):1–18, 2018.

92. Dagogo-Jack, I. and Shaw, A. T.: Tumour heterogeneity and resistance to cancer therapies. Nature reviews Clinical oncology, 15(2):81, 2018.

93. Sasaki, H., Hikosaka, Y., Kawano, O., Moriyama, S., Yano, M., and Fujii, Y.: Evaluation of kras gene mutation and copy number gain in non-small cell lung cancer. Journal of Thoracic Oncology, 6(1):15–20, 2011.

94. Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., Stephens, P. J., Daniels, G. A., and Kurzrock, R.: Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Molecular cancer therapeutics, 16(11):2598–2608, 2017.

95. Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B., Veeriah, S., Shafi, S., Johnson, D. H., Mitter, R., Rosenthal, R., et al.: Tracking the evolution of non–small-cell lung cancer. New England Journal of Medicine, 376(22):2109–2121, 2017.

96. Alizadeh, A. A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., et al.: Toward understanding and exploiting tumor heterogeneity. Nature medicine, 21(8):846, 2015.

97. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., and Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data, 4:170117, 2017.

98. Kwon, S. H., Yoon, J.-K., An, Y.-S., Shin, Y. S., Kim, C.-H., Lee, D. H., Jo, K. S., and Lee, S. J.: Prognostic significance of the intratumoral heterogeneity of 18f-fdg uptake in oral cavity cancer. Journal of surgical oncology, 110(6):702–706, 2014.

99. Lima, M., Kataoka, M., Honda, M., et al.: Time-varying diffusion patterns in breast cancer linked to prognostic factors. In 27th annual meeting & exhibition of the international society of magnetic resonance in medicine, 2019.

100. Naber, S. P., Tsutsumi, Y., Yin, S., Zolnay, S. A., Mobtaker, H., Marks, P. J., Mckenzie, S. J., Delellis, R. A., and Wolfe, H. J.: Strategies for the analysis of oncogene

overexpression: studies of the neu oncogene in breast carcinoma. American journal of clinical pathology, 94(2):125–136, 1990.

101. Davnall, F., Yip, C. S., Ljungqvist, G., Selmi, M., Ng, F., Sanghera, B., Ganeshan, B., Miles, K. A., Cook, G. J., and Goh, V.: Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights into imaging, 3(6):573–589, 2012.

102. Siegel, R. L., Miller, K. D., and Jemal, A.: Cancer statistics, 2020. CA: a cancer journal for clinicians, 70(1):7–30, 2020.

103. Khan, S. A., Thomas, H. C., Davidson, B. R., and Taylor-Robinson, S. D.: Cholangiocarcinoma. The Lancet, 366(9493):1303–1314, 2005.

104. Smith, M. and McCall, J.: Systematic review of tumour number and outcome after radical treatment of colorectal liver metastases. British journal of surgery, 96(10):1101–1113, 2009.

105. Dawson, L. A.: Overview: where does radiation therapy fit in the spectrum of liver cancer local-regional therapies? In Seminars in Radiation Oncology, volume 21, pages 241–246. Elsevier, 2011.

106. DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., Jemal, A., and Siegel, R. L.: Breast cancer statistics, 2019. CA: a cancer journal for clinicians, 69(6):438–451, 2019.

107. Szabo, B. K., Aspelin, P., Kristoffersen Wiberg, M., and Boné, B.: Dynamic mr imaging of the breast: analysis of kinetic and morphologic diagnostic criteria. Acta radiologica, 44(4):379–386, 2003.

108. Mussurakis, S., Buckley, D., and Horsman, A.: Dynamic mr imaging of invasive breast cancer: correlation with tumour grade and other histological factors. The British journal of radiology, 70(833):446–451, 1997.

109. Bone, B., Aspelin, P., Bronge, L., and Veress, B.: Contrast-enhanced mr imaging as a prognostic indicator of breast cancer. Acta radiologica, 39(3):279–284, 1998.

110. Fan, M., Wu, G., Cheng, H., Zhang, J., Shao, G., and Li, L.: Radiomic analysis of dce-mri for prediction of response to neoadjuvant chemotherapy in breast cancer patients. European journal of radiology, 94:140–147, 2017.

111. Fan, M., Li, H., Wang, S., Zheng, B., Zhang, J., and Li, L.: Radiomic analysis reveals dce-mri features for prediction of molecular subtypes of breast cancer. PLoS One, 12(2):e0171683, 2017.

112. Braman, N. M., Etesami, M., Prasanna, P., Dubchuk, C., Gilmore, H., Tiwari, P., Plecha, D., and Madabhushi, A.: Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast dce-mri. Breast Cancer Research, 19(1):1–14, 2017.

113. Vidić, I., Egnell, L., Jerome, N. P., Teruel, J. R., Sjøbakk, T. E., Østlie, A., Fjøsne, H. E., Bathen, T. F., and Goa, P. E.: Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: Preliminary study. Journal of Magnetic Resonance Imaging, 2018.

114. Horvat, J. V., Bernard-Davila, B., Helbich, T. H., Zhang, M., Morris, E. A., Thakur, S. B., Ochoa-Albiztegui, R. E., Leithner, D., Marino, M. A., Baltzer, P. A., Clauser, P., Kapetas, P., Bago-Horvath, Z., and Pinker, K.: Diffusion-weighted imaging (DWI) with apparent diffusion coefficient (ADC) mapping as a quantitative imaging biomarker for prediction of immunohistochemical receptor status, proliferation rate, and molecular subtypes of breast cancer. Journal of Magnetic Resonance Imaging, 2019.

115. Iima, M., Honda, M., Sigmund, E. E., Ohno Kishimoto, A., Kataoka, M., and Togashi, K.: Diffusion MRI of the breast: Current status and future directions, 2020.

116. Karaman, M. M., Sui, Y., Wang, H., Magin, R. L., Li, Y., and Zhou, X. J.: Differentiating low- and high-grade pediatric brain tumors using a continuous-time random-walk diffusion model at high b-values. Magnetic Resonance in Medicine, 76(4):1149–1157, 2016.

