8-2018

PREDICTION OF 1P/19Q CODELETION STATUS IN DIFFUSE GLIOMA PATIENTS USING PREOPERATIVE MULTIPARAMETRIC MAGNETIC RESONANCE IMAGING

Donnie Kim

PREDICTION OF 1P/19Q CODELETION STATUS IN DIFFUSE GLIOMA PATIENTS USING PREOPERATIVE MULTIPARAMETRIC MAGNETIC RESONANCE IMAGING

By

Donnie Duckyeom Kim, B.S.

APPROVED:

_____
Arvind Rao, Ph.D.
Advisory Professor

_____
Richard Wendt, Ph.D.
On-site Advisory Professor

_____
Ankit Patel, Ph.D.

_____
Ganesh Rao, M.D.

_____
David Fuentes, Ph.D.

APPROVED:

_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

PREDICTION OF 1P/19Q CODELETION STATUS IN DIFFUSE GLIOMA PATIENTS USING PREOPERATIVE MULTIPARAMETRIC MAGNETIC RESONANCE IMAGING

A

THESIS

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

In Partial Fulfillment

Of the Requirements

For the Degree of

MASTER OF SCIENCE

by

Donnie Duckyeom Kim, B.S.

Houston, Texas

August, 2018

Acknowledgement

First off, I would like to thank Dr. Mattox for supporting me financially until my graduation.

Also, I would like to thank my friends, especially Laura Bennett and Constance Owens.

I would like to thank every committee member who guided me through my thesis work. Especially, I would like to thank Dr. Wendt for discussing my thesis problems one on one with me on a weekly basis.

A special thanks to Dr. Ma who came to my defense as a substitute with such a short notice.

Thank you to Dr. Arvind Rao for mentoring me for the past two years. Not only did you help me throughout this project, but you also taught me to become a better scientist in general.

Lastly, but most importantly, I want to thank my parents, my brother, and sister-in-law. They have been a tremendous support.

Abstract

PREDICTION OF 1P/19Q CODELETION STATUS IN DIFFUSE GLIOMA PATIENTS USING PREOPERATIVE MULTIPARAMETRIC MAGNETIC RESONANCE IMAGING

Donnie Duckyeom Kim, B.S.

Advisory Professor: Arvind Rao, Ph.D.

A complete codeletion of chromosome 1p/19q is strongly correlated with better overall survival of diffuse glioma patients, hence determining the codeletion status early in the course of a patient's disease would be valuable in that patient's care. The current practice requires a surgical biopsy in order to assess the codeletion status, which exposes patients to risks and is limited in its accuracy by sampling variations. To overcome such limitations, we utilized four conventional magnetic resonance imaging sequences to predict the 1p/19q status. We extracted three sets of image-derived features, namely texture-based, topology-based, and convolutional neural network (CNN)-based, and analyzed each feature's prediction performance. The topology-based model (AUC = 0.855 +/- 0.079) performed significantly better compared to the texture-based model (AUC = 0.707 +/- 0.118) while comparably against the CNN-based model (0.787 +/- 0.195). However, none of the models performed better than the baseline model that is built with only clinical variables, namely, age, gender, and Karnofsky Performance Score (AUC = 0.703 +/- 0.256). In summary, predicting 1p/19q chromosome codeletion status via MRI scan analysis can be a viable non-invasive assessment tool at an early stage of gliomas and in follow-ups although further investigation is needed to improve the model performance.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 New glioma classification in 2016

Prior to 2016, gliomas, which are the most common primary brain tumors growing from glial cells, were classified simply based on their histologic phenotypes. The World Health Organization (WHO) classified oligodendrogliomas, astrocytomas, and oligoastrocytomas as lower grade gliomas (LGG) with grades II and III. Among these LGG patients, about 25 to 50% would experience recurrence or develop into a high-grade gliomas (HGG), known as glioblastoma with WHO grade IV (1, 2). Compared to LGG, HGG is much more aggressive and it has a poorer prognosis. For LGG cases, the five year survival rates range from 29.8% to 81.3% depending on subtypes whereas, for HGG, the five year survival rate is only 5.5% (3) .

In 2016, WHO announced its new classification. For the diagnosis of LGG, the isocitrate dehydrogenase (IDH) mutation status and 1p/19q codeletions status are now considered alongside of the histologic phenotype (4) to reflect the recent reports showing that gliomas may have very different clinical responses and behaviors based on their molecular marker status (5, 6). Figure 1 demonstrates how this new classification works based on histological and genetic features. For both the LGG and HGG cases, the IDH mutation status is first confirmed. For LGGs, the IDH mutation status is further categorized based on the 1p/19q chromosome codeletion status. In sum, the molecular subtypes of LGGs can be divided into three classes: IDH wild type, IDH1 mutated and no 1p/19q codeletion, and IDH mutated and 1p/19q codeletion.

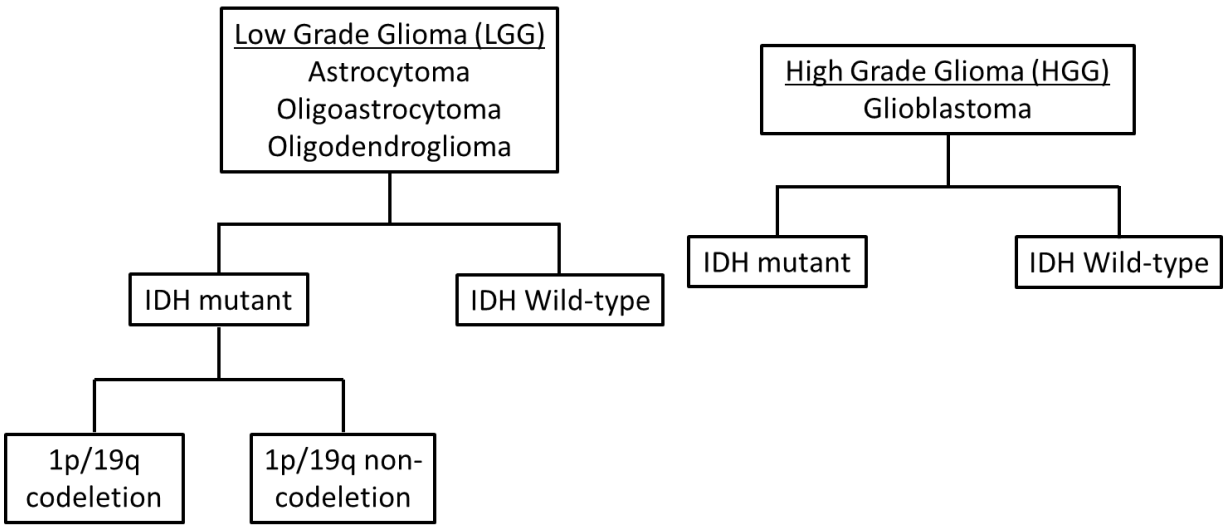


Figure 1 New classification of diffuse glioma. First, they are classified by IDH mutation status. Then, for IDH mutant LGG cases, they are further classified by their 1p/19q chromosome codeletion status. Redrawn with permission. D. N. Louis et al., "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," Acta Neuropathologica, vol. 131, no. 6. Springer Berlin Heidelberg, pp. 803–820, 09-Jun-2016e.

1.2 1p/19q chromosomal codeletion

1p/19q codeletion is defined as a complete deletion of both the short arm of chromosome 1, or 1p, and the long arm of chromosome 19, or 19q, hence 1p/19q codeletion. As described in the previous section, 1p/19q codeletion is a molecular genetic feature of LGG that accounts for approximately 10 to 15% of adult patients (3). As a prognostic biomarker in glioma, the presence of the 1p/19q codeletion is strongly associated with better survival (5, 7–10). According to The Cancer Genome Atlas Research Network (TCGA), LGG patients with the IDH mutation and a 1p/19q codeletion status had a median overall survival (OS) of 8.0 years, whereas it was only 6.3 OS years in patients with an IDH mutation and *no* 1p/19q codeletion (5) (Figure 2). As a predictive biomarker, 1p/19q codeletion status is a good indicator for chemotherapy responses. According to Cairncross et al., patients with the 1p/19q codeletetion had twice the median OS of those with the wildtype gene (14.7 years vs. 7.3 years) under the treatment of procarbazine/lomustine/vincristine (PCV) chemotherapy in combination with radiotherapy (7).

Overall, the status of 1p/19q codeletion as a prognostic or predictive biomarker is crucial for patients' treatment and thereby their OS.

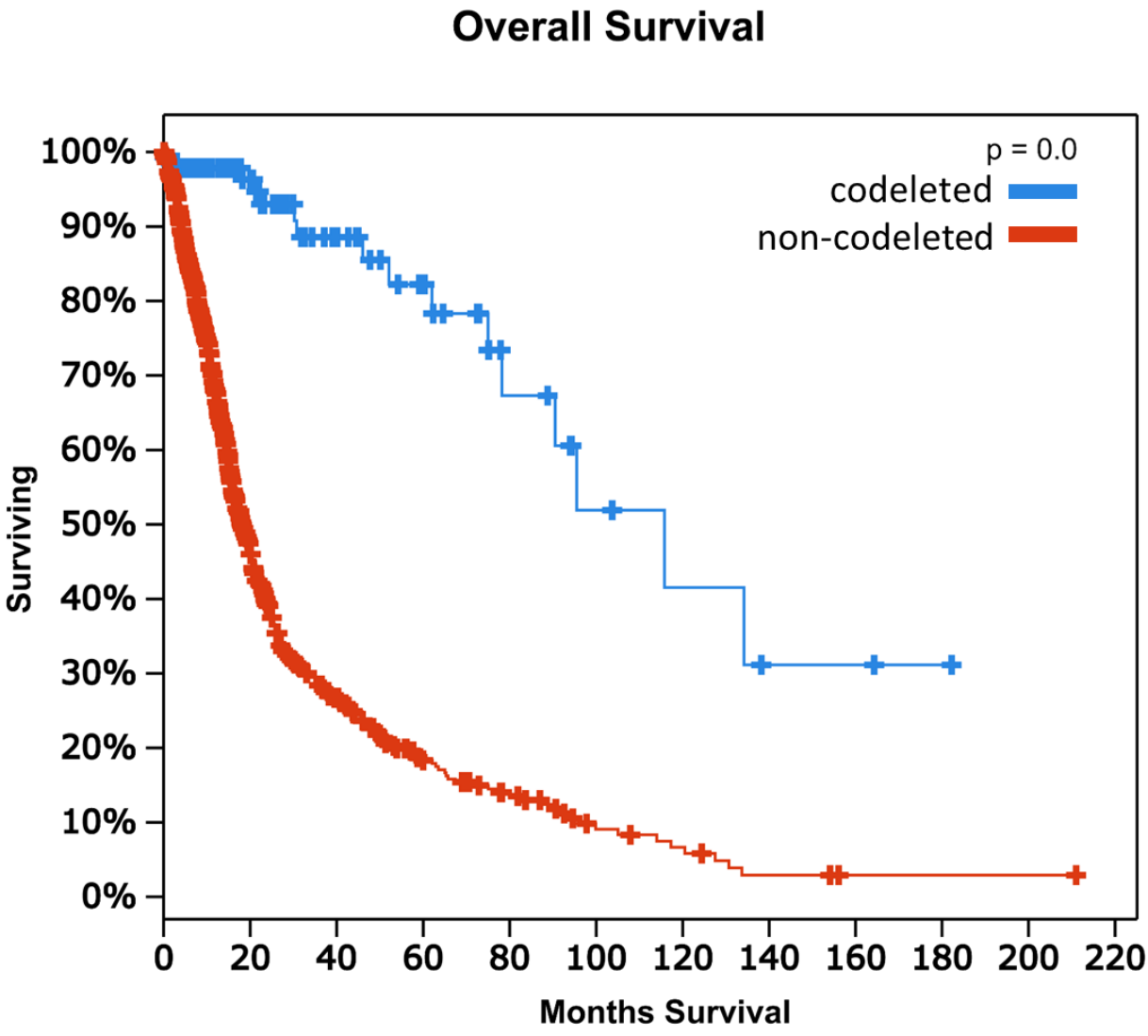


Figure 2 Survival analysis between 1p/19q codeletion and non-codeletion. This graph was generated via CBioPortal using the TCGA-LGG and TCGA-GBM cases.

1.3 Determining 1p/19q codeletion status and Radiogenomics

Currently, determining a patient's 1p/19q codeletion status requires a surgical biopsy followed by fluorescence in-situ hybridization (FISH) analysis of the specimen (11). Such practice, however, exposes patients to risks and is limited by sampling variations, which suggests the need for a non-invasive way of assessing 1p/19q codeletion status.

As a non-invasive determination of 1p/19q codeletion status, radiogenomics, a field of study investigating the quantitative relationship between medical image features and genomic markers, has been investigated. In 2012, Jansen et al. reported that 18-F fluoroethyltyrosine-positron emission tomography (FET-PET) image feature analysis did not reliably predict the 1p/19q codeletion status for WHO grade II and III patients (12). Fella et al. attempted to predict 1p/19q codeletion status for WHO grade II and III oligodendroglioma and oligoastrocytoma patients by analyzing magnetic resonance (MR) imaging, including T1-weighted (T1), T2-weighted (T2), T2*-weighted (T2*), fluid-attenuated inversion recovery (FLAIR), and T1-weighted post-contrast (T1-post), diffusion-weighted imaging (DWI), perfusion-weighted imaging (PWI), and MR spectroscopy. In their analysis, they reported that the DWI, PWD, and MR spectroscopy showed little improvement in determining 1p/19q codeletion status compared to the more conventional MR sequences, namely T1, T1post, T2, T2*, and FLAIR sequences alone (13). In addition, Park et al. reported that whole tumor histogram and texture analyses of diffusion tensor imaging (DTI) can predict the 1p/19q codeletion status in grade II gliomas (14). Lastly, in 2017, Akkus et al. demonstrated that T1-post and T2 images can be analyzed to determine the 1p/19q codeletion status using machine intelligence (15).

In light of these previous radiogenomics studies, this thesis investigates the feasibility of assessing 1p/19q codeletion status by analyzing multimodal MR sequence scans.

1.4 Central Hypothesis

There exists an underlying image level signature on multi-parametric MR scans (T1, T1post, T2, and FLAIR modalities) that distinguishes 1p/19q co-deleted from non-codeleted chromosomes.

Specific Aim 1: Develop a pipeline that quantitatively extracts MR scans image features

To achieve this aim, we extract three set of distinctive image features, namely texture-based, topology-based, and convolutional neural network based, and evaluate their performances in predicting 1p/19q codeletion status.

Specific Aim 2: Assess the feasibility of conventional MR sequences in prediction task of 1p/19q codeletion status

Hypothesis: Image feature analysis using conventional and universally performed MR sequences is not inferior to that using advanced MR sequences in predicting 1p/19q status.

To achieve this aim, we compare our analysis results with those of reports in the literature.

1.6 Organization of this thesis

The overall goal of this thesis is to compare and contrast three different image features from multimodal MRI scans in the context of determining a patient's 1p/19q codeletion status. First, Chapter 2 discusses the data acquisition and pre-processing steps. Then, Chapters 3, 4, and 5 address texture, topological (and specifically, persistent homology), and CNN feature acquisition procedures respectively. Then, Chapter 6 discusses the feature selection, feature reduction, and training procedures using those three set of features. Chapter 7 reports the results from Chapter 6. Finally, Chapter 8 concludes this thesis by comparing and contrasting the three different image analysis results.

Chapter 2: Data acquisition and pre-processing

2.1 BRATS 2017 dataset

The multimodal Brain Tumor Image Segmentation Benchmark (BRATS) 2017 dataset was originally designed for the brain tumor segmentation challenge (16, 17) and comprises pathologically confirmed LGG (n = 65) and HGG (n = 102) cases from The Cancer Imaging Archive (TCIA) (18). The dataset contains pre-operative multi-modal MRI sequences, namely T1, T1post, T2, and FLAIR, and was acquired with differing imaging/clinical protocols and scanners from 19 different institutions (Figure 3). These four sequences were co-registered rigidly to the T1post sequence as it had the highest spatial resolution, and were then resampled to 1mm x 1mm x 1mm isotropically in an axial orientation by using a linear interpolation algorithm. Then, all images were skull-stripped to anonymize the patient information. All tumor volumes in the imaging dataset have been segmented manually by one to four different experienced neuro-radiologists: The pixels in necrotic and non-enhancing tumor (NCR/NET) were labeled 1, those in peritumoral edema (ED) were labeled 2, and those in gadolinium-enhancing tumors (Gd-ET) were labeled 4 (16, 17) so that each type is identified by a unique bit when these labels are viewed as binary numbers.

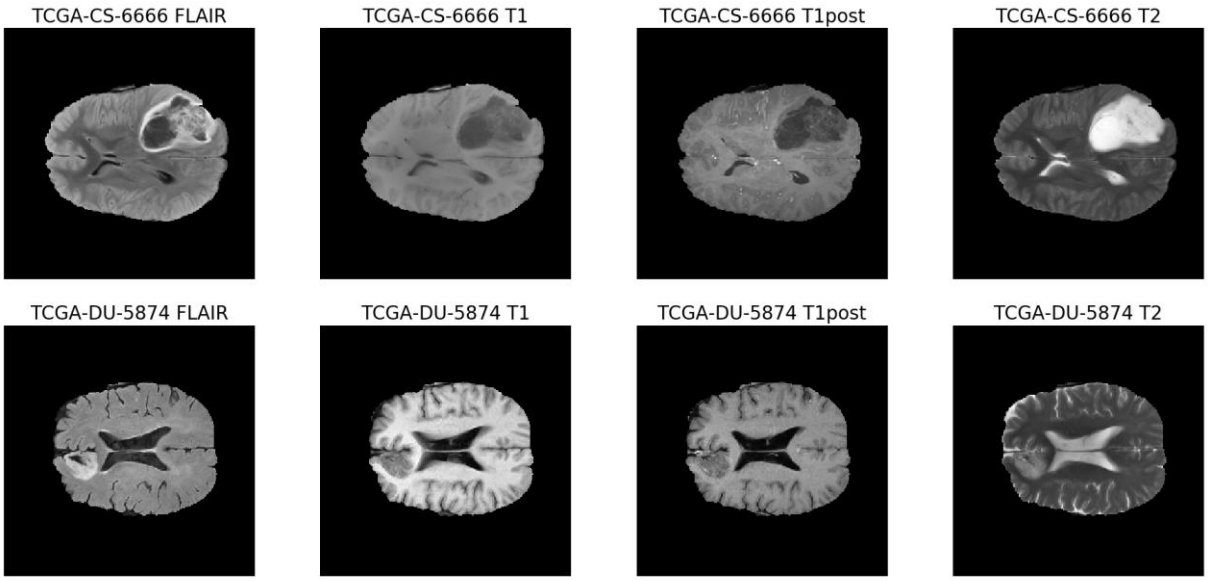


Figure 3 Some sample images from BRATS data. TCGA-CS-6666 is a non-codeleted case while TCGA-DU-5874 is codeleted.

2.2 1p/19q status and patient information from TCGA

Since the BRATS 2017 dataset comes from TCIA, one can also find the corresponding molecular/genetic dataset from The Cancer Genome Atlas (TCGA): https://cancergenome.nih.gov/. Under TCGA/TCIA data use agreements, analysis of this cohort was exempt from IRB approval. We first retrospectively identified the patients with histologically confirmed WHO grade II-IV gliomas (n=1122) and their corresponding 1p/19q chromosome codeletion statuses (after surgical biopsy). In addition, the patients' age, gender, Karnofskyi Performance Score (KPS) were collected as clinical variables.

Considering both the imaging dataset from BRATS 2017 and the 1p/19q statuses from TCGA, only 143 patients had both. Figure 4 illustrates the data acquisition process in a diagram.

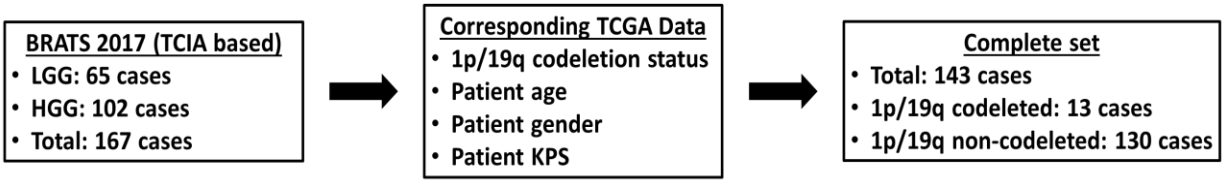


Figure 4 Flow diagram of data collection. We first obtained the BRATS 2017 dataset that contains T1, T1post, T2, and FLAIR sequence images with tumor region mask. Then, we retrospectively

collected patient 1p/19q codeletion status, age, gender, and KPS from the TCGA dataset. Considering those two, we obtained a cohort of 143 patients.

2.3 Patient Characteristics

Table 1 summarizes the patient demographics and 1p/19q status. The median age, KPS, and male population proportion of the patient cohort was 54 years, 80, and 48.25% respectively. For WHO glioma grades, 44.76% (n=64) of the cohort was LGG cases while the remainder was HGG (n=79). For 1p/19q codeletion status, only 13 patients were codeleted and all came from the LGG cases while the remaining 130 patients were non-codeleted and came from both LGG (n=51) and HGG (n=79). There was thus a very large imbalance in the ratio (1:10) between the codeleted and non-codeleted patients.

Age	54 (18-84)
KPS	80 (50-100)
Sex (% male)	69 (48.3 %)
1p/19q codeletion rate	13 (9.1 %)
Grade and 1p/19q codeletion status	
LGG (II and III)	51 (44.8 %)
II non-codeleted	21 (14.7 %)
II codeleted	6 (4.2 %)
III non-codeleted	30 (21.0 %)
III codeleted	7 (4.9 %)
HGG (IV)	79 (55.2 %)
IV non-codeleted	0 (0 %)
IV codeleted	79 (55.2 %)
Note: Age and KPS shown as median (min-max)	

Table 1 Patient demographics, KPS, 1p/19q codeletion status, and glioma grade.

2.4 Train/test dataset split

When sampling the patient cohorts into either the train or the test datasets, the ratio between codeleted and non-codeleted must be kept the same. In order to do so, we divided the patient cohort randomly into 13 groups such that each group contained 1 codeleted and 10 non-codeleted cases. Among these 13 groups, 2 random groups were assigned to be the testing dataset (2 codeleted/ 20 non-codeleted) while the remaining 11 groups were assigned to be the

training dataset (11 codeleted/ 110 non-codeleted). This process was repeated 10 times such that there were 10 independent train/test splits of the data. Figure 5 illustrates this process in a diagram.

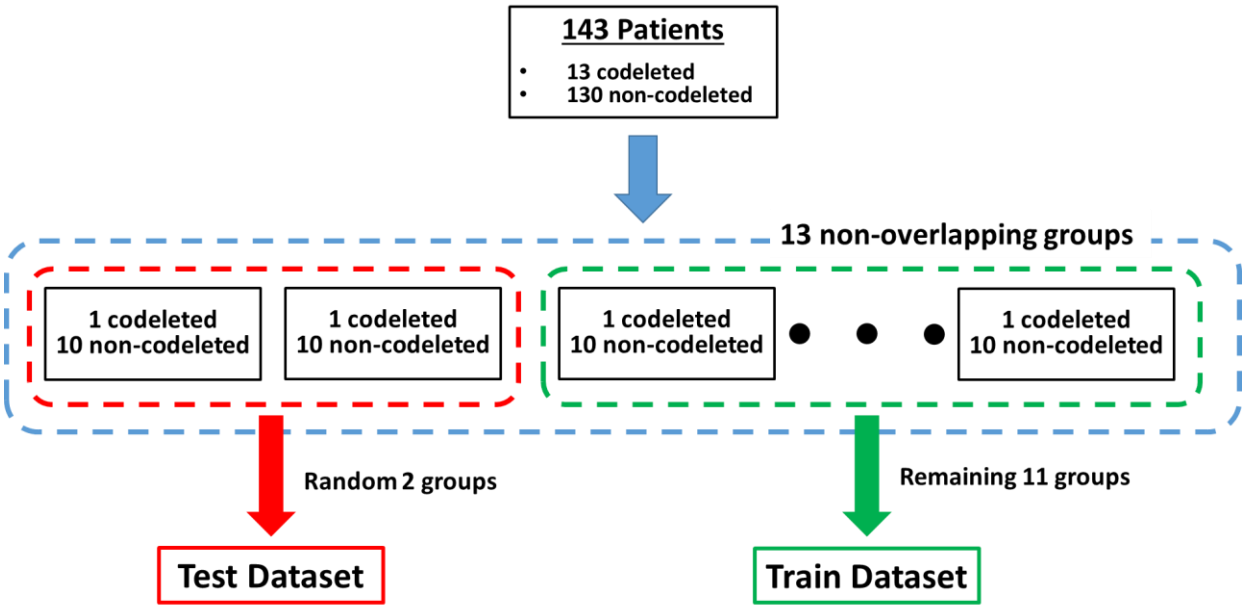


Figure 5 A diagram depicting train/test dataset split. First, we performed stratified random sampling such that each group contains 1 codeleted and 10 non-codeleted cases. Then, we randomly assigned 2 groups to test dataset while the remaining belongs to the training dataset. We repeated this process 10 times to obtain 10 train/test splits.

From now on, unless otherwise specified, all of the figures and tables are from train/test dataset split number 0 for simplicity.