117. Berkowitz, D., Hushon, J. M., Whitfield, H. J., Roth, J., and Ames, B. N.: Procedure for identifying nonsense mutations. Journal of Bacteriology, 96(1):215–220, 1968.

118. Huang, N., Shah, P. K., and Li, C.: Lessons from a decade of integrating cancer copy number alterations with gene expression profiles. Briefings in bioinformatics, 13(3):305–316, 2012.

119. Lloyd, S.: Least squares quantization in pcm. IEEE transactions on information theory, 28(2):129–137, 1982.

120. Tan, X. and Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE transactions on image processing, 19(6):1635–1650, 2010.

121. Dhawan, A. P., Chitre, Y., and Kaiser-Bonasso, C.: Analysis of mammographic microcalcifications using gray-level image structure features. IEEE Transactions on medical imaging, 15(3):246–259, 1996.

122. Al-Kadi, O. S., Watson, D., et al.: Texture analysis of aggressive and nonaggressive lung tumor ce ct images. IEEE transactions on biomedical engineering, 55(7):1822–1830, 2008.

123. Ganeshan, B., Miles, K., Young, R., and Chatwin, C.: Hepatic entropy and uniformity: additional parameters that can potentially increase the effectiveness of contrast enhancement during abdominal ct. Clinical radiology, 62(8):761–768, 2007.

124. Eliat, P.-A., Olivié, D., Saïkali, S., Carsin, B., Saint-Jalmes, H., and de Certaines, J. D.: Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumors from other glioblastoma? Neurology research international, 2012.

125. Haralick, R. M., Shanmugam, K., and Dinstein, I. H.: Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6):610–621, 1973.

126. Ondimu, S. and Murase, H.: Effect of probability-distance based markovian texture extraction on discrimination in biological imaging. Computers and Electronics in Agriculture, 63(1):2–12, 2008.

127. Kim, J.-H., Ko, E. S., Lim, Y., Lee, K. S., Han, B.-K., Ko, E. Y., Hahn, S. Y., and Nam, S. J.: Breast cancer heterogeneity: Mr imaging texture analysis and survival outcomes. Radiology, 282(3):665–675, 2017.

128. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I.: A survey on deep learning in medical image analysis. Medical image analysis, 42:60–88, 2017.

129. Gatys, L., Ecker, A. S., and Bethge, M.: Texture synthesis using convolutional neural networks. In Advances in neural information processing systems, pages 262–270, 2015.

130. LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning. nature, 521(7553):436–444, 2015.

131. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. C., Darzi, A., et al.: International evaluation of an ai system for breast cancer screening. Nature, 577(7788):89–94, 2020.

132. Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012, 2015.

133. Dollár, P., Wojek, C., Schiele, B., and Perona, P.: Pedestrian detection: A benchmark. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 304–311. IEEE, 2009.

134. Haralick, R. M., Joo, H., Lee, C.-N., Zhuang, X., Vaidya, V. G., and Kim, M. B.: Pose estimation from corresponding point data. IEEE Transactions on Systems, Man, and Cybernetics, 19(6):1426–1446, 1989.

135. Tian, Y., Pei, K., Jana, S., and Ray, B.: Deeptest: Automated testing of deep-neural-network-driven autonomous cars. In Proceedings of the 40th international conference on software engineering, pages 303–314, 2018.

136. Sadjadi, F. A. and Hall, E. L.: Three-dimensional moment invariants. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2):127–136, 1980.

137. Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D.: Matching 3d models with shape distributions. In Proceedings International Conference on Shape Modeling and Applications, pages 154–166. IEEE, 2001.

138. Shilane, P., Min, P., Kazhdan, M., and Funkhouser, T.: The princeton shape benchmark. In Proceedings Shape Modeling Applications, 2004., pages 167–178. IEEE, 2004.

139. Bronstein, A., Bronstein, M., Guibas, L., and Ovsjanikov, M.: Shape google: Geometric words and expressions for invariant shape retrieval. ACM Trans. Graph, 30:1–20.

140. Qi, C. R., Su, H., Mo, K., and Guibas, L. J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 652–660, 2017.

141. Qi, C. R., Yi, L., Su, H., and Guibas, L. J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In Advances in neural information processing systems, pages 5099–5108, 2017.

142. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B.: Pointcnn: Convolution on x-transformed points. In Advances in neural information processing systems, pages 820–830, 2018.

143. Mnih, A. and Salakhutdinov, R. R.: Probabilistic matrix factorization. In Advances in neural information processing systems, pages 1257–1264, 2008.

144. Reynolds, D. A.: Gaussian mixture models. Encyclopedia of biometrics, 741, 2009.

145. Griffiths, T. L. and Ghahramani, Z.: The indian buffet process: An introduction and review. Journal of Machine Learning Research, 12(4), 2011.

146. Allison, D. B., Cui, X., Page, G. P., and Sabripour, M.: Microarray data analysis: from disarray to consolidation and consensus. Nature reviews genetics, 7(1):55–65, 2006.

147. Hotelling, H.: Analysis of a complex of statistical variables into principal components. Journal of educational psychology, 24(6):417, 1933.

148. Tipping, M. E. and Bishop, C. M.: Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3):611–622, 1999.

149. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N.: Deep generative modeling for single-cell transcriptomics. Nature methods, 15(12):1053–1058, 2018.

150. Ingraham, J., Garg, V., Barzilay, R., and Jaakkola, T.: Generative models for graph-based protein design. In Advances in Neural Information Processing Systems, pages 15820–15831, 2019.

151. Knowles, D. and Ghahramani, Z.: Infinite sparse factor analysis and infinite independent components analysis. In International Conference on Independent Component Analysis and Signal Separation, pages 381–388. Springer, 2007.

152. Ferguson, T. S.: A bayesian analysis of some nonparametric problems. The annals of statistics, pages 209–230, 1973.

153. Rasmussen, C. E.: The infinite gaussian mixture model. In Advances in neural information processing systems, pages 554–560, 2000.

154. Kallenberg, O.: Foundations of modern probability. Springer Science & Business Media, 2006.

155. Kingman, J.: Completely random measures. Pacific Journal of Mathematics, 21(1):59–78, 1967.

156. Escobar, M. D. and West, M.: Bayesian density estimation and inference using mixtures. Journal of the american statistical association, 90(430):577–588, 1995.