2.5 Image pre-processing pipeline

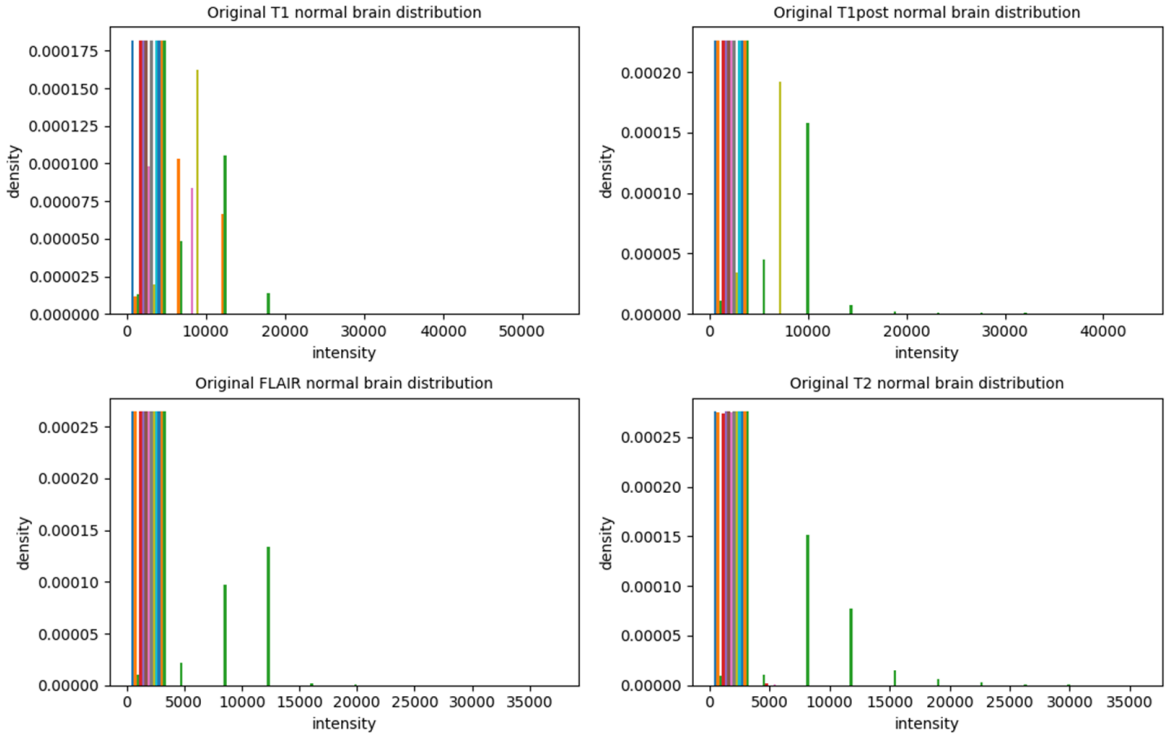
2.5.1 N4 bias correction

One of the main assumptions of MRI scanning is a uniform magnetic field in the center of a scanner, which in reality is often not achieved. Consequently, some MRI scans can present low frequency intensity inhomogeneity, which can be detected as shading artifacts (19–21). In

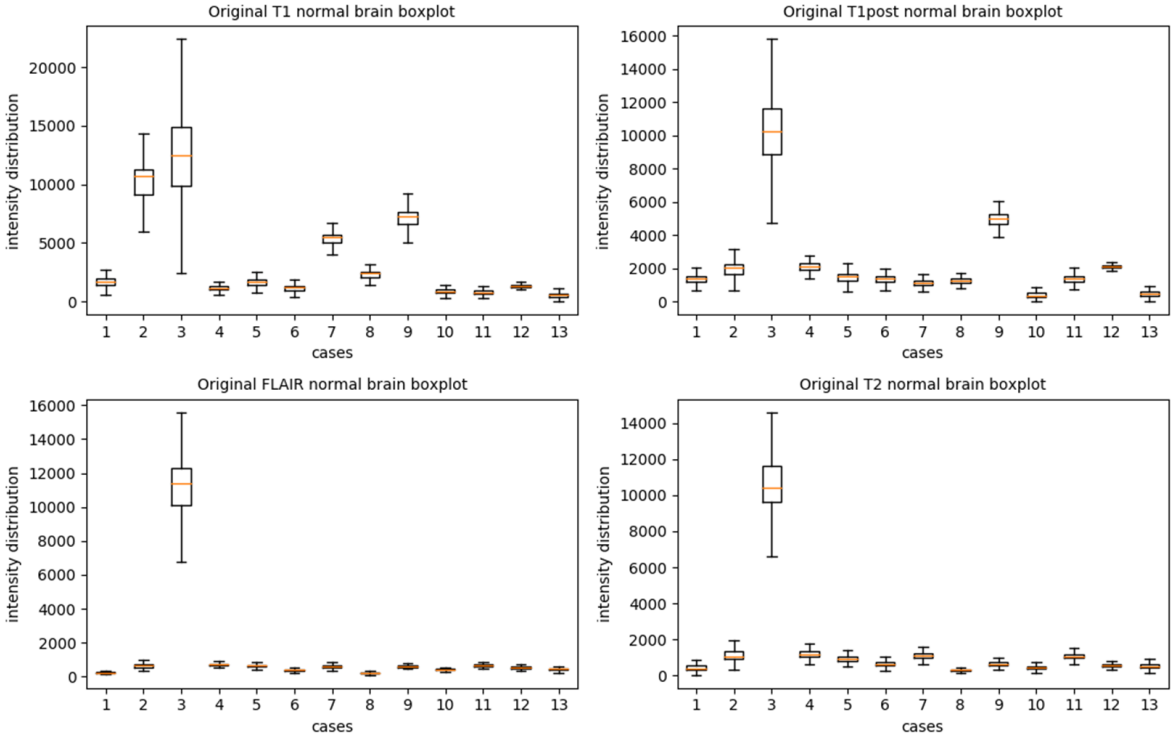
order to correct this phenomenon, N4 bias correction was applied to remove any low frequency intensity inhomogeneity (22), (23).

2.5.2 Intensity Normalization

Another problem with MRI scans is the arbitrary scaling of the voxel intensity. Computerized tomography (CT) scans have voxel intensity values that are a linear mapping of Hounsfield Unit (HU), which is derived from the linear attenuation coefficients of water and of air. In other words, CT voxel values can be mapped back to some “physical meaning” that is based upon the attenuation of the material in the voxel. However, MRI scans have no clear, consistent mapping from the voxel intensity values back to a simple “physical meaning”. Consequently, MRI scans are expressed in arbitrary units that are difficult to understand in an absolute sense and that may well differ among scanners, protocols, and institutions. Figure 6. showcases the intensity distributions that may be encountered among patients. For this illustration, 13 patients were chosen at random. In both the normal regions (Figures 6(a) and (b)) and normal tumor regions (Figures 6(c) and (d)), there exists some variability among the patients in the intensity distributions.


Original T1 normal brain distribution
density
intensity
Original T1post normal brain distribution
density
intensity
Original FLAIR normal brain distribution
density
intensity
Original T2 normal brain distribution
density
intensity

(a)


Original T1 normal brain boxplot
intensity distribution
cases
Original T1post normal brain boxplot
intensity distribution
cases
Original FLAIR normal brain boxplot
intensity distribution
cases
Original T2 normal brain boxplot
intensity distribution
cases

(b)

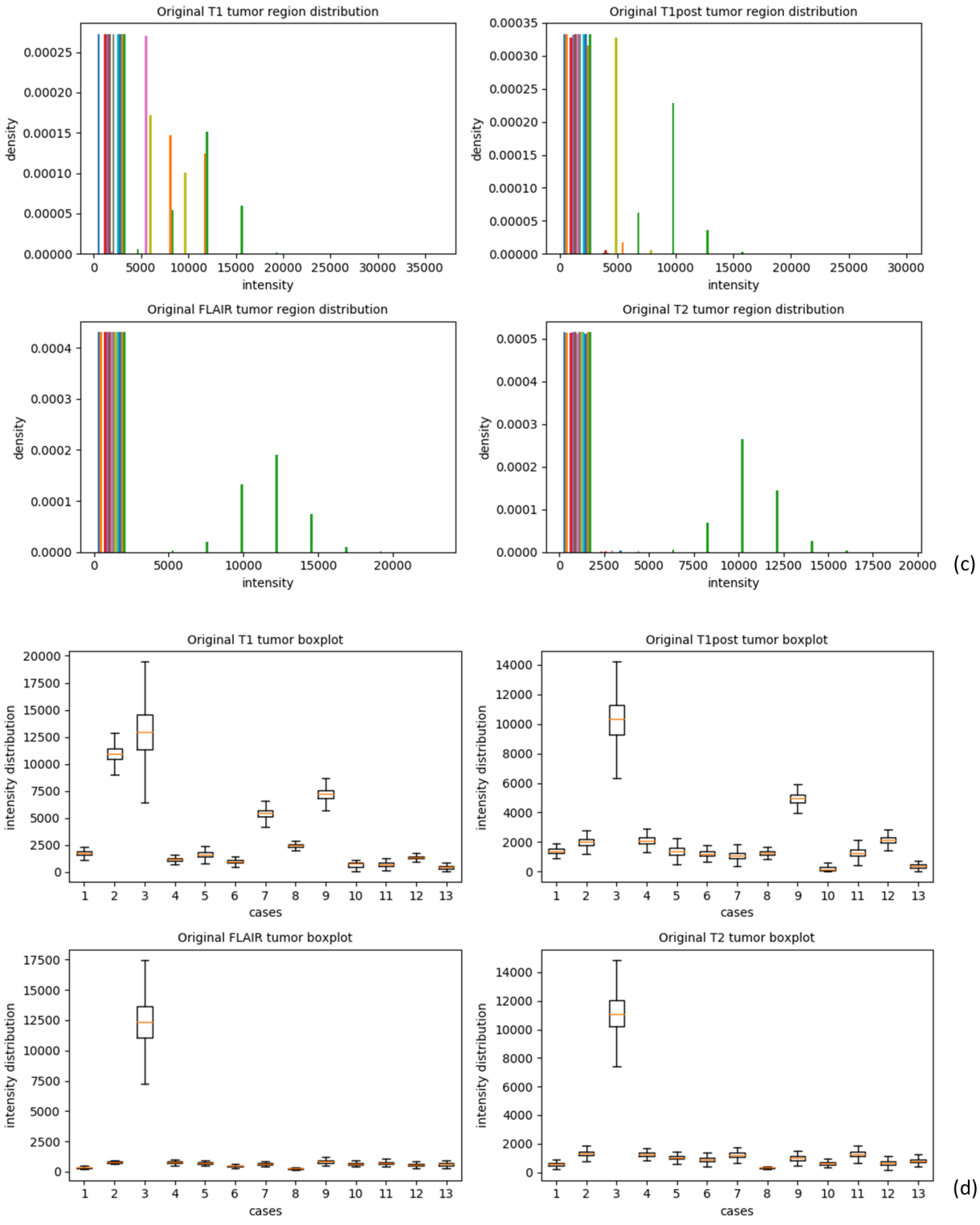
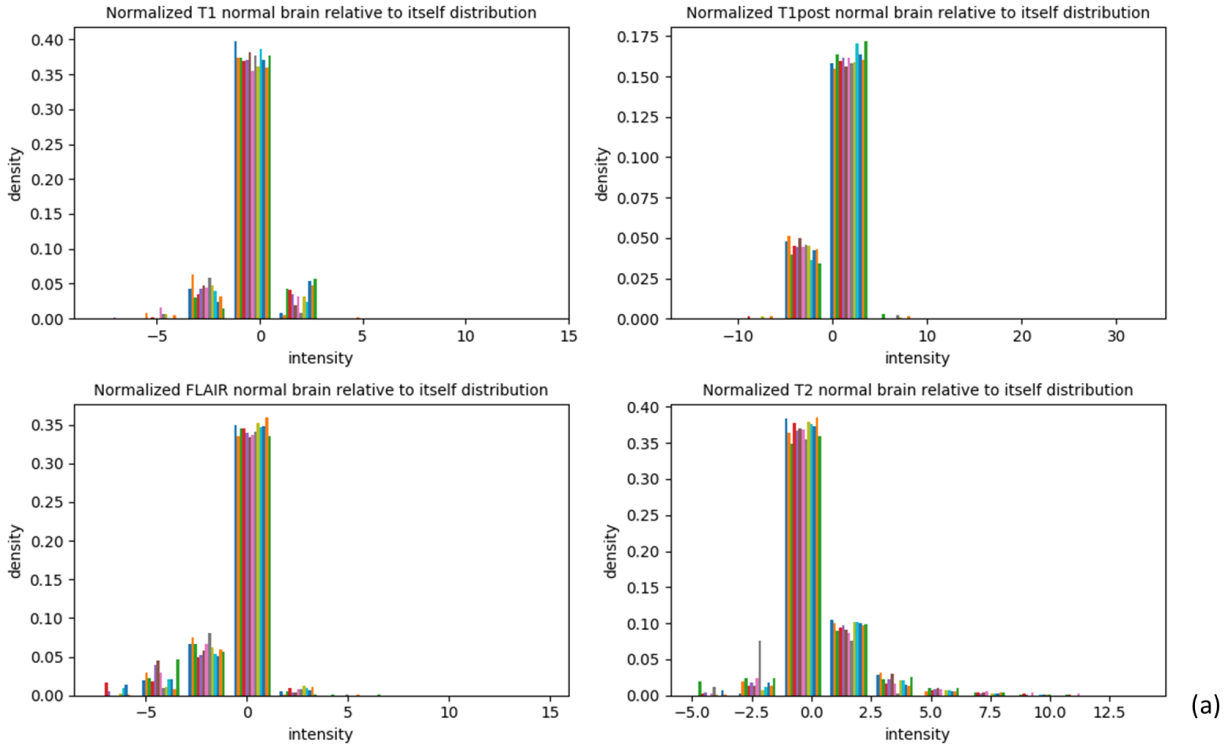
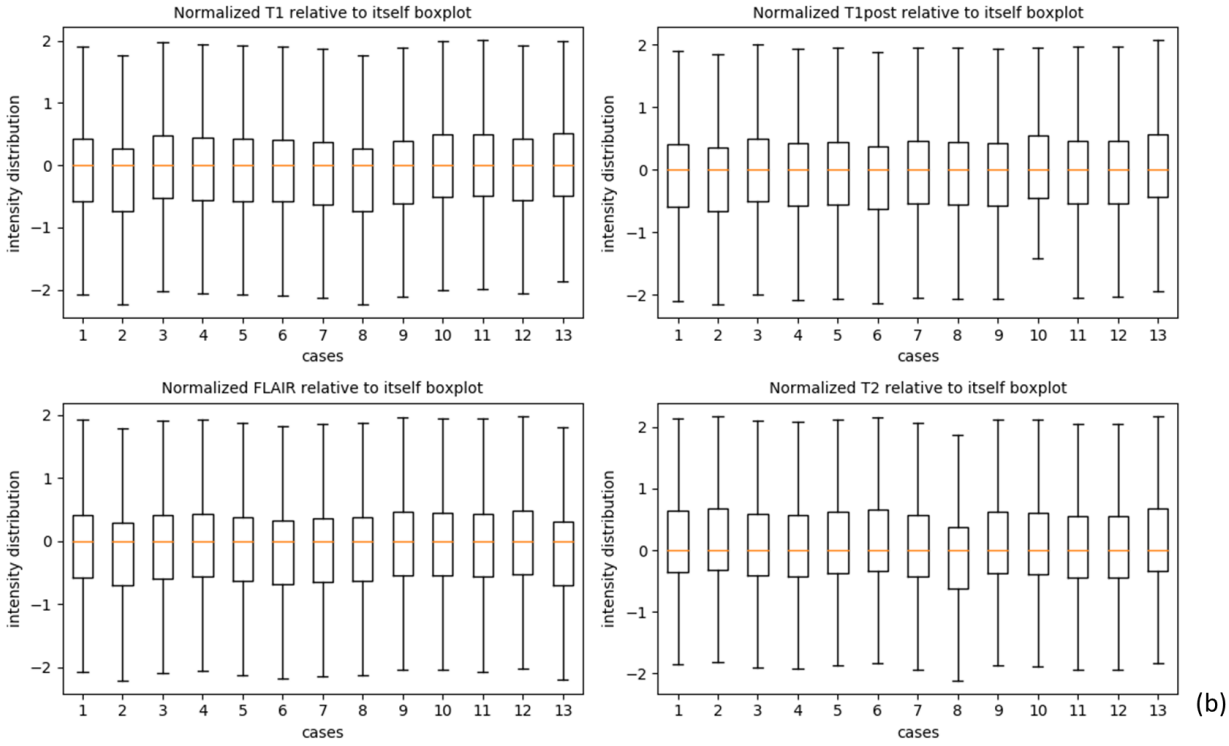


Figure 6 (a) shows the intensity distribution histogram while (b) shows the intensity distribution boxplot across 4 MRI sequences of normal brain region for 13 randomly chosen patients. There are some clear discrepancies in the intensity distribution ranges across patients. The same can be observed for the tumor region histogram (c) and the tumor region boxplot (d).

There have been several methods proposed (24–26) to deal with such a problem, but the simplest approach is normalizing relative to “normal brain”. For such an approach, two assumptions are made. First, there exists a linear mapping that maps the “true” or “physical meaning” distribution to what is seen in the MRI scan intensity distribution (i.e. the arbitrary unit distribution). Second, in this “true” or “physical meaning” distribution, the normal brain distribution should be consistent or at least similar across patients and scanners. This is a fair assumption since it is a normal brain, and hence the patient-to-patient distribution variability should be minimal in this “true” or “physical meaning” unit. Also, just as the normal brain intensity distributions across patients are similar, so should be the interquartile intensity values. Now, we realize that the interquartile intensity of normal brain scales according to which space it lies within, that is, whether it is in the “true”/ “physical meaning” distribution or the transformation of that (i.e. arbitrary unit distribution), since the transformation is a linear mapping. Therefore, we normalize image intensities in the following fashion: first, we subtract the median intensity of normal brain from the image, which would zero center the normal brain intensity distribution, and then divide by the interquartile intensity of normal brain. By so doing, in this “normalized space”, both the “physical meaning” distribution and the “arbitrary unit” distribution should map to the same distribution. Figures 7(a) and (b) show the normalized normal brain distribution and the corresponding boxplot. As expected, it is zero-centered with a nice normal distribution. Also, the boxplot of the normalized normal brain shows that the distributions in the individual patients overlap with each other in all four sequences. This indicates that our two major assumptions, a linear transformation and consistent normal brain distribution across patients, hold true. Figures 7(c) and (d) show the normalized tumor distribution and the corresponding boxplot. A big difference between normal brain and tumor is the intensity distribution consistency. It is well known that gliomas exhibit heterogeneity [32-35], and if we consider our first assumption, the heterogeneity would still exist in the “true” intensity space and so also in the “normalized space”, which one can observe in Figures 7(c) and (d).


Normalized T1 normal brain relative to itself distribution
Normalized T1post normal brain relative to itself distribution
Normalized FLAIR normal brain relative to itself distribution
Normalized T2 normal brain relative to itself distribution
density
intensity

(a)


Normalized T1 relative to itself boxplot
Normalized T1post relative to itself boxplot
Normalized FLAIR relative to itself boxplot
Normalized T2 relative to itself boxplot
intensity distribution
cases

(b)

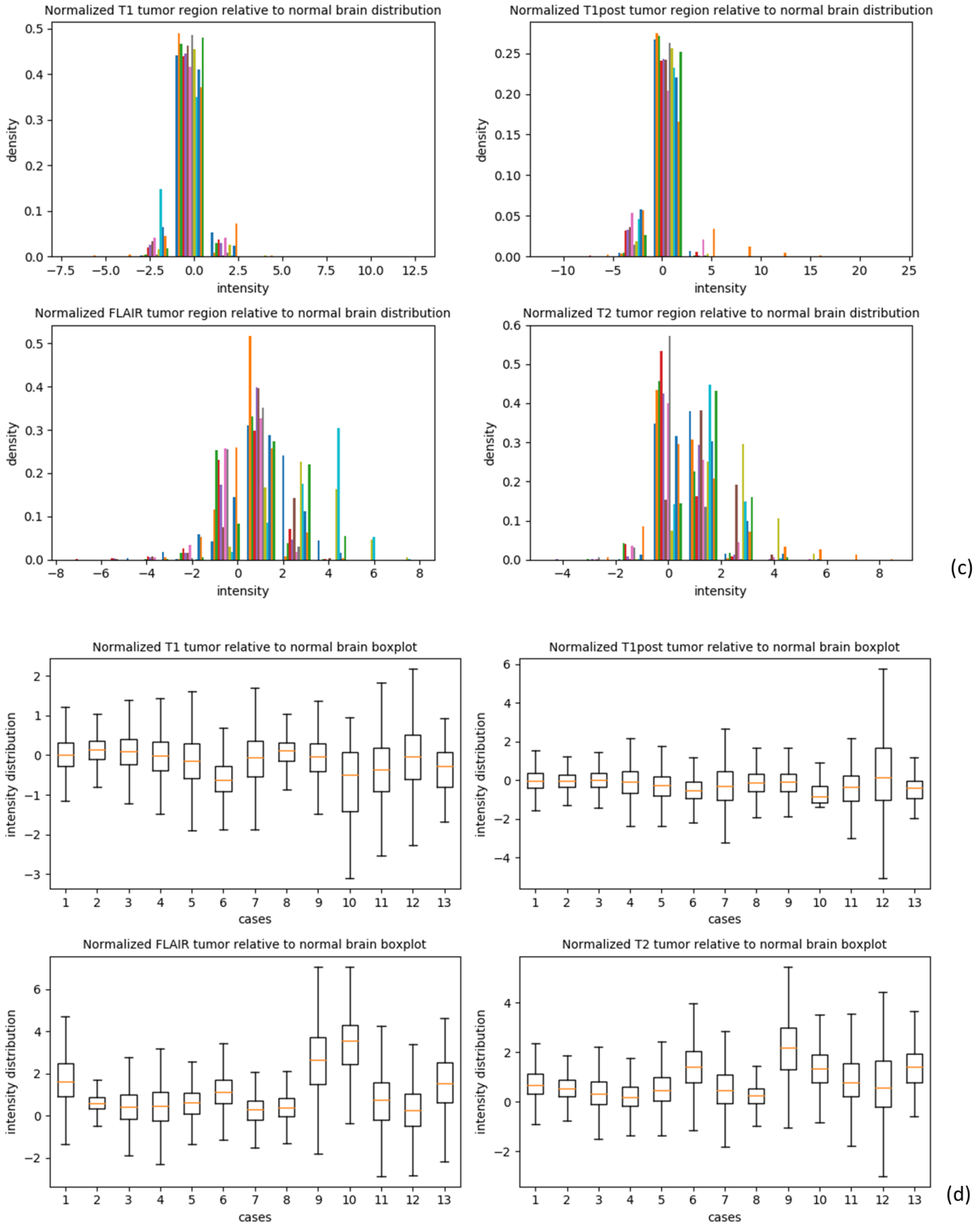


Figure 7 (a) shows the normalized intensity distribution histogram while (b) shows the normalized intensity distribution boxplot across four MRI sequences of normal brain region for 13 randomly chosen patients. In the normalized space, the histogram is now in a nice bell shape and zero-centered. Also, the boxplot median and interquartile ranges line up well with each other across patients. On the other hand, for normalized tumor regions, there are still some discrepancies

across patients in intensity distribution ranges across patients for the histograms (c) and the boxplots (d). This reflects the patients' individual tumor region heterogeneity.

2.5.3 Intensity Discretization – Texture analysis only

Once the intensity distribution is normalized relative to the normal brain region, intensity discretization is implemented in the following fashion:

$$X_{b,i} = \left\lfloor \frac{X_{gl,i}}{W} \right\rfloor - \left\lfloor \frac{min(X_{gl})}{W} \right\rfloor + 1$$

- $X_{gl,i}$ = gray level before discretization
- $X_{b,i}$ = gray level after discretization
- W = bin width specified by user

Notice that the minimum gray level is 1 such that the minimum intensity value may be distinguished from the background, which is coded as 0. Also, it is important that the user specify the bin width because this bin width is the quantum of measurement in the gray level quantity. By supplying the user-defined bin width, the definition of a gray level unit of 1 is consistent, and hence the definition of "contrast" is consistent across patients. This process is crucial, especially when it comes to creating a matrix representation of the gray levels in the ROI, such as the gray level co-occurrence matrix (GLCM), or the gray level run length matrix (GLRLM), because such matrices assume a well-defined and consistent meaning of the gray level values.

Our reasoning is perhaps best illustrated by considering the following example: assume that patient A has an intensity distribution that ranges from -5 to 5 whereas patient B has the distribution ranging from -10 to 10 after the normalization step. If we specify the bin width to be 0.5, then patient A will uniquely yield 20 different gray levels whereas patient B will yield 40. They might have different gray level ranges, but that simply reflects the fact that patient B has the wider range of intensity distributions compare to patient A from the beginning. On the other hand, the definition of 1 gray level in patient A is same as that of in patient B, i.e. bin width of 0.5, hence two patients have the same physical meaning in 1 gray level unit (27) .

In our study, we chose the bin width of 0.0625 such that the resulting amount of bins fell between 30 and 130 bins following the recommendation from PyRadiomics FAQs

2.5.4 Intensity Linearization – TDA and CNN analysis

Unlike texture analysis where 1 gray level unit is important, TDA and CNN analysis are less concerned with the definition of a gray level unit. This is the case because of their nature of feature extraction. Unlike texture-based features where gray level matrix representation must have the same physical meaning in gray level across patients, both topological features and CNN features are more concerned with relative rankings in voxel intensities. In other words, as long as the rankings of voxel intensities are preserved in a consistent manner, then we can enforce intensity distributions into the same window:

$$X_{b,i} = \left\lfloor \frac{X_{gl,i} - min(X_{gl})}{max(X_{gl}) - min(X_{gl})} \right\rfloor + 1$$

- $X_{gl,i}$ = gray level before discretization
- $X_{b,i}$ = gray level after discretization

This is a simple min-max scaling scheme, but adding 1 on top of it so that every voxel intensity ranges from 1 to 2. Again, it starts from 1 to distinguish the minimum intensity value from background, which has the value 0.

2.5.5 Tumor Patch Extraction

In order to extract a tumor region slice, the smallest bounding rectangle was drawn around the tumor to capture the tumor region, and then the image within that rectangle was resized to 142 by 142 pixels. Note that for texture and topological features, we extracted only the patches in the axial plane. On the other hand, for the CNN features, we extracted in all three

dimensions, namely in the coronal, sagittal, and axial planes, so that we could meet the input dimensions of the CNN architecture (Chapter 5 explains this in more detail).