157. Thibaux, R. and Jordan, M. I.: Hierarchical beta processes and the indian buffet process. In Artificial Intelligence and Statistics, pages 564–571, 2007.

158. Hjort, N. L. et al.: Nonparametric bayes estimators based on beta processes in models for life history data. the Annals of Statistics, 18(3):1259–1294, 1990.

159. Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I.: An introduction to mcmc for machine learning. Machine learning, 50(1-2):5–43, 2003.

160. Kennedy, A., Nag, S., Salem, R., Murthy, R., McEwan, A. J., Nutting, C., Espat, J., Bilbao, J. I., Sharma, R. A., Thomas, J. P., et al.: Recommendations for radioembolization of hepatic malignancies using yttrium-90 microsphere brachytherapy: a consensus panel report from the radioembolization brachytherapy oncology consortium. International Journal of Radiation Oncology* Biology* Physics, 68(1):13–23, 2007.

161. Lewandowski, R., Memon, K., Mulcahy, M., Hickey, R., Marshall, K., Williams, M., and Salzig, K.: Twelve-year experience of radioembolization for colorectal hepatic metastases in 214 patients: survival by era and chemotherapy. Eur J Nucl Med Mol Imaging. doi: 1810.1007/s00259-00014-02799-00252. Epub 02014 Jun 00257.

162. Paprottka, P., Hoffmann, R., Haug, A., Sommer, W., Raessler, F., Trumm, C., and Schmidt, G.: Radioembolization of symptomatic, unresectable neuroendocrine hepatic metastases using yttrium-90 microspheres. Cardiovasc Intervent Radiol, 35:334–342.

163. Salem, R., Lewandowski, R., Mulcahy, M., Riaz, A., Ryu, R., Ibrahim, S., and Atassi, B.: Radioembolization for hepatocellular carcinoma using yttrium-90 microspheres: a comprehensive report of long-term outcomes. Gastroenterology, 138:52–64.

164. Rafi, S., Piduru, S., El-Rayes, B., Kauh, J., Kooby, D., Sarmiento, J., and Kim, H.: Yttrium-90 radioembolization for unresectable standard-chemorefractory intrahepatic cholangiocarcinoma: survival, efficacy, and safety study. Cardiovasc Intervent Radiol, 36(440-448):410 1007 00270–00012–00463–00274. Epub 02012 Sep 00277.

165. Young, H., Baum, R., Cremerius, U., Herholz, K., Hoekstra, O., Lammertsma, A., and Pruim, J.: Measurement of clinical and subclinical tumour response using [18f]-fluorodeoxyglucose and positron emission tomography: review and 1999 eortc recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. Eur J Cancer, 35:1773–1782.

166. Wahl, R., Jacene, H., Kasamon, Y., and Lodge, M.: From recist to percist: Evolving considerations for pet response criteria in solid tumors. J Nucl Med, 1(122).

167. Eisenhauer, E., Therasse, P., Bogaerts, J., Schwartz, L., Sargent, D., Ford, R., and Dancey, J.: New response evaluation criteria in solid tumours: revised recist guideline (version 1.1. Eur J Cancer, 45:228–247.

168. Fendler, W., Philippe Tiega, D., Ilhan, H., Paprottka, P., Heinemann, V., Jakobs, T., and Bartenstein, P.: Validation of several suv-based parameters derived from 18f-fdg pet for prediction of survival after sirt of hepatic metastases from colorectal cancer. J Nucl Med. doi: 1210.2967/jnumed.1112.116426. Epub 112013 May 116431.

169. Zalom, M., Yu, R., Friedman, M., Bresee, C., and Waxman, A.: Fdg pet/ct as a prognostic test after 90y radioembolization in patients with metastatic hepatic disease. Clin Nucl Med, 37:862–865.

170. Sabet, A., Ahmadzadehfar, H., Bruhman, J., Sabet, A., Meyer, C., Wasmuth, J., and Pieper, C.: Survival in patients with hepatocellular carcinoma treated with 90y-microsphere radioembolization. In Prediction by 18F-FDG PET. Nuklearmedizin 2014;53:39-45.

171. Filippi, L., Pelle, G., Cianni, R., Scopinaro, F., and Bagni, O.: Change in total lesion glycolysis and clinical outcome after y radioembolization in intrahepatic cholangiocarcinoma. Nucl Med Biol, 2014(011).

172. Pugachev, A., Ruan, S., Carlin, S., Larson, S., Campa, J., Ling, C., and Humm, J.: Dependence of fdg uptake on tumor microenvironment. International Journal of Radiation Oncology* Biology* Physics, 62:545–553.

173. Esfahani, S., Heidari, P., Halpern, E., Hochberg, E., Palmer, E., and Mahmood, U.: Baseline total lesion glycolysis measured with (18)f-fdg pet/ct as a predictor of progression-free survival in diffuse large b-cell lymphoma: a pilot study. American Journal of Nuclear Medicine and Molecular Imaging, 3:272–281.

174. Itti, E., Juweid, M., Haioun, C., Yeddes, I., Hamza-Maaloul, F., El Bez, I., and Evangelista, E.: Improvement of early 18f-fdg pet interpretation in diffuse large b-cell lymphoma: importance of the reference background. Journal of Nuclear Medicine, 51:1857–1862.

175. Tan, S., Kligerman, S., Chen, W., Lu, M., Kim, G., Feigenberg, S., and D'Souza, W.: Spatial-temporal [(1)(8)f]fdg-pet features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. Int J Radiat Oncol Biol Phys, 85:1375–1382.

176. Salem, R. and Thurston, K.: Radioembolization with 90yttrium microspheres: a state-of-the-art brachytherapy treatment for primary and secondary liver malignancies. part 1: Technical and methodologic considerations. J Vasc Interv Radiol, 17:1251–1278.

177. Miller, F. H., Keppke, A. L., Reddy, D., Huang, J., Jin, J., Mulcahy, M. F., and Salem, R.: Response of liver metastases after treatment with yttrium-90 microspheres: role of size, necrosis, and pet. American Journal of Roentgenology, 188(3):776–783, 2007.