As we already found out, there exists a very high imbalance in the dataset (which has only 13 codeleted vs 130 non-codeleted). To combat this imbalance as well as the small amount of available data, we extracted 20 slices from the codeleted cases but only 2 slices from the non-codeleted cases when training. For codeleted patch extraction, the 20 largest tumor areas were selected. On the other hand, for non-codeleted, 100^{th} and 75^{th} percentile tumor areas were used. In total, we obtained 220/220 patches between codeleted and non-codeleted for the training dataset but 4/40 patches between codeleted and non-codeleted for the test dataset.

The process of this image pre-processing pipeline is depicted in Figure 8 as a diagram.

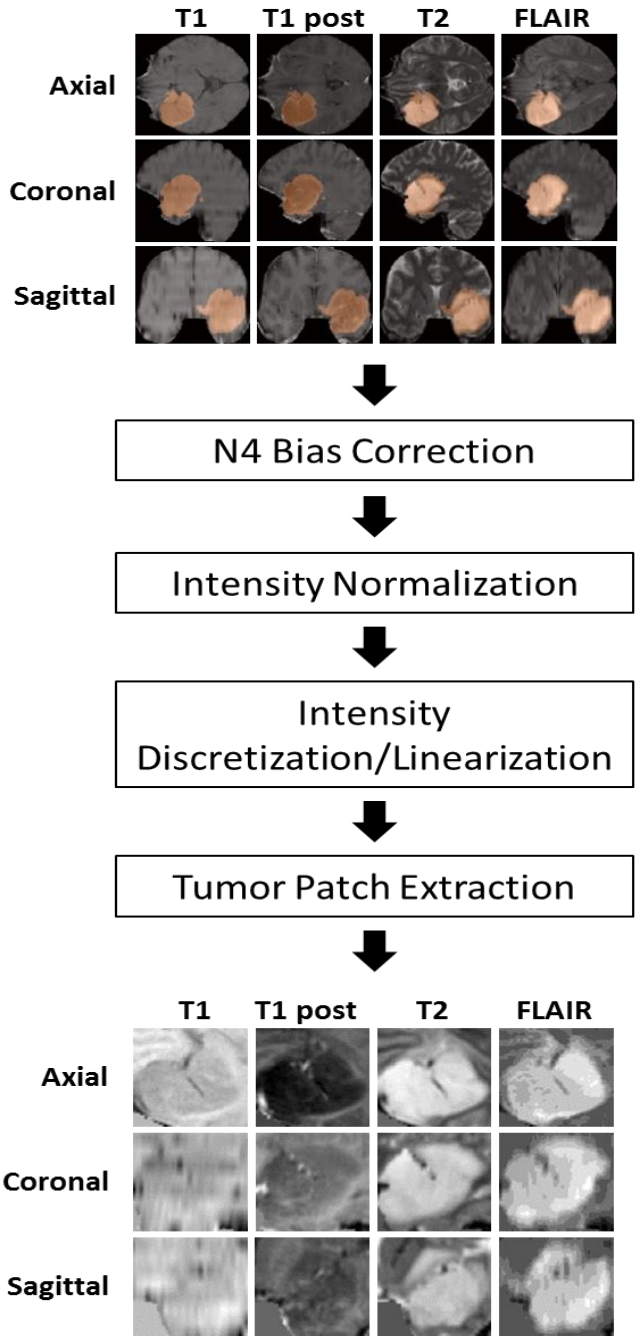


Figure 8 A flow chart of image pre-processing. From 4 MRI sequences, first the N4 bias correction was applied to remove low-level noises. Then, intensity normalization was applied to put the patients' intensity ranges into the similar range. Then, for texture features, intensity discretization was applied while for topological and CNN features, intensity linearization was implemented. Then, the smallest bounding rectangle was drawn around the tumor to capture the tumor region, and the image within that rectangle was resized to 142 by 142 pixels. For the codeleted cases, the top 20 largest tumor areas were extracted, while for the non-codeleted cases, 100^{th} and 75^{th} percentile tumor areas were extracted for training dataset. In total we obtained 240/240, or 480 tumor patches for training dataset. For test, we extracted 100^{th} and 75^{th} percentile tumor areas for both codeleted (4) and non-codeleted (40), resulting in 44 tumor patches.

Chapter 3: Texture features

3.1 Introduction

The main goal of image analysis, whether it be of a medical image or a natural scene, is to “quantitatively” assess the image information based upon some metrics. For instance, in human words, one might describe an image as “smooth”, “rough”, or “honeycomb shaped”, but such descriptions lack consistency in interpretation and cannot produce numerical features that can be statistically analyzed by machines. Texture analysis attempts to solve such challenges by quantifying those qualitative descriptions as a function of spatial variation in pixel or voxel intensities, and has been shown to be successful in radiogenomics studies. For instance, texture features have shown promise for characterizing cancer genetics regardless of medical image modalities such as MRI (27), PET (28), and CT (29, 30). Also, texture analysis has been reported to be effective in characterizing a variety of cancer types including breast tumors (31), lung tumors (32, 33), and head and neck tumors (32). Considering these past reports on their successes on characterizing numerous cancer types with differing image modalities, it is not unreasonable to pursue such methods to predict the 1p/19q codeletion status.

3.3 Texture Feature extraction

Here, we defined six different subtypes of texture features which produced 91 features for a single modality. Since there are four modality sequences, we obtained a total of 364 features.

3.3.1 Texture Feature Calculation Software

The PyRadiomics software package version 1.3.0 was used for the calculation of all of the features in this chapter of study (27). This software is an open source Python based package that is freely available online at https://pyradiomics.readthedocs.io/en/latest/.

Below, we defined six unique texture feature groups. For the exact definition of the calculation of each feature, please refer to the PyRadiomics website.

3.3.2 First order statistics based features

First order statistics describes the voxel intensity distributions within the image region defined by the mask. Under this category, the following statistics were computed:

1. Energy
2. Total Energy
3. Entropy
4. Minimum
5. 10th percentile
6. 90th percentile
7. Maximum
8. Mean
9. Median
10. Interquartile Range
11. Range
12. Mean Absolute Deviation (MAD)
13. Robust Mean Absolute Deviation (rMAD)
14. Root Mean Squared (RMS)
15. Standard Deviation
16. Skewness
17. Kurtosis
18. Variance
19. Uniformity

Most of these features were calculated directly from the original intensity distributions (i.e.no intensity discretization took in place) except for entropy and uniformity. For these, we used the bin width of 0.0625 (as specified in section 2.5.4).

3.3.3 Shape based features

Shape-based features describe the size and shape of the ROI. The following features were computed:

1. Volume
2. Surface Area
3. Surface Area-to-Volume ratio
4. Sphericity
5. Compactness 1
6. Compactness 2
7. Spherical Disproportion
8. Maximum 3D diameter
9. Maximum 2D diameter (Slice)
10. Maximum 2D diameter (Column)
11. Maximum 2D diameter (Row)
12. Major Axis
13. Minor Axis
14. Elongation

3.3.4 Gray Level Co-occurrence Matrix (GLCM) based features

The co-occurrence matrix (GLCM) is the matrix that counts the frequency of a specified gray level intensity of a pixel of interest in relation to its neighboring pixel/voxel in a specified

direction (34). Here, the intensity discretization method was applied (as it was as well for all of the other matrices defined below) to ensure that 1 gray level unit has the same physical meaning. The (*i,j*)th element of this matrix represents the number of times that the combination of gray levels *i* and *j* occur in two neighboring pixels in the image for a specified direction. For the neighbors, we considered all 8 directions (0º, 45º, 90º …, and 315 º). From this matrix, the following features were computed:

1. Autocorrelation
2. Joint Average
3. Cluster Prominence
4. Cluster Shade
5. Cluster Tendency
6. Contrast
7. Correlation
8. Difference Average
9. Difference Entropy
10. Difference Variance
11. Joint Energy
12. Joint Entropy
13. Informal Measure of Correlation (IMC) 1
14. Informal Measure of Correlation (IMC) 2
15. Inverse Difference Moment (IDM)
16. Inverse Difference Moment Normalized (IDMN)
17. Inverse Difference (ID)
18. Inverse Difference Normalized (IDN)
19. Inverse Variance
20. Maximum Probability
21. Sum Average

22. Sum Entropy

23. Sum of Squares

Once all of the features for all eight directions were calculated, then the distance-weighted average was calculated. For neighbors in 0º, 90º, and 270º, the weight was 1 whereas for the diagonal angles (i.e. 45º, 135º, 225º, and 315º) the weight was $\sqrt{2}$ to account for the distance difference.

3.3.5 Gray Level Run Length Matrix (GLRLM) based features

The gray level run length matrix (GLRLM) quantifies gray level runs, which are defined as the length in the number of pixels that have the same gray level for a given direction. The element (*i,j*) in GLRLM describes the number of runs for a gray level *i* with length *j* that occur in the image in the specified direction. The following features were computed under this category (35–38):

1. Short Run Emphasis (SRE)
2. Long Run Emphasis (LRE)
3. Gray Level Non-Uniformity (GLN)
4. Gray Level Non-Uniformity Normalized (GLNN)
5. Run Length Non-Uniformity (RLN)
6. Run Length Non-Uniformity Normalized (RLNN)
7. Run Percentage (RP)
8. Gray Level Variance (GLV)
9. Run Variance (RV)
10. Run Entropy (RE)
11. Low Gray Level Run Emphasis (LGLRE)
12. High Gray Level Run Emphasis (HGLRE)

13. Short Run Low Gray Level Emphasis (SRLGLE)
14. Short Run High Gray Level Emphasis (SRHGLE)
15. Long Run Low Gray Level Emphasis (LRLGLE)

3.3.6 Gray Level Size Zone Matrix (GLSZM) based features

The gray level size zone matrix (GLSZM) quantifies gray level zones in an image. A gray level zone is defined as the number of connected voxels for a given gray level. In a GLSZM, the (*i,j*)th element represents the number of zones with gray level *i* and size *j* appear in image. The advantage of GLSZM over GLCM and GLRLM is that it is direction-independent by considering all of the neighbors when connecting voxels (34). From the GLSZM, we computed the following features:

1. Small Area Emphasis (SAE)
2. Large Area Emphasis (LAE)
3. Gray Level Non-Uniformity (GLN)
4. Gray Level Non-Uniformity Normalized (GLNN)
5. Size-Zone Non-Uniformity (SZN)
6. Size-Zone Non-Uniformity Normalized (SZNN)
7. Zone Percentage (ZP)
8. Gray Level Variance (GLV)
9. Zone Variance (ZV)
10. Zone Entropy (ZE)
11. Low Gray Level Zone Emphasis (LGLZE)
12. High Gray Level Zone Emphasis (HGLZE)
13. Small Area Low Gray Level Emphasis (SALGLE)
14. Small Area High Gray Level Emphasis (SAHGLE)
15. Large Area Low Gray Level Emphasis (LALGLE)

16. Large Area High Gray Level Emphasis (LAHGLE)

3.3.7 Neighboring Gray Tone Difference Matrix (NGTDM) based features

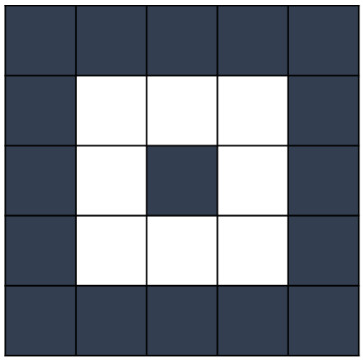
A Neighboring Gray Tone Difference Matrix (NGTDM) is a not really a matrix but rather is a vector that quantifies the difference between a gray value and the average gray value of its neighbors, which we defined as the eight adjacent pixels. Hence, an element *i* in the NGTDM represents the sum of absolute differences. Again, this matrix is direction-independent and can be used to compute the following terms (39):

1. Coarseness
2. Contrast
3. Busyness
4. Complexity
5. Strength

Chapter 4: Topological features: Persistent homology

4.1 Introduction

While many of the texture features were designed to quantify spatial variations in pixel or voxel intensities by assessing the relationship between the pixel/voxel of interest and its neighboring pixels/voxels, such features *can capture only local intensity variations and are directionally dependent*. For example, in the construction of a GLCM, every element (i,j) in the matrix represents only how many times a neighboring pixel has a certain gray level value in relation to the pixel of interest in a specified direction. Or, one might consider with GLZSM-based features where the matrix elements are direction-independent, yet it still accounts only for how many times a certain gray level occurs. For instance, consider the following binary figure (Figure 9).



GLZSM

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 9 An exemplary binary image. One can see that there is a point at the center and a rectangle surrounding it. However, GLZSM fails to detect such topological features but only returns the value of how many pixels are connected with the same gray level.

It is quite easy for us humans to tell that there is a point at the center and a rectangle surrounding it. If we construct a GLZSM, we get a very simple matrix showing size zones of 1 and 16 at gray level 1. However, it does not tell us any information about the geometrical shape, other than how many pixels are connected with the same gray level. In other words, while texture features are good for capturing pixel intensity distributions for a given gray level and a direction, it might fail to capture a geometrical or *topological* representation in the image. Hence, to better

capture the topological features, this chapter investigates topological data analysis in medical images.

4.2 Theory

A review of the deep theoretical frame work of topology and persistent homology is beyond the scope of this study. Therefore, our focus will be on a graphical explanation of the application of computational topology to medical images following (40, 41).

4.2.1. Definition of topology

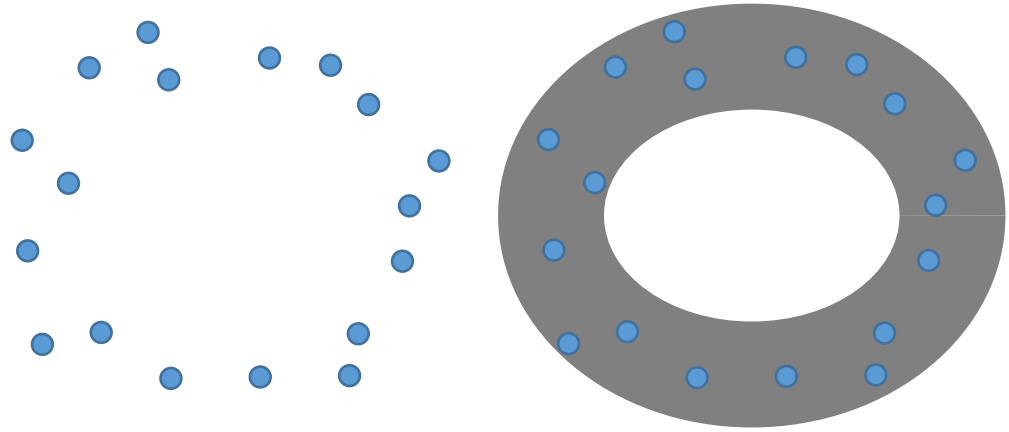


Figure 10 An example of point cloud data. One can easily tell that those points form a ring.

Topology is the field of mathematics that is concerned with the properties of geometric spaces that remain invariant under certain transformations, such as bending or stretching. Persistent homology is an algebraic method for detecting these topological features of data, such as components, clusters, holes, graph structures, etc. For example, consider the example of a point cloud (Figure 10). It is quite easy to detect that the point cloud forms a ring. However, in a machine's perspective, just purely from discrete points alone, it is hard to detect the ring, as the discrete points have a trivial topology, namely dots.

4.2.2 Detecting topology

One way to detect some meaningful topological features is to connect neighboring data points by drawing circles around the points with a specific diameter, “d” (Figure 11).

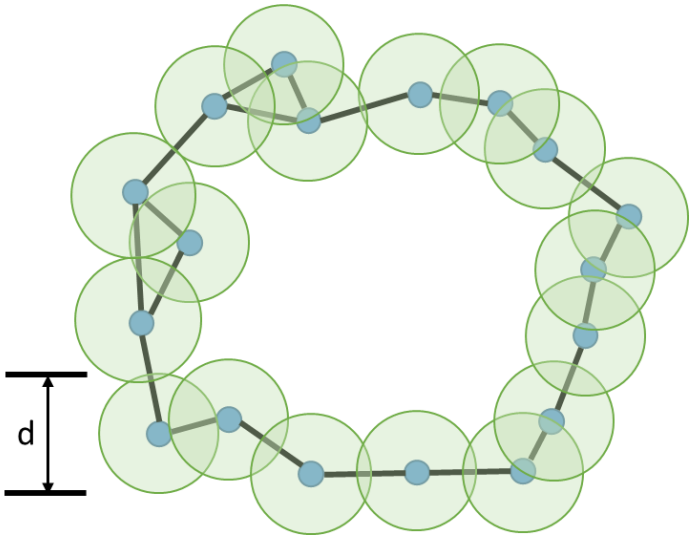


Figure 11 An example of detecting point cloud data. By drawing circles around the points with a specified diameter, “d”, we can connect dots and form a single cluster. However, it still fails to inform that there exists a tunnel within the cluster.

Now the connected dots show that these points form a single “cluster” at diameter d, but it still does not tell us that there exists a tunnel (or hole) at the center. To detect the tunnel, or some “higher-order” features, first, we need to define the following terms.

4.2.3 Simplicial complex and homology

A *simplicial complex* is an object built from combinations of points, edges, triangular faces, etc. For example, a simple point is a 0-dimensional simplex, an edge between two points is a 1-dimensional simplex, a triangular face is a 2-dimensional simplex, and so on (Figure 12). If more than 2 simplices are connected with either the same or differing dimensionality, it is still a simplex and called a *simplicial complex*. These simplicial complexes are considered “trivial” and do not count towards *Homology*.

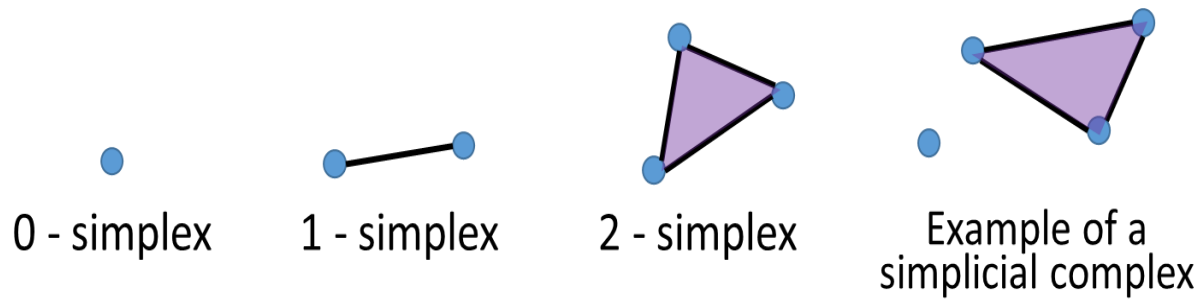


Figure 12 Some exemplary simplexes. A point forms a 0-simplex. A line that connects two points make 1-simplex, and a triangular face formed by 3 connected points make 2-simplex, and so on. A simplicial complex is an object built from combinations of these simplexes.

Homology, in a very crude definition, is the term that counts connected components, holes, voids, etc. For example, in our simple example in Figure 13, homology counts 1 connected component and 1 tunnel at the center. In this study, since we deal with 2-D MRI tumor image patch data, we consider only connected components and tunnels.

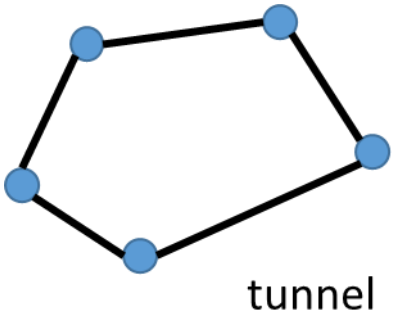


Figure 13 A graphical description of homology. Homology is the term that counts the number of connected components and holes. Here, we count 1 connected component and 1 tunnel.

Now back to our original example. This time, we connect the dots that are within distance d from circles and also fill in the simplexes if any exists. Again, as a reminder, simple complexes do not count towards homology. Hence, we detect both a hole at the center and the connected components surrounding it (Figure 14). However, even if we detected some meaningful topological features now, we still have a problem: the distance, “d” is some arbitrary measure and needs to be specified. Hence, we consider every diameter d, namely from 0 to infinity.

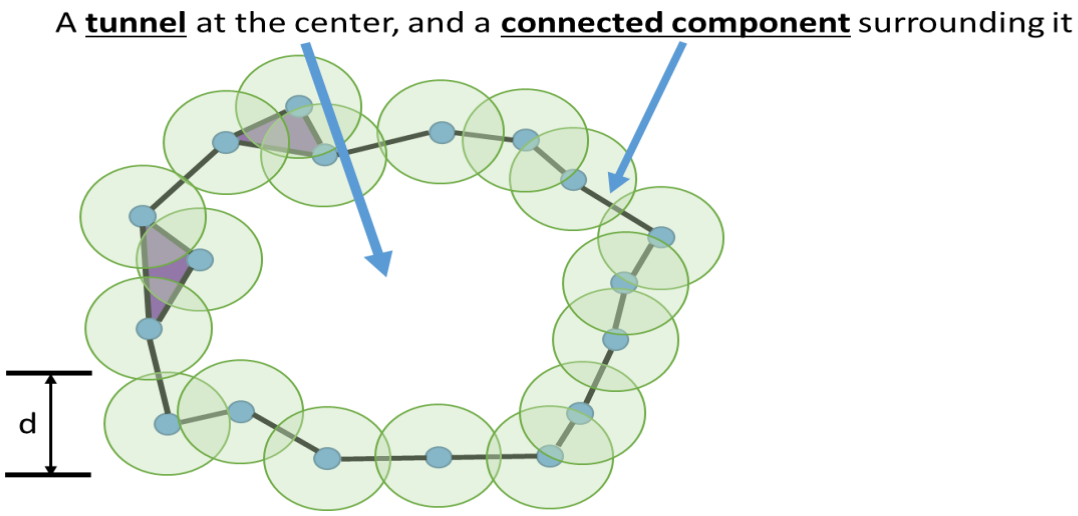


Figure 14 From the cloud of points, using homology, we can count the number of tunnels. There is one tunnel at the center and one connected component surrounding it.

4.2.4 Persistent Homology and Betti number

Here, we consider a new example shown in Figure 15. At diameter d1, we get a tunnel (Figure 15 (a)). But as we increase the diameter to d2, we lose the tunnel (Figure 15 (b)). This creation and destruction (or we could call them *birth* and *death*) of the tunnel can be visualized as a *barcode* (Figure 15(c)). This barcode represents the *persistence* of the tunnel as a pair (d1,d2).

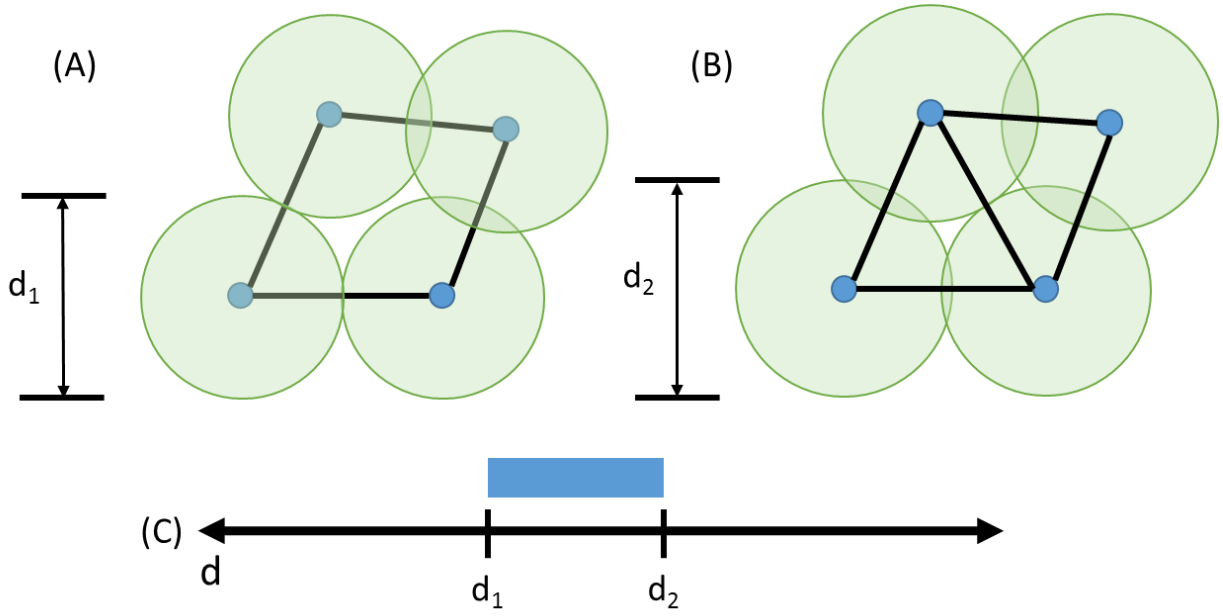


Figure 15 (a) At diameter d1, we create a tunnel. (b) As we increase the dimeter of circles to d2, we now destruct a tunnel. (c) This creation and destruction of the tunnel can be visualized as a barcode.

The *Betti number (β)* counts the number of these persistent homologies. Considering the example from Figure 15, we can see that there is one connected component that persists from d1 to infinity (β0). On the other hand, a single tunnel (β1) appears at d1 and then disappears at d2 This barcode representation is depicted in Figure 16.

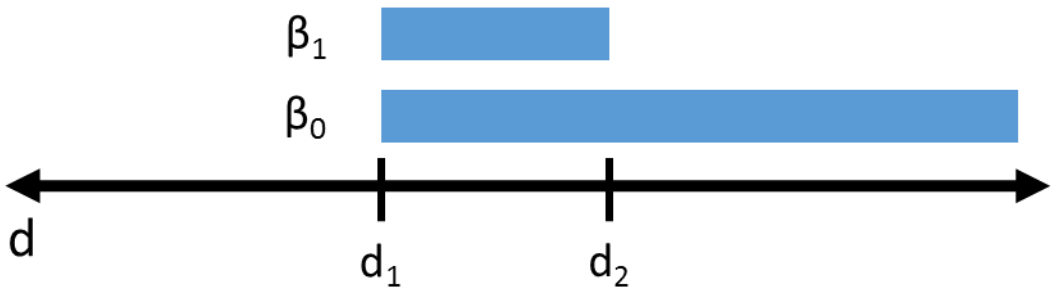


Figure 16 Betti number β counts the number of persistent homology. Considering the example from Figure 15, β0 represents the persistence of the connected components while β1 represents the persistence of the tunnel.

4.3 Why persistent homology?

It is well known that MRI scans suffer from different sources of degradation, including patient motion, inhomogeneous magnetic field (which has been addressed with N4 correction in Chapter 2), aliasing, etc. Hence, the integrity of our image data (i.e., the accuracy of the pixel intensities) always carries some uncertainty. However, it turns out that the persistent barcodes are stable under perturbations of the data (42), which allows one to analyze data within the margin of error. As such, persistent homology has been widely applied in many different fields, including machine learning (43), analysis of remote sensing data (44), and percolating surfaces and porous materials (45). It has also been utilized in medical datasets, such as brain networks (46), brain arterial tree structure (47), protein binding (48), orthodontics (44), gastrointestinal tract endoscopy images (49), and CT hepatic lesions (40).

Overall, persistent homology analysis has been highlighted as one of the most novel and widely accepted mathematical approaches in analyzing complex datasets, and we attempt to analyze glioma MRI scans in light of persistent homology features.

4.4 Applying Persistent Homology to image data

This part of the work closely follows that of Adcock et al. (40).

So far, we have dealt only with a cluster of data points. In order to convert the cloud of data points into images, we first organize them into a 2D matrix and consider these points as vertices at the center pixels in an image. Then, instead of measuring distance as our metric to check the neighboring points, we can measure intensity values as our new metric. Specifically, we can construct persistent barcodes in the following procedures:

1. Given an image and an ROI, set vertices at each ROI pixel (excluding the background pixels, which are labeled NA)

2. Increase the intensity from 0 to infinity (just as we did with distance for the cloud data points)
3. If any neighboring pixels (among the eight adjacent pixels) meet the intensity threshold, then we connect them
4. Count connected components and tunnels (but do not count simplex complexes)
5. Draw persistent barcodes for different Betti numbers

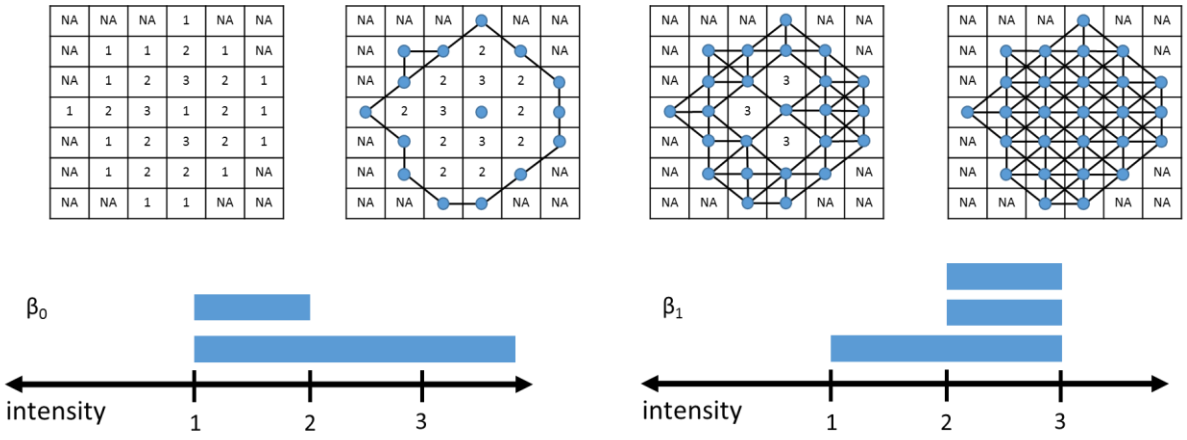


Figure 17 A process of creating persistent barcodes using image data. First we set vertices at each ROI pixel. Then, we increase the intensity from 0 to infinity. As we increase the intensity level, if any neighboring pixels meet the intensity threshold, we connect them and count the number of connected components and tunnels. For each Betti number, we draw persistent barcodes. This figure was redrawn with permission from A. Adcock, D. Rubin, and G. Carlsson, "Classification of Hepatic Lesions using the Matching Metric," Oct. 2012.

An example is demonstrated in Figure 17. First, we increase our intensity value from 0 to 1. When the intensity value is equal to 1, we see that there are two connected components. Notice that we also created a 2-dimensional simplex (i.e., the triangular facet), which does not count as a tunnel. Also, we see a tunnel (or rather a ring) appears at intensity 1, which is reflected on β1. If we increase the intensity to 2, one of the connected components, specifically the point at the center, is now incorporated into the surrounding connected component, and hence disappears. This observation is reflected on the β0 barcode plot. Concurrently, we now obtain three tunnels, one from the original tunnel at intensity value 2, and two new tunnels. If we

increase the intensity to 3, we destroy all the tunnels (β1 barcode), and only the connected component persists to infinity.

4.5 Persistent homology and barcode generation software

The Geometry Understanding in Higher Dimensions (GUDHI) package version 2.1.0 was used for the calculation of persistent homology and for barcode generation. This software is an open source Python based package that is freely available online at http://gudhi.gforge.inria.fr/.

4.6 Barcode generation from multimodal MRI tumor patches

In order to apply this approach to our four sequences of MRI tumor patches (T1, T1post, T2, and FLAIR), first we scaled the patch intensity distribution into the range of values from 1 to 2. Please refer to section 2.5.4. for more information. Also, if the barcodes went to infinity, we terminated at 2.1. This occurred for the connected component (Betti number 0), as all of the higher-order features (i.e., the tunnels) eventually died off and get incorporated into connected components as the intensity value went up.

Below, some sample barcodes and their corresponding images are presented (Figure 18).

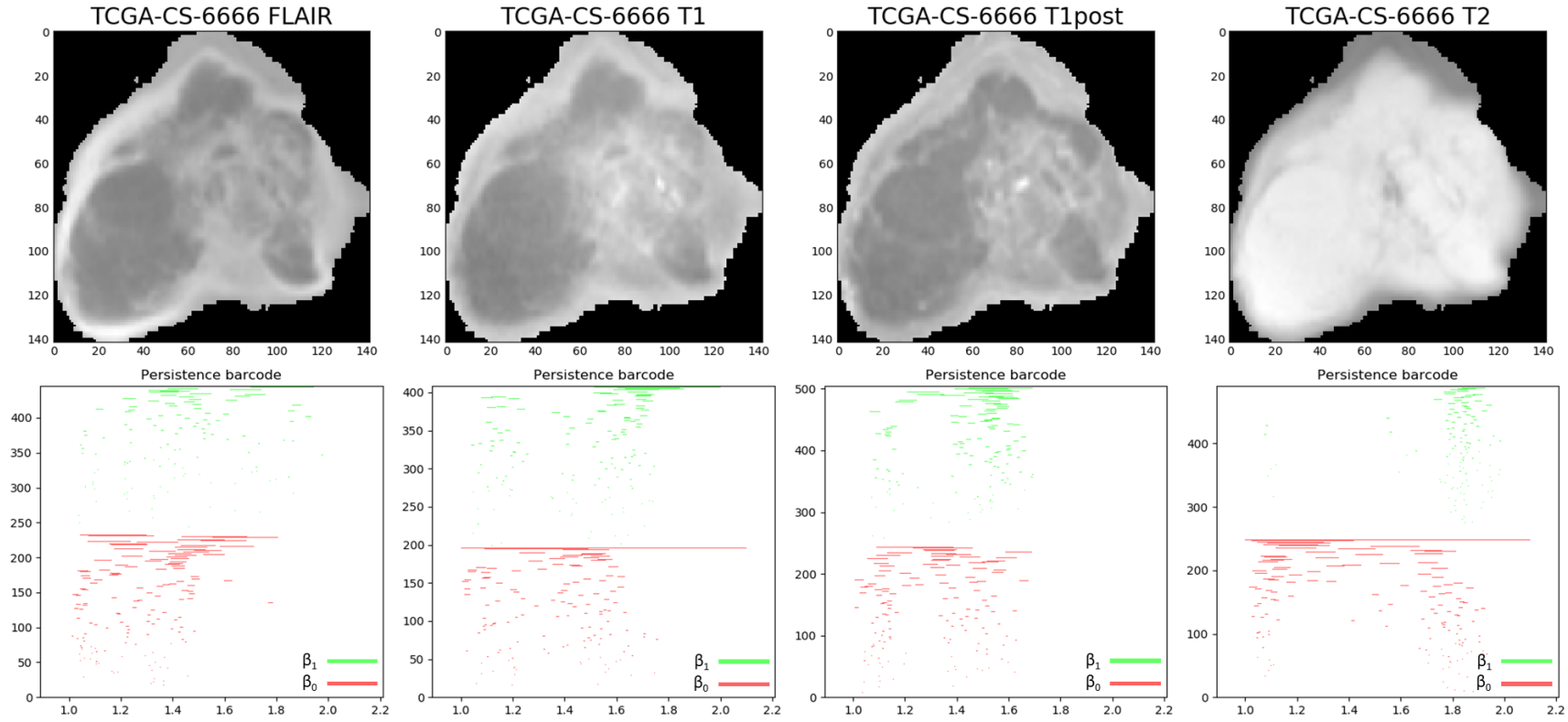


Figure 18 A sample patient tumor patch (TCGA-CS-6666) for all four MRI sequences with their corresponding persistent barcodes.

4.7 Barcode Feature Extraction

For the barcode feature extraction, we referred to Adcock et al. (41) and Giansiracuas et al. (50). The following features were computed:

- Polynomial feature 1 = $\sum_i^n b_i(d_i - b_i)/n$
- Polynomial feature 2 = $\sum_i^n (d_{max} - d_i)(d_i - b_i)/n$
- Polynomial feature 3 = $\sum_i^n b_i^2(d_i - b_i)^4/n$
- Polynomial feature 4 = $\sum_i^n (d_{max} - d_i)^2(d_i - b_i)^4/n$
- Mean/median/std. of b_i (i.e. birth intensity)
- Mean/median/std. of d_i (i.e. death intensity)
- Mean/median/std. of barcode length (i.e. death intensity – birth intensity)
- Mean/median of $(d_{max} - d_i)$

Where

- $b_i = i^{th}\ birth$
- $d_i = i^{th}\ death$
- $d_{max} = \ maximum\ death$
- $n = total\ number\ of\ bars$

In sum, we extracted 4 polynomial + 11 statistics based = 15 features for 2 Betti numbers (connected components and tunnels) for 4 modalities, which resulted in 120 (15 x 2 x 4) features in total.

Chapter 5: Convolutional Neural Network Features

5.1 Introduction

So far, we have explored two very different features, texture-based and topology (or specifically, persistent homology) based, that capture different image information. While such features might be useful in classifying a patient's 1p/19q codeletion status, there are some inherent limitations. To understand these limitations, one must first understand the radiomics study pipeline. In general, a radiomics study consists of the following steps: image acquisition, image segmentation, feature extraction, feature selection and statistical analysis (51). Other than the errors produced in the image acquisition and segmentation steps, the first problem arises from the feature extraction step. Because the features are human-engineered, depending on which features are being investigated in a radiomics study, the statistical analysis would produce very different results. There are some "popular" sets of features, such as GLCM, first order statistics, etc, that are preferred by the radiomics community, but none is standardized yet, which leaves the door open as to which features to utilize. Second, because there are often more radiomic features being computed than the number of patient cohorts in the radiomics study, feature selection and reduction steps generally take place to reduce the dimensionality of the feature space. However, such procedures invite another source of human bias as they require more manual engineering, such as deciding how many features to select or to what lower dimension that feature space would be transformed into. Lastly, there is no standard metric for image feature evaluation such that no one can verify its accuracy. Since it is hard to verify the reproducibility of the computed image features, there is always room for unintended errors (50) .

To overcome some of the aforementioned shortcomings of radiomics methods, a convolutional neural network (CNN) was investigated in this study. A CNN is a subset of deep learning algorithms (or more broadly, machine learning algorithms) that allows hierarchical feature extraction and transformation of image data. The input image is passed through a series of convolutional layers with non-linear activation function units, and each layer learns its own

feature representation of the data from simple to complex as the network gets deeper (52). At the bottom layer, based on the learned hierarchical feature representation, the network produces a vector of probabilities for target labels. Unlike the conventional radiomics study, which requires a “manual” feature extraction and feature selection, a CNN performs both of them automatically, and hence no extra errors come into play due to feature calculations.

5.2 Previous work

CNN has been a buzzword for the past 6-7 years, especially in computer vision, due to its superhuman accuracy in object recognition tasks, such as the ImageNet challenge (53). It has also been used in the medical field including tumor segmentation (54), image registration (55), image denoising (56), and radiogenomics (57). In particular, Akkus et al. demonstrated that a multi-scale CNN can predict 1p/19q codeletion status (15). However, their study only included axial plane T1post and T2 MR scans, which limited the utilization of other sequences such as T1 and FLAIR as well as dimensional information. In our approach, we utilized four sequences that are routinely acquired in most MR studies of the brain, namely FLAIR, T1, T1-post, and T2, in all three planes: axial, coronal, and sagittal.

5.3 Residual Convolutional Neural Network

Here, we considered a residual CNN (ResNet), which won the 2015 ImageNet challenge by introducing “residual connections” (58). Instead of simply stacking up non-linear layers in which the network is trained to learn the original mapping, by passing an identity mapping from one layer to those that are higher up, the network is optimized to learn the “residuals”.

Our network was derived from a 34-layer ResNet. As with the original residual network architecture. Batch normalization was used after every convolutional layer to reduce internal covariate shift and overfitting (59). The top two layers of the original ResNet were modified as the size of our input (142 × 142) is smaller than that of the original one (224 × 224).

5.4 Transfer Learning: off the shelf approach

Transfer learning is a technique in which a pre-developed model for a source dataset is re-used to improve generalization in a target dataset. There are three possible benefits if the transfer learning is done right: A higher performance starting point, a higher learning slope, and a higher performance score (60). Transfer learning works because the learned features from the source task are "general" such that these features are suitable to both the source and target task, not just specific to the source task (61).

For this study, we utilized a pre-trained ResNet-34 from Ken Chang's work on IDH-1 mutation (57). His model was particularly well-suited to this investigation because the network was trained with the very same set of MRI pulse sequences as those in the present work (i.e. FLAIR, T1, T1-post, and T2). Since the target dataset (i.e., the 1p/19q dataset) is small but of the same fundamental nature as the original (i.e., Chang's dataset), one would expect his pre-trained network would produce features that are relevant to determining the 1p/19q codeletion status.

5.5 CNN Feature Extraction

From the pre-trained ResNet-34, we took out the last classifying layer (or sigmoid layer). Then we fed our input images into the network, and extracted features from the flattened average pooling layer with 512 nodes. Hence, we obtained a vector of 512 features for all four modalities, resulting in 2048 features per sample. This process is depicted in Figure 19.

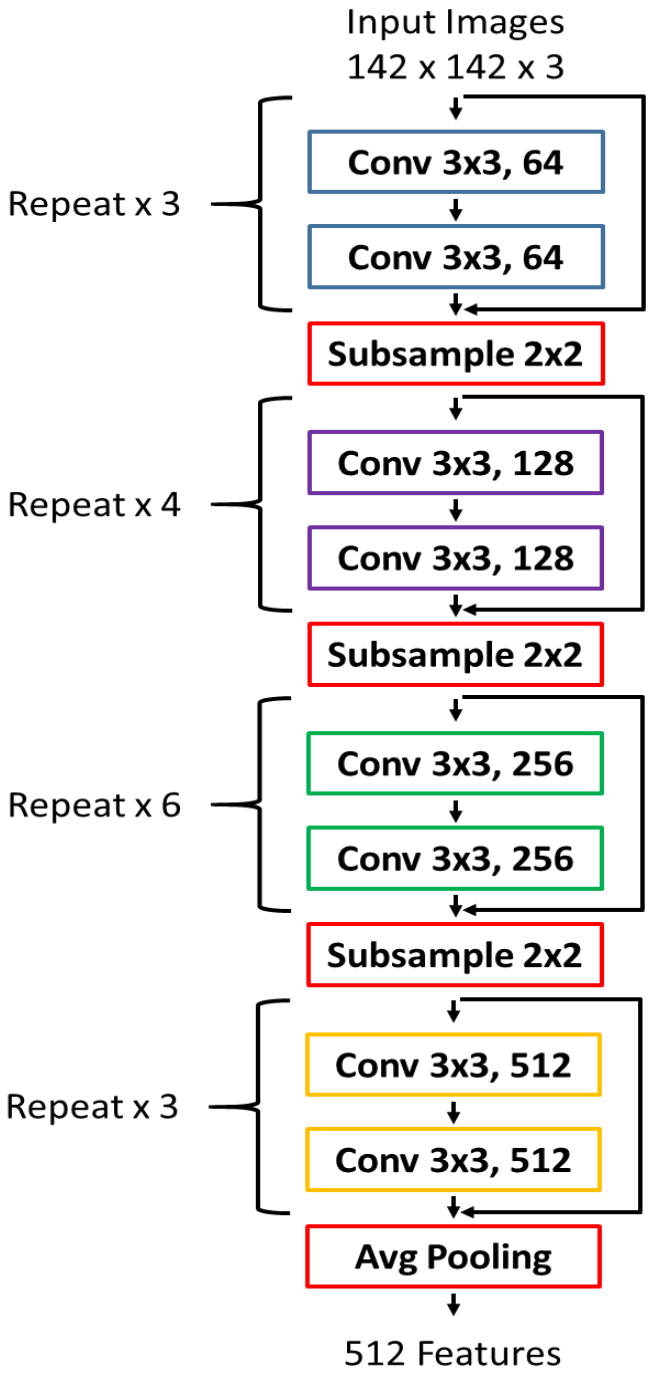


Figure 19 A modified ResNet-34. We took the pre-trained ResNet-34 as a feature extractor from Ken Chang's work on predicting IDH mutation status for glioma patients. At the last flattened out average pooling layer, we obtained a vector of 512 features for each image. Since there are four MRI sequences, we obtained a total of 2048 features. Permission obtained from K. Chang et al., "Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging," Clin. Cancer Res., vol. 24, no. 5, pp. 1073–1081, Mar. 2018.

Chapter 6. Feature Selection, Feature Reduction, and Modeling

Up until this point, we have discussed data acquisition, pre-processing, and feature acquisition processes. In this chapter, we will discuss feature selection and reduction and modeling processes.

6.1 Feature Selection: Correlation filtration and Recursive Feature Elimination (RFE)

In order to select a subset of given features, correlation filtration and recursive feature elimination were implemented. As the first step, we filtered out features that are linearly correlated. If the observed features are too correlated with each other, then several problems can arise: 1. the interpretation of individual predictor variables is no longer reliable. Most of the known linear regression or linear models assume uncorrelated features, hence the interpretation of the coefficient of a certain variable estimates the effect of that particular variable in predicting the target (while holding the other variables unchanged). If, however, two variables are correlated, and the uncorrelated feature assumption no longer holds, it is hard to estimate the true effect of an independent change in one variable. 2. As mentioned earlier, having more features than is necessary, especially when only one variable is enough to describe certain characteristics of the data, puts us into the curse of a high dimensionality problem. To overcome these two problems, we measured the Pearson correlation coefficient in a pairwise fashion. The coefficient value lies between -1 and 1, with -1 meaning perfect negative correlation while +1 means perfect positive correlation. A value of 0 means that there is no linear correlation between the two observed variables, which is what we want. Here, we used the absolute value of 0.8 as a threshold to filter out the correlated features.

Once we removed the linearly correlated features, we then implemented recursive feature elimination (RFE) (62) to further reduce the number of features to select. The way in which RFE works is the following: First, the user specifies an external estimator, which we chose to be a random forest classifier. Then, the classifier is trained with the initial set of features and calculates the importance of the individual feature, which is gini impurity in our case. Then, the least important features are pruned from the initial set of features, and this procedure is recursively repeated on the pruned set until the user-specified number of features is all that remains. The problem with such a method, however, is that the user has to specify the number of features, and often, that number is empirically determined by cross validation. Here, we measured the area under the curve (AUC) of the Receiver Operator Characteristic (ROC) curve across all possible numbers of features with 11-fold leave one group out (LOGO) cross validation. LOGO was implemented to prevent patches from one patient from spilling over into both the training and the validation datasets. In other words, the tumor region patches from one patient must belong to either the training or the validation set, but not to both.

This feature selection was only applied to texture features and topological features because we could not find a single reference in the literature that applied feature selection on CNN features that had been produced by transfer learning.

6.2 Feature Reduction: Principal Component Analysis (PCA)

Unlike feature selection where a subset of features is chosen and retains its original feature characteristics, feature reduction aims to reduce the dimensionality of features by projecting the original features into a lower dimensional space. Principal component analysis (PCA) achieves such a projection via Singular Value Decomposition (SVD) of the data while maximizing the total variance of the projection. By doing so, one also benefits from converting possibly correlated variables into a set of linearly independent variables (i.e. principal components). Here, we first standardized (i.e. transformed to zero mean and unit variance) our

training dataset, applied PCA, and looked for the number of PCs that explain 95% of the data variance. Once the number of PCs was fixed, we repeated PCA on the training dataset with the number of PCs specified. Then, using the same PCA statistics from the training dataset, we applied them to the test dataset.

6.3 Modeling: Logistic Regression and Random Forest

Two classifiers were considered in this study: logistic regression and random forest (RF). Table 2 summarizes the hyper-parameters that were optimized during training. Again, we performed LOGO to ensure that there is no information spillage from training to validation or vice versa. Also, we considered the cases in which we provided the clinical variables, including KPS, age, and gender, or not. In total, we built eight different classifiers depending on the classifier type (RF vs Logistic Regression), the feature type (CORR + RFE selected vs PCA features), and clinical variables (either considered or not) (2 x 2 x 2 = 8). In addition, we built a logistic regression with clinical variables only (i.e., age, gender, and KPS) as our baseline. According to Boots-Sprenger et al., age is correlated with 1p/19q codeletion status (63), hence it is not a bad choice of features around which to build a classifier.

For the assessment of each classifier, we computed the AUC from the ROC, accuracy, sensitivity, and specificity using Youden's J statistic.

Random Forest Hyper-parameters	
Number of trees in the forest	200,400, …, 2000
Number of features to consider when looking for the best split	sqrt(number of features)
Maximum depth of the tree	10,20, …, 100
Minimum number of samples required to split an internal node	2,5,10
Minimum number of samples required to be at a leaf node	1,2,4

Logistic Regression Hyper-parameters	
Regularization Type	L1, L2
Regularization Strength	10E-4, 10E-3, … 10E4

Table 2 Hyper-parameters for random forest and logistic regression.

6.3.1 Feature Importance

In order to understand the contribution of each feature in predicting 1p/19q status, we extracted the feature importance from the classifiers. For the random forest, Gini impurity (in magnitude) was used as the feature importance. For logistic regression, the top three negative and the top three positive coefficients (or weights) of the feature variables were considered to be important features.

The reason that we retained the top three negative and the top three positive coefficients of the feature variables is the following. Recall that a basic form of the logistic regression equation is:

$$P(Y = 1|x) = \frac{1}{1 + e^{-z}} \; where \; z = \theta^T x$$

- θ : A coefficient vector in front of the feature variables where $\theta = [\theta_0, \theta_1, \dots, \theta_n]$. n is the number of feature variables. Note that θ_0 is a real number that acts like a y-intercept (i.e. it is not multiplied by a feature variable), hence $\theta \in \mathbb{R}^{n+1}$.
- x : A feature variable vector where $x = [x_1, x_2, \dots, x_n]$. n is the number of feature variables. $x \in \mathbb{R}^n$.

From the equation, we can see that as $z \to \infty$, $P \to 1$. On the other hand, if $z \to -\infty$, $P \to 0$. Now, we realize that it is the $\theta^T x$ that determines the overall magnitude of z. We also realize that depending on the "sign" of θ, or coefficients, z can either lean towards negative or positive. In other words, we can say that negative coefficients push towards the probability being 0, or non-codeleted, whereas positive coefficients push towards the probability being 1, or codeleted. Understanding of the meaning of negative and positive coefficients in logistic regression becomes important as we later on try to visualize the features back onto the image.

Also, note that only the classifiers that were trained with *non-PCA reduced features* were considered for this analysis, as PCA destroys the meaning of the original features as it projects the feature space into a lower dimensional subspace with possibly different basis vectors than the original feature space.

Chapter 7. Results

7.1. Feature selection results

In theory, for CNNs, feature selection and reduction are performed automatically during the learning process, and hence only the “relevant” features are present at the final layer of the architecture. As such, feature selection was only performed on the texture and topological features, but not on the CNN features.

7.1.1 Correlation filtering

Appendix A shows heatmaps of the Pearson coefficient of features after thresholding at an absolute value of 0.9. Recall that the lower this value is, the less correlated are the features.

From correlation filtering, we could retain 19.0% of linearly correlated data for the texture features (n=69) and 26.7% for the topological features (n=32). Still, it seems that there are just too many variables to consider given the number of training data, both for the texture and the topological features.

7.1.2 RFE

Below, we present the mean AUC across 10 train/test splits vs. the number of features selected for 11-fold LOGO cross validation (Figure 20 (a) for the texture and (b) for the topological features). In practice, we want to see an increase in AUC initially, and then saturation as we consider more features. By inspecting the graphs, such a trend seems to be the case. For the optimal number of features, we simply selected the one that produced the highest AUC score (n=17 for texture and n=22 for topology). Once the number of features to select was fixed, we re-ran RFE and retained those as our final set of features to train with. For each train/test split, the final feature sets that were selected are listed in Appendix B.