178. Abdullah, A., Hirayama, A., Yatsushiro, S., Matsumae, M., and Kuroda, K.: Cerebrospinal fluid image segmentation using spatial fuzzy clustering method with improved evolutionary expectation maximization. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, page 3359–3362.

179. Rupa, S., Mohan, V., and Venkataramani, Y.: Mri brain image compression using spatial fuzzy clustering technique. In Communications and Signal Processing (ICCSP), 2014 International Conference on; 2014, page 915–919. IEEE.

180. Shyu, C.-R., Brodley, C., Kak, A., Kosaka, A., Aisen, A., and Broderick, L.: Local versus global features for content-based image retrieval. In Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on; 1998, page 30–34. IEEE.

181. Venkatraman, V., Sael, L., and Kihara, D.: Potential for protein surface shape analysis using spherical harmonics and 3d zernike descriptors. Cell biochemistry and biophysics,

54:23–32.

182. Amadasun, M. and King, R.: Textural features corresponding to textural properties. Systems, Man and Cybernetics, IEEE Transactions on, 19:1264–1274.

183. Aubry, M., Schlickewei, U., and Cremers, D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, page 1626–1633.

184. Novotni, M. and Klein, R.: 3d zernike descriptors for content based shape retrieval. In Proceedings of the eighth ACM symposium on Solid modeling and applications, page 216–225, Seattle, Washington, USA. ACM.

185. Lazebnik, S., Schmid, C., and Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In 2006: IEEE, page 2169–2178.

186. Kibriya, A., Frank, E., Pfahringer, B., and Holmes, G.: Multinomial naive bayes for text categorization revisited. In AI 2004: Advances in Artificial Intelligence, eds. G. Webb and X. Yu, page 488–499. Cairns, Australia, Springer.

187. Ridgeway, G., Madigan, D., Richardson, T., and O'Kane, J.: Interpretable boosted naive bayes classification. In Proceedings of the 4th international conference on knowledge discovery in database, page 101–104.

188. Tixier, F., Le Rest, C., Hatt, M., Albarghach, N., Pradier, O., Metges, J., and Corcos, L.: Intratumor heterogeneity characterized by textural features on baseline 18f-fdg pet images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med, 52:369–378.

189. Waclaw, B., Bozic, I., Pittman, M., Hruban, R., Vogelstein, B., and Nowak, M.: A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature, 525:261–264.

190. Sardanelli, F., Boetes, C., Borisch, B., Decker, T., Federico, M., Gilbert, F. J., Helbich, T., Heywang-Köbrunner, S. H., Kaiser, W. A., Kerin, M. J., Mansel, R. E., Marotti, L., Martincich, L., Mauriac, L., Meijers-Heijboer, H., Orecchia, R., Panizza, P., Ponti, A., Purushotham, A. D., Regitnig, P., Del Turco, M. R., Thibault, F., and Wilson, R.: Magnetic resonance imaging of the breast: Recommendations from the EUSOMA working group. European Journal of Cancer, 46:1296–1316, 2010.

191. Lehman, C. D., Gatsonis, C., Kuhl, C. K., Hendrick, R. E., Pisano, E. D., Hanna, L., Peacock, S., Smazal, S. F., Maki, D. D., Julian, T. B., DePeri, E. R., Bluemke, D. A., and Schnall, M. D.: MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. New England Journal of Medicine, 2007.

192. Saslow, D., Boetes, C., Burke, W., Harms, S., Leach, M. O., Lehman, C. D., Morris, E., Pisano, E., Schnall, M., Sener, S., Smith, R. A., Warner, E., Yaffe, M., Andrews, K. S., and Russell, C. A.: American Cancer Society Guidelines for Breast Screening with MRI as an Adjunct to Mammography. CA: A Cancer Journal for Clinicians, 57:75–89, 2007.

193. Partridge, S. C. and McDonald, E. S.: Diffusion weighted MRI of the breast: Protocol optimization, guidelines for interpretation, and potential clinical applications. Magnetic Resonance Imaging Clinics of North America, 21(3):601–624, 2013.

194. Partridge, S. C., Nissan, N., Rahbar, H., Kitsch, A. E., and Sigmund, E. E.: Diffusion-weighted breast MRI: Clinical applications and emerging techniques, 2017.

195. Dietzel, M., Ellmann, S., Schulz-Wendtland, R., Clauser, P., Wenkel, E., Uder, M., and Baltzer, P. A.: Breast MRI in the era of diffusion weighted imaging: do we still need signal-intensity time curves? European Radiology, 30(1):47–56, 2020.

196. Wu, L.-M., Hu, J.-N., Gu, H.-Y., Hua, J., Chen, J., and Xu, J.-R.: Can diffusion-weighted MR imaging and contrast-enhanced MR imaging precisely evaluate and predict pathological response to neoadjuvant chemotherapy in patients with breast cancer? Breast Cancer Research and Treatment, 135(1):17–28, 2012.

197. Partridge, S. C., DeMartini, W. B., Kurland, B. F., Eby, P. R., White, S. W., and Lehman, C. D.: Quantitative diffusion-weighted imaging as an adjunct to conventional breast MRI for improved positive predictive value. American Journal of Roentgenology, 193(6):1716–1722, 2009.

198. Anastasio, M. A. and La Riviere, P.: Emerging imaging technologies in medicine. 2012.

199. Le Bihan, D., Breton, E., Lallemand, D., Aubin, M. L., Vignaud, J., and Laval-Jeantet, M.: Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology, 168(2):497–505, 1988.

200. Tang, L. and Zhou, X. J.: Diffusion mri of cancer: From low to high b-values. Journal of Magnetic Resonance Imaging, 49(1):23–40, 2019.

201. Song, Y. S., Park, C. M., Lee, S. M., Park, S. J., Cho, H. R., Choi, S. H., Lee, J. M., Kiefer, B., and Goo, J. M.: Reproducibility of histogram and texture parameters derived from intravoxel incoherent motion diffusion-weighted mri of fn13762 rat breast carcinomas. Anticancer Research, 34(5):2135–2144, 2014.

202. Iima, M. and Le Bihan, D.: Clinical intravoxel incoherent motion and diffusion MR imaging: past, present, and future. Radiology, 278(1):13–32, 2016.

203. Iima, M., Yano, K., Kataoka, M., Umehana, M., Murata, K., Kanao, S., Togashi, K., and Le Bihan, D.: Quantitative Non-Gaussian Diffusion and Intravoxel Incoherent Motion Magnetic Resonance Imaging. Investigative Radiology, 50(4):205–211, apr 2015.

204. Suo, S., Lin, N., Wang, H., Zhang, L., Wang, R., Zhang, S., Hua, J., and Xu, J.: Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: Comparison of different curve-fitting methods. Journal of Magnetic Resonance Imaging, 42(2):362–370, 2015.

205. Cho, G. Y., Moy, L., Zhang, J. L., Baete, S., Lattanzi, R., Moccaldi, M., Babb, J. S., Kim, S., Sodickson, D. K., and Sigmund, E. E.: Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer. Magnetic Resonance in Medicine, 74(4):1077–1085, 2015.

206. Sigmund, E. E., Cho, G. Y., Kim, S., Finn, M., Moccaldi, M., Jensen, J. H., Sodickson, D. K., Goldberg, J. D., Formenti, S., and Moy, L.: Intravoxel incoherent motion imaging of tumor microenvironment in locally advanced breast cancer. Magnetic Resonance in Medicine, 65(5):1437–1447, 2011.

207. Bokacheva, L., Kaplan, J. B., Giri, D. D., Patil, S., Gnanasigamani, M., Nyman, C. G., Deasy, J. O., Morris, E. a., and Thakur, S. B.: Intravoxel incoherent motion diffusion-weighted MRI at 3.0 T differentiates malignant breast lesions from benign lesions and breast parenchyma. Journal of Magnetic Resonance Imaging, 40(4):813–23, 2014.

208. Liu, C., Liang, C., Liu, Z., Zhang, S., and Huang, B.: Intravoxel incoherent motion (IVIM) in evaluation of breast lesions:Comparison with conventional DWI. European Journal of Radiology, 2013.

209. Kim, H. J. and Kim, W.: Method of tumor volume evaluation using magnetic resonance imaging for outcome prediction in cervical cancer treated with concurrent chemotherapy and radiotherapy. Radiation Oncology Journal, 2012.

210. Karaman, M. M., Sui, Y., Wang, H., Magin, R. L., Li, Y., and Zhou, X. J.: Differentiating low- and high-grade pediatric brain tumors using a continuous-time random-walk diffusion model at high b-values. Magnetic Resonance in Medicine, 2016.

211. Zhong, Z., Merkitch, D., Karaman, M., Zhang, J., Sui, Y., Goldman, J., and Zhou, X.: High-spatial-resolution diffusion MRI in Parkinson disease: Lateral asymmetry of the substantia nigra. Radiology, 291(1):149–157, 2019.

212. Zhou, X. J., Gao, Q., Abdullah, O., and Magin, R. L.: Studies of anomalous diffusion in the human brain using fractional order calculus. Magnetic Resonance in Medicine, 2010.

213. Magin, R. L., Abdullah, O., Baleanu, D., and Zhou, X. J.: Anomalous diffusion expressed through fractional order differential operators in the Bloch-Torrey equation. Journal of Magnetic Resonance Imaging, 190:255–270, 2008.

214. Ingo, C., Magin, R. L., Colon-Perez, L., Triplett, W., and Mareci, T. H.: On random walks and entropy in diffusion-weighted magnetic resonance imaging studies of neural tissue. Magnetic Reson Med, 71(2):617–627, 2014.

215. Sui, Y., Xiong, Y., Xie, K. L., Karaman, M. M., Zhu, W., and Zhou, X. J.: Differentiation of low- and high-grade gliomas using high b-value diffusion imaging with a non-Gaussian diffusion model. American Journal of Neuroradiology, 37:1643–1649, 2016.

216. Sui, Y., He, W., Damen, F. W., Wanamaker, C., Li, Y., and Zhou, X. J.: Differentiation of low- and high-grade pediatric brain tumors with high b-value diffusion weighted MR imaging and a fractional order calculus model. Radiology, 277(2):489–496, 2015.

217. Tang, L., Sui, Y., Zhong, Z., Damen, F. C., Li, J., Shen, L., Sun, Y., and Zhou, X. J.: Non-Gaussian diffusion imaging with a fractional order calculus model to predict response of gastrointestinal stromal tumor to second-line sunitinib therapy. Magnetic Resonance in Medicine, 79(3):1399–1406, 2018.

218. Bedair, R., Priest, A. N., Patterson, A. J., McLean, M. A., Graves, M. J., Manavaki, R., Gill, A. B., Abeyakoon, O., Griffiths, J. R., and Gilbert, F. J.: Assessment of early treatment response to neoadjuvant chemotherapy in breast cancer using non-mono-exponential diffusion models: a feasibility study comparing the baseline and mid-treatment MRI examinations. European Radiology, 27(7):2726–2736, 2017.

219. Dietrich, O., Biffar, A., Baur-Melnyk, A., and Reiser, M. F.: Technical aspects of MR diffusion imaging of the body. European Journal of Radiology, 76(3):314–322, 2010.

220. Şerifoğlu, İ., Oz, İ. İ., Damar, M., Tokgöz, Ö., Yazgan, Ö., and Erdem, Z.: Diffusion-weighted imaging in the head and neck region: usefulness of apparent diffusion coefficient values for characterization of lesions. Diagnostic and Interventional Radiology, 21(3):208, 2015.

221. Pereira, F. P. A., Martins, G., Figueiredo, E., Domingues, M. N. A., Domingues, R. C., Da Fonseca, L. M. B., and Gasparetto, E. L.: Assessment of breast lesions with diffusion-weighted MRI: Comparing the use of different b values. American Journal of Roentgenology, 2009.

222. Min, Q., Shao, K., Zhai, L., Liu, W., Zhu, C., Yuan, L., and Yang, J.: Differential diagnosis of benign and malignant breast masses using diffusion-weighted magnetic resonance imaging. World Journal of Surgical Oncology, 2015.