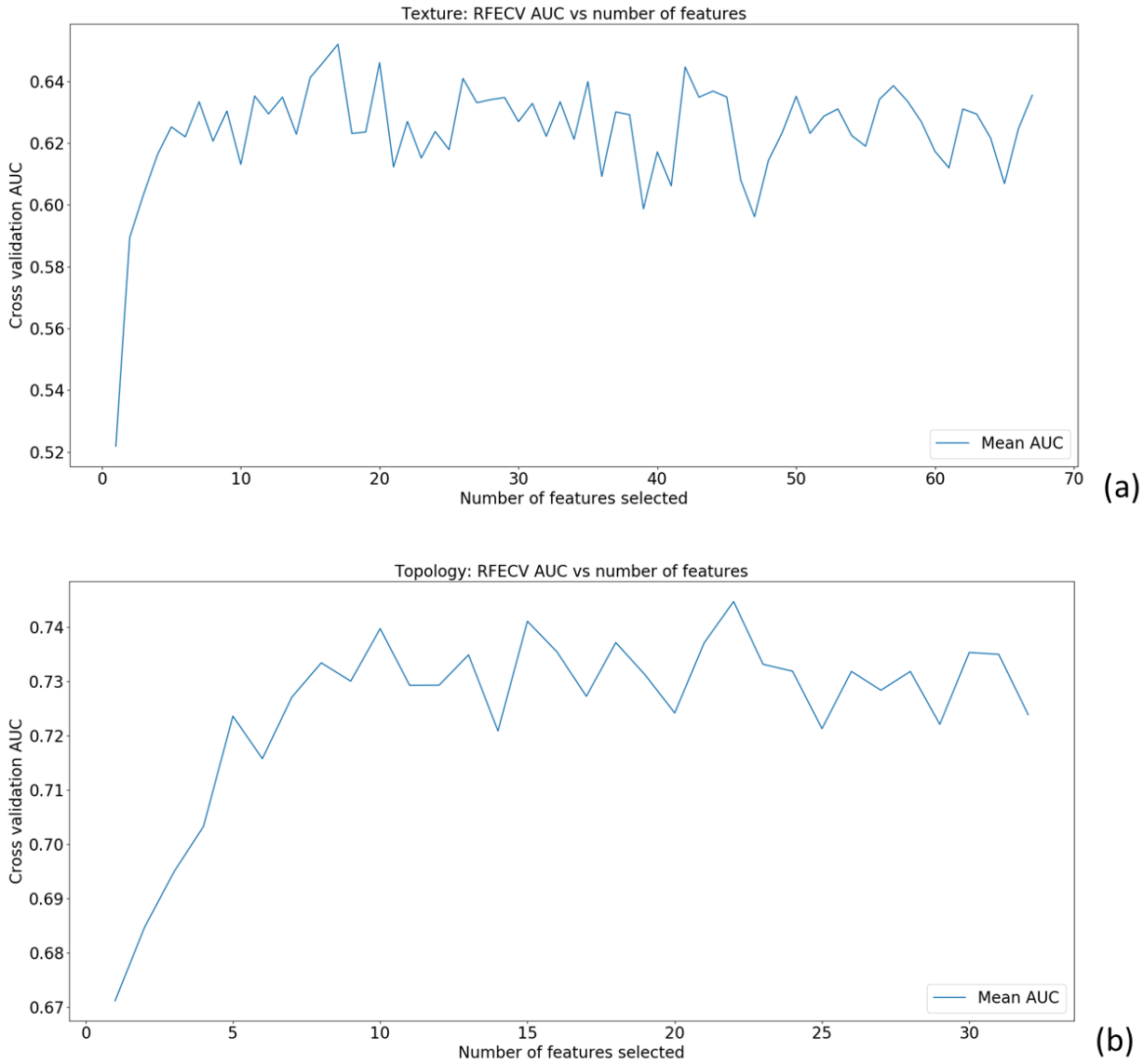


Figure 20 The mean AUC across 10 train/test splits as a function of the number of features selected for 11-fold LOGO cross validation. We took the highest mean AUC yielding number of features as our optimal number of features to select. (a) is for texture features while (b) is for topological features.

7.2. Feature reduction results

Figure 21 shows the percentage of the variance that is explained vs Principal Component number on the training data set.

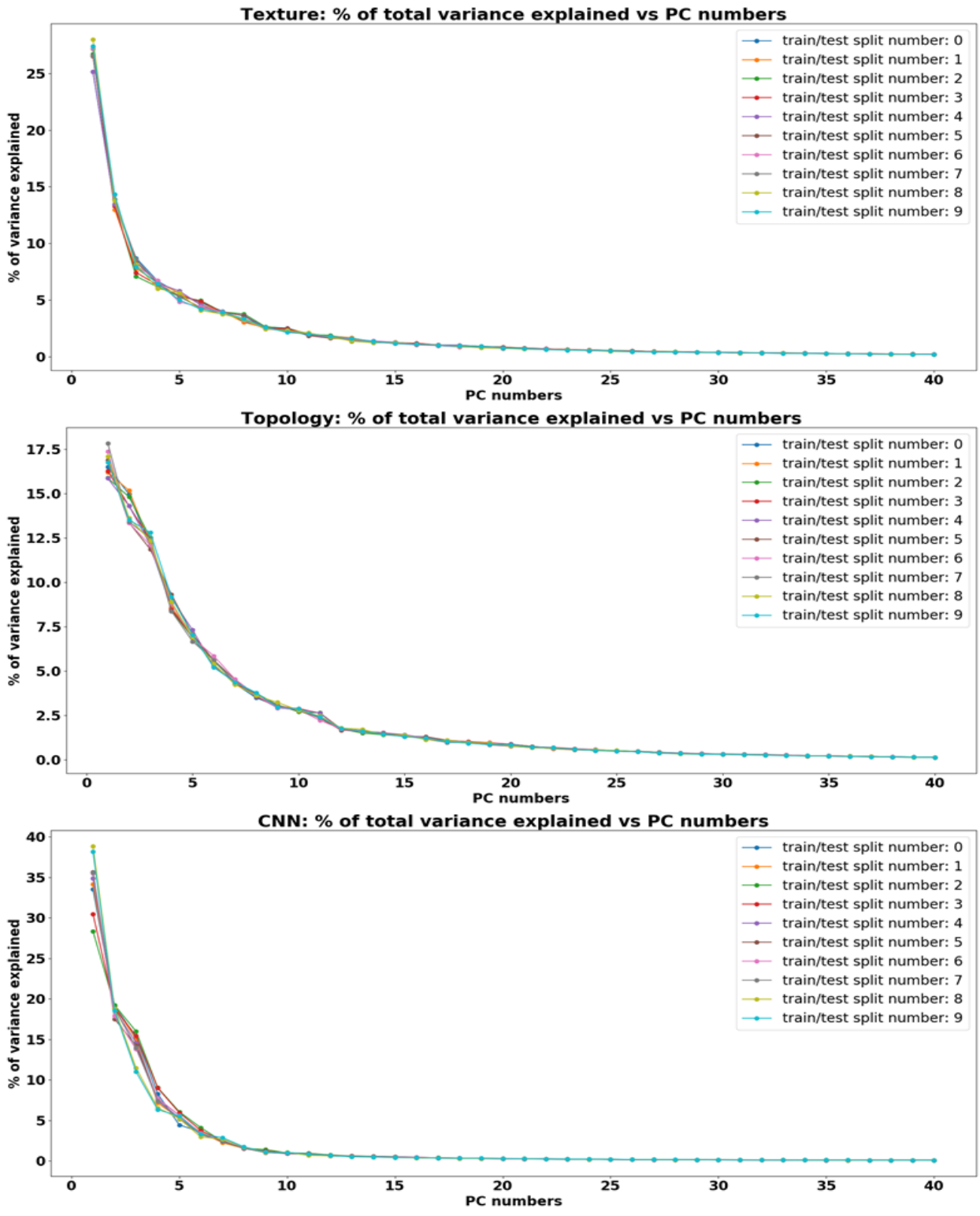


Figure 21 The percentage of the variance explained vs. Principal Component number of the training data set for (a) texture-, (b) topological-, (c) CNN-based features.

The following figure shows the cumulative percentage of the variance that is explained over Principal Components (Figure 22).

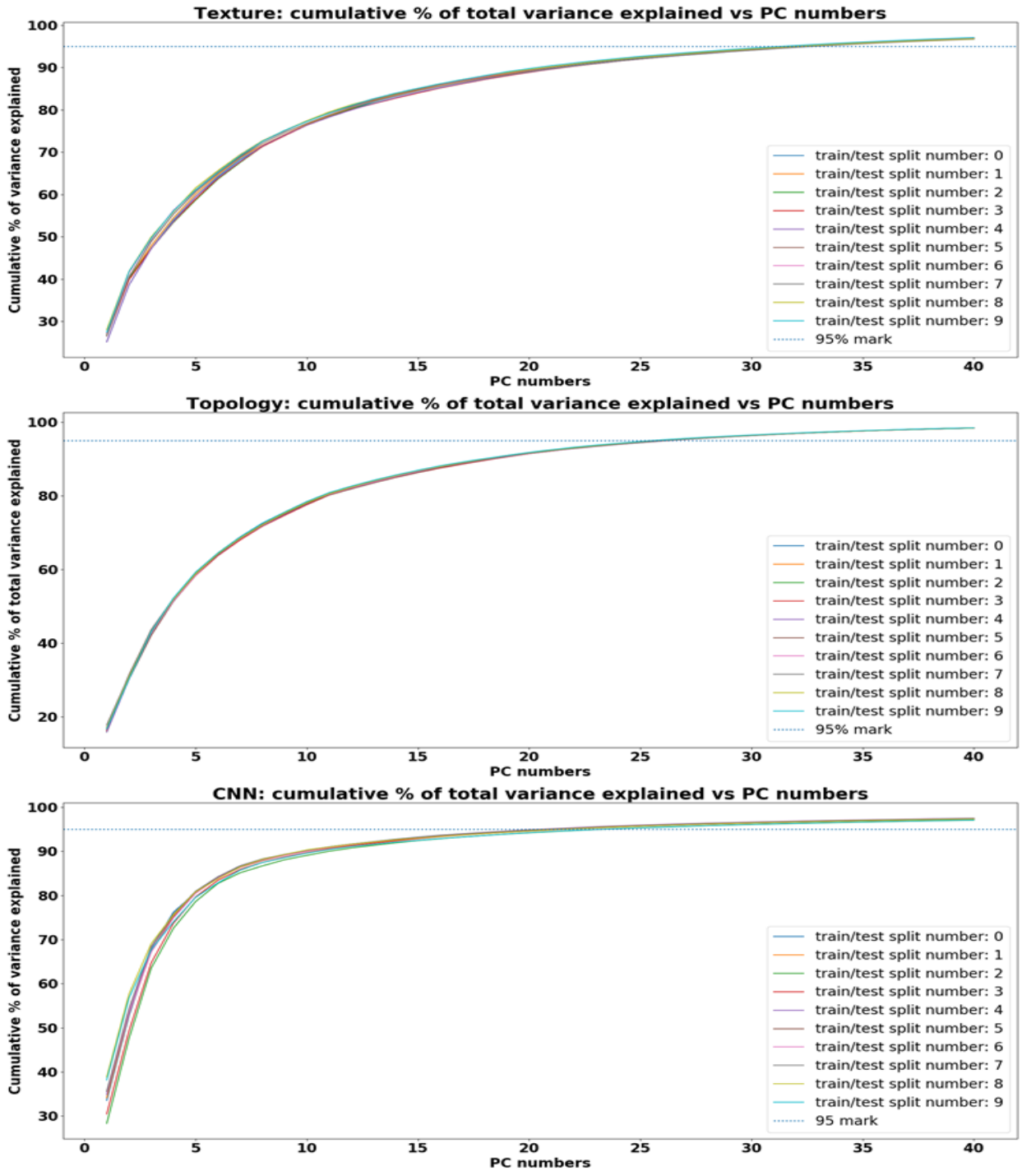


Figure 22 The cumulative percentage of the variance explained over Principal Components for (a) texture-, (b) topological-, (c) CNN-based features.

To explain 95% of the total variance, we need 33, 26, and 22 PCs for texture, topology, and CNN features respectively.

7.3 Modeling Results

7.3.1 Model performance in predicting 1p/19q status

Here, we are reporting mean statistics (AUC, accuracy, sensitivity, and specificity) ± standard deviations across 10 different test sets (Table 3, Figure 23 and 24).

Classifier Type		Logistic				Random Forest			
Metric		AUC	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity
Features	texture	0.593 +/- 0.222	0.643 +/- 0.285	0.800 +/- 0.230	0.627 +/- 0.328	0.707 +/- 0.118	0.609 +/- 0.161	0.950 +/- 0.105	0.575 +/- 0.186
	topology	0.855 +/- 0.079	0.820 +/- 0.107	0.925 +/- 0.169	0.810 +/- 0.126	0.854 +/- 0.116	0.825 +/- 0.133	0.900 +/- 0.175	0.818 +/- 0.148
	CNN	0.787 +/- 0.195	0.905 +/- 0.102	0.775 +/- 0.249	0.917 +/- 0.121	0.503 +/- 0.204	0.652 +/- 0.260	0.725 +/- 0.249	0.645 +/- 0.305
Features + Clin. Var	texture	0.645 +/- 0.173	0.666 +/- 0.203	0.875 +/- 0.212	0.645 +/- 0.238	0.747 +/- 0.154	0.659 +/- 0.197	1.000 +/- 0.000	0.625 +/- 0.216
	topology	0.861 +/- 0.119	0.811 +/- 0.137	0.975 +/- 0.079	0.795 +/- 0.151	0.863 +/- 0.127	0.845 +/- 0.176	0.925 +/- 0.169	0.838 +/- 0.199
	CNN	0.782 +/- 0.183	0.891 +/- 0.145	0.775 +/- 0.249	0.903 +/- 0.169	0.489 +/- 0.212	0.664 +/- 0.285	0.700 +/- 0.258	0.660 +/- 0.332
PCA featues	texture	0.594 +/- 0.248	0.695 +/- 0.285	0.725 +/- 0.275	0.693 +/- 0.328	0.576 +/- 0.181	0.541 +/- 0.224	0.875 +/- 0.212	0.507 +/- 0.260
	topology	0.811 +/- 0.103	0.773 +/- 0.101	0.925 +/- 0.169	0.758 +/- 0.117	0.763 +/- 0.158	0.705 +/- 0.209	0.925 +/- 0.169	0.682 +/- 0.238
	CNN	0.461 +/- 0.207	0.439 +/- 0.217	0.925 +/- 0.169	0.390 +/- 0.251	0.526 +/- 0.138	0.427 +/- 0.124	0.975 +/- 0.079	0.372 +/- 0.138
PCA features + Clin. Var	texture	0.617 +/- 0.209	0.666 +/- 0.303	0.800 +/- 0.258	0.652 +/- 0.350	0.672 +/- 0.197	0.700 +/- 0.208	0.850 +/- 0.211	0.685 +/- 0.238
	topology	0.754 +/- 0.135	0.691 +/- 0.149	0.925 +/- 0.169	0.667 +/- 0.170	0.819 +/- 0.135	0.757 +/- 0.144	0.925 +/- 0.121	0.740 +/- 0.154
	CNN	0.665 +/- 0.198	0.620 +/- 0.214	0.950 +/- 0.158	0.588 +/- 0.244	0.697 +/- 0.309	0.734 +/- 0.317	0.900 +/- 0.211	0.718 +/- 0.355

* Mean value +/- standard deviation

** Accuracy, sensitivity, and specificity computed with Youden's J index

Table 3 AUC, accuracy, specificity, and sensitivity for three different feature types.

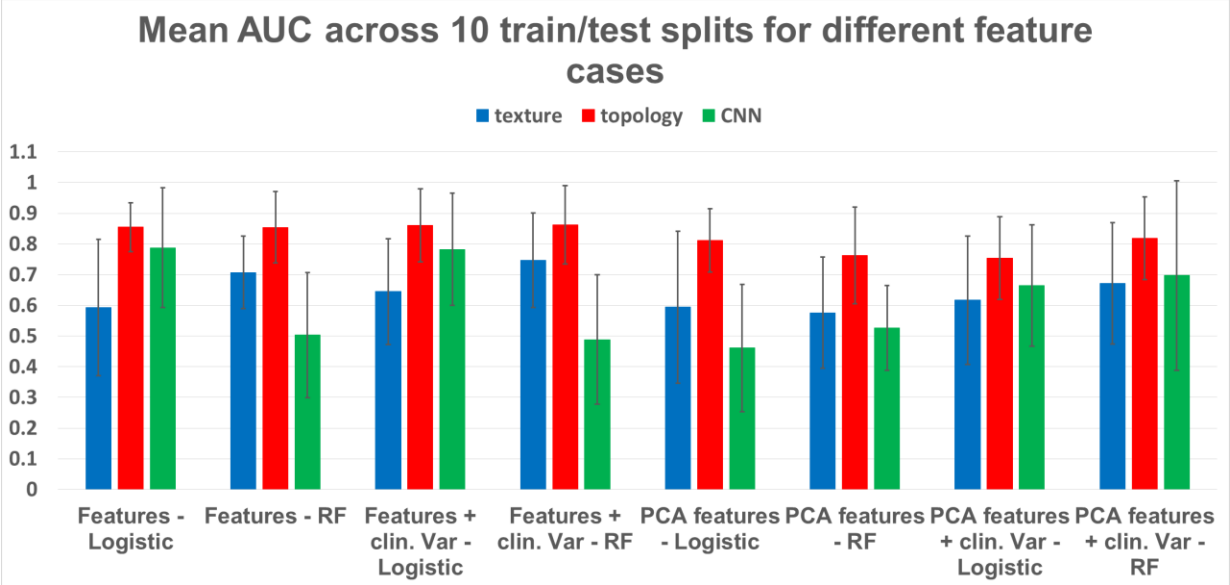


Figure 23 Mean AUC scores across 10 train/test split for different classifier cases. Error bars represent the standard deviation from the mean.

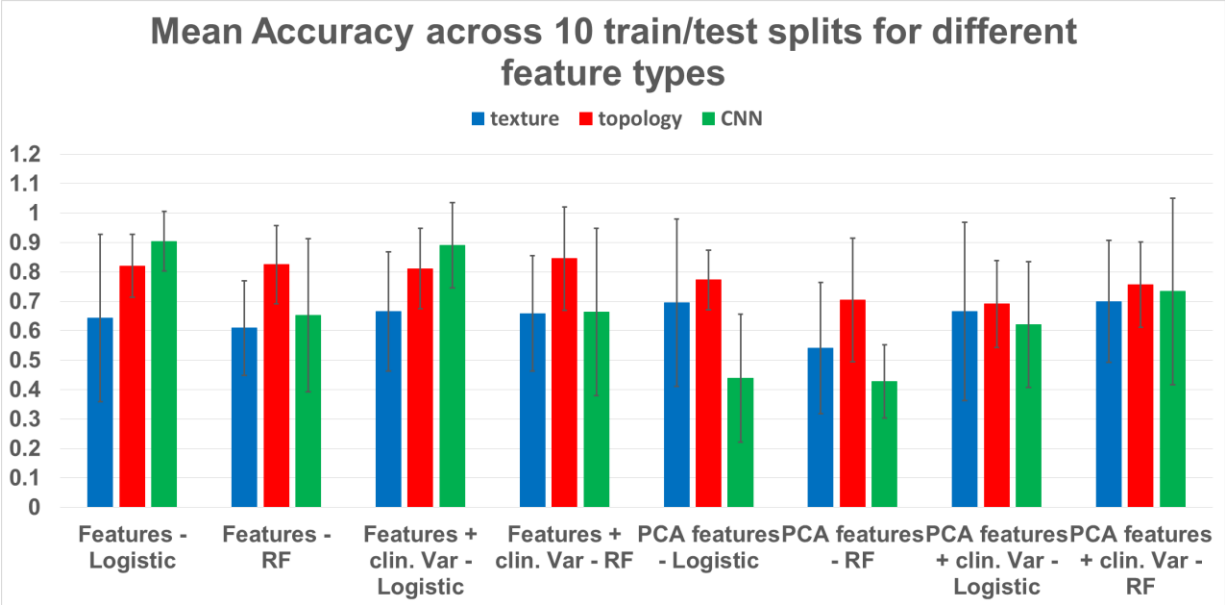


Figure 24 Mean Accuracies across 10 train/test splits for different classifier cases. Error bars represent the standard deviation from the mean.

At first glance, it is probably natural to select models that yield the highest AUC scores. For instance, for topology-based features, a Random Forest with clinical variables considered model yielded the best AUC (0.863 +/- 0.127). However, a logistic regression without clinical variables considered (in other words, a much simpler model) yielded a very similar AUC within the standard deviation (0.855 +/- 0.079). In other words, selecting a model simply based on the

best AUC score is such a short-sighted approach. In order to simplify the complexity of the model while maintaining a reasonable AUC score, we considered the following questions:

1. Do clinical variables add any additional value in modeling?
2. Which dimensionality reduction method yielded the better performance (i.e. Correlation filtration + RFE vs. PCA)?
3. Which model to consider (i.e., Logistic regression or Random Forest)for each feature?
4. Does combining all of the feature types help to improve the model performance?

7.3.2 Assessing Clinical Variables impact on modeling

To assess if clinical variables added any additional value in modeling performance, we performed pair-wise Wilcoxon signed rank tests (Table 4). The null hypothesis was that the mean scores of the AUC between the model with clinical variables and the one without are the same. If the p-value is lower than 0.05, we rejected the null hypothesis and concluded that the means are significantly different.

		(texture, texture + clin. Var)	(texture PCA, texture PCA+ clin. Var)	(topology, topology+ clin. Var)	(topology PCA, topology PCA + clin. Var)	(CNN, CNN + clin. Var)	(CNN PCA, CNN PCA + clin. Var)
AUC	Logistic	0.386	0.161	0.646	0.141	0.833	**0.007**
	Random Forest	0.285	0.139	0.515	0.333	0.445	0.114

Table 4 Pair-wise Wilcoxon signed rank test between the models with clinical variables considered and the ones without. P-values on AUC scores are reported.

From Table 4, one can conclude that considering the clinical variables did not improve the modeling performance in terms of AUC and accuracy. The only exception is the case of the CNN PCA features with a logistic regression for AUC (p-value <0.05). However, from Table 3, CNN PCA features with a logistic regression achieved AUC of 0.461 +/- 0.207, indicating that the classifier itself was worse than pure chance (which would have an AUC of 0.5). Hence, we concluded that clinical variables did not add any additional value in the modeling process and disregarded all models that incorporated clinical variables.

7.3.2 Feature selection vs. Feature reduction

To assess which feature dimensionality reduction method performed better in terms of AUC, we plotted the AUC scores between feature-selected models and feature-reduced (PCA). Also, we performed the Wilcoxon test between pairs of cases and reported their p-values.

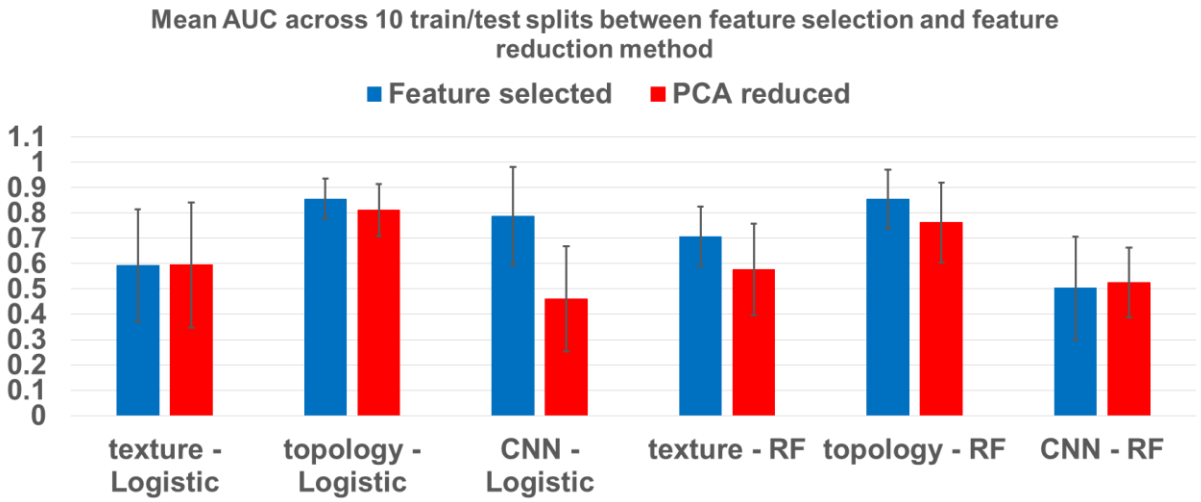


Figure 25 Mean AUC bar graphs between the feature-selected models and the feature-reduced models. Error bars represent the standard deviation. In general, the feature-selected models performed on a par with if not better than the feature-reduced models.

		(texture, texture PCA)	(topology, topology PCA)	(CNN, CNN PCA)
AUC	Logistic	0.959	0.059	**0.013**
	Random Forest	**0.012**	**0.037**	0.445

Table 5 Wilcoxon signed rank test between features selected and features reduced (PCA). P-values of AUC scores are reported here.

Regardless whether it was a logistic regression or a random forest classifier, the feature-selected models performed on a par with, if not better than, the feature-reduced cases. Hence, we conclud that feature selection is the better approach to reducing the feature dimensionality.

7.3.3 Final model choices

After filtering out all of the clinical variables added models as well as PCA feature models, the remainder are feature-selected (Correlation filtration + RFE selected) models only (Figure 26). We also performed the pair-wise Wilcoxon signed rank test between the logistic regression and random forest model for each feature type, and reported the p-values (Table 6)

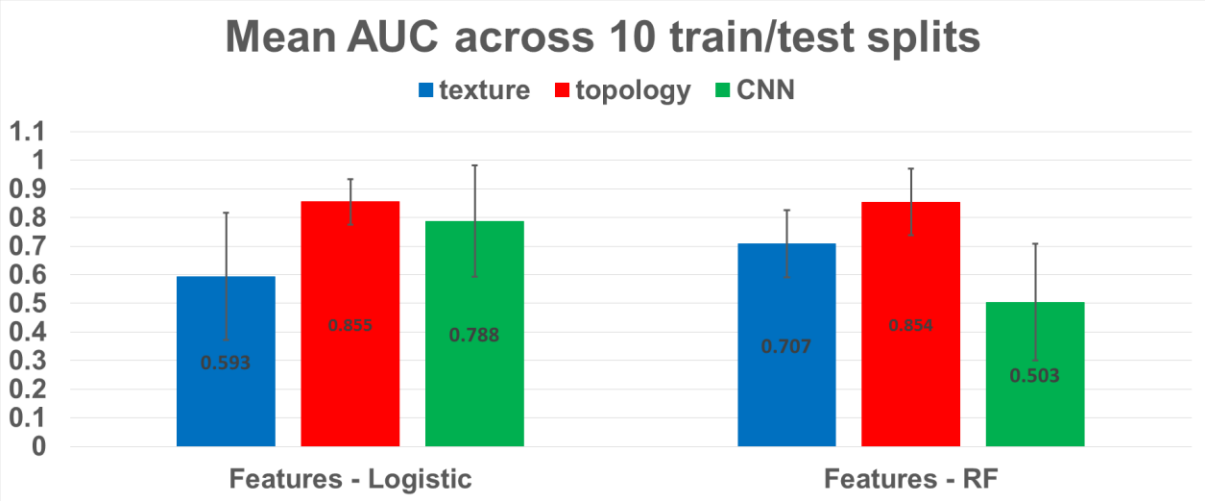


Figure 26 Mean AUC score across 10 train/test splits for each feature type. The error bar represents the standard deviation from the mean. For texture features, we selected the random forest. For topology and CNN features, we selected the logistic regression.