223. Hu, Q., Whitney, H. M., Giger, M. L., Edwards, A., and Papaioannou, J.: Radiomics and deep learning of diffusion-weighted MRI in the diagnosis of breast cancer. 2019.

224. Cho, G. Y., Moy, L., Kim, S. G., Baete, S. H., Moccaldi, M., Babb, J. S., Sodickson, D. K., and Sigmund, E. E.: Evaluation of breast cancer using intravoxel incoherent motion (IVIM) histogram analysis: comparison with malignant status, histological subtype, and molecular prognostic factors. European Radiology, 26(8):2547–2558, 2016.

225. Chen, F., Chen, P., Muhammed, H. H., and Zhang, J.: Intravoxel Incoherent Motion Diffusion for Identification of Breast Malignant and Benign Tumors Using Chemometrics. BioMed Research International, 2017.

226. Fan, Y. and Gao, J. H.: Fractional motion model for characterization of anomalous diffusion from NMR signals. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 2015.

227. Miller, A. J. and Joseph, P. M.: The use of power images to perform quantitative analysis on low SNR MR images. Magnetic Resonance Imaging, 1993.

228. Luciani, A., Vignaud, A., Cavet, M., Van Nhieu, J. T., Mallat, A., Ruel, L., Laurent, A., Deux, J. F., Brugieres, P., and Rahmouni, A.: Liver cirrhosis: Intravoxel incoherent motion MR imaging-pilot study. Radiology, 2008.

229. Kursa, M. B. and Rudnicki, W. R.: Feature selection with the boruta package. Journal of Statistical Software, 2010.

230. Burges, C. J.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998.

231. Seeger, M.: Gaussian processes for machine learning., 2004.

232. Safavian, S. R. and Landgrebe, D.: A Survey of Decision Tree Classifier Methodology. IEEE Transactions on Systems, Man and Cybernetics, 1991.

233. Breiman, L.: Random forests. Machine Learning, 2001.

234. Friedman, J. H.: Greedy Function Approximation : A Gradient Boosting Machine 1 Function estimation 2 Numerical optimization in function space. North, 1999.

235. Schapire, R. E.: A brief introduction to boosting. In IJCAI International Joint Conference on Artificial Intelligence, 1999.

236. Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. International Joint Conference of Artificial Intelligence, 1995.

237. Bedard, P. L., Hansen, A. R., Ratain, M. J., and Siu, L. L.: Tumour heterogeneity in the clinic, 2013.

238. Murdoch, D. and Sager, J.: Will targeted therapy hold its promise? an evidence-based review. Current opinion in oncology, 20(1):104–111, 2008.

239. Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., Dekker, A., et al.: Radiomics: extracting more information from medical images using advanced feature analysis. European journal of cancer, 48(4):441–446, 2012.

240. Nair, V. S., Gevaert, O., Davidzon, G., Napel, S., Graves, E. E., Hoang, C. D., Shrager, J. B., Quon, A., Rubin, D. L., and Plevritis, S. K.: Prognostic pet 18f-fdg uptake imaging features are associated with major oncogenomic alterations in patients with resected non–small cell lung cancer. Cancer research, 72(15):3725–3734, 2012.

241. Bai, H. X., Lee, A. M., Yang, L., Zhang, P., Davatzikos, C., Maris, J. M., and Diskin, S. J.: Imaging genomics in cancer research: limitations and promises. The British journal of radiology, 89(1061):20151030, 2016.

242. Pharoah, P. D., Dunning, A. M., Ponder, B. A., and Easton, D. F.: Association studies for finding cancer-susceptibility genetic variants. Nature Reviews Cancer, 4(11):850–860, 2004.

243. Balak, M. N., Gong, Y., Riely, G. J., Somwar, R., Li, A. R., Zakowski, M. F., Chiang, A., Yang, G., Ouerfelli, O., Kris, M. G., et al.: Novel d761y and common secondary t790m mutations in epidermal growth factor receptor–mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. Clinical cancer research, 12(21):6494–6501, 2006.

244. Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., and Campbell, P. J.: Universal patterns of selection in cancer and somatic tissues. Cell, 171(5):1029–1041, 2017.

245. Wang, S. and Summers, R. M.: Machine learning and radiology. Medical image analysis, 16(5):933–951, 2012.

246. Sun, B., Feng, J., and Saenko, K.: Return of frustratingly easy domain adaptation. In Thirtieth AAAI Conference on Artificial Intelligence, 2016.

247. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.

248. Rezende, D. J. and Mohamed, S.: Variational inference with normalizing flows. arXiv preprint arXiv:1505.05770, 2015.

249. Wang, L., Li, Y., and Lazebnik, S.: Learning deep structure-preserving image-text embeddings. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5005–5013, 2016.

250. Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 2223–2232, 2017.

251. Dinh, L., Sohl-Dickstein, J., and Bengio, S.: Density estimation using real nvp. arXiv preprint arXiv:1605.08803, 2016.

252. Kingma, D. P. and Dhariwal, P.: Glow: Generative flow with invertible 1x1 convolutions. In Advances in Neural Information Processing Systems, pages 10215–10224, 2018.

253. Tran, E., Robbins, P. F., and Rosenberg, S. A.: 'final common pathway'of human cancer immunotherapy: targeting random somatic mutations. Nature immunology, 18(3):255, 2017.

254. Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., Barron, D. A., Zehir, A., Jordan, E. J., Omuro, A., et al.: Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nature genetics, 51(2):202–206, 2019.

255. Brock, A., Lim, T., Ritchie, J. M., and Weston, N.: Generative and discriminative voxel modeling with convolutional neural networks. arXiv preprint arXiv:1608.04236, 2016.

256. Ziegler, Z. M. and Rush, A. M.: Latent normalizing flows for discrete sequences. arXiv preprint arXiv:1901.10548, 2019.

257. Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E. W., Klessen, R. S., Maier-Hein, L., Rother, C., and Köthe, U.: Analyzing inverse problems with invertible neural networks. arXiv preprint arXiv:1808.04730, 2018.

258. Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., and Welling, M.: Improved variational inference with inverse autoregressive flow. In Advances in neural information processing systems, pages 4743–4751, 2016.

259. Krishnan, R. G., Liang, D., and Hoffman, M.: On the challenges of learning with inference networks on sparse, high-dimensional data. arXiv preprint arXiv:1710.06085, 2017.

260. Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., et al.: Comprehensive characterization of cancer driver genes and mutations. Cell, 173(2):371–385, 2018.

261. Zhao, H., Rai, P., Du, L., Buntine, W. L., and Zhou, M.: Variational autoencoders for sparse and overdispersed discrete data. CoRR, abs/1905.00616, 2019.

262. Rasiwasia, N., Costa Pereira, J., Coviello, E., Doyle, G., Lanckriet, G. R., Levy, R., and Vasconcelos, N.: A new approach to cross-modal multimedia retrieval. In Proceedings of the 18th ACM international conference on Multimedia, pages 251–260, 2010.

263. Tomczak, J. M. and Welling, M.: Vae with a vampprior. arXiv preprint arXiv:1705.07120, 2017.

264. Sohn, K., Lee, H., and Yan, X.: Learning structured output representation using deep conditional generative models. In Advances in neural information processing systems, pages 3483–3491, 2015.

265. Mahajan, S., Gurevych, I., and Roth, S.: Latent normalizing flows for many-to-many cross domain mappings. In International Conference on Learning Representations, 2020.

266. Xie, Z. and Ma, S.: Dual-view variational autoencoders for semi-supervised text matching. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, pages 5306–5312. AAAI Press, 2019.

267. Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., Kovatich, A. J., Benz, C. C., Levine, D. A., Lee, A. V., et al.: An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell, 173(2):400–416, 2018.

268. Amidror, I.: Scattered data interpolation methods for electronic imaging systems: a survey. Journal of electronic imaging, 11(ARTICLE):157–76, 2002.

269. Blei, D. M.: Probabilistic topic models. Communications of the ACM, 55(4):77–84, 2012.

270. Sønderby, C. K., Raiko, T., Maaløe, L., Sønderby, S. K., and Winther, O.: Ladder variational autoencoders. In Advances in neural information processing systems, pages 3738–3746, 2016.

271. Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B.: Adversarial autoencoders. arXiv preprint arXiv:1511.05644, 2015.

272. Achlioptas, P., Diamanti, O., Mitliagkas, I., and Guibas, L.: Learning representations and generative models for 3d point clouds. arXiv preprint arXiv:1707.02392, 2017.

273. He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

274. Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., Chmielecki, J., et al.: Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome medicine, 9(1):34, 2017.

275. Zhou, M. and Carin, L.: Negative binomial process count and mixture modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(2):307–320, 2013.

276. Ducreux, M., Chamseddine, A., Laurent-Puig, P., Smolenschi, C., Hollebecque, A., Dartigues, P., Samallin, E., Boige, V., Malka, D., and Gelli, M.: Molecular targeted therapy of braf-mutant colorectal cancer. Therapeutic advances in medical oncology, 11:1758835919856494, 2019.

277. van de Haar, J., Canisius, S., Michael, K. Y., Voest, E. E., Wessels, L. F., and Ideker, T.: Identifying epistasis in cancer genomes: a delicate affair. Cell, 177(6):1375–1383, 2019.

278. Bauer, B., Siebert, R., and Traulsen, A.: Cancer initiation with epistatic interactions between driver and passenger mutations. Journal of Theoretical Biology, 358:52–60, 2014.

279. Leiserson, M. D., Vandin, F., Wu, H.-T., Dobson, J. R., Eldridge, J. V., Thomas, J. L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., et al.: Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nature genetics, 47(2):106, 2015.

280. Zhang, J. and Zhang, S.: The discovery of mutated driver pathways in cancer: Models and algorithms. IEEE/ACM transactions on computational biology and bioinformatics, 15(3):988–998, 2016.

281. Kanehisa, M. and Goto, S.: Kegg: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1):27–30, 2000.

282. Cowen, L., Ideker, T., Raphael, B. J., and Sharan, R.: Network propagation: a universal amplifier of genetic associations. Nature Reviews Genetics, 18(9):551, 2017.

283. Yin, H. and Gai, K.: An empirical study on preprocessing high-dimensional class-imbalanced data for classification. In 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems, pages 1314–1319, 2015.

284. Morgillo, F., Della Corte, C. M., Fasano, M., and Ciardiello, F.: Mechanisms of resistance to egfr-targeted drugs: lung cancer. ESMO open, 1(3):e000060, 2016.

285. Zhang, A. and Paisley, J.: Random function priors for correlation modeling. arXiv preprint arXiv:1905.03826, 2019.

286. Ranganath, R. and Blei, D. M.: Correlated random measures. Journal of the American Statistical Association, 113(521):417–430, 2018.

287. Paisley, J., Wang, C., and Blei, D.: The discrete infinite logistic normal distribution for mixed-membership modeling. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 74–82, 2011.

288. Anders, S. and Huber, W.: Differential expression analysis for sequence count data. Nature Precedings, pages 1–1, 2010.

289. Dadaneh, S. Z., Qian, X., and Zhou, M.: Bnp-seq: Bayesian nonparametric differential expression analysis of sequencing count data. Journal of the American Statistical Association, 113(521):81–94, 2018.

290. Liu, B., Liu, L., Tsykin, A., Goodall, G. J., Green, J. E., Zhu, M., Kim, C. H., and Li, J.: Identifying functional mirna–mrna regulatory modules with correspondence latent dirichlet allocation. Bioinformatics, 26(24):3105–3111, 2010.

291. Teh, Y. W. and Jordan, M. I.: Hierarchical bayesian nonparametric models with applications. Bayesian nonparametrics, 1:158–207, 2010.

292. Lafferty, J. D. and Blei, D. M.: Correlated topic models. In Advances in neural information processing systems, pages 147–154, 2006.

293. Carlin, B. P. and Louis, T. A.: Bayes and empirical Bayes methods for data analysis. Chapman and Hall/CRC, 2010.

294. Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J.: Stochastic variational inference. The Journal of Machine Learning Research, 14(1):1303–1347, 2013.