	Texture	Topology	CNN
p-values	0.074	0.859	**0.005**

Table 6 Pair-wise Wilcoxon signed rank test between logistic regression and Random forest. P-values on AUC scores are reported here.

For CNN features, it is clearly better to select the logistic regression model as it yielded better AUC score significantly (p-value < 0.05). For topology features, there is really no AUC score difference between the two models (p-value >0.05), hence we selected the logistic regression as it is a simpler model. For texture features as well, it seems that there is no difference between two models (p-value >0.05). However, for the logistic regression case, the standard deviation is 0.222, which would put the mean AUC score under the 0.5 mark in the worst case. Hence, we decided to select the random forest model.

From now on, unless stated otherwise, for CNN and topology features, a logistic regression model is used as our best model. On the other hand, for texture features, a random forest is used as the best classifier.

7.3.4 Individual feature and Combined features model analysis

Here, we wanted to assess which feature type performed the best among the three cases, namely texture-based, topology-based, and CNN-based. In addition, we built a logistic regression using all those three features to see if combining the features yields a better AUC score. To compare among these four different types of features, again, we performed a pairwise Wilcoxon signed rank test and assessed their p-values. Our null hypothesis was that the mean AUC scores are the same with our significance level at 0.05.

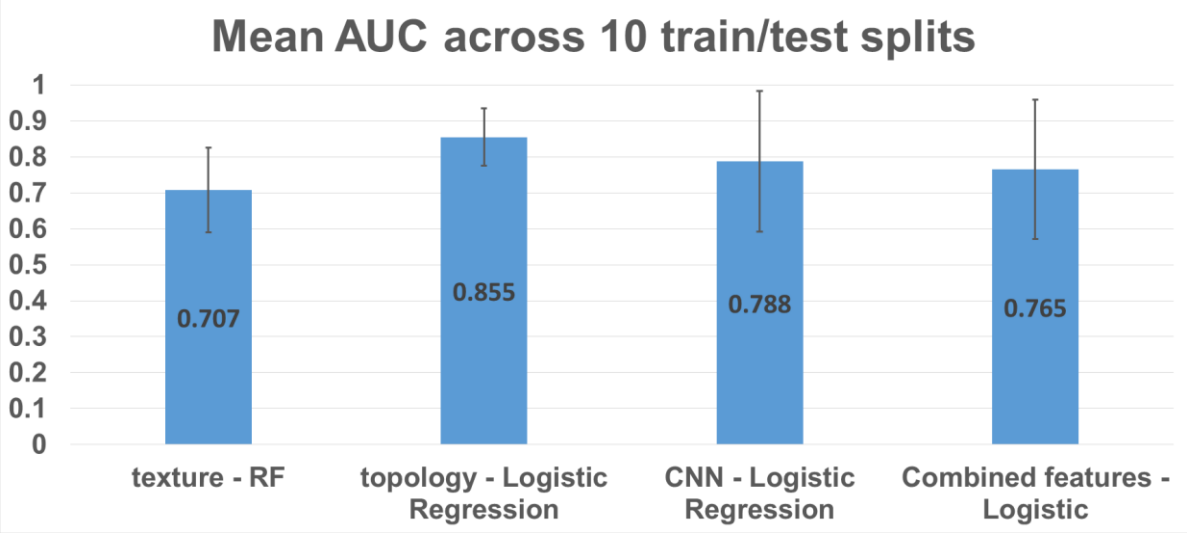


Figure 27 Mean AUC across 10 train/test splits among three feature types and combined features models.

	(texture, topology)	(texture, CNN)	(texture, Combined features)	(topology, CNN)	(topology, Combined features)	(CNN, Combined features)
p-values	**0.017**	0.203	0.203	0.386	0.386	1.0
* Combined features: texture + topology + CNN features						

Table 7 Pair-wise Wilcoxon signed rank test among individual feature models and combined features model. P-values on AUC scores are reported here.

From the pair-wise Wilcoxon signed rank test, we can see that the topology-based model outperformed the texture-based model significantly (p-value < 0.05). On the other hand,

combining all feature types did not help improving AUC score as none of the p-values was less than 0.05.

7.3.5 Comparing against the clinical variables only model

As our baseline, we built a logistic regression with clinical variables only (KPS, age, and gender) and compared it to our image features-derived models. To test our specific aim 2, we compared the AUC scores among the models and performed a pair-wise Wilcoxon signed rank test (Figure 28, Table 8). Also, Table 9 summarizes the AUC, accuracy, sensitivity, and specificity for each feature type using its best model case (that we selected from section 7.3.3).

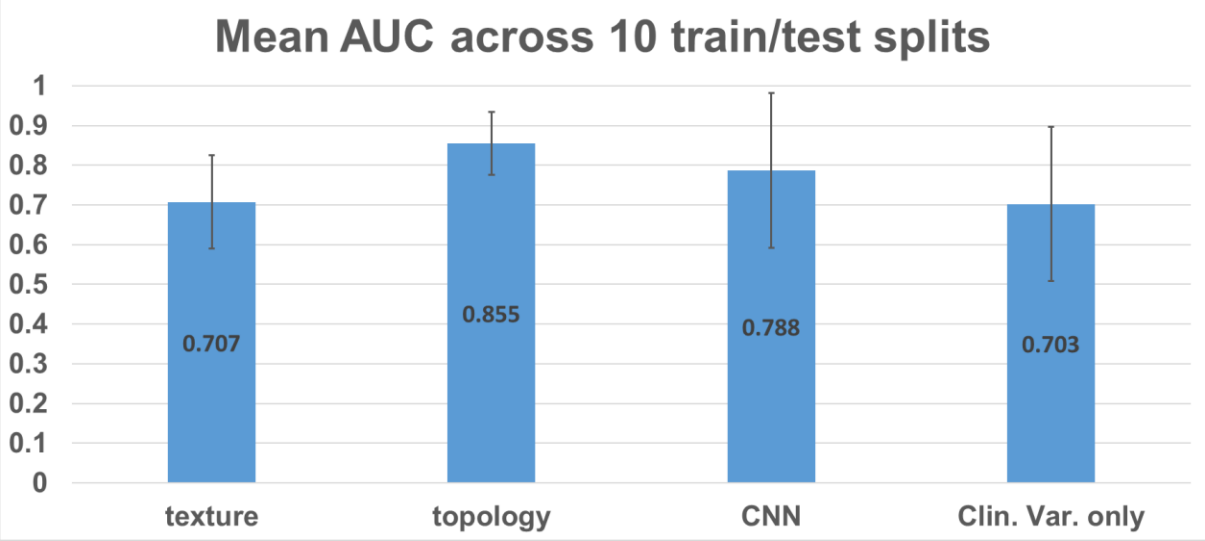


Figure 28 Mean AUC across 10 train/test splits among three feature types and clinical variables only model.

	(texture, baseline)	(topology, baseline)	(CNN, baseline)
AUC	1.000	0.173	0.445

* texture: Random Forest
* topology: Logistic Regression
* CNN: Logistic Regression
* baseline: Logistic Regression with clinical variables only (KPS, age, gender)

Table 8 Pair-wise Wilcoxon signed rank test between three image-derived feature models and the baseline. None of the image derived feature models significantly differed from the baseline in mean AUC score.

Features	AUC	Accuracy	Sensitivity	Specificity
texture (Random Forest)	0.707 +/- 0.118	0.609 +/- 0.161	0.950 +/- 0.105	0.575 +/- 0.186
topology (Logistic Regression)	0.855 +/- 0.079	0.820 +/- 0.107	0.925 +/- 0.169	0.810 +/- 0.126
CNN (Logistic Regression)	0.787 +/- 0.195	0.905 +/- 0.102	0.775 +/- 0.249	0.917 +/- 0.121
Clin. Var. only (Logistic Regression)	0.703 +/- 0.256	0.673 +/- 0.206	0.950 +/- 0.158	0.665 +/- 0.272

* Mean value +/- standard deviation
** Sensitivity and Specificity computed with Youden's J index

Table 9 AUC, accuracy, sensitivity, and specificity for each feature type and clinical variables only model. Accuracy, sensitivity, and specificity were computed using Youden's J index.

Surprisingly, none of the image-derived feature models significantly differed from the baseline in their mean AUC scores (p-value > 0.05).

7.4 Feature importance analysis

7.4.1 Modality-wise contributions to important features

For logistic regression (topology and CNN features), the top three negative and the top three positive coefficients (or weights) of the features were considered important. For the random forest (texture feature), the top six features that yielded the highest gini impurity were considered to be important. Appendix C. lists these top six important features across 10 different train/test splits. From this table, we investigated the contributions of each modality in producing the important features (Figure 29).

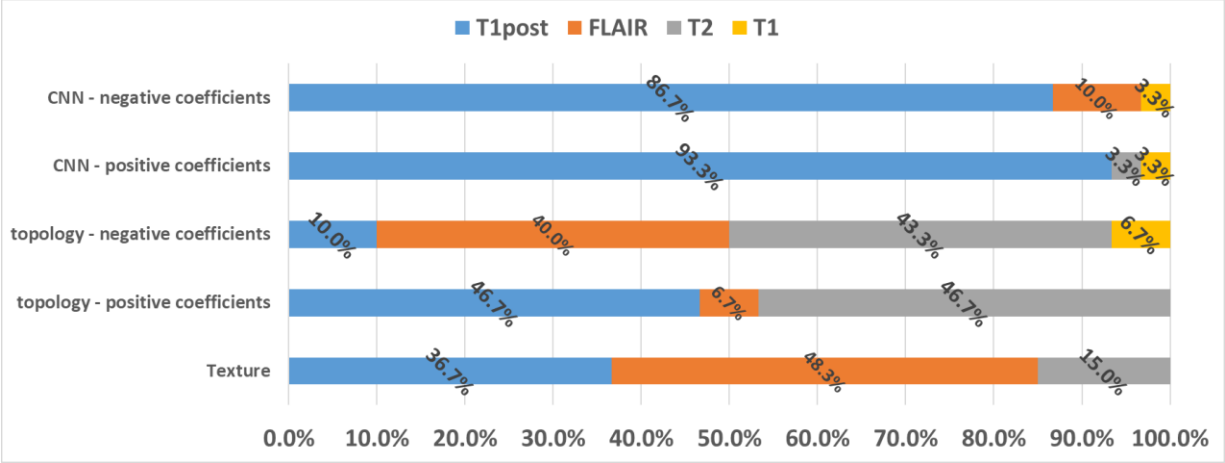


Figure 29 Contributions of each modality in producing top 6 important features

It is noteworthy that for CNN, for both negative and positive coefficients, most of the important features came from the T1post sequence. For topology, the T2 sequence contributes almost half of the top six important features in both positive and negative coefficients. However, the contribution of T1post and FLAIR swaps depending on the "sign" of the coefficients (10% and 40.0% for T1post and FLAIR in negative coefficients compared to 46.7% and 6.7% in positive coefficients). For texture, most of the features came from the T1post and FLAIR sequences (85%). In addition, regardless of the feature types, the T1 sequence contributed either very little (a maximum of 6.7%) or none at all.

7.4.2 Visualizing texture features back onto the image

For texture features, the feature importance was evaluated by the magnitude in the Gini impurity. From Appendix C, for train/test split: 6, Large Area High Gray Level Emphasis from GLSZM for the T1post sequence turned out to be the most important feature. This particular feature yields a high number if there exists a large region with a high gray level intensity. Since this particular feature was deemed to be important, one would find this to be discriminatory between the codeleted and non-codeleted cases. Hence, we picked a tumor patch for codeleted and non-codeleted from a test dataset (from train/test split: 6), and obtained its corresponding value for large area gray level emphasis from the T1 post sequence (Figure 30).

	name	Importance (Gini Impurity)	codel	non-codel
train/test split: 6	glszm_Large Area High Gray Level Emphasis_T1post	0.094	12537.5117	8802.837395
	firstorder_10Percentile_T2	0.091	-0.215062	-0.507398
	firstorder_10Percentile_FLAIR	0.086	-1.304937	-0.104678
	glcm_Imc2_T1post	0.081	0.979217	0.992707
	ngtdm_Coarseness_T2	0.065	0.005571	0.00668
	glszm_Large Area High Gray Level Emphasis_FLAIR	0.058	15628.61786	9717.615768

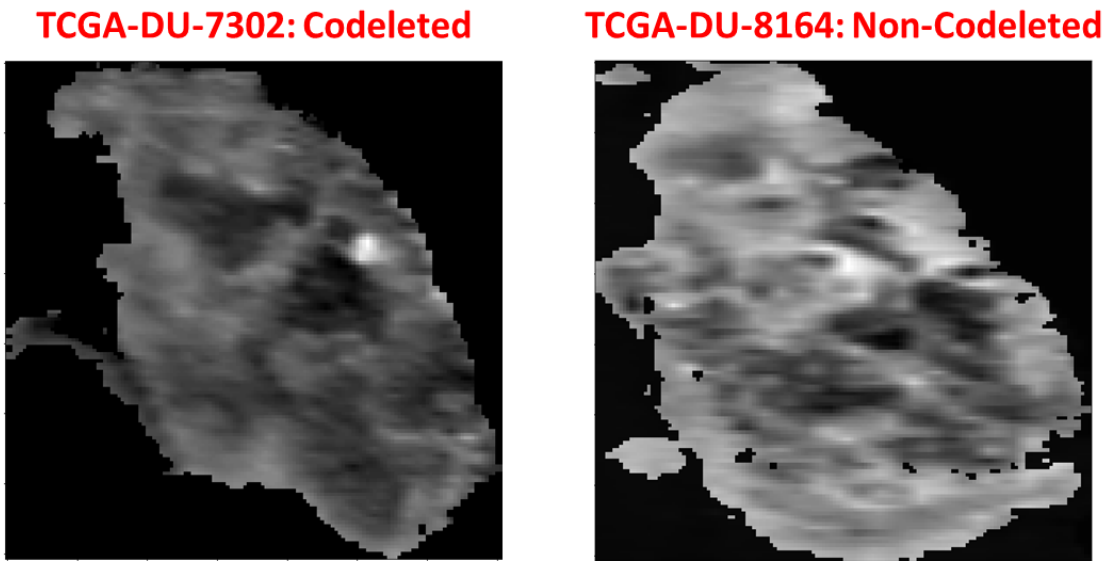


Figure 30 A sample tumor patch for codeleted and non-codleted case. By directly looking at the images and its important feature value, it is hard to understand why this particular feature turned out to be discriminatory between codeleted and non-codeleted.

By directly assessing the images and their corresponding values, it is indeed difficult to comprehend. To the naked eye, it is actually the non-codeleted tumor patch that seems to have more bright regions than does the codeleted tumor patch, but somehow the former obtained a lower value. Also, it is hard to comprehend which region in the image would be discriminatory between the two labels by a visual inspection. Nonetheless, a machine somehow learned a very subtle difference between codeleted and non-codeleted tumor patches and was able to distinguish between the two labels.

7.4.3 Visualizing topology features back onto the image

Unlike texture features, some of the topological features can be actually visualized and *localized* back onto the image. In order to do so, first, we considered the train/test split: 6 case from Appendix C. For this particular split, the median bar length for tunnels (a feature that indicates how long a tunnel persists given the intensity ranges) from theT2 sequence was found to be the most important. It has a positive coefficient, indicating that this particular feature favors the codeleted case. Then, we picked a tumor patch for the codeleted case (from train/test split: 6), obtained the median value of the bar length for tunnels as well as the intensity range over which the bar persisted from the persistence barcode plot. Finally, using the intensity range, we could visualize and localize the birth and death of tunnels back on the actual image (Figure 31).

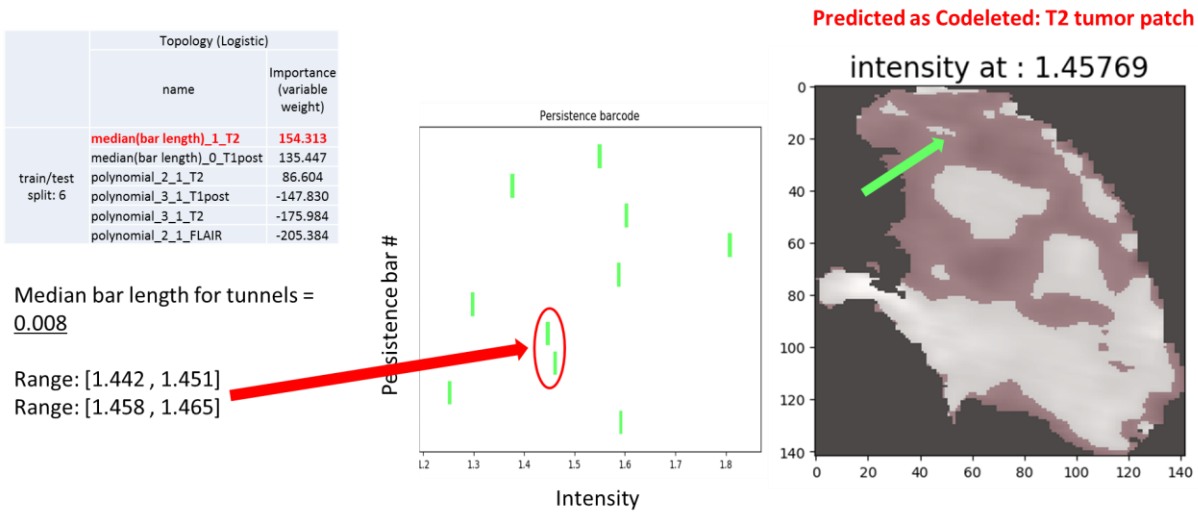


Figure 31 Visualizing topological features back onto the image. First, we identify an important feature (for example, the median bar length for tunnels in the T2 sequence) based on the coefficients of the logistic regression. Then, for a given test tumor patch, we obtained the median bar length for tunnels, traced back to the persistence barcode plot, and then localized on the image.

For instance, in figure 31, for our test tumor patch, the median bar length for tunnels is 0.008. Hence, we identified the persistent bars that yielded the median value of 0.008 from the persistent barcode plot, and obtained the intensity range for those bars. Here, there were two bars (indicated by the red arrow) that yielded a median bar length value of 0.008, and thus

there are two intensity ranges to consider, namely from 1.442 to 1.451 and from 1.458 to 1.465. Then, by sliding through the intensities ranging from 1.442 to 1.465 and saturating the image pixels that meet that intensity range, one could visualize where the tunnels were born and died. Here, the green arrow indicates where those tunnels are.

Chapter 8. Discussion

Pretreatment identification of a patient's 1p/19q codeletion status is important in clinical decision making as it can aid treatment planning on the choice of chemotherapy, radical resection, radiation therapy, etc. In this study, we demonstrated the utility of three distinctive MRI scan features: texture-based, topology-based, and CNN-based, to predict the 1p/19q codeletion status non-invasively using the BRATS dataset (which is based on TCIA) and its corresponding TCGA dataset. To our knowledge, this is the only study that has comprehensively compared three different unique image-derived features in predicting the 1p/19q codeletion status from conventional MR imaging. Moreover, our approach has broad applicability by utilizing only conventional MRI scans, as advanced MR modality sequences, such as DWI, PWI, or functional MRI (fMRI), may not be so readily available.

We did not include WHO grade information or IDH-1/2 mutation status in our prediction model since such information would not be available prior to a histopathological assessment after an invasive surgical biopsy. The whole point of this investigation was to non-invasively predict the 1p/19q codeletion status by analyzing readily available conventional MRI sequences. Further, our study meets the emphasis on molecular genotype from the 2016 WHO classification algorithm of diffuse gliomas (4).

Park et al. reported that the histogram and GLCM features of the apparent diffusion coefficient (ADC) and fractional anisotropy maps achieved an AUC of 0.807 in determining the 1p/19q codeletion status (14). Considering our texture feature-based model, which obtained an AUC of 0.707 +/- 0.118, one might be able to achieve the same level of AUC score by simply analyzing the conventional MRI sequences instead of more advanced ones. Also, we tried to visualize the most important features back on the image. But with the naked eye, differentiating 1p/19q codeleted from non-codeleted was a challenging task purely based on feature numbers (Figure 30).

To the best of our knowledge, this study is the first study to analyze the 1p/19q codeletion status using topological features. Not only did we find that topological features could reliably classify the 1p/19q codeletion status (AUC = 0.855 +/- 0.079), we also demonstrated how to visualize and localize important features back on the image. Such a visualization technique can actually be utilized in clinical settings too because finding such features is very intuitive. One simply has to pay attention where the tunnels or connected components appear and disappear as one slides through the intensity values.

For the CNN features, we could confirm that a pre-trained model for the task of predicting IDH mutation status using conventional MRI scans can also produce features that are relevant to the task of predicting the 1p/19q codeletion status. This finding suggests that there are some deep underlying image-level signals that are shared by both IDH mutation and 1p/19q codeletion. Also, Akkus et al. reported an accuracy of 87.7% in classifying the 1p/19q status using T1-post and T2 images via a CNN (14). Our best performing CNN-based model achieved an AUC of 0.787 +/- 0.195 and an accuracy of 0.905 +/- 0.102, suggesting that our model performed comparably to theirs, which had been trained from the ground up. Considering the amount of time and resources that are needed to train a CNN from scratch, this suggests that it is time- and cost-efficient to simply utilize the off-the-shelf approach by the process known as transfer learning.

It is interesting that almost all of the important features, whether they have negative or positive coefficients, came from the T1post sequence for CNN features (Figure 29). On the other hand, for topological features, the T2 sequence contributed 43.3% and 46.7% for the negative and positive coefficients respectively. Additionally, depending on the “sign” of the coefficients, the contribution of T1post and FLAIR drastically changed. For the features with a negative coefficient, the T1post and FLAIR contributions were 10.0% and 40.0% whereas for those with a positive coefficient, the T1post and FLAIR contributions were 46.7% and 6.7%. Considering that negative coefficients bring the logistic regression probability to 0, i.e., non-codeleted, and positive coefficients bring the probability to 1, i.e., codeleted, such a change in

contribution might be correlated with 1p/19q codeletion status. Additionally, Patel et al. found that the T2-FLAIR mismatch sign is a highly specific biomarker for IDH-mutant, 1p/19q non-codeleted gliomas (64). This finding was further supported by Lasocki et al (65) who reported that the presence of calcifications suggests 1p/19q non-codeleted gliomas. Our findings from topological feature analysis agree with their claims, as T2 and FLAIR sequences contribute significantly (83.3% in total) to the non-codeleted important features.

In this study, we analyzed texture-, topological-, and CNN-based features from conventional MRI sequences: T1, T1post, T2, and FLAIR. Among those three, the topological features predicted the 1p/19q status significantly better than did the texture features. Between the topological features and the CNN features, however, there was no clear difference in the prediction of 1p/19q status in terms of the AUC score. Also, we found that combining all three features did not help our classification of the 1p/19q status. We suspect that as we combined all these features together, we might have overfitted as the number of features increased, but further investigation of this conjecture is needed. In comparison with the baseline model (clinical variables only), none of the image-derived models outperformed the baseline. However, all of the image-derived models had AUC of 0.707 +/- 0.118, 0.855 +/- 0.079, and 0.787 +/- 0.195 (in order of texture, topology, and CNN), suggesting that there exist some underlying image level signals that may be exploited to classify the 1p/19q status.

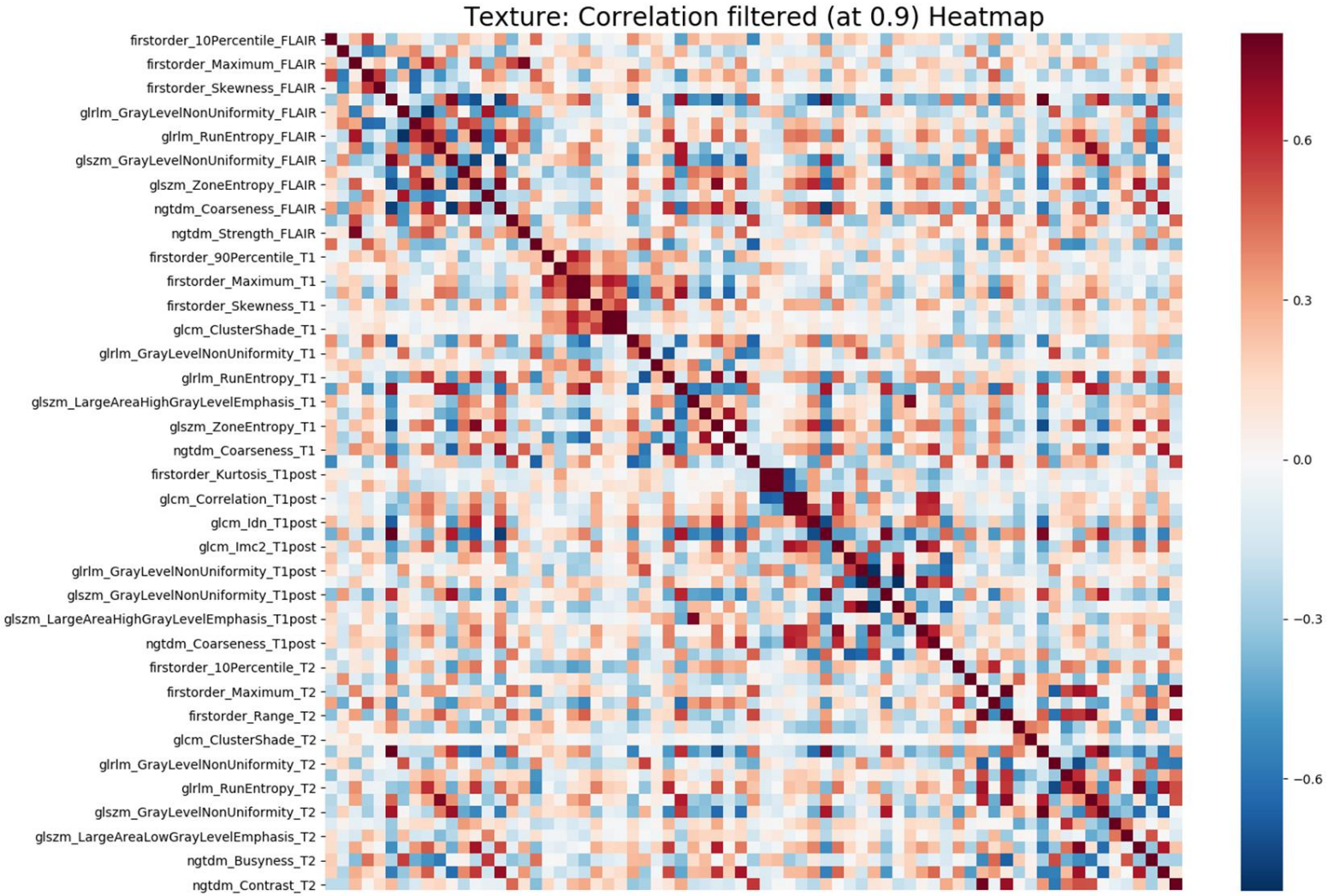
There are several potential improvements to this study that could be incorporated into future work. First, the MRI intensity normalization method could have benefited from more advanced mechanisms, such as, patch based (24), histogram matching (25), white stripe normalization (26). We did not explore the impact of different normalization methods on the prediction of the 1p/19q status, but a future study could assess such implications. Also, we could have extended our study by investigating the relationship between sub-tumor regions (e.g., necrotic and non-enhancing tumor, peritumoral edema, and gadolinium-enhancing tumor) and the 1p/19q codeletion status. Additionally, a small cohort size and a high imbalance between codeletion statuses were limiting factors in our study. We tried to overcome this by oversampling

the minority label by a factor of 10 to balance out the ratio between codeleted and non-codeleted when training, but having more patient data would have helped our modeling process significantly. Lastly, we analyzed our data only in 2D even though 3D image volume data were available. We tried to do 3D analysis for the CNN case by feeding all three planes of the MRI tumor patches as input, but it was not truly a 3D volume image. Hence, extending our study into 3D analysis remains as future work.

Overall, this thesis investigated the utilization of three distinct image features in predicting a patient's 1p/19q codeletion status from standard MR images, and has demonstrated the noninferiority to the analysis of clinical indicators of a method for non-invasive assessment at an earlier stage of gliomas and in follow-ups.

Appendix

Appendix A



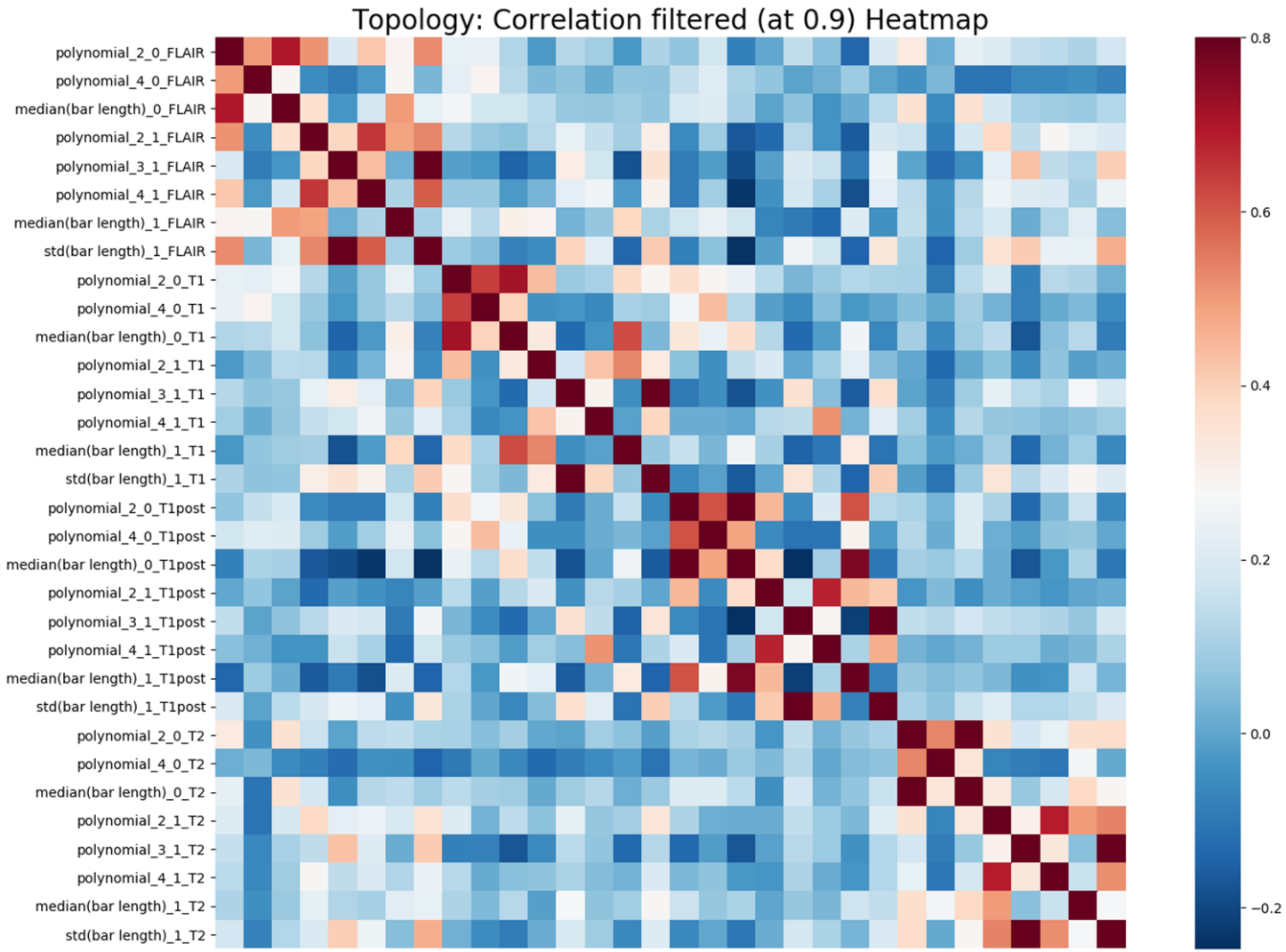


Figure 32 Appendix A figure. A heatmap of spearman correlation filtered features (texture and topology in order). Any features that had a correlation coefficient of greater than absolute value of 0.9 were removed

Appendix B

	train/tesst split: 0	train/tesst split: 1	train/tesst split: 2	train/tesst split: 3	train/tesst split: 4
Texture Feature Names	firstorder_10Percentile_FLAIR	firstorder_10Percentile_FLAIR	firstorder_10Percentile_FLAIR	firstorder_10Percentile_FLAIR	firstorder_10Percentile_FLAIR
	firstorder_Kurtosis_FLAIR	firstorder_Kurtosis_FLAIR	firstorder_Kurtosis_FLAIR	firstorder_Kurtosis_FLAIR	firstorder_Minimum_FLAIR
	firstorder_Skewness_FLAIR	firstorder_Skewness_FLAIR	firstorder_Minimum_FLAIR	firstorder_Minimum_FLAIR	firstorder_Skewness_FLAIR
	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	firstorder_90Percentile_T1	firstorder_Skewness_FLAIR	firstorder_Skewness_FLAIR	ngtdm_Coarseness_FLAIR
	firstorder_90Percentile_T1	glrlm_GrayLevelNonUniformity_T1	glszm_GrayLevelNonUniformity_FLAIR	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	glrlm_RunEntropy_T1
	glszm_LargeAreaHighGrayLevelEmphasis_T1	glszm_LargeAreaHighGrayLevelEmphasis_T1	glrlm_LongRunHighGrayLevelEmphasis_T1	glrlm_LongRunHighGrayLevelEmphasis_T1	ngtdm_Coarseness_T1
	ngtdm_Coarseness_T1	glcm_Imc2_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1	ngtdm_Coarseness_T1	ngtdm_Contrast_T1
	glcm_Imc2_T1post	glszm_GrayLevelNonUniformity_T1post	ngtdm_Coarseness_T1	ngtdm_Contrast_T1	glcm_Correlation_T1post
	glszm_GrayLevelNonUniformity_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1post	glcm_Imc2_T1post	glcm_Imc2_T1post	glcm_Imc2_T1post
	glszm_LargeAreaHighGrayLevelEmphasis_T1post	glszm_ZoneEntropy_T1post	glszm_GrayLevelNonUniformity_T1post	glszm_GrayLevelNonUniformity_T1post	glszm_GrayLevelNonUniformity_T1post
	glszm_ZoneEntropy_T1post	ngtdm_Coarseness_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1post
	ngtdm_Coarseness_T1post	firstorder_10Percentile_T2	ngtdm_Coarseness_T1post	firstorder_10Percentile_T2	firstorder_10Percentile_T2
	firstorder_10Percentile_T2	firstorder_Minimum_T2	firstorder_10Percentile_T2	firstorder_Kurtosis_T2	firstorder_Kurtosis_T2
	firstorder_Kurtosis_T2	glcm_ClusterShade_T2	firstorder_Kurtosis_T2	glcm_ClusterShade_T2	firstorder_Minimum_T2
	firstorder_Minimum_T2	glszm_GrayLevelNonUniformity_T2	firstorder_Minimum_T2	glszm_GrayLevelNonUniformity_T2	firstorder_Skewness_T2
	glszm_LargeAreaHighGrayLevelEmphasis_T2	glszm_LargeAreaHighGrayLevelEmphasis_T2	glcm_ClusterShade_T2	glszm_LargeAreaHighGrayLevelEmphasis_T2	glcm_ClusterShade_T2
	ngtdm_Coarseness_T2	ngtdm_Coarseness_T2	glszm_LargeAreaHighGrayLevelEmphasis_T2	ngtdm_Coarseness_T2	ngtdm_Coarseness_T2

train/tesst split: 5	train/tesst split: 6	train/tesst split: 7	train/tesst split: 8	train/tesst split: 9
firstorder_Minimum_FLAIR	firstorder_10Percentile_FLAIR	firstorder_10Percentile_FLAIR	firstorder_10Percentile_FLAIR	firstorder_10Percentile_FLAIR
firstorder_Skewness_FLAIR	firstorder_Kurtosis_FLAIR	firstorder_Kurtosis_FLAIR	firstorder_Kurtosis_FLAIR	firstorder_Kurtosis_FLAIR
glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	firstorder_Minimum_FLAIR	firstorder_Minimum_FLAIR	firstorder_Minimum_FLAIR	firstorder_Minimum_FLAIR
firstorder_90Percentile_T1	firstorder_Skewness_FLAIR	firstorder_Skewness_FLAIR	firstorder_Skewness_FLAIR	firstorder_Skewness_FLAIR
firstorder_Kurtosis_T1post	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	glszm_GrayLevelNonUniformity_FLAIR	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR
firstorder_Skewness_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1	ngtdm_Coarseness_FLAIR	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	ngtdm_Coarseness_FLAIR
glcm_Correlation_T1post	glcm_Imc2_T1post	firstorder_90Percentile_T1	ngtdm_Coarseness_FLAIR	firstorder_90Percentile_T1
glcm_Idmn_T1post	glszm_GrayLevelNonUniformity_T1post	glrlm_LongRunHighGrayLevelEmphasis_T1	glcm_ClusterShade_T1	glrlm_RunEntropy_T1
glcm_Imc2_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1post	glszm_GrayLevelNonUniformity_T1post	glrlm_LongRunHighGrayLevelEmphasis_T1	glcm_Imc2_T1post
glszm_GrayLevelNonUniformity_T1post	glszm_ZoneEntropy_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1post	glszm_LargeAreaHighGrayLevelEmphasis_T1	glszm_GrayLevelNonUniformity_T1post
glszm_LargeAreaHighGrayLevelEmphasis_T1post	ngtdm_Coarseness_T1post	glszm_ZoneEntropy_T1post	glcm_Imc2_T1post	glszm_ZoneEntropy_T1post
glszm_ZoneEntropy_T1post	firstorder_10Percentile_T2	firstorder_10Percentile_T2	glszm_GrayLevelNonUniformity_T1post	ngtdm_Coarseness_T1post
ngtdm_Coarseness_T1post	firstorder_Kurtosis_T2	firstorder_Kurtosis_T2	glszm_LargeAreaHighGrayLevelEmphasis_T1post	firstorder_Minimum_T2
firstorder_10Percentile_T2	firstorder_Minimum_T2	firstorder_Minimum_T2	ngtdm_Coarseness_T1post	firstorder_Skewness_T2
glcm_ClusterShade_T2	glcm_ClusterShade_T2	firstorder_Skewness_T2	glcm_ClusterShade_T2	glcm_ClusterShade_T2
glszm_LargeAreaHighGrayLevelEmphasis_T2	glszm_LargeAreaHighGrayLevelEmphasis_T2	glcm_ClusterShade_T2	glszm_LargeAreaHighGrayLevelEmphasis_T2	glszm_GrayLevelNonUniformity_T2
ngtdm_Coarseness_T2	ngtdm_Coarseness_T2	ngtdm_Coarseness_T2	glszm_ZoneEntropy_T2	glszm_ZoneEntropy_T2

	train/tesst split: 0	train/tesst split: 1	train/tesst split: 2	train/tesst split: 3	train/tesst split: 4
Topology Feature Names	polynomial_4_0_FLAIR	polynomial_3_0_FLAIR	polynomial_2_0_FLAIR	polynomial_4_0_FLAIR	polynomial_4_0_FLAIR
	polynomial_2_1_FLAIR	polynomial_4_0_FLAIR	polynomial_4_0_FLAIR	polynomial_2_1_FLAIR	polynomial_2_1_FLAIR
	polynomial_3_1_FLAIR	std(bar length)_0_FLAIR	polynomial_2_1_FLAIR	polynomial_3_1_FLAIR	polynomial_3_1_FLAIR
	polynomial_4_1_FLAIR	polynomial_3_1_FLAIR	polynomial_3_1_FLAIR	polynomial_4_1_FLAIR	median(bar length)_1_FLAIR
	median(bar length)_1_FLAIR	polynomial_4_1_FLAIR	polynomial_4_1_FLAIR	std(bar length)_1_FLAIR	std(bar length)_1_FLAIR
	polynomial_2_0_T1	std(bar length)_1_FLAIR	median(bar length)_1_FLAIR	polynomial_4_0_T1	median(bar length)_0_T1
	polynomial_4_0_T1	polynomial_2_0_T1	std(bar length)_1_FLAIR	median(bar length)_0_T1	polynomial_3_1_T1
	polynomial_2_1_T1	polynomial_4_0_T1	polynomial_2_1_T1	polynomial_2_1_T1	median(bar length)_1_T1
	polynomial_3_1_T1	median(bar length)_1_T1	polynomial_3_1_T1	polynomial_3_1_T1	std(bar length)_1_T1
	median(bar length)_1_T1	std(bar length)_1_T1	median(bar length)_1_T1	median(bar length)_1_T1	polynomial_2_0_T1post
	std(bar length)_1_T1	polynomial_2_0_T1post	std(bar length)_1_T1	std(bar length)_1_T1	polynomial_4_0_T1post
	polynomial_2_0_T1post	polynomial_4_0_T1post	polynomial_2_0_T1post	polynomial_2_0_T1post	median(bar length)_0_T1post
	median(bar length)_0_T1post	median(bar length)_0_T1post	polynomial_4_0_T1post	polynomial_4_0_T1post	polynomial_2_1_T1post
	polynomial_3_1_T1post	polynomial_2_1_T1post	median(bar length)_0_T1post	median(bar length)_0_T1post	median(bar length)_1_T1post
	median(bar length)_1_T1post	polynomial_4_1_T1post	median(bar length)_1_T1post	polynomial_2_1_T1post	std(bar length)_1_T1post
	std(bar length)_1_T1post	median(bar length)_1_T1post	std(bar length)_1_T1post	std(bar length)_1_T1post	polynomial_2_0_T2
	polynomial_4_0_T2	polynomial_2_0_T2	polynomial_2_0_T2	polynomial_2_0_T2	polynomial_4_0_T2
	median(bar length)_0_T2	polynomial_2_1_T2	polynomial_4_0_T2	polynomial_4_0_T2	polynomial_2_1_T2
	polynomial_2_1_T2	polynomial_3_1_T2	polynomial_2_1_T2	median(bar length)_0_T2	polynomial_3_1_T2
	polynomial_3_1_T2	polynomial_4_1_T2	polynomial_3_1_T2	polynomial_2_1_T2	polynomial_4_1_T2
	median(bar length)_1_T2	median(bar length)_1_T2	median(bar length)_1_T2	polynomial_3_1_T2	median(bar length)_1_T2
	std(bar length)_1_T2	std(bar length)_1_T2	std(bar length)_1_T2	median(bar length)_1_T2	std(bar length)_1_T2

train/tesst split: 5	train/tesst split: 6	train/tesst split: 7	train/tesst split: 8	train/tesst split: 9
polynomial_2_0_FLAIR	polynomial_4_0_FLAIR	polynomial_4_0_FLAIR	polynomial_2_0_FLAIR	polynomial_2_0_FLAIR
polynomial_4_0_FLAIR	polynomial_2_1_FLAIR	polynomial_2_1_FLAIR	polynomial_4_0_FLAIR	polynomial_3_0_FLAIR
polynomial_2_1_FLAIR	polynomial_3_1_FLAIR	polynomial_3_1_FLAIR	polynomial_2_1_FLAIR	std(bar length)_0_FLAIR
polynomial_3_1_FLAIR	polynomial_4_1_FLAIR	polynomial_4_1_FLAIR	polynomial_3_1_FLAIR	polynomial_2_1_FLAIR
polynomial_4_1_FLAIR	median(bar length)_1_FLAIR	median(bar length)_1_FLAIR	polynomial_4_1_FLAIR	polynomial_3_1_FLAIR
std(bar length)_1_FLAIR	std(bar length)_1_FLAIR	std(bar length)_1_FLAIR	std(bar length)_1_FLAIR	polynomial_4_1_FLAIR
polynomial_2_0_T1	polynomial_2_0_T1	polynomial_2_0_T1	polynomial_4_0_T1	std(bar length)_1_FLAIR
median(bar length)_0_T1	median(bar length)_0_T1	polynomial_3_1_T1	polynomial_2_1_T1	polynomial_3_1_T1
polynomial_3_1_T1	polynomial_3_1_T1	std(bar length)_1_T1	polynomial_3_1_T1	median(bar length)_1_T1
median(bar length)_1_T1	std(bar length)_1_T1	polynomial_2_0_T1post	median(bar length)_1_T1	std(bar length)_1_T1
std(bar length)_1_T1	polynomial_2_0_T1post	polynomial_4_0_T1post	std(bar length)_1_T1	polynomial_2_0_T1post
polynomial_2_0_T1post	polynomial_4_0_T1post	median(bar length)_0_T1post	polynomial_2_0_T1post	polynomial_4_0_T1post
polynomial_4_0_T1post	median(bar length)_0_T1post	polynomial_2_1_T1post	polynomial_4_0_T1post	median(bar length)_0_T1post
median(bar length)_0_T1post	polynomial_2_1_T1post	polynomial_4_1_T1post	median(bar length)_0_T1post	polynomial_4_1_T1post
polynomial_2_1_T1post	polynomial_3_1_T1post	median(bar length)_1_T1post	polynomial_2_1_T1post	median(bar length)_1_T1post
median(bar length)_1_T1post	polynomial_4_1_T1post	std(bar length)_1_T1post	polynomial_3_1_T1post	std(bar length)_1_T1post
polynomial_2_0_T2	median(bar length)_1_T1post	polynomial_2_0_T2	median(bar length)_1_T1post	polynomial_2_0_T2
polynomial_2_1_T2	polynomial_2_0_T2	median(bar length)_0_T2	polynomial_2_0_T2	polynomial_2_1_T2
polynomial_3_1_T2	polynomial_2_1_T2	polynomial_2_1_T2	polynomial_4_0_T2	polynomial_3_1_T2
polynomial_4_1_T2	polynomial_3_1_T2	polynomial_3_1_T2	polynomial_2_1_T2	polynomial_4_1_T2
median(bar length)_1_T2	median(bar length)_1_T2	median(bar length)_1_T2	median(bar length)_1_T2	median(bar length)_1_T2
std(bar length)_1_T2	std(bar length)_1_T2	std(bar length)_1_T2	std(bar length)_1_T2	std(bar length)_1_T2

Table 10 Appendix B Table. A table that lists out the RFE selected features for each train/test split. Note that only texture and topology features were filtered via RFE method as there was no precedent work that has been applied for CNN features.

Appendix C

	Texture (RF)		Topology (Logistic)		CNN (Logistic)	
	name	Importance (Gini Impurity)	name	Importance (variable weight)	name	Importance (variable weight)
train/test split: 0	glcm_Imc2_T1post	0.109	median(bar length)_1_T2	177.177	CNN_feature_445_T1post	4.107
	glszm_GrayLevelNonUniformity_T1post	0.103	median(bar length)_1_T1post	122.000	CNN_feature_474_T1post	2.866
	firstorder_Skewness_FLAIR	0.102	polynomial_2_1_T2	110.683	CNN_feature_256_T2	1.842
	firstorder_Kurtosis_FLAIR	0.096	polynomial_3_1_T2	-368.371	CNN_feature_203_T1post	-1.904
	glszm_LargeAreaHighGrayLevelEmphasis_T1post	0.089	polynomial_3_1_FLAIR	-431.239	CNN_feature_118_T1post	-3.031
	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	0.066	polynomial_2_1_FLAIR	-473.890	CNN_feature_25_T1post	-3.112
train/test split: 1	glszm_GrayLevelNonUniformity_T1post	0.129	median(bar length)_0_T1post	71.906	CNN_feature_445_T1post	5.122
	glcm_Imc2_T1post	0.091	polynomial_2_0_T1post	64.799	CNN_feature_142_T1post	2.759
	firstorder_10Percentile_FLAIR	0.077	polynomial_2_1_T2	64.490	CNN_feature_474_T1post	2.307
	firstorder_Skewness_FLAIR	0.071	median(bar length)_1_T1	-33.262	CNN_feature_157_T1post	-2.250
	firstorder_Kurtosis_FLAIR	0.071	polynomial_2_0_T2	-60.670	CNN_feature_118_T1post	-2.834
	glszm_LargeAreaHighGrayLevelEmphasis_T1post	0.070	std(bar length)_1_FLAIR	-64.507	CNN_feature_25_T1post	-3.949
train/test split: 2	firstorder_10Percentile_FLAIR	0.125	median(bar length)_1_T2	133.858	CNN_feature_445_T1post	4.241
	glszm_GrayLevelNonUniformity_T1post	0.098	polynomial_2_1_T2	129.330	CNN_feature_142_T1post	2.612
	glcm_Imc2_T1post	0.079	median(bar length)_0_T1post	122.065	CNN_feature_474_T1post	2.344
	glszm_LargeAreaHighGrayLevelEmphasis_T1post	0.072	polynomial_2_0_T2	-117.942	CNN_feature_411_T1post	-1.520
	firstorder_10Percentile_T2	0.060	polynomial_2_1_FLAIR	-171.795	CNN_feature_157_T1post	-1.639
	firstorder_Skewness_FLAIR	0.056	polynomial_3_1_T2	-218.561	CNN_feature_25_T1post	-2.817
train/test split: 3	firstorder_10Percentile_FLAIR	0.099	median(bar length)_0_T1post	87.566	CNN_feature_474_T1post	4.028
	firstorder_Skewness_FLAIR	0.085	polynomial_2_0_T1post	71.439	CNN_feature_445_T1post	3.408
	firstorder_Kurtosis_T2	0.078	median(bar length)_1_T2	69.237	CNN_feature_251_T1post	2.298
	glszm_GrayLevelNonUniformity_T1post	0.072	polynomial_2_1_FLAIR	-32.113	CNN_feature_25_T1post	-1.819
	glcm_Imc2_T1post	0.072	std(bar length)_1_FLAIR	-50.021	CNN_feature_83_T1post	-2.209
	firstorder_Minimum_FLAIR	0.067	polynomial_2_0_T2	-61.505	CNN_feature_411_T1post	-3.187
train/test split: 4	firstorder_10Percentile_FLAIR	0.101	polynomial_4_0_FLAIR	3250.809	CNN_feature_445_T1post	4.146
	glszm_GrayLevelNonUniformity_T1post	0.099	polynomial_3_1_FLAIR	497.213	CNN_feature_474_T1post	3.158
	ngtdm_Coarseness_FLAIR	0.091	median(bar length)_1_T2	189.798	CNN_feature_196_T1post	1.763
	firstorder_Minimum_FLAIR	0.084	polynomial_3_1_T1	-564.474	CNN_feature_181_FLAIR	-1.248
	firstorder_Skewness_FLAIR	0.084	polynomial_3_1_T2	-1540.448	CNN_feature_457_T1post	-1.248
	firstorder_Kurtosis_T2	0.074	polynomial_4_0_T1post	-6952.187	CNN_feature_25_T1post	-2.341

train/test split: 5	firstorder_10Percentile_T2	0.083	median(bar length)_1_T2	143.913	CNN_feature_445_T1post	4.019
	glszm_LargeAreaHighGrayLevelEmphasis_T1post	0.081	polynomial_2_1_T2	130.548	CNN_feature_474_T1post	2.540
	firstorder_Minimum_FLAIR	0.081	median(bar length)_0_T1post	116.540	CNN_feature_196_T1post	2.286
	firstorder_Skewness_FLAIR	0.077	polynomial_2_0_T2	-125.704	CNN_feature_181_FLAIR	-0.977
	firstorder_Skewness_T1post	0.069	polynomial_2_1_FLAIR	-297.064	CNN_feature_213_T1	-1.672
	glcm_ClusterShade_T2	0.069	polynomial_3_1_T2	-547.659	CNN_feature_25_T1post	-3.724
train/test split: 6	glszm_LargeAreaHighGrayLevelEmphasis_T1post	0.094	median(bar length)_1_T2	154.313	CNN_feature_474_T1post	1.110
	firstorder_10Percentile_T2	0.091	median(bar length)_0_T1post	135.447	CNN_feature_153_T1post	1.049
	firstorder_10Percentile_FLAIR	0.086	polynomial_2_1_T2	86.604	CNN_feature_263_T1post	1.029
	glcm_Imc2_T1post	0.081	polynomial_3_1_T1post	-147.830	CNN_feature_110_T1post	-0.965
	ngtdm_Coarseness_T2	0.065	polynomial_3_1_T2	-175.984	CNN_feature_118_T1post	-1.046
	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	0.058	polynomial_2_1_FLAIR	-205.384	CNN_feature_411_T1post	-1.175
train/test split: 7	glszm_LargeAreaHighGrayLevelEmphasis_T1post	0.104	median(bar length)_1_T2	151.954	CNN_feature_474_T1post	1.160
	glszm_GrayLevelNonUniformity_T1post	0.102	polynomial_2_1_T2	126.964	CNN_feature_153_T1post	1.124
	ngtdm_Coarseness_FLAIR	0.096	median(bar length)_0_T1post	89.157	CNN_feature_251_T1post	1.066
	firstorder_10Percentile_T2	0.075	polynomial_2_0_T2	-74.813	CNN_feature_110_T1post	-0.922
	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	0.069	polynomial_3_1_T2	-189.826	CNN_feature_63_T1post	-0.969
	glszm_ZoneEntropy_T1post	0.064	polynomial_2_1_FLAIR	-193.277	CNN_feature_411_T1post	-1.297
train/test split: 8	ngtdm_Coarseness_FLAIR	0.116	median(bar length)_1_T1post	69.456	CNN_feature_445_T1post	1.299
	glszm_LargeAreaHighGrayLevelEmphasis_T1post	0.088	median(bar length)_0_T1post	68.937	CNN_feature_474_T1post	1.104
	glcm_Imc2_T1post	0.081	median(bar length)_1_T2	63.820	CNN_feature_251_T1post	1.051
	firstorder_Skewness_FLAIR	0.080	polynomial_3_1_T1post	-45.060	CNN_feature_157_T1post	-1.001
	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	0.071	polynomial_2_0_T2	-46.818	CNN_feature_118_T1post	-1.041
	firstorder_10Percentile_FLAIR	0.067	polynomial_2_1_FLAIR	-58.550	CNN_feature_411_T1post	-1.084
train/test split: 9	ngtdm_Coarseness_FLAIR	0.117	median(bar length)_0_T1post	62.584	CNN_feature_445_T1post	5.900
	firstorder_Skewness_FLAIR	0.114	median(bar length)_1_T1post	62.200	CNN_feature_474_T1post	2.899
	glcm_ClusterShade_T2	0.071	polynomial_2_0_T1post	59.973	CNN_feature_206_T1	2.448
	firstorder_Minimum_FLAIR	0.070	polynomial_2_1_FLAIR	-42.167	CNN_feature_190_T1post	-2.563
	glszm_LargeAreaHighGrayLevelEmphasis_FLAIR	0.069	polynomial_2_0_T2	-46.629	CNN_feature_181_FLAIR	-2.866
	glcm_Imc2_T1post	0.067	std(bar length)_1_FLAIR	-51.703	CNN_feature_25_T1post	-4.858

Table 11 Appendix C Table. A table that lists top 6 important features for each feature type. Note that for topology and CNN features, top three negative and top three positive features were selected as their models were logistic regressions. For texture features, a random forest was used, hence top 6 gini impurity features were selected.