295. Srivastava, A. and Sutton, C.: Autoencoding variational inference for topic models. arXiv preprint arXiv:1703.01488, 2017.

296. Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., and Scholkopf, B.: Support vector machines. IEEE Intelligent Systems and their applications, 13(4):18–28, 1998.

297. Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

298. Network, C. G. A. R. et al.: Comprehensive molecular profiling of lung adenocarcinoma. Nature, 511(7511):543, 2014.

299. Cisowski, J. and Bergo, M. O.: What makes oncogenes mutually exclusive? Small GTPases, 8(3):187–192, 2017.

300. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., et al.: The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, 2012.

301. Helfrich, O., Puech, P., Betrouni, N., Pinçon, C., Ouzzane, A., Rizk, J., Marcq, G., Randazzo, M., Durand, M., Lakroum, S., et al.: Quantified analysis of histological components and architectural patterns of gleason grades in apparent diffusion coefficient restricted areas upon diffusion weighted mri for peripheral or transition zone cancer locations. Journal of Magnetic Resonance Imaging, 46(6):1786–1796, 2017.

302. Yang, C.-H., Chou, H.-C., Fu, Y.-N., Yeh, C.-L., Cheng, H.-W., Chang, I.-C., Liu, K.-J., Chang, G.-C., Tsai, T.-F., Tsai, S.-F., et al.: Egfr over-expression in non-small cell lung cancers harboring egfr mutations is associated with marked down-regulation of cd82. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1852(7):1540–1549, 2015.

VITA

Rahul Mehta /

Education

2013–Present **Ph.D in Bioengineering**, *University of Illinois at Chicago*, Chicago,IL,
Dissertation: Patterns Among and Between Somatic Mutations and Medical Imaging in Human Cancers.
Committee: Muge Karaman, Jie Liang, Yang Dai, Yang Lu

2005–2011 **Bs.C in Electrical Engineering**, *University of Illinois at Urbana Champaign*, Urbana-Champaign,IL.

Academic Experience

Research

2019–Present **Center for Magnetic Resonance Research**, *UIC*, Chicago.
Research topics: Designed machine learning models for predicting lesion type and it's treatment response in Diffusion Weighted MRI images.

2014–Present **Laboratory of Computational Proteomics**, *UIC*, Chicago.
Research topics: Probabilistic models with efficient and flexible Bayesian inference algorithms for learning the underlying structure of cancerous somatic mutation data. Built a multi-domain learning model for translation between cancerous 3D lesion images and their corresponding somatic mutation data.

Teaching

2014–2016 **Imaging Informatics**, *UIC*, Chicago.
Designed a course teaching extraction of information from medical images for machine learning tasks. Topics included: Image Processing and Computer Vision

2014–2016 **Biodatamining**, *UIC*, Chicago.
Designed a course teaching data-mining in biological and medical sets. Topics included: Dimensionality Reduction, Text Mining, Neural Networks, and Visualization Techniques

Papers

Journal and Conference Papers

2020 Mehta, R. and M. Karaman. "Correlated Mixed Membership Modeling of Somatic Mutation Profiles". *In International Joint Conference on Neural Networks*. IEEE. 2020.

Mehta, R., M. Karaman, and Y. Lu. "Mapping of Lesion Images to Somatic Mutations". *In Data and Text Mining in Biomedical Informatics*. ACM. 2020.

2017 Mehta, R., K. Cai, N. Kumar, M. G. Knuttinen, T. M. Anderson, H. Lu, and Y. Lu. "A lesion-based response prediction model using pretherapy PET/CT image features for Y90 radioembolization to hepatic malignancies". *In Technology in cancer research & treatment* 16.5 (2017), pp. 620–629.

Workshop Papers

2018 Mehta, R. and H. Lu. "Bayesian Power Law Models for Somatic Mutation Profiles". *In the ISMB workshop on Machine Learning in Computational and Systems Biology*. 2018.

Mehta, R. and H. Lu. "Normalized Random Measure Mixture Models in Variational Autoencoders". *In the NeurIPS workshop on Advances in Approximate Bayesian Inference*. 2018. URL: `http://approximateinference.org/2018/accepted/MehtaLu2018.pdf`.

Mehta, R. and H. Lu. "Power Law Models in Somatic Mutation Profiles". *In the IJCAI workshop on Biomedical Informatics with Optimization and Machine Learning*. 2018.

Manuscripts in Progress

2020 Mehta, R., M. Karaman, Y. Bu, Z. Zhong, W. Shiwei, C. Zhou, H. Weihong, X. Maosheng, and Z. J. Xiaohong. "Discrimination of Malignant and Benign Breast Lesions Using Machine Learning on Multi-modal Diffusion MRI Parameters". *In* 2020.

Presentations

2020 Correlated Mixed Membership Modeling of Somatic Mutations.
International Joint Conference on Neural Networks, 2020. **Contributed Talk**

2020 Diffusion-Weighted MRI-Based Quantitative Markers for Characterizing Breast Cancer Lesions Using Machine Learning.
The International Society for Magnetic Resonance in Medicine, 2020. **Poster**

2019 Non-parametric Models of Somatic Mutation Profiles.
IEEE Biomedical Health Informatics Special Session on Nonparametric Statistics in Omics Applications, 2019. **Invited Talk**

2018 Power Law Models in Somatic Mutation Profiles.
IJCAI workshop on Biomedical Informatics with Optimization and Machine Learning, 2018. **Contributed Talk**

2018 Bayesian Power Law Models for Somatic Mutation Profiles.
Intelligent Systems for Molecular Biology, 2018. **Poster**

2018 Normalized Random Measure Mixture Models in Variational Autoencoders.
NeurIPS workshop on Advances in Approximate Bayesian Inference, 2018. **Poster**

2014 Computer Aided Response Prediction Based on Pre-therapy FDG PET/CT Imaging Biomarkers of Y90-SIRT Therapy in Patients with Primary and Metastatic Liver Cancers.
Radiological Society of North America, 2014. **Contributed Talk**

Research Support

2013 UIC National Institute of Health T32 Training Grant

2016 NVIDIA Hardware Grant