References

1. Chaichana, K. L., M. J. McGirt, J. Laterra, A. Olivi, and A. Quiñones-Hinojosa. 2010. Recurrence and malignant degeneration after resection of adult hemispheric low-grade gliomas. *J. Neurosurg.* 112: 10–17.

2. Sakarunchai, I., R. Sangthong, N. Phuenpathom, and M. Phukaoloun. 2013. Free survival time of recurrence and malignant transformation and associated factors in patients with supratentorial low-grade gliomas. *J. Med. Assoc. Thai.* 96: 1542–9.

3. Ostrom, Q. T., H. Gittleman, P. Liao, T. Vecchione-Koval, Y. Wolinsky, C. Kruchko, and J. S. Barnholtz-Sloan. 2017. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro. Oncol.* 19: v1–v88.

4. Louis, D. N., A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison. 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* 131: 803–820.

5. Cancer Genome Atlas Research Network, D. J. Brat, R. G. W. Verhaak, K. D. Aldape, W. K. A. Yung, S. R. Salama, L. A. D. Cooper, E. Rheinbay, C. R. Miller, M. Vitucci, O. Morozova, A. G. Robertson, H. Noushmehr, P. W. Laird, A. D. Cherniack, R. Akbani, J. T. Huse, G. Ciriello, L. M. Poisson, J. S. Barnholtz-Sloan, M. S. Berger, C. Brennan, R. R. Colen, H. Colman, A. E. Flanders, C. Giannini, M. Grifford, A. Iavarone, R. Jain, I. Joseph, J. Kim, K. Kasaian, T. Mikkelsen, B. A. Murray, B. P. O'Neill, L. Pachter, D. W. Parsons, C. Sougnez, E. P. Sulman, S. R. Vandenberg, E. G. Van Meir, A. von Deimling, H. Zhang, D. Crain, K. Lau, D. Mallery, S. Morris, J. Paulauskis, R. Penny, T. Shelton, M. Sherman, P. Yena, A. Black, J. Bowen, K. Dicostanzo, J. Gastier-Foster, K. M. Leraas, T. M. Lichtenberg, C. R. Pierson, N. C. Ramirez, C. Taylor, S. Weaver, L. Wise, E. Zmuda, T. Davidsen, J. A. Demchok, G. Eley, M. L. Ferguson, C. M. Hutter, K. R. Mills Shaw, B. A. Ozenberger, M. Sheth, H. J. Sofia, R.

Tarnuzzer, Z. Wang, L. Yang, J. C. Zenklusen, B. Ayala, J. Baboud, S. Chudamani, M. A. Jensen, J. Liu, T. Pihl, R. Raman, Y. Wan, Y. Wu, A. Ally, J. T. Auman, M. Balasundaram, S. Balu, S. B. Baylin, R. Beroukhim, M. S. Bootwalla, R. Bowlby, C. A. Bristow, D. Brooks, Y. Butterfield, R. Carlsen, S. Carter, L. Chin, A. Chu, E. Chuah, K. Cibulskis, A. Clarke, S. G. Coetzee, N. Dhalla, T. Fennell, S. Fisher, S. Gabriel, G. Getz, R. Gibbs, R. Guin, A. Hadjipanayis, D. N. Hayes, T. Hinoue, K. Hoadley, R. A. Holt, A. P. Hoyle, S. R. Jefferys, S. Jones, C. D. Jones, R. Kucherlapati, P. H. Lai, E. Lander, S. Lee, L. Lichtenstein, Y. Ma, D. T. Maglinte, H. S. Mahadeshwar, M. A. Marra, M. Mayo, S. Meng, M. L. Meyerson, P. A. Mieczkowski, R. A. Moore, L. E. Mose, A. J. Mungall, A. Pantazi, M. Parfenov, P. J. Park, J. S. Parker, C. M. Perou, A. Protopopov, X. Ren, J. Roach, T. S. Sabedot, J. Schein, S. E. Schumacher, J. G. Seidman, S. Seth, H. Shen, J. V Simons, P. Sipahimalani, M. G. Soloway, X. Song, H. Sun, B. Tabak, A. Tam, D. Tan, J. Tang, N. Thiessen, T. Triche, D. J. Van Den Berg, U. Veluvolu, S. Waring, D. J. Weisenberger, M. D. Wilkerson, T. Wong, J. Wu, L. Xi, A. W. Xu, L. Yang, T. I. Zack, J. Zhang, B. A. Aksoy, H. Arachchi, C. Benz, B. Bernard, D. Carlin, J. Cho, D. DiCara, S. Frazer, G. N. Fuller, J. Gao, N. Gehlenborg, D. Haussler, D. I. Heiman, L. Iype, A. Jacobsen, Z. Ju, S. Katzman, H. Kim, T. Knijnenburg, R. B. Kreisberg, M. S. Lawrence, W. Lee, K. Leinonen, P. Lin, S. Ling, W. Liu, Y. Liu, Y. Liu, Y. Lu, G. Mills, S. Ng, M. S. Noble, E. Paull, A. Rao, S. Reynolds, G. Saksena, Z. Sanborn, C. Sander, N. Schultz, Y. Senbabaoglu, R. Shen, I. Shmulevich, R. Sinha, J. Stuart, S. O. Sumer, Y. Sun, N. Tasman, B. S. Taylor, D. Voet, N. Weinhold, J. N. Weinstein, D. Yang, K. Yoshihara, S. Zheng, W. Zhang, L. Zou, T. Abel, S. Sadeghi, M. L. Cohen, J. Eschbacher, E. M. Hattab, A. Raghunathan, M. J. Schniederjan, D. Aziz, G. Barnett, W. Barrett, D. D. Bigner, L. Boice, C. Brewer, C. Calatozzolo, B. Campos, C. G. Carlotti, T. A. Chan, L. Cuppini, E. Curley, S. Cuzzubbo, K. Devine, F. DiMeco, R. Duell, J. B. Elder, A. Fehrenbach, G. Finocchiaro, W. Friedman, J. Fulop, J. Gardner, B. Hermes, C. Herold-Mende, C. Jungk, A. Kendler, N. L. Lehman, E. Lipp, O. Liu, R. Mandt, M. McGraw, R. Mclendon, C. McPherson, L. Neder, P. Nguyen, A. Noss, R. Nunziata, Q. T. Ostrom, C. Palmer, A. Perin, B. Pollo, A. Potapov, O. Potapova, W. K.

Rathmell, D. Rotin, L. Scarpace, C. Schilero, K. Senecal, K. Shimmel, V. Shurkhay, S. Sifri, R. Singh, A. E. Sloan, K. Smolenski, S. M. Staugaitis, R. Steele, L. Thorne, D. P. C. Tirapelli, A. Unterberg, M. Vallurupalli, Y. Wang, R. Warnick, F. Williams, Y. Wolinsky, S. Bell, M. Rosenberg, C. Stewart, F. Huang, J. L. Grimsby, A. J. Radenbaugh, and J. Zhang. 2015. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* 372: 2481–2498.

6. Eckel-Passow, J. E., D. H. Lachance, A. M. Molinaro, K. M. Walsh, P. A. Decker, H. Sicotte, M. Pekmezci, T. Rice, M. L. Kosel, I. V. Smirnov, G. Sarkar, A. A. Caron, T. M. Kollmeyer, C. E. Praska, A. R. Chada, C. Halder, H. M. Hansen, L. S. McCoy, P. M. Bracci, R. Marshall, S. Zheng, G. F. Reis, A. R. Pico, B. P. O'Neill, J. C. Buckner, C. Giannini, J. T. Huse, A. Perry, T. Tihan, M. S. Berger, S. M. Chang, M. D. Prados, J. Wiemels, J. K. Wiencke, M. R. Wrensch, and R. B. Jenkins. 2015. Glioma Groups Based on 1p/19q, *IDH* , and *TERT* Promoter Mutations in Tumors. *N. Engl. J. Med.* 372: 2499–2508.

7. Cairncross, G., B. Berkey, E. Shaw, R. Jenkins, B. Scheithauer, D. Brachman, J. Buckner, K. Fink, L. Souhami, N. Laperierre, M. Mehta, W. Curran, and W. Curran. 2006. Phase III Trial of Chemotherapy Plus Radiotherapy Compared With Radiotherapy Alone for Pure and Mixed Anaplastic Oligodendroglioma: Intergroup Radiation Therapy Oncology Group Trial 9402. *J. Clin. Oncol.* 24: 2707–2714.

8. van den Bent, M. J., A. F. Carpentier, A. A. Brandes, M. Sanson, M. J. B. Taphoorn, H. J. J. A. Bernsen, M. Frenay, C. C. Tijssen, W. Grisold, L. Sipos, H. Haaxma-Reiche, J. M. Kros, M. C. M. van Kouwenhoven, C. J. Vecht, A. Allgeier, D. Lacombe, and T. Gorlia. 2006. Adjuvant Procarbazine, Lomustine, and Vincristine Improves Progression-Free Survival but Not Overall Survival in Newly Diagnosed Anaplastic Oligodendrogliomas and Oligoastrocytomas: A Randomized European Organisation for Research and Treatment of Cancer Phase III Trial. *J. Clin. Oncol.* 24: 2715–2722.

9. Zhao, J., W. Ma, and H. Zhao. 2014. Loss of heterozygosity 1p/19q and survival in glioma: a meta-analysis. *Neuro. Oncol.* 16: 103–112.

10. Jenkins, R. B., H. Blair, K. V. Ballman, C. Giannini, R. M. Arusell, M. Law, H. Flynn, S. Passe, S. Felten, P. D. Brown, E. G. Shaw, and J. C. Buckner. 2006. A t(1;19)(q10;p10) Mediates the Combined Deletions of 1p and 19q and Predicts a Better Prognosis of Patients with Oligodendroglioma. *Cancer Res.* 66: 9852–9861.

11. Woehrer, A., P. Sander, C. Haberler, S. Kern, H. Maier, M. Preusser, C. Hartmann, J. M. Kros, J. A. Hainfellner, and Research Committee of the European Confederation of Neuropathological Societies. FISH-based detection of 1p 19q codeletion in oligodendroglial tumors: procedures and protocols for neuropathological practice - a publication under the auspices of the Research Committee of the European Confederation of Neuropathological Societies (Euro-CNS). *Clin. Neuropathol.* 30: 47–55.

12. Jansen, N. L., C. Schwartz, V. Graute, S. Eigenbrod, J. Lutz, R. Egensperger, G. Pöpperl, H. A. Kretzschmar, P. Cumming, P. Bartenstein, J.-C. Tonn, F.-W. Kreth, C. la Fougère, and N. Thon. 2012. Prediction of oligodendroglial histology and LOH 1p/19q using dynamic [(18)F]FET-PET imaging in intracranial WHO grade II and III gliomas. *Neuro. Oncol.* 14: 1473–80.

13. Fellah, S., D. Caudal, A. M. De Paula, P. Dory-Lautrec, D. Figarella-Branger, O. Chinot, P. Metellus, P. J. Cozzone, S. Confort-Gouny, B. Ghattas, V. Callot, and N. Girard. 2013. Multimodal MR Imaging (Diffusion, Perfusion, and Spectroscopy): Is It Possible to Distinguish Oligodendroglial Tumor Grade and 1p/19q Codeletion in the Pretherapeutic Diagnosis? *Am. J. Neuroradiol.* 34: 1326–1333.

14. Park, Y. W., K. Han, S. S. Ahn, Y. S. Choi, J. H. Chang, S. H. Kim, S.-G. Kang, E. H. Kim, and S.-K. Lee. 2018. Whole-Tumor Histogram and Texture Analyses of DTI for Evaluation of IDH1-Mutation and 1p/19q-Codeletion Status in World Health Organization Grade II Gliomas.

AJNR. Am. J. Neuroradiol. 39: 693–698.

15. Akkus, Z., I. Ali, J. Sedlář, J. P. Agrawal, I. F. Parney, C. Giannini, and B. J. Erickson. 2017. Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence. *J. Digit. Imaging* 30: 469–476.

16. Bakas, S., H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos. 2017. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* 4: 170117.

17. Menze, B. H., A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput. 2015. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* 34: 1993–2024.

18. Clark, K., B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior. 2013. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* 26: 1045–1057.

19. Belaroussi, B., J. Milles, S. Carme, Y. M. Zhu, and H. Benoit-Cattin. 2006. Intensity non-uniformity correction in MRI: Existing methods and their validation. *Med. Image Anal.* 10: 234–246.

20. Vovk, U., F. Pernus, and B. Likar. 2007. A Review of Methods for Correction of Intensity

Inhomogeneity in MRI. *IEEE Trans. Med. Imaging* 26: 405–421.

21. Hanson', L. G., and T. Dyrby2. 2002. RF inhomogeneity correction: Validity of the smooth-bias approximation. *Int. Soc. Mag. Reson. Med. (ISMRM 2002)* 2316.

22. Tustison, N. J., B. B. Avants, P. A. Cook, Y. Yuanjie Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* 29: 1310–1320.

23. Gorgolewski, K., C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. S. Ghosh. 2011. Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Front. Neuroinform.* 5: 13.

24. Roy, S., A. Carass, and J. L. Prince. 2013. Patch based intensity normalization of brain MR images. In *Proceedings - International Symposium on Biomedical Imaging* IEEE. 342–345.

25. Sun, X., L. Shi, Y. Luo, W. Yang, H. Li, P. Liang, K. Li, V. C. T. Mok, W. C. W. Chu, and D. Wang. 2015. Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. *Biomed. Eng. Online* 14: 73.

26. Shinohara, R. T., E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, C. M. Crainiceanu, Australian Imaging Biomarkers Lifestyle Flagship Study of Ageing, and Alzheimer's Disease Neuroimaging Initiative. 2014. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage Clin.* 6: 9–19.

27. van Griethuysen, J. J. M., A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. W. L. Aerts. 2017. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 77: e104–e107.

28. Cheng, N.-M., Y.-H. Dean Fang, J. Tung-Chieh Chang, C.-G. Huang, D.-L. Tsan, S.-H. Ng, H.-M. Wang, C.-Y. Lin, C.-T. Liao, and T.-C. Yen. 2013. Textural Features of Pretreatment 18F-FDG PET/CT Images: Prognostic Significance in Patients with Advanced T-Stage

Oropharyngeal Squamous Cell Carcinoma. *J. Nucl. Med.* 54: 1703–1709.

29. Miles, K. A., B. Ganeshan, M. R. Griffiths, R. C. D. Young, and C. R. Chatwin. 2009. Colorectal Cancer: Texture Analysis of Portal Phase Hepatic CT Images as a Potential Marker of Survival. *Radiology* 250: 444–452.

30. Ghosh, P., P. Tamboli, R. Vikram, and A. Rao. 2015. Imaging-genomic pipeline for identifying gene mutations using three-dimensional intra-tumor heterogeneity features. *J. Med. imaging (Bellingham, Wash.)* 2: 041009.

31. Grimm, L. J., J. Zhang, and M. A. Mazurowski. 2015. Computational approach to radiogenomics of breast cancer: Luminal A and luminal B molecular subtypes are associated with imaging features on routine breast MRI extracted using computer vision algorithms. *J. Magn. Reson. Imaging* 42: 902–907.

32. Parmar, C., R. T. H. Leijenaar, P. Grossmann, E. Rios Velazquez, J. Bussink, D. Rietveld, M. M. Rietbergen, B. Haibe-Kains, P. Lambin, and H. J. W. L. Aerts. 2015. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Sci. Rep.* 5: 11044.

33. Aerts, H. J. W. L., E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin. 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 5: 4006.

34. Haralick, R. M., K. Shanmugam, and I. Distein. 1973. Textural Features for Image Classification. *IEEE Trans. Syst. man, cybertetics* 3: 610–621.

35. Galloway, M. M. 1975. Texture analysis using gray level run lengths. *Comput. Graph. Image Process.* 4: 172–179.

36. Xiaoou Tang. 1998. Texture information in run-length matrices. *IEEE Trans. Image Process.* 7: 1602–1609.

37. Xu, D.-H., A. S. Kurani, J. D. Furst, and D. S. Raicu. 2004. Run-Length Encoding for Volumetric Texture. *Heart* 27: 452–458.

38. Chu, A., C. M. Sehgal, and J. F. Greenleaf. 1990. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognit. Lett.* 11: 415–419.

39. Amadasun, M., and R. King. 1989. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man. Cybern.* 19: 1264–1274.

40. Adcock, A., D. Rubin, and G. Carlsson. 2012. Classification of Hepatic Lesions using the Matching Metric. *Comput. Vis. image Underst.* 121: 36–42.

41. Adcock, A., E. Carlsson, and G. Carlsson. 2013. The Ring of Algebraic Functions on Persistence Bar Codes. *arXiv Prepr. arXiv1304.0530* .

42. Cohen-Steiner, D., H. Edelsbrunner, and J. Harer. 2007. Stability of Persistence Diagrams. *Discrete Comput. Geom.* 37: 103–120.

43. Reininghaus, J., S. Huber, U. Bauer, and R. Kwitt. 2015. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vol. 07–12–June. 4741–4748.

44. Gamble, J., and G. Heo. 2010. Exploring uses of persistent homology for statistical analysis of landmark-based shape data. *J. Multivar. Anal.* 101: 2184–2199.

45. Robins, V., M. Saadatfar, O. Delgado-Friedrichs, and A. P. Sheppard. 2016. Percolating length scales from topological persistence analysis of micro-CT images of porous materials. *Water Resour. Res.* 52: 315–329.

46. Hyekyoung Lee, H., H. Hyejin Kang, M. K. Chung, B.-N. Bung-Nyun Kim, and D. S. Dong

Soo Lee. 2012. Persistent Brain Network Homology From the Perspective of Dendrogram. *IEEE Trans. Med. Imaging* 31: 2267–2277.

47. Bendich, P., J. S. Marron, E. Miller, A. Pieloch, and S. Skwerer. 2016. Persistent homology analysis of brain artery trees. *Ann. Appl. Stat.* 10: 198–218.

48. Holland, D. O., B. H. Shapiro, P. Xue, and M. E. Johnson. 2017. Protein-protein binding selectivity and network topology constrain global and local properties of interface binding networks. *Sci. Rep.* 7: 5631.

49. Dunaeva, O., H. Edelsbrunner, A. Lukyanov, M. Machin, and D. Malkova. 2014. The Classification of Endoscopy Images with Persistent Homology. In *2014 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing* IEEE. 565–570.

50. Giansiracusa, N., R. Giansiracusa, and C. Moon. 2017. Persistent homology machine learning for fingerprint classification. *arXiv Prepr. arXiv1711.09158* .

51. Kumar, V., Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. W. L. Aerts, A. Dekker, D. Fenstermacher, D. B. Goldgof, L. O. Hall, P. Lambin, Y. Balagurunathan, R. A. Gatenby, and R. J. Gillies. 2012. Radiomics: the process and the challenges. *Magn. Reson. Imaging* 30: 1234–1248.

52. Zeiler, M. D., and R. Fergus. 2014. Visualizing and understanding convolutional networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 8689 LNCS. 818–833.

53. Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* 115: 211–252.

54. Dong, H., G. Yang, F. Liu, Y. Mo, and Y. Guo. 2017. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. In *Medical Image*

Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 506–517.

55. Yang, X., R. Kwitt, M. Styner, and M. Niethammer. 2017. Quicksilver: Fast predictive image registration – A deep learning approach. *Neuroimage* 158: 378–396.

56. Yang, Q., P. Yan, M. K. Kalra, and G. Wang. 2017. CT Image Denoising with Perceptive Deep Neural Networks. *arXiv Prepr. arXiv1702.07019* .

57. Chang, K., H. X. Bai, H. Zhou, C. Su, W. L. Bi, E. Agbodza, V. K. Kavouridis, J. T. Senders, A. Boaro, A. Beers, B. Zhang, A. Capellini, W. Liao, Q. Shen, X. Li, B. Xiao, J. Cryan, S. Ramkissoon, L. Ramkissoon, K. Ligon, P. Y. Wen, R. S. Bindra, J. Woo, O. Arnaout, E. R. Gerstner, P. J. Zhang, B. R. Rosen, L. Yang, R. Y. Huang, and J. Kalpathy-Cramer. 2018. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from mr imaging. *Clin. Cancer Res.* 24: 1073–1081.

58. He, K., X. Zhang, S. Ren, and J. Sun. 2015. Deep Residual Learning for Image Recognition. *Arxiv.Org* 7: 171–180.

59. Ioffe, S., and C. Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv Prepr. arXiv1502.03167* .

60. Torrey, L., and J. Shavlik. 2010. Transfer Learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* IGI Global. 242–264.

61. Yosinski, J., J. Clune, Y. Bengio, and H. Lipson. 2014. How transferable are features in deep neural networks? In *Advances in neural information processing systems* 3320–3328.

62. Guyon, I., J. Weston, S. Barnhill, and V. Vapnik. 2002. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* 46: 389–422.

63. Boots-Sprenger, S. H. E., A. Sijben, J. Rijntjes, B. B. J. Tops, A. J. Idema, A. L. Rivera, F. E. Bleeker, A. M. Gijtenbeek, K. Diefes, L. Heathcock, K. D. Aldape, J. W. M. Jeuken, and P.

Wesseling. 2013. Significance of complete 1p/19q co-deletion, IDH1 mutation and MGMT promoter methylation in gliomas: Use with caution. *Mod. Pathol.* 26: 922–929.

64. Patel, S. H., L. M. Poisson, D. J. Brat, Y. Zhou, L. Cooper, M. Snuderl, C. Thomas, A. M. Franceschi, B. Griffith, A. E. Flanders, J. G. Golfinos, A. S. Chi, and R. Jain. 2017. T2-FLAIR Mismatch, an Imaging Biomarker for IDH and 1p/19q Status in Lower-grade Gliomas: A TCGA/TCIA Project. *Clin. Cancer Res.* 23: 6078–6085.

65. Lasocki, A., F. Gaillard, A. Gorelik, and M. Gonzales. 2018. MRI Features Can Predict 1p/19q Status in Intracranial Gliomas. *Am. J. Neuroradiol.* 39: 687–692.

Vita

Donnie Kim received the degree of Bachelor of Science in Physics from Rice University in May, 2015. In August of 2015, he entered the University Of Texas MD Anderson Cancer Center UTHealth Graduate School Of Biomedical Sciences.

Permanent address:
8181 Fannin St Apt 2129
Houston, Texas 77